IIUM ENGINEERING JOURNAL

Volume 26 Number 2

May 2025

IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

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

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

IIUM ENGINEERING JOURNAL

Volume 26, Issue 2, May 2025

https:/doi.org/10.31436/iiumej.v26i2

Copyright Notice

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

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

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

The IIUM Engineering Journal follows the open access policy.

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

IIUM ENGINEERING JOURNAL EDITORS

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 with in 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

Published by:

IIUM Press

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

ISSN 1511 - 788X

9 771511 788008

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

https://doi.org/10.31436/iiumej.v26i2.3726

EDITORIAL MESSAGE

The IIUM Engineering Journal Vol. 26 No. 2 continues its mission of advancing scholarly excellence by featuring 22 high-quality papers that span a wide spectrum of engineering disciplines. This issue brings together cutting-edge research showcasing the dynamic and multidisciplinary nature of engineering in addressing real-world problems.

Researchers explore sustainable and biologically driven innovations in the realm of Chemical and Biotechnology Engineering. Notably, identifying Rhizopus sp. fungi as an alternative lactic acid source and the ozonation of vegetable oils underscore efforts toward environmentally friendly biochemical processes. Civil and Environmental Engineering contributions highlight using natural materials for infrastructure, focusing on the feasibility of treated natural bitumen as a replacement for petroleum-based asphalt—an approach aligned with sustainable construction practices.

The most significant representation comes from Electrical, Computer, and Communications Engineering, where numerous works delve into AI, signal processing, secure communications, and sensor networks. From deep learning-based anomaly detection and vision transformer analysis for driver fatigue, to latency evaluations of LEO and GEO networks under tropical conditions, these studies reflect the growing importance of digital intelligence and communication resilience. Contributions on flexible antenna designs, mobile applications for dyslexic learners, and modified grey relational analysis further showcase how electronics and AI are increasingly embedded in daily societal functions.

In Materials and Manufacturing Engineering, the research on MSW-derived biochar for iron production and voltage-induced void formation in boiler tubes exemplifies innovation in resource utilization and reliability of industrial systems. Additionally, enhancing photocatalytic materials for water treatment presents promising strides in environmental engineering. Mechatronics and Automation Engineering advances are evident in developing regenerative braking systems for electric motorcycles and smart IoT-based solutions for e-bike sharing systems, both addressing urban mobility and energy efficiency. Finally, contributions in Engineering Mathematics and Applied Science provide theoretical underpinnings crucial for innovation, such as the study on the elliptic drum of vertical spindle cotton pickers, demonstrating the importance of mathematical modeling in optimizing mechanical design.

This issue affirms the journal's commitment to featuring impactful, interdisciplinary research that bridges theory and application. The editorial team extends sincere appreciation to all contributing authors, dedicated reviewers, and section editors whose efforts ensure the continued quality and relevance of the IIUM Engineering Journal.

We hope this collection will inspire further research, inform policy and industrial practices, and foster collaborations across academia and industry to pursue sustainable and inclusive technological development..

We extend our gratitude to the authors, reviewers, and editorial board for their unwavering commitment to excellence. Your contributions ensure this journal thrives as a beacon of scholarly inquiry and practical impact. Let us embrace these breakthroughs as stepping stones to a brighter, more sustainable future.

Prof. Ir. Ts. Dr. Teddy Surya Gunawan

Executive Editor

IIUM Engineering Journal

IIUM ENGINEERING JOURNAL

Volume 26, Issue 2, May 2025
https:/doi.org/10.31436/iiumej.v26i2

Table of Contents

Blank page.

https://doi.org/10.31436/iiumej.v26i2.3293

IDENTIFICATION OF THE RHIZOPUS SP. FUNGI AS AN ALTERNATIVE LACTIC ACID PRODUCTION SOURCE

AINIL HAWA JASNI*, AZLIN SUHAIDA AZMI, NOOR ILLI MOHAMAD PUAD, FATHILLAH ALI, YUSILAWATI AHMAD NOR

Department of Chemical Engineering and Sustainability, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak 53100 Kuala Lumpur, Malaysia

**Corresponding author: azlinsu76@iium.edu.my*

(Received: 10 May 2024; Accepted: 7 April 2024; Published online: 15 May 2025)

ABSTRACT: The search for eco-friendly alternatives to conventional petroleum-based materials has intensified in an era marked by a growing global awareness of environmental sustainability. This study addresses the critical need for molecular identification and characterization of fungi sourced from a tempeh commercial starter culture for their potential role in fungal-based polymer production. The problem is the limited knowledge and understanding of the genetic composition of these fungi and their suitability for lactic acid (LA) production, which is a crucial component of fungal-based polymer manufacturing. This research examined the tempeh starter culture fungi to identify suitable strains for LA production. The fungi were genotyped by DNA sequencing of the ITS region. The study revealed that the Ragi tempeh commercial starter culture contained only one strain of *Rhizopus* (*R. microsporus*), which was verified through ITS rRNA sequencing with 99.8% similarity to the GenBank database, simplifying control over fungal growth and potentially leading to consistent biomaterial yields. The method employed, involving DNA PCR (and sequencing of the ITS region, proved to be accurate, straightforward, and not excessively labor-intensive. The PCR conditions were as follows: initial denaturation at 98°C for 2 min, followed by 25 cycles of denaturation (98°C for 15 seconds), annealing (60°C for 30 seconds), and elongation (72°C for 30 seconds), with a final extension at 72°C for 10 min. Consequently, the consistent presence of only one *Rhizopus* species in commercial starter cultures of tempeh presents a promising avenue for sustainable biomaterial production, particularly in LA production. The pilot flask setup at 1×10^7 spores/mL was inoculated into 150 mL shake flasks with 1.2 g/mL glucose, incubated at 37°C for 1 to 7 days with 100 rpm shaking, yielding 1.037 g/g after 5 days, demonstrating the feasibility of using this strain for industrial applications.

ABSTRAK: Dalam era yang semakin menekankan kesedaran global terhadap kelestarian alam sekitar, pencarian alternatif mesra alam kepada bahan berasaskan petroleum konvensional semakin giat dijalankan. Kajian ini menangani keperluan kritikal untuk mengenal pasti dan mencirikan kulat secara molekul daripada kultur pemula komersial tempeh bagi potensi penggunaannya dalam penghasilan asid laktik (LA), komponen penting dalam pembuatan polimer berasaskan kulat. Masalah utama yang dibincangkan ialah kekurangan pengetahuan dan pemahaman mengenai komposisi genetik kulat ini serta kesesuaiannya untuk sintesis LA. Kajian ini menumpukan kepada pemeriksaan kulat dari kultur pemula tempeh bagi mengenal pasti strain yang sesuai untuk pengeluaran LA. Penjujukan DNA kawasan *internal transcribed spacer* (ITS) digunakan untuk mengenal pasti genotip kulat tersebut. Hasil kajian menunjukkan bahawa kultur pemula komersial Ragi tempeh hanya mengandungi satu strain sahaja, iaitu *Rhizopus microsporus* (disahkan melalui penjujukan ITS rRNA dengan 99.8% kesamaan dengan pangkalan data GenBank). Kehadiran satu spesies ini memudahkan kawalan pertumbuhan kulat dan berpotensi meningkatkan konsistensi hasil pengeluaran. Kaedah yang digunakan melibatkan PCR DNA dan penjujukan

kawasan ITS, yang terbukti tepat, mudah, serta tidak terlalu memerlukan tenaga kerja yang banyak. Keadaan PCR adalah seperti berikut: penyahdenaturan awal pada suhu 98°C selama 2 minit, diikuti 25 kitaran yang terdiri daripada penyahdenaturan (98°C, 15 saat), pengannealan (60°C, 30 saat), dan pemanjangan (72°C, 30 saat), dengan pemanjangan akhir pada 72°C selama 10 minit. Kehadiran spesies *Rhizopus* yang konsisten dalam kultur pemula tempeh komersial membuka peluang yang menjanjikan untuk pengeluaran asid laktik yang mampan. Penggunaan susunan flask perintis pada suhu 30°C menghasilkan 1.037 g/g selepas 5 hari, membuktikan potensi strain ini untuk aplikasi industri.

KEYWORDS: *DNA sequencing of the ITS region, DNA PCR sequencing, fungal polymer production, lactic acid production, molecular identification of fungi.*

1. INTRODUCTION

There has been growing interest in bio-based polymer production compared to traditional petroleum-based polymers. This shift is primarily driven by bio-based polymers' renewable and eco-friendly nature, particularly considering the declining petroleum reserves [1, 2]. Among the prominent bio-based polymers, polylactic acid (PLA), often called bioplastic, has garnered significant attention. Polylactic acid (PLA), commonly called bioplastic, has attracted considerable attention among these polymers. Lactic acid (LA), a key precursor of PLA, is produced through fermentation and exists as two enantiomers, D-LA and L-LA. While D-LAs are considered hazardous, L-LAs are deemed safe and widely used in the food and beverage industries [3, 4].

Both lactic acid bacteria (LAB) and fungi produce LA. Still, fungi are preferred for biomaterial production because of their adaptability to a wide range of renewable substrates, such as sugar and starch, compared with LAB [4]. Additionally, purification technology for fungal-derived LA is more mature and cost-effective. Fungi are abundant, easy to cultivate, and require minimal nutrients. Among the various fungal species capable of LA production, *Rhizopus sp*., including *R. oryzae*, *R. arrhizus*, and *R. microsporus*, have emerged as promising biomaterial sources [5].

Several studies have demonstrated the ability of *Rhizopus* spp. to produce LA. For instance, [6] demonstrated how well *R. oryzae* produces LA using a range of agro-industrial residues as substrates. Similarly, [7] observed that *R. arrhizus* produced a high yield and productivity of LA when the fermentation conditions were optimized. Using R. microsporus and corn stover as substrates [8] demonstrated high LA production. Furthermore, [9] highlighted the potential of food waste as a cost-effective substrate for LA production in *R. oryzae*. A study by [10] successfully demonstrated LA production using *R. arrhizus* with cassava pulp as the substrate (Table 1). In this context, high production refers to efficiently generating large amounts of LA. This is typically indicated by high yield (≥1.0 g LA/g substrate), fast production rate, and high final concentration in the fermentation broth.

Despite these advances, the potential of fungi from tempeh starter cultures to produce LA remains to be explored. Traditional identification methods have primarily focused on LAB, overlooking the efficiency and sustainability of the fungal species. DNA PCR and sequencing are the key molecular techniques used for genotyping and studying fungal sources [11-13]. This study aimed to bridge this gap by focusing on the molecular identification and characterization of fungi derived from tempeh starter culture. Specifically, the internal transcribed spacer (ITS) region, a widely used fungal DNA barcode, was analyzed to achieve precise species identification and to characterize inter- and intraspecific variations. The findings of this study could contribute to a more sustainable and efficient bio-based polymer

production. The LA generation route begins with glucose traveling via glycolysis, which breaks it down into pyruvic acid, while producing ATP and NADH. Lactate dehydrogenase (LDH) converts pyruvic acid to LA under anaerobic conditions by oxidizing NADH to NAD^+, a necessary step for glycolysis. Microorganisms such as *Lactobacillus* and *Rhizopus sp*. efficiently ferment sugars from various renewable substrates, facilitating this process [14]. The ability of fungi to exploit lignocellulosic biomass and other agro-industrial leftovers makes them a possible source for long-term LA production [15].

Table 1. Recent Similar Studies on Fungal-Based Lactic Acid Production

Fungal Species	Substrate Used	Yield (g/g)	Key Findings	Ref
R. oryzae	Agro-industrial residues	0.9	Efficient conversion of agro-industrial residues to LA	[6]
R. arrhizus	Optimized fermentation media	1.2	High yield and productivity under optimized conditions	[7]
R. microsporus	Corn stover	1.05	High LA production using lignocellulosic biomass	[8]
R. oryzae	Food waste	1.15	Potential of food waste as a cost-effective LA substrate	[9]
R. arrhizus	Cassava pulp	1.0	Efficient LA production from cassava pulp	[10]

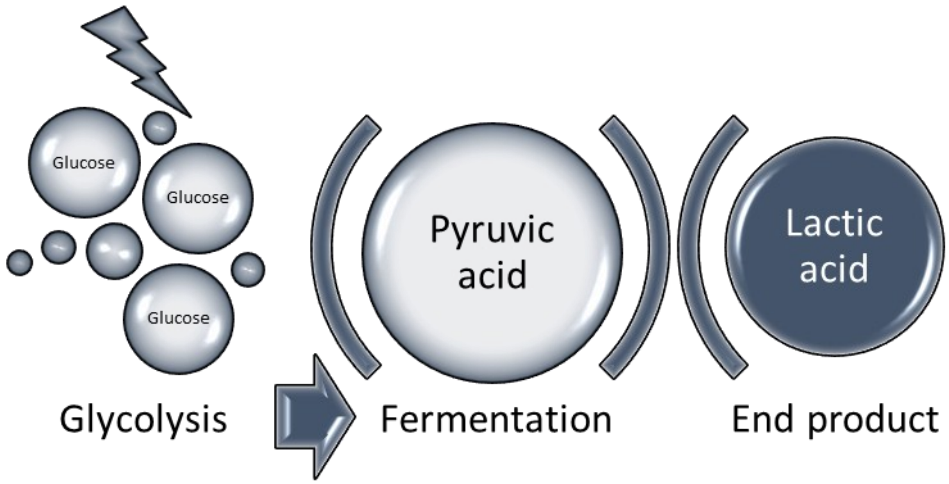


Figure 1. LA Production Pathway

2. MATERIALS AND METHODS

2.1. Yeast growth and genomic DNA extraction

The Ragi tempeh commercial starter culture, sourced from PT Aneka Fermentasi Industri in Indonesia and known to contain *Rhizopus sp.,* was cultured, isolated, and cultivated on potato dextrose agar (PDA) at an incubation temperature of 37ºC for a period of 7 days. Subsequently, *Rhizopus* growth on the agar was excised (0.5 cm x 0.5 cm) using a sterilized blade and transferred into 1.5 mL centrifuge tubes. This process was performed in triplicate to collect the yeast samples. Genomic DNA (gDNA) of the *Rhizopus* isolates was extracted using the Presto™ Mini gDNA Yeast Kit (Geneaid, GBYB100), followed by PCR amplification of the Internal Transcribed Spacer (ITS) gene using the universal primers ITS1 and ITS4.

2.2. Application of the ITS region

Amplification of the fungal ITS gene was accomplished using a pair of universal primers, ITS1 (5'-TCCGTAGGTGAACCTGCGG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3'). Each reaction mixture, totaling 25 µL in volume, consisted of purified gDNA, 0.5 pmol DNA template, deoxynucleotide triphosphates (dNTPs, 200 µM each), thermostable DNA polymerase (0.5 U), PCR buffer, and dH_2O. The PCR procedure included an initial denaturation cycle at 98°C for 2 min followed by 25 cycles of denaturation at 98°C for 15 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s. The final extension step was performed at 72°C for 10 min. The PCR products were visualized on a 1% TAE agarose gel at 100 V for 60 minutes [17]. The amplified PCR products were purified and subjected to bidirectional sequencing using M13F (-20) and M13R-pUC (-26) primers and the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing was performed by First BASE Laboratories (Apical Scientific Sdn. Bhd.).

2.3. BLAST and phylogenetic tree of the ITS region

The resulting ITS sequences (~700 bp) were analyzed and compared against the NCBI nucleotide database (nr/nt) using BLAST for fungal identification, accessible on the National Centre for Biotechnology Information (NCBI) website [18]. Neighbor-joining analysis was conducted to create an unrooted phylogenetic tree. This analysis was performed using the NCBI BLAST tree method available on the NCBI website [18].

2.4. Culture *Rhizopu*s in the fermentation medium

The fungi were cultured on potato dextrose agar (PDA) at 37°C for one week in Petri dishes with replicates. After sporulation and proliferation, 10 mL of distilled water was poured aseptically onto each agar plate, which was then scraped to release spores. The spore suspension was centrifuged for 10 min at 4,000 rpm, washed, and resuspended in 1 mL of distilled water. Then, a 1 mL spore suspension containing approximately 1 × 107 spores/mL was utilized to provide a constant quantity of spores for each experiment. The spores were inoculated and grown in each medium within a shake flask (150 mL) containing 1.2 g/mL glucose at 37°C, for 1–7 days, and at an agitation speed of 100 rpm in an incubator shaker. Fungal spore concentration was determined using a hemocytometer (Neubauer improved chamber) under a light microscope (40× magnification). After culturing the fungi on potato dextrose agar (PDA) at 37°C for one week, spores were harvested by adding 10 mL of sterile distilled water to the petri dish and scraping the surface. The spore suspension was centrifuged at 4,000 rpm for 10 min, washed, and resuspended in 1 mL sterile distilled water. A 10 µL aliquot of the well-mixed spore suspension was loaded onto the hemocytometer, and spores were counted in four large squares of the grid. The spore concentration was calculated using the following formula:

$$Spores\ concentration\ \left(\frac{Spores}{mL}\right) = \frac{Average\ spore\ count\ per\ square \times Dilution\ factor \times 10^4}{Volume\ of\ counted\ samples\ (mL)} \quad (1)$$

The final concentration was adjusted to 1×10^7 spores/mL by dilution with sterile distilled water or concentrating via centrifugation before inoculation into the fermentation medium.

2.5. LA Concentration Analysis

The LA concentration was determined using a colorimetric assay with iron (III) chloride ($FeCl_3$) at a wavelength of 540 nm using a DR5000™ Hach UV-VIS spectrophotometer. A stock LA solution with a concentration of 89 g/L was prepared by dissolving 1.2 g/mL of LA (89% purity) in distilled water to prepare the LA standard curve. Serial two-fold dilutions were

https://doi.org/10.31436/iiumej.v26i2.3293

performed to generate six LA solutions with concentrations ranging from 0 to 5 wt.%. A 0.2% $FeCl_3$ nanoparticle solution was prepared by dissolving 0.2 g of $FeCl_3$ nanoparticle in a 100 mL volumetric flask and diluting it with distilled water to the required volume. The solution was stored at 25°C (room temperature) and swirled until thoroughly mixed. For the absorbance measurements, 2 mL of 0.2% $FeCl_3$ solution was mixed with 50 µL of each LA standard solution. These solutions' optical density (O.D.)/absorbance ratio was measured at 540 nm using a UV-VIS spectrophotometer, and a standard calibration curve was constructed. The equation of the standard curve was determined as follows:

$$y = 0.9939x \tag{2}$$

where y represents the absorbance value (a.u.) and x is the LA concentration (wt.%). The coefficient of determination for the calibration curve was $R^2 = 0.99$, ensuring the accuracy of the concentration calculations. To determine the LA concentration in the fermentation samples, 2 mL of 0.2% $FeCl_3$ solution was mixed with 50 µL of the experimental sample, and the absorbance was measured at 540 nm. The LA concentration was calculated using Eq. (2) based on a plotted standard curve.

3. RESULTS AND DISCUSSION

3.1. Yeast ITS Region PCR Amplicon Result

DNA extraction was successfully performed. Assessment of DNA purity through spectrophotometric measurements within the 240–260 nm range indicated a ratio greater than 1.8, signifying good quality [19]. Moreover, amplification of the *Rhizopus* isolate ITS amplicon was successfully achieved, resulting in a single band of approximately 750 bp (Fig. 2). Notably, a positive control band appeared at approximately 700 bp. This variation is anticipated because the positive control employed a universal ITS, which may have slight differences in the length of its ITS region compared to that of *Rhizopus sp.*

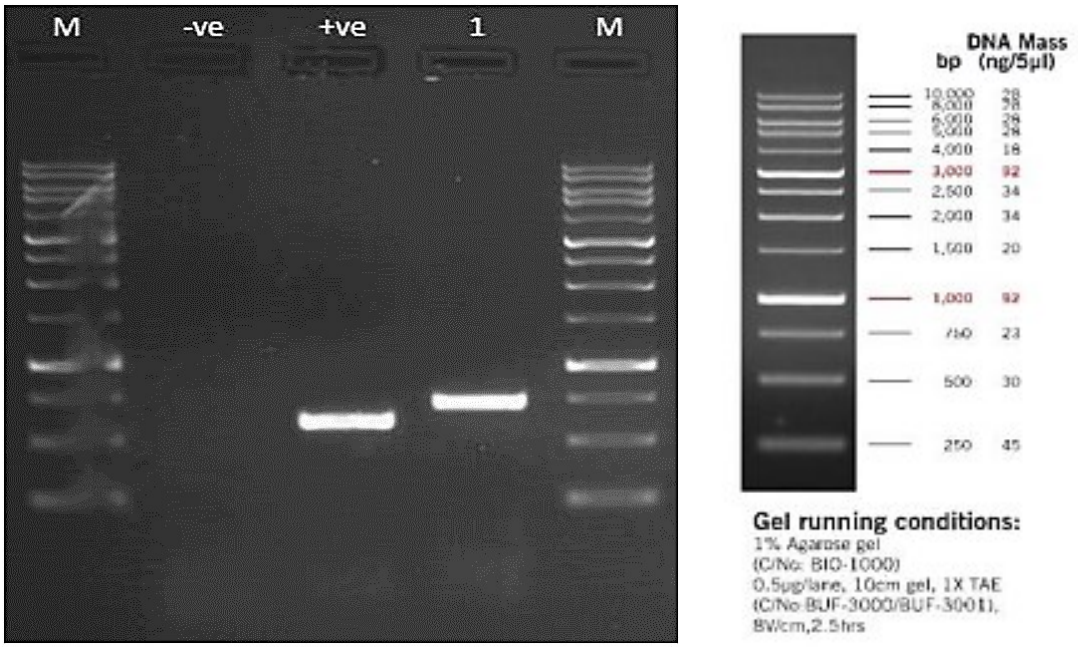


Figure 2. Gel electrophoresis image showing PCR amplicons. M, DNA ladder; −ve, no template control; +ve, positive control (purified plasmid with ITS region insert, 1 ng); 1: Rhizopus isolate from Ragi tempeh.

https://doi.org/10.31436/iiumej.v26i2.3293

3.2. DNA Sequencing and BLAST Analysis

Purified ITS amplicons from the *Rhizopus* isolate were successfully processed and sequenced. The sequencing outcomes obtained using the ITS region's forward and reverse primers yielded pure, high-quality, and easily readable DNA sequences (Fig. 3. Table 2 provides the complete ITS DNA sequence assembled from *Rhizopus* isolate sequencing. Subsequently, the ITS DNA sequence (Table 2) was compared to the GenBank database using BLASTN [18]. The phylogenetic tree results, illustrated in Figure 3, revealed a single cluster, all indicating the *Rhizopus microsporus* strain.

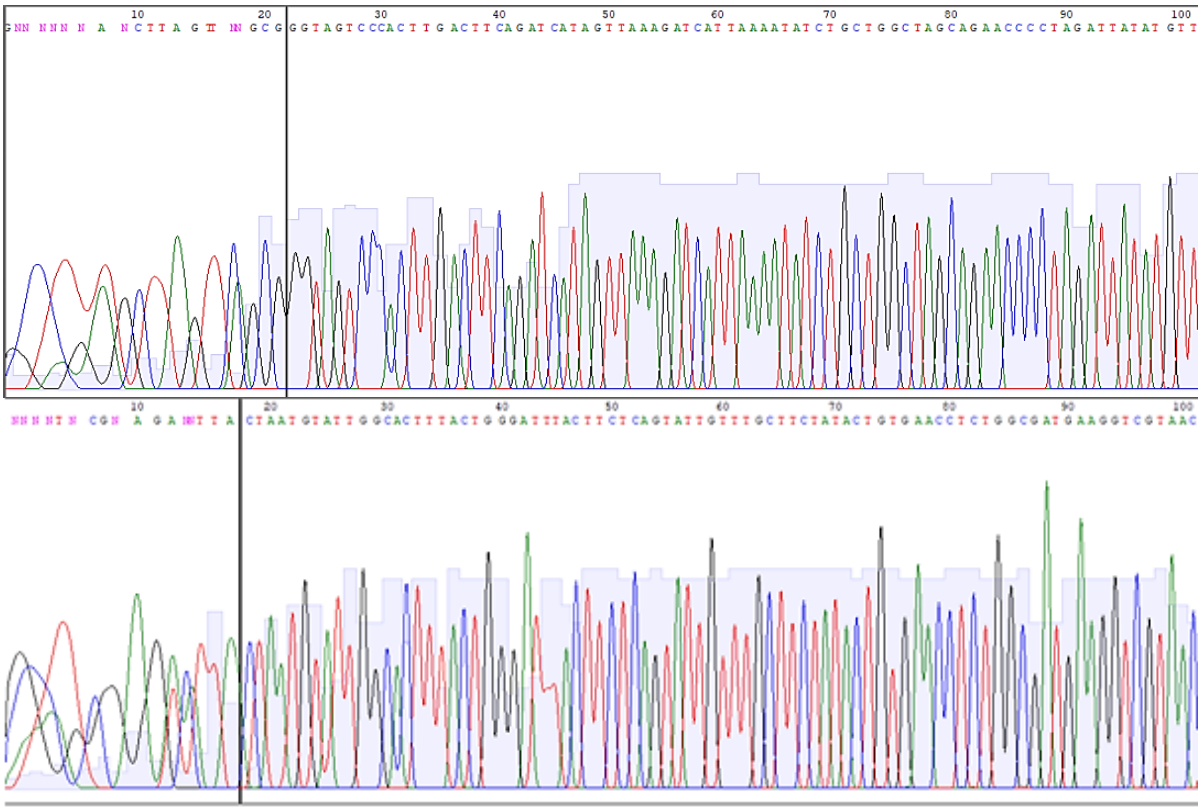


Figure 3**.** DNA sequence of the first 100 bp obtained from the Ragi tempeh *Rhizopus* isolate. Amplicons were obtained using an ITS reverse primer. The black line indicates the start of a good, readable sequence.

Table 3. ITS sequence identified from Ragi tempeh *Rhizopus* isolate.

Amplicon Sample	ITS Sequence Identify
Ragi tempeh *Rhizopus* isolate	CCAGTTCCGTAGGTGAACCTGCGGAAGGATCATTAACTAATGTATTGGCACTTTA CTGGGATTTACTTCTCAGTATTGTTTGCTTCTATACTGTGAACCTCTGGCGATGAA GGTCGTAACTGACCTTCGGGAGAGACTCAGGACATATAGGCTATAATGGGTAGG CCTGTTCTGGGGTTTGATCGATGCCAATCAGGATTACCTTTCTTCCTTTGGGAAGG AAGGTGCCTGGTACCCTTTACCATATACCATGAATTCAGAATTGAAAGTATAATA TAATAACAACTTTTAACAATGGATCTCTTGGTTCTCGCATCGATGAAGAACGTAG CAAAGTGCGATAACTAGTGTGAATTGCATATTCGTGAATCATCGAGTCTTTGAAC GCAGCTTGCACTCTATGGATCTTCTATAGAGTACGCTTGCTTCAGTATCATAACCA ACCCACACATAAAATTTATTTTATGTGGTGATGGACAAGCTCGGTTAAATTTAAT TATTATACCGATTGTCTAAAATACAGCCTCTTTGTAATTTTCATTAAATTACGAAC TACCTAGCCATCGTGCTTTTTTGGTCCAACCAAAAAACATATAATCTAGGGGTTCT GCTAGCCAGCAGATATTTTAATGATCTTTAACTATGATCTGAAGTCAAGTGGGAC TACCCGCTGAACTTAAGCATATCAATAAGCGGAGGAGTCAT

This study successfully identified a specific *Rhizopus* strain found in the Ragi tempeh commercial starter culture. According to the latest *Rhizopus* taxonomy, our findings strongly suggest the exclusive presence of a single *Rhizopus* species within the Ragi Tempeh commercial starter culture. The ITS rRNA sequence obtained from the Ragi Tempeh commercial starter culture was confirmed to belong to the species *R. microsporus.* The obtained ITS rRNA sequence from the Ragi tempeh commercial starter culture was analyzed using BLAST, revealing 99.8% similarity with *R. microsporus* strain. The highest identity matches confirmed a 627 bp sequence length, aligning with reference ITS sequences for these species. The substantial genetic similarity supports the isolate classification as *Rhizopus microsporus*, as it shares identical ITS rRNA sequences with known *R. microsporus* strains. Conserved ITS regions further validate this identification, confirming its role as a dominant species in tempeh fermentation.

This study also highlighted the effectiveness of ITS sequences as DNA barcodes for identifying fungal species. Identifying a singular *Rhizopus sp.* species was well-founded, considering that tempeh commercial starter cultures have long been recognized to contain only one *Rhizopus sp*. strain. The reduction in *Rhizopus* diversity in tempeh, linked to the widespread use of commercial tempeh starters in Indonesia since the early 2000s, is evident [20]. This phenomenon was clearly illustrated in the phylogenetic analysis of the ITS sequence results obtained from *Rhizopus* isolated from a commercial starter culture (see Figure 4). ITS sequences, which are shared by numerous fungal kingdoms, including *Rhizopus sp.,* have been extensively documented and continue to be updated [13].

Adopting DNA PCR and sequencing of the ITS region proved accurate and user-friendly, significantly reducing the labor involved compared to alternative methods, such as microscopic morphological characterization or random amplified polymorphic DNA analysis [21]. This advantage holds particular value at the industrial level, especially during quality control assessments of *Rhizopus* growth.

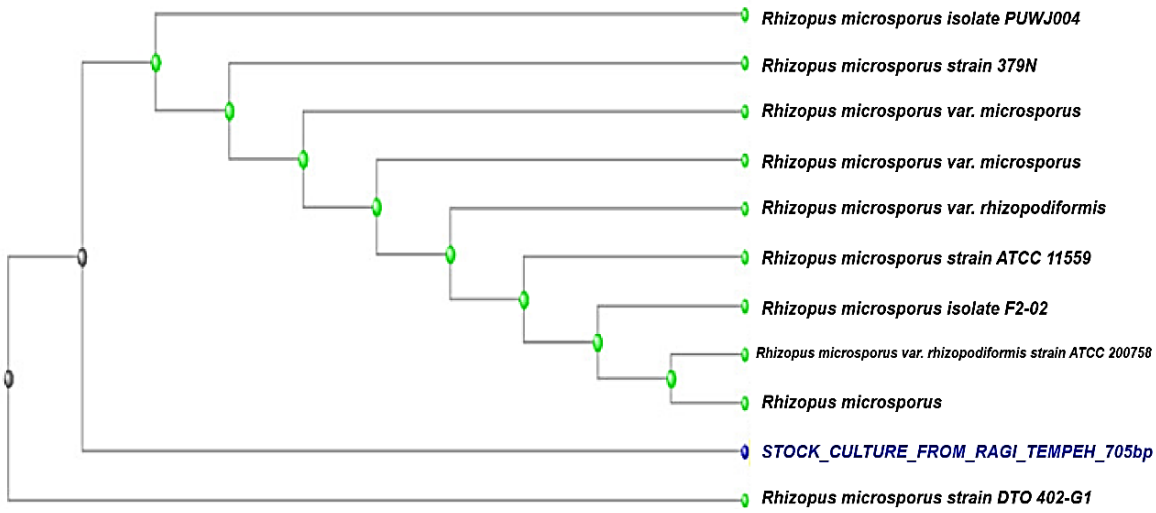


Figure 4. The BLAST results of the query sequence to the most similar biological sequences are available in the NCBI database.

The outcomes of the molecular identification and the resulting phylogenetic tree analysis, based on the ITS region of the isolated *Rhizopus* strain from the Ragi tempeh, confirm that the prolonged use of tempeh commercial starter cultures has led to the decline of traditional *Rhizopus* strains used in tempeh production [22] Traditionally, the inoculum for tempeh fermentation, known as "Usar," is typically derived from the leaves of *Hibiscus tiliaceaus* [23].

https://doi.org/10.31436/iiumej.v26i2.3293

It has been documented that when Usar is the source of the initial tempeh inoculum, multiple strains of various *Rhizopus* species, such as *R. delamar, R. oryzae*, and *R. arrhizus*, have been identified [22]. The reduction in *Rhizopus* species diversity within tempeh could potentially affect its status as a healthy food product, although it may find applications in other industries**.**

Nevertheless, a single *Rhizopus sp.* species within the Ragi Tempeh commercial starter culture could offer advantages for LA production. This is because various *Rhizopus* species have been reported to exhibit distinct nutritional preferences for their growth [22]. Consequently, when fungi with complex nutritional requirements are used, it becomes challenging to streamline LA production because multiple factors, such as nitrogen substrate, temperature, and pH, need to be meticulously controlled. Furthermore, LA production efficiency is highly susceptible to fluctuations when different fungi exhibit varying growth rates [1]. Notably, interest in this field declined from 2017 onwards [23]. In contrast, only one strain of *Rhizopus* (*R. microsporus*) within the Ragi Tempeh commercial starter culture simplifies the control of fungal growth, potentially leading to consistent biomaterial yields. Fig. 5 shows the species image under ten times magnification of the light contrast microscope.

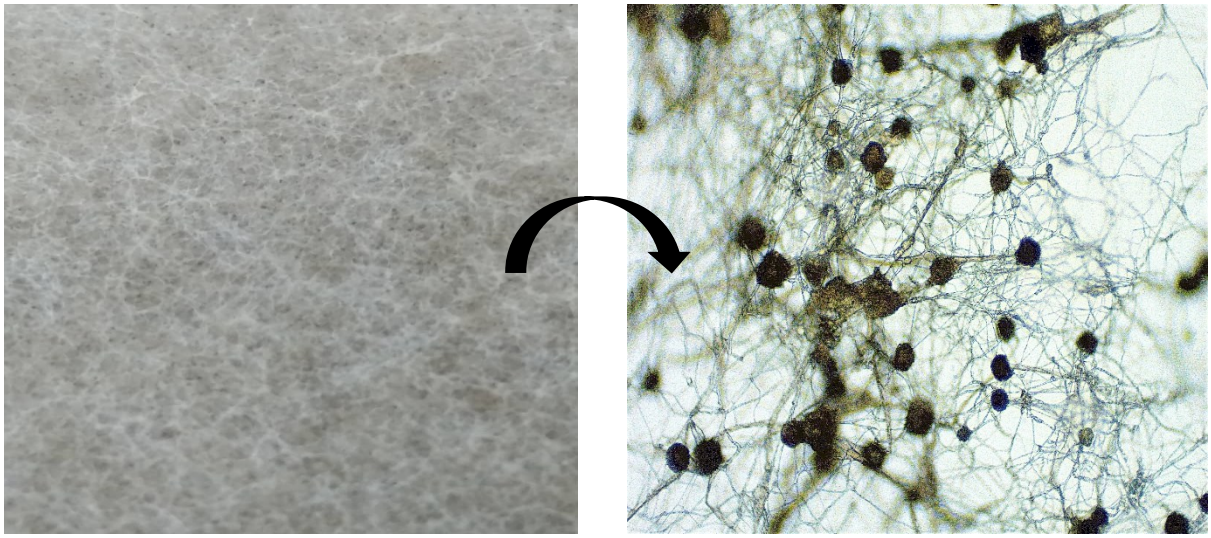


Figure 5**.** *R. microsporus* fungi under the naked eye on PDA. The image on the right is the zoom view of the species under 10x magnification of the light microscope.

3.3. LA Yield from the identified strain

To prove the effect of fungal strain cell metabolism on the production of LA, fermentation reactions using glucose yielded LA. Thus, the experimental yield was 1.037 g/g after 5 days in the pilot flask size setup at 30 °C, as shown in Figure 6. The maximum LA yield showed significantly higher results than a study done by [24], which recorded 0.95 g/g LA yield, respectively.

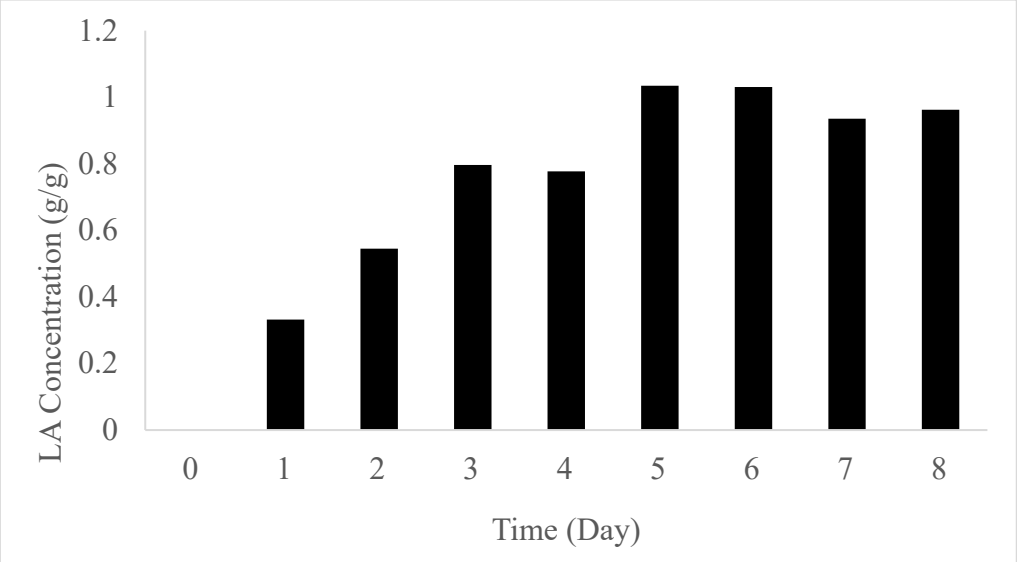


Figure 6. LA concentration produced by *Rhizopus microsporus.* over 7 days of incubation time

4. CONCLUSIONS

Molecular analysis confirmed that the *Rhizopus* strain isolated from the Ragi Tempeh commercial starter culture falls under the classification of *R. microsporus.* This molecular identification method, which involves DNA PCR and ITS region sequencing, has proven precise, straightforward, and efficient. Moreover, the consistent presence of a single *Rhizopus* species within tempeh commercial starter cultures offers a promising opportunity for venturing into biomaterial production. Specifically, it presents a potential avenue for LA production using the tempeh commercial starter culture, as the LA yield was 1.037 g/g.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education (MOE) Malaysia through Fundamental Research Grant Scheme (FRGS/1/2021/TK0/UIAM/02/19). All authors thank Apical Scientific Sdn. Bhd., which is responsible for the fungal DNA sequencing service.

REFERENCES

[1] Andrew JJ, Dhakal HN. (2022) Sustainable biobased Composites for Advanced Applications: Recent Trends and Future Opportunities: A Critical Review. *Composites Part C: Open Access*, 7:100220. https://doi.org/10.1016/j.jcomc.2021.100220

[2] Asadollahzadeh M, Mahboubi A, Taherzadeh MJ, Åkesson D, et al. (2022) Application of fungal biomass for the development of new polylactic acid-based biocomposites. Polymers, 14(9):1738. https://doi.org/10.3390/polym14091738

[3] Huang LP. (2005) Simultaneous saccharification and fermentation of potato starch wastewater to lactic acid by Rhizopus oryzae and Rhizopus arrhizus. Biochemical Engineering Journal, 23:265–276. https://doi.org/10.1016/j.bej.2005.01.009

[4] Yuwa-amompitak T, Chookietwattana K. (2014) L-lactic acid production from cassava starch by thermotolerant Rhizopus microsporus LTH23. Journal of Biological Sciences, 14(4):284–291. https://doi.org/10.3923/jbs.2014.284.291

[5] Kitpreechavanich V, Maneeboon T, Kayano Y, Sakai K. (2008) Comparative characterization of L-lactic acid-producing thermotolerant Rhizopus fungi. Journal of Bioscience and Bioengineering, 106(6):541–546. https://doi.org/10.1263/jbb.106.541

[6] Ge X, Luo S, Tan S, Li Y, Li D, Chen Z, Zhang K. (2015) Efficient production of L-lactic acid from agro-industrial residues by Rhizopus oryzae. Bioresource Technology, 190:378–385. https://doi.org/10.1016/j.biortech.2015.04.004

[7] Zhang Y, Zhang S, Zhang Y, Li Y. (2016) High-yield production of L-lactic acid from corn stover hydrolysate by Rhizopus arrhizus. Bioresource Technology, 207:269–275. https://doi.org/10.1016/j.biortech.2016.02.099

[8] Li X, Liu D. (2018) Efficient production of L-lactic acid by Rhizopus microsporus using corn stover hydrolysate as the substrate. Biotechnology Letters, 40(1):135–141. https://doi.org/10.1007/s10529-017-2445-0

[9] Chen Y, Xiao Y. (2019) Utilization of food waste for L-lactic acid production by Rhizopus oryzae. Food Science and Biotechnology, 28(2):577–585. https://doi.org/10.1007/s10068-018-0518-y

[10] Wang L, Zhang H, Liu J, Zhang J. (2020) Production of L-lactic acid from cassava pulp hydrolysate by Rhizopus arrhizus. Preparative Biochemistry & Biotechnology, 50(2):181–187. https://doi.org/10.1080/10826068.2019.1697767

[11] Iwen PC, Hinrichs SH, Rupp ME. (2002) Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. Medical Mycology, 40(1):87–109. https://doi.org/10.1080/mmy.40.1.87.109

[12] Lott T, Burns B, Zancope-Oliveira R. (1998) Sequence analysis of the internal transcribed spacer 2 (ITS2) from yeast species within the genus Candida. Current Microbiology, 36:63–69. https://doi.org/10.1007/s002849900280

[13] Schoch CL, Seifert KA, Huhndorf S, Robert V, et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proceedings of the National Academy of Sciences, 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

[14] Giuffrè D, Giuffrè AM. (2024) Fermentation technology and functional foods. Frontiers in Bioscience (Elite Edition), 16(1):8.

[15] Yin F, Sun X, Zheng W, Yin L, Luo X, Zhang Y, Wang Y, Fu Y. (2022) Development of a strategy for L-lactic acid production by Rhizopus oryzae using Zizania latifolia waste and cane molasses as carbon sources. Molecules, 28(17):6234.

[16] Geneaid. (2023) PrestoTM Mini gDNA Yeast Kit Instruction Manual. https://www.geneaid.com/data/files/1605664221308055331

[17] Fisher T. (2016) BigDye Terminator v3.1 Cycle Sequencing Kit Quick Reference. https://assets.thermofisher.com/TFSAssets/LSG/manuals/MAN0015666_BigDyeTerminator_V3_1_CycleSeq_QR.pdf

[18] NCBI National Center for Biotechnology Information. (1988) https://www.ncbi.nlm.nih.gov/

[19] Glasel JA. (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. BioTechniques, 18(1):62–63.

[20] Sjamsuridzal W, Khasanah M, Febriani R, Vebliza Y, Oetari A, et al. (2021) The effect of the use of commercial tempeh starter on the diversity of Rhizopus tempeh in Indonesia. Scientific Reports, 11(1):23932. https://doi.org/10.1038/s41598-021-03308-6.

[21] Barus TS. (2020) Genotypic characterization of Rhizopus species from tempeh and usar: Traditional inoculum of tempeh in Indonesia. Microbiology Indonesia, 14(3):101–107. https://doi.org/10.5454/mi.14.3.3

[22] Kaerger K, Schwartze VU, Dolatabadi S, Nyilasi I, et al. (2015) Adaptation to thermotolerance in Rhizopus coincides with virulence as revealed by avian and invertebrate infection models, phylogeny, physiological and metabolic flexibility. Virulence, 6(4):395–403. https://doi.org/10.1080/21505594.2015.1029219

[23] Syakirah MR, Azlin SA, Ali F, Mohamad Puad NI. (2021) Mapping literature review on lactic acid production by Rhizopus. Chemical and Natural Resources Engineering Journal, 5(2):1–27.

[24] Azmi AS, Yusuf N, Jimat DN, Mohamad Puad NI. (2016) Co-production of lactic acid and ethanol using Rhizopus sp. from hydrolyzed inedible cassava starch and leaves. IIUM Engineering Journal, 17(2):1–10. https://doi.org/10.31436/iiumej.v17i2.610

https://doi.org/10.31436/iiumej.v26i2.3410

OZONATION OF VEGETABLE OILS AND STUDY ON THEIR PHYSICOCHEMICAL AND BIOLOGICAL CHARACTERISTICS

NUR AMIRA ROSDI, YUSILAWATI AHMAD NOR *, DZUN NORAINI JIMAT, AZLIN SUHAIDA AZMI, BASSAM MOHAMED JAWAHAR

Department of Chemical Engineering and Sustainability, Kulliyyah of Engineering, International Islamic University Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia

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

(Received: 20 August 2024; Accepted: 21 March 2025; Published online: 15 May 2025)

ABSTRACT: Free ozone offers significant benefits in biological applications due to its efficacy as a disinfectant, but toxicity and instability are associated with it. Hence, producing ozonated vegetable oil (OVO) has been explored as a potential solution, yielding stable ozonation by-products with medical potential, such as antimicrobial activity. Several studies have explored OVO's characteristics and biological effects, including olive oil, sunflower oil, and canola oil. However, optimizing ozonation conditions is still lacking, with many other types of vegetable oils yet to be studied. This research comprises three phases: i) ozonation of selected oils: red palm oil (RPO), rice bran oil (RBO), peanut oil (PO), and virgin coconut oil (VCO), ii) screening for the most effective OVO against three bacteria (*Staphylococcus aureus, Bacillus subtilis,* and *Escherichia coli*), and iii) physicochemical testing. Results show increased peroxide and acidity values in most OVO and a decrease in iodine value compared to untreated oil. Ozonated virgin coconut oil (OVCO) exhibits the highest antibacterial activity by showing a zone of inhibition of 11.3 mm and 84.35% killing rate at 30 minutes incubation time, particularly against *S. aureus*. Further optimization using Design Expert®6.0.8 software identifies the most effective ozonation conditions for OVCO, achieving a peak killing rate of 100% against *S. aureus* with 360 mins of ozone exposure and ozone flow rates of 1 l/min. Kinetic studies confirm rapid bacterial eradication, with over 90% of *S. aureus* killed by OVCO within 2 mins. Moreover, OVCO proved to be non-toxic to human foreskin fibroblast (HFF1) cells, maintaining 80% viability even after exposure to 1 mg/ml OVCO treated with ozone for 120 and 240 mins. These findings underscore the promising medical potential of OVCO, particularly in treating skin diseases.

ABSTRAK: Ozon bebas menawarkan manfaat signifikan dalam aplikasi biologi disebabkan keberkesanannya sebagai bahan pembasmi kuman, namun ia turut dikaitkan dengan ketoksikan dan ketidakstabilan. Oleh itu, penghasilan minyak sayuran berozon (OVO) telah diteroka sebagai potensi penyelesaian, menghasilkan hasil sampingan ozonasi yang stabil dengan potensi perubatan seperti aktiviti antimikrob. Beberapa kajian telah meneliti ciri-ciri dan kesan biologi OVO termasuk minyak zaitun, minyak bunga matahari, dan minyak kanola. Namun, proses pengoptimuman keadaan ozonasi masih belum lengkap dan banyak lagi jenis minyak sayuran belum dikaji. Kajian ini terdiri daripada tiga fasa: i) ozonasi minyak terpilih iaitu minyak sawit merah (RPO), minyak dedak padi (RBO), minyak kacang tanah (PO), dan minyak kelapa dara (VCO), ii) saringan keberkesanan OVO terhadap tiga jenis bakteria (*Staphilokokus aureus*, *Basillus subtilis*, dan *Escherichia coli*), dan iii) ujian fisikokimia. Keputusan menunjukkan peningkatan nilai peroksida dan keasidan dalam kebanyakan OVO serta penurunan nilai iodin berbanding minyak yang tidak dirawat. Minyak kelapa dara berozon (OVCO) menunjukkan aktiviti antibakteria tertinggi dengan zon perencatan berdiameter 11.3 mm dan kadar pembunuhan bakteria sebanyak 84.35% dalam masa inkubasi

30 minit, khususnya terhadap *S. aureus*. Pengoptimuman lanjut menggunakan perisian Design Expert®6.0.8 mengenal pasti keadaan ozonasi paling berkesan bagi OVCO, dengan pencapaian kadar pembunuhan maksimum 100% terhadap *S. aureus* pada pendedahan ozon selama 360 minit dan kadar aliran ozon 1 l/min. Kajian kinetik mengesahkan penghapusan bakteria yang pantas, dengan lebih 90% *S. aureus* dibunuh oleh OVCO dalam masa 2 minit. Tambahan, OVCO terbukti tidak toksik terhadap sel fibroblas kulit manusia (HFF1), dengan mengekalkan 80% daya hidup walaupun selepas pendedahan kepada 1 mg/ml OVCO yang dirawat ozon selama 120 dan 240 minit. Penemuan ini menekankan potensi perubatan OVCO, khususnya dalam merawat penyakit kulit.

KEYWORDS: *Ozonated vegetable oil, Cytotoxicity, Antibacterial, Killing rate*

1. INTRODUCTION

Ozone is acknowledged for its powerful disinfectant properties in therapeutic and biological applications, but its free form is harmful to living cells and the lungs when inhaled [1]. To overcome this limitation, ozonated vegetable oil (OVO) has emerged as an alternative, containing therapeutic by-products resulting from the alteration of unsaturated fatty acids through ozonation to produce active oxygen and other active substances such as ozonides, aldehydes, and peroxides. Moreover, unlike gaseous ozone, OVO offers enhanced stability and efficacy [2]. OVO contains a series of ozonation by-products that remain active for years when properly stored [3]. OVO possesses diverse biological properties, functioning as bactericidal and antifungal agents, and facilitating wound healing. Thus, it is a potential substitute for traditional antibiotics and antifungals, particularly in dermatology, dental care, ophthalmology, and gynaecology.

Vegetable oils are rich in saturated fatty acids, unsaturated fatty acids, antioxidants, fat-soluble vitamins, and various minor components. These oils have low allergenic potential, are non-toxic, and are utilized in fuel, food, and functional products due to their natural therapeutic properties [4]. Considering the benefits associated with vegetable oils, numerous studies have been undertaken to explore the attributes and potential biological activities of diverse ozonated vegetable oils, encompassing olive oil, grapeseed oil [2], sunflower oil [5], and soybean oil [2]. Studies have demonstrated that ozonation significantly enhances the antibacterial properties of vegetable oils [6]. These studies also highlight variations in physicochemical properties, such as density, viscosity, and ozonation by-products, which arise due to differences in oil composition, including the levels of saturated and unsaturated fatty acids and the degree of ozonation applied [7]. However, despite these findings, a comprehensive comparison of various types of OVO under standardized ozonation conditions has not yet been conducted. This gap limits the ability to determine which vegetable oils yield the most effective antibacterial properties under similar conditions. With many vegetable oils having unique compositions and origins, it is crucial to study their ozonated properties systematically. This would facilitate the identification of the most potent OVO with antibacterial properties, providing valuable insights for therapeutic and industrial applications. Additionally, while the benefits of ozonation are well-documented, there has been limited focus on optimizing ozonation parameters to maximize antibacterial activity. Another key gap in current research is the limited investigation of potential toxicity linked to prolonged ozonation treatment times. Addressing these gaps is essential to ensure the practical and safe application of OVOs in the medical and pharmaceutical fields.

Thus, this study aims to address these gaps by investigating the physicochemical parameters (peroxide values, acidity values, iodine values, and functional groups) and

antibacterial properties of several other types of OVO including red palm oil (RPO), rice bran oil (RBO), peanut oil (PO), and virgin coconut oil (VCO). OVO with the best antibacterial activity against several different bacterial strains will be screened first before determining the optimal ozonation conditions to achieve optimal antibacterial activity. The time-dependent effect of the optimized OVO against the bacteria, as well as the toxicity effect of the OVO against normal skin cells, was also investigated in this study. The comprehensive analysis obtained from this study will contribute to understanding OVO's properties and investigate its reasonable toxicity towards normal skin cell lines to support its utilization as a sustainable and effective alternative for combating bacterial infections.

2. MATERIALS AND METHODOLOGY

2.1. Chemical and Materials

Vegetable oils (RPO, RBO, PO, and VCO) were obtained from a local supermarket, and a NANO ozone generator from Absolute Ozone for ozonation was used in the ozonation process. To analyze the OVO, potassium hydroxide (0.1 M), sodium thiosulfate ($Na_2S_2O_3$), sodium hydroxide (NaOH), Wijs solution, phenolphthalein, chloroform ($CHCl_3$), and deionized water were obtained from Chemiz and Sigma. The bacteria used in the study are *E. coli*, *B. subtilis,* and *S. aureus* obtained from INHART, IIUM.

2.2. Ozonation of Vegetable Oils for Screening Process

The experiment started with 20 ml of each vegetable oil being ozonated in a Dresher bottle (Quickfit®) connected to an ozone generator. Ozone is produced by the silent discharge of oxygen (purity of 99.9%) from the ozone generator with a 30 mg/l ozone concentration. The oils were exposed to the ozone at room temperature for 120 minutes with a 1.5 l/min ozone flow rate.

2.3. Characterization of OVO

2.3.1. Density Value and Color Observation

Density represents the ratio between the mass of a substance and its volume at a determined temperature, expressed in grams per milliliter (g/ml). The calculation involved determining the ratio of the mass of a given volume of the ozonated vegetable oil at 20°C to the mass of an equivalent volume of water at 4°C. All measurements were performed in triplicate [6]. Color observation was conducted using visual inspection. Equal amounts of untreated vegetable oil and OVO were placed in separate beakers. The samples were examined under consistent lighting conditions, and any color differences were observed.

2.3.2. Peroxide Value

The peroxide value (PV) is a number that quantifies the number of peroxides present in 1000 g of a substance in 1 ml equivalents of active oxygen. In a 250 ml conical flask, 5 g of the OVO was weighed and mixed with 30 ml of acetic acid/chloroform [18 ml Acetic Acid: 12 ml chloroform]. Then, 0.5 ml of saturated potassium iodide solution was added once the material was completely dissolved. The solution was stirred and left for 1 min. Subsequently, 30 ml of distilled water was added to the solution, followed by titration with sodium thiosulfate solution (0.05 M) and 1 ml of starch solution as an indicator. The steps were repeated without oil for the blank sample. Each sample was analyzed in triplicate. Subsequently, the peroxide value was calculated using Eq. (1).

https://doi.org/10.31436/iiumej.v26i2.3410

$$PV = 1000\ (V_1 - V_0).\frac{c}{w} \quad (1)$$

where PV is the peroxide value, V_1 is the volume of the thiosulfate solution used for the titration, V_0 is the volume of the thiosulphate solution used to carry out a blank sample, c is the sodium thiosulfate normality, and w is the mass of the oil sample (g).

2.3.3. Acidity Value

Acidity value (AV) is the amount of sodium hydroxide needed to neutralize the free acids in 1.0 grams of the substance. It acts as an indicator of the oil's oxidation, reflecting how much the triglycerides of the OVO have broken down to produce free fatty acids. The AV of the oil samples was determined using ethanol titration with phenolphthalein as an indicator.

The experiment began by dissolving 1 g of each oil in 15 mL of ethanol in a measuring flask. This resulted in a solution titrated with sodium hydroxide (NaOH), using phenolphthalein as an indicator [6]. Three replications were made for each of the samples. Eq. (2) was used to calculate the acid value.

$$AV = \frac{McV}{w} \quad (2)$$

where AV is the acid value (mg NaOH/g fat), M is the molar mass of NaOH (g/mol), c is the concentration of NaOH (M), V is the volume of NaOH utilized (ml), and w is the mass of the oil sample (g).

2.3.4. Iodine Value

The iodine values (IV) can be used to determine the presence of double bonds in vegetable oils. It measures the amount of iodine that can react with the oils' carbon-carbon double bonds, indicating their unsaturation rate. When the ozonation process is successful, the iodine value of the vegetable oil decreases.

The experiment started by dissolving 2.5 g of the OVO sample in 12.5 ml of carbon tetrachloride (CCl4), then 12.5 ml of Wijs solution was added to the mixture along with 1 ml of potassium iodide, and the mixture was kept in the dark for approximately 1 h to complete the reaction. Subsequently, 50 ml of deionized water was added and was titrated with 0.05 M sodium thiosulfate ($Na_2S_2O_3$) until reaching the endpoint (turning into a milky white color solution). The same procedure above was repeated for a blank sample without the OVO. Each sample was analyzed in triplicate. Subsequently, the peroxide value was calculated using Eq. (3).

$$IV = \frac{12.69(V_1 - V_0)N}{w} \quad (3)$$

where IV is the iodine value, V_1 is the volume of the thiosulfate solution used for the titration, V_0 is the volume of the thiosulphate solution used to carry out a blank sample, N is the sodium thiosulfate normality, and w is the mass of the oil sample (g).

2.4. Screening of OVO for Antibacterial Activity

2.4.1. Seed Culture Antibacterial Testing

A single colony of bacteria was taken from the culture plate using an inoculum loop and dipped into 10 mL sterile LB broth in a 50 mL centrifuge tube. The cultures were incubated overnight at 37°C with 150 rpm in the incubator shaker. After overnight incubation, the cultures were collected and centrifuged at 3500 rpm for 5 minutes. The supernatant was then discarded,

and the pellet was dissolved in PBS. The optical density (OD) of the solution was adjusted to a specific value of 0.1 using a spectrophotometer.

2.4.2. Antibacterial Testing

The study used two methods to investigate the antibacterial properties of the untreated vegetable oil and OVO: the agar diffusion method and the plating method. The agar diffusion method is a common technique to study the antibacterial properties of OVO. The procedure described involves mixing 0.5 ml of DMSO with 4 ml of oil in a 15 ml centrifuge tube, incubating the mixture for 30 minutes at 37°C, and then spreading 0.1 ml of the seed culture sample on LB agar. Discs are impregnated with 0.04 ml of the incubated solutions and dropped on the agar plate containing the seed culture. Each sample was analyzed in triplicate. The agar plates were incubated overnight at 37°C, and the inhibition diameter was calculated to determine the antibacterial activity of the OVO.

For the plating method, 0.5 ml DMSO with 4 ml oil and 0.5 ml seed culture is added to a 15 ml centrifuge tube. The samples were vortexed and incubated for 30 minutes at 37°C. Afterward, 0.1 ml of the sample was plated onto LB agar and continued incubating overnight at 37°C. Three replications were made for each of the samples. Then, the number of colonies was calculated using Eq. (4) to obtain the percentage of the killing rate as follows:

$$Killing\ Rate\ (\%) = \frac{(control\ colony\ number - tested\ colony\ number)}{control\ colony\ number} \times 100 \quad (4)$$

2.5. Optimization of the ozonation condition for antibacterial activity

The optimization study used Design Expert® 6.0.8 software, employing the Response Surface Method (RSM) with a Face-Centered Central Composite Design (FCCD). The best OVO found after the screening process on the efficacy of antibacterial properties using the plating method was chosen for the optimization study. In this experiment, two independent variables were selected: ozone exposure time (120 mins to 360 mins) and ozone flow rate (1 l/min to 3 l/min). The range was selected based on a previous study [2][3].

2.6. Kinetic Study

A kinetic study on the antibacterial activity was conducted using the best OVO from the optimization study at 120 mins and 360 mins ozone exposure time with a flow rate of 1 l/min. The procedure was the same as the plating method. The cultures were incubated separately at various incubation periods before being plated onto LB agar and incubated overnight at 37°C. Then, the number of colonies was calculated to obtain the percentage of the killing rate using Eq. (4). All samples were performed in triplicate. Graphs of the percentage killing rate of the OVO against time were plotted to obtain the kinetic pattern of antibacterial activity over time.

2.7. Cytotoxicity Test

Cell viability of HFF1 (normal fibroblast skin cells) upon exposure to the OVO was assessed by performing the MTT assay. Cells were cultivated in 96-well plates at 50,000 cells per well. Subsequently, the cells were treated with variable concentrations of OVO produced at different ozonation times. This treatment spanned a period of 48 h within a 5% CO_2 incubator, maintaining a temperature of 37°C. After the treatment period, the cells were rinsed and exposed to 0.01 ml of MTT solution, following which the plates were incubated at 37°C in a dark environment for 4 h. Once formazan crystals had formed, 0.1 ml of DMSO was added, and the absorbance was quantified at 570 nm utilizing a microtiter plate reader [7].

https://doi.org/10.31436/iiumej.v26i2.3410

3. RESULTS AND DISCUSSIONS

3.1. Physical Characterization of Ozonated Vegetable Oils Illustrations

Physical characterization studies the changes in density and color of vegetable oils before and after ozonation. Table 1 shows that all vegetable oils showed a slight increment in their density after the ozonation process. ORPO has the highest density at 0.962 g/ml, reflecting an 8.09% increase, followed by OVCO at 0.945 g/ml, with a 3.85% increase. OPO has a density of 0.935 g/ml, showing a 1.63% increase, while RBO has the lowest increase of 1.13%, with a density of 0.895 g/ml. A higher percentage of density increment for ORPO is primarily due to its high levels of unsaturated fatty acids (50%), such as oleic and linoleic acids, as well as reactive components like carotenoids and tocopherols. These substances undergo significant ozonation, resulting in the creation of heavier oxidation products and polymerized structures. Conversely, RBO with the lowest increment at only 1.13%, likely has its unique antioxidant profile, including oryzanol and tocotrienols, which helps stabilize the oil during ozonation and reduces the formation of denser oxidation products [8].

The ozonation of vegetable oils transforms their original color, yellow, to a lighter yellowish and eventually a colorless appearance for RBO, PO, and VCO. In contrast, untreated RPO has a distinctive red color because it contains high carotene and low free fatty acid levels. It shows a significant change in color to deep yellow, losing its original color as ozonation time increases. Additionally, RPO is known for its balanced fatty acid composition, consisting of 50% saturated fats (palmitic and stearic acids), 40% monounsaturated fats (oleic acid), and 10% polyunsaturated fats (linoleic acid). This particular composition allows it to remain semi-solid after ozonation [1]. Compared to untreated vegetable oil, OVO is more transparent and has a thicker texture with a tendency to solidify.

3.2. Chemical Characterization of Ozonated Vegetable Oils

Chemical characterization of oils involves assessing key parameters like peroxide, acidity, and iodine values to evaluate the oil changes after ozone exposure. The ozonation process can modify the composition of linoleic acid in unsaturated fatty acids by breaking double bonds, forming various degradation products such as ozonides, peroxide, and aldehydes with different chemical properties [9].

Table 2 presents untreated and ozonated oils' peroxide, acid, and iodine values. RPO, RBO, and PO exhibit higher peroxide values after ozonation than VCO, which has a much lower peroxide value. This is due to their higher unsaturated fatty acid content, approximately 50% for RPO, 70% for RBO, and 80% for PO. Unsaturated fatty acids are more reactive with ozone, forming hydroperoxides, peroxides, diperoxides, and polyperoxides [10].

Despite its relatively low level of unsaturated fatty acids (9%), OVCO shows a significant increase in peroxide value from 3.78 meq/kg to 26.31 meq/kg, a rise of 596.03%. This notable increase is likely due to the high concentration of lauric acid (40%-60%) in VCO, which is reactive with ozone. The ozonation process generates various reactive intermediates that contribute to increased peroxide values. A higher peroxide number indicates more ozone reacted with fatty acids, reflecting the effectiveness of the ozonation process. Measuring peroxide levels is crucial, as peroxides are indicators of oxidation and contribute to the biological activity of oils [10,11].

https://doi.org/10.31436/iiumej.v26i2.3410

Table 1. Physical properties of the vegetable oil

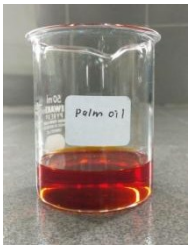
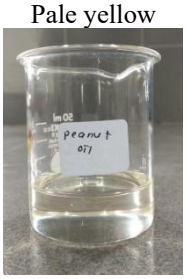
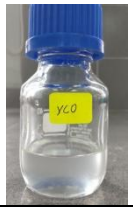
Types of Oil	Color Changes		Density (g/ml)	
	Before	After	Before	After
RPO	Red 	Deep Yellow 	0.890	0.962
RBO	Yellow 	Yellow 	0.885	0.895
PO	Pale yellow 	Clear Off white 	0.920	0.935
VCO	Off white 	Clear Off white 	0.910	0.945

Table 2. Chemical characterization of untreated vegetable oils and OVO

Types of Oil	Peroxide Value		Acidity Value		Iodine Value	
	Untreated vegetable oils	OVO	Untreated vegetable oils	OVO	Untreated vegetable oils	OVO
RPO	38.34	116.16	2.39	3.52	3.74	1.69
RBO	34.87	71.52	0.75	2.13	2.06	1.82
PO	13.91	63.01	1.09	7.86	3.52	1.38
VCO	3.78	26.31	0.30	0.91	1.37	0.32

The acid value measures the amount of free fatty acid in oil, reflecting its degree of oxidative degradation. Table 2 shows that the acid values of oils generally increase after ozonation, signifying higher acidity. OPO exhibits the most significant rise in acid value, increasing from 1.09 mg NaOH/g to 7.86 mg NaOH/g, representing a 621.1% increase. This substantial increase is likely due to the breakdown of triglycerides into free fatty acids during ozonation, a process accelerated by the high reactivity of ozone with double bonds in unsaturated fatty acids, such as oleic acid or linoleic acid [12]. This reaction forms peroxides,

https://doi.org/10.31436/iiumej.v26i2.3410

aldehydes, and carboxylic acids, ultimately contributing to higher acidity levels. In contrast, RPO shows the smallest increase in acid value, rising by only 47.28% from 2.39 mg NaOH/g to 3.52 mg NaOH/g. This is likely due to antioxidants like tocopherols and carotenoids, which can neutralize free radicals and reduce the formation of acidic compounds during ozonation [1].

The iodine value measures the degree of unsaturation in oils, indicating the number of double bonds available to react with iodine. Ozonation reduces iodine values across all oil types due to ozone's cleavage of double bonds [4]. Thus, all types of OVO have decreased iodine values as depicted in Table 2, which suggests that ozonation succeeds in attacking and lowering the double bond on the vegetable oils. OVCO experiences the highest reduction in iodine value by 76%, attributed to its high saturation and susceptibility to ozone. Conversely, ORBO demonstrates the lowest reduction in iodine value due to its unique antioxidant content, including γ-oryzanol, tocopherols, and tocotrienols. These antioxidants shield the double bonds in unsaturated fatty acids from oxidation by neutralizing reactive intermediates, thereby preserving a higher degree of unsaturation [13].

3.3. FT-IR Spectroscopy

The FT-IR spectroscopy is used to highlight differences in the functional groups before and after the oil ozonation, particularly the decrease of the bands corresponding to both C = C and =C–H stretching and the increase of the band corresponding to ozonide C-O and peroxides stretching [14].

From Figure 1 (a-c), the conversion of double bonds to ozonides was observed in ORPO, ORBO and OPO as shown by the expected decreases in the characteristic wave numbers for C=C double bond stretching (1500 cm^{-1}) and =C–H stretching (3000 cm^{-1}), along with an increase in the ozonide C-O stretching signal (1100 cm^{-1}), which aligns with Criegee's mechanism during ozonation [5]. Additionally, ozone reacts with the linoleic and oleic acids in all vegetable oils under study, producing peroxides at 2920 cm^{-1} [15]. The FT-IR spectrum of each OVO consistently shows a reduction in the frequency of the characteristic wave associated with double bonds. However, in OVCO, the presence of ozonides is not visible and remains undetected, though a peak at C-H is observed (Fig.1d). This absence of detectable ozonides may be due to the lower amount of unsaturated fatty acids in OVCO [12], which leads to ozone reacting more with saturated fatty acids like lauric acid. This reaction could result in the formation of fewer aldehydes. In the IR spectra of an aldehyde, a peak typically appears around 2720 cm^{-1}. It often manifests as a shoulder peak adjacent to the alkyl C–H stretching vibrations compared to more unsaturated oils. The resulting ozonide intermediates may be unstable or transient, quickly decomposing or reacting with other constituents in the oil, which could explain their undetectable levels. The chemical changes are further supported by the data in Table 2, where OVCO exhibits the lowest number of peroxide values compared to other OVOs.

https://doi.org/10.31436/iiumej.v26i2.3410

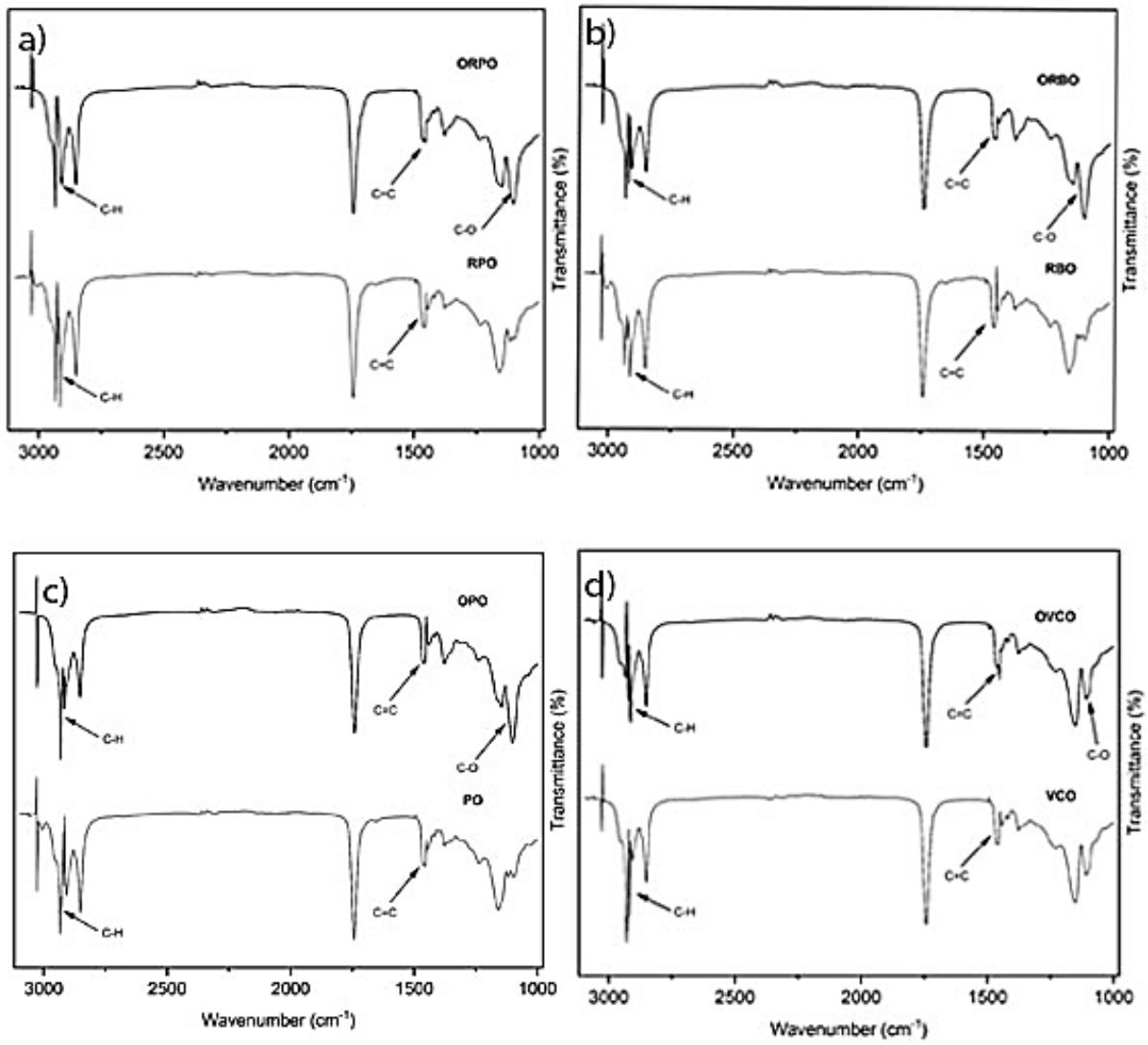


Figure 1. FT-IR Spectroscopy of the untreated and OVO; a) RPO, b) RBO, c) PO, d) VCO

3.4. Antibacterial screening of different types of OVO

OVO exhibits significant activity against the tested bacterial strains, as its ozonation by-products (such as peroxides and ozonides) contribute to its ability to destroy bacterial membranes [16]. In agar diffusion, the size of the zone of inhibition diameter in the antibacterial studies indicated that OVO has significant activity against the bacteria tested under study, as shown in Table 3. Gram-positive bacteria: *B. subtilis* and *S. aureus* are the strains most sensitive to OVO and have a larger inhibitory zone diameter than *E. coli*. The thick peptidoglycan layer in *B. subtilis* and *S. aureus* makes them more susceptible to certain antibacterial substances like OVO. The high peroxide value of OVO is the main reason for its high antibacterial activity. The previous study's findings [17] confirm this hypothesis, as they discovered a link between the peroxide value of OVO and its antibacterial efficacy against gram-positive bacteria.

In the zone of inhibition test, it was observed that both ORPO and OVCO exhibited significant antibacterial activity compared to the other types of OVO studied. The inhibition zone diameters for ORPO against E. coli, B. subtilis, and S. aureus are 3.7, 8.0, and 10.2 mm, respectively. Meanwhile, for OVCO, inhibition zone diameters are 6.6, 8.4, and 11.3 mm,

https://doi.org/10.31436/iiumej.v26i2.3410

respectively, the highest among the other tested OVOs. This finding shows that peroxide is not the only ozonation by-product toxic to the bacteria, as OVCO has the lowest peroxide value. A plating method was used to confirm the result further and identify which specific bacteria were most sensitive to OVCO and ORPO.

Table 3. Antibacterial activities of vegetable oils using the zone of inhibition test

Types of Oil	Zone of Inhibition Diameter (mm)		
	Bacteria		
	E. coli	*B. subtilis*	*S. aureus*
RPO	4.5	6.0	5.0
ORPO	3.7	8.0	10.2
RBO	5.4	5.0	5.0
ORBO	3.9	6.1	9.7
PO	3.8	4.2	6.0
OPO	5.2	7.3	9.5
VCO	5.5	7.0	5.5
OVCO	6.6	8.4	11.3

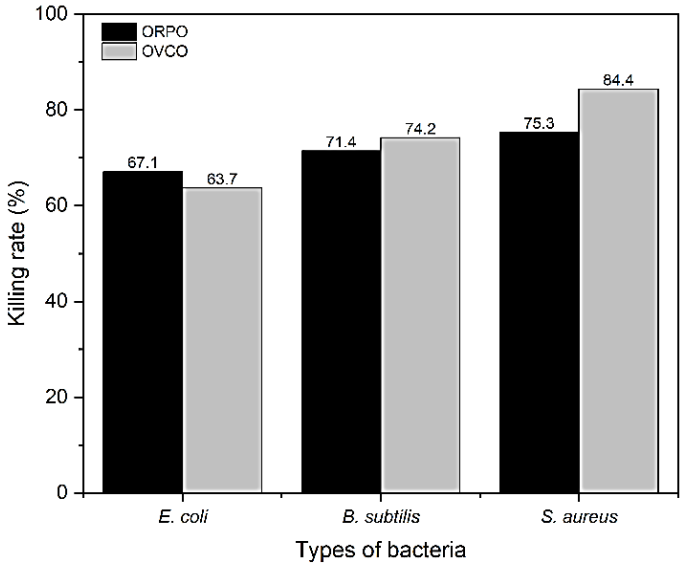


Figure 2. Antibacterial activities of OVO using a plating method.

Figure 2 shows that *B. subtilis* and *S. aureus* are the most sensitive towards OVCO and ORPO. The results indicate that OVCO can kill up to 74.2% of *B. subtilis* and up to 84.35% of *S. aureus* within a 30-minute incubation period compared to ORPO. Hence, OVCO has been identified as having the highest antibacterial activity against the gram-positive bacteria *S. aureus* compared to *B. subtilis* and *E. coli*. This high level of effectiveness is likely due to the presence of lauric acid in VCO (40-60%) compared to RPO (0.2%). Lauric acid is known to have antibacterial, antiviral, antifungal, and antiprotozoal properties and can also enhance the immune system [18]. Studies have shown that ozonation of oils by reacting with saturated fatty acids like lauric acid can enhance their antibacterial activity [17]. Thus, the high lauric acid content of VCO and the potential benefits of ozonation enhanced the synergistic effect of OVCO for antibacterial activity. Therefore, OVCO was chosen to be further optimized for its antibacterial properties against *S. aureus* by varying the ozone treatment time and flow rates.

https://doi.org/10.31436/iiumej.v26i2.3410

3.5. Optimization of the OVO conditions

The effects of two parameters (A: ozonation exposure time and B: ozone flow rate) for treating OVCO were analyzed using a percentage of killing rate towards *S. aureus* as the response test to determine their bactericidal activities. The ozonation conditions (120–360 mins and 1–3 l/min) used in this study are consistent with previous research on vegetable oil ozonation. These parameters balance effective ozone interaction with the oil's unsaturated compounds while minimizing oxidative degradation and unwanted by-products. In this design, a total of 13 runs were carried out in triplicate, and the experimental results are shown in Table 4.

Table 4. Response Surface Method with Face-Centered Central Composite Design and Responses

Run	Factor 1: Ozone Exposure Time (mins)	Factor 2: Ozone Flow Rate (l/min)	Response: Killing Rate (%)
1	120	3	78.3
2	360	2	100
3	360	1	100
4	240	3	87.01
5	240	1	99.78
6	240	2	91
7	120	1	90.33
8	240	2	92.7
9	240	2	98
10	120	2	90.3
11	240	2	93.4
12	360	3	97.8
13	240	2	93

Table 4 suggests that runs 1 through 13 of an experiment show significant antibacterial activity against *S. aureus*. The highest percentage of killing rate (100%) was achieved in runs 2 and 3 with 360 mins of ozone exposure time and 2 l/min and 1 l/min flow rate, respectively. Run 1, with 120 mins of ozone exposure time and 3 l/min ozone flow rate, had the lowest percentage of killing rate at 78.3%. The findings indicate that longer ozone exposure times with lower flow rates improve the bactericidal activity of OVCO. Longer reaction times allow for more complete interactions between ozone and the double bonds in the oil. In contrast, higher ozone flow rates may decrease the contact time and reduce the effective ozone concentration available for reaction [19].

Table 5. Analysis of Variance (ANOVA) for killing rate (R1)

Source	Sum of Squares	Degree of Freedom, df	Mean Square	F-value	p-value	Impact
Model	373.31	2	186.66	21.02	0.0003	Significant
A: Ozone Exposure Time	251.81	1	251.81	28.36	0.0003	
B: Ozone Flow Rate	121.50	1	121.50	13.68	0.0041	
Residual	88.81	10	8.88			
Lack of Fit	61.48	6	10.25	1.50	0.3620	Not significant
Pure Error	27.33	4	6.83			
Cor Total	462.12	12				

The ANOVA analysis in Table 5 suggests that the model and selected parameters used in the Design Expert software are linear and significant, with a p-value less than 0.05. The optimal

https://doi.org/10.31436/iiumej.v26i2.3410

ozone conditions for the antibacterial study were 357.52 mins of ozonation with a 1.39 l/min ozone flow rate.

The ANOVA analysis in Table 5 tests the significance of the ratio of mean square variation due to the regression coefficient and residual error. The ANOVA result for R1 shows an F-value of 21.02, implying that the terms in the model significantly affect the responses. A large value of F indicates that most of the variation in the response can be explained by the regression models. The associated p-value estimates whether F is large enough to indicate statistical significance. A p-value of 0.0003, lower than 0.05, indicates that the model is statistically significant. Results show that both selected parameters in the optimization study are significant for antibacterial properties, which is crucial for the potential use of OVCO in skincare and medical treatments.

3.6. Kinetic study of OVCO against *S. aureus*

Throughout the study, OVCO inhibits a significant antibacterial effect against *S. aureus*, and the killing rate increases as the ozonation time increases and the flow rate decreases. The growth pattern for the killing rate of *S. aureus* when exposed to OVCO ozonated for 120 mins and 360 mins with a flow rate of 1 l/min was examined.

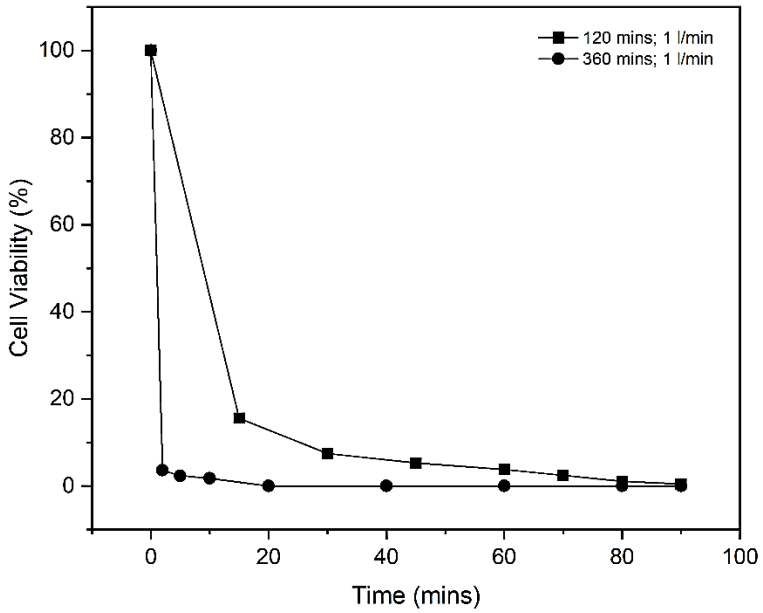


Figure 3. Kinetic study of OVCO against *S. aureus* at different times

Figure 3 depicts the quantitative evaluation of antibacterial properties for OVCO using a biocidal kinetic test. At the start of the incubation phase, the cell survival rate remained constant at 100%. After 60 mins of incubation, the residual live *S. aureus* was detected at 3.78% and 0.01% for 120 mins and 360 mins of OVCO ozonation, respectively. The killing rate for OVCO ozonated at 360 mins was very fast, which killed 96.4% of the bacterial cells within 2 2-minute incubation period. The graph shows a sharp decline in cell viability to 3.6%. The difference in the kinetic pattern of the killing rate proves that the ozone exposure time significantly affects the bactericidal properties. However, if a shorter ozonation period of vegetable oil can still result in strong antibacterial properties, it would be more efficient and cost-effective for manufacturers to produce the OVO along with environmental benefits, since 120 mins of OVCO can kill almost 100% of bacteria within 60 mins.

https://doi.org/10.31436/iiumej.v26i2.3410

3.7. Cytotoxicity Test

The cytotoxicity of OVCO was further evaluated using HFF1 fibroblasts, a model for normal skin cells. Based on Figure 4, it was observed that OVCO demonstrated minimal toxicity and can be safely administered, as 100% cell viability was achieved at a concentration of 1 mg/ml for samples exposed to ozone for 120 mins. This suggests that OVCO treated with 120 mins of ozone exposure is the safest formulation among the tested samples. However, extended ozone exposure times resulted in reduced cell viability. Specifically, samples exposed to ozone for 240 mins displayed slightly higher than 80% cell viability at 1 mg/ml, while samples exposed for 360 mins showed less than 80% cell viability.

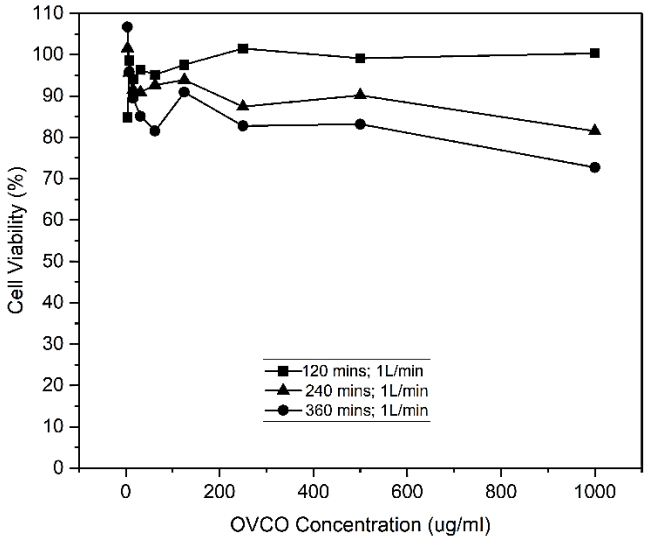


Figure 4. Dose-dependent cytotoxic test of the OVCO at different treatment times

Despite the observed decline in cell viability at higher ozone exposure times, OVCO consistently maintained over 80% cell viability at concentrations up to 500 mg/ml, indicating low cytotoxicity. These findings align with previous studies on ozonated oils, such as ozonated olive oil [3] and sunflower oil [2], demonstrating low toxicity to normal skin cells while retaining therapeutic benefits, including antimicrobial and wound-healing properties. The findings demonstrate the potential of OVCO as a therapeutic agent, particularly for dermatological and medical applications. The combination of low toxicity at therapeutic concentrations and strong antibacterial properties makes it a promising option for skincare and medical treatments. The peroxide value of OVCO is significantly lower than that of other OVOs, which enhances its safety profile. Maintaining a low peroxide value is essential to minimize oxidative damage to healthy cells, making this formulation highly suitable for managing chronic wounds, burns, and various dermatological conditions where protecting healthy tissue is a priority. These characteristics support its role as a safe and practical choice for skin tissue care [20].

4. CONCLUSION

Ozonation of vegetable oils (RPO, RBO, PO, and VCO) resulted in increased antibacterial properties against *E. coli, B. subtilis*, and *S. aureus* compared to untreated samples. From the screening processes, OVCO was found to have the highest bactericidal activity among the studied OVOs, with notable potency against gram-positive bacteria. The high bactericidal activity of OVCO is believed to result from the combined effects of lauric acid and ozonation by-products such as peroxides and aldehydes. Furthermore, optimization of ozonation

conditions through DOE achieved a 100% bacterial killing rate within 2 minutes, using an ozone exposure time of 360 minutes and a flow rate of 1 l/min. The kinetic study confirmed that longer ozone exposure times reduce the time needed to reach a 100% bacterial killing rate. The cytotoxic test also demonstrated that OVCO is non-toxic to normal skin cells. However, despite its strong antibacterial properties, OVO requires the addition of a surfactant as a solubilizing agent to improve its bioavailability. Future studies could focus on enhancing the properties of OVO formulations without chemical surfactants to increase their efficacy further. These findings provide a strong foundation for exploring the pharmaceutical potential of OVO.

REFERENCES

[1] Tan, C. H., Lee, C. J., Tan, S. N., Poon, D. T. S., Chong, C. Y. E., & Pui, L. P. (2021) Red Palm Oil: A review on processing, health benefits and its application in food. *Journal of Oleo Science.* Retrieved January 20, 2023, from https://doi.org/10.5650/jos.ess21108

[2] Tavassoli-Kafrani, M. H., Foley, P., Kharraz, E., & Curtis, J. M. (2016). Quantification of Nonanal and Oleic Acid Formed during the Ozonolysis of Vegetable Oil Free Fatty Acids or Fatty Acid Methyl Esters. JAOCS, Journal of the American Oil Chemists' Society, 93(3), 303–310. https://doi.org/10.1007/s11746-015-2780-7

[3] Puxeddu, S., Scano, A., Scorciapino, M. A., Delogu, I., Vascellari, S., Ennas, G., Manzin, A., & Angius, F. (2024). Physico-Chemical Investigation and Antimicrobial Efficacy of Ozonated Oils: The Case Study of Commercial Ozonated Olive and Sunflower Seed Refined Oils. Molecules, 29(3). https://doi.org/10.3390/molecules29030679

[4] Lenart-Boroń, A., Stankiewicz, K., Bulanda, K., Czernecka, N., Heliasz, M., Hunter, W., Ratajewicz, A., Khachatryan, K., & Khachatryan, G. (2024). In Vitro Antibacterial Activity of Ozonated Olive Oil against Bacteria of Various Antimicrobial Resistance Profiles Isolated from Wounds of Companion Animals. International Journal of Molecular Sciences, 25(6). https://doi.org/10.3390/ijms25063557

[5] Guerra-Blanco, P., Chairez, I., Poznyak, T., & Brito-Arias, M. (2021). Kinetic Analysis of Ozonation Degree Effect on the Physicochemical Properties of Ozonated Vegetable Oils. Ozone: Science and Engineering, 43(6), 546–561. https://doi.org/10.1080/01919512.2020.1868972

[6] Semeniuc, C. A., & Mureşan, V. (2023). Spectroscopic, Chromatographic, and Chemometric Techniques Applied in Food Products Characterization. In Separations (Vol. 10, Issue 1). MDPI. https://doi.org/10.3390/separations10010055

[7] Colombo, M., Ceci, M., Felisa, E., Poggio, C., & Pietrocola, G. (2018). Cytotoxicity evaluation of a new ozonized olive oil. *European journal of dentistry*, *12*(4), 585–589. https://doi.org/10.4103/ejd.ejd_422_18

[8] Ugazio, E., Tullio, V., Binello, A., Tagliapietra, S., & Dosio, F. (2020). Ozonated oils as antimicrobial systems in topical applications. Their characterization, current applications, and advances in improved delivery techniques. In Molecules (Vol. 25, Issue 2). MDPI AG. https://doi.org/10.3390/molecules25020334

[9] Absalome, M. A., Massara, C. C., Alexandre, A. A., Gervais, K., Chantal, G. G. A., Ferdinand, D., Rhedoor, A. J., Coulibaly, I., George, T. G., Brigitte, T., Marion, M., & Jean-Paul, C. (2020). Biochemical properties, nutritional values, health benefits and sustainability of palm oil. In Biochimie (Vol. 178, pp. 81–95). Elsevier B.V. https://doi.org/10.1016/j.biochi.2020.09.019

[10] Salsabila, N., Moulydia, F., & Bismo, S. (2018). Formulation of oleozon withPhaleria macrocarpa and Cinnamomum burmanii extract for diabetic wound treatment. IOP Conference Series: Materials Science and Engineering 334, 012069. doi:10.1088/1757-899x/334/1/012069

[11] Silva, V., Peirone, C., Amaral, J. S., Capita, R., Alonso-Calleja, C., Marques-Magallanes, J. Poeta, P. (2020). High Efficacy of OVO on the Removal of Biofilms Produced by Methicillin-

Resistant Staphylococcus aureus (MRSA) from Infected Diabetic Foot Ulcers. Molecules, 25(16), 3601. doi:10.3390/molecules25163601

[12] Siddiqua, T., Uddin, I., Hasan, M. R., & Begum, R. (2024). Effect of heating and Re-heating on physico-chemical properties of rice bran oil and soybean oil. Journal of Agriculture and Food Research, 15. https://doi.org/10.1016/j.jafr.2024.100979

[13] Petani, L., Jung, A. M., Frietsch Musulin, R. R., Sturm, G., Kaster, A. K., & Pylatiuk, C. (2024). Exploring the Antibacterial Effects of Ozonated Oils in Medicine: A Study on Escherichia coli Inhibition. Ozone: Science and Engineering, 46(2), 174–181. https://doi.org/10.1080/01919512.2023.2233746

[14] Travagli, V., Zanardi, I., Valacchi, G., & Bocci, V. (2010). Ozone and OVO in Skin Diseases: A Review. Mediators of Inflammation, 2010, 1–9. doi:10.1155/2010/610418

[15] Enjarlis, Handayani, S., & Anwar, Y. (2018). Synthesis and Characterization of Cocozone Oil as Skin Care Ingredient. Retrieved April 21, 2022, from doi:10.14419/ijet.v7i3.32.18415

[16] Amri, Z., Ben Hamida, S., Dbeibia, A., Ghorbel, A., Mahdhi, A., Znati, M., … Hammami, M. (2020). Physico-chemical Characterization and Antibacterial Activity of Ozonated Pomegranate Seeds Oil. Ozone: Science & Engineering, 1–8. https://doi.org/10.1080/01919512.2020.1735993

[17] Slavinskien˙, G. E., Grigonis, A., Ivaškien˙, M., Sinkevičien˙ E, I., Andruleviči, V., Ut˙ E, ¯, Ivanauskas, L., Juodžent˙, D., Ramanauskien˙, K., Daunoras, G., & Kriaučeli, D. L. (2024). A Comparative Study of the Chemical Properties and Antibacterial Activity of Four Different Ozonated Oils for Veterinary Purposes. https://doi.org/10.3390/vetsci

[18] Widianingrum, D. C., Noviandi, C. T., & Salasia, S. I. O. (2019). Antibacterial and immunomodulator activities of virgin coconut oil (VCO) against Staphylococcus aureus. Heliyon, 5(10), e02612. https://doi.org/10.1016/j.heliyon.2019.e02612

[19] Arif Malik, M. (2023.). INTRODUCTION TO ORGANIC AND BIOCHEMISTRY. https://LibreTexts.org

[20] Balea, A., Ciotlăuș, I., Pojar-Fenești, M., & Carpa, R. (2023). Comparative Chemical And Antimicrobial Characterization Of Non-Ozonated And Ozonated Vegetable Oils. Studia Universitatis Babes-Bolyai Chemia, 2023(1), 285–301. https://doi.org/10.24193/subbchem.2023.1.21

https://doi.org/10.31436/iiumej.v26i2.3452

NATURAL BITUMEN IN HOT ASPHALT MIXTURE: SUITABILITY OF USING TREATED NATURAL BITUMEN INSTEAD OF PETROLEUM ASPHALT BINDER

HANEEN KAREEM MOHSIN, ROAA HAMED LATIEF*

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

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

(Received: 23 September 2024; Accepted: 17 February 2025; Published online: 15 May 2025)

ABSTRACT: In recent years, the search for economic and environmentally friendly alternatives has become a global necessity to achieve sustainability and preserve raw materials. From this concept, natural bitumen (NB) derived from sulphur springs is now one of the most promising alternative energy resources for many applications, especially in asphalt pavement construction. Its low price and abundance characterise NB since sulphur springs produce thousands of tonnes of NB annually and are used in very limited fields. Two main objectives were adopted for this work. The first objective is to examine the virgin NB properties from five sulphur springs and compare them with petroleum asphalt. The second objective is to enhance NB properties by applying heat treatment. The experimental results reveal that heat treatment is highly effective in improving the NB properties. This was confirmed by conducting many tests, such as asphalt, Marshall, SEM-EDX, and indirect tensile strength tests. Based on the experimental test outcomes, virgin NB properties do not conform to asphalt specification limits and are unsuitable for flexible roads. Also, MS-NB significantly affected the mechanical properties of the asphalt mixture, as Marshall's stability increased by 41.3% compared to the conventional mixture. In addition, SS-NB was more effective in moisture damage resistance due to increasing the tensile strength ratio by 5.72 % and provided a higher stiffness index than the conventional mixture by 40.36%. In conclusion, the treated NB can successfully be used as a binder material in hot asphalt mixtures.

ABSTRAK: Beberapa tahun kebelakangan, pencarian alternatif baharu ekonomi dan mesra alam telah menjadi satu keperluan global bagi mencapai kemampanan dan memelihara bahan mentah. Melalui konsep ini, bitumen asli (NB) yang diperoleh daripada mata air sulfur kini merupakan satu sumber tenaga alternatif yang berguna bagi digunakan dalam pelbagai aplikasi, terutamanya dalam pembinaan turapan asfalt. NB dicirikan oleh harga yang rendah dan kelimpahannya kerana mata air sulfur menghasilkan beribu-ribu tan NB setiap tahun dan digunakan dalam bidang terhad. Dua objektif utama telah dipakai untuk kajian ini. Objektif pertama adalah memeriksa sifat NB dara daripada lima mata air sulfur dan membandingkannya dengan asfalt petroleum. Objektif kedua adalah meningkatkan sifat NB dengan menggunakan rawatan haba. Dapatan eksperimen mendedahkan bahawa rawatan haba sangat berkesan dalam meningkatkan sifat NB. Ini disahkan dengan menjalankan pelbagai ujian, seperti ujian asfalt, ujian Marshall, ujian SEM-EDX, dan ujian kekuatan tegangan tidak langsung. Berdasarkan dapatan eksperimen, sifat NB dara tidak mematuhi had spesifikasi asfalt dan tidak sesuai untuk jalan fleksibel. Selain itu, MS-NB mempunyai kesan ketara pada sifat mekanikal campuran asfalt kerana kestabilan Marshalls meningkat sebanyak 41.3% berbanding campuran konvensional. Tambahan, SS-NB lebih berkesan dalam rintangan kerosakan lembapan kerana meningkatkan nisbah kekuatan tegangan sebanyak 5.72 % dan memberikan indeks kekukuhan yang lebih tinggi daripada campuran konvensional sebanyak 40.36%. Kesimpulannya, NB yang dirawat boleh berjaya digunakan sebagai bahan pengikat dalam campuran asfalt panas.

KEYWORDS: *Natural bitumen, heat treatment, petroleum asphalt, Marshall, moisture damage.*

1. INTRODUCTION

Road construction is crucial for improving connectivity by creating efficient, safe, and durable routes for traffic movement, while also enhancing the quality of life for communities and promoting economic growth. Therefore, ensuring high performance and quality of paved roadways is essential to maximize progress across the industrial, agricultural, and commercial sectors. More specifically, high-performance roads are necessary for the industrial sector to accommodate larger wagons safely. For the agricultural sector, roads enable easier market access to preserve perishable products and reduce transportation costs. For the commercial sector, roads offer faster routes for transporting goods over short distances, thereby minimizing costs and reducing delays. This requires continuous road development efforts by developing maintenance programs, expanding road networks, and, most importantly, enhancing road pavement quality to prevent failures. The evolving landscape of technologies and materials supports the development of a global social economy, leading to the adoption of innovative approaches and materials in road construction [1-5].

The global issue of high-viscosity oil and natural bitumen (NB) commercial development is influenced by increasing hydrocarbon consumption, oil reserve structure changes, heavy crude dominance, and innovative technologies for bitumen production and processing [6], [7]. Oil remains a significant global resource, representing over 30% of energy consumption. In 2019, global oil stockpiles decreased by 0.1%, leading to decreased quantities of asphalt binder produced from refinery oil, as part of the crude oil is converted to asphalt binder during the distillation process. Furthermore, Venezuela, Saudi Arabia, Canada, Iran, and Iraq have the largest proven reserves [8-9]. This has prompted road engineers to search for alternatives, as asphalt affects the properties of the mixture and is considered 7-8 times more expensive than aggregate. The high cost and lack of resources have prompted the search for alternatives to industrial asphalt. NB exists in bitumen deposits, asphalt lakes, asphalt rocks, etc. It improves pavement performance and reduces hot mixture asphalt (HMA) prices [10].

The first major discovery of NB in the UK was at Alphalt in 1877, which became the first oil shale mine. However, oil extraction and refining became more critical than the mine, and it closed in 1919 as the shale oil industry phased out. Despite its numerous advantages, NB remains a lesser-known material globally because it is found in specific places worldwide and cannot be used in pavement production without treatment [11].

NB is a complex mixture of organic compounds with a chemical composition that includes long hydrocarbon chains and intense, polar aromatic rings. This gives asphalt a wide range of properties, from solid to liquid, encouraging researchers to study and improve its properties for use as an alternative in paving operations. NB is a unique material found in Trinidad and the Dead Sea. It has been used for centuries in various applications, including bath construction, embalming, shipbuilding, and waterproofing [12-13].

Bitumen, a natural asphalt found in Iraq, is formed from decomposing living organisms and solids. It is a petroleum derivative that remains on the Earth's surface after the evaporation of light elements. Bitumen is found in geological areas affected by faults or fissures, causing groundwater to flow from the ground to the surface. It contains salts and sulphur in high percentages. The Heet region is oil-rich, as evidenced by bitumen springs [14].

https://doi.org/10.31436/iiumej.v26i2.3452

Bitumen is a petroleum-derived, non-crystalline solid or viscous substance that dissolves in carbon disulfide. Its composition consists of carbon (87%), hydrogen (11%), oxygen (2%), and trace amounts of nitrogen, iron, and nickel [15]. It has a brown or black appearance. It is a suitable option for construction materials, as it is porous, brittle, and prone to oxidation. Although flammable, it is difficult to ignite [16-17]. NB is a black, hard, and brittle material made from a mixture of high molecular weight hydrocarbons, with many carbons, up to 150 atoms. It has a viscous fluid consistency, a special smell, and is easily soluble. There are two forms: solid asphalt rocks and sticky materials from warm sulphur springs [18].

Many physical and chemical processes have been used to treat and develop NB and meet the requirements of industrial use. Shlimon et al. [19] evaluated oil wells in the Kurdistan region (North of Iraq) to determine the origin of NB and its components. Also, the chemical properties of NB from the springs in the Heet region are investigated by Farhan et al. [20] by conducting many chemical tests and identifying the chemical elements of NB. Also, Muttar et al. [21] studied the NB and water of the three springs in Heet: Al-Khader, Al-Khalidiyah, and Al-Shuhada. In this study, the NB was dried, separated from the water, and treated by heating to enhance its properties. The physical and chemical properties were examined, and improvement was shown. Furthermore, the study revealed a chemical similarity between NB and conventional asphalt.

The investigation by Al-Dulaymie et al. [14] studied the components of NB from the Abu Al-Jeer spring. It evaluated the characteristics of the water from this spring and its components. In this study, a ranking and screening technique was developed to assess preferred sulphur springs specified for natural therapy, and the application of this technique demonstrates four graded, consequent preferred sulphur springs for balneotherapeutic investment. In Abdul-Jaleel and Najres [22] investigation, oxidation treatment was applied to enhance the properties of NB produced from the Abu Al-Jeer spring. In this treatment, NB was heated in an oven at 160°C for different periods: 5, 10, 25, 35, and 55 hours. The chemical test results indicated that the aromatic ratio was decreased, whereas the asphaltenes increased due to oxidation. This leads to the transformation of the NB to a gel type. Also, increasing viscosity and ductility, along with reducing penetration, are other effects of oxidation treatment on the physical properties of NB.

The extraction and examination of NB for use as an alternative to refinery asphalt were studied by Abdul-Jaleel et al. [23]. There were two methods to conduct the study: the first, using liquid chromatography, the asphalt was separated, then Sarah N-pentane was used as a solvent to examine the chemical properties, and Fourier Transform Infrared (FT-IR) analysis was performed to analyze the components thoroughly. Improving the rheological properties of NB derived from the Abu-Aljeer sulphur spring was the second strategy by combining NB after removing water in specific proportions (5, 15, 35, and 35) % with natural, readily available, in large quantities in the Al-Anbar governorate, and low-cost minerals (limestone), due to the great effectiveness of calcium oxide CaO (limestone) and its ability to create bonds of adsorption Chemo–Physics to form the connection between asphalt hydrocarbon chains. The outcomes revealed that the softening point, flash point, fire point, and viscosity are increased, and the penetration is decreased with the addition of limestone; moreover, the optimal ratio of limestone addition is 35%.

Mahmood et al. [24] studied and evaluated the NB derived from Heet city and used polyvinyl chloride (PVC) pipes as a modifier to enhance the properties of NB and lower the harmful impact of this waste on the environment. PAV powder was added to NB in three percentages (10, 20, and 30%) by the weight of NB. The findings indicated that adding PVC to NB produced a notable decrease in penetration and a significant increase in viscosity. This

enhancement in the physical properties of NB is due to the chemical compatibility between NB and PVC, which provides more adhesion and cohesion between the two, as both PVC and NB are considered hydrocarbon materials. According to local and international specifications, NB modified with 45% to 60% of PVC is suitable for roofing works such as waterproofing and insulation purposes. Meanwhile, Ahmed et al. [25] used heat treatment to enhance the properties of NB from the Abu Al-Jeer spring to use it in asphalt pavement. NB is heated at a temperature of 163°C for various periods (5, 10, 15, 20, and 25 hours) to improve its physical properties. The results showed improved penetration, ductility, softening point, and other properties. The Marshall test for asphalt mixtures containing NB demonstrated an increase of 17.8% in Marshall stability values compared to asphalt mixtures made from petroleum asphalt, as heat treatment increased the stiffness of NB more than that of petroleum asphalt. To avoid the external heat treatment costs, Ahmed et al. [16] studied the process of mixing NB with petroleum asphalt in different proportions (20, 40, 60, and 80) %. The outcomes of the experimental tests revealed that the combination of NB and refinery asphalt leads to the production of a new asphalt that meets the requirements of the local standard for asphalt used in road construction. The mixtures that contain 80% NB have greater Marshall stability values than traditional mixtures by 23.5%.

In the present research, NB was supplied from five sulphur springs: Al-Mamora, Al-Jabal, Atatt, Al-Atffa, and Al-Askaree, which were located in the city of Heet, Anbar Governorate. Previous studies have been conducted to evaluate the NB from these springs and have focused on its physical properties. Therefore, an extensive and deep analysis is needed to understand the nature of NB produced from these springs. This research has two main aims. The first aim is to investigate the suitability of virgin NB in producing HMA within the specification limits. The second aim is to improve the physical properties of NB by applying a heat treatment for different periods depending on the spring type. Also, the chemical composition and microstructure scanning of treated NB and petroleum asphalt are examined. In addition, the current study assesses NB's ability to improve mechanical properties and resistance to water damage for asphalt pavement. Finally, the experimental results for the virgin and treated NB are compared to those of petroleum asphalt from crude oil refineries.

2. STUDY AREA

The sulphur springs are the primary source of NB in Iraq. The research area comprised many bitumen springs in Heet, inside the Anbar Governorate, approximately 190 km west of Baghdad, Iraq, as shown in Figure 1(a). The city of Heet was situated between longitudes 15°42'-15°43' to the east and latitudes 15°33'-38°34' to the north. Five of the numerous springs in the study region were selected for this study, as indicated in Figure 1(b). The selected springs are Al-Mamora, Al-Jabal, Atatt, Al-Atffa, and Al-Askaree.

https://doi.org/10.31436/iiumej.v26i2.3452

(a) Anbar governorate, Iraq (b) Heet city, Anbar governorate

Figure 1. Sulphur spring sites in Heet city, Anbar governorate.

3. MATERIALS

This project used materials locally in Iraq to achieve economic benefit and take advantage of national resources. The materials are listed in the following sections.

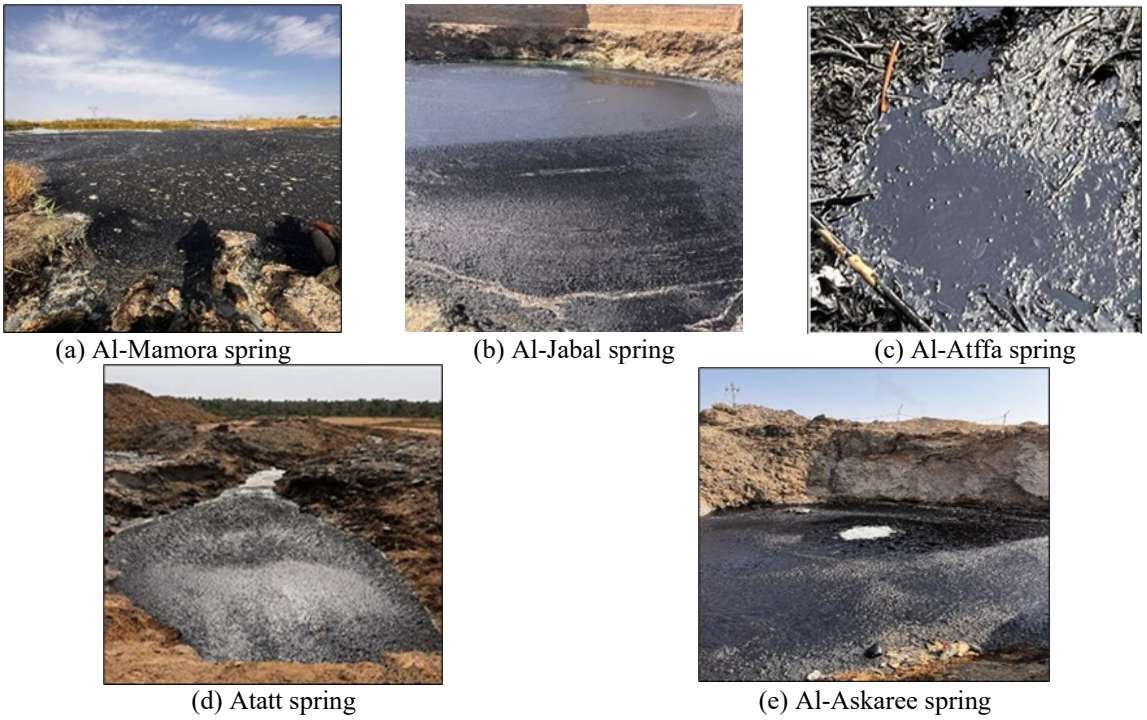
3.1. Petroleum Asphalt

It is derived from the distillation process of crude oil in the Doura refinery and asphalt approved for use in central and southern Iraq for paving AC (40-50), following the specifications of the Department of Roads and Bridges [26].

3.2. Natural Bitumen

Natural bitumen (NB) was provided from five springs in different areas of Anbar Governorate in Iraq. As illustrated in Figure 2, NB extracted from these springs differs in physical and chemical properties from the synthetic asphalt resulting from the destructive distillation of crude oil, whose characteristics are largely controlled by the Doura refinery. NB contains a large water concentration and pollutants, so it cannot be used directly. Therefore, NB must be dried and treated before being used in paving operations.

https://doi.org/10.31436/iiumej.v26i2.3452



(a) Al-Mamora spring (b) Al-Jabal spring (c) Al-Atffa spring

(d) Atatt spring (e) Al-Askaree spring

Figure 2. Sulphur springs in Heet city.

3.3. Fine and Coarse Aggregate

The fine and coarse aggregates in this work were provided by the Al-Nibaie quarry. Aggregate retained between sieve No. ¾" and sieve No. 4 are categorized as coarse aggregates. Meanwhile, the aggregate retained between sieve No. 4 and sieve No. 200 is categorized as fine aggregates. In the road laboratory, fine and coarse aggregates were sieved and combined to produce the selected gradation within the limits of the local specification [26], as illustrated in Figure 3. The physical characteristics of fine and coarse aggregates were tested and listed in Table 1.

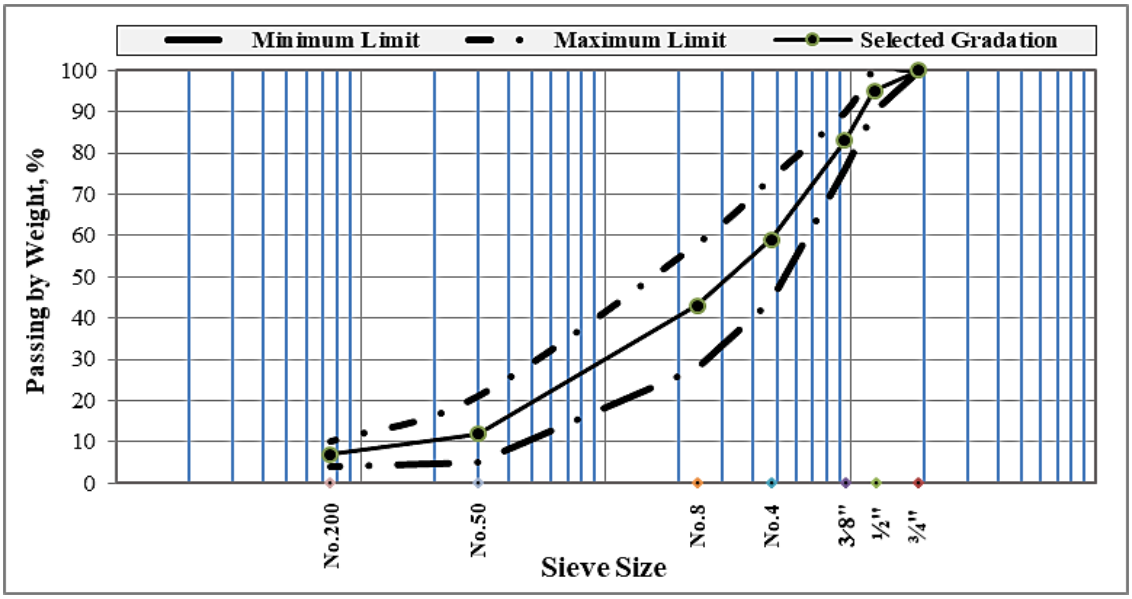


Figure 3. Aggregate gradation curve for wearing layer.

Table 1. Fine and coarse aggregates properties

Property	Test Method	Test Result	Specification Limit [26]
Coarse aggregate			
Wearing (Los Angeles Abrasion), %	ASTM C-131	11.6	30 Max.
Bulk Specific Gravity	ASTM C-127	2.63	-
Water Absorption, %	ASTM C-127	0.19	-
Fractured Pieces, %	ASTM D-5821	96.3	90 Min.
Flat and Elongated Particles, %	ASTM D-4791	3.75	10 Max.
Fine aggregate			
Water Absorption, %	ASTM C-128	0.93	-
Bulk Specific Gravity	ASTM C-128	2.661	-

3.4. Mineral Filler

The limestone filler was prepared from the city of Karbala in southeastern Iraq. The properties of this filler were tested, and the outcomes are presented in Table 2.

Table 2. Limestone properties

Property	Result
Specific Gravity	2.73
Percentage Passing Sieve No. 200 by Weight, %	94

4. WORK PROGRAM

NB's properties were investigated through experimental testing both before and after heat treatment. In addition, many tests were applied to investigate the suitability of using the treated NB as a binder material in preparing HMA instead of petroleum asphalt.

4.1. Natural Bitumen Processing

Drying and isolating the NB extracted from the sulphur springs is the initial action in the treatment process. This is accomplished by heating the NB in a convection oven to a temperature of 110 °C for a variable amount of time, depending on the spring's type and water content, while keeping in mind that asphalt does not burn [27]. The second stage of treatment involves heating the NB in a convection oven to a temperature of 163 °C, with the required time varying for each sample, depending on the confirmation that the bitumen has been separated from the water and making it conform to the local specifications [26]. NB is obtained from five springs and treated as follows:

4.1.1. Al-Mamora Spring (MS)

Al-Mamora Spring is located on the Heet-Kabisa road. Because of its liquid nature and high water content, the NB recovered from it requires two hours to dry in an oven at 110 °C. Next, to align with the properties of asphalt AC (40-50) used in paving operations, following the local specification [26], the NB is heated in a convection oven at 163 °C for 17 hours, divided into periods of 5, 10, 15, and 17 hours.

4.1.2. Al-Jabal Spring (JS)

Al-Jabal Spring is in the Jabal neighborhood of Heet-Anbar, also known as Ein Al-Khader. The liquid NB contains a significant amount of water, which is removed by drying in the oven at 110 °C for approximately two and a half hours. Then, it is equipped with heating in the oven

at 163 °C for 25 hours and separated into periods ranging from 5 to 25 hours to match the asphalt used for pavement operations according to the local specification [26].

4.1.3. Atatt Spring (TS)

Atatt Spring is located in Heet-Anbar, close to the Al-Mamora Spring. The drying process for the NB and removing water takes two hours in an oven at 110 °C. The heat treatment lasts 21 hours at 163°C, with the time divided into periods 5, 10, 15, 20, and 21 hours, ensuring compatibility with the local specification [26].

4.1.4. Al-Atffa Spring (AS)

Al-Atffa Spring is located in Heet-Anbar, in the fire pit neighborhood. Drying the NB from the water took an hour and a half in an oven at 110 °C because the bitumen extracted from this spring is solid and has a low degree of penetration compared to other sulphur springs. The heat treatment takes two hours to meet the limits of the local specification [26].

4.1.5. Al-Askaree Spring (SS)

Al-Askaree Spring is located in the Askaree neighborhood of Heet-Anbar. It is also called the Siali Spring, as the NB found here is of the Siali type, which is very liquid and has a high degree of penetration. The drying process took two and a half hours in an oven at 110 °C to get rid of the water and separate the bitumen. The NB was then treated by heating for 26 hours, divided into periods of 5, 10, 15, 20, 25, and 26 hours, to conform to the local specification [26].

4.2. Asphalt Cement Tests

As shown in Table 3, many tests were conducted on petroleum asphalt, untreated NB, and treated NB from the five springs, heated for varying durations according to each sample.

Table 3. Asphalt binder tests

Test	Methods
Penetration	ASTM D-5
Ductility	ASTM D-113
Softening Point	ASTM D-36
Flash Point	ASTM D-92
Specific Gravity	ASTM D-70

4.3. Scanning Electron Microscopy-Energy Dispersive X-Ray Analysis (SEM-EDX)

The chemical composition of bitumen's molecular components significantly impacts the material's internal structure. Bitumen is a complex blend composed mainly of hydrocarbon compounds, with some structurally similar heterocyclic species, as well as functional groups comprising oxygen, nitrogen, and sulphur atoms. In addition, bitumen contains trace amounts of metals such as vanadium, nickel, iron, aluminium, and calcium, which exist as inorganic salts, oxides, or porphyrin complexes.

Energy dispersive X-ray analysis (EDX) in Figure 4 is an analytical technique used to analyze a sample's elements or chemical characterization based on the interaction of specific sources. The device is manufactured in the Netherlands. The EDX device's operation principle analyzes the X-rays emitted from the sample when the electron beam collides and interacts with the sample atoms. Each element has a distinct X-ray energy, and the EDX detector records

https://doi.org/10.31436/iiumej.v26i2.3452

the emitted X-rays, distinguishing the elements present and their concentrations. This technique is used to study environmental pollution, where EDX microanalysis can greatly detect heavy metal pollution [28].

In addition to these detectors, modern SEM tools are equipped with an inbuilt EDX sensor, which allows for compositional analysis of materials, called scanning electron microscopy (SEM) with EDX spectroscopy (SEM-EDX). EDX testing can assist road engineers in understanding the physical-chemical behavior of asphalt pavements through chemical element analysis. On the other hand, SEM images can provide information about elemental analysis in conjunction with EDX, as they help analyze the surface and interactions and understand the surface properties of compounds. Also, SEM-EDX testing helps to examine changes occurring at the atomic and molecular levels [29]. The principle of operation of SEM involves generating a beam of electrons directed to the sample through a column containing electromagnetic lenses. The electrons then interact with the surface of the sample. As a result of this interaction, different signals are emitted. The detectors collect these resulting signals to generate three- or two-dimensional images with fine details, such as surface composition.

Figure 4. SEM-EDX device.

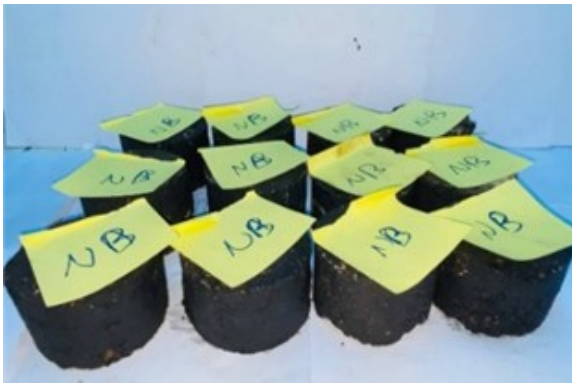
4.4. Marshall Test

As specified in ASTM D6927-15, the Marshall test is employed to arrange, compress, and verify the samples. The asphalt and aggregate are heated separately for a couple of hours, with temperatures reaching 135 °C and 150 °C, respectively. Avoiding excessive heating and preheating of the asphalt is essential, as this can alter its properties and negatively affect the overall mixture. According to the local specification [26], the asphalt binder content for the wearing layer must be between 4% and 6%. Therefore, five asphalt proportions with a constant increment of 0.5% (4.0%, 4.5%, 5.0%, 5.5%, and 6.0% based on the aggregate weight) are combined with the heated aggregate to find the optimum binder content (OBC). The mixture must be agitated for at least three minutes to ensure consistent application of the asphalt cement on every aggregate particle.

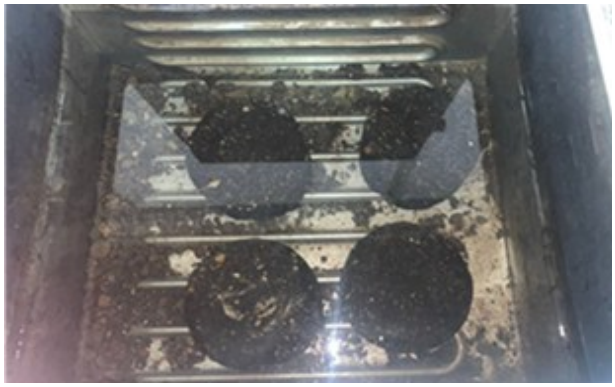
https://doi.org/10.31436/iiumej.v26i2.3452

The conventional cylindrical Marshall test moulds possess a diameter of 102 mm and a height of 64 mm. Before use, these moulds are heated to a temperature of 130 °C. The mould is then filled with the heated mixture and compressed using a Marshall hammer, applying 75 impacts from each side for heavy traffic roads. The compressed Marshall sample is permitted to cool. After a duration of 24 hours, the mould is unsealed, and the specimen is immersed in a water bath at a temperature of 60 °C for a period of 30 minutes before examination.

The mixture characteristics of petroleum asphalt and NB from the five springs post-treatment are evaluated. These characteristics include flow value, Marshall stability value, air void ratio, asphalt-filled voids, voids in mineral aggregate, and OBC. Figure 5 illustrates the detailed Marshall test procedures.



(a) Specimen preparation



(b) Specimens in a water bath

(c) Specimen testing

(d) Fractured specimens

Figure 5. Marshall test.

4.5. Moisture Damage Test

Moisture damage refers to the progressive decline in the performance of the pavement mixture caused by the weakening of the adhesive link between the bitumen and the surface and the reduction in cohesive resistance. Moisture leads to a loss of adhesion between the bitumen surfaces and aggregates. Consequently, the mixture undergoes separation, resulting in a decline in stability, strength, and hardness due to reduced cohesion.

The compacted mixtures' tensile strength ratio (TSR) was tested according to ASTM D 4867 M-96 to evaluate their susceptibility to water harm. This test is indicated to determine whether water reduces the adhesion between asphalt particles and aggregate. To achieve this,

https://doi.org/10.31436/iiumej.v26i2.3452

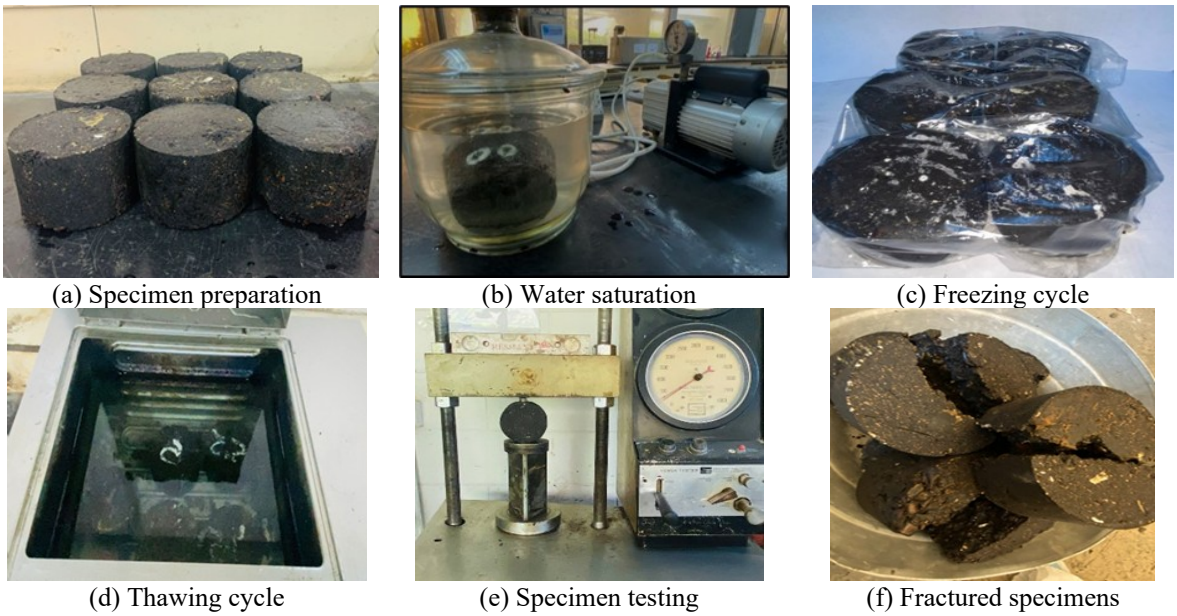
for the petroleum asphalt and NB after undergoing specific treatment durations for each spring, Marshall samples were prepared using the optimum binder content (OBC). These samples were categorized into two groups: the unconditioned group, which was placed in a water bath at a temperature of 25 °C for 20 minutes, and the conditioned group, which underwent one freeze-thaw cycle followed by immersion for one hour at the same temperature. The minimum requirement for HMA (Hot Mix Asphalt) TSR is more than 80%. TSR is calculated using the following equation:

$$TSR = \frac{IDT_{wet}}{IDT_{dry}} \times 100 \quad (1)$$

$$IDT = \frac{2000 \times P}{\pi t D} \quad (2)$$

where TSR is the tensile strength ratio (%), IDT_{wet} is the average indirect tensile strength of the conditioned groups-wet (kPa), IDT_{dry} is the average indirect tensile strength of the unconditioned groups-dry (kPa), P is the peak load, t is the specimen thickness (mm), and D is the specimen diameter (mm).

Figure 6 shows the procedures for conducting the TSR test as stated previously. Each specimen was tested by applying a rate of 50.8 mm per minute. After the load reached its maximum value, the specimen was fractured entirely, and this load was recorded.



(a) Specimen preparation (b) Water saturation (c) Freezing cycle

(d) Thawing cycle (e) Specimen testing (f) Fractured specimens

Figure 6. Moisture damage test.

5. RESULTS AND DISCUSSION

5.1. Asphalt Properties

The physical properties of untreated NB derived from springs and petroleum asphalt are shown in Table 4. Physical testing shows that petroleum asphalt meets all requirements stipulated in the local specifications [26], including penetration, flash point, ductility, and other tests. Compared to petroleum asphalt, the test results reveal that NB has a much higher penetration value and does not meet the specification requirements [26]. It needs to be treated to improve its qualities and make it conform to the local specifications for paving asphalt.

https://doi.org/10.31436/iiumej.v26i2.3452

Table 4. Physical properties of petroleum, asphalt, and untreated NB

Test	Results						Specification Requirements [26] for AC (40-50)
	Petroleum Asphalt	MS-NB	JS-NB	TS-NB	AS-NB	SS-NB	
Penetration, 1/10 mm	42	130	143	120	58	142	40-50
Ductility, cm	>100	68	63.7	66.7	65	64	>100
Softening point, ^{0}C	53.5	39	35	43.2	42.5	36.2	-
Flashpoint, ^{0}C	324	180	158	175	175	175	232 min
Specific Gravity	1.02	1.04	1.052	1.03	1.028	1.04	-

Table 5. Physical properties of treated NB

Test	Heating Period in Hours							Specification Requirements [26] for AC (40-50)
	0	5	10	15	17	-	-	
MS-NB Results								
Penetration	130	95	72	58.6	45	-	-	40-50
Ductility	68	82.9	94	>100	>100	-	-	>100
Softening point	39	45	48.7	50.4	52	-	-	-
Flash point	180	216	224	240	248	-	-	232 min
Specific Gravity	1.04	1.0415	1.042	1.043	1.0431	-	-	-
JS-NB Results								
	0	**5**	**10**	**15**	**20**	**25**	-	
Penetration	143	97	88.8	72	66	44	-	40-50
Ductility	63.7	75	80	88	>100	>100	-	>100
Softening point	35	41	46.7	48	51	53.5	-	-
Flash point	158	183	197	213	224	232	-	232 min
Specific Gravity	1.052	1.0524	1.0528	1.053	1.0532	1.054	-	-
TS-NB Results								
	0	**5**	**10**	**15**	**20**	**21**	-	
Penetration	120	83	72.6	65	50.8	44	-	40-50
Ductility	66.7	74	86	91	>100	>100	-	>100
Softening point	43.2	44.6	47	48.7	51	53.5	-	-
Flash point	175	188	190	210	230	240	-	232 min
Specific Gravity	1.03	1.031	1.0316	1.032	1.0323	1.0327	-	-
AS-NB Results								
	0	**2**	**5**	**-**	**-**	**-**	**-**	
Penetration	58	45	40	-	-	-	-	40-50
Ductility	65	>100	>100	-	-	-	-	>100
Softening point	42.5	52.3	54	-	-	-	-	-
Flash point	175	238	242	-	-	-	-	232 min
Specific Gravity	1.028	1.03	1.033	-	-	-	-	-
SS-NB Results								
	0	**5**	**10**	**15**	**20**	**25**	**26**	
Penetration	142	115	90.6	83	69	52	45	40-50
Ductility	64	70	80	88	97	>100	>100	>100
Softening point	36.2	40	43	47.6	50.1	53.5	56	-
Flash point	177	187	198	213	224	232	245	232 min
Specific Gravity	1.04	1.042	1.0426	1.043	1.0434	1.044	1.0441	-

Table 5 displays the physical characteristics of NB after treatment and reveals improvements in these properties. MS-NB is classified as liquid, with a penetration of 130 (0.1 mm). Also, it requires a heat treatment period of about 17 hours. Heating the asphalt for a longer period, for example, 20 hours, leads to poor results and increases its hardness due to a high rate of oxidation, which in turn affects the asphalt mixture and causes issues. Therefore, the ideal penetration rate was 45 (0.1 mm) after just 17 hours. Meanwhile, JS-NB has a high degree of penetration and was treated for 25 hours, after which the penetration became 44 (0.1

mm). This is within the limits of the local specifications [26] and is close to the penetration of the petroleum asphalt. Continuous mixing and heating cause the oxidation of all parts of the asphalt, which alters its properties, making it harder and giving it a dark black color. Thus, this affects the mechanical performance of the asphalt mixture.

In addition, the heating treatment for TS-NB took a total of 21 hours; if the treatment were carried out for 25 hours, the degree of penetration would be very low and not meet the limits of the standard specification [26]. At 21 hours, the penetration was 44 (0.1 mm), which complies with the specification because the heat increases the softening point of asphalt and enhances its hardness. Moreover, AS-NB is considered the best of the five types of NB because the bitumen extracted from this spring is considered solid, with a low penetration of 58 (0.1 mm), and contains a very small amount of water. Therefore, the treatment period was only two hours, the penetration became 45 (0.1 mm), and the ductility was greater than 100 cm. Therefore, this bitumen type will increase the resistance against cracking formation and extend the lifespan of the asphalt mixture. It conforms to the specification limits, and the improvement in properties is attributed to the heating process. The longer the heat treatment, the higher the percentage of sulphur, which improves the properties of asphalt.

Due to its fluid nature and high water content, SS-NB took a long time to dry. The longest heat treatment period required 26 hours to reach the required degree of penetration, which is 45 (0.1 mm). The material has a ductility exceeding 100 cm, a softening point of 56 °C, and a flash point of 245 °C, all meeting the specification limits.

The NB of Abu Al-Jeer spring was treated to improve the physical properties through the same heat treatment process for 25 hours, as conducted in the previous study by Ahmed et al. [25]. This is approximately compatible with the treatment time in this study for NB from JS and SS. In addition, Altameemi et al. [27] treated NB-MS by applying heat treatment for 20 hours to enhance its properties, which is not consistent with the treatment period in the current study, which is 17 hours. The variance in treatment periods for the same spring may be related to changes in environmental conditions over the years, especially temperature fluctuations and rainfall rate reductions, which lead to the production of different NB.

5.2. Morphology and Element Analysis Properties

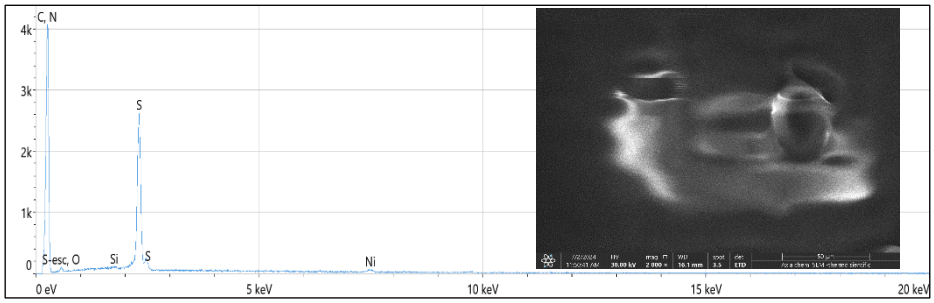
Hydrocarbons, the main component of asphalt, have paraffinic, naphthenic, and aromatic structures. They also include nitrogen, sulphur, and cyclic or non-cyclic oxygen molecules. In addition, atoms of nickel, iron, and vanadium are present. Compounds containing sulphur, nitrogen, and oxygen react with these elements. Various component fractions affect bitumen compatibility and durability. However, the balance of these components gives bitumen its unique viscoelastic characteristics, which are crucial for its use as an asphalt binder in paving works. Thus, a lack of balance or compatibility between components leads to component phase separation and undesirable characteristics [30]. Minerals significantly impact the physical properties of asphalt because non-polar atoms cause molecular interference, which alters the material's solubility, boiling temperature, and viscosity. The SEM-EDX analysis for petroleum asphalt and treated NB is shown in Figure 7.

The microstructure of the asphalt binder significantly affects its properties, as this microstructure is closely linked to asphalt chemical composition [31]. It explains the changes that occur in asphalt due to chemical and environmental factors and helps detect defects and impurities within the asphalt that affect its properties and quality. It also helps determine the components' distribution, their homogeneity with the asphalt when applied, and their effect on performance. From the SEM images, petroleum asphalt is a mass of homogeneous

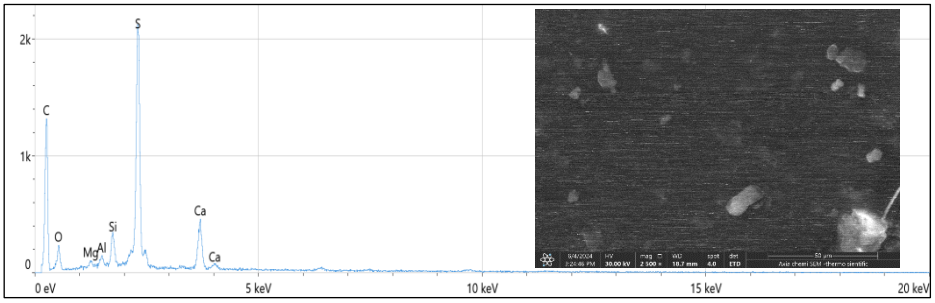
hydrocarbons, and the surface of this asphalt appeared primarily plain and clear of any noticeable features in surface morphology. This is because the chemical composition of petroleum asphalt is less diverse in its compounds than that of NB samples. It can be seen that SS-NB is approximately similar to petroleum asphalt in morphology structure compared to NB from other springs. This may be due to its high carbon content (88.7%) and lower component diversity, with no trace elements in this NB. The percentage of chemical components in petroleum asphalt and NB (after treatment) for each of the five sulphur springs is illustrated in Figure 7 and listed in Table 6. Evaluating the structure information and molecular type is important for a fundamental understanding of how the element compositions affect the chemical reactivity and physical properties of bitumen.

Besides sulphur, most of the NB from sulphur springs consists of organic components such as carbon and oxygen, small amounts of trace components such as iron, magnesium, and silicon, and other contaminants. Also, NB is free of toxic elements such as lead, mercury, and fluoride, which is advantageous for environmental preservation.

SS-NB had the highest carbon content, at 88.7. Increasing carbon has a positive effect on the asphalt mixture, improving its flexibility, which decreases the failure stress of asphalt at low temperatures. This is because carbon improves asphalt properties, especially anti-aging behavior, and thermal and electrical conductivity, enhancing the behavior of asphalt roads under different climatic conditions. Also, it enhances the asphalt's ability to resist deformation at high temperatures and cracking under cold conditions, ultimately improving the durability of the asphalt [32]. Asphalt contains sulphur, nitrogen, and oxygen, known as heteroatoms that encourage reactions in materials. Compared to the hydrocarbon moiety, heteroatoms are a minor component, and their concentrations are not constant and vary depending on the source of bitumen. This is attributed to heteroatom moieties generally imparting polarity and functionality to the molecules, contributing to differences in the physical properties between asphalt binders supplied from different sources. Furthermore, sufficient amounts of these heteroatoms made the hydrocarbon molecular structures more complicated; usually, one or more heteroatoms per molecule may be present [33].

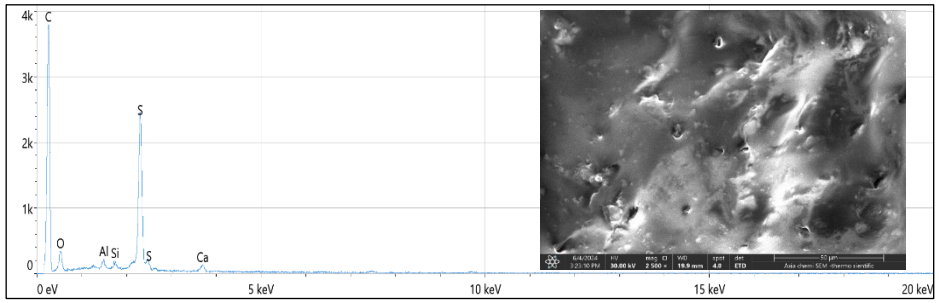


(a) Petroleum Asphalt

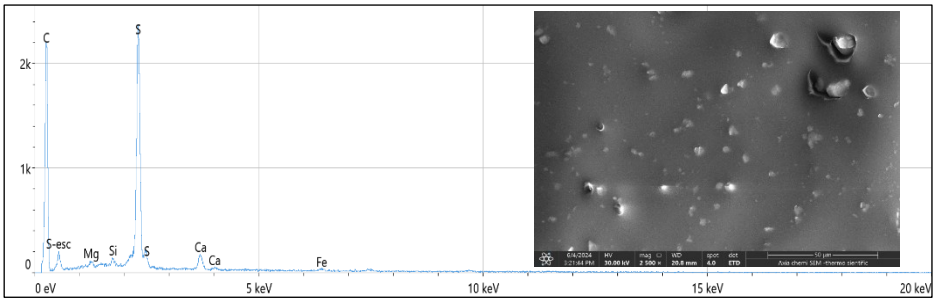


(b) MS-NB

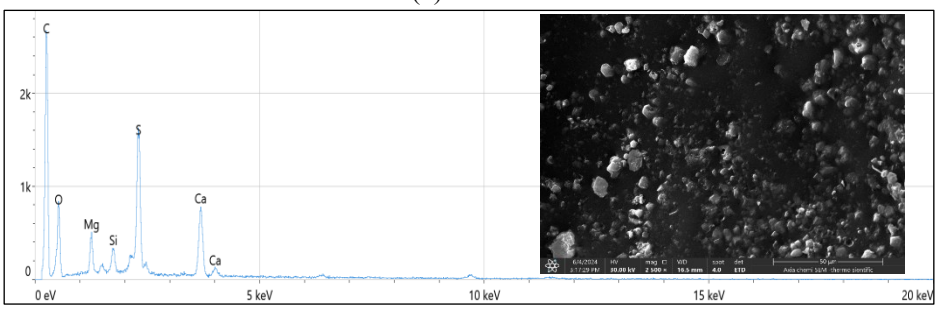
https://doi.org/10.31436/iiumej.v26i2.3452



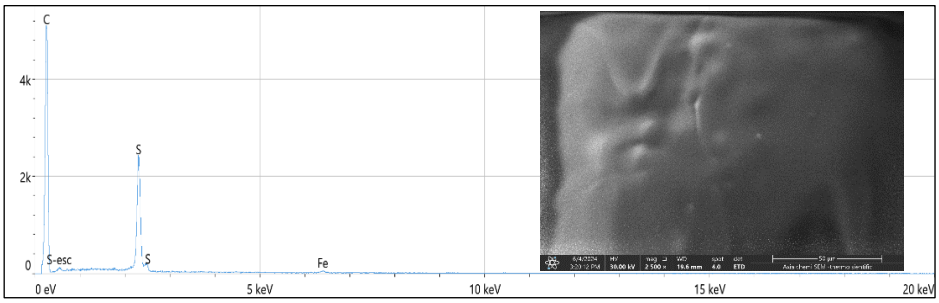
(c) JS-NB



(d) TS-NB



(e) AS-NB



(f) SS-NB

Figure 7. SEM-EDX analysis for petroleum asphalt and NB from five sulphur springs.

For each of the six NB samples, the percentage of sulphur in Table 6 varied and was not the same. AS-NB had the lowest content (6.2%), while TS-NB had the highest sulphur content (16.4 %). The high density of the sulphur elements increases the density of the binder, which, in turn, increases the service life of asphalt pavement and saves costs. For more clarification, sulphur reacts with indole to form polysulfides. It leads to a change in its colloidal structure, which leads to an increase in the proportion of asphaltenes/resins, which causes the structure to change to a more stable and cohesive gel structure, increasing the density. Additionally, the softening point rises with increased sulphur content, and this outcome aligns with the results of the previous study [34]. NB from the five springs does not contain nitrogen and nickel elements in its structure, compared to petroleum asphalt. Nitrogen increases the affinity of asphalt with aggregate, improving the resistance of the asphalt mixture against weathering,

https://doi.org/10.31436/iiumej.v26i2.3452

especially stripping and oxidation [25]. Meanwhile, nickel is one of the trace elements, and when heating the asphalt, higher concentrations of nickel can cause higher emissions of pollutants [35]. Furthermore, the presence of calcium and magnesium in NB positively affects its properties, as it increases hardness and improves stability [36].

Table 6. Elemental composition of petroleum asphalt and treated NB

Asphalt Type	Element, % *									
	C	O	S	Si	Fe	Ca	Mg	Al	N	Ni
Petroleum Asphalt	83.4	1.2	11.7	0.1	-	-	-	-	2.6	1.0
MS-NB	62.0	16.5	14.2	1.8	-	4.5	0.4	0.6	-	-
JS-NB	71.6	16.2	10.6	0.4	-	0.7	-	0.5	-	-
AS-NB	54.4	31.2	6.2	1.2	-	4.6	2.4	-	-	-
SS-NB	88.7	-	10.8	-	0.5	-	-	-	-	-
TS-NB	80.4	-	16.4	0.4	0.6	1.8	0.4	-	-	-

* Carbon (C), Oxygen (O), Sulphur (S), Silicon (Si), Iron (Fe), Calcium (Ca), Magnesium (Mg), Aluminium (Al), Nitrogen (N), and Nickel (Ni)

5.3. Marshall Properties

The OBC is 5% for a conventional mixture prepared with petroleum asphalt. After applying heat treatment to the NB samples, the OBC values are 5.2%, 4.9%, 5.3%, 4.93%, and 5.1% (by weight of the aggregate) for mixtures prepared with NB from AS, JS, MS, SS, and TS, respectively. The Marshall properties of both combinations (conventional mixture and NB mixtures) are shown in Figures 8 and 9. In the Marshall test, flow and stability are the most important performance indicators of asphalt pavement.

The results in Figure 8 showed that the stability of NB mixtures, except for the JS-NB mixture, was better than that of the conventional mixture. The results were (13, 10.8, 12.6, and 11.8) kN for mixtures with NB from MS, AS, TS, and SS, respectively. In contrast, the stability of the conventional mixture was 9.2 kN. Compared to the conventional mixtures' stability, NB mixtures' stability increased by 17.3%, 28.2%, 36.9%, and 41.3% for NB from AS, SS, TS, and MS, respectively. This might be because NB contains a high percentage of sulphur, which increases the hardness of asphalt by raising the temperature between 112 and 170 °C. Sulphur consists of chemical chains that bind with the organic molecules of the compound, thus forming polysulfates [16, 25, and 37]. However, the stability of the JS-NB mixture is 8.8 kN, which is 4.4% less than the stability of the conventional mixture. This may be because JS-NB is more affected by high oxidation rates than other NB samples. Generally, all mixtures' stability values meet the specification limits (8 kN minimum limit for wearing layer) [26].

Figure 9 shows the Marshall flow values of the conventional and NB mixtures, according to the Marshall test outcome. The flow values are close, except for the MS-NB mixture, which increases by 12.1% compared to the conventional mixture flow. According to the specification limits, the flow should range between 2 and 4 mm, and all the mixtures comply with this requirement.

https://doi.org/10.31436/iiumej.v26i2.3452

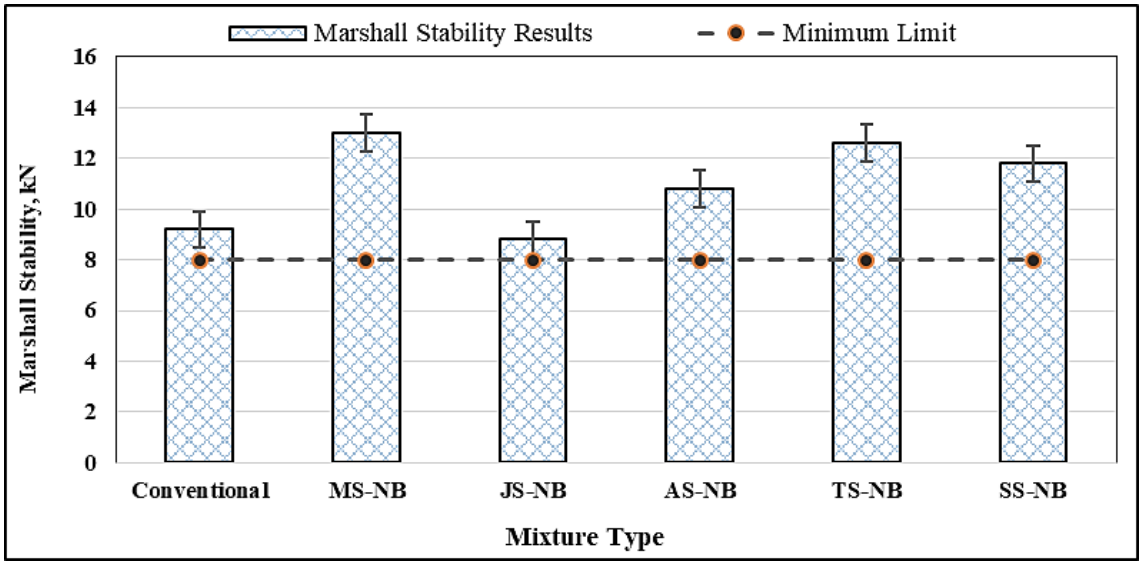


Figure 8. Marshall stability results.

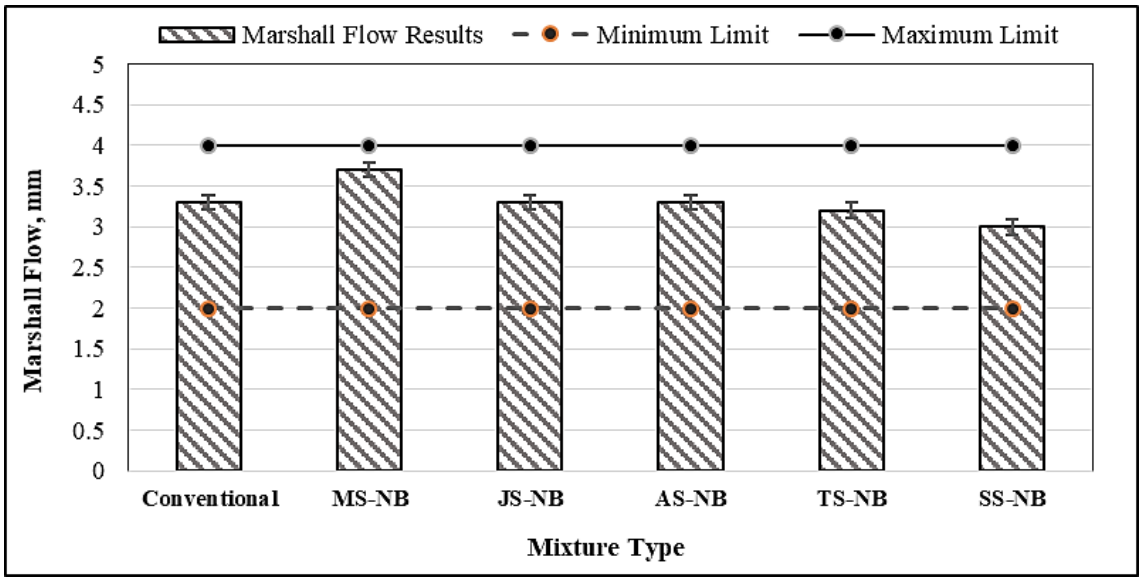


Figure 9. Marshall flow results.

Figure 10 shows conventional and NB mixtures' Marshall Quotient (MQ), or stiffness index. MQ refers to the proportion of Marshall stability to Marshall flow and is considered an indicator of the asphalt mixture's rutting resistance. Higher MQ values indicate better stability and more resistance to deformation. Therefore, an increase in MQ means the stability of the asphalt mixture has improved regarding its flow behavior. This is desirable for the durability of high-quality asphalt pavements [38].

The volumetric properties for conventional and NB mixtures are listed in Table 7. Based on local specifications [26], the air void (Vv) should be between three and five percent. An increase in voids leads to rutting of the road, rapid oxidation of the asphalt, and higher permeability, which leads to problems that shorten the road's lifespan. A decrease in the percentage of voids can also lead to creep and eruption of asphalt [39]. The Vv % in conventional and NB mixtures are similar, and both fall within the limits of the local specification.

https://doi.org/10.31436/iiumej.v26i2.3452

The performance of a mixture depends on the voids in mineral aggregate (VMA), which should be sufficiently high to ensure adequate bitumen content to meet durability requirements. However, as VMA increases, the mixture becomes more susceptible to stability issues. The results indicated that the VMA values for NB mixtures slightly differ from those of conventional mixtures.

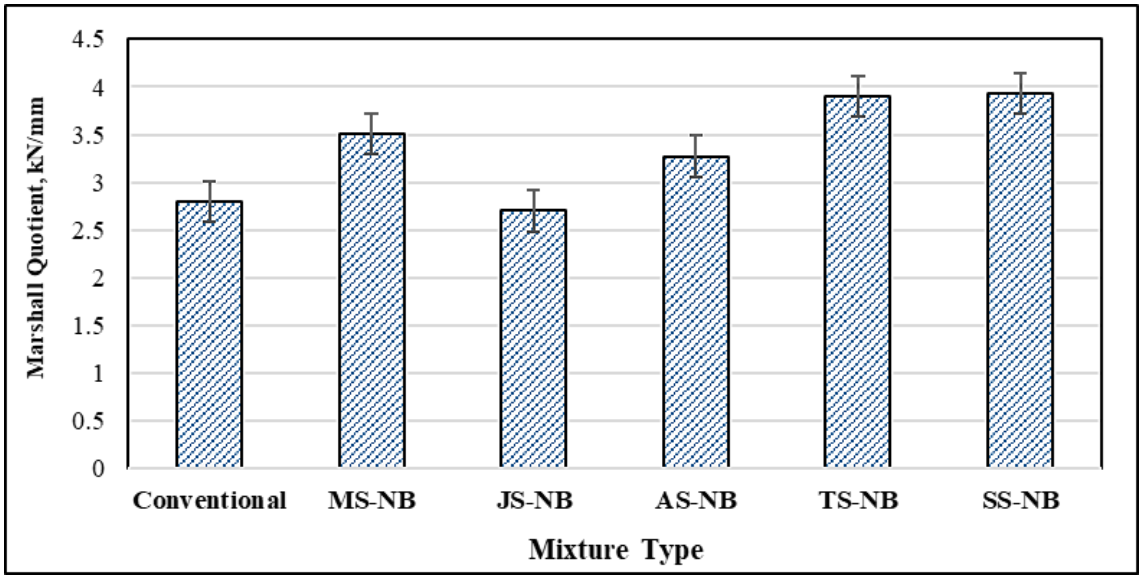


Figure 10. Marshall quotient results.

Table 7. Volumetric properties for conventional and NB mixtures

Property		Bulk Density g/cm^3	Vv (%)	VMA (%)	VFB (%)
Conventional Mixture		2.323	3.8	16.4	76.8
NB-Mixture	MS	2.340	3.6	16.1	77.6
	JS	2.330	4.0	16	75
	AS	2.325	3.5	16.5	78
	TS	2.335	3.4	16.1	78.9
	SS	2.332	3.6	16.1	77.6
Specification Requirements [26]		-	3-5	14 Minimum	-

The voids filled with binder (VFB) refer to the distances among aggregate particles in a compacted mixture filled with bitumen. These voids impact the durability of the mixtures. The findings reveal that VFB percentages for conventional and NB mixtures are either similar to or lower than conventional values, except for TS-NB, which showed a 2.7% increase compared to the conventional mixtures.

5.4. Moisture Damage Resistance

As a result, it is necessary to recognize the harmful effects of water and humidity and take measures to mitigate them, while also ensuring the mixture's hardness, durability, and resilience under various conditions [40]. TSR can be defined as a coefficient of moisture durability index, and its value should be at least 80%. TSR depends on the adhesion between the aggregate surface and bitumen and the cohesion properties of the asphalt binder. Cohesion refers to the overall integrity of the material when subjected to load and stress, and it is influenced by several factors, including the viscosity and chemical components of the asphalt binder, attraction inside the asphalt binder, and water penetration. Water can affect cohesion by saturating and expanding the system of voids due to freezing and thawing cycles. Asphalt adhesion to water can be affected by separation or displacement mechanisms.

https://doi.org/10.31436/iiumej.v26i2.3452

Adhesion depends on the aggregate gradation, Vv, and permeability. Adhesion under moisture circumstances, adhesion may vary if any of these parameters change [41].

Figure 11 displays the TSR values for conventional and NB mixtures. The test findings show that the NB mixture had the highest resistance to moisture and water conditions. The mixture with SS-NB exhibited 5.72% greater resistance to moisture damage than the conventional mixture. This may be attributed to the large amount of carbon in SS-NB, since the high content of carbon elements increases the adhesion properties of the asphalt binder, which increases durability and provides better resistance to moisture damage. Specifically, carbon plays a crucial role. It is present in high concentrations and forms chains that help create strong bonds within the NB structure, increasing the stability of the asphalt and its ability to adhere to the aggregate. Furthermore, asphalt with high carbon content is an absorbent material, which helps trap moisture and enhances the performance and moisture resistance of the asphalt [32].

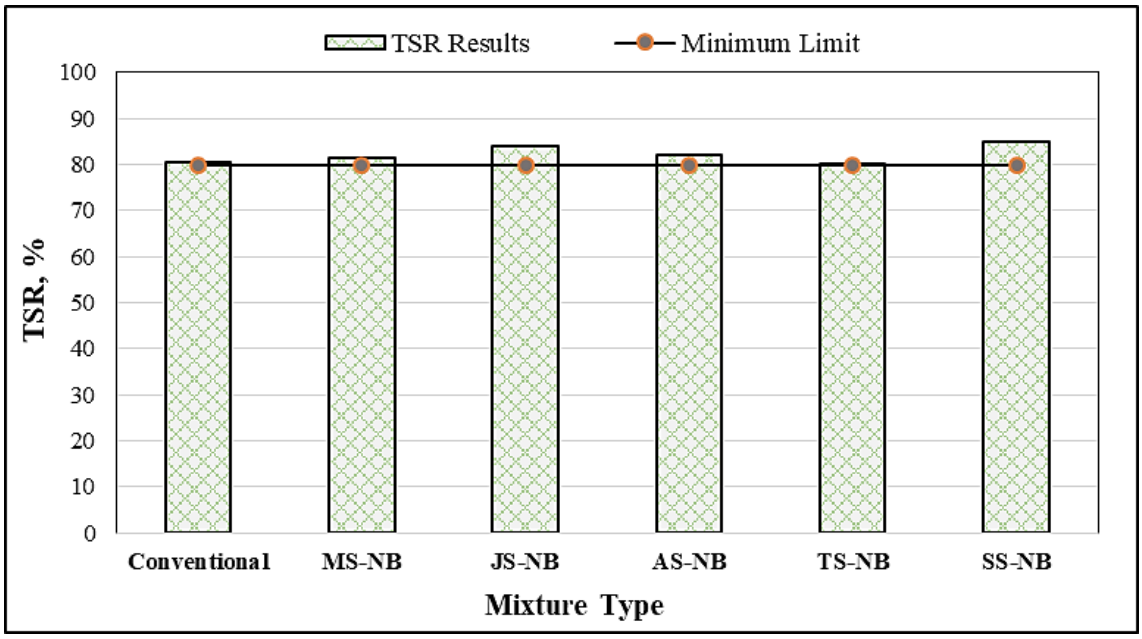


Figure 11. TSR for conventional and NB mixtures.

6. COST ANALYSIS

Developing innovative processes and product models that simultaneously deliver economic, technical, and environmental benefits is essential to ensure that policies effectively target sustainable development goals. Environmentally friendly and more economical alternatives must be discovered to achieve sustainable development. NB is considered one of the natural resources available in abundance in Iraq, and it can be exploited across various governorates, from north to south. Economic gains can be achieved by investing in NB or incorporating it into road paving operations. This is because NB is more economical than petroleum asphalt, which is derived from the distillation processes of crude oil, and it is a national resource. Oil, a major energy source, represents over 30% of global consumption. Countries such as Venezuela, Saudi Arabia, Canada, Iran, Iraq, and Russia possess large proven oil reserves. During the distillation process in oil refineries, around 10% of crude oil is converted into asphalt materials and other products through asphalt manufacture. This asphalt produces various retail products, including asphalt emulsions, paving mixtures, paints, tar roofing, and other products [42].

Due to the continuous change in oil prices over the past six decades, which affects asphalt prices, asphalt costs have increased by a factor of 20 due to oil refining operations. Asphalt prices follow crude oil prices; consequently, the quantity and quality of asphalt are impacted. For every 1% increase in crude oil, asphalt prices rise by 0.7%. The decisions made by the Organization of the Petroleum Exporting Countries (OPEC), which consists of 12 countries, have played an important role in determining the cost of asphalt [25, 43]. Some specialists have stated that the oil peak occurred between 2010 and 2020 and will decrease over the coming years due to the decrease in production quantities, the occurrence of a crisis at the global level, and a significant rise in prices. Meanwhile, geologist Marion King Hubbert, in the 1950s, predicted that the global productivity peak would occur between 2007 and 2037, according to estimates from primary energy sources, after which production would decline to less than a third of its previous levels. Therefore, the price of oil has a significant impact on asphalt production and its prices, prompting road engineers to study alternative synthetic asphalt materials to use instead that are more economical and environmentally friendly [44]. The environment is given serious attention during the asphalt production process. There must be strict regulations on emissions from asphalt producers and oil refineries. Reducing carbon emissions during the transportation and production of asphalt pavement has become a global priority. Finding more environmentally friendly natural materials is essential to managing and minimizing greenhouse gas emissions as much as possible. The use of low-carbon technology has become vital to achieving this goal. For this reason, there is an urgent need for future investigations to study the carbon emissions rate and the economic benefits of using NB in paving works.

7. CONCLUSIONS

This study conducted a series of experiments on natural bitumen (NB) derived from sulphur springs, petroleum asphalt, and their mixtures to evaluate the suitability of using NB in hot asphalt mixtures. Physical tests revealed that untreated NB does not meet the local Standard Specification limits, making it unsuitable for direct use in paving. However, when NB is properly dried and heat-treated at 163 °C for a duration dependent on the characteristics of each spring and the penetration level, its properties significantly improve. Overheating beyond 163 °C is discouraged due to potential degradation of performance. Post-treatment physical and chemical analyses showed enhanced characteristics, including reduced penetration, increased ductility and softening point, and improved flash point, bringing NB's properties closer to those of petroleum asphalt. Morphological observations highlighted that the surface microstructure of petroleum asphalt closely resembles that of SS-NB, differing notably from other NB types due to variations in elemental composition. Treated NB demonstrated higher stability than petroleum asphalt, likely due to its elevated sulphur content, which enhances mixture hardness and durability, with calcium and oxygen contributing further to performance. Nonetheless, the JS-NB mixture did not significantly improve Marshall stability due to its high oxidation rate, but did show a slight enhancement in water damage resistance with a 4.35% increase. Notably, the SS-NB mixture achieved the highest tensile strength ratio (TSR) at 85%, surpassing the conventional mixture by 5.7% and exceeding the 80% minimum threshold, indicating NB's strong potential in mitigating moisture damage. In conclusion, NB is a valuable and promising alternative for improving asphalt mixture performance. Economically, it is highly advantageous, being widely available nationwide and costing at least five times less than petroleum asphalt. As a national asset, NB warrants further economic and engineering research to optimize its use in infrastructure applications.

8. RECOMMENDATIONS

Future studies in several key areas may provide deeper insights and a more comprehensive understanding of the current findings. Further investigation is needed to evaluate the intrinsic properties of natural bitumen (NB), including its asphaltene origin, extraction process, and influencing factors such as sediment age, extraction depth, and proximity to sulphur sources. Environmental variables—such as rainfall, temperature fluctuations, and the geographical location of the spring—may also significantly impact NB's characteristics. Additional studies are encouraged to explore alternative treatment methods, such as enhancing NB with inexpensive, readily available materials like fillers or utilizing solar energy by exposing NB to sunlight over extended periods and periodically testing its properties for changes. Furthermore, it is crucial to assess the rutting resistance of NB mixtures under repeated loading and elevated temperatures and evaluate their fatigue cracking behavior to ensure performance under real-world conditions. From a sustainability perspective, economic and environmental assessments of NB usage should be conducted to determine its cost-effectiveness and environmental impact as a substitute for petroleum asphalt, including strategies to minimize emissions associated with its application.

REFERENCES

[1] Ng CP, Law TH, Jakarni FM, Kulanthayan S. (2019) Road infrastructure development and economic growth. IOP Conference Series: Materials Science and Engineering, 512(1):12045. https://iopscience.iop.org/article/10.1088/1757-899X/512/1/012045.

[2] Latief RH. (2019) Evaluation of the performance of glasphalt concrete mixtures for binder course. International Journal on Advanced Science, Engineering and Information Technology, 9(4): 1251-1259. https://doi.org/10.18517/ijaseit.9.4.5858.

[3] Chen J, Dan H, Ding Y, Gao Y, Guo M, Guo S, et al. (2021) New innovations in pavement materials and engineering: A review on pavement engineering. Journal of Traffic and Transportation Engineering, 8(6): 815–999. https://doi.org/10.1016/j.jtte.2021.10.001.

[4] Kołodziej K, Bichajło L, Siwowski T. (2021) Effects of aging on the physical and rheological properties of trinidad lake asphalt modified bitumen. Journal of Materials,14(10): 2532. https://doi.org/10.3390/ma14102532.

[5] Al-Azawee ET, Latief RH. (2020) The feasibility of using styrene-butadiene-styrene (SBS) as modifier in Iraqi bituminous binder. Journal of Engineering Science and Technology, 15(3): 1596–1607.

[6] Petrov SM, Kayukova GP, Goncharova IN, Safiulina AG, Lakhova AI. (2018) High-quality asphalt binders produced by deasphalting of natural bitumen. Section Oil and Gas Exploration, 14(1.4): 455-460. https://doi.org/10.5593/sgem2018/1.4/S06.060.

[7] Zhang XF, Zeng Y, Feng YN, Zhang CX, Zhang L. (2023) Carbon emissions analysis of producing modified asphalt with natural asphalt. Green Processing and Synthesis, 12(1): 20228146. https://doi.org/10.1515/gps-2022-8146.

[8] Khusnutdinov I, Goncharova I, Safiulina A. (2021) Extractive deasphalting as a method of obtaining asphalt binders and low-viscosity deasphalted hydrocarbon feedstock from natural bitumen. Egyptian Journal of Petroleum, 30(2): 69-73. https://doi.org/10.1016/j.ejpe.2021.03.002.

[9] Anupam K, Akinmade D, Kasbergen C, Erkens S, Adebiyi F. (2023) A state-of-the-art review of Natural bitumen in pavement: Underlining challenges and the way forward. Journal of Cleaner Production, 382(8): 134957. https://doi.org/10.1016/j.jclepro.2022.134957.

[10] Kök BV, Yílmaz M, Turgut P, Kulog N. (2012) Evaluation of the mechanical properties of natural asphalt-modified hot mixture. International Journal of Materials Research, 103(4): 506-512. https://doi.org/10.3139/146.110654.

[11] Nalbandian KM, Carpio M, González A. (2021) Analysis of the scientific evolution of self-healing asphalt pavements. Journal of Cleaner Production, 293: 126107. https://doi.org/10.1016/j.jclepro.2021.126107.

[12] Mahmood NS. (2011) Sand dunes modification by natural asphalt with cement and lime. Muthanna Journal of Engineering and Technology, 1(1): 66-75. https://muthjet.mu.edu.iq/wp-content/uploads/2022/08/Sand-dunes-modification-by-natural-asphalt-with-cement-and-lime.pdf.

[13] Bilski M. (2018) Natural asphalts – properties and use. Archives of Institute of Civil Engineering, 27(27): 25-36. DOI:10.21008/j.1897-4007.2018.27.02.

[14] Al Dulaymie AS, Hussien B, Gharbi MA, Mekhlif HN. (2013) Balneological study based on the hydrogeochemical aspects of the sulfate springs water (Hit-Kubaiysa region), Iraq. Arabian Journal of Geosciences, 6(3): 801-816. https://doi.org/10.1007/s12517-011-0385-5.

[15] Sobhi S, Yousefi A, Behnood A. (2020) The effects of Gilsonite and Sasobit on the mechanical properties and durability of asphalt mixtures. Construction and Building Materials, 30(238): 117676. https://doi.org/10.1016/j.conbuildmat.2019.117676.

[16] Mohammed FH, Latief RH, Albayati AH. (2024) Assessment of traditional asphalt mixture performance using natural asphalt from sulfur springs. Journal of Engineering, 30(1): 54-73. https://doi.org/10.31026/j.eng.2024.01.04.

[17] Mahmood BA. (2021) Influence of waste engine oil addition on the properties of natural asphalt. In Proceedings of the Second International Virtual Conference on Pure Science: 21-22 April 2020; Diwaniyah. pp 1-11. https://dx.doi.org/10.1088/1742-6596/1999/1/011001.

[18] Koshkarov V, Shunyaev I, Koshkarov E. (2023) Characteristics evaluation of modern bituminous binders in group hydrocarbon composition and their application in construction. Materials Sciences and Applications, 14(8): 407-415. https://doi.org/10.4236/msa.2023.148026.

[19] Shlimon AG, Mansurbeg H , Othman RS, Gittel A, Aitken CM , Head IM, et al. (2020) Microbial community composition in crude oils and asphalts from the kurdistan region of Iraq. Geomicrobiol Journal, 37(7): 635-652. https://doi.org/10.1080/01490451.2020.1753131.

[20] Farhan MM, Rabeea MA, Muslim RF, Zidan TA. (2021) Chemical composition (saturate fraction) of western Iraq natural bitumen. Journal of Materials, 42(5): 2527–2533. https://doi.org/10.1016/j.matpr.2020.12.574.

[21] Muttar AA, Zedan TA, Mahmood BA. (2009) Analytical comparison study for asphalt and water of Heet sulphurous springs. Journal of University of Anbar for Pure Science, 3(1): 45-56. https://doi.org/10.37652/juaps.2009.15453.

[22] Abdul-Jaleel T, Najres M. (2012) Oxidation influence on the chemical content and some rheological properties for natural asphalt to Abu Al Jeer-Al Anbar. Iraqi Journal of Desert Studies, 4(1): 1-15. DOI:10.36531/ijds.2012.68137.

[23] Abdul-Jaleel T, Salem JW, Najres AM. (2016) A new separation, fractionation and improving of Abu- Aljeer asphalt. Anbar Journal of Engineering Science, 7(1): 31-41. https://doi.org/10.37649/aengs.2016.124363.

[24] Mahmood BA, Al-Ani Y, Mohammed KA, Eyada SO. (2018) The feasibility of using plastic wastes to improve the properties of natural asphalt. Journal of Engineering and Applied Science, 13(21): 8929–8934. https://doi.org/10.3923/jeasci.2018.8929.8934.

[25] Ahmed F, Latief R, Albayati A. (2024) Effect of natural bitumen on the performance of hot asphalt mixture. Gazi University Journal of Science, 37(3): 1154–1170. https://doi.org/10.35378/gujs.1282239.

[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] Altameemi AA, Al-Jumaili MA, Mohammed AJ, Abdul karim A. (2016) Studying the possibility of utilizing the vast deposits of natural asphalt in Iraq in specific uses using as little as possible manufacturing treatment. The Iraqi Journal for Mechanical and Material Engineering, 1(1): 462-488.

[28] Tosun HB, Alver A, Baştürk E. (2022) Removal of exhaust gas with advanced solar photocatalytic asphalt applications. KSCE Journal of Civil Engineering, 26(1): 13-24. https://doi.org/10.1007/s12205-021-0654-0.

[29] Mithil M, Raju A, Ahmed W, Soon J. (2018) SEM and ESEM techniques used for analysis of asphalt binder and mixture. Journal of Construction and Building Materials, 186(1): 313-329. https://doi.org/10.1016/j.conbuildmat.2018.07.126.

[30] Liang T, Wang J, Kou X, Peng P. (2024) Influence mechanism of original components on mechanical properties and transformation behaviors of natural bitumen. Fuel, 371 (Part A): 131924. https://doi.org/10.1016/j.fuel.2024.131924.

[31] Zhang M, Hao P, Dong S, Li Y, Yuan G. (2020) Asphalt binder micro-characterization and testing approaches: A review. Measurement, 151: 107255. https://doi.org/10.1016/j.measurement.2019.107255.

[32] Yamaguchi K, Sasaki I, Nishizaki I, Meiarashi S, Moriyoshi A. (2005) Reinforcing effects of carbon black on asphalt binder for pavement. Journal of the Japan Petroleum Institute, 48(6): 373−379. https://doi.org/10.1627/jpi.48.373.

[33] Yao H, Liu J, Xu M, Bick A, Xu Q, Zhang J. (2021) Generation and properties of the new asphalt binder model using molecular dynamics (MD). Scientific Reports, 11: 9890. https://doi.org/10.1038/s41598-021-89339-5

[34] Khedaywi T, Haddad M, Mujalli R, Shareef S. (2023) Effect of sulfur on the asphalt cement and asphalt concrete mixture. Innovative Infrastructure Solutions, 8(11): 286. https://doi.org/10.1007/s41062-023-01264-w.

[35] Awadh SM, Al-Mimar HS. (2023) The effect of nickel, vanadium, asphaltene, NSO and sulfur on crude oil quality. Iraqi Geological Journal, 56(2B): 137-144. https://doi.org/10.46717/igj.56.2B.10ms-2023-8-19.

[36] Yuan Y, Chen H, Wang Y, Xu S, Xue B, Zou G, Wang B. (2023) How calcium and magnesium ions affect the stability of asphalt with anionic emulsifiers: A molecular dynamics study. Construction and Building Materials, 12(10): 132770. https://doi.org/10.1016/j.conbuildmat.2023.132770.

[37] Sakib N, Bhasin A, Islam MK, Khan K, Khan MI. (2021) A review of the evolution of technologies to use sulphur as a pavement construction material. International Journal of Pavement Engineering, 22(3): 392–403. https://doi.org/10.1080/10298436.2019.1612064.

[38] Sanja D, Miroslav S, Martina Z, Tatjana R. (2021) Laboratory evaluation of the properties of asphalt mixture with wood ash filler. Journal of Materials, 14(3): 575. https://doi.org/10.3390/ma14030575.

[39] Abd NI, Latief RH. (2024) Assessment of rutting resistance for fiber-modified asphalt mixtures. Journal of Engineering, 30(5): 98–113. https://doi.org/10.31026/j.eng.2024.05.07.

[40] Omar HA,Yusoff NI, Mubaraki M, Ceylan H. (2020) Effects of moisture damage on asphalt mixtures. Journal of Traffic and Transportation Engineering, 7(5): 600–628. https://doi.org/10.1016/j.jtte.2020.07.001.

[41] Al-Saadi AA, Ismael MQ. (2023) Improvement of moisture susceptibility for asphalt mixture with ceramic fiber. Journal of Engineering, 29(4): 78–91. https://doi.org/10.31026/j.eng.2023.04.05.

[42] Yildiz I. (2018) Chapter 1.12 Fossil Fuels. In Comprehensive Energy Systems, 521-567. https://doi.org/10.1016/B978-0-12-809597-3.00111-5.

[43] Kaylyn MC, Khalafalla M, Benavides JR. (2021) Protocol to assess the impact of crude oil price fluctuations on future asphalt prices. Journal of the Transportation Research Board, 2675(6): 294-305. https://doi.org/10.1177/0361198121992072.

[44] U.S. Bureau of Labor Statistics, U.S. Department of Labor. Website Available: https://www.bls.gov/data/.

https://doi.org/10.31436/iiumej.v26i2.2988

CRACK TRACKING IN SMALL-DIAMETER METAL PIPE USING CONTROLLED MOTOR VIBRATIONS AND FLEXIBLE SENSOR

MD RABIUL AWAL[1*], NURUL ATIQAH BINTI TAJULMAR[1], MUHAMMAD SYARIFUDDIN YAHYA[1], NURAFNIDA AFRIZAL[1], WAN HAFIZA WAN HASSAN[1], NURUL ADILAH ABDUL LATIFF[1], SHAKIR SAAT[2]

[1]Faculty of Ocean Engineering Technology, Universiti Malaysia Terengganu (UMT), Kuala Nerus 21300, Terengganu, Malaysia
[2]Faculty of Electronics and Computer Technology and Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

**Corresponding author: rabiulawal1@gmail.com*

(Received: 12 February 2024; Accepted: 7 November 2024; Published online:15 May 2025)

ABSTRACT: Metal pipes are the most integral part of transporting water, gas, and other petrochemical substances over long distances. Higher strength, durability (along with wear and corrosion resistance), and lower cost make these pipes suitable for extreme weather conditions and hostile environments. Over time, these pipes experience significant impacts that may lead to defects such as holes, cracks, bends, corrosion, and finally component failure and property losses. Therefore, early detection of the defects in pipes is crucial to prevent such failures. There are several methods to detect defects in metal pipes, including non-destructive testing (NDT). However, high costs and declining performance are existing concerns for those NDTs. A motor-induced vibration source is more robust and reliable than a conventional vibration sensor. Thus, the feasibility of using a motor as a vibration source for metal pipe crack detection is studied in this work. To achieve this, a DC motor is placed on one side of the metal pipe and used as the vibration source. These vibrations are collected by a piezoelectric polymer, specifically a Polyvinylidene fluoride (PVDF) sensor, on the other side of the pipe. This work considers three types of pipe conditions: healthy pipe, bent pipe, and cracked pipe. Additionally, two different sensor locations (180-degree rotation) and sensor patterns (bent and not bent) are studied. From the studies, we can see that there are significant differences in pressure responses for healthy pipe and cracked pipe conditions. The maximum pressure response for a cracked pipe is 783 a.u. (intensity) whereas it is just 262 a.u. for a healthy pipe. Thus, the difference is sufficient to set a threshold margin. We have set 300 a.u. as the threshold margin and applied it to an algorithm. The algorithm can successfully detect a healthy or cracked pipe. However, it is very tricky in the case of a bent pipe, as the pressure differences are less than 300 a.u. for three conditions and above for only one. Hence, it might provoke an incorrect decision when detecting a bent pipe.

ABSTRAK: Paip logam adalah bahagian utama dalam mengangkut air, gas, dan bahan petrokimia lain dalam jarak jauh. Kekuatan dan ketahanan tinggi (bersama rintangan hakisan dan penggunaan), dan kos lebih rendah menjadikan paip logam sesuai bagi keadaan cuaca dan persekitaran melampau. Walau bagaimanapun, dari masa ke masa, paip logam mengalami kesan ketara seperti berlubang, retak, bengkok, hakisan dan akhirnya kegagalan komponen dan kehilangan harta benda. Oleh itu, pengesanan awal kecacatan pada paip adalah sangat penting bagi mengelakkan kegagalan tersebut. Terdapat kaedah tidak merosakkan (NDT) bagi mengesan kecacatan pada paip logam. Walau bagaimanapun, kos yang tinggi dan prestasi merosot adalah kebimbangan sedia ada pada NDT. Sumber getaran dari motor adalah lebih berdaya tahan dan lebih dipercayai berbanding pengesan getaran konvensional. Oleh itu,

https://doi.org/10.31436/iiumej.v26i2.2988

kebolehlaksanaan motor sebagai sumber getaran bagi mengesan paip logam yang retak dikaji dalam kajian ini. Bagi tujuan ini, motor DC diletakkan pada satu sisi paip logam dan digunakan sebagai sumber getaran. Getaran ini dikumpul oleh pengesan polimer piezoelektrik Poliviniliden Fluorida (PVDF) pada bahagian lain paip. Tiga jenis keadaan paip, iaitu paip sihat, paip bengkok dan paip retak dipertimbangkan dalam kajian ini. Tambahan, dua lokasi pengesan berbeza (pada putaran 180 darjah) dan corak pengesan (bengkok dan tidak bengkok) dikaji. Dapatan kajian menunjukkan terdapat perbezaan ketara dalam tindak balas tekanan bagi paip berkeadaan sihat dan retak. Malah, tindak balas tekanan maksimum bagi paip retak adalah 783 a.u. (intensiti) sedangkan hanya 262 a.u. bagi paip sihat. Oleh itu, perbezaan ini cukup bagi menetapkan margin ambang. Kajian ini telah menetapkan 300 a.u. sebagai margin ambang dan menggunakannya pada algoritma. Algoritma ini berjaya mengesan paip sihat atau retak. Walau bagaimanapun, adalah sangat rumit bagi mengesan paip bengkok kerana perbezaan tekanan adalah kurang daripada 300 a.u. bagi tiga syarat tersebut. Oleh itu, ia mungkin mencetuskan keputusan yang salah bagi mengesan paip bengkok.

KEYWORDS: *Guided wave; Pipe inspection; Metallic pipe structures; Non-destructive evaluation (NDE); Remaining useful life (RUL)*

1. INTRODUCTION

Pipelines are very crucial for resources like oil, gas, and water. Extracting these from underground to transporting them to the consumer ends requires a steady and reliable medium, and pipelines are the best solution. Resources can easily be transported over long distances or between countries using pipelines [1,2]. However, despite the improved quality and reliability, pipes are not meant to last a lifetime. Hence, quality and reliability both will degrade over time or due to influence. This can cause accidents with serious consequences and can cause fatalities, injuries, economic losses, and environmental damages [3]. According to the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA) data, during 2002-2021, 680 pipeline incidents were recorded in the USA, with 260 fatalities, 1109 injuries, and over a billion dollars in damages [4]. Major reasons for these incidents are external interferences, corrosion, construction defects/material failures, hot taps, ground movements (earthquakes, landslides), and deformations [5].

Therefore, inspecting and monitoring the pipeline to maintain structural integrity is crucial. Various solutions are available to detect the unusual structural deformations of the pipes. However, sensor-based approaches have become the most convenient choice, along with magnetic flux leakage (MFL) testing, ultrasonic testing (UT), electromagnetic acoustic technology (EMAT), eddy current testing (EC), Electric Field Mapping (EFM), Eddy Current Inspections, sonar mapping, and guided waves [6-9]. These sensors are already available in the market and ready for use. Nevertheless, their functions are complex and multi-directional. Moreover, their larger geometry makes longer pipes necessary, impacting lab-based quality enhancements. On the contrary, we aim to investigate a more cost-effective and simpler solution.

Guided waves are a widely used technique for nondestructive testing (NDT). In this technique, waves produced by a vibration source are guided through the pipe to be inspected, and a sensor collects the waves. The pattern of the received waves can define the deformity in the pipe. These vibration sources (technically transducers) usually experience a decline in performance over time and during their life cycle. These factors make their performance unpredictable in the long run, especially when consistently influenced by temperature, humidity, vibration, and shock [10,11]. Declined performance can lead to incorrect results and, consequently, wrong recommendations [12]. Consequently, this makes them unsuitable as a

https://doi.org/10.31436/iiumej.v26i2.2988

vibration source for a sensory system, where the troubleshooting recommendation depends entirely on the received results.

A motor, on the other hand, is more reliable compared to a conventional vibration sensor. A motor's output, represented as Revolutions Per Minute (RPM), is fixed based on the input rating. Therefore, applying a specific voltage can only achieve a specific RPM. The combination of the motor's produced torque and RPM will create vibrations due to its rotation. These vibrations can be propagated to an attached surface. Hence, the motor acts as a vibration source. Motivated by this, we intend to investigate the feasibility of guided waves from motor vibrations. Therefore, we propose using a DC motor as a wave source. These motor vibrations can be guided through a metal pipe from one end and collected by a vibration sensor on the other. We have used a flexible piezoelectric polymer, Polyvinylidene fluoride (PVDF), as a sensor to collect the vibrations from the DC motor. In this work, we will focus on using guided waves to detect a crack in a metal pipe. To achieve this, we consider two pipe conditions: a healthy pipe and a deformed or cracked pipe. The characteristics of the received waves will indicate any deformations on the surface of the pipe.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 describes the experimental design, while Section 4 presents the results and evaluation. Lastly, Sections 5 and 6 conclude this paper with some prospective future agendas and the limitations of the work.

2. RELATED WORKS

The best way to monitor pipelines is to use cost-effective and less invasive screening inspections that could provide a more global perspective of the pipeline while also suggesting which areas may require additional attention [13]. One very simplistic example of said screening inspections is sensors. Sensor technologies can include instruments that cover a wide range of physical principles, including electrical, optical, radiographic, chemical, and acoustic domains.

2.1. Magnetic Flux Leakage (MFL)

MFL inspection starts by saturating a magnetic field with a metallic surface, such as the pipe. Defects on the pipe's surface will disrupt the magnetic field's flux, leading to an aberration in the field outside the pipe surface. The aberration will indicate a pipe leakage, which is measured by the Hall effect sensor [13].

2.2. Electric Field Mapping (EFM)

The EFM method uses two electrodes in contact with the pipe. Current will pass through the pipe between the two electrodes, and the voltage drops between the two electrodes are measured. Any anomalies on the surface of the pipe, like cracks and corrosion, will alter that area of the pipe and thus the measured potential drop within the two electrodes [14, 15, 16].

2.3. Eddy Current Inspections

Eddy currents are structured electric currents that develop in a conductor due to changes in the magnetic field. During eddy current inspections, a magnetic field can be passed through the pipe, penetrating its surface, inducing current, and generating eddy currents on its surface. Any cracks and anomalies on the pipe's surface will be identified by the disturbance in the formation of eddy currents on the surface [17, 18].

https://doi.org/10.31436/iiumej.v26i2.2988

2.4. Sonar Mapping

Sonar technology uses sound waves underwater to detect objects, aid navigation, and facilitate mapping. Multi-beam bathymetry and Side Scan Sonar (SSS) are commonly preferred methods for subsea pipeline inspections. By analyzing the seabed with sound waves, one can detect the return of the sonar pulse, the sea depth, and the position of anomalies by examining the amplitude of the sound wave reflections [19-21].

Moreover, guided waves benefit pipe inspections for various reasons [22-25]. This can be attributed to their long propagation and high sensitivity, but primarily, the interaction between guided waves and materials can effectively identify cracks in a pipe. Guided waves are particularly utilized in situations requiring high-sensitivity detection techniques. This includes microscale damage such as the initiation of fatigue cracks, early-stage corrosion, and material degradation. The excitation wavelength constrains the guided wave inspection technique, which can only detect significant damage.

3. SYSTEM DESIGN AND WORKING METHOD

Flexible sensors typically exhibit a broad range of pressure responses. Depending on the product definition and material volume, these responses range from very low to very high. A monitoring platform can also detect these responses. We chose the Arduino-based temporary monitoring system for its flexibility. Nevertheless, the limitation with Arduino is its inability to detect piezoelectric reactions in terms of voltages. Instead, it presents the responses in terms of sensor values. The highest value indicates the maximum voltage generated by the sensor. These sensor values are adjustable. We divided the total sensor values into 1024 units for our system. This implies that 1024 units will represent the highest pressure effect as the maximum sensor value. The fluctuation in these values will help identify the patterns for the pipe crack conditions.

Table 1. Experimental Setups Based on Pipe Conditions

Pipe condition	Sensor location	Sensor pattern
Healthy pipe	Top	Not bent
	Top	Bent
	Bottom (+180 deg)	Not bent
	Bottom (+180 deg)	Not bent
Bent pipe	Top	Not bent
	Top	Bent
	Bottom (+180 deg)	Not bent
	Bottom (+180 deg)	Not bent
Cracked pipe	Top	Not bent
	Top	Bent
	Bottom (+180 deg)	Not bent
	Bottom (+180 deg)	Not bent

Table 2. Mechanical properties of the pipe

Mechanical Properties	Metric
Ultimate Tensile Strength	310 MPa
Tensile Yield Strength	276 MPa
Modulus of Elasticity	68.9 GPa
Poisson's Ratio	0.33
Fatigue Strength	96.5 MPa
Shear Modulus	26 GPa

https://doi.org/10.31436/iiumej.v26i2.2988

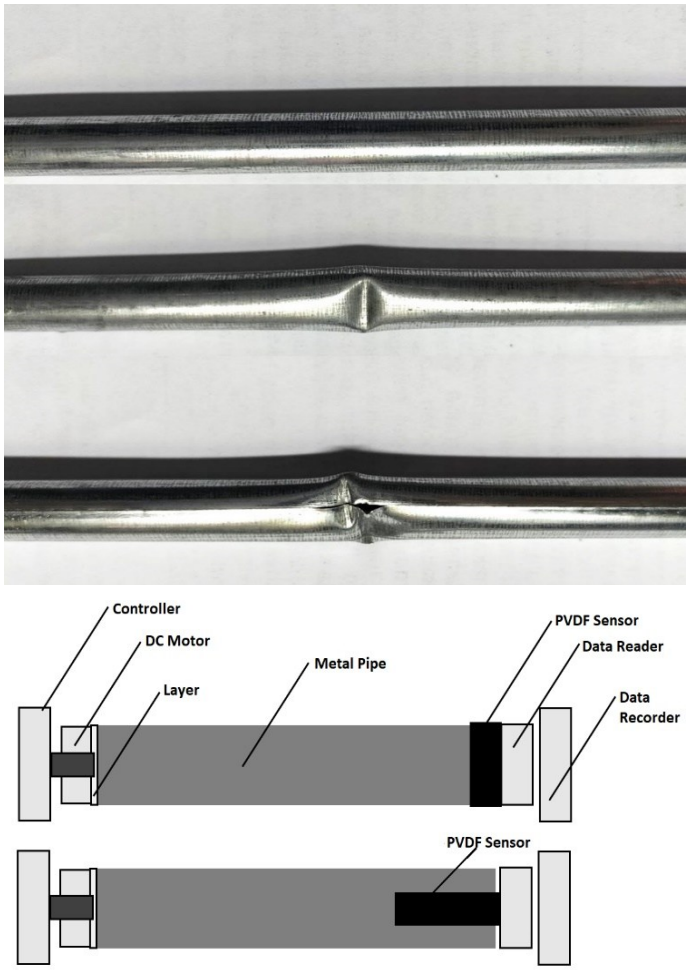


Figure 1. Considered cases of metal pipes. (a) a healthy pipe without any damage, (b) a bent pipe without an existing air hole, and (c) a cracked pipe with an existing air hole.

These fluctuations need to be stored as well for the sake of investigation. The data can be stored using Arduino since it connects to a data reader. Thus, the data is stored in memory.

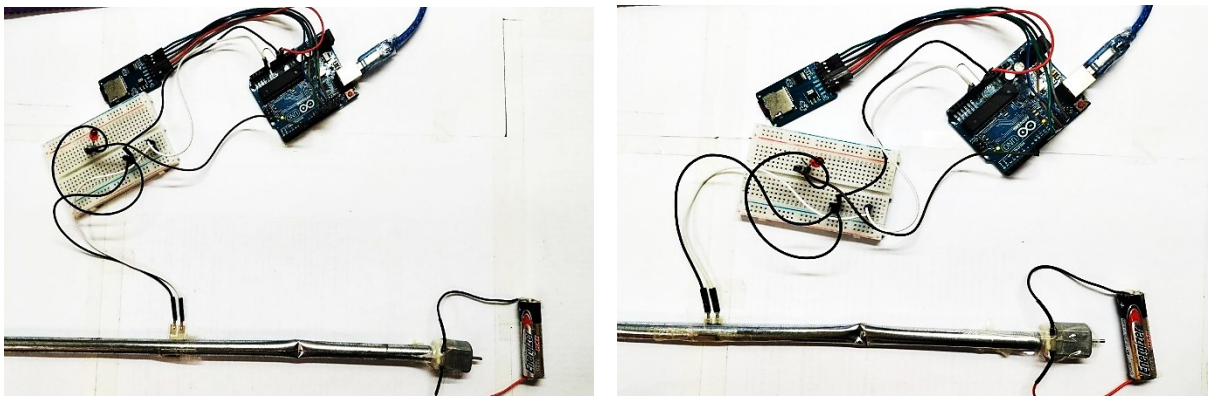


Figure 2. Experimental setup with a bent (left) and a cracked (right) pipe with damage.

3.1. Experimental Setups

Three types of pipe conditions are considered for the experiment. We have used aluminum pipes (6061 aluminum alloy tube (aluminum magnesium silicon alloy)) with an approximate diameter of 9 mm and 0.5 mm. Mechanical properties are mentioned in Table 2. First, we take

https://doi.org/10.31436/iiumej.v26i2.2988

a healthy pipe without deformation to find the threshold response values. These values will later be compared with the other cases. A bent pipe without any holes is then experimented on. Lastly, a cracked pipe is investigated to determine distinguishable differences. The pipe's bent length is approximately 7 mm. The open hole in the pipe is approximately 8.5 mm. The experimental setups are shown in Table 1 and demonstrated in Figure 2. All experiments are performed using two criteria: sensor location and sensor pattern. Sensor location reflects the position of the sensor with respect to the crack. Therefore, the top sensor location indicates that the sensor is positioned at the end of the pipe and aligned with the crack. Additionally, the sensor is attached to the pipe either parallel or bent according to the pipe surface, as illustrated in Table 1.

3.2. Components

This investigation utilizes three main component blocks: vibration sources, reading platforms, and recorders.

- *Vibration Source*: A vibration source will generate the prescribed vibrations. It is placed at one end of the investigated pipe. As previously mentioned, a DC motor has been utilized as the vibration source for this study. A power source is used to drive the vibration source. We have applied 3V DC to run the source; the rated RPM of the motor is approximately 13000 RPM, 3V.
- *Reading Platform*: A reading platform can read traveling vibrations through the pipe as responses. We have used Arduino Uno to read the pressure responses.
- *Recorder*: The pressure responses can be stored in memory. For this work, we have used a 16-gigabyte memory.
- *Definition*: The voltage presented in Section 4 (Results and Discussion) does not represent the real-time measured voltage. As previously mentioned, Arduino Uno cannot display the sensor output in volts. Therefore, we measured the supplied source voltage and distributed it based on the sensor value range. The sensor value range is 1-1024, and the measured voltage is 2 volts. Therefore, each sensor value represents (2000/1024) = 1.953 mV; the 1024 sensor value corresponds to 2 V.

3.3. Cases

Three types of cases are examined in the experiment based on the pipe's condition. In case 1, an intact pipe without any deformities is investigated. In cases 2 and 3, bent and cracked pipes are analyzed in the investigation. The bent and cracked pipes are intentionally deformed, not naturally occurring deformities. For all cases, 4 conditions are considered, with 2 kinds of sensor locations and 2 types of bending patterns. The conditions are described as follows:

- *Condition 1*: In this condition, the sensor is not bent and is attached in a parallel position with respect to the pipe. The sensor is located at the end of the pipe.
- *Condition 2*: In this condition, the sensor is bent and is attached to a vertical position with respect to the pipe. The sensor is located at the end of the pipe.
- *Condition 3*: Similar to condition 1, the sensor is not bent and is attached in a parallel position with respect to the pipe. However, compared to condition 1, the sensor is located at the end of the pipe with a 180-degree rotation.

https://doi.org/10.31436/iiumej.v26i2.2988

- *Condition 4*: Similar to condition 2, the sensor is not bent and is attached vertically to the pipe as well. However, compared to condition 2, the sensor is located at the end of the pipe with a 180-degree rotation.

4. RESULTS AND DISCUSSION

4.1. Case 1

In the first case, the pipe had no external damage. From the results, we can see that conditions 1 and 4 both have the highest sensor value of 244 a.u. (intensity) with averages of 179.33 a.u. and 175.51 a.u., respectively. Conditions 2 and 3 both show the highest value of 262 a.u., with average sensor values of 179.11 a.u. and 172.45 a.u., respectively. Hence, we can summarize that for case 1, the highest sensor value obtained was 262 a.u. The results are summarized in Figure 3.

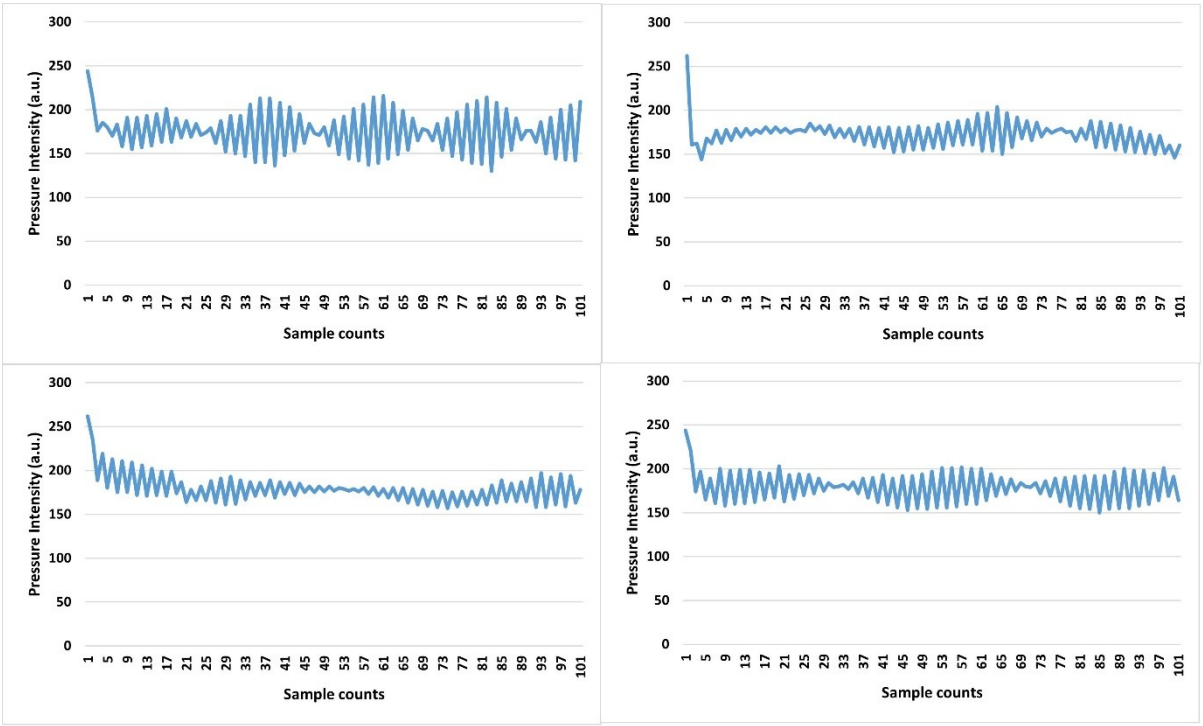


Figure 3. Pressure responses from pipe with no external damage (case 1)

4.2. Case 2

In the second case, the pipe is bent, so it has existing damage. As Figure 4 shows, condition 1 exhibits 300 a.u. of maximum values with an average of 177.21 a.u., while condition 2 has a maximum of 311 a.u. and an average of 190.54 a.u. It is 321 a.u. maximum and an average of 175.69 a.u., for condition 3 and a maximum of 590 a.u. with an average of 242.18 a.u. for condition 4. Hence, the highest sensor value recorded for this case is 590 a.u.

https://doi.org/10.31436/iiumej.v26i2.2988

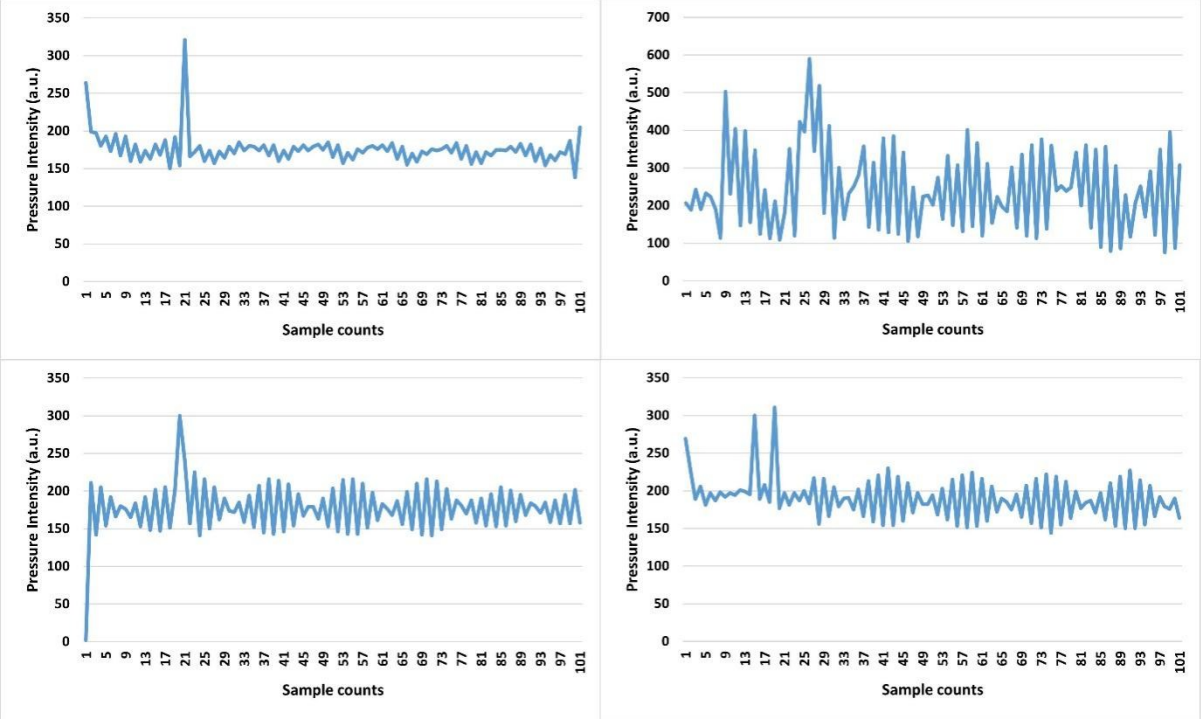


Figure 4. Pressure responses from a bent pipe without an existing air hole (case 2)

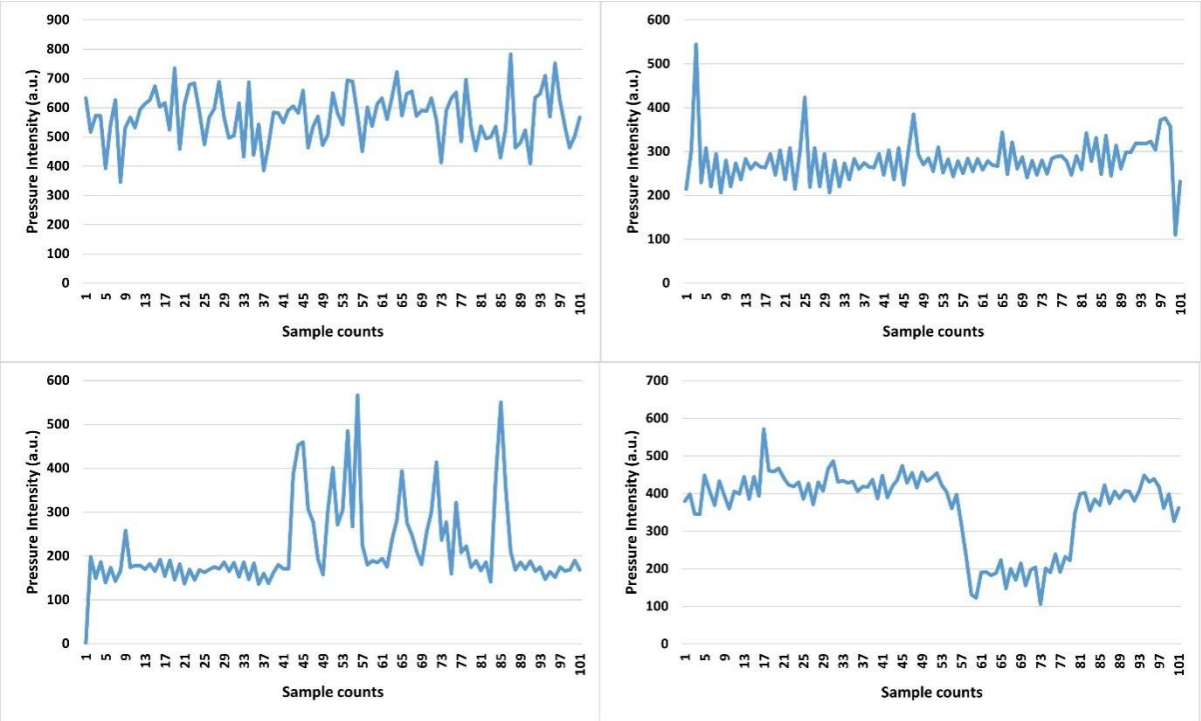


Figure 5. Pressure responses from the pipe with a crack and an existing air hole (case 3)

4.3. Case 3

In this case, the pipe is damaged with a crack. The results are depicted in Figure 5. In this case, condition 1 exhibits 566 a.u. as the maximum values with an average of 216.65 a.u., while condition 2 has a 572 a.u. and 363.49 a.u. average value. It is 783 a.u. and an average of 568.63

https://doi.org/10.31436/iiumej.v26i2.2988

a.u. for condition 3 and a maximum of 544 a.u. with an average of 279.13 a.u. for condition 4. Hence, the highest sensor value recorded for this case is 783 a.u.

4.4. Discussion

Since we have the required values, we can now make the decisions from the aforementioned discussions and results presented in Figures 3, 4, and 5. We can see a maximum sensor value of 262 a.u. with an average of 179.11 a.u. indicates a healthy pipe response, as shown in Figure 3. Comparatively, a bent pipe will experience a maximum of 590 a.u. with an average of 242.18 a.u. of pressure responses. But for the case of a cracked pipe, a 783 a.u. maximum value and average of 568.63 a.u. is found.

The compiled results are presented in Figure 6. From the figure, we can see a better picture of the cases regarding distinguished responses. Figure 6 presents the received pressure responses for conditions 1, 2, 3, and 4, where the sensor is either not bent or bent, either aligned to the crack or not. Therefore, this number can be taken as a threshold value for the considered cases.

However, the differences between a healthy and a bent pipe are very difficult to distinguish. The maximum pressure responses are very close for conditions 1, 2, and 3, which are 56 a.u., 49 a.u., and 59 a.u. Interestingly, condition 4 exhibits 346 a.u. pressure differences, which are over the threshold value. Therefore, it might not be wise to consider all four conditions when evaluating the metal pipe deformation.

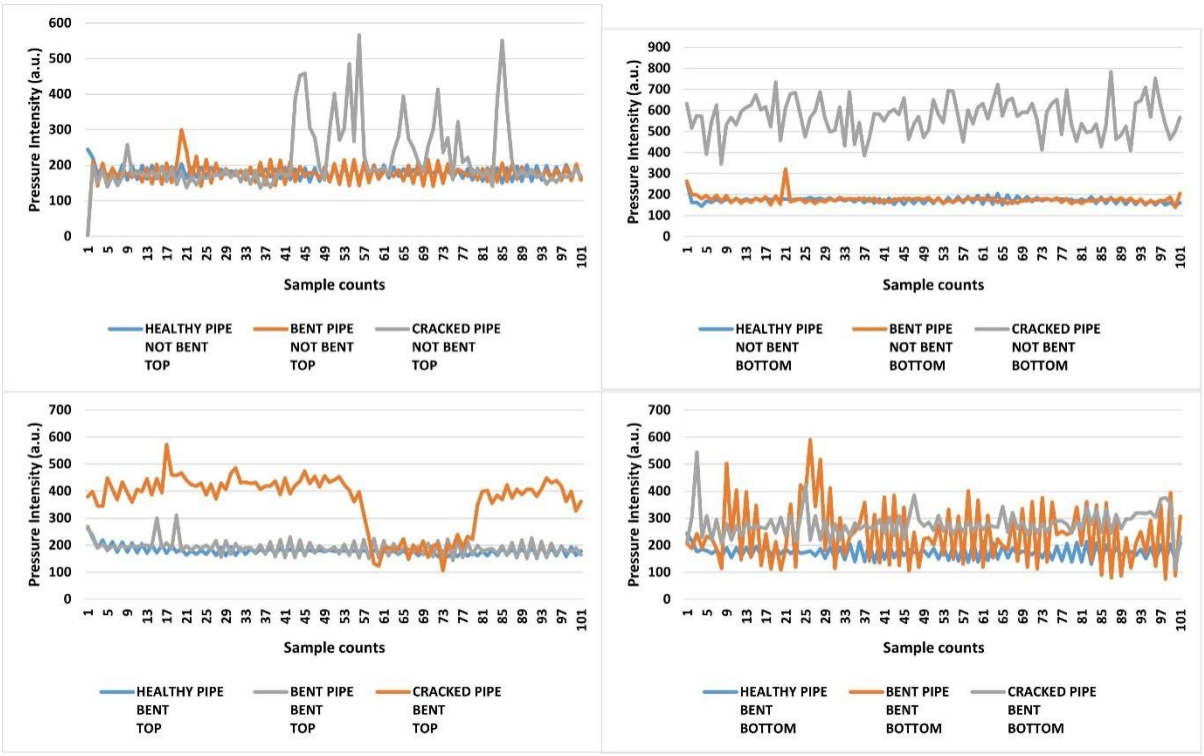


Figure 6. Cases and conditions comparison

We can now introduce an algorithm to determine the pipe health cases based on the aforementioned results and data summary from Table 2. The cases are differentiable based on the maximum sensor values in different conditions. Algorithm 1 presents the distinguishable differences among the three cases.

https://doi.org/10.31436/iiumej.v26i2.2988

Table 2. Summary of all experiments

Experimental Setup	Maximum Sensor Value (a.u.)	Minimum Sensor Value (a.u.)	Threshold Value	Damage
Healthy pipe	262	244	No	No
Bent pipe without an air hole	590	300	Yes	Yes
Cracked pipe with an air hole	783	544	Yes	Yes

Algorithm 1 Pipe Health Algorithm

Pipe Deformation (Threshold Pressure, Max Pressure)
Pressure Values, ⌊0⌋ → ⌈1023⌉
Received Response, R_R= ⌊0⌋ → ⌈*Sensor Values*⌉;
Max Pressure, R_{max}= Max [R_{R1}, R_{R2}, R_{R3}, ...R_{Rn}]; where n = # of cases
Threshold Pressure, $T_R = \exists!|\lceil R_{max} \rceil|$;
Condition, $C = C_N$; where N = # of conditions
while $R^{CN}_{max} \geq T_R$ **do**
    Pipe Deformation = EXISTS
    if $R^{CN}_{max} \geq T_R$ && $\forall N$ **then**
        STATUS = Deformation Identified [CRACK]
    else if $R^{CN}_{max} \geq T_R$ && $\neg \forall N$ **then**
        STATUS = Deformation Identified [OTHERWISE]
    end if
end while
if $R^{CN}_{max} < T_R$ **then**
    STATUS = HEALTHY PIPE
else
    STATUS = ERROR
end if

Algorithm 1 depicts the decision logic for the induced pressure in the range of 0 to 1023, as we designed it for the Arduino. Thus, the received responses will be recorded and compared later to the threshold values. From the four different experimental setups, it is noticeable that the sensor exhibits a distinct pattern when placed at the bottom of the existing crack. It remains above the threshold values for all the samples for a cracked pipe. Therefore, this case will be identified as a cracked pipe. However, all other cases for the bent pipe will show a significantly lesser number of samples below the threshold. This scenario will distinguish the bent pipe from the cracked pipe.

5. CONCLUSION

Non-destructive testing (NDT) is significant in structural health monitoring systems, especially in fluid transportation using metal pipes. However, the decaying performance of sensor-based systems at high costs poses a drawback. Therefore, a system with low cost and minimal performance decay is desirable. We have utilized a DC motor as a wave source, which offers significant advantages over traditional sensors. The primary objective of this study is to assess the viability of using a motor as a vibration source for detecting potential cracks in a metal pipe. A simple PVDF sensor was used at the pipe's other end to monitor the motor's pressure responses. An Arduino platform is used to read and record the collected responses afterward. Three types of scenarios are analyzed in this study. These scenarios include a healthy pipe, a bent pipe, and a cracked pipe. Furthermore, four different conditions are also considered, involving two sensor locations (180 degrees rotation) and sensor patterns (bent and

not bent). The results reveal significant differences in pressure responses between healthy and cracked pipes. The maximum pressure response for a cracked pipe is 783 a.u. whereas it is just 262 a.u. for a healthy pipe. Hence, the differences are sufficient to set a threshold margin. We have set 300 a.u. as the threshold margin and applied it in an algorithm. However, it is very tricky in the case of a bent pipe, as the pressure differences are less than 300 a.u. for three conditions and above for only one. This could potentially result in an inaccurate decision when detecting a bent pipe.

As previously mentioned, the primary objective of this study is to assess the feasibility of a motor as a vibration source. Hence, no comparisons were made with existing NDTs in this study. Consequently, only open-ended pipes were tested. The prepared pipe damages are artificial (human-made) and not natural damages; hence, the crack pattern will be distinguishable.

ACKNOWLEDGEMENT

This work is partially supported by Universiti Malaysia Terengganu, under the fundamental research grant scheme (Ref: FRGS/1/2019/TK04/UMT/03/1), Ministry of Higher Education, Government of Malaysia.

REFERENCES

[1] Zakikhani K, Nasiri F, Zayed T. A review of failure prediction models for oil and gas pipelines. Journal of Pipeline Systems Engineering and Practice. 2020 Feb 1;11(1):03119001.

[2] Lu H, Behbahani S, Azimi M, Matthews JC, Han S, Iseley T. Trenchless construction technologies for oil and gas pipelines: State-of-the-art review. Journal of Construction Engineering and Management. 2020 Jun 1;146(6):03120001.

[3] Coramik M, Ege Y. Discontinuity inspection in pipelines: A comparison review. Measurement. 2017 Dec 1;111:359-73.

[4] http://phmsa.dot.gov/pipeline/library/data-stats (Available 26.12.2016).

[5] Gas pipeline incidents, 9th report of the European Gas Pipeline Incident Data Group, (period 1970 – 2013), Doc. Number EGIG 25714.R.0403, February 2015.

[6] Kishawy HA, Gabbar HA. Review of pipeline integrity management practices. International Journal of Pressure Vessels and Piping. 2010 Jul 1;87(7):373-80.

[7] Tehranchi MM, Ranjbaran M, Eftekhari H. Double core giant magneto-impedance sensors for the inspection of magnetic flux leakage from metal surface cracks. Sensors and Actuators A: Physical. 2011 Nov 1;170(1-2):55-61.

[8] Alobaidi WM, Alkuam EA, Al-Rizzo HM, Sandgren E. Applications of ultrasonic techniques in oil and gas pipeline industries: A review. American Journal of Operations Research. 2015;5(04):274.

[9] P.E. Mix, Introduction to Nondestructive Testing a Training Guide, Wiley Inter Science, U.S.A, 2005.

[10] Awal, M.R., Jusoh, M., Yahya, M.S., Rahman, S.A., Dagang, A.N., Latiff, N.A.A., Zakaria, H.A. and Saat, S., 2021. Acoustic wave propagation in high scale impedance mismatch mediums. IIUM Engineering Journal, 22(2), pp.1-9.

[11] Awal MR, Jusoh M, Sabapathy T, Ahmad RB, Kamarudin MR, Osman MN, Ahmad MF, Rahman SA, Dagang AN. A Patternless Piezoelectric Energy Harvester for Ultra Low Frequency Applications. International Journal of Integrated Engineering. 2020 Oct 26;12(7):187-94.

[12] Awal MR, Zakaria NA, Jusoh M, Abdelghany MA, Yahya MS, Zakaria HA. Warranty Seal Deformation Identification for Product Warranty Violation. Sensors. 2022 Jun 21;22(13):4688

[13] Ho M, El-Borgi S, Patil D, Song G. Inspection and monitoring systems subsea pipelines: A review paper. Structural Health Monitoring. 2020 Mar;19(2):606-45.

[14] Wang Q, Bi K, Hao Y, Guo L, Dong G, Wu H, Lei M. High-sensitivity dielectric resonator-based waveguide sensor for crack detection on metallic surfaces. IEEE Sensors Journal. 2019 Mar 25;19(14):5470-4.

[15] Zhou X, Chen H, Hao T. Efficient detection of buried plastic pipes by combining GPR and electric field methods. IEEE Transactions on Geoscience and Remote Sensing. 2019 Feb 8;57(6):3967-79.

[16] Memon, A.M., Imran, I.H. and Alhems, L.M., 2023. Neural network based corrosion modeling of Stainless Steel 316L elbow using electric field mapping data. Scientific Reports, 13(1), p.13088.

[17] Rifai D, Abdalla AN, Razali R, Ali K, Faraj MA. An eddy current testing platform system for pipe defect inspection based on an optimized eddy current technique probe design. Sensors. 2017 Mar 13;17(3):579.

[18] Santos, D., Machado, M.A., Monteiro, J., Sousa, J.P., Proença, C.S., Crivellaro, F.S., Rosado, L.S. and Santos, T.G., 2023. Non-Destructive Inspection of High Temperature Piping Combining Ultrasound and Eddy Current Testing. Sensors, 23(6), p.3348.

[19] Guo W, Zhang X, Zhang D, Chen Z, Zhou B, Huang D, Li Q. Detection and classification of pipe defects based on pipe-extended feature pyramid network. Automation in Construction. 2022 Sep 1;141:104399.

[20] Zhang, H., Ding, Z., Zhou, L. and Wang, D., 2023. Particle Filtering SLAM algorithm for urban pipe leakage detection and localization. Wireless Networks, pp.1-12.

[21] Ékes C. New Developments in Multi-Sensor Condition Assessment Technologies for Large Diameter Pipe Infrastructure. InPipelines 2018 2018 Jul 12 (pp. 142-148). Reston, VA: American Society of Civil Engineers.

[22] Lowe MJ, Alleyne DN, Cawley P. Defect detection in pipes using guided waves. Ultrasonics. 1998 Feb 1;36(1-5):147-54.

[23] Guan R, Lu Y, Wang K, Su Z. Fatigue crack detection in pipes with multiple mode nonlinear guided waves. Structural Health Monitoring. 2019 Jan;18(1):180-92.

[24] Zhao J, Durham N, Abdel-Hadi K, McKenzie CA, Thomson DJ. Acoustic guided wave techniques for detecting corrosion damage of electrical grounding rods. Measurement. 2019 Dec 1;147:106858.

[25] Durham N, Zhao J, Bridges G, Thomson D. Acoustic guided wave detection of grounding rod corrosion: equivalent circuit model and implementation. Smart Materials and Structures. 2020 Apr 8;29(5):055040.

https://doi.org/10.31436/iiumej.v26i2.3089

SCALABILITY AND COST OPTIMIZATION IN LOAD-BALANCED MICROSERVICE SCHEDULING SYSTEM

SHAMSUDDEEN RABIU [1], CHAN HUAH YONG [1*], SHARIFAH MASHITA SYED-MOHAMAD [2]

[1]*School of Computer Sciences, Universiti Sains Malaysia (USM), Penang, Malaysia*
[2]*Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

**Corresponding author: hychan@usm.my*

(Received: 31 October 2023; Accepted: 22 April 2025; Published online: 15 May 2025)

ABSTRACT: Microservice, a widely adopted architectural paradigm to overcome monolithic limitations, faces difficulties in efficient load balancing, scalability, and cost-effective deployment. To address these issues, we introduce a Container Microservice Load Balanced (CMLB) framework, which integrates the novel OEPTA algorithm. This framework aims to optimize microservice-based applications deployed on Docker within cloud environments. Common microservices scheduling strategies often grapple with load distribution challenges, resulting in suboptimal resource utilization. Concurrently, traditional containerization methods face difficulties reconciling trade-offs between scalability, deployment cost, and execution time. Our primary goal is to present a comprehensive solution that enhances the scalability, cost efficiency, and execution time of microservices deployment. This paper introduces a novel deployment framework for microservices, leveraging Docker for decentralized resource allocation across Microservice Controllers (MSCs). Additionally, a specialized algorithm is introduced to evaluate the cost, execution time, and availability aspects of microservice applications, enabling optimized resource allocation in a distributed manner. The evaluation results demonstrate that the CMLB framework, driven by the OEPTA algorithm, surpasses existing algorithms in achieving optimal scalability, cost efficiency, and execution times. This research provides a robust solution to enhance microservices deployment in cloud environments, effectively addressing key challenges in the field.

ABSTRAK: Mikroservis, sebuah paradigma seni bina yang diadaptasi secara meluas untuk mengatasi keterbatasan monolitik, menghadapi kesulitan dalam penyeimbangan beban yang cekap, skalabiliti, dan penyebaran kos efektif. Untuk mengatasi masalah ini, kami memperkenalkan rangka kerja Container Microservice Load Balanced (CMLB), yang mengintegrasikan algoritma OEPTA yang baru. Rangka kerja ini bertujuan untuk mengoptimumkan aplikasi berasaskan perkhidmatan mikroservis yang digunakan pada Docker dalam persekitaran awan. Strategi penjadualan mikroservis umumnya bergelut dengan cabaran pengagihan beban, yang menghasilkan penggunaan sumber daya yang kurang optimal. Pada masa yang sama, kaedah pengkontenaan tradisional menghadapi kesulitan dalam menyeimbangkan pertukaran antara skalabiliti, kos penggunaan, dan masa pelaksanaan. Matlamat utama kami adalah untuk membentangkan penyelesaian komprehensif yang meningkatkan skalabiliti, kos kecekapan, dan masa pelaksanaan dalam penggunaan mikroservis. Dalam makalah ini, kami memperkenalkan rangka kerja penggunaan yang baru untuk perkhidmatan mikroservis, dengan memanfaatkan Docker untuk peruntukan sumber terdesentralisasi merentas Pengawalan Perkhidmatan Mikroservis (MSCs). Selain itu, algoritma khusus diperkenalkan untuk menilai kos, masa pelaksanaan, dan ketersediaan aplikasi mikroservis, membolehkan peruntukan sumber dioptimumkan dalam cara yang diedarkan. Keputusan penilaian menunjukkan bahawa rangka kerja CMLB, didorong oleh algoritma OEPTA, mengatasi algoritma sedia ada dalam mencapai skalibiliti

optimum, kecekapan kos, dan masa pelaksanaan. Penyelidikan ini memberikan penyelesaian yang teguh untuk meningkatkan penggunaan mikroservis dalam persekitaran awan, menangani cabaran utama dalam lapangan dengan berkesan.

KEYWORDS: *Algorithm, Cloud-based, Container, Docker, Load balancing, Microservice.*

1. INTRODUCTION

Monolithic architecture refers to an application with a unified code base housing multiple services that interact with external systems or consumers through diverse interfaces like Web services, HTML pages, or REST API [1]. In this architectural model, all functionalities are consolidated within a single application, resulting in modules that cannot function independently [2]. This inherent characteristic of tight coupling means that all logic for handling a request operates within a single process [3]. Despite the initial advantages of ease in development, testing, and deployment for simpler applications, the drawbacks of monolithic architecture become apparent as the application becomes more intricate. The monolith's structure expands in size, transforming into a cumbersome and challenging piece of software to manage and scale [2]. As the application size and team grow, the limitations of this architecture become increasingly significant [3]:

- Complexity in understanding and modifying the application, leading to a deceleration in development speed.
- Difficulty in continuous deployment, where even minor changes necessitate the complete rebuilding and redeployment of the entire monolith.
- Challenges in scaling the application, restricted to horizontal scaling within the confines of monolithic architecture.

To overcome the challenges associated with monolithic applications and harness the benefits of Service-Oriented Architecture (SOA), the microservices architecture pattern has emerged as a lightweight subset of SOA, as exemplified by companies like Amazon [4]. In recent years, microservices have gained significant traction in the business landscape, representing an enhanced and streamlined version of SOA [5]. According to [6], microservices architecture is a specific implementation approach within SOA that facilitates the creation of flexible and independently deployable software systems. This approach typically involves breaking down a software application into smaller components that collaborate to achieve a specific, complex task, thereby facilitating easier development and maintenance [7].

The Microservices architecture represents a paradigm shift away from traditional development methods, placing emphasis on building the applications through small, autonomous services that communicate via lightweight mechanisms [6]. It's essential to note that the term "micro" in Microservices refers to the size of their contribution to the application, rather than the lines of code they encompass [2]. Therefore, the architecture can be understood as a set of small services with precise tasks that interact to achieve users' goals through standard communication channels [8, 9, 10]. This architectural approach offers a strategy for developing a cohesive application as a collection of small services. Each service operates independently in its dedicated process and communicates through lightweight mechanisms, such as HTTP APIs [10]. These Microservices are designed around specific business capabilities, enabling them to be independently deployable through automated deployment processes [3]. Lately, there has been a surge in using containers to distribute microservices across various cloud locations [11].

Containers, an emerging virtualization technology, are gaining popularity over virtual machines (VMs) due to their superior performance, lightweight design, and enhanced scalability [12]. These containers act as comprehensive packages, bundling applications with their dependencies, facilitating easy and consistent deployment across diverse environments. Dependencies, including binaries, libraries, and configuration files, are essential for the application's functionality [13]. Containers encapsulate self-contained, deployable components of applications, and may include middleware and business logic in the form of binaries and libraries [10]. Container engines, such as Docker, leverage containers as portable entities for packaging applications. This shift emphasizes the critical need to manage container dependencies [14]. The application consists of individual, self-contained services that operate in their own processes and communicate through a lightweight mechanism [14].

To mitigate the risk of a single instance becoming a bottleneck or a potential point of failure [16], a load-balanced microservice scheduling system distributes requests for a specific service across multiple instances of that service. Typically, a load balancer, positioned in front of the service instances, achieves this distribution by directing incoming traffic to the least busy instance [17]. The primary objective of load balancing is to optimize resource utilization, enabling the system to handle increased traffic levels with minimal downtime or performance slowdowns [18]. In microservices, load balancing plays a crucial role in maintaining uninterrupted services even if one or more components fail, utilizing failover [17]. This involves adding and removing instances of applications in a balanced manner, preventing failures [19]. Numerous research studies have aimed to improve Quality of Service (QoS) in Container Microservices through load balancing [20]. However, many of these studies have found existing methods ineffective in enhancing user QoS as the methods often rely on queuing systems, leading to issues like increased network traffic, longer processing times, server overloads, and high deployment costs [6,17,18,19,21]. This results in sudden load spikes, disrupting system balance and degrading performance. While load balancing traditionally considers factors like traffic and performance constraints, adjusting computational resources dynamically to optimize costs based on load changes is also crucial [17]. Hence, applying load balancing strategies becomes crucial for optimizing performance in such scenarios.

To address the challenges associated with load-balanced microservice scheduling systems, we introduced a pioneering framework called CMLB. This framework is designed to optimize server overloads, manage traffic spikes, and enhance the cost efficiency of microservices. We developed a load balancing algorithm to determine the service deployment cost, improve reliability, and ensure the availability of microservice applications. In our proposed system, the allocation and management of resources for applications are decentralized and orchestrated by the Master Load Balancer (MLB), operating through Local Load Balancers (LLB) on Microservice Controllers (MSCs). The MSCs play a vital role in decision-making related to resource allocation, requesting resources for Execution Containers (ECs), monitoring task progress on ECs, and overseeing the life cycle of ECs. Simultaneously, the ECs are responsible for executing assigned tasks, providing progress updates to the MSCs, and comparing their performance against expected outcomes. The primary contributions of this paper are outlined as follows:

- Proposed a novel Container Microservice Load Balanced (CMLB) framework, designed for deploying microservice-based applications with Docker, incorporating autonomous resource allocation on each controller in a distributed manner.
- Presented an OEPTA algorithm to optimize the cost, execution time, and scalability aspects of microservice applications.

- Implemented and evaluated the OEPTA algorithm's effectiveness, showcasing its capacity to deliver positive outcomes regarding traffic spike management and server overload minimization.

The remaining sections of this paper are organized as follows: Section 2 presents a literature review on related works. Section 3 offers an overview of the methodology. Section 4 presents results and discussions. Section 5 provides a summary of the conclusion and outlines future work. Table 1 highlights key processes, components, and challenges within the system. Figure 1 summarizes the deployment of microservice applications with Docker Container, while Figure 2 describes the elements of the cluster and introduces the concept of a Global ResourceSpace. Figure 3 illustrates the workflow for application deployment and introduces the Optimize EC Placement and Task Assignment Algorithm (OEPTA) for resource allocation decisions. The presented frameworks aim to address the challenge of deploying and managing microservice applications efficiently within clusters. The key problems include optimal resource allocation, load balancing, and timely task execution. The primary objectives are to ensure minimal delays, maximize resource utilization, and maintain overall system performance. The introduced OEPTA algorithm balances application demands, available resources, and cluster configurations to make informed decisions for microservice deployment and execution. Table 1 lists the notations used in this section to aid understanding of the algorithm.

Table 1. Summary of the Notations used and their descriptions.

Notations	Descriptions
Adr	Application Deployment Request
MS	Microservice
$MSCs$	Microservice Controllers
ECs	Execution Containers
PM	Physical Machine
N^p	Set of physical nodes
Λ	Set of applications needed to be deployed
G^a	Set of all clusters in which the application is deployed.
M^a	Set of microservice M^a for application G^a
$R^a(u)$	Total resources available on PM 'i' for microservice 'u' in application 'a'.
$CR^a(u)$	Resource requirement of microservice 'u' in application 'a'.
$CN^{node}(i,a)$	Cost associated with deploying application 'a' on physical machine (PM) 'i' in the cluster.
C^0	Cost of using resources
$x(i,a,u)$	Resource usage of microservice 'u' in application 'a' on PM 'i'.
$l^a(k,u)$	Processing time of task 'k' for microservice 'u' in application 'a'
$l^p(k,i)$	Predefined execution time requirement for task 'k' on PM 'i'.
$l(k,i)$	Communication time between tasks on PM 'i'.
$RR^p(i)$	Predefined resource requirement for cluster 'i'.
$C^p.s_i$	Processing cost of the PM 'i'.
$C^l(k)$	Cost of accessing the PM, which may depend on the microservice 'k'.

2. LITERATURE REVIEW

In recent years, an expanding body of research has focused on deploying and managing microservice containers [17, 19, 20, 22]. The Microservices architecture has garnered significant attention for its adaptability, cost-effectiveness, and scalability, leading large enterprises to deploy microservices across diverse cloud locations [17, 20, 22]. Containers, known for their lightweight nature compared to virtual machines, enable easy downloads and swift deployments [11,23]. Adopting a microservices architecture presents several advantages, including reduced interdependence between services, faster recovery in the face of catastrophic events,

and heightened reliability achieved by isolating the impact of failures to a small service segment [24]. However, as applications scale, challenges arise, notably in increased API calls. This necessitates the implementation of effective load balancing solutions to manage API calls across the architecture.

Contributing to this domain, [17] developed a microservices architecture utilizing Docker containers to enhance scalability and elasticity in the cloud computing environment. Their approach aims to reduce deployment and operational costs while meeting service delay requirements. Expanding upon their work, our research enhances their framework by incorporating additional parameters, including execution time, traffic spikes, and cost considerations. We further augment the framework by distributing requests evenly across all PMs through a master load balancer on the client side. This extension is designed to ensure Quality of Service (QoS) for developers, responding to the growing demand for new software [25].

A key concern in container microservices cloud-based systems revolves around the intricacies of load balancing. This process entails the equal distribution of workloads across servers to forestall service failures, minimize response time, alleviate downtime, and safeguard against data loss [17, 26, 27]. Effective load balancing is paramount for averting resource overload, enhancing performance, handling unforeseen traffic spikes, curtailing response time, and optimizing resource utilization [27]. In a correlated investigation, an inventive Load Balancing Ant Colony Optimization (LBACO) algorithm was devised to distribute workloads throughout the entire system, thereby reducing the makespan [21]. Meanwhile, [28] introduces a groundbreaking approach to enhance performance and diminish latency by leveraging SmartNICs on edge servers for middlebox processing. Their SmartLB methodology deploys a load balancer and an auto scaler entirely on the SmartNIC, resulting in judicious decisions and decreased CPU load. Despite numerous existing methodologies for optimizing load balancing, container placement, application deployment costs, operational costs, service failure, and traffic issues, many treat these challenges as knapsack problems, often neglecting essential load balancing features. Therefore, integrating these crucial features into our system will significantly enhance its overall performance.

3. METHODOLOGY

This section outlines the systematic scalability and cost optimization approach in the proposed Load-Balanced Microservice Scheduling System. It details the overarching strategy for developing and implementing the microservices framework with Docker containers and evaluates its performance. The methodology involves creating a simulation model, implementing a novel scheduling algorithm, and thoroughly evaluating results based on predefined criteria.

3.1. System Model

To achieve a balanced distribution of requests across microservice instances, we defined this process in a structured manner using our equations. These equations serve as a foundation for the OEPTA algorithm, enabling it to make well-informed choices on resource distribution, task delegation, and load equilibrium. The algorithm aims to minimize costs, ensure scalability, and fulfill execution time criteria for deploying applications.

Our system model analyzes a physical network comprising a collection of Physical Machines (PMs). Each PM possesses finite physical resources, with our focus being on computational resources as a representative illustration of the resource allocation challenge. We assume uniformity in capacity and pricing among the PMs. However, the PMs may install different libraries in advance to support different microservices for applications.

https://doi.org/10.31436/iiumej.v26i2.3089

In the OEPTA algorithm, there are three equations used to guide the decision-making process:

a. Cost

$$CN^{node}(i,a) = C^p.s(i) + \sum_k C^l(k) + \sum_{u \in M^a} C^0.\frac{x(i,a,u).CR^a(u)}{R^a(u)} \quad (1)$$

Eq. (1) calculates the cost of deploying an application on a specific physical machine (PM) in a cluster. It considers various factors such as the processing cost of the PM ($C^p.s(i)$), the cost of accessing the PM ($C^l(k)$), and the cost of using resources on the PM ($C^0.\frac{x(i,a,u).CR^a(u)}{R^a(u)}$). By evaluating the cost for each PM in the cluster, the algorithm can select the PM with the lowest cost for the given application.

b. Scalability

$$\sum_{G^a \in \Lambda} \sum_{u \in M^a} \frac{x(i,a,u).CR^a(u)}{R^a(u)} < RR^p(i), \forall i \in N^p \quad (2)$$

Eq. (2) is the scalability equation, which assesses the overall resource utilization across all clusters for a specific application. It calculates the ratio of resource usage $\frac{x(i,a,u).CR^a(u)}{R^a(u)}$ for each microservice) u in the application across all clusters (G^a) and compares it to a predefined resource requirement ($RR^p(i)$). The algorithm proceeds to the next step if the total resource utilization meets the scalability requirement. Otherwise, it revisits the allocation and assignment process for better resource utilization.

c. Execution Time

$$\sum_{G^a \in \Lambda} \frac{l^a(k,u).x(i,a,u)}{R^a(u)} - l(k,i) \leq l^p(k,i), \forall i \in N^p, u \in M^a \quad (3)$$

Eq. (3) evaluates the expected execution time for each task (u) of an application on a specific $PM_{i,}$. It considers the processing time for the task on the PM ($l^a(k,u)$ and the communication time between tasks ($l(k,i)$). The equation sets a constraint ($\leq l^p(k,i)$) on the execution time, ensuring that the execution time for each task meets the predefined requirement. If the execution time constraint is satisfied for all tasks, the deployment is considered successful. Otherwise, the algorithm revisits the allocation and assignment process to optimize task execution time.

These equations help the OEPTA algorithm make informed decisions regarding resource allocation, task assignment, and load balancing, aiming to minimize costs, ensure scalability, and meet application deployment execution time requirements.

3.2. A Framework for Deploying Cloud-Based Microservice Container Applications with Docker

This section introduces an adapted Container Microservice Load Balanced (CMLB) framework that integrates the innovative OEPTA algorithm. This framework aims to improve the performance of microservice-based applications deployed on Docker in cloud environments. The system uses microservices on ECs to process application requests within a framework. Resource allocation and application management are decentralized, with Registry and Service Discovery coordinating the process through a Load balancer on MSCs. The MSCs are

https://doi.org/10.31436/iiumej.v26i2.3089

responsible for making decisions on resource allocation, requesting resources for ECs, monitoring task progress on ECs, and managing the life cycle of ECs. Once a task is complete, the EC reports back to the MSC on the progress compared to what was expected.

Fig. 1 shows the deployment of microservice applications with Docker containers. The process begins with the user sending an application deployment request ($Adr1$) to the gateway. The gateway then sends the request to the registry and Service Discovery (RSD) system, which registers and assigns the request to the first cluster in the resource table (since the cluster placement is done automatically, with a priority given to the least busy). Once the cluster receives the registration request, the load balancer is notified and will register and update the status of the microservices controllers and execution containers to execute the job. If more resources are needed, the load balancer will select the best candidate for the available cluster resource status in the global resourcespace to avoid delays or traffic. When a resource is picked from the global resourcespace, it disappears, and the following available resource is released for use. While the load balancer manages jobs, the other jobs in the queue will be assigned to subsequent clusters that are available concurrently, following the same process.

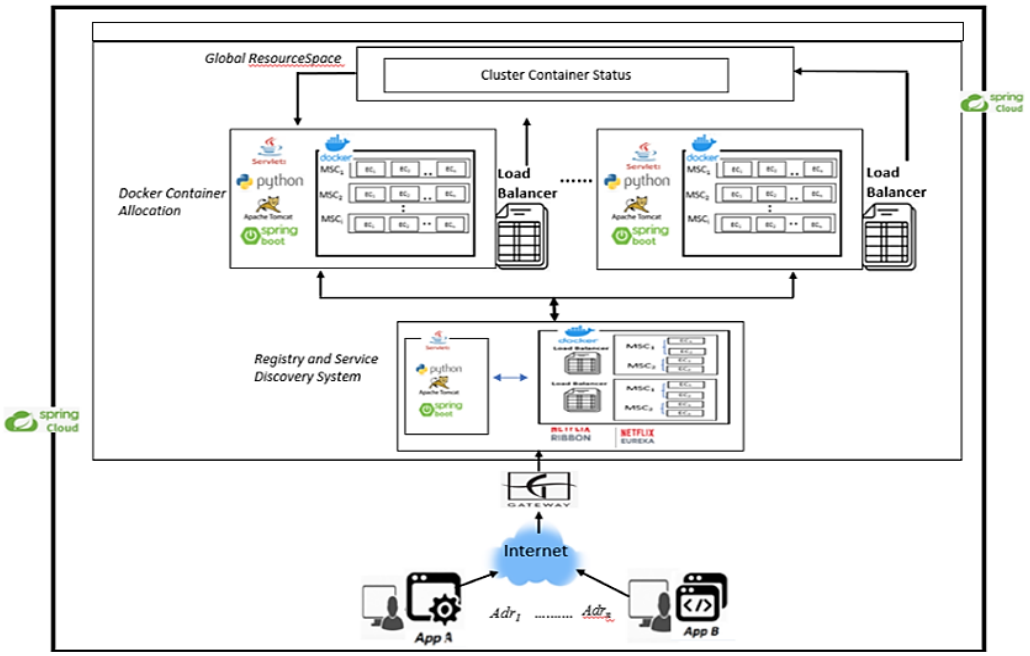


Figure 1. Framework for deploying microservice applications with Docker containers

Fig. 2 illustrates the main components of PM in a Cluster. Each PM within the cluster has a host operating system that runs Docker. This Docker engine is responsible for maintaining the container's operating environment, embedding containers, and isolating containers operating on the same PM. The Registry and Service Discovery system that registers and directs requests to the PM is also introduced. Each PM includes a load balancer that registers and updates the status of the MSCs and ECs operating on that PM. Furthermore, a Global ResourceSpace is introduced, allowing for the scaling up or down of the resources if needed.

The workflow of deploying the application deployment requests is illustrated in Figure 3. When an application request is sent to the registry and service discovery system, the registry and Service Discovery (RSD) system will register and assign the request to the load balancer to manage the microservice controllers and execution containers in the physical machine within the clusters. If more resources are needed, the load balancer will select the best candidate from the available global resourcespace to avoid delays or traffic. When a resource is selected from

https://doi.org/10.31436/iiumej.v26i2.3089

the global resourcespace, it disappears, and the following available resource is released from the cluster. While the load balancer manages cluster one, the other jobs in the queue will be assigned to subsequent clusters that are available concurrently, following the same process. Based on the application's requirements and available resources on PMs in the clusters, the load balancer makes resource allocation decisions using the proposed Optimize EC Placement and Task Assignment Algorithm (OEPTA).

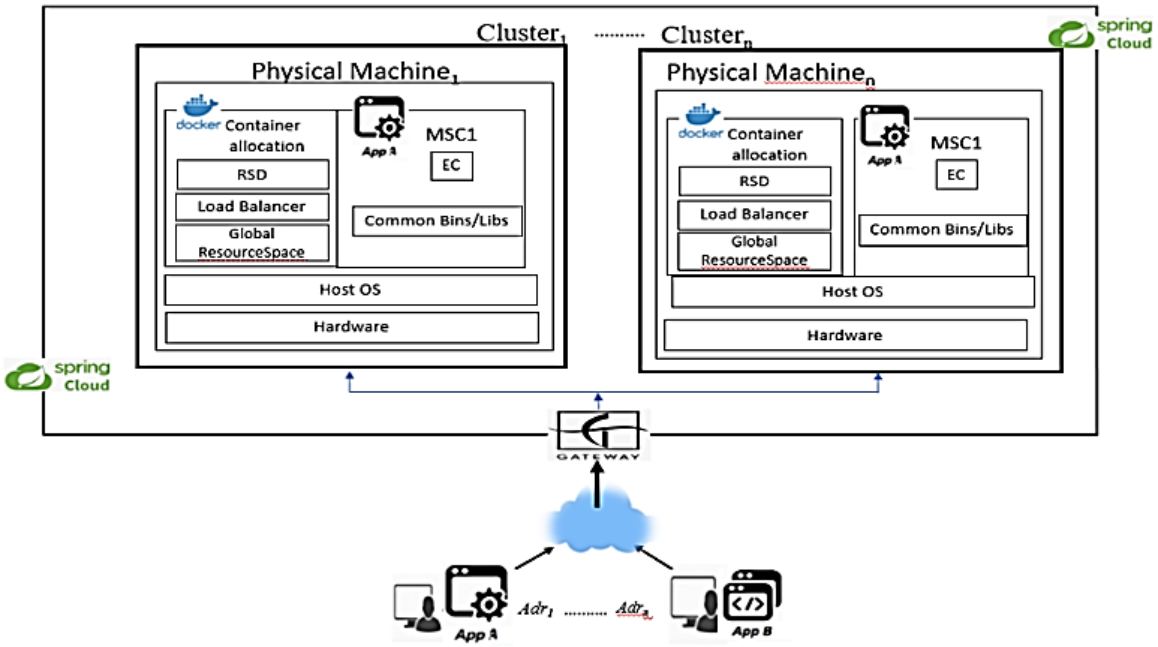


Figure 2. Components of physical machines ($PM_1, \ldots, PM_i$) in the clusters embed microservice controllers and execution containers.

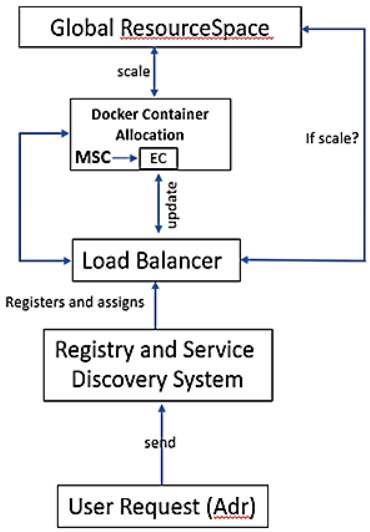


Figure 3. Workflow for deploying the application deployment requests.

3.3. Develop a Simulation Model and Implement the Proposed Algorithm.

This section outlines the overall design of the cloud-based container microservices framework and the specific implementation of the microservices function. The simulation model is established following the previously mentioned methodology. To assess the efficiency of the suggested algorithm, Netflix, Spring Boot, and Spring Cloud are employed as microservice

https://doi.org/10.31436/iiumej.v26i2.3089

implementations and communication simulation tools. To address the scalable distributed-systems problems, the registry and service discovery system, load balancer, and global resource space have all been utilized, resulting in practical solutions for scheduling requests. The design framework is anticipated to boost the performance of the CBCM system.

3.3.1. Algorithm Implementation

A comparison of the algorithms under consideration reveals some of their strengths and weaknesses. All of the algorithms are effective at resolving microservice scheduling issues [24, 26, 27]Although they are time-consuming and have a slow convergence speed for solving complex scheduling problems, the algorithms have been modified to address more specific issues such as cost, server overload, traffic spikes, and performance issues.

The OEPTA (Optimized Executive Containers Placement Task Assignment) algorithm optimizes the placement and assignment of executive containers in a distributed computing environment. It initializes variables and data structures, sets up clusters and physical machines, and receives application deployment requests. The algorithm registers and assigns requests, calculates costs for each physical machine, and selects the best candidate based on cost and other factors. Tasks are assigned to the selected machine, and load balancing is performed. A scalability check ensures resource utilization meets requirements, and execution time is calculated for each task. The system is updated by releasing resources, marking them as unavailable, and updating load balancing. The process is repeated for the remaining deployment requests. Overall, OEPTA aims to achieve efficient task execution through optimized placement and assignment of executive containers.

Algorithm: (Optimized Executive Containers Placement Task Assignment)

Input: User Application deployment requests $\left[Adr\middle|1, Adr_{2,} \dots \dots \dots \dots \dots \dots Adr_{n,}\right]$

Output: $min\{NT, Cost, Overloading\}$

1: Initialization:

2: Initialize necessary variables, data structures, and parameters.

3: Set up the clusters and their associated physical machines PMs.

4: Set up the global resourcespace.

5: For each application deployment request $Adr_{i,}$ in $\left[Adr\middle|1, Adr_{2,} \dots \dots \dots \dots \dots \dots Adr_{n,}\right]$:

6: Application Deployment Request:

7: Receive $Adr_{i,}$ and its associated parameters.

8: Register and Assign Request:

9: Send $Adr_{i,}$to the registry and service discovery system (RSD).

10: RSD registers and assigns $Adr_{i,}$ to the load balancer (LB).

11: Resource Allocation and Task Assignment:

12: Calculate the cost for each PM in the cluster for $Adr_{i,}$ using Eq. 1:

13: $$CN^{node}(i,a) = C^{p}.s(i) + \sum_{k} C^{l}(k) + \sum_{u \in M^{a}} C^{0}.\frac{x(i,a,u).CR^{a}(u)}{R^{a}(u)}$$

14: Select the best candidate from the available resources in the global resourcespace, considering cost and other factors.

15: Assign $Adr_{i,}$'s tasks to the selected PM, updating the resource allocation.

16: Release the next available resource from the cluster's resource pool.

17: Cluster Management and Load Balancing:

18: Manage the microservice controllers and execution containers on the assigned PM.

19: If there are other jobs in the queue, assign them to subsequent clusters that are available concurrently, following the same process.

20: Scalability Check:

21: Calculate the total resource utilization across all clusters for Adr(i) using Eq. 2:

22: $$\sum_{G^a \in \Lambda} \sum_{u \in M^a} \frac{x(i,a,u).CR^a(u)}{R^a(u)} < RR^p(i), \forall i \in N^p$$

23: If the condition in Eq. 2 is satisfied, proceed to the next step. Otherwise, revisit the allocation and assignment process.

24: Execution Time Calculation:

25: Calculate the execution time for each task on the assigned PM using Eq. 3:

26: $$\sum_{G^a \in \Lambda} \frac{l^a(k,u).x(i,a,u)}{R^a(u)} - l(k,i) \leq l^p(k,i), \forall i \in N^p, u \in M^a$$

27: If the condition in Eq. 3 is satisfied for all tasks, the deployment is successful. Otherwise, revisit the allocation and assignment process.

28: Update the System:

29: Release the assigned resources from the cluster and update the global resourcespace accordingly.

30: Mark the assigned resources as unavailable in the cluster.

31: Update the load balancer and cluster management for the released resources.

32: Repeat the process for the remaining application deployment requests in the queue.

3.3.2. Algorithm Description

As shown in the algorithm, Steps 1 to 5 involve initializing necessary variables, data structures, and parameters required for the algorithm. Includes setting up the clusters and their associated physical machines (PM) and establishing the global resource space. Step 6 is the Application Deployment Request. The algorithm receives an application deployment request and associated parameters in these steps. The request contains information about the application that needs to be deployed. Steps 8-10 are the Register and Assign Request. The algorithm sends the deployment request to the registry and service discovery system (RSD) for registration and assignment in these steps. The RSD registers the request and assigns it to the load balancer, which handles the distribution of tasks. Steps 11-16, Resource Allocation and Task Assignment; These steps involve calculating the cost for each PM in the cluster for the given application using the cost equation (Eq. 1). The algorithm selects the best candidate PM from the

https://doi.org/10.31436/iiumej.v26i2.3089

available resources in the global resource space based on factors like cost and other considerations. And the application's tasks are then assigned to the selected PM, and the resource allocation is updated. The following available resource is also released from the cluster's resource pool. Steps 17-19, Cluster Management and Load Balancing; In these steps, the algorithm manages the microservice controllers and execution containers on the assigned PM. If there are other jobs in the queue, the algorithm assigns them to subsequent clusters that are available concurrently, following the same process. Steps 20-23, Scalability Check; The algorithm calculates the total resource utilization across all clusters for the assigned application using the scalability equation (Eq. 2). It checks if the resource utilization meets the predefined scalability requirement $RR^{p(i)}$ for all clusters. If the condition is satisfied, the algorithm proceeds to the next step. Otherwise, it revisits the allocation and assignment process to optimize resource utilization. Steps 24-27, Execution Time Calculation; These steps involve calculating the execution time for each task on the assigned PM using the execution time equation (Eq. 3). The algorithm considers the processing time for the task and the communication time between tasks. It checks if the execution time constraint is satisfied for all functions. If yes, the deployment is considered successful. Otherwise, the algorithm revisits the allocation and assignment process to optimize task execution time. Steps 28-31, Update the System; Once the deployment is successful, this step involves releasing the assigned resources from the cluster and updating the global resource space accordingly. The assigned resources are marked as unavailable in the cluster, and the load balancer and cluster management are updated for the released resources. Step 32, Repeat for Remaining Deployment Requests: The algorithm repeats the process for the remaining application deployment requests in the queue. It continues deploying applications and optimizing resource allocation until all requests have been processed.

Overall, the OEPTA algorithm performs a systematic and optimized placement and assignment of application tasks on available physical machines in clusters, considering cost, scalability, and execution time requirements. It aims to efficiently utilize resources and ensure the successful and optimized deployment of applications.

3.4. Evaluation

Through trace-driven simulation studies, we evaluate the performance of our OEPTA algorithm in various contexts. All evaluations are based on actual Google Cluster Traces. In various aspects, we compare the OEPTA algorithm to four strategies implemented in our paper, including EPTA, Spread, Binpack, and Random. The allocation and management of resources for applications are decentralized and performed by the load balancer on MSCs via the registry and service discovery system. An MSC decides on resource allocation, requests resources for ECs, monitors task status on ECs, and manages the life cycle of ECs. ECs complete the assigned tasks and report to the MSC on the status of their task execution in comparison to the expected progress. We evaluate and compare the performance of our microservice by creating job requests at random using non-load balancing and load balancing methods to display the execution time that results from the system's random distribution.

3.4.1. Load Balanced Vs Non-Load Balanced

Load-balanced microservice scheduling systems improve scalability, availability, and resilience, but add complexity and infrastructure. Non-load-balanced systems are less complicated, but they can introduce a single point of failure and do not provide the same level of scalability and resilience [30]. The choice between the two approaches will be determined by the system's specific needs and the acceptable trade-offs between complexity and availability.

https://doi.org/10.31436/iiumej.v26i2.3089

We first evaluate our algorithms randomly using non-load-balancing methods, then with a load-balancing system, to examine the algorithm's behavior in both cases.

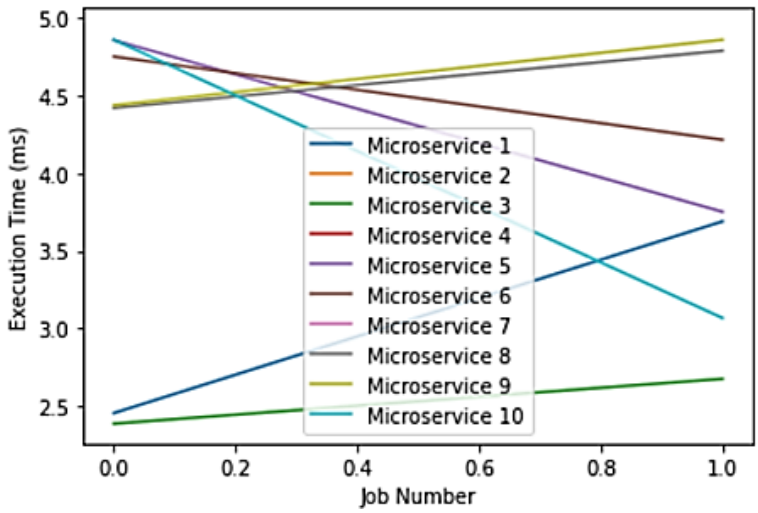


Figure 4. Microservice Non-load Balanced execution time with varied number of jobs.

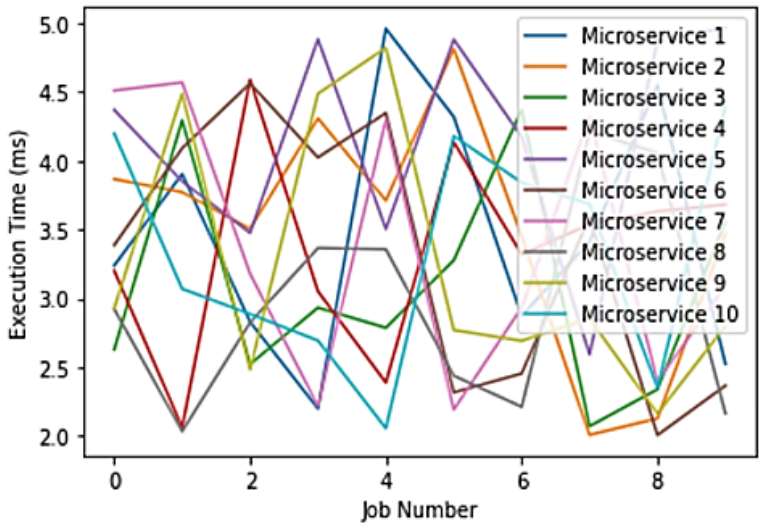


Figure 5. Microservice Load-balanced execution time with a varied number of jobs.

Fig. 4 shows the non-load-balanced algorithm used to test the number of microservices jobs against the execution time to see how the system behaved and changed compared to the load-balanced algorithm in Fig. 5.

Fig. 6 depicts the difference between non-load-balanced and load-balanced microservice job execution. We compared the number of microservices against the execution time of each microservice to see how the CBCM system behaved when load-balanced or not. Based on the figures above, load and non-load balancing are two distinct approaches to scheduling micro-services in a distributed system. It also shows that we require a load balancing system to reduce application deployment costs and execution time.

https://doi.org/10.31436/iiumej.v26i2.3089

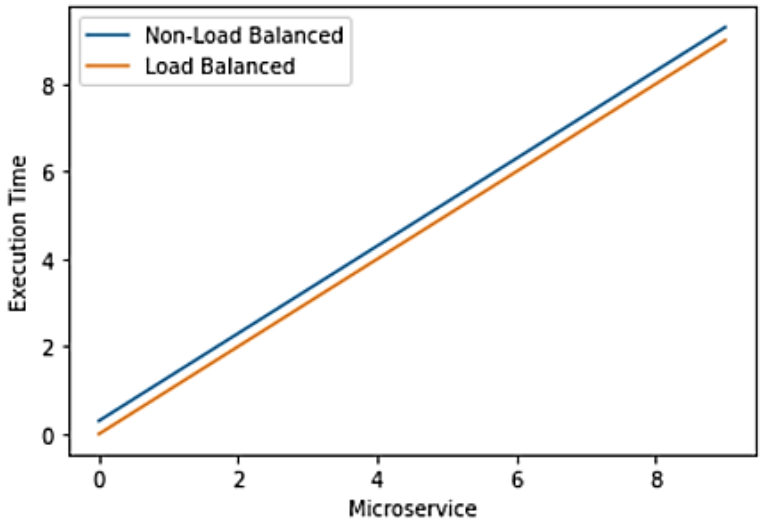


Figure 6. Microservice Load Balanced Vs Non-Load Balanced execution time with varied microservice

3.4.2. On the Number of Scaled Microservices

We assumed we had 200 microservices and used only 150 microservices (Active). The inactive will be pushed to cache load balancing as shown in Fig. 3. In case the computing resource requests overload the active microservices and have high traffic, then the system will up-scale from the cache memory (inactive) automatically. The higher the number of microservices available, the higher the active nodes. Our OEPTA algorithm's number of scaled microservices is being compared to four strategies. We looked at the active rate in the network and the total number of active microservices, while the number of available microservices ranges from 60 to 140. Fig. 7 demonstrates that OEPTA outperforms other algorithms regarding microservice scalability, whereas Spread is the most expensive. The expenses associated with deploying applications using the three Docker swarm strategies are somewhere between those of OEPTA and EPTA. When the number of scaled microservices increases from 60 to 140, the total number of active microservices used by OEPTA decreases slightly. With more microservices available, finding a better PM to handle requests becomes more difficult. However, as illustrated in the figure, the number of scaled microservices used by other strategies and algorithms increases since they occupy more microservices.

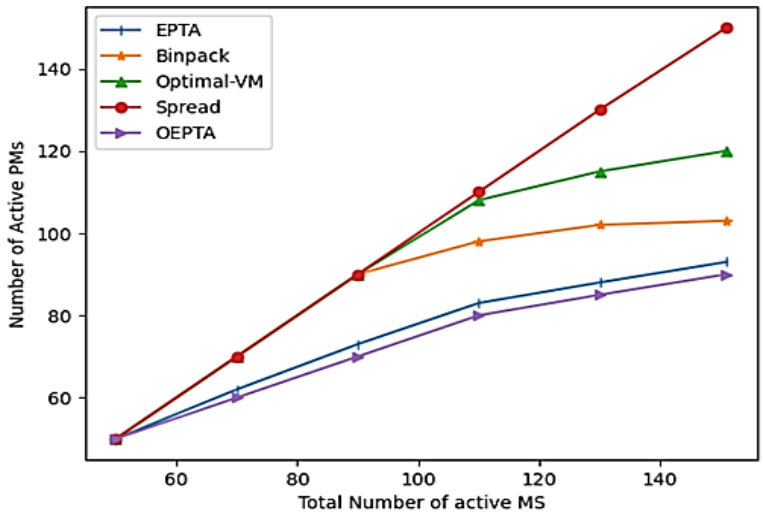


Fig. 7. Number of scaled microservices with a varied number of active microservices

https://doi.org/10.31436/iiumej.v26i2.3089

3.4.3. On the Number Application Deployment Cost

Fig. 8 shows the costs of deploying applications using five strategies and algorithms. The data reveals that OEPTA has the lowest cost, while ETPA has a considerably higher cost. Binpack has the lowest deployment cost among the three Docker Swarm strategies, while spread has the highest. This is because Binpack places microservices from the same application on the same physical machine, whereas Spread and Random distribute microservices across the network for load balancing. As a result, Binpack reduces the communication costs between microservices. The data in the figure indicates that OEPTA outperforms other algorithms in terms of total deployment cost, whereas Optimal-VM has the highest deployment cost. The application deployment costs of the three Docker Swarm strategies fall between those of OEPTA and EPTA. As the number of microservices increases from 60 to 140, the total deployment costs of OEPTA slightly decrease. This is because a larger pool of microservices increases the likelihood of finding a better microservice to place. Conversely, the total deployment costs of other strategies and algorithms increase because they utilize more microservices.

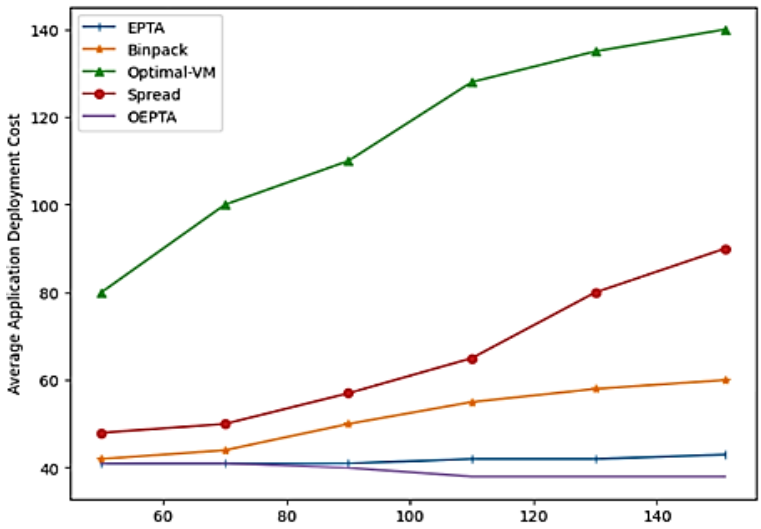


Figure 8. Average application deployment cost with a varied number of microservices.

3.4.4. On the number of Execution time

As depicted in Fig. 9, utilizing the linear programming (LP) solver causes a rise in the execution time of EPTA and Optimal-VM as the number of microservices within an application grows. To improve performance, we have devised a specialized solver customized to their problem and replacing the typical LP solver used in the EPTA algorithm. This new solver is controlled by a load balancer, resulting in optimized execution time. Fig. 9 compares the execution times of the five strategies and algorithms. The figure indicates that as the number of microservices grows, Optimal-VM and EPTA algorithms exhibit slightly better performance than OEPTA. Among the four, Spread has the shortest execution time. Interestingly, as the execution time increases from 0.2 to 2.0, the total number of microservices in OEPTA decreases compared to the other algorithms. This is because Optimal-VM and EPTA employ linear problems and the LP solver. However, OEPTA incrementally expands the search area instead of taking the entire physical network as input, resulting in a time complexity that does not exponentially increase as the network scales.

https://doi.org/10.31436/iiumej.v26i2.3089

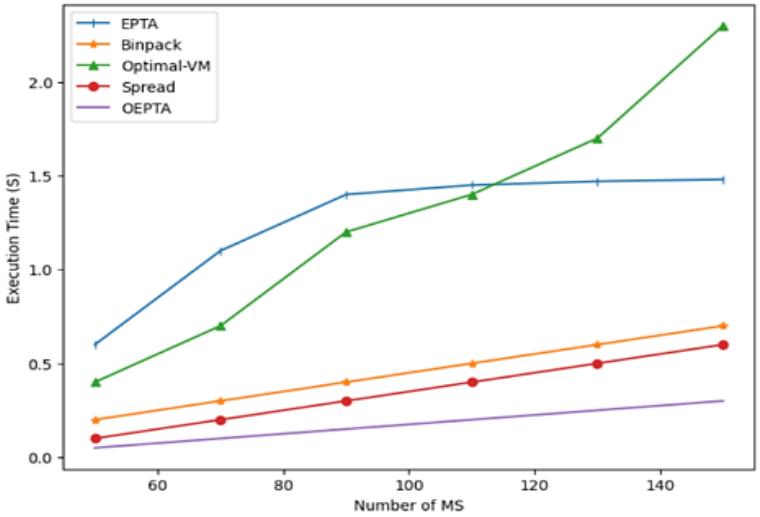


Figure 9. Average execution time with a varied number of microservices.

The paper employs a benchmarking approach to evaluate the proposed methodology against four Docker Swarm strategies and a Hypervisor-based VM embedding algorithm. It emphasizes state-of-the-art results in container microservice cloud-based systems, covering deployment costs, scaled microservices, and execution time (Wan et al. 2018). Detailed comparisons with a VM placement algorithm and three Docker Swarm strategies, utilizing real data traces, validate the proposed schemes. The paper explores strategies and algorithms, assessing deployment costs and overall performance. The OEPTA algorithm outperforms EPTA and other strategies in deployment cost, scaled microservices, and execution time, as shown in Table 2.

Table 2. Algorithm Comparison Results

Algorithms	Microservice Scalability	Deployment Cost	Execution Time
OEPTA	Outperforming EPTA slightly decreased as the total number of active microservices increased (see Figure 7).	It demonstrates the lowest cost compared to all strategies. Deployment costs decrease slightly with more microservices (refer to Figure 8).	Demonstrates competitive execution time performance, showcasing optimized execution time and successful deployment with efficient task execution (refer to Figure 9).
EPTA	Demonstrates higher microservice scalability costs than OEPTA (refer to Figure 7).	It is considered to have a higher deployment cost than OEPTA (refer to Figure 9).	It demonstrates higher execution time, which is attributed to the utilization of a linear programming (LP) solver compared to OEPTA (refer to Figure 9).
Optimal-VM	Demonstrates the highest microservice scalability among the three strategies (Binpack, EPTA, and OEPTA), but is still lower than spread. Consequently, the number of active microservices in the OEPTA algorithm surpasses all four strategies (refer to Figure 7).	Proven to have the highest total deployment cost; however, OEPTA surpasses other strategies, including Optimal-VM (refer to Figure 8).	Deploys linear problems and an LP solver, causing an escalation in execution time with the growing number of microservices, demonstrating slightly inferior performance compared to OEPTA (see Figure 9).
Binpack	Compared to EPTA and OEPTA strategies, more scaled microservices with	Demonstrates optimization of deployment costs by consolidating microservices	It exhibits the shortest execution time compared to EPTA and Optimal-VM strategies

	varying active microservices were demonstrated. Nevertheless, it exhibits lower scalability than the Optimal-VM and Spread strategies (see Figure 7).	from the same application on the same physical machine, leading to a lower deployment cost than Spread and Optimal-VM but higher than OEPTA (see Figure 8).	while displaying higher execution time than Spread and OEPTA (see Figure 9).
Spread	It demonstrates higher microservices scalability among all four strategies (see Figure 7). It distributes microservices across the network for load balancing, increasing deployment costs between microservices.	Compared to Binpack, higher deployment costs are the highest among the three Docker Swarm strategies (see Figure 8).	Demonstrates the shortest execution time compared to Binpack, EPTA, and Optimal-VM, whereas OEPTA emerges as the shortest among these compared to Spread (see Figure 9).

Table 2 presents a comparison with Wan et al. (2018), illustrating that our newly developed optimized algorithm (OEPTA) surpasses the adapted (EPTA) algorithm in terms of microservice scalability, deployment cost, and execution time. This highlights the efficacy of the OEPTA algorithm, emphasizing its superior scalability, optimized execution time, and successful deployment for efficient executional tasks, surpassing other strategies and algorithms in this domain.

4. CONCLUSION AND FUTURE WORK

We developed the OEPTA algorithm and a cutting-edge framework that minimizes the cost of deploying microservice applications while addressing the issue of load-balanced microservice scheduling systems. We use load balancing tools and libraries to describe our framework thoroughly, and we described our improvement in terms of pertinent metrics like execution time, application deployment cost, and scalability. We created a load balancing algorithm for container microservice scheduling optimization to ascertain the service deployment cost, reliability, and availability of the microservice application. The distribution and control of resources for applications occur in a decentralized manner. By conducting a comparative analysis, we confirmed the effectiveness of the proposed strategies. We found that the OEPTA algorithm delivered good results in terms of optimizing costs, traffic spikes, and server overload. These measures can effectively balance user requests for deploying applications and enhance the performance of the cloud-based container microservice system. In future work, we will incorporate load balancing and auto-scaling features by leveraging a multi-objective algorithm. This approach will consider additional optimization objectives, including latency and CPU utilization. By doing so, we aim to address more refined challenges and enhance the overall performance of container-based microservice cloud systems, ultimately improving users' quality of service (QoS).

ACKNOWLEDGMENT

The authors would like to acknowledge and thank the Research University Grant (1011/PKOMP/8014076) from Universiti Sains Malaysia (USM) and the Talent and Publications Enhancement Research Grant (TAPERG/2023/UMT/2223) titled “Empirical Analysis of Software Maintainability Metrics in DevOps Environments” from Universiti Malaysia Terengganu (UMT) for supporting this publication.

REFERENCES

[1] O. Al-Debagy and P. Martinek, "A Comparative Review of Microservices and Monolithic Architectures," *18th IEEE Int. Symp. Comput. Intell. Informatics, CINTI 2018 - Proc.*, pp. 149–154, 2018, doi: 10.1109/CINTI.2018.8928192.

[2] L. De Lauretis, "From monolithic architecture to microservices architecture," *Proc. - 2019 IEEE 30th Int. Symp. Softw. Reliab. Eng. Work. ISSREW 2019*, pp. 93–96, 2019, doi: 10.1109/ISSREW.2019.00050.

[3] F. Ponce, G. Marquez, and H. Astudillo, "Migrating from monolithic architecture to microservices: A Rapid Review," *Proc. - Int. Conf. Chil. Comput. Sci. Soc. SCCC*, vol. 2019-Novem, 2019, doi: 10.1109/SCCC49216.2019.8966423.

[4] M. Villamizar, O. Garcés, H. Castro, M. Verano, L. Salamanca, and S. Gil, "Evaluating the Monolithic and the Microservice Architecture Pattern to Deploy Web Applications in the Cloud Evaluando el Patrón de Arquitectura Monolítica y de Micro Servicios Para Desplegar Aplicaciones en la Nube," *10th Comput. Colomb. Conf.*, pp. 583–590, 2015.

[5] T. Erl, J. Fontenla, M. Caeiro, and M. Llamas, "Web Services and Contemporary SOA," *Serv. A rch itectu re Connceptsts , Technol. , Des.*, pp. 25–81, 2005.

[6] D. Bhamare, M. Samaka, A. Erbad, R. Jain, and L. Gupta, "Exploring microservices for enhancing internet QoS," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 11, 2018, doi: 10.1002/ett.3445.

[7] A. Sundberg, "A study on load balancing within microservices architecture," 2019, [Online]. Available: https://www.mendeley.com/catalogue/a4814e2d-827e-3e18-93b5-3f91efa6d98b/?utm_source=desktop&utm_medium=1.19.4&utm_campaign=open_catalog&userDocumentId=%7B0ed39c18-97ea-4c9c-994a-2dd841333607%7D

[8] J. A. Valdivia, X. Limon, and K. Cortes-Verdin, "Quality attributes in patterns related to microservice architecture: a Systematic Literature Review," pp. 181–190, 2020, doi: 10.1109/conisoft.2019.00034.

[9] Z. Ding, S. Wang, and M. Pan, "QoS-Constrained Service Selection for Networked Microservices," *IEEE Access*, vol. 8, pp. 39285–39299, 2020, doi: 10.1109/ACCESS.2020.2974188.

[10] M. Villamizar *et al.*, "Cost comparison of running web applications in the cloud using monolithic, microservice, and AWS Lambda architectures," *Serv. Oriented Comput. Appl.*, vol. 11, no. 2, pp. 233–247, 2017, doi: 10.1007/s11761-017-0208-y.

[11] N. Viennot, M. Lécuyer, J. Bell, R. Geambasu, and J. Nieh, "Synapse: A microservices architecture for heterogeneous-database web applications," *Proc. 10th Eur. Conf. Comput. Syst. EuroSys 2015*, 2015, doi: 10.1145/2741948.2741975.

[12] V. Singh and S. K. Peddoju, "Container-based microservice architecture for cloud applications," *Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2017*, vol. 2017-Janua, pp. 847–852, 2017, doi: 10.1109/CCAA.2017.8229914.

[13] J. Mathenge, "Containers vs Microservices: What's The Difference?," *Https://Www.Bmc.Com/Blogs/Containers-Vs-Microservices/*, 2021.

[14] C. Pahl, A. Brogi, J. Soldani, and P. Jamshidi, "Cloud container technologies: A state-of-the-art review," *IEEE Trans. Cloud Comput.*, vol. 7, no. 3, pp. 677–692, 2019, doi: 10.1109/TCC.2017.2702586.

[15] E. Casalicchio and V. Perciballi, "Auto-Scaling of Containers: The Impact of Relative and Absolute Metrics," *Proc. - 2017 IEEE 2nd Int. Work. Found. Appl. Self* Syst. FAS*W 2017*, pp. 207–214, 2017, doi: 10.1109/FAS-W.2017.149.

[16] M. D. Cojocaru, A. Oprescu, and A. Uta, "Attributes assessing the quality of microservices automatically decomposed from monolithic applications," *Proc. - 2019 18th Int. Symp. Parallel Distrib. Comput. ISPDC 2019*, no. June, pp. 84–93, 2019, doi: 10.1109/ISPDC.2019.00021.

[17] S. N. Srirama, M. Adhikari, and S. Paul, "Application deployment using containers with auto-scaling for microservices in cloud environment," *J. Netw. Comput. Appl.*, vol. 160, no. August 2019, 2020, doi: 10.1016/j.jnca.2020.102629.

[18] S. Jain and A. K. Saxena, "A survey of load balancing challenges in cloud environment," *Proc.*

5th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2016, pp. 291–293, 2017, doi: 10.1109/SYSMART.2016.7894537.

[19] X. Wan, X. Guan, T. Wang, G. Bai, and B. Choi, "Journal of Network and Computer Applications Application deployment using Microservice and Docker containers : Framework and optimization," vol. 119, no. December 2017, pp. 97–109, 2018.

[20] M. Lin, J. Xi, W. Bai, and J. Wu, "Ant Colony Algorithm for Multi-Objective Optimization of Container-Based Microservice Scheduling in Cloud," *IEEE Access*, vol. 7, pp. 83088–83100, 2019, doi: 10.1109/ACCESS.2019.2924414.

[21] K. Li, G. Xu, G. Zhao, Y. Dong, and D. Wang, "Cloud task scheduling based on load balancing ant colony optimization," *Proc. - 2011 6th Annu. ChinaGrid Conf. ChinaGrid 2011*, pp. 3–9, 2011, doi: 10.1109/ChinaGrid.2011.17.

[22] X. Guan, X. Wan, B. Choi, S. Song, and J. Zhu, "Application Oriented Dynamic Resource Allocation for Data Centers Using Docker Containers," vol. 1, no. c, pp. 1–4, 2016, doi: 10.1109/LCOMM.2016.2644658.

[23] A. Balalaie, A. Heydarnoori, and P. Jamshidi, "Migrating to Cloud-Native architectures using microservices: An experience report," *Commun. Comput. Inf. Sci.*, vol. 567, pp. 201–215, 2016, doi: 10.1007/978-3-319-33313-7_15.

[24] M. Bravetti, S. Giallorenzo, J. Mauro, I. Talevi, and G. Zavattaro, *Optimal and automated deployment for microservices*, vol. 11424 LNCS. Springer International Publishing, 2019. doi: 10.1007/978-3-030-16722-6_21.

[25] P. Stefanic, M. Cigale, A. Jones, and V. Stankovski, "Quality of Service Models for Microservices and Their Integration into the SWITCH IDE," *Proc. - 2017 IEEE 2nd Int. Work. Found. Appl. Self* Syst. FAS*W 2017*, no. September, pp. 215–218, 2017, doi: 10.1109/FAS-W.2017.150.

[26] B. Stevant, J. L. Pazat, and A. Blanc, "Optimizing the Performance of a Microservice-Based Application Deployed on User-Provided Devices," *Proc. - 17th Int. Symp. Parallel Distrib. Comput. ISPDC 2018*, pp. 133–140, 2018, doi: 10.1109/ISPDC2018.2018.00027.

[27] T. Tupid, "Basic Guide: Load Balancing and Auto-Scaling in Cloud Computing," 2019. https://medium.com/@tudip/basic-guide-load-balancing-and-auto-scaling-in-cloud-computing-219a5f0768a

[28] Z. Ni, C. Wei, T. Wood, and N. Choi, "A SmartNIC-based Load Balancing and Auto Scaling Framework for Middlebox Edge Server," pp. 21–27, 2022, doi: 10.1109/nfv-sdn53031.2021.9665167.

[29] I. Lera and C. Juiz, "Genetic Algorithm for Multi-Objective Optimization of Container Allocation in Cloud Architecture," 2017.

[30] D. A. Shafiq, N. Z. Jhanjhi, and A. Abdullah, "Load balancing techniques in cloud computing environment: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, pp. 3910–3933, 2022, doi: 10.1016/j.jksuci.2021.02.007.

https://doi.org/10.31436/iiumej.v26i2.3175

MODELLING OF THE PI-SHAPE LOW CONCENTRATING PHOTOVOLTAIC SOLAR CELLS

AINUR KAPPAROVA, SAYAT ORYNBASSAR, GULBAKHAR DOSYMBETOVA*, DINARA ALMEN, EVAN YERSHOV, AHMET SAYMBETOV, MADIYAR NURGALIYEV, NURZHIGIT KUTTYBAY, NURDAULET ALGAZIN

Faculty of Physics and Technology, Al-Farabi Kazakh National University, 71 Al-Farabi, Almaty 050040, Kazakhstan

**Corresponding author: dossymbetova.g@kaznu.kz*

(Received: 19 January 2024; Accepted: 19 March 2025; Published online: 15 May 2025)

ABSTRACT: One of the cheapest ways to improve photovoltaic (PV) systems is to create LCPV systems with polycrystalline silicon solar cells, which require less cost and have high optical efficiency. Additional reflective mirrors were added to improve the optical efficiency of the low concentrating photovoltaic (LCPV) system based on a Fresnel lens. Pi-shaped LCPV cells were obtained and compared with an ordinary LCPV based on a Fresnel lens. The proposed LCPV shows high optical efficiency even at 50 mm of cell-lens distance, while the ordinary LCPV presents a maximum of 30% of optical efficiency. The concentration ratio of 8 suns can be achieved at 150 mm of cell-lens distance at the range ±20° of the incidence angle. When the cell-lens distance is 100 mm or 125 mm, the optical efficiency is more than 80%, and the concentration ratio (CR) is more than 2 suns at the range of incidence angle ±25°. The proposed LCPV design helps to work the system at ±25° without the help of a solar tracking system. Hence, when developing the LCPV system, increasing the acceptance angle might reduce the work of solar tracking systems and the tracking errors. Good irradiance uniformity can be achieved, and the acceptance angle can be increased.

ABSTRAK: Salah satu cara termurah bagi menambah baik sistem fotovoltaik (PV) adalah dengan mencipta sistem LCPV menggunakan sel solar silikon polihabluran, di mana kos lebih rendah dan kecekapan optik tinggi. Bagi meningkatkan kecekapan optik fotovoltaik rendah tumpuan (LCPV) berasaskan kanta Fresnel, cermin pantulan tambahan diperlukan dan bentuk Pi LCPV diperolehi dan dibanding dengan LCPV biasa berasaskan kanta Fresnel. LCPV yang dicadangkan ini menunjukkan kecekapan optik tinggi walau pada jarak 50 mm antara sel dan kanta, manakala LCPV biasa mencapai kecekapan optik maksimum sebanyak 30%. Nisbah tumpuan sebanyak 8 kali pencahayaan matahari dapat dicapai pada jarak 150 mm antara sel dan kanta dalam julat sudut kejadian ±20°. Apabila jarak antara sel dan kanta sebanyak 100 - 125 mm, kecekapan optik adalah melebihi 80% dan nisbah tumpuan (CR) melebihi 2 pencahayaan matahari dalam julat sudut kejadian ±25°. LCPV yang dicadangkan ini dapat membantu sistem beroperasi pada julat sudut ±25° tanpa bantuan sistem penjejak suria. Oleh itu, dengan meningkatkan sudut penerimaan sistem LCPV semasa mencipta sistem, ini berkemungkinan mengurangkan keperluan sistem penjejak suria dan mengurangkan ralat penjejak, mencapai taburan pencahayaan seragam, serta meningkatkan sudut penerimaan secara keseluruhan.

KEY WORDS: *LCPV, Optical efficiency, Fresnel lens, Concentration ratio, Incidence angle*

1. INTRODUCTION

Solar energy is becoming significantly crucial among renewable sources due to its low price and affordability [1, 2]. However, the electrical efficiency of photovoltaic systems today is insufficient to switch to green energy massively, and high-efficiency solar panels are expensive [3]. Researchers are working on increasing the generated energy by PV panels. Several factors affect the performance of PV efficiency, and they are increasing the amount of incident sunlight, cooling the solar cells, and using solar cells with high efficiency [4, 5]. The solar cell's temperature is reduced by applying a cooling system [6]. Using high-efficiency multi-junction solar cells is more expensive and requires a high-precision solar tracking system. There are many types of solar cells [7], and about 90% of the solar cells worldwide are crystalline silicon solar cells [8]. It is considered that one of the cheapest solar cells is polycrystalline silicon solar cells [9]. However, increasing the solar cells' generated energy is a significant task, and several methods exist to solve the problem.

One of the practical methods to increase the incident solar radiation is using solar concentrator optics [10]. Many optical designs for solar concentrators might use mirrors, lenses, and other reflective or refractive optical elements. The paper [11] improved the PV panel efficiency using three mirrors and a cooling system. It proved that the efficiency of solar cells is greatly affected by the amount of solar irradiance. The work [12] modeled the performance of a V-trough concentrating system with a dual-axis solar tracking system. As a result, the overall energy output was increased by 86%, with contributions to both tracking (25%) and concentration (50%). It shows that applying a concentrating optical element improves the system's performance.

According to the concentrating degree, current concentrating photovoltaic (CPV) cells are divided into low, medium, and high CPV cells [13]. Optical elements of high-concentration photovoltaic (HCPV) systems concentrate sunlight into one small point to achieve a higher concentration ratio (more than 300 suns), which leads to the use of small solar cells, such as multi-junction solar cells, which are not affordable. Low concentrating photovoltaic (LCPV) and medium concentrating photovoltaic (MCPV) systems can be designed with cheaper silicon solar cells [14]. LCPV systems benefit from their simplicity and affordability.

Low concentrating photovoltaic (LCPV) systems, which have a 1:10 concentration ratio, are considered simple in optics, can be made with affordable materials, and do not require a high-accuracy solar tracking system. Currently known LCPV optics [15] are V-trough [16], parabolic concentrators [17], cylindrical troughs, and linear Fresnel reflectors [18]. The paper [19] has designed a 7-sun low-concentration CPV system based on compound parabolic concentrators. The work [20] showed that 2x V-Trough LCPV with a seasonal tracking system can generate 44% higher energy output than a flat PV panel. Much research has been done on low-concentration photovoltaic systems with different designs of concentrating elements [21, 22].

One of the concentrating elements for LCPV is Fresnel lenses, and today, some work is being done on improving LCPVs based on Fresnel lenses. The works [23, 24] showed that when using a Fresnel lens as a concentrator for polycrystalline solar cells, it is possible to get an LCPV panel that can generate 27% more energy than a non-concentrated silicon solar panel. It is certain that when using a Fresnel lens, the hot spot occurs in the center of the solar cell, which degrades its reliability and conversion efficiency. To mitigate damage caused by hot spots and enhance the acceptance angle, Fresnel concentrators typically incorporate an additional concentrator called a secondary optical element (SOE). This integration aims to

https://doi.org/10.31436/iiumej.v26i2.3175

improve the acceptance angle and achieve a more uniform energy distribution, thereby addressing hot spots and ensuring efficient sunlight concentration [25].

Another paper [26] presented a complete optical modeling to improve the simulation of an ordinary Fresnel lens with a refractive secondary optical element (SOE) and found that the pyramid has high optical efficiency and a more uniform irradiance distribution. The CPC has a large acceptance angle but shows the least uniformity. However, most of the work is done for HCPV systems with multi-junction solar cells, and those systems need a solar tracker because they cannot work at high incidence angles [27]. Their optical efficiency is 90% up to an incident angle of 0.4°, then reduced to 80% at 0.6°, and then to 10% at 1° [28]. This work will use polycrystalline silicon solar cells, which are several times cheaper than multi-junction solar cells. Reflective surfaces will be used as additional optical elements.

Increasing the acceptance angle refers to the ability of the concentrator to capture sunlight from a wider range of incidence angles [29]. The problem is that when the incidence angle increases, the sunlight flux on the solar cell decreases, reducing the number of photons on the solar cell and the output current of the cell array. To overcome the problem, a solar tracking system is usually applied, and the concentrating PV systems will be oriented to the Sun. However, the solar tracking systems require external power consumption, affecting system costs. Therefore, we propose reducing the work of solar tracking by concentrating optics. To do this, optimal parameters of the optical system must be found. This work proposes a design of LCPV based on a Fresnel lens and Pi-shaped reflective surfaces with nine silicon solar cells, which can reduce the work of a solar tracking system at a specific range of incidence angles. The optimum distance between the lens and the solar cell is significantly affected by the relation between the concentrating optics and the solar cell [30]. The proposed system is compared with an LCPV with only a Fresnel lens without mirrors and a single solar cell. The novelty of the work is to increase the optical efficiency and widen the incidence angle of ordinary Fresnel-based LCPV using reflective mirrors. One of the study's aims is to identify the optimal cell-lens distance and calculate the optical efficiency at different incident angles. Due to the proposed design, the low-concentration optical system can show good optical performance at higher incidence angles.

2. METHODOLOGY

The solar cell and the optics must be compatible within the module [30]. The solar cell used in this work is a polycrystalline silicon solar cell. The commercial silicon solar cells are used under a concentration from 2 suns to 10 suns, as the electrical performance of a polycrystalline silicon solar cell is limited by the recombination losses, which limit the open-circuit voltages. The losses increase proportionately at higher illumination [31]. Our previous works [23, 24] proved that a polycrystalline silicon solar cell can work with a Fresnel lens at an optimal distance shown in Figure 1(a) and convert 27% more electrical energy than a non-concentrated solar cell. In this work, the modified structure of the system is proposed.

Figure 1 (b) shows this work's proposed Pi-shape LCPV design. It consists of a Fresnel lens and four (4) reflective surfaces. The LCPV system has nine (9) solar cells, which can help harvest more sunlight from inclined rays compared to a single solar cell at wider incidence angles.

https://doi.org/10.31436/iiumej.v26i2.3175

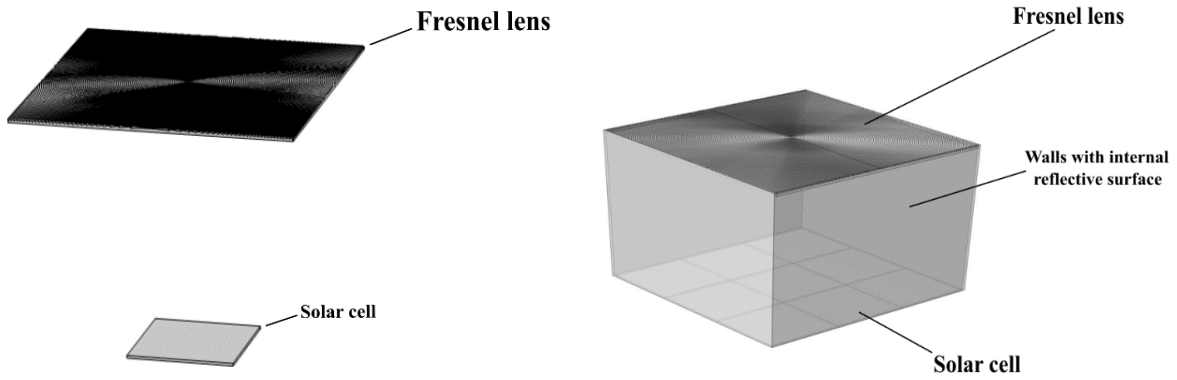


Figure 1. (a) A solar cell with a Fresnel lens, (b) the Pi-shape system based on a Fresnel lens and a reflective surface with nine solar cells.

LCPV systems with concentrating optical systems based on Fresnel lenses require a solar tracking system because Fresnel lenses have a point focus. This means that the systems cannot work from a wider angle of incidence. The concentrated sunlight spot is to the side of the solar cell; therefore, adding additional solar cells around the main solar cell helps prevent the concentrated sunlight at high incidence angles. The cost of polycrystalline solar cells is not expensive; therefore, using nine solar cells instead of one doesn't affect the total cost of the system much, since using a single solar cell with a Fresnel lens requires a high-accuracy solar tracking system, which costs more than fixed solar panels [32].

Moreover, reflective mirrors do not show chromatic aberration, and high optical efficiencies can be achieved [33]. Therefore, reflective surfaces redirect the shifted sunlight flux to the solar cells. The size of the reflective mirrors depends on the cell-lens distance of the system.

COMSOL Multiphysics is used in simulation to assess the proposed system. The geometrical parameters of the system used in COMSOL Multiphysics are presented in Table 1.

Table 1. Parameters of the LCPV system

Types of Vehicle	Value
Size of the Fresnel lens, mm^2	150×150
Reflective index of the mirror	1
Size of the Solar cell, mm^2	50×50

The simulation illuminated the CPV system under a DNI of 1000 W⁄m2. A 50×50 mm polycrystalline silicon solar cell is used. The size of the Fresnel lens was 150×150 mm, and the optimal size of reflective surfaces had to be found depending on the system's optical performance at different cell-lens distances. To simulate the model, 5776 rays were set to produce the results. The ray-tracing method was used to simulate and optimize lenses and reflectors.

Figure 2 presents the work process in COMSOL Multiphysics as a block diagram. In this work, a ray tracing simulation was conducted using the Geometrical Optics module of the COMSOL Multiphysics software package. In the Geometrical Optics module, electromagnetic waves are treated as rays, and ordinary differential systems are solved, determining each ray's location and wave vector for individual parts of the 3-D model.

https://doi.org/10.31436/iiumej.v26i2.3175

The boundary conditions for the simulation were: 1) for the Fresnel lens, it was a refractive interface that allows light refraction and focusing and works according to Snell's Law; 2) mirror reflection was used for reflective mirrors, and the purpose was to reflect rays toward the cells; 3) an absorbing surface was used for detectors to capture the rays. Moreover, a steady-state method was used for ray-tracing simulations, which does not include any dynamic changes in optical properties. Ray tracing uses numerical calculations to model and is performed by solving Hamilton's equations.

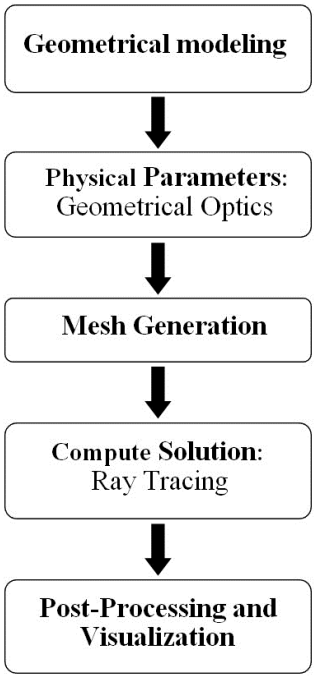


Figure 2. Block diagram of COMSOL Multiphysics.

To evaluate optical performance, the optical system is tested at different incidence angles, cell-lens distances, and geometrical parameters [34]. Finding optimal optical efficiency and concentration ratio at different incidence angles is a significant task; therefore, to fully characterize a given optical system, it is necessary to assess optical efficiency and concentration ratio parameters. The optical efficiency η_{op} of a lens can be defined as the ratio of radiant power at its input aperture P_{in} which reaches its output P_{out} or, in our case, the ratio of the number of entered rays to the number of incident rays on the surface of the solar cells [33].

$$\eta_{op} = \frac{P_{in}}{P_{out}} = \frac{N_{ent}}{N_{exit}} \tag{1}$$

The geometrical concentration ratio is defined as the ratio of the input area A_{in} (in our case, the area evaluated at the lens's input aperture) to the output area A_{out} (i.e., the area evaluated at the receiver area) [34].

$$C_g = \frac{A_{in}}{A_{out}} \tag{2}$$

The optical concentration ratio of the system is calculated by the following equation and is measured with the sun:

$$CR = \eta_{op} \cdot C_g \tag{3}$$

https://doi.org/10.31436/iiumej.v26i2.3175

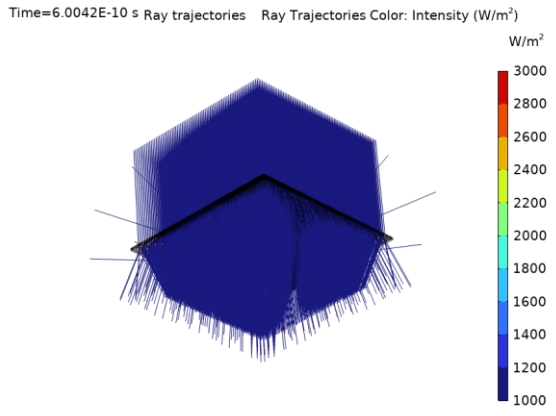
The optical parameters of the optical system will be calculated and analyzed using Eqs. (1)-(3).

3. RESULTS AND DISCUSSION

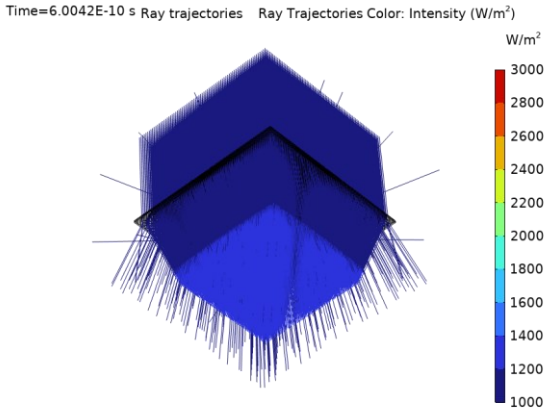
3.1. Results

Simultaneously, simulations in COMSOL Multiphysics for both systems were conducted to evaluate the proposed LCPV system and compare it with an ordinary LCPV with just a Fresnel lens. The results from the ray tracing simulation were plotted, as shown in Figure 3.

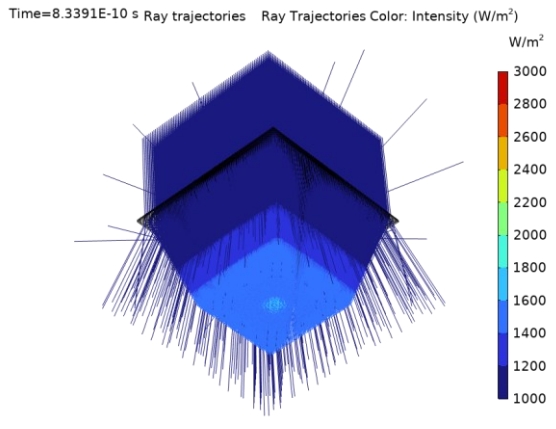
The incidence angle of the rays was zero degrees in this simulation. The results show that if the rays are perpendicular to the Fresnel lens surface with reflective mirrors, the central solar cell receives more concentrated sunlight as the cell-lens distance increases. The length of the light spot increases by decreasing the cell-lens distance. In the simulation, the results of an ordinary LCPV with only a Fresnel lens show that some rays are not concentrated. In contrast, in the proposed system, the rays are concentrated and shaped rectangularly due to reflective mirrors. The proposed LCPV system can collect all rays perpendicularly and at some incidence angles. Figure 4(e) shows that the Fresnel lens at 150 mm from the solar cell increases the solar radiation from 1000 W⁄m^2 to 3000 W⁄m^2. It means the optical concentrator can help to increase the sunlight power 3 times compared to without the concentrating element, since the proposed concentrating system gathers sunlight from a larger area and directs it to a smaller solar cell surface.



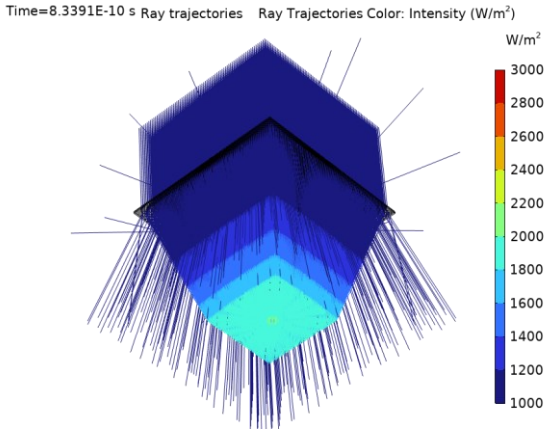
(a) 50 mm



(b) 75 mm

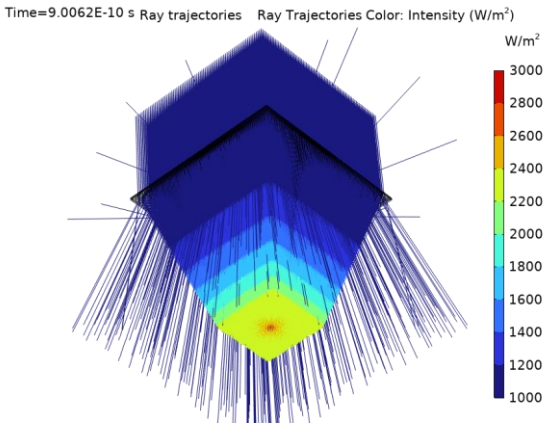


(c) 100 mm



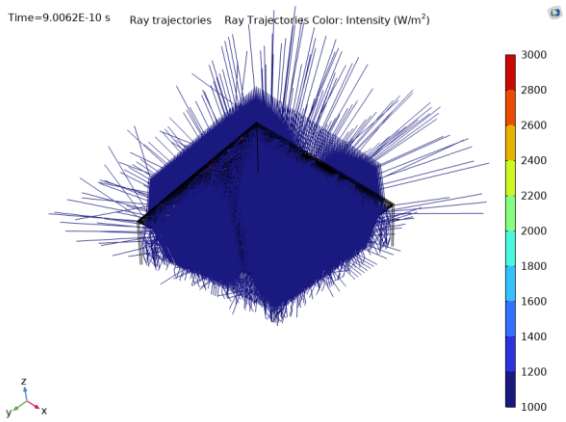
(d) 100 mm

https://doi.org/10.31436/iiumej.v26i2.3175

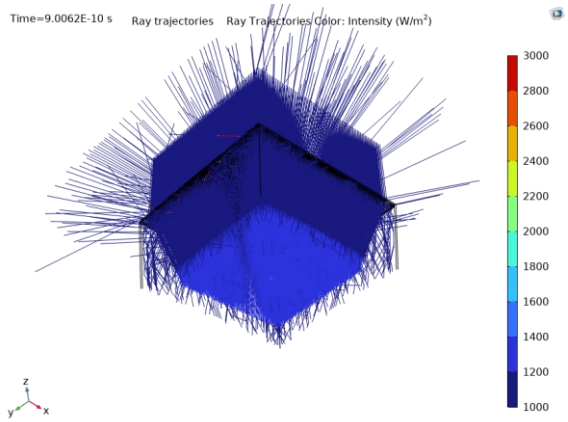


(e) 150 mm

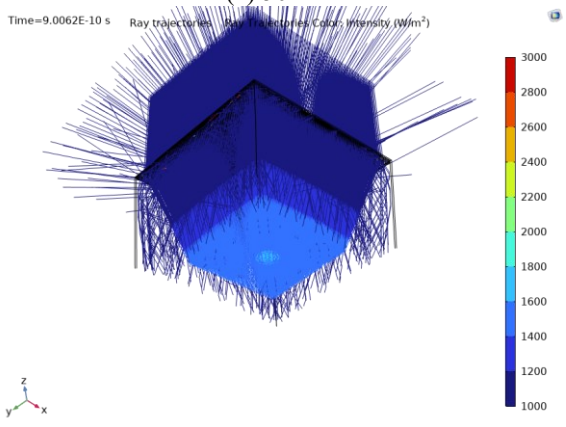
Figure 3. Ray trajectories at different cell-lens distances of ordinary optical systems



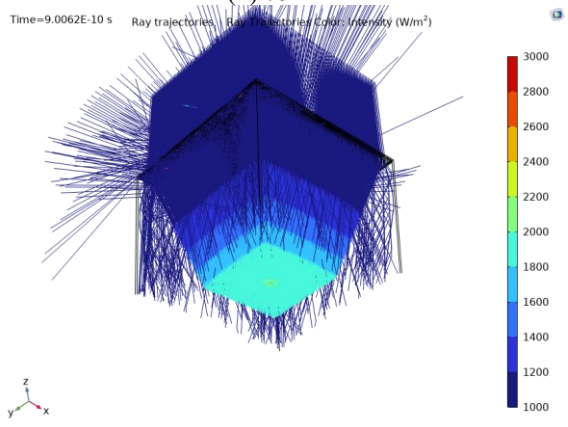
(a) 50 mm



(b) 75 mm

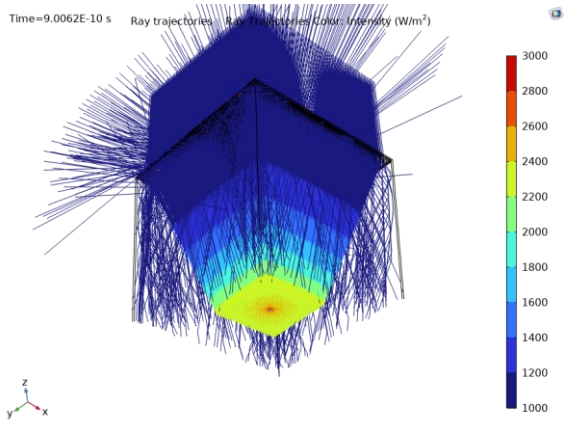


(c) 100 mm



(d) 100 mm

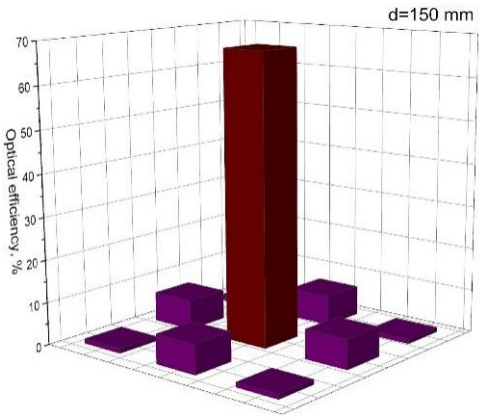
https://doi.org/10.31436/iiumej.v26i2.3175



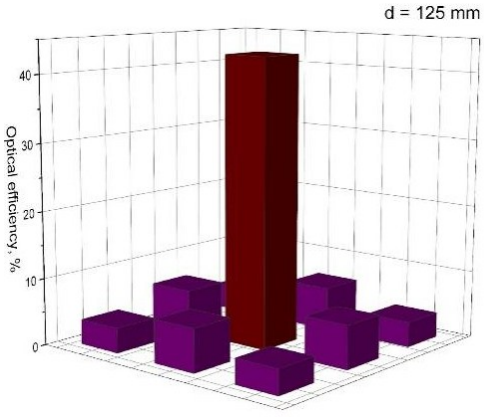
(e) 150 mm

Figure 4. Ray trajectories at different cell-lens distances of the proposed LCPV design.

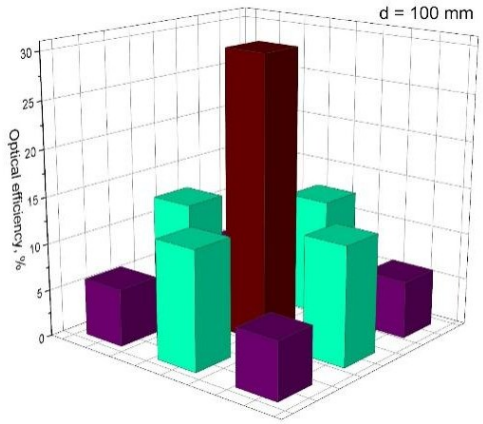
As the Fresnel lens has a point focus, the sunlight is concentrated more on the central part of the solar cells. Figure 5 shows the distribution of incident sunlight flux on the surface of solar cell arrays at a zero incidence angle. When the Fresnel lens is closer to the array, the sunlight is concentrated on the central solar cell and the other eight solar cells. The closer to the center, the more optical efficiency they gain. Solar radiation is mostly concentrated on the central solar cell at 150 mm (Figure 5 (a)) since the optical efficiency of the central solar cell is higher than that of the other solar cells. At 125 mm (Figure 5 (b)), the optical efficiency of the central solar cell is still much higher than that of the other cells, but lower than at 150 mm. At 100, 75, and 50 mm, the optical efficiency of the four solar cells around the central one is higher than that of the four solar cells at the four corners of the module (Figure 5 (c, d, e)).



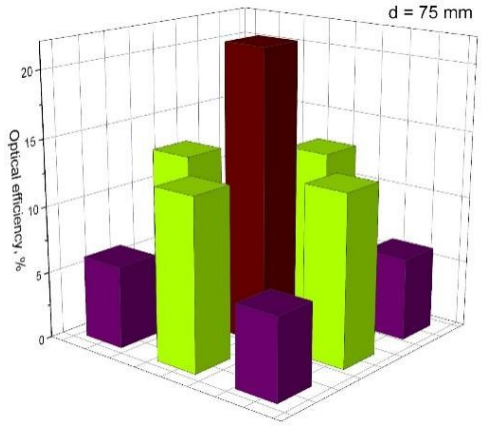
(a) 50 mm



(b) 75 mm

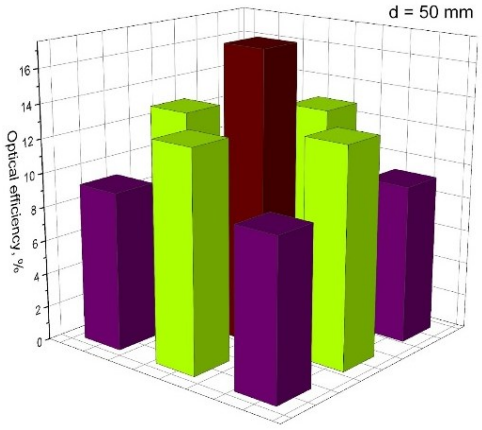


(c) 100 mm



(d) 100 mm

https://doi.org/10.31436/iiumej.v26i2.3175

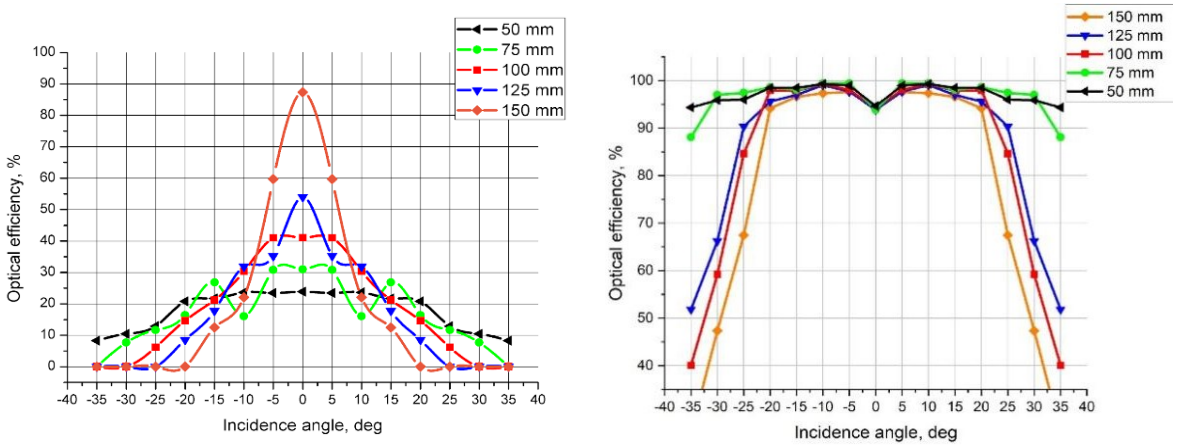


(e) 150 mm

Figure 5. Optical efficiency in the area of nine solar cells at different cell-lens distances.

As the solar cell is located at the center of the Fresnel lens, there is always high optical efficiency at a zero incidence angle. Both ordinary and proposed systems were simulated at a range of ±35° of the incidence angle. In Figure 6a, it can be seen that the LCPV with an ordinary design has approximately 90% optical efficiency at only a zero incidence angle, and when the incidence angle is ± 5°, the optical efficiency falls to 60% at 150 mm cell-lens distance. Moreover, at other cell-lens distances and incidence angles, the optical efficiency is very low. The proposed LCPV design shows good optical performance at wider incidence angles (Figure 6 (b)).

Increasing the cell-lens distance decreases the width of the concentrated light spot. When the incidence angle increases, there is a significant increase in optical efficiency from approximately 95% to 99% (Figure 6 (b)), which means redirecting rays using reflective mirrors can improve the optical efficiency at a certain range of incidence angles. The optical system has a range of incidence angles for which the optical efficiency of the optical system is higher than in the case of a zero incidence angle. For nonzero incidence angles, the incoming sunlight undergoes side reflections, which increase the chances of light being redirected towards the solar cells, meaning that the reflective mirrors help trap more sunlight, reducing optical losses.



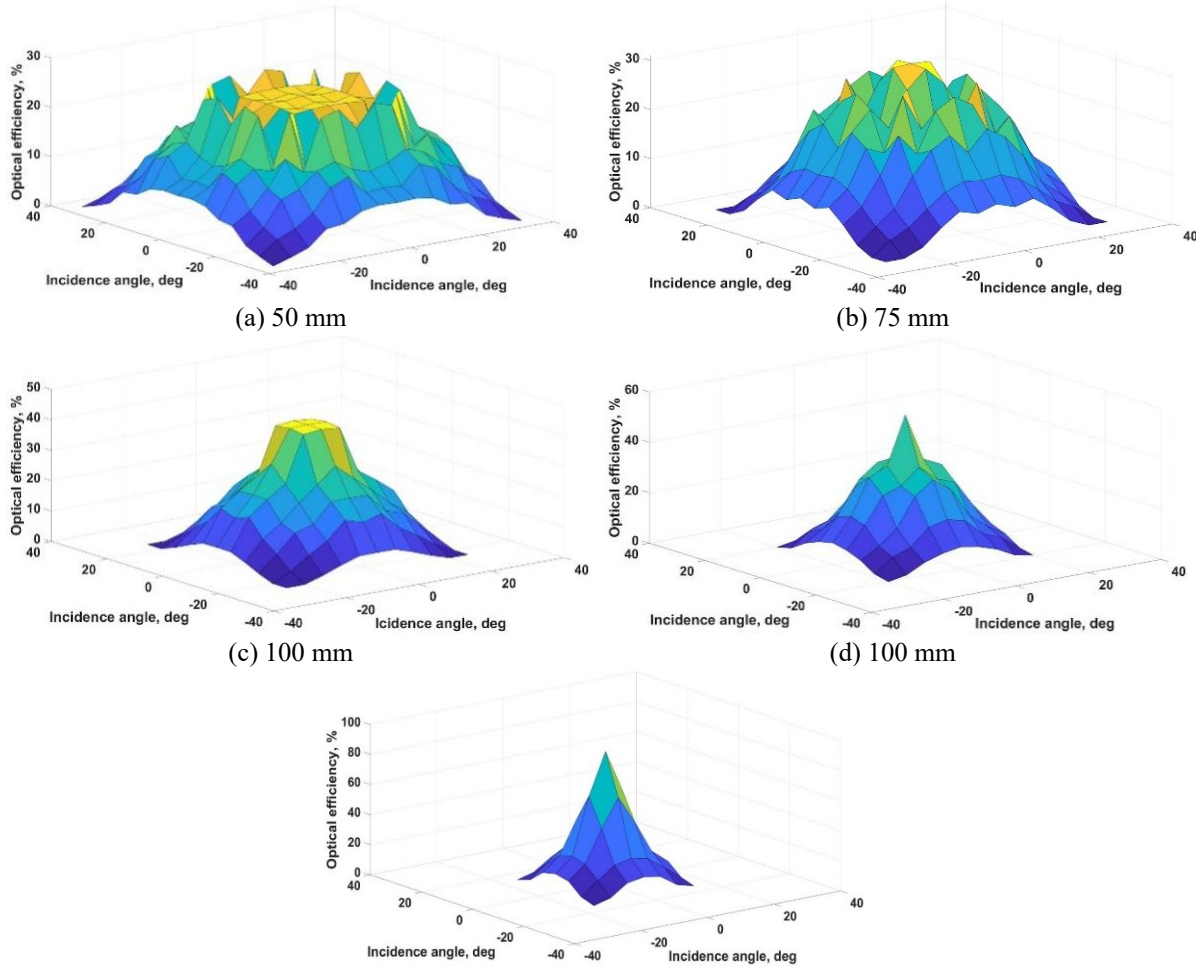
(a) Ordinary Design (b) Proposed Design

Figure 6. The dependence of optical efficiency on the incidence angles at different cell-lens distances.

https://doi.org/10.31436/iiumej.v26i2.3175

Figures 7 and 8 show the optical efficiencies of ordinary LCPV and proposed LCPV systems at different incidence angles in 3-D. The proposed LCPV shows high optical efficiency even at 50 mm of cell-lens distance, thanks to the reflective mirrors and nine solar cells. At the same time, the ordinary LCPV presents a maximum of 30% optical efficiency (Figure 7 (e), 8 (e)).

It can be seen in Figures 6 and 8 that the proposed optical system can achieve high optical efficiency in the range of ±20° of the incidence angle. However, optical efficiency is also affected by cell-lens distance. It is possible to achieve more than 85% optical efficiency when the cell-lens distance is 75mm, 50 mm in the range of ±40° of incidence angle (Figure 8 (d, e)). In the case of cell-lens distances are 100, 125, 150 mm, at more than ±20° As the incidence angle increases, the system's optical efficiency decreases gradually (Figure 8 (a, b, c)). The results suggest that the higher the optical efficiency at wider incidence angles, the shorter the cell-lens distance. A larger cell-lens distance makes the system more sensitive to the incidence angle, as the concentrated light spot may shift away from the solar cell area.



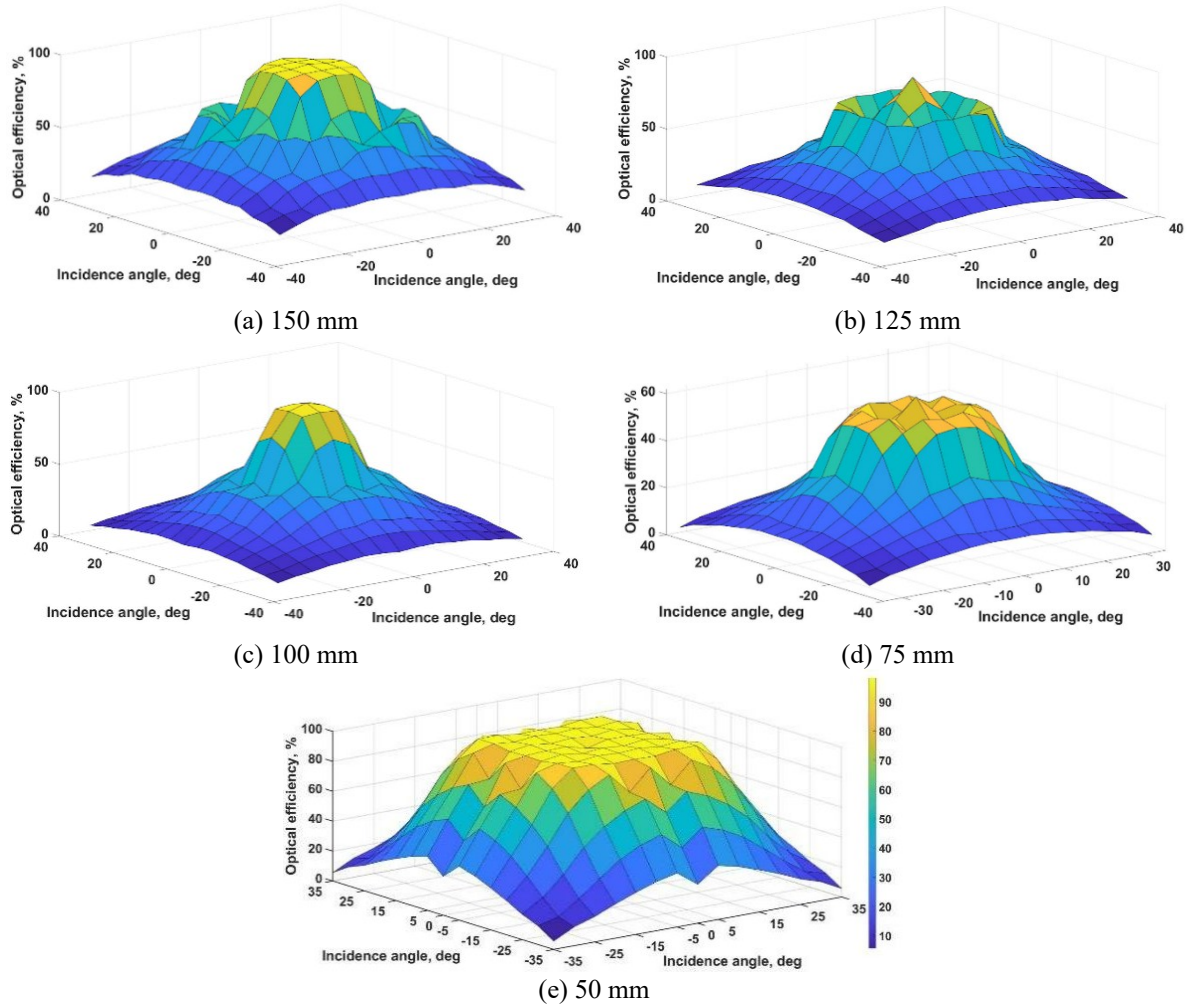
(a) 50 mm (b) 75 mm

(c) 100 mm (d) 100 mm

(e) 150 mm

Figure 7. The three-dimensional dependence of optical efficiency on the incidence angle at different cell-lens distances of the ordinary LCPV.

https://doi.org/10.31436/iiumej.v26i2.3175

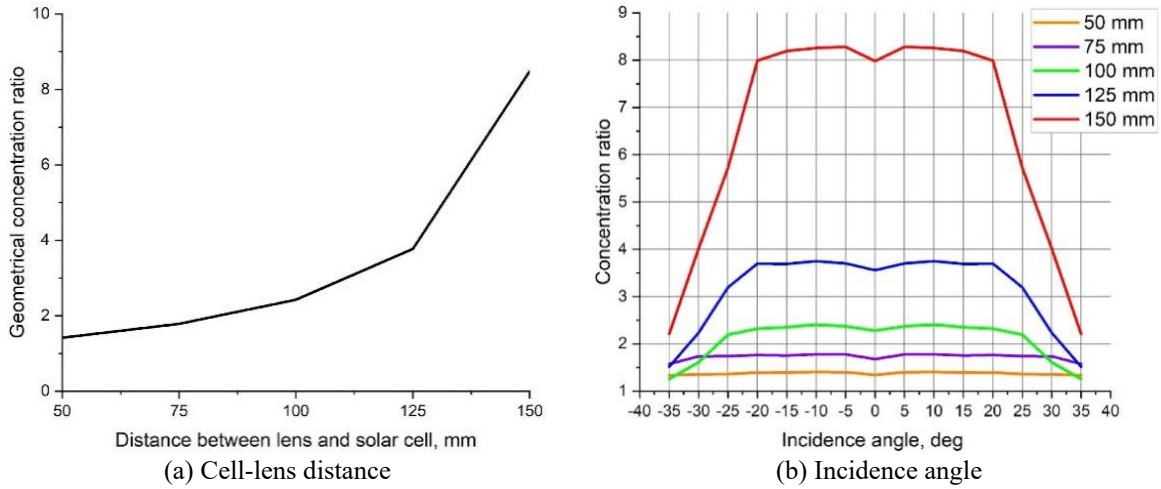


(a) 150 mm (b) 125 mm

(c) 100 mm (d) 75 mm

(e) 50 mm

Figure 8. The three-dimensional dependence of optical efficiency on the incidence angle at different cell-lens distances of the proposed LCPV.



(a) Cell-lens distance (b) Incidence angle

Figure 9. Dependence of the concentration ratio

Figure 9 shows the calculated results by Eq. (2), and if the Fresnel lens is closer to the solar cells, the geometrical concentration ratio reduces. In Figure 9a, the geometrical concentration ratio reaches 8.4 if the cell-lens distance is 150 mm, but at this distance, the

https://doi.org/10.31436/iiumej.v26i2.3175

optical efficiency reduces sharply if the incidence angle is more than ±20°. Figure 9(b) shows the results calculated by Eq. (3).

A concentration ratio of 8 suns can be achieved at 150 mm of cell-lens distance at the range ±20° of the incidence angle. At 50, 75 mm, the system cannot reach 2 suns, but optical efficiency is always high at ±35° of the incidence angle. When the cell-lens distance is 100, 125 mm, the optical efficiency is more than 80% and CR is more than 2 suns in the range of incidence angle ±25°. The proposed LCPV design helps to work the system at ±25° without the help of a solar tracking system.

3.2. Discussion

Many Fresnel lens-based concentrating photovoltaics cannot work at incidence angles more than ±1.5° and work with small multi-junction solar cells [35]. Therefore, they require a high-accuracy solar tracking system, which might incur extra costs. Another main cost of the CPV systems comes from the solar cells [36]. Many Fresnel-based concentrating photovoltaics use expensive multi-junction solar cells [37, 38] while in this work, polycrystalline silicon solar cells were used, which are the cheapest solar cells. These solar cells can work at low and medium concentrations [14], and in this work, low concentration is achieved using a Fresnel lens and reflective mirrors. Our previous works achieved low concentration [23, 24] and could get 27% more energy, but the proposed LCPV has more advantages than our ordinary system. In the proposed LCPV, the rays are concentrated and shaped as a rectangle due to reflective mirrors, and the geometrical concentration ratio reached 8.4 when the cell-lens distance was 150 mm. One of the advantages of the proposed system is that when the incidence angle increases, there is a significant increase in optical efficiency from approximately 95% to 99% (Figure 6(b)), it means redirecting rays by using the reflective mirrors can improve the optical efficiency at a certain range of incidence angles. Moreover, the proposed LCPV shows high optical efficiency at 50 mm of cell-lens distance, while the ordinary LCPV presents a maximum of 30% of optical efficiency. The next advantage of the proposed system is that the proposed optical system can have high optical efficiency at the range ±20° of the incidence angle. Subsequently, the concentration ratio of 8 suns can be achieved at 150 mm of cell-lens distance at the range ±20° of incidence angles. When the cell-lens distance is 100, 125 mm, the optical efficiency is more than 80%, and CR is more than 2 suns at the range of incidence angle ±25°. Finally, the proposed LCPV design helps to work the system at ±25° without the help of a solar tracking system.

Hence, by increasing the acceptance angle of the LCPV system, it might be possible to reduce the work of solar tracking systems, reach irradiance uniformity, and increase the acceptance angle. This system can reduce the maintenance of CPVs, extend the effective operating hours of the LCPV, as it can capture tilted incoming sunlight, and with ±25° tolerance, the system may perform better in partially cloudy weather conditions.

4. CONCLUSION

LCPV was designed, simulated, and tested using a Fresnel lens and reflective surfaces with nine polycrystalline silicon solar cells. The proposed optical system is called a Pi-shape concentrating optical design. The proposed LCPV design was compared with an ordinary Fresnel lens-based LCPV without mirrors. Both systems are made with polycrystalline silicon solar cells, which are more affordable than multi-junction solar cells. The proposed system doesn't lose rays since the reflective mirrors redirect the sunlight, and the redirected rays hit any of the nine solar cells. Ordinary LCPV can show high optical efficiency only when the cell-lens distance is 150 mm, and when the incidence angle increases, there is a sharp fall in

optical efficiency. The proposed optical system could improve the optical efficiency approximately from 95% to 99% at the range ±15° of the incidence angle. Optical efficiency of the system is more than 85% when the cell-lens distances are 75mm and 50 mm, and the incidence angle is in the range of ±40°. When the cell-lens distance is 100 or 125 mm, the optical efficiency is more than 80%. CR is more than 2 suns at the range of incidence angle ±25°. Due to high optical efficiencies at higher incidence angles, the proposed LCPV with mirrors can work without a solar tracking system, reducing the system's manufacturing cost in the future.

ACKNOWLEDGEMENT

The Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan funded the research (Grant No. AP19574454).

REFERENCES

[1] Al-Shahri, O.A., Ismail, F.B., Hannan, M.A., Lipu, M.H., Al-Shetwi, A.Q., Begum, R.A., Al-Muhsen, N.F. and Soujeri, E. (2021) Solar photovoltaic energy optimization methods, challenges and issues: A comprehensive review. Journal of Cleaner Production, 284, 125465. https://doi.org/10.1016/j.jclepro.2020.125465

[2] Emami Razavi, S.M., Jahangir, M.H. and Mousavi, S. (2019) A review study about photovoltaic systems and the energy payback time calculation for selected modules. Journal of Renewable Energy and Environment, 6(3): 38-49.

[3] Obeidat, F., 2018. A comprehensive review of future photovoltaic systems. Solar Energy, 163:545-551. https://doi.org/10.1016/j.solener.2018.01.050

[4] Kabeel, A.E., Abdelgaied, M., Sathyamurthy, R. and Kabeel, A. (2021) A comprehensive review of technologies used to improve the performance of PV systems in a view of cooling mediums, reflectors design, spectrum splitting, and economic analysis. Environmental Science and Pollution Research, 28:7955-7980. https://doi.org/10.1007/s11356-020-11008-3

[5] Alam, M.R., Bhuiyan, I.U. and Mohammad, N. (2023) Variability Assessment of Solar Irradiance for the Safe Grid Integration of Solar Photovoltaic Power Plants. Journal of Renewable Energy and Environment.

[6] Xiao, M., Tang, L., Zhang, X., Lun, I.Y.F. and Yuan, Y. (2018) A review on recent development of cooling technologies for concentrated photovoltaics (CPV) systems. Energies, 11(12):3416. https://doi.org/10.3390/en11123416

[7] Nath Tiwari, G., Singh, S. and Singh, Y.K. (2023) Effect of Solar Cell Materials on Energy Matrices of GISPVT System. Journal of Renewable Energy and Environment.

[8] Hosenuzzaman, M., Rahim, N.A., Selvaraj, J., Hasanuzzaman, M., Malek, A.A. and Nahar, A. (2015) Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation. Renewable and sustainable energy reviews, 41:284-297. https://doi.org/10.1016/j.rser.2014.08.046

[9] Bagher, A.M., Vahid, M.M.A. and Mohsen, M. (2015) Types of solar cells and application. American Journal of optics and Photonics, 3(5):94-113. https://doi.org/10.11648/j.ajop.20150305.17

[10] RTIMI, B., RAJAB, R., BENHMIDENE, A. and CHAOUACHI, B. (2020) October. A review on a comparative study between concentrator panel CPV and conventional panel PV. In 2020 11th International Renewable Energy Congress (IREC) 1-6. IEEE. https://doi.org/10.1109/IREC48820.2020.9310375

[11] Arshad, R., Tariq, S., Niaz, M.U. and Jamil, M. (2014) April. Improvement in solar panel efficiency using solar concentration by simple mirrors and by cooling. In 2014 international conference on robotics and emerging allied technologies in engineering (iCREATE) 292-295. IEEE. https://doi.org/10.1109/iCREATE.2014.6828382

[12] Reis, F., Brito, M.C., Corregidor, V., Wemans, J. and Sorasio, G. (2010) Modeling the performance of low concentration photovoltaic systems. Solar energy materials and solar cells, 94(7), pp.1222-1226. https://doi.org/10.1016/j.solmat.2010.03.010

[13] Cole, A., Baistow, I., Brown, L., Devenport, S., Heasman, K.C., Morrison, D., Whyte, G. and Bruton, T.M. (2009) March. Technological and financial aspects of laser grooved buried contact silicon solar cell based concentrator systems. In 2nd International Workshop on Concentrating Pholtovoltaic Power Plants: Optical Design and Grid Connection 9-10.

[14] Baig, H., Heasman, K.C., Sarmah, N. and Mallick, T. (2012)October. Solar cells design for low and medium concentrating photovoltaic systems. In AIP conference proceedings 1477:98-101. American Institute of Physics. https://doi.org/10.1063/1.4753843

[15] Amanlou, Y., Hashjin, T.T., Ghobadian, B., Najafi, G. and Mamat, R. (2016) A comprehensive review of uniform solar illumination at low concentration photovoltaic (LCPV) systems. Renewable and Sustainable Energy Reviews, 60:1430-1441. https://doi.org/10.1016/j.rser.2016.03.032

[16] Alnajideen, M. and Gao, M. (2022) A new configuration of V-trough concentrator for achieving improved concentration ratio of> 3.0 x. Solar Energy Materials and Solar Cells, 245:111877. https://doi.org/10.1016/j.solmat.2022.111877

[17] Parupudi, R.V., Singh, H. and Kolokotroni, M. (2020) Low Concentrating Photovoltaics (LCPV) for buildings and their performance analyses. Applied Energy, 279:115839. https://doi.org/10.1016/j.apenergy.2020.115839

[18] Barbón, A., Ayuso, P.F., Bayón, L. and Fernández-Rubiera, J.A. (2022) Non-uniform illumination in low concentration photovoltaic systems based on small-scale linear Fresnel reflectors. Energy, 239:122217. https://doi.org/10.1016/j.energy.2021.122217

[19] Schuetz, M. A., Shell, K. A., Brown, S. A., Reinbolt, G. S., French, R. H., & Davis, R. J. (2012). Design and construction of a∼ 7× low-concentration photovoltaic system based on compound parabolic concentrators. IEEE journal of photovoltaics, 2(3):382-386. https://doi.org/10.1109/JPHOTOV.2012.2186283

[20] Sangani, C. S., & Solanki, C. S. (2007). Experimental evaluation of V-trough (2 suns) PV concentrator system using commercial PV modules. Solar energy materials and solar cells, 91(6):453-459. https://doi.org/10.1016/j.solmat.2006.10.012

[21] Yadav, P., Tripathi, B., Lokhande, M., & Kumar, M. (2013). Effect of temperature and concentration on commercial silicon module based low-concentration photovoltaic system. Journal of Renewable and Sustainable Energy, 5(1). https://doi.org/10.1063/1.4790817

[22] Zahedi, A. (2011). Review of modelling details in relation to low-concentration solar concentrating photovoltaic. Renewable and Sustainable Energy Reviews, 15(3):1609-1614. https://doi.org/10.1016/j.rser.2010.11.051

[23] Dosymbetova, G., Mekhilef, S., Saymbetov, A., Nurgaliyev, M., Kapparova, A., Manakov, S., ... & Koshkarbay, N. (2022). Modeling and Simulation of Silicon Solar Cells under Low Concentration Conditions. Energies, 15(24):9404. https://doi.org/10.3390/en15249404

[24] Dosymbetova, G., Mekhilef, S., Orynbassar, S., Kapparova, A., Saymbetov, A., Nurgaliyev, M., ... & Koshkarbay, N. (2023). Neural Network based Active Cooling System with IoT Monitoring and Control for LCPV Silicon Solar Cells. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3280265

[25] Pan, J. W., Huang, J. Y., Wang, C. M., Hong, H. F., & Liang, Y. P. (2011). High concentration and homogenized Fresnel lens without secondary optics element. Optics communications, 284(19):4283-4288. https://doi.org/10.1016/j.optcom.2011.06.019

[26] El Himer, S., Ahaitouf, A., El-Yahyaoui, S., Mechaqrane, A., & Ouagazzaden, A. (2012). A comparative of four secondary optical elements for CPV systems. In AIP Conference Proceedings (Vol. 30003, No. 2018). https://doi.org/10.1063/1.5053502

[27] Renno, C. (2022). Characterization of spherical optics performance compared to other types of optical systems in a point-focus CPV system. Thermal Science and Engineering Progress, 29:101201. https://doi.org/10.1016/j.tsep.2022.101201

[28] Akisawa, A., Hiramatsu, M., & Ozaki, K. (2012). Design of dome-shaped non-imaging Fresnel lenses taking chromatic aberration into account. Solar Energy, 86(3):877-885. https://doi.org/10.1016/j.solener.2011.12.017

[29] Chouidira, M., Ihaddadene, N., Ihaddadenea, R., Jed, M. E. H., & Kherbiche, Y. (2023). An Experimental Study on the Effect of Surface Orientation and Inclination on Incident Solar Irradiation: Application to Buildings. Journal of Renewable Energy and Environment.

[30] Ullah, I. (2021). Optical design of centered-receiver trough-based CPV system. Journal of Photonics for Energy, 11(3):035502-035502. https://doi.org/10.1117/1.JPE.11.035502

[31] Yadav, P., Tripathi, B., Pandey, K., & Kumar, M. (2015). Investigating the charge transport kinetics in poly-crystalline silicon solar cells for low-concentration illumination by impedance spectroscopy. Solar Energy Materials and Solar Cells, 133:105-112. https://doi.org/10.1016/j.solmat.2014.10.031

[32] Kuttybay, N., Mekhilef, S., Koshkarbay, N., Saymbetov, A., Nurgaliyev, M., Dosymbetova, G., ... & Bolatbek, A. (2024). Assessment of solar tracking systems: A comprehensive review. Sustainable Energy Technologies and Assessments, 68, 103879. https://doi.org/10.1016/j.seta.2024.103879

[33] Wiesenfarth, M., Anton, I., & Bett, A. W. (2018). Challenges in the design of concentrator photovoltaic (CPV) modules to achieve highest efficiencies. Applied Physics Reviews, 5(4). https://doi.org/10.1063/1.5046752

[34] Victoria, M., Askins, S., Herrero, R., Antón, I., & Sala, G. (2016). Assessment of the optical efficiency of a primary lens to be used in a CPV system. Solar Energy, 134:406-415. https://doi.org/10.1016/j.solener.2016.05.016

[35] Cvetkovic, A., Mohedano, R., Gonzalez, O., Zamora, P., Benitez, P., Fernandez, P. M., ... & Miñano, J. C. (2011, December). Performance modeling of Fresnel-based CPV systems: effects of deformations under real operation conditions. In AIP Conference Proceedings 1407:74-78. American Institute of Physics. https://doi.org/10.1063/1.3658298

[36] Ullah, I. (2014). Development of Fresnel-based concentrated photovoltaic (CPV) system with uniform irradiance. Journal of Daylighting, 1(1):2-7. https://doi.org/10.15627/jd.2014.1

[37] Mahalakshmi, K., Reddy, K. S., & Subrahmanyam, A. (2022). Outdoor degradation evaluation of multi-junction solar cell for four Fresnel concentrated photovoltaic systems. International Journal of Sustainable Energy, 41(11):1958-1972. https://doi.org/10.1080/14786451.2022.2125517

[38] Shanks, K., Senthilarasu, S., & Mallick, T. K. (2016). Optics for concentrating photovoltaics: Trends, limits and opportunities for materials and design. Renewable and Sustainable Energy Reviews, 60:394-407. https://doi.org/10.1016/j.rser.2016.01.089

https://doi.org/10.31436/iiumej.v26i2.3287

ENHANCING ANOMALY DETECTION PERFORMANCE: DEEP LEARNING MODELS EVALUATION

YUNUSA MOHAMMED JEDDAH[1*], AISHA HASSAN ABDALLA HASHIM[1,2*], OTHMAN OMRAN KHALIFA[1], KHMAIES OUAHADA[2]

[1]*Dept. of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia*
[2]*Dept. of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa*

**Corresponding authors: yunusmj2@hotmail.com, aisha@iium.edu.my*

(Received: 8 May 2024; Accepted: 10 March 2025; Published online: 15 May 2025)

ABSTRACT: Detection of anomalies within video streams continues to be challenging, mostly due to the complexities involved in distinguishing abnormal activities from normal ones. This study aimed to enhance anomaly detection performance by evaluating different deep learning models and optimizers. Utilizing the Keras framework and Python on a Kaggle notebook, the experiment explored the effectiveness of DenseNet121, VGG19, ResNet50, and InceptionV3 models in conjunction with Adam, SGD, RMSprop, and Adagrad optimizers. A UCF Crimes dataset subset focused on Accuracy, F1 Score, and AUC evaluation metrics. The results establish that the InceptionV3 model paired with the Adam optimizer outperforms the other combinations, attaining AUC scores of 0.9918. In contrast to other state-of-the-art models such as DenseNet121 and ResNet50, InceptionV3 presents enhanced precision and adaptability in handling the variability found in video anomaly datasets. This study enhances security by providing insights into enhanced model-optimizer combinations, advancing video surveillance approaches, and providing support for developing robust anomaly detection systems.

ABSTRAK: Pengesanan anomali dalam strim video terus mencabar, kebanyakan disebabkan oleh kerumitan yang terlibat dalam membezakan aktiviti tidak normal dari biasa. Kajian ini cuba meningkatkan prestasi pengesanan anomali dengan menilai model dan pengoptimum pembelajaran mendalam yang berbeza. Menggunakan rangka kerja Keras dan Python pada komputer riba Kaggle, eksperimen ini meneroka keberkesanan model DenseNet121, VGG19, ResNet50 dan InceptionV3 bersama pengoptimum Adam, SGD, RMSprop dan Adagrad. Subset data Jenayah UCF digunakan, memfokuskan pada ketepatan, Skor F1 dan metrik penilaian AUC. Dapatan kajian menunjukkan bahawa model InceptionV3 bersama pengoptimum Adam, mengatasi kombinasi lain, mencapai skor AUC 0.9918. Berbeza dengan model canggih lain seperti DenseNet121 dan ResNet50, InceptionV3 mempunyai ketepatan dan kebolehsuaian yang tinggi dalam mengendalikan kebolehubahan yang terdapat dalam set data anomali video. Kajian ini menyumbang kepada peningkatan keselamatan dengan memberi gabungan pengoptimum bersama model yang dipertingkatkan, memajukan pendekatan pengawasan video dan menyediakan sokongan bagi pembangunan sistem pengesanan anomali yang teguh.

KEYWORDS: *Deep learning models, Video anomaly detection, Optimization Techniques, Video Surveillance, Performance Evaluation*

1. INTRODUCTION

Modern surveillance systems have become increasingly indispensable for protecting the public, especially in anomaly detection or suspicious behavior in videos. Technological

advancements have transformed video surveillance into a component of security systems, providing critical functions in public places, transportation hubs, and other critical infrastructure. Anomalies in video streams, like unattended baggage or abnormal human behaviours [1], can hint at potential threats that demand immediate attention. Swift and precise detection of these abnormalities can prevent minor disruptions from significant safety breaches. However, real-time monitoring of video feeds is an overwhelming undertaking for human operatives, especially in dynamic and crowded environments. Automating video anomaly detection systems enhances security personnel's ability to identify potential threats in real time, where manual monitoring is challenging [2].

Despite its cybersecurity roots, the term "anomaly detection" has been widely employed in video analysis. Numerous studies have been conducted on anomaly detection in videos, highlighting its applicability in circumstances such as detecting suspicious behaviour [3], monitoring traffic violations [4], and identifying harmful objects in sensitive places [5]. However, video anomaly detection faces unique challenges, such as anomalies' rarity, appearance variability, and the complexity of describing abnormal behaviours [[6]]. These challenges necessitate employing adaptable and robust techniques to address the complexity and imbalances in video data.

This study addresses these concerns by evaluating the performance of four cutting-edge deep learning models — ResNet50, DenseNet121, VGG19, and InceptionV3 — for video anomaly detection. In addition, we evaluate the effects of four optimization algorithms — SGD, RMSprop, Adam, and Adagrad—on the models' performance. Using Python and the Keras framework on the Kaggle notebook, this study analyses key evaluation metrics, such as accuracy, F1 Score, and AUC, to reveal the most efficient model-optimizer pairings. By investigating these combinations, we aim to provide insights into enhancing the reliability and accuracy of video anomaly detection systems.

This study's primary contributions are as follows:

- A comparative assessment of four popular deep learning models for video anomaly recognition.
- An investigation of the impact of four different optimization algorithms on these models' performance.
- Optimal model-optimizer combinations identification to address variability and imbalance issues in video data.
- Providing insights to develop robust anomaly detection systems that can handle real-world scenarios.

The remaining sections of this paper are arranged as follows: The Related Works section explores current studies, focusing on advancements in research in video anomaly detection. The Method section covers the experimental setup, dataset preparation, preprocessing, and model selection. The Results section provides the findings and compares the models' performance. The Discussion section presents the interpretation of the results in context, and the Conclusion section highlights key insights, limitations, and proposes future work.

2. RELATED WORKS

Video anomaly detection has attracted much attention thanks to its significance in surveillance and security systems. Recent advances in deep learning enable the development of powerful models trained to recognize anomalies in complex video data. However, the choice

of optimization techniques has major effects on their performance. This section reviews recent studies evaluating model-optimizer combinations for video anomaly detection.

2.1. Technology in Anomaly Detection Using Deep Learning

Deep learning models, especially CNNs, have successfully detected video anomalies. Pre-trained models such as ResNet [17], DenseNet [16], and InceptionV3 [20] have been widely used to extract hierarchical features from video frames. Wu et al. [13] assessed pre-trained CNNs on the UCSD dataset, revealing that transfer learning notably enhances anomaly detection efficiency. Wang et al. [9] also offered spatiotemporal improvements to correct the data imbalance, attaining state-of-the-art results on benchmark datasets such as UCSD Ped2 and Avenue.

Optimization methods are essential for training deep learning models. Stochastic Gradient Descent (SGD) is a primary optimization technique; however, it frequently has difficulties with noisy gradients and poor convergence. Adaptive optimizers such as Adam, RMSprop, and Adagrad mitigate these restrictions by dynamically modifying learning rates. Lydia and Francis [14] extensively assessed optimization approaches, emphasizing Adam's advantages in managing large-scale datasets with sparse gradients. Pawar and Attar [8] compared Adam and RMSprop for video anomaly detection, indicating that Adam attains rapid convergence and better accuracy.

Recent studies demonstrated how important it is to pair specific models with optimizers suited for their designs. For example, InceptionV3, with factorized convolutions and supplementary classifiers, gains substantial benefits from Adam's adaptive learning rates [20]. Conversely, ResNet50, which utilizes residual connections, demonstrates effective performance with both Adam and RMSprop, thanks to its ability to cope with vanishing gradients [17]. DenseNet121, characterized by its dense connections, demonstrates uniform performance across various optimizers, attaining optimal results with RMSprop [16]. These outcomes demonstrate reasons for the rigorous evaluation of model-optimizer pairings to optimize performance.

Numerous other recent studies have explored various methods for detecting unusual or suspicious activity in video streams, establishing a foundation for automated security solutions. [7] reviewed about 290 articles, claiming that unsupervised learning is the most frequently adopted method. Deep learning techniques, in particular, have demonstrated strong promise. For instance, in [8], they investigated deep learning applications for video-based anomaly detection, analyzing deep learning methods. The authors introduced a graphical taxonomy, addressed spatial anomalies, and compared frameworks. Other studies, such as [9] and [10], they proposed enhancements in spatiotemporal relations to tackle data imbalance, attaining significant results on popular UCSD Ped1, Avenue, and UCSD Ped2 datasets.

Traditional anomaly detection techniques, such as classical and statistical machine learning algorithms, still have a key role in certain scenarios, even though they often depend on manually crafted features. On the contrary, deep learning models can autonomously learn features from large datasets; however, they have higher computational requirements. Research on video-based anomaly detection, like [11] and [12], presented a framework for understanding the strengths and weaknesses of these methods, which we employ in this study. These studies are valuable resources for researchers and experts seeking to understand and apply deep learning techniques for anomaly detection.

Pre-trained deep learning models have proven to be useful in anomaly detection. For instance, [13] confirmed the efficiency of pre-trained convolutional neural networks (CNN) for

video-based anomaly detection on the UCSD dataset. Others, such as [14], emphasize the significance of optimizers in realizing high performance, which we further investigated in this study by evaluating four optimizers across four pre-trained deep-learning models.

2.2. Open Challenges in Video-Based Anomaly Detection

Deep learning techniques have demonstrated outstanding possibilities in detecting anomalies, showing the ability to understand complex patterns and relationships within video data. Nonetheless, notable limitations and obstacles exist in using deep learning to detect human anomalies.

Scalability presents a challenge, for example. Deep learning models typically demand significant processing power, especially with video data. The more the data grows or varies exponentially, the more deep learning systems face the challenge of efficiently handling the processing, especially regarding computing resources. The more complex the models become, the longer the training time, the greater the need for more hardware, and the higher the energy demand. Ensuring that low-latency and real-time processing capabilities are taken care of while preserving the performance and accuracy of the model is a decisive scalability issue.

Another constraint is the interpretability of deep learning models. Despite their outstanding performance, these models are often naturally considered "black boxes," making them difficult to interpret. This can impede practical deployment and regulatory compliance. Understanding the rationale behind a model's decision is essential for establishing trust and obtaining actionable insights. However, most deep learning models do not inherently provide this interpretability.

Likewise, deep learning models characteristically require large, labeled datasets for training. In the context of human anomaly detection, this involves having video data where anomalous and normal behaviors are correctly labeled. However, obtaining such labeled data can be difficult and time-consuming. Likewise, the labeling process can be subjective and susceptible to errors.

As these deficiencies persist in generalizing to real-world surveillance contexts, this study attempts to address these drawbacks by assessing model-optimizer synergies using the UCF Crime dataset, highlighting adaptation to real-world unpredictable conditions. This will culminate in figuring out which of the different ways of enhancing anomaly detection techniques proves economically and realistically viable. In this study, we investigated by pairing four pre-trained deep learning models and four optimizers to determine which combinations enhance anomaly detection.

3. METHODOLOGY

This study employs a quantitative experimental approach, investigating model-optimizer pairings via empirical metrics (Accuracy, F1 Score, AUC).

3.1. Datasets and Preprocessing

3.1.1. Data Structure and Subfolders

The UCF Crime Dataset [15] serves as the study's foundation, consisting of 1,900 real-world surveillance videos, with a total runtime of 128 hours. These videos are uncut and categorized into thirteen (13) distinct types of realistic anomalies, such as Arrest, Abuse, Arson,

https://doi.org/10.31436/iiumej.v26i2.3287

Assault, Robbery, Road Accident, Explosion, Burglary, Stealing, Fighting, Shooting, Shoplifting, and Vandalism.

The dataset used in this study is a subset of this dataset, which was divided into Train and Test subfolders, each containing seven different subfolders. The subfolders in the Train and Test folders represent distinctive classes representing individual criminal activities and normal behaviours ('Abuse', 'Arson', 'Assault', 'Fighting', 'NormalVideos', 'Robbery', and 'Shooting').

3.1.2. Preprocessing Steps

As shown in Figure 1, the preprocessing process ensures that raw video data is prepared for model evaluation. The workflow comprises the following blocks:

1. *Video Frames and Feature Extraction*: The videos were pre-processed, extracted frames, and converted to .png format. The frames were resized to 64×64 pixels for DenseNet121, VGG19, and ResNet50 models, and 75×75 pixels for the InceptionV3 model.
2. *Model Training*: This process comprised integrating pre-trained CNNs for fine-tuning, testing individual optimizers, and iteratively training the models with the selected optimizers to improve performance and minimize loss.
3. *Anomaly Detection*: The trained models were employed to predict anomaly scores for input videos, and a threshold was established to determine detected patterns as normal or abnormal.
4. *Model Evaluation*: Performed using AUC, accuracy, and F1 Score, followed by a comparison of model-optimizer combo to discover the best-performing configurations.

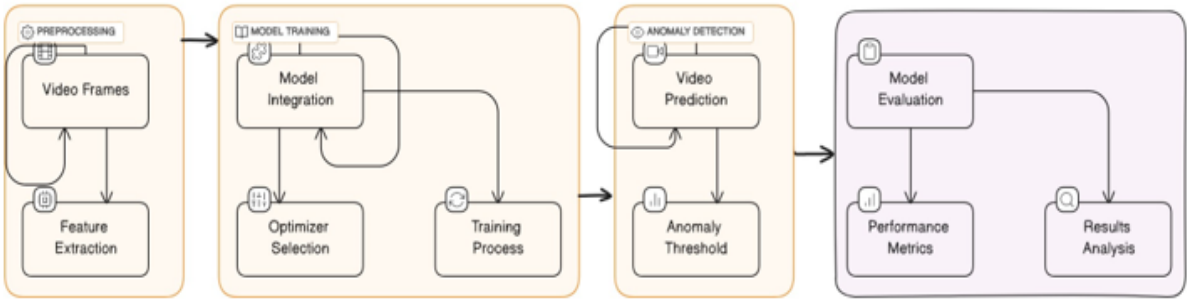


Figure 1. Anomaly Detection Process

3.2. Experimental Setup and Design

The experimental setup involves a structured workflow to evaluate the effectiveness of pre-trained deep learning models and optimization methods in video anomaly detection. The design of the experiment, as depicted in Figure 2, outlines the flowchart of the methodology:

1. *Model Selection*: This stage involves choosing the pre-trained CNN models (ResNet50, DenseNet121, VGG19, and InceptionV3) for their capabilities in image-based classification tasks.
2. *Optimizer Variation*: This stage tests four optimization algorithms—Adam, SGD, RMSProp, and Adagrad—on the selected models to analyze their impact on convergence and performance.
3. *Training on Keras API*: Leveraging the Kagle framework to streamline the training of models with diverging optimizers.

https://doi.org/10.31436/iiumej.v26i2.3287

4. *Performance Evaluation*: The evaluation metrics include Accuracy, F1 Score, and AUC to evaluate the effectiveness of the model-optimizer combinations.
5. *Best Model Found (Objectives Achieved)*: Confirm whether the study objectives, such as identifying optimal combinations, are met.
6. *Results and Insights*: Reveal which model-optimizer combinations yield the highest performance metrics and explain conclusions on model behaviors and optimization methods that enhance anomaly detection.

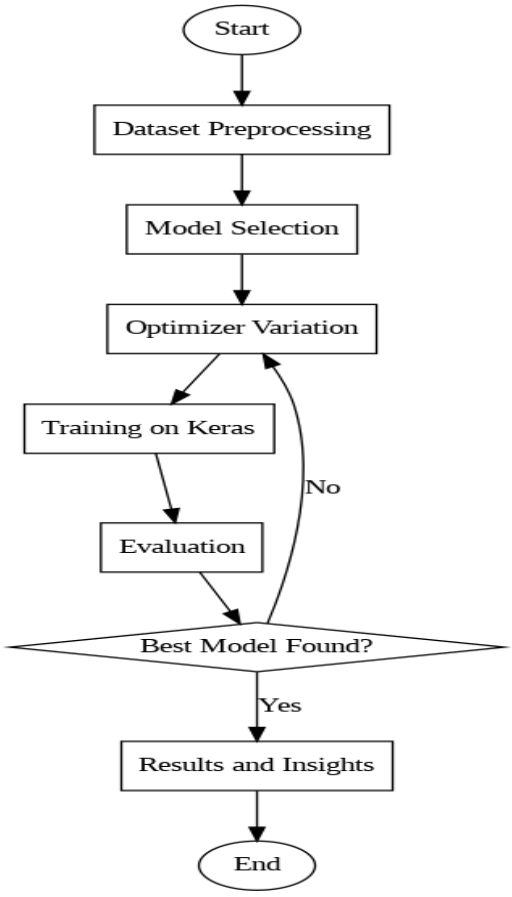


Figure 2. Flowchart illustrating the methodology: (1) Dataset preprocessing, (2) Model selection (DenseNet121, ResNet50, VGG19, InceptionV3), (3) Optimizer variation (Adam, SGD, RMSprop, Adagrad), (4) Training on Keras, (5) Evaluation using AUC, Accuracy, F1 Score.

3.3. Description of Deep Learning Models

In this study, we employed four major pre-trained deep learning models.

3.3.1. DenseNet121

DenseNet (Densely Connected Convolutional Networks) [16] is a pre-trained CNN model that achieved recognition due to its effectiveness in image-related classification. It gets its name by forward propagation by connecting every layer to another layer. There are 121 layers in this version. The bottleneck layer comprises two convolutional layers with a batch normalization layer in the middle. The initial convolutional layer is a 1x1 convolution, which decreases the

https://doi.org/10.31436/iiumej.v26i2.3287

number of input feature mappings for the bottleneck layer. The following convolutional layer is a 3x3 convolution that generates the output feature maps. The feature maps generated by both layers in a dense block are determined by the hyperparameter growth_rate, which is set to four times the inner_channel parameter. DenseNet's transition layer comprises a 1x1 convolutional layer, a 2x2 average pooling layer, and a batch normalization layer. The batch normalization layer normalizes the activations, the 1x1 convolutional layer decreases the feature maps, while the spatial dimensions are reduced by the feature maps' average pooling layer. The global average pooling layer output is passed through a fully connected (FC) layer with an activation function (ReLU), and a small number of neurons (e.g., 256). This layer decreases the feature vector dimensionality and extracts higher-level features. The FC layer's output is passed through the activation function (softmax), which gives class probabilities.

3.3.2. ResNet50

ResNet-50 (residual neural network, ResNet variant) [17] is a pre-trained CNN that is 50 layers deep (i.e., 48 convolution layers, 1 average pool layer, and 1 MaxPool layer). ResNet is an artificial neural network (ANN) that piles residual blocks on top of each other to make up a network. The model has 50 layers, comprising convolutional, batch normalization, activation, pooling, and FC layers. Residual connections are used to enable easier training of deep neural networks. This model is among the most popular variants of the ResNet architecture, having 50 layers that show impressive performance on a range of classification tasks (images). The model comprises five stages, each with convolution and identity blocks. Each convolution block has three convolutional layers, and each identity block has three. The trainable parameters of ResNet50 are more than 23 million parameters.

3.3.3. VGG19

Created by Zisserman and Simonyan of Oxford University, VGG19 has 19 layers (16 convolutions and 3 fully connected layers) [18]. It is a CNN model that uses strictly 3×3 filters with a stride of 1 and padding, alongside 2×2 max-pooling layers with a stride of 2. The model is deeper and has more layers than AlexNet. To lower the parameter count in such deep networks, it utilizes small 3×3 filters in all convolutional layers and is best used with its 7.3% error rate. This model uses (3 × 3) kernels with a 1-pixel stride size, and spatial padding is applied to preserve the image's spatial resolution. Furthermore, a 2 × 2-pixel window with a 2-pixel stride is used for max pooling. The model in question is quite complex, having undergone training on millions of photos with complicated classification tasks. With an enormous 19.6 billion FLOPs [19], it is a potent tool for picture classification and recognition.

3.3.4. InceptionV3

As a member of the Inception family, InceptionV3 [20] is an architecture of CNN introduced by Google. Compared to its predecessors, the architecture is more technologically advanced and optimized. It includes several techniques to enhance model adaptation. Among such techniques is Label Smoothing, which provides regularization and keeps the model from becoming unduly confident in its class assignments. To lower the number of parameters and computational expense, factorized 7×7 convolutions are also utilized. Similarly, an auxiliary classifier is used to convey label information to lower layers of the network. This increases the gradient signal and provides regularization. Batch normalization is also implemented in InceptionV3 at the network's side head. InceptionV3 retains excellent efficiency without sacrificing speed, even though it is deeper than its versions. Deeper networks can be created thanks to their design, which also limits parameter expansion for an increasingly effective

model. It works better than another well-known convolutional neural network, AlexNet, with fewer than 25 million parameters compared to 60 million.

3.4. Optimizers Used in the Study

Optimization is a crucial aspect of deep learning, influencing model performance. In this study, we investigated the deep learning models with the following optimizers:

1. Adam (Adaptive Moment Estimation) is an efficient optimization algorithm for large-scale problems with extensive parameters and/or data. It utilizes adaptive estimates of lower-order moments and first-order gradients to optimize stochastic objective functions. It is easy to implement and has been empirically proven effective.
2. RMSProp (Root Mean Square Propagation) was introduced by Geoff Hinton. It is an adaptive learning rate method that adjusts weight updates based on a moving average of squared gradients. It converges to a stationary point for realizable problems and a bounded region for non-realizable problems.
3. SGD (Stochastic Gradient Descent) is a fundamental set of rules that combines classical gradient descent with random subsampling to optimize the objective function. It's commonly used for neural network optimization.
4. A stochastic optimization technique called Adagrad (Adaptive Gradient Algorithm) modifies the learning rate in response to parameters. It makes smaller updates for features that occur frequently, while for features that occur infrequently, it makes more significant updates.

The optimizers were employed to train each model, ensuring consistent hyperparameters throughout trials to maintain fairness. By weighing their impact on the evaluation metrics, the study identifies the optimal model-optimizer combinations for the challenges of video anomaly detection.

3.5. Evaluation Metrics

We evaluated each of the models with each of the four optimizers using three performance evaluation metrics.

1. Accuracy: This evaluation metric measures the proportion of correctly classified samples:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP is the True Positive, TN is the True Negative, FP is the False Positive, and FN is the False Negative.

Accuracy, also referred to as top-1 accuracy, is a statistical measure that is used to demonstrate how accurately a binary classification test recognizes or rules out a condition. In other words, accuracy is described as the percentage of true positives and true negatives across every instance investigated that was predicted correctly. Although it gives a broad picture of the model's effectiveness, imbalanced datasets might not be a good fit for it. When there is a large imbalance in class in the dataset, accuracy alone may be deceiving because if the model consistently predicts the majority class, it could achieve a prominent level of accuracy.

2. F1 Score: This is simply the harmonic mean of precision and recall. It is the combination of recall and precision, delivering a single score, in which precision is the division of the number of true positive values by all sample numbers predicted as positive, comprising

those not correctly identified. The recall is calculated through the division of the total number of true positive outcomes by the total number of samples that ought to have been recognized as positive.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

3. AUC Score: Assesses the area under the receiver operating characteristic (ROC) curve, which represents the model's ability to differentiate between classes.

$$\text{AUC} = \int_0^1 \text{TPRd(FPR)} \quad (3)$$

A receiver operating characteristic (ROC) plot is a two-dimensional representation of the classifier's performance. Multiclass AUCs are calculated by generating each class's ROC curve, determining the AUC, and then tallying the AUCs weighted by the reference class's prevalence in the video footage. The AUC performance metric is closely related to the Gini coefficient, which is occasionally employed as an alternative. The most common definition of this is double the area between the diagonal and the ROC curve. In simple geometry, $Gini + 1 = 2 \times AUC$. It is a plot of the true-positive rate (TPR) compared to the false-positive rate (FPR). A good deep-learning model will have an AUC value of nearly 1, but a random model could have a 0.5 AUC value.These metrics ensure a robust evaluation of models, especially in imbalanced datasets (such as the one used in this study) where accuracy alone is inadequate.

4. EXPERIMENTAL RESULTS

This section provides an experimental illustration of the capabilities of the chosen models and how they vary based on the optimizer. We presented the evaluation metrics used to evaluate the models' performance in tabular form.

4.1. Tabular Representation of Evaluation Metrics

Models and their performances with different optimizers are presented in tabular form.

Table 1. DenseNet121 Model

Optimizer	Accuracy	F1 Score	AUC
SGD	0.8079	0.8079	0.9307
Adam	0.8079	0.8079	0.9499
RMSprop	0.8079	0.8079	0.9537
Adagrad	0.8079	0.8079	0.9345

Table 2. ResNet50 Model

Optimizer	Accuracy	F1 Score	AUC
SGD	0.8079	0.8079	0.9307
Adam	0.8147	0.8147	0.9683
RMSprop	0.8079	0.8079	0.9701
Adagrad	0.8079	0.8079	0.9301

Table 3. VGG19 Model

Optimizer	Accuracy	F1 Score	AUC
SGD	0.8079	0.8079	0.9405
Adam	0.7583	0.7583	0.9905
RMSprop	0.7838	0.7838	0.9872
Adagrad	0.8079	0.8079	0.9391

https://doi.org/10.31436/iiumej.v26i2.3287

Table 4. InceptionV3 Model

Optimizer	Accuracy	F1 Score	AUC
SGD	0.8079	0.8079	0.9393
Adam	0.7641	0.7641	0.9918
RMSprop	0.7889	0.7889	0.9886
Adagrad	0.8079	0.8079	0.9401

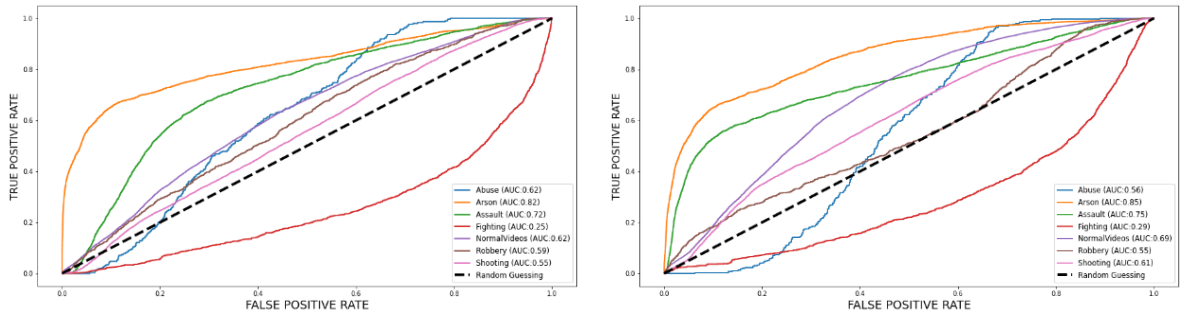
4.2. Comparative Model and Optimizer Analysis

The DenseNet121 model's performance was consistent and robust across our selected optimizers in the anomaly detection analysis. The optimizers impact the performance metrics, maintaining accuracy and F1 Score. SGD demonstrated an AUC value of 0.9307. Adam and RMSprop optimizers perform competitively, with Adam slightly outperforming RMSprop with an AUC of 0.9499.

The ResNet50 model provides consistent and competitive performance, with SGD achieving an AUC of 0.9307. Adam and RMSprop, on the other hand, outperform SGD with higher AUCs of 0.9683 and 0.9701. These optimizers improve accuracy and F1 scores, implying enhanced model correctness. Adagrad slightly lags with an AUC of 0.9301.

With regard to VGG19, SGD accomplished quite a good AUC of 0.9405. However, both Adam and RMSprop perform better than SGD, with remarkable respective AUCs of 0.9905 and 0.9872. These optimizers particularly improved the accuracy and F1 score of the model, underlining the strength of VGG19 in recognizing anomalies. While maintaining a competitive AUC value of 0.9391, Adagrad falls marginally behind Adam and RMSprop.

For the InceptionV3 model, Adam and RMSprop optimizers demonstrate considerable anomaly detection capabilities. SGD yielded an AUC of 0.9393, with constant accuracy and F1 score performances. However, Adam and RMSprop perform beyond SGD, with AUCs of 0.9918 and 0.9886, respectively. These optimizers also improve the accuracy and F1 Score of the InceptionV3 model, highlighting their ability to fine-tune the model for improved anomaly detection. Adagrad retains an AUC value of 0.9401, trailing behind the Adam and RMSprop optimizers.



(a) InceptionV3 with Adam Optimizer (b) InceptionV3 with RMSprop Optimizer

Figure 3. Visualization of the InceptionV3 Model of ROC-AUC

5. DISCUSSION

This study demonstrates that the InceptionV3 model, when optimized with Adam, surpasses other model-optimizer combinations in video anomaly detection. This conclusion is backed by its superior AUC scores of 0.9918 (Adam) and consistently high accuracy and F1 scores. InceptionV3's architecture incorporates advanced techniques like factorized convolutions and auxiliary classifiers, enabling it to capture complex spatiotemporal patterns

https://doi.org/10.31436/iiumej.v26i2.3287

effectively. This capability addresses the variability and imbalance inherent in video anomaly datasets. In comparison, the other models, however robust in their own right, had lower performance due to their less optimized handling of spatial and temporal dependencies in video data. Likewise, VGG19's modest architecture and dependence on smaller filters caused slower convergence and reduced precision, further highlighting InceptionV3's advantage.

The performance observed with InceptionV3 is further enhanced by its combination with the Adam optimizer. Adam's momentum and adaptive learning rates allow for efficient parameter updates, which ensure faster convergence and improved generalization. RMSprop complements InceptionV3 by stabilizing training through adaptive learning rates tailored to sparse or noisy gradients. In comparison, Adagrad and SGD were less effective, with Adagrad exhibiting diminishing learning rates over extended training and SGD struggling to adapt to the dynamic nature of video data. This comparative study emphasizes the importance of selecting both high-performing models and appropriate optimizers to address the challenges of video anomaly detection. In this study, InceptionV3, combined with Adam, is a balanced solution that achieves high results, especially in terms of AUC, and thus emerges as the most efficient model-optimizer combination.

6. CONCLUSION

6.1. Key Insights and Observations

The distinct responses presented by each model underscore the importance of tailoring optimization methods to accommodate the distinct attributes of each CNN model. This highlights the complex relationships between models and optimizers and emphasizes the need for adaptability in optimization methods. Likewise, optimizers are important in deciding how well an anomaly detection algorithm performs. The choice of an optimizer enhances models' capabilities to detect anomalies in video footage.

As mentioned earlier, DenseNet121 demonstrated exciting strength and consistency when run in conjunction with different optimizers, suggesting that it may have flexibility in independent optimization strategies. Due to its steady performance, DenseNet121 is a reliable option for anomaly detection tasks, providing consistency even in different optimization approaches. In addition, optimizers and model architecture have a complex interaction that requires careful direction finding, as demonstrated by fine-tuning models for anomaly detection. This highlights the importance of utilizing hyperparameters to achieve the best performance results possible.

Furthermore, the synergistic impact of model-optimizer combinations can be found in the superior detection of anomaly capabilities of pairings like InceptionV3 with Adam and RMSprop optimizers. Maximizing the effectiveness of the models requires leveraging these synergies. On the other hand, choosing optimizers enforces striking a balance among many performance assessment standards, leading to variations in performance results between models and measures. Therefore, making well-informed decisions requires a thorough assessment considering variations in performance metrics.

In addition, these trends have real-world applications in anomaly detection, offering practitioners recommendations of approaches to adopt depending on their datasets. Examining these observations would greatly benefit the design and implementation of algorithms in real-world situations. Eventually, the experiment tends to spark debates on model-optimization approaches and other factors affecting performance variability. Evaluating these provides more insights into trends that have been observed and serves as a guide for further research.

6.2. Limitations and Future Directions

The experimental assessment conducted in this study encountered some limitations worth mentioning and needs insights into future studies. The most challenging constraint faced is the insufficient computing capacity, which necessitated experimenting on the Kaggle notebook, restricting the model training to a single epoch.

In the future**,** computing capacity can be addressed using stronger hardware or cloud-based platforms. This will enable extensive training sessions, allowing for many epochs and a thorough assessment of the dynamic range between the model and optimizer. Moreover, holding out the experiments with more epochs may better evaluate model convergence, stability, and long-term performance. This approach may yield valuable insights into the model's capabilities over extended training durations.

With regard to the proposed research directions, widening the scope of the investigation to involve separate deep learning models employed here would improve the overall understanding of the interactions between models and optimizers. Still, evaluating the tendency to deploy optimized models in real-world settings and their effectiveness in dynamic circumstances will be significant for their practical application.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS22-264-0873) (Grant No: FRGS/1/2022/ICT11/UIAM/01/1).

REFERENCES

[1] P. G. I. M. Chandrasekara, L. L. G. Chathuranga, K. A. A. Chathurangi, and D. M. K. N. Seneviratna, "Intelligent Video Surveillance Mechanisms for Abnormal Activity Recognition in Real- Time: A Systematic Literature Review," vol. 5, no. 1, pp. 26–40, 2023.

[2] B. Gayathri, A. Abhinav, C. H. Reddy, and G. Praneeth, "ANOMALY XPERT: A Deep Learning Approach," no. March 2024.

[3] N. Gupta, C. Science, B. B. Agarwal, and C. Science, "Recognition of Suspicious Human Activity in Video Surveillance: A Review," vol. 13, no. 2, pp. 10529–10534, 2023.

[4] A. Tonge, S. Chandak, R. Khiste, U. Khan, and L. A. Bewoor, "Traffic Rules Violation Detection using Deep Learning," in 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 1250–1257. doi: 10.1109/ICECA49313.2020.9297495.

[5] U. Y. Reddy, M. S. Nikhil, P. S. S. Krishna, and S. S., "Systematic Harmful Signs Detection for Women's Safety Using Neural Networks," in 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023, pp. 1–5. doi: 10.1109/ICCCI56745.2023.10128298.

[6] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," Image Vis Comput, vol. 106, p. 104078, Feb. 2021, doi: 10.1016/J.IMAVIS.2020.104078.

[7] S. Bhakat and G. Ramakrishnan, "Anomaly detection in surveillance videos," ACM International Conference Proceeding Series, pp. 252–255, 2019, doi: 10.1145/3297001.3297034.

[8] K. Pawar and V. Attar, "Deep learning approaches for video-based anomalous activity detection," World Wide Web, vol.22, no.2, pp. 571–601, 2019, doi:10.1007/s11280-018-0582-1.

[9] J. Wang, D. Jia, Z. Huang, M. Zhang, and X. Ren, "Normal Spatio-Temporal Information Enhance for Unsupervised Video Anomaly Detection," Neural Processing Letters, 2023, doi: 10.1007/s11063-023-11347-5.

[10] H. Lv, Z. Cui, B. Wang, and J. Yang, "Spatio-Temporal Relation Learning for Video Anomaly Detection," Journal of Latex Class Files, vol. 14, no. 8, 2022, doi: 10.48550/arXiv.2209.13116.

[11] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," pp. 1–50, 2019.

[12] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep Learning for Anomaly Detection: A Review," ACM Computing Surveys, vol. 54, no. 2, pp. 1–36, 2020, doi: 10.1145/3439950.

[13] C. Wu, S. Shao, C. Tunc, and S. Hariri, "Video Anomaly Detection using Pre-Trained Deep Convolutional Neural Nets and Context Mining," Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, vol. 2020-Novem, 2020, doi: 10.1109/AICCSA50499.2020.9316538.

[14] A. Lydia and F. Sagayaraj Francis, "A Survey of Optimization Techniques for Deep Learning Networks," International Journal for Research in Engineering Application & Management (IJREAM), vol. 05, no. August 2020, p. 2, 2019, doi: 10.35291/2454-9150.2019.0100.

[15] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 6479–6488, 2018, doi: 10.1109/CVPR.2018.00678.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–14, 2015.

[19] M. F. Amin, Z. Othman, S. S. S. Ahmad, and F. Kasmin, "Analysis on the impact of imagenet preprocessing image mode using VGG19 pre-trained model in plant disease classification," Proceedings of Mechanical Engineering Research Day 2022, no. August, pp. 154–155, 2022.

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

https://doi.org/10.31436/iiumej.v26i2.3393

CLOUD COMPUTING-BASED SECURITY ANALYSIS ON WIRELESS SENSOR NODES CLUSTER USING PREDICTIVE TECHNIQUE

MUHAMMED ZAHARADEEN AHMED[1,2*], AISHA HASSAN ABDALLAH HASHIM[1,3], OTHMAN OMRAN KHALIFA[1,4], ALIYU MUHAMMAD WAKIL[2], ZEINAB E. AHMED[1], KHMAIES OUAHADA[3]

[1]*Dept. of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia*
[2]*Dept. of Computer Engineering, University of Maiduguri, Borno State, Nigeria*
[3]*Dept. of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa.*
[4]*Libyan Center for Engineering Research and Information Technology, Bani Waleed, Libya*

**Corresponding author: zaharadeenccna@gmail.com*

(Received: 6 August 2024; Accepted: 4 February 2025; Published online: 15 May 2025)

ABSTRACT: Rapid technological advancements have led to the widespread deployment of wireless sensor networks (WSNs) in industrial environments, making cybersecurity a critical concern in cloud computing. This paper presents a predictive framework for cloud-based intrusion detection and prevention for WSNs. It integrates machine learning models—Multilayer Perceptron (MLP), Decision Tree, and Autoencoder—to precisely classify and mitigate various impacts of cyber intrusions on a cluster of wireless sensors. An intelligent prioritization and prevention system is also proposed, categorizing attacks—blackhole, grayhole, flooding, and scheduling—based on their impact on industrial processes. Experimental results indicate robust detection capabilities, with the Decision Tree achieving 99.48% accuracy, slightly outperforming MLP at 99.37%. The Autoencoder demonstrated superior binary classification, distinguishing between normal and anomalous instances with high precision and recall rates. This framework leverages the WSN-DS dataset to simulate and validate its efficiency in mitigating real-time threats. Future work will focus on refining the prioritization model and integrating advanced machine learning techniques for enhanced adaptability and resilience.

ABSTRAK: Kemajuan pesat dalam teknologi telah membawa kepada penggunaan meluas rangkaian penderia wayarles (WSN) dalam persekitaran industri, menjadikan keselamatan siber sebagai kebimbangan kritikal dalam pengkomputeran awan. Kajian ini membentangkan rangka kerja ramalan bagi mengesan dan mencegah pencerobohan berasaskan awan untuk WSN. Ia menyepadukan model pembelajaran mesin—Perseptron Berbilang Lapis (MLP), Pokok Keputusan (Decision Tree) dan Enkoder Automatik (Autoencoder)—bagi klasifikasi tepat dan pengurangan pelbagai kesan pencerobohan siber pada kelompok penderia wayarles. Sistem keutamaan dan pencegahan pintar turut dicadangkan, mengkategorikan serangan—lubang hitam, lubang kelabu, banjir dan penjadualan—berdasarkan kesan terhadap proses industri. Dapatan eksperimen menunjukkan keupayaan pengesanan yang mantap dengan Decision Tree mencapai ketepatan 99.48%, sedikit mengatasi prestasi MLP pada 99.37%. Autoencoder menunjukkan klasifikasi binari yang unggul, membezakan antara kejadian biasa dan anomali dengan ketepatan tinggi dan kadar ingatan semula. Rangka kerja ini memanfaatkan set data WSN-DS bagi simulasi dan pengesahan kecekapan dalam mengurangkan ancaman masa nyata. Kajian akan menumpukan pada memperhalusi model keutamaan dan menyepadukan teknik pembelajaran mesin lanjutan pada masa hadapan bagi kebolehsuaian dan daya tahan yang tinggi.

https://doi.org/10.31436/iiumej.v26i2.3393

KEYWORDS: *Wireless Sensor Networks, Cloud, Security, Deep learning, and Predictive technique.*

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have emerged as a critical technology in various domains, including environmental monitoring, healthcare, industrial automation, and military applications. However, the clusters of wireless sensor nodes are vulnerable to multiple security threats due to their inherent characteristics, such as limited computational power, energy constraints, and wireless communication. With the increasing adoption of cloud computing, it becomes essential to explore its potential for enhancing the security of WSNs. Technological innovations like artificial intelligence (AI), machine learning, and augmented reality are integral to Industry 4.0. These technologies provide intelligent assistance, streamline processes, and increase productivity across various industrial tasks [1]. A Wireless Sensor Network (WSN) is the interconnection of wireless sensors and devices. These networks are vital for collecting, transmitting, and analyzing real-time data essential for process optimization and decision-making. WSNs act as the industrial setup's nerve center, gathering data on motion, temperature, pressure, and other parameters. They enable seamless communication between devices, systems, and people, offering greater scalability and flexibility than traditional wired systems. WSNs can be easily expanded, reconfigured, and adapted to various industrial environments [2]. However, WSNs face several Internet security challenges. The increased number of device connections raises vulnerability to cyberattacks, necessitating strong cybersecurity measures to protect sensitive data. Signal interference and reliability issues in complex industrial environments affect network performance. Using battery-powered sensors poses challenges, as maintaining prolonged battery life and energy efficiency is crucial for consistent and reliable operation.

Due to the convergence of digital technologies, WSNs are crucial for real-time data transfer. This research is motivated by the dynamic and evolving nature of cyber threats. This requires framework development with well-defined priorities. Several Internet attacks pose a greater danger and have a more significant impact on the cloud environment. This research addresses the dynamic nature of cyber threats in WSNs, proposing a framework that leverages predictive techniques to prioritize and mitigate attacks based on their severity and impact. Unlike traditional reactive systems, this framework aims to proactively detect and neutralize threats, ensuring minimal disruption to industrial processes.

Integrating cloud computing with wireless sensor networks (WSNs) has become increasingly prevalent among researchers due to the enhanced processing capabilities and scalable resources that the cloud infrastructure provides. This integration simplifies collecting, storing, and analyzing large volumes of data generated by sensor nodes. However, the simplification also introduces additional security challenges, especially in situations involving limited computational resources and energy constraints of sensor nodes. To ensure data transmission security, storage, and processing, it is critical to have high integrity, confidentiality, and availability of sensitive information. This study leverages predictive techniques like machine learning and data analytics to identify potential security threats and vulnerabilities within a wireless sensor network. These algorithms can improve security by predicting and mitigating attacks before a system is compromised. However, implementing such an algorithm involves examining complex issues, such as trade-offs between security and performance, the accuracy of predictive models, and the efficient management of limited resources.

https://doi.org/10.31436/iiumej.v26i2.3393

There are numerous open issues related to this research. These include the resource constraints of sensor nodes. The sensor nodes are challenged by processing power, memory, and battery life in this context. These constraints must be considered when implementing security mechanisms and predictive techniques to avoid node overload and resource depletion. Also, data storage and transmission challenges must be addressed by proposing an effective strategy or algorithm to minimize energy consumption while ensuring data integrity and security. The security of the internet cloud is paramount. This means data integrity and confidentiality must be assured when they are generated from the sensors. They must not be tampered with or be vulnerable to any threats during transmission. A Robust Access Control management strategy must be implemented to prevent unauthorized access to sensitive data stored in the cloud. The network system must have a better detection and mitigation strategy against attacks, such as denial-of-service (DoS), eavesdropping, and data tampering. Addressing these open issues in this paper will develop secure and efficient cloud computing-based systems that manage wireless sensor node clusters. Using predictive techniques can be a promising avenue to enhance security.

2. LITERATURE REVIEW

This section reviews relevant literature on cloud computing-based security analysis for wireless sensor node clusters using predictive techniques. Integrating cloud computing with WSNs offers numerous benefits, including scalable storage, robust computational capabilities, and centralized management. However, the distributed nature and resource limitations of WSNs introduce significant security challenges. Recent studies highlight applying predictive techniques, such as machine learning and data analytics, to provide proactive security measures by identifying potential threats before they cause harm. This review also examines current trends, challenges, and solutions associated with leveraging cloud computing and predictive approaches for WSN security.

2.1. Security Threats in WSN

Wireless Sensor Networks (WSNs) are inherently vulnerable to various cybersecurity threats due to their distributed nature, limited computational resources, energy constraints, and typical deployment in hostile or unattended environments. Common threats include eavesdropping, where unauthorized parties intercept data transmissions, potentially leading to privacy breaches or data manipulation [3]. Another significant threat is node capture, where attackers physically access nodes to extract sensitive information, such as cryptographic keys, which can then be used to launch further attacks or manipulate network operations [4], [5]. Denial-of-Service (DoS) attacks overload networks or specific nodes with traffic, disrupting services and depleting resources, rendering nodes unavailable for legitimate tasks.

More advanced threats include blackhole attacks, where compromised nodes drop all received packets, causing data loss and potential network isolation, and grayhole attacks, which selectively alter or drop packets, complicating threat detection [6]. Sybil attacks involve a malicious node presenting multiple identities to disrupt data aggregation, voting, or routing protocols. Wormhole attacks tunnel messages between network locations to create false views of the network, misleading routing protocols and disrupting operations. Sinkhole attacks occur when a compromised node falsely advertises an optimal route to attract traffic, leading to data interception or denial of service. Additionally, Hello flood attacks exploit routing protocols by sending or replaying numerous "Hello" packets, causing energy depletion as nodes attempt to respond.

https://doi.org/10.31436/iiumej.v26i2.3393

Mitigation strategies for these threats include encryption, authentication, intrusion detection systems (IDS), secure routing protocols, redundancy, and multipath routing. However, given the resource constraints in WSNs, these solutions must be lightweight and efficient to ensure their practicality [6].

2.2. Cloud Computing in WSNs

Integrating cloud computing with Wireless Sensor Networks (WSNs) brings significant advantages, opening new possibilities for data management, analysis, and application deployment [7]. By leveraging the cloud's computational power and storage capabilities, this integration addresses WSNs' inherent limitations, such as constrained processing power, memory, and energy efficiency. One of the key benefits of cloud computing in WSNs is its scalability, enabling administrators to efficiently handle the growing volumes of data generated by large-scale WSN deployments without overwhelming local resources.

Cloud platforms also provide robust storage and data management capabilities, allowing for long-term storage of sensor data and facilitating historical data analysis using tailored applications [8]. Furthermore, cloud-based solutions enhance data processing and analytics, offering powerful computational resources for real-time processing, machine learning, and big data analytics to extract actionable insights. The cloud’s remote accessibility further enables the centralized management and monitoring of WSNs without requiring physical proximity. However, these benefits come with potential security vulnerabilities, such as unauthorized access and data breaches, highlighting the necessity for secure transmission protocols and reliable intrusion detection mechanisms to safeguard sensitive data.

2.3. Predictive Security Techniques

Predictive security techniques utilize advanced analytics, machine learning (ML), and artificial intelligence (AI) to anticipate and mitigate cyber threats before they occur. These techniques identify vulnerabilities, anomalous patterns, and emerging threats by analyzing historical and real-time data [9]. To identify potential security risks, threat intelligence and analysis aggregate data from various sources, such as threat feeds, dark web monitoring, and incident reports. In application, threat intelligence helps create predictive models recognizing early indicators of potential attacks or vulnerabilities.

Machine learning and AI-based anomaly detection algorithms identify deviations from normal network traffic and system process behavior. This approach is effective in detecting zero-day attacks, insider threats, and other unknown risks by recognizing unusual patterns that may signal malicious activity. Specifically, network traffic patterns are analyzed to detect anomalies, suspicious communications, or signs of a security breach [10]. This helps identify command-and-control communications, data exfiltration, and other malicious activities.

In predictive security, machine learning models, such as neural networks, decision trees, and autoencoders, are particularly effective in identifying abnormal patterns in network traffic. These models enable proactive defenses, reducing the likelihood of successful cyberattacks. Research into predictive techniques for security analysis, especially in wireless sensor networks (WSNs), has shown promising results for threat identification and mitigation. For example, a machine learning-based approach for anomaly detection in WSNs, using supervised learning algorithms to classify normal and abnormal behaviors, is presented in [14]. The study demonstrates the high accuracy of predictive models in detecting intrusions. Additionally, [15] examines neural networks for intrusion detection in WSNs, proposing a hybrid model that combines feature selection with neural network classifiers to enhance detection performance while minimizing computational overhead. Furthermore, predictive analytics in cloud

environments, focusing on cloud-based machine learning models to analyze large volumes of sensor data, is demonstrated in [15], aiming to predict and prevent security incidents in real-time.

2.4. Summary of Related Work

Cloud computing-based analysis of security has been leveraged in WSNs to overcome resource constraints. It is also leveraged to mitigate scalable storage and enable sophisticated data processing on the internet. In [11], the authors develop a cloud-based architecture to manage WSNs. It also highlights how cloud services can centralize data storage and processing. In addition, it examines offloading computational tasks from resource-limited sensor nodes. Research in [12] discusses the advantages of using cloud computing for data aggregation and analytics. This facilitates additional efficiency in network management and decision-making processes. However, considering security threats and challenges in WSNs, some vulnerabilities of sensors with limited energy resources, inefficient computational capabilities, and susceptibility to physical tampering have been extensively studied. Some common security threats in WSNs are examined in [13]. These include eavesdropping, data injection, node compromise, and denial-of-service (DoS) attacks. In the research, more emphasis is placed on lightweight security mechanisms to operate within WSN constraints.

Table 1. Summary of Related Works

Ref	Contribution	Results	Limitation
[1]	The paper examines potential replay-attack vulnerabilities in SECS/GEM systems by proposing a detection and prevention system against attacks.	Achieved an effective mechanism that identifies and prevents replay attacks on SECS/GEM communications.	However, the authors were unable to propose an effective mechanism that resists security attacks such as DoS, forgery, modification, and man-in-the-middle attacks. These are suggested for future work.
[2]	The paper proposed a hierarchical intrusion detection algorithm to group sensor nodes based on functional assignment to lessen energy consumption during threat detection.	Presented an ideal system for WSNs with limited resources. The detection interval is moderate, and detection accuracy is high.	However, the authors were unable to enhance the detection of multiple intrusion patterns.
[3]	The paper presented a framework that combines preventative and deterrent strategies to reduce the danger of insider attacks.	Achieved a better regulatory framework on information security ethics.	However, the authors were unable to examine intrusion detection and prevention. The focus was only centered on the human and behavioral aspects of cyber misconduct and security.
[4]	The paper presented a detection strategy for creating an active defense system using deception technology. The framework can be used to conceptualize a Hybrid Threat Model.	Illustrated how deception is used to validate network resilience. Using in-network deception for threat detection exhibits how attack information can be generated to accelerate incident response and strengthen network defenses.	However, the authors were unable to manage the issues of reactive defenses, such as intrusion detection systems, which are prone to false positives and potentially lead to analyst alert fatigue and decreased effectiveness.

2. METHODOLOGY

This section provides a framework design for developing a cloud-based architecture that uses predictive techniques to integrate WSN with the Internet cloud service. This includes a

https://doi.org/10.31436/iiumej.v26i2.3393

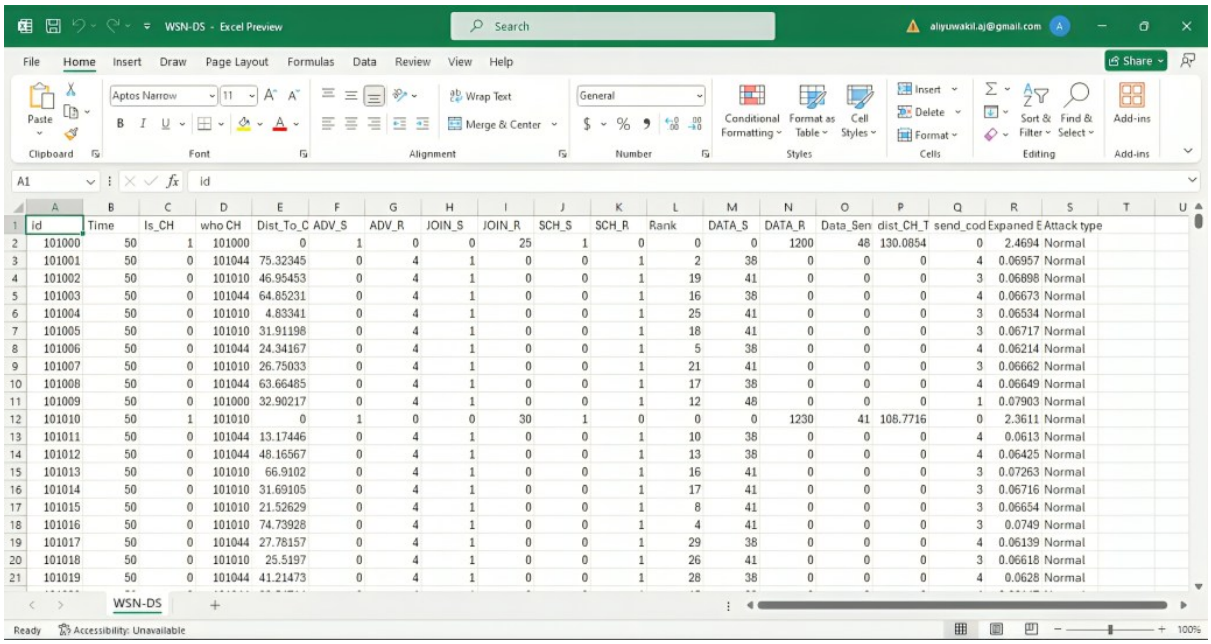
dataset aggregation design pattern using a communication protocol to effectively transmit data from a sensor node to the Internet cloud and prevent attacks and intrusions. During implementation, integrating machine learning algorithms such as decision trees, support vector machines, multilayer perceptrons (MLP), autoencoders, and neural network models to accurately detect and classify different types of attacks, including blackhole, grayhole, flooding, and scheduling attacks, is also conducted. This can potentially predict security threats on sensor data. In terms of implementation, we deployed the designed predictive model in a testbed environment using a cluster of wireless sensor nodes and the Internet cloud.

2.1. Design of the Proposed Methodology

The design describes the dataset, system architecture, block diagram of the system, flow chart, and components required for successful implementation.

2.1.1. Dataset

The models implemented in this framework will be evaluated with WSN-DS. This open-source dataset from Kaggle is for intrusion detection systems in wireless sensor networks. This dataset simulates various denial-of-service (DoS) attacks in WSNs using the LEACH (Low Energy Adaptive Clustering Hierarchy) protocol. We further determine the impact of the framework dataset by dividing it into two sections. These are: the training set and the test set. The training set comprises 80% of the total records in the dataset and will be used to train our models. However, the test set comprises 20% of the records and will be used to test and validate the model.



id	Time	Is_CH	who CH	Dist_To_C	ADV_S	ADV_R	JOIN_S	JOIN_R	SCH_S	SCH_R	Rank	DATA_S	DATA_R	Data_Sen	dist_CH_T	send_cod	Expaned E	Attack type
101000	50	1	101000	0	1	0	0	25	1	0	0	0	1200	48	130.0854	0	2.4694	Normal
101001	50	0	101044	75.32345	0	4	1	0	0	1	2	38	0	0	0	4	0.06957	Normal
101002	50	0	101010	46.95453	0	4	1	0	0	1	19	41	0	0	0	3	0.06898	Normal
101003	50	0	101044	64.85231	0	4	1	0	0	1	16	38	0	0	0	4	0.06673	Normal
101004	50	0	101010	4.83341	0	4	1	0	0	1	25	41	0	0	0	3	0.06534	Normal
101005	50	0	101010	31.91198	0	4	1	0	0	1	18	41	0	0	0	3	0.06717	Normal
101006	50	0	101044	24.34167	0	4	1	0	0	1	5	38	0	0	0	4	0.06214	Normal
101007	50	0	101010	26.75033	0	4	1	0	0	1	21	41	0	0	0	3	0.06662	Normal
101008	50	0	101044	63.66485	0	4	1	0	0	1	17	38	0	0	0	4	0.06649	Normal
101009	50	0	101000	32.90217	0	4	1	0	0	1	12	48	0	0	0	1	0.07903	Normal
101010	50	1	101010	0	1	0	0	30	1	0	0	0	1230	41	108.7716	0	2.3611	Normal
101011	50	0	101044	13.17446	0	4	1	0	0	1	10	38	0	0	0	4	0.0613	Normal
101012	50	0	101044	48.16567	0	4	1	0	0	1	13	38	0	0	0	4	0.06425	Normal
101013	50	0	101010	66.9102	0	4	1	0	0	1	16	41	0	0	0	3	0.07263	Normal
101014	50	0	101010	31.69105	0	4	1	0	0	1	17	41	0	0	0	3	0.06716	Normal
101015	50	0	101010	21.52629	0	4	1	0	0	1	8	41	0	0	0	3	0.06654	Normal
101016	50	0	101010	74.73928	0	4	1	0	0	1	4	41	0	0	0	3	0.0749	Normal
101017	50	0	101044	27.78157	0	4	1	0	0	1	29	38	0	0	0	4	0.06139	Normal
101018	50	0	101010	25.5197	0	4	1	0	0	1	26	41	0	0	0	3	0.06618	Normal
101019	50	0	101044	41.21473	0	4	1	0	0	1	28	38	0	0	0	4	0.0628	Normal

Figure 1. Intrusion Detection Systems Dataset for Wireless Sensor Networks

2.1.2. Block Diagram

The block diagram provides a high-level overview of the components' interactions within the wireless sensor network system. The primary components include the Raspberry Pi 3 microcontroller, sensors, power source, and the MCP 3008 Analog-to-Digital converter.

https://doi.org/10.31436/iiumej.v26i2.3393

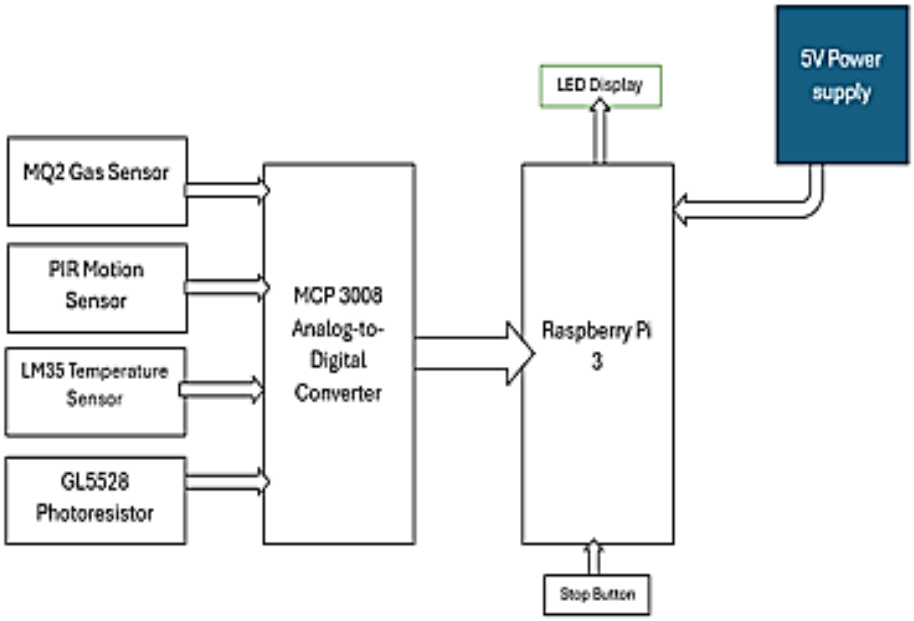


Figure 2. Block Diagram for Cloud-based WSN using Deep learning

Based on Figure 2, the Raspberry Pi microcontroller implements the intrusion detection and prevention logic. The Raspberry Pi is ideal for this project due to its affordability, versatility, and built-in connectivity, which make it suitable for prototyping and deploying IoT-based systems. It has extensive community support that provides valuable resources for troubleshooting and development. The Raspberry Pi has built-in Wi-Fi and Ethernet connectivity, which is essential for creating networked systems. This feature is vital for simulating real-world Industry 4.0 environments, where seamless communication between devices is a cornerstone. The Raspberry Pi can also run various AI and machine learning frameworks, such as TensorFlow and scikit-learn. This compatibility is essential for implementing the AI-driven detection and classification models integral to the project. Using the Raspberry Pi allows for the development of a prototype that can be scaled and implemented in real-world industrial settings. Its portability and ease of use make it an ideal choice for testing and demonstrating the proposed cybersecurity framework.

For the MCP 3008 Analog-to-Digital Converter, since the Raspberry Pi is fundamentally a digital device, any I/O done through its GPIO pins will happen through high (one) and low (zero) states. When input signals are analog (as in the case of our WSNs), they need to be converted to the digital domain so the Raspberry Pi can understand them. The MCP3008 is a 10-bit 8-channel analog-to-digital converter chip that performs this operation.

The Stop Button is connected to one of the Raspberry Pi's GPIO pins. Its function serves as an interrupt/end button for the code's implementation. The algorithm is set to run continuously until a high-priority threat is detected and preventive actions are implemented. The user can end the program/system at their convenience by providing a stop button.

Based on the prototype sensors, the WSN comprises four different sensor types, each of which outputs an analog signal. The choice of these four specific sensors was made with several considerations. Each sensor is meant to simulate one of the four attack types. This framework is trained to conduct smart detection, which means blackhole, flooding, gray hole, and scheduling attacks.

https://doi.org/10.31436/iiumej.v26i2.3393

2.1.3. Flowchart

The flowchart outlines the sequential steps in the system's operation, from data acquisition to decision-making and action execution, as shown in Figure 3.

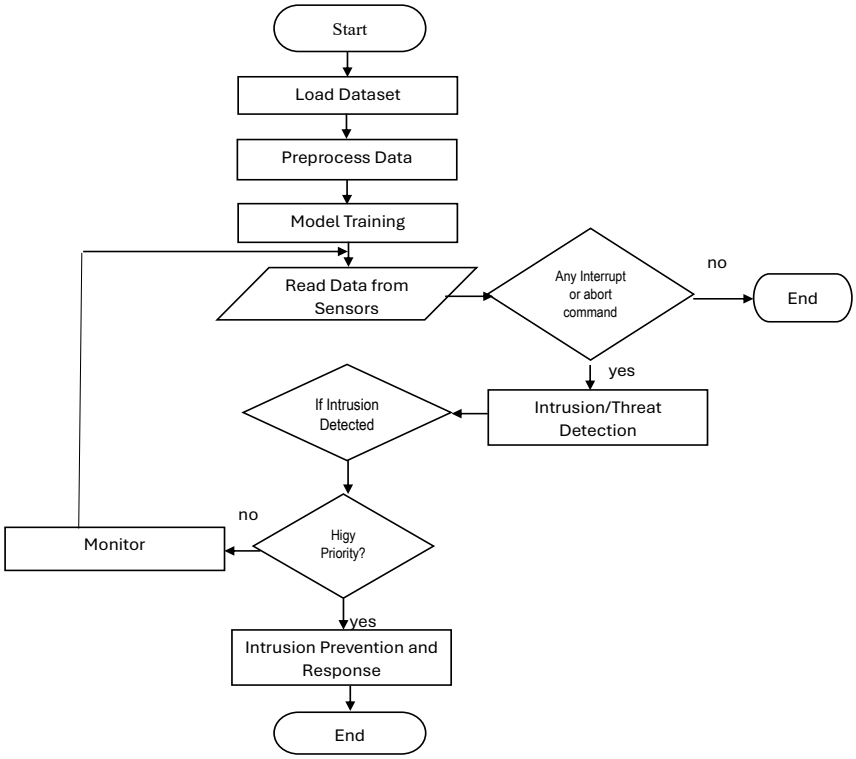


Figure 3. Flowchart for cloud-based WSN using deep learning

We initialize the Raspberry Pi microcontroller and other design components in the simulation process. The dataset from a different Excel application is loaded and uploaded using the import command on the WSN-DS dataset. For data preprocessing, we run commands such as clean, normalize, and prepare a dataset for analysis. This includes extracting features (X) and labels (y) from the data frame. Data is split into training and testing sets, and standardization is applied. The next process is model training. This is where the core logic of the framework is implemented. Three models are used to train the framework comprehensively: the autoencoder, decision tree, and multilayer perceptron. We present additional information for the model training discussion in this paper's implementation section. Also, data is read from the sensors and preprocessed. If there is no command from the user to halt execution, then proceed to intrusion detection; otherwise, terminate. We then check input from sensors against the trained data from the models. If the result comes back as positive, i.e., intrusion/threat is detected, then check for the type of attack and the priority assigned to it. The priority of an attack depends on what kind of industry the system is being used in, and as such, is dynamic. Also, if the attack is of high priority, preventive actions should be immediately implemented, and the program should end. Then, if the attack is of lower priority, raise a warning and closely monitor the inputs from that sensor while continuing to take inputs.

2.2. Implementation

Implementing the proposed methodology involves using a simulation tool to model the cloud-based intrusion detection and prevention framework using predictive techniques and a

https://doi.org/10.31436/iiumej.v26i2.3393

deep learning algorithm on wireless sensors. We also conduct performance evaluations of some parameters to deploy on a functional system prototype.

2.2.1. Simulation Tool

The implementation utilizes a QEMU-based simulation environment to replicate real-world scenarios. Python libraries, including TensorFlow and scikit-learn, support developing and testing machine learning models. The simulation evaluates the system's performance under various cyberattack scenarios, ensuring its robustness and reliability.

2.2.2. Data Collection & Preprocessing

The WSN-DS dataset underwent preprocessing steps to improve data quality and consistency. These steps included:

- Normalization: Ensuring uniform data scales.
- Feature Extraction: Identifying relevant attributes for model training.
- Splitting: Dividing the dataset into 80% training and 20% testing sets.

This is done to have reliable information supplied on the wireless sensors fully secured in the cloud environment. Both real-time and historical data are used. The process of loading the dataset and preprocessing for the wireless sensors is presented based on the script in Figure 4.

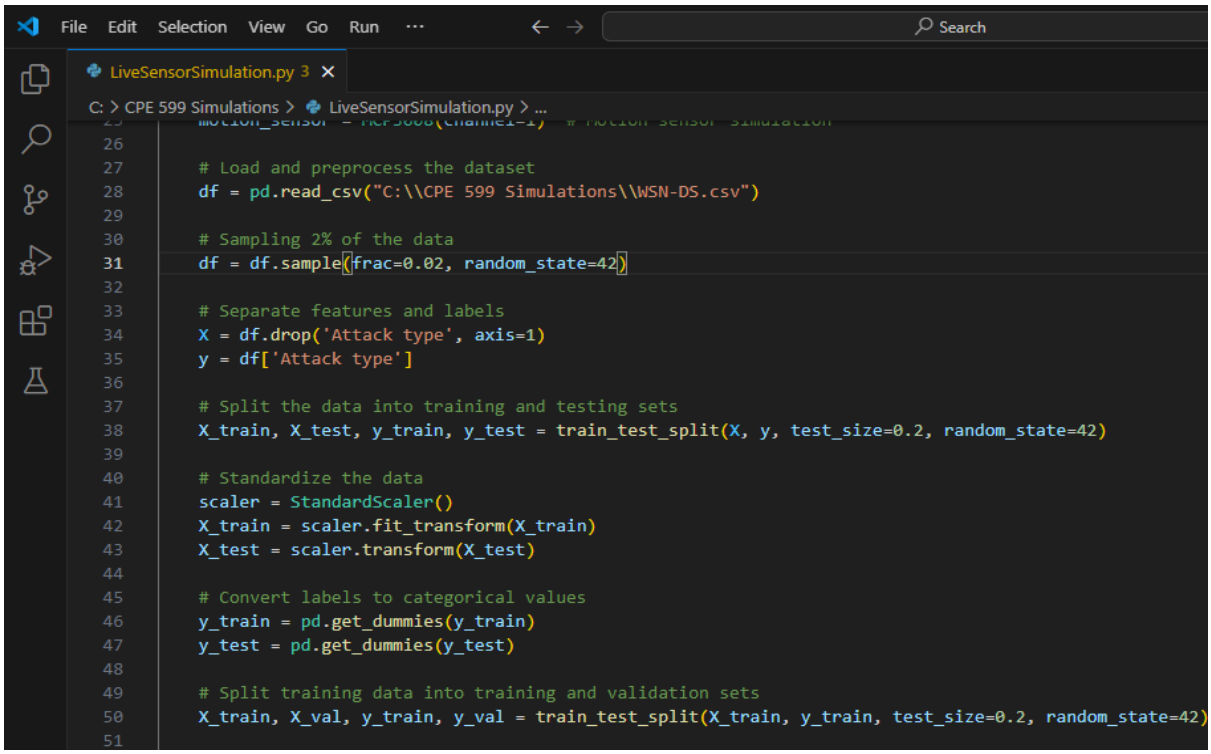


Figure 4. Loading the WSN-DS dataset and preprocessing

The historical data helps understand patterns, while real-time data detects ongoing threats in the cloud environment. Data preprocessing includes cleaning, filling gaps, removing duplicates, correcting inconsistencies, and normalizing data for machine learning. Based on cybersecurity techniques and cloud services, extracting relevant features from raw data involves various algorithm choices and identifying potential threats or anomalies.

https://doi.org/10.31436/iiumej.v26i2.3393

2.2.3. AI-Based Detection and Classification for Cloud Intrusions

An AI-driven intrusion detection and classification system safeguards Industry 4.0 WSNs. Three different AI models, namely, Decision Tree, Multilayer Perceptron (MLP), and Autoencoder, are implemented using a publicly available WSN dataset. Cybersecurity intrusions aredetected and classified with a particular emphasis on flooding, scheduling, black hole, and gray hole attacks. Each selected model has a distinct function in recognizing and categorizing cyberattacks, as shown in Figure 5. The Python library and dynamic AI libraries were leveraged due to their strong AI modeling support, particularly (sklearn, Tensorflow, Pandas, Numpy, and Keras).

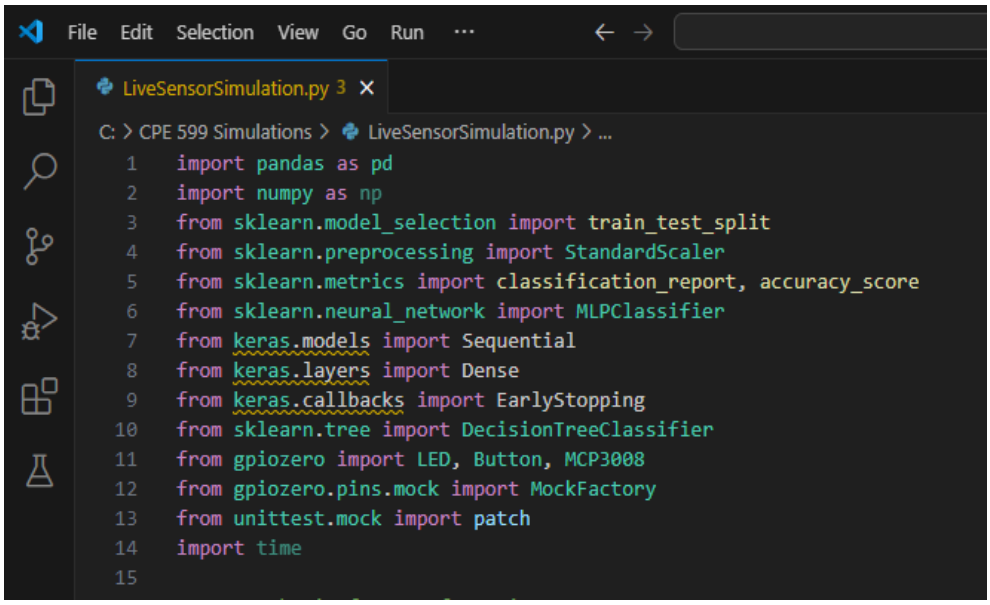


Figure 5. Importing AI and Machine Learning Libraries

2.2.4. The Autoencoder

This unsupervised learning approach is applied to data compression and feature learning.It is a powerful feature that presents adequate results for anomaly detection in the cloud environment. It also recognizes anomalies or deviations from the learned patterns by recreating the input data, as adopted in [1].

https://doi.org/10.31436/iiumej.v26i2.3393

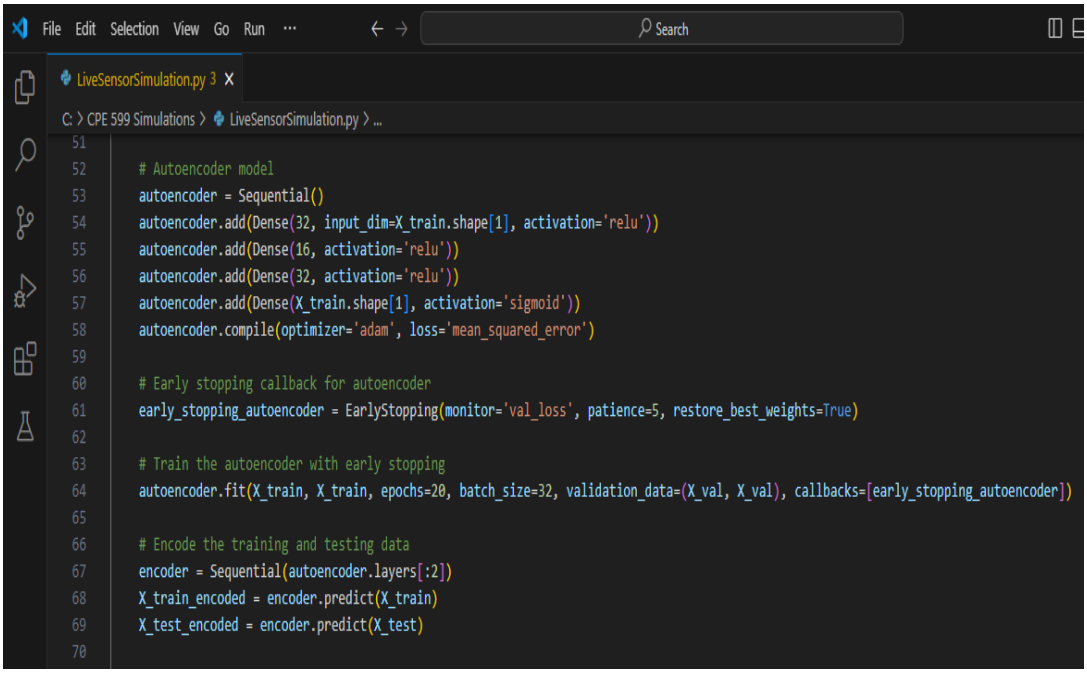


Figure 6. The Autoencoder Model Definition and Training

2.2.5. Autoencoder Model Definition:

`Sequential()` initializes a linear stack of layers.

`Dense()` layers are fully connected layers.

`input_dim=X_train.shape[1]` specifies the input dimension for the first layer based on the number of features in X_train.

`activation='relu'` uses Rectified Linear Unit activation function for hidden layers.

`activation='sigmoid'` uses Sigmoid activation function for the output layer to reconstruct input data between 0 and 1.

`compile (optimizer='adam', loss='mean_squared_error')` configures the model for training with Adam optimizer and Mean Squared Error loss function.

`EarlyStopping` is a callback that stops training when a monitored metric has stopped improving.

`monitor='val_loss'` monitors validation loss.

`patience=5` waits for 5 epochs after the validation loss has stopped improving.

`restore_best_weights=True` restores model weights from the epoch with the best value of the monitored quantity.

2.2.6. Training the Autoencoder:

`fit()` trains the autoencoder model on X_train with itself (X_train) as both input and output.

`epochs=20` specifies the number of training epochs.

`batch_size=32` determines the number of samples per gradient update.

`validation_data=(X_val, X_val)` uses X_val for validation during training.

`callbacks=[early_stopping]` applies early stopping during training to prevent overfitting.

2.2.7. Encoding Data:

`Sequential(autoencoder.layers[:2])` creates an encoder model using the first two layers of the trained autoencoder.

https://doi.org/10.31436/iiumej.v26i2.3393

`encoder.predict(X_train)` and `encoder.predict(X_test)` encode `X_train` and `X_test` data into compressed representations (`X_train_encoded` and `X_test_encoded`) using the trained encoder model.

2.2.8. Multilayer Perceptron (MLP)

A Multilayer Perceptron is a type of neural network that is highly accurate at finding complicated patterns and connections in data. Its ability to learn from both structured and unstructured data makes it useful for finding simple and complicated attackpatterns [3]. As shown in Figure 7, the MLP Classifier Implementation is explained as follows.

2.2.9. MLP Classifier Definition:

`MLPClassifier` is a multi-layer perceptron classifier from `sklearn.neural_network` module.

`hidden_layer_sizes=(64, 32)` defines two hidden layers with 64 and 32 neurons, respectively.

`max_iter=200` sets the maximum number of iterations for training.

`activation='relu'` uses Rectified Linear Unit activation function for hidden layers.

`solver='adam'` uses Adam optimizer for training.

`random_state=42` sets the random seed for reproducibility.

2.2.10. Training the MLP Classifier:

`mlp.fit(X_train_encoded, y_train)` trains the MLP classifier on the encoded training data (`X_train_encoded`) and corresponding labels (`y_train`).

2.2.11. Evaluating the MLP Classifier:

`mlp.predict(X_test_encoded)` predicts labels for the encoded test data (`X_test_encoded`).

`accuracy_score(y_test, mlp_predictions)` computes the accuracy of predicted labels (`mlp_predictions`) compared to true labels (`y_test`).

`classification_report(y_test, mlp_predictions)` generates a detailed classification report including precision, recall, F1-score, and support.

2.2.12. Decision Tree

We use this model to organize data into a tree-like structure to conduct AI-based decisions on the smart sensors using predetermined conditions. The decision tree classification model is intuitive and can handle numerical and categorical data, making it suitable for classifying different types of cloud intrusion and detection attacks based on the specific parameters and characteristics we mentioned. This research presents the decision tree for this cloud service as follows.

https://doi.org/10.31436/iiumej.v26i2.3393

```
# MLP Classifier
mlp = MLPClassifier(hidden_layer_sizes=(64, 32), max_iter=200, activation='relu', solver='adam', random_state=42)
mlp.fit(X_train_encoded, y_train)

# Evaluate the MLP on the test set
mlp_predictions = mlp.predict(X_test_encoded)
print("MLP Test Accuracy:", accuracy_score(y_test, mlp_predictions))
print(classification_report(y_test, mlp_predictions))

# Decision Tree Classifier
dt = DecisionTreeClassifier(random_state=42)
dt.fit(X_train, y_train)

# Evaluate the Decision Tree on the test set
dt_predictions = dt.predict(X_test)
print("Decision Tree Test Accuracy:", accuracy_score(y_test, dt_predictions))
print(classification_report(y_test, dt_predictions))
```

Figure 7. MLP and Decision Tree Model Implementation

The Decision Tree Classifier Implementation in Figure 7 is explained below:

2.2.13. Decision Tree Classifier Definition:

`DecisionTreeClassifier` is a classifier from `sklearn.tree` module.

`random_state=42` sets the random seed for reproducibility.

2.2.14. Training the Decision Tree Classifier:

`dt.fit(X_train, y_train)` trains the Decision Tree classifier on the training data (`X_train`) and corresponding labels (`y_train`).

2.2.15. Evaluating the Decision Tree Classifier:

`dt.predict(X_test)` predicts labels for the test data (`X_test`).

`accuracy_score(y_test, dt_predictions)` computes the accuracy of predicted labels (`dt_predictions`) compared to true labels (`y_test`).

`classification_report(y_test, dt_predictions)` generates a detailed classification report including precision, recall, F1-score, and support.

2.2.16. Intelligent Prioritization and Prevention System

This paper implemented a smart prioritization and prevention system framework in a cloud environment, as shown in Figure 8. Based on their significance, this system categorizes various types of attacks (blackhole, grayhole, flooding, and scheduling). Blackhole attacks, for instance, are deemed critical in sectors like robotics control and energy management due to their potential to disrupt vital systems. Grayhole attacks, which selectively alter packets, are common in quality assurance and asset tracking, affecting data accuracy. Flooding attacks, prevalent in smart logistics and supply chain visibility, overload networks with excessive traffic. Scheduling attacks target manufacturing processes and healthcare equipment, impacting timing and scheduling systems. To proactively prevent such threats, attacks are prioritized based on their impact on Industry 4.0 scenarios. Prevention measures include setting up verification environments, utilizing validation tools, deploying traffic analysis and rate-limiting methods, and employing time synchronization procedures. These tailored prevention plans aim to enhance the safety and reliability of Wireless Sensor Networks (WSNs).

https://doi.org/10.31436/iiumej.v26i2.3393

3. RESULTS AND DISCUSSION

This paper uses deep learning to implement a predictive framework for cloud-based intrusion detection and prevention in wireless sensor networks. The concept of the deep learning algorithm used has significantly improved the accuracy and efficiency of cloud intrusion detection and prevention in WSNs. This simulation will demonstrate three distinct machine learning models implemented based on the cloud-based intrusion and prevention system for wireless sensors. These are: Decision Tree, Multilayer Perceptron (MLP), and Autoencoder. The evaluation results are highlighted below:

3.1. Multilayer Perceptron (MLP)

MLP presented the best classification approach for cloud-based WSN intrusions. As evidenced in Figure 9, this research achieves 99.37% simulation accuracy.

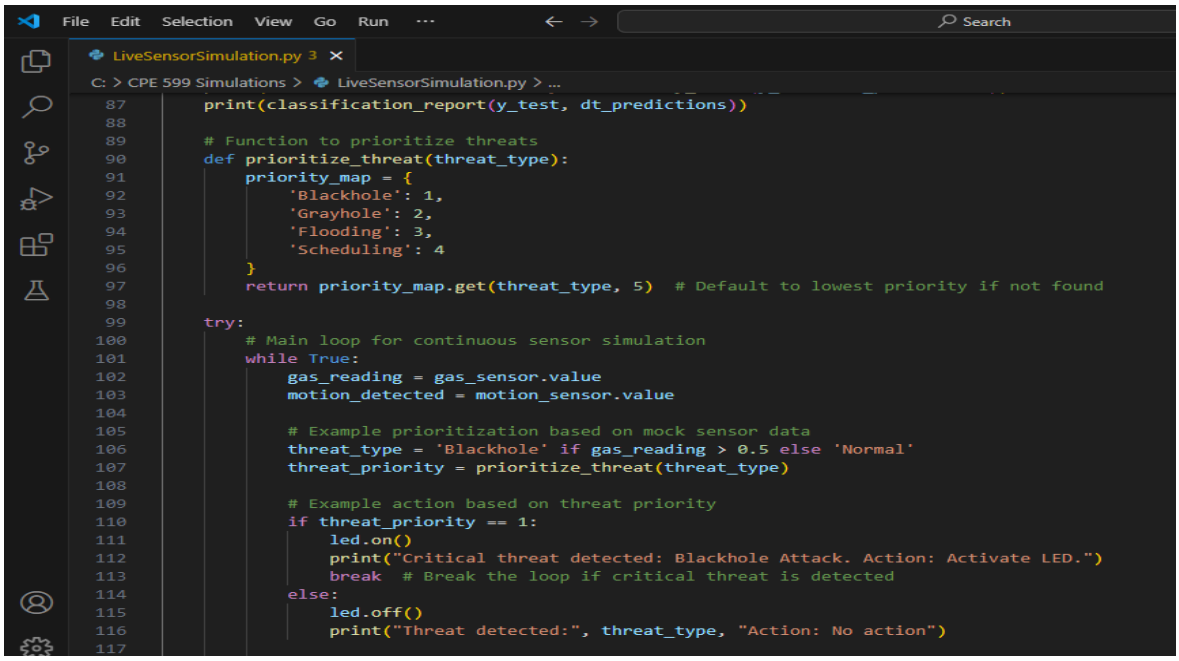


Figure 8. Intelligent Prioritization and Prevention Logic

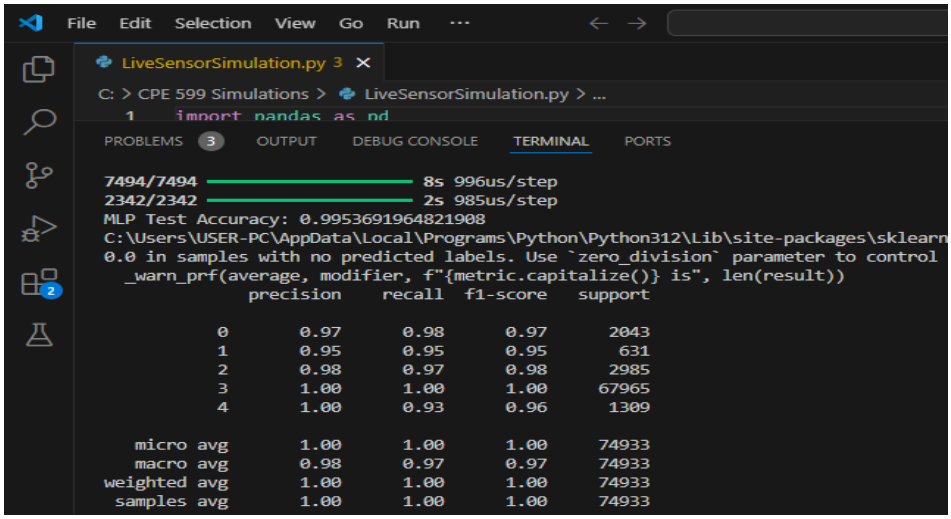


Figure 9. Accuracy, precision, recall, and F1 score for the MLP model

https://doi.org/10.31436/iiumej.v26i2.3393

Based on Figure 9, a simulation is conducted for accuracy, precision, recall, and F1 score for different attack types in the cloud computing environment. In the MLP model, Normal and TDMA classes were examined. This is to indicate high accuracy in positive predictions. However, the TDMA class has a slightly lower recall rate, suggesting that some real TDMA instances are missed. The Blackhole, Flooding, and Grayhole classes show a good balance between precision and recall, reflected in their high F1-Scores. The 'Macro Avg' and 'Weighted Avg' rows highlight the MLP model's consistently high performance across the dataset. This indicates the robust classification capabilities of the MLP model in this multiclass problem. In addition to the MLP model's precision, recall, and F1 scores per class, Figure 9 presents the Normal class with an exceptional precision of 99.7%, and the TDMA class also achieves 99.7%, demonstrating the model's effectiveness in reducing false positives. The Grayhole class has a precision of 94%, while the Blackhole and Flooding classes show strong precisions of 93% and 99.9%, respectively, indicating effective reduction of false positives.

3.2. Decision Tree (DT)

DT in this simulation presents effective accuracy during attack identification and prediction in cloud-based networks, which is crucial for cybersecurity (see Figure 10). The macro average and weighted average metrics demonstrate our model's overall high performance across the dataset, underscoring the robust capability of DT handling multiclass classification problems.

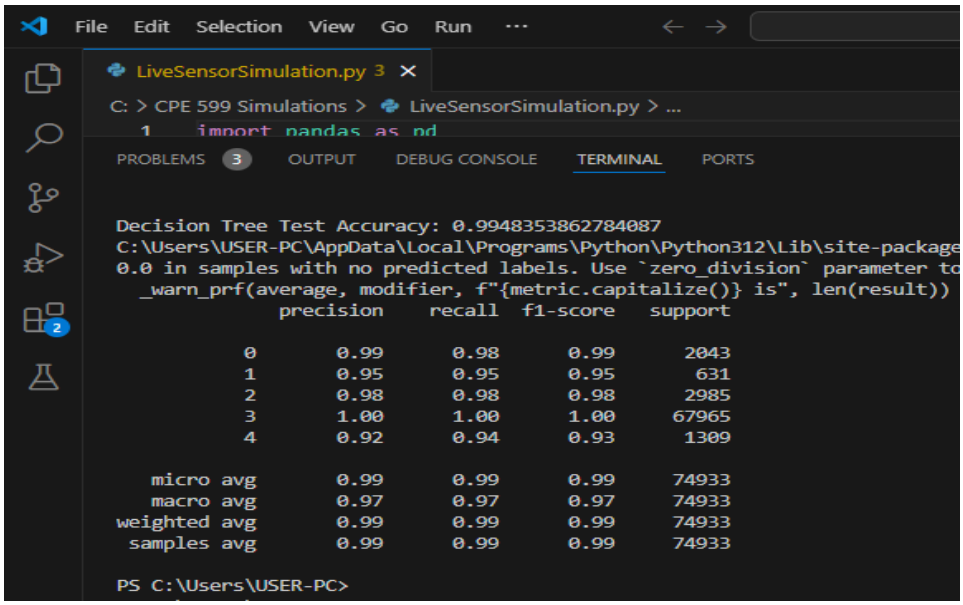


Figure 10. Impact of Decision Tree Classifier

Figure 10 shows a simulation based on accuracy, precision, recall, and F1 score for different attack types in the cloud environment. The results measure the precision score per class, where the Decision Tree is robustly classified in all instances. For precision, 99% is achieved for the "Normal" class. This indicates high accuracy in detecting typical occurrences and reducing false positives. For Blackhole and Grayhole classes, we achieved precisions of 95.82% and 98.89%. This shows its effectiveness in classifying all kinds of intrusions. The flooding class achieves a precision of 92%. This suggests a slightly higher chance of false positives. Overall, the Decision Tree model shows strong performance across all classes. In addition, the recall score per class is presented to indicate the model's ability to classify each class accurately. With high recall values for "Normal" (98%) and "Blackhole" (95%), the model

https://doi.org/10.31436/iiumej.v26i2.3393

effectively captures most real occurrences in these classes. For the Grayhole class, we achieved a recall of 98%. That means the model minimizes false negatives by efficiently identifying instances of network intrusions.

3.3. Autoencoder

As shown in Table 2, Autoencoder results highlight the performance metrics of a binary classification model that distinguishes between normal and anomalous instances. This simulation's precision for normal and anomalous instances is high, at 0.95 and 0.88. This indicates the model predicts an instance as normal or anomalous. The recall for normal and anomalous instances is 0.88 and 0.95, showing the model captures 88% of actual normal instances and 95% of actual anomalous instances. The F1-score for the Autoencoder model is 0.91 for normal instances and 0.92 for anomalous instances. This indicates enhanced balance between precision and recall. Therefore, the Autoencoder model outperforms other models during instance identification. Overall accuracy of the Autoencoder model is 0.91. This means it can correctly identify 91% of the total instances. For sensitivity, the Autoencoder model is 88%, and thus performs well in identifying positive instances out of all positive ones. For specificity, the Autoencoder model, being 95%, performs well in identifying negative instances out of all actual negative instances. Considering sensitivity and specificity, the Autoencoder model effectively identifies positive and negative instances. These results are presented in Table 2.

Table 2. Training and Validation Results of Autoencoder

Class	Precision	Recall	F1-Score	Accuracy	Sensitivity	Specificity
Normal	99%	88%	91%	91%	88%	95%
Anomaly	88%	95%	92%			
Macro Average	92%	91%	91%			
Weighted Average	92%	91%	91%			

As shown in Table 2, the training and validation of the Autoencoder model are conducted across 20 epochs, along with the reconstruction error report. An enhancement is noted for accuracy from 84.90% in the first epoch to 87.81% in the final epoch, with corresponding decreases in loss values. The validation accuracy mirrors this trend, reaching 87.81% by the end of the training period (this is only one of many iterations). This indicates that the model learns effectively and generalizes well to unseen data. The mean reconstruction error is 0.250247, which is presented using a standard deviation of 0.238110, suggesting that most reconstruction errors are low, though there is some variability. The minimum error is 0.038404, and the maximum error is 26.071553. The 25th, 50th, and 75th percentiles are 0.094400, 0.128891, and 0.233322, respectively, indicating that most reconstruction errors are below 0.233322. Furthermore, the Autoencoder Reconstruction Error Report is presented in Figure 11.

The performance of the Multilayer Perceptron (MLP) model significantly enhances the effectiveness of the Decision Tree model, resulting in improved accuracy, precision, recall, and F1 scores. This improvement is particularly evident in detecting and classifying cloud-based cybersecurity intrusions, underscoring the models' potential for robust threat identification. The findings indicate optimizing the MLP model yields better precision in detecting cloud-based intrusions within wireless sensor networks. Additionally, the results for the Decision Tree model are analyzed using various averaging methods, including micro, macro, weighted, and sample averages, to comprehensively evaluate its performance. These insights highlight the

https://doi.org/10.31436/iiumej.v26i2.3393

complementary strengths of the MLP and Decision Tree models in addressing cybersecurity challenges in cloud and wireless sensor network environments.

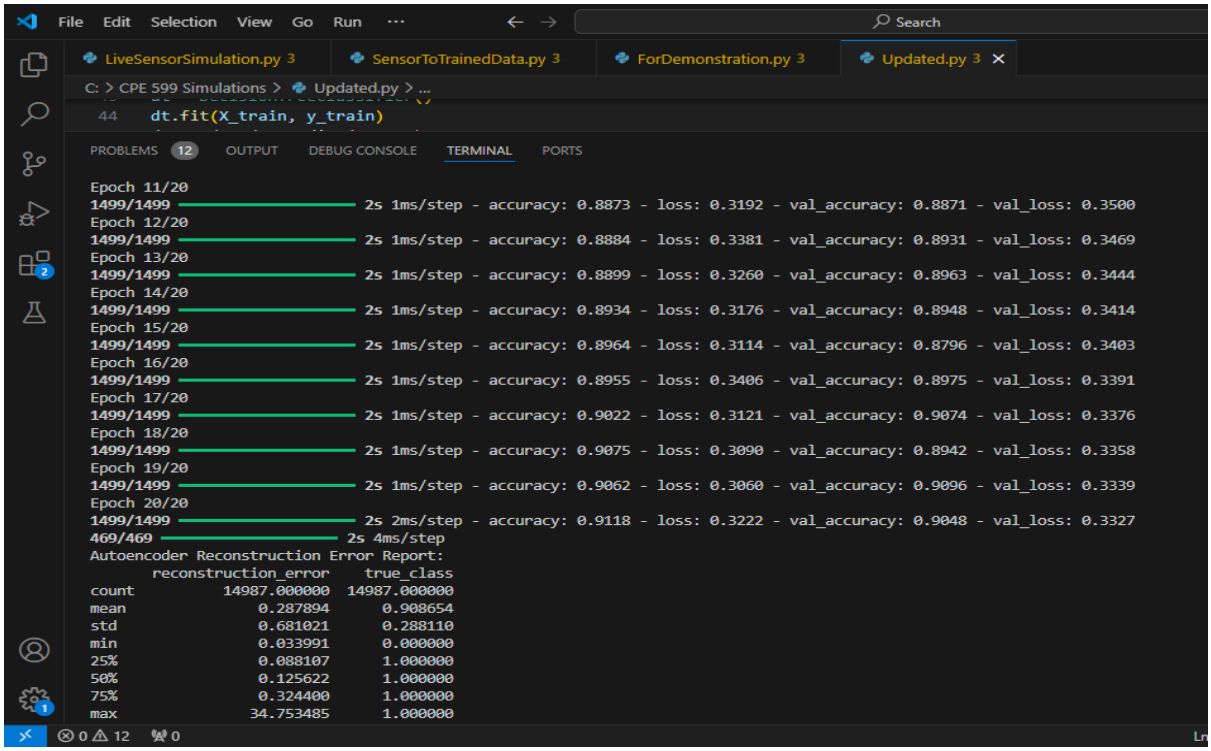


Figure 11. Autoencoder Reconstruction Error Report

3.4. Limitations

Observed limitations include increased computational delays on the Raspberry Pi during high-traffic scenarios. Addressing these limitations through hardware upgrades or algorithmic optimization will enhance the system's efficiency.

3.5. TESTING AND EVALUATION

Testing is conducted to evaluate the precision of the implemented predictive system. The evaluation considers a cloud computing scenario involving wireless sensor clusters in the cloud during intrusion detection and prevention actions. The technique involves, first, the use of an Internet cloud-based framework to secure and analyze real-time wireless sensor clusters. This includes detecting and classifying a wide range of cloud-based cybersecurity intrusions, such as data exfiltration, denial-of-service attacks, and unauthorized access attempts. Second, we perform intelligent prediction and prioritization to assess the framework's accuracy. This ensures that the model can accurately identify potential threats in the Internet cloud under specific algorithmic conditions. Finally, proactive measures are implemented to prevent attacks on the system's capacity, mitigating the impact of cloud-based breaches and enhancing real-time defense mechanisms through adaptive security approaches.

4. CONCLUSION AND FUTURE WORK

This paper outlined a comprehensive research study investigating cloud computing-based security analysis on wireless sensor node clusters using predictive techniques. It actively implements an intelligent and robust predictive framework for cloud-based intrusion detection

and prevention for Industry 4.0 using wireless sensor networks as follows. Firstly, an AI-based detection mechanism is implemented to recognize and classify wireless sensors using a cloud computing strategy and cybersecurity intrusions. Three distinct machine learning models (multilayer perceptron, autoencoder, and decision tree) are implemented for cybersecurity intrusion detection and classification within WSNs. Secondly, an intelligent prioritization model is implemented to give priorities to cyber threats based on their nature and impact. Finally, a prevention system is implemented to mitigate the impact of cybersecurity intrusions on the cloud effectively. The result presents significant advancements in WSN security, with implications for various applications. Leveraging advanced algorithms such as machine learning algorithms, intelligent prioritization mechanisms, and proactive safety measures enhances the framework's security and resilience against evolving cyber threats. Future research will explore advanced ensemble techniques to improve model performance, adaptive learning algorithms to enhance resilience against evolving threats, and real-time deployment scenarios to validate the framework's effectiveness in industrial environments.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS22-264-0873) (Grant No: FRGS/1/2022/ICT11/UIAM/01/1).

REFERENCES

[1] Dash, Sanjit Kumar, Subasish Mohapatra, and Prasant Kumar Pattnaik. "A survey on applications of wireless sensor network using cloud computing." *International Journal of Computer Science & Emerging Technologies* 1, no. 4 (2010): 50-55.

[2] John, Ayuba, Ismail Fauzi Bin Isnin, Syed Hamid Hussain Madni, and Muhammed Faheem. "Cluster-based wireless sensor network framework for denial-of-service attack detection based on variable selection ensemble machine learning algorithms." *Intelligent Systems with Applications* 22 (2024): 200381.

[3] Sareen, Sanjay, Sandeep K. Sood, and Sunil Kumar Gupta. "An automatic prediction of epileptic seizures using cloud computing and wireless sensor networks." Journal of medical systems 40 (2016): 1-18.

[4] Mohammed, Ahmed. "The Web Technology and Cloud Computing Security based Machine Learning Algorithms for Detect DDOS Attacks." *Journal of Information Technology and Informatics* 3, no. 1 (2024).

[5] Gayathri, S., and D. Surendran. "Unified ensemble federated learning with cloud computing for online anomaly detection in energy-efficient wireless sensor networks." *Journal of Cloud Computing* 13, no. 1 (2024): 49.

[6] Luo, Haonan, Jing Wang, Deyu Lin, Linghe Kong, Yufei Zhao, and Yong Liang Guan. "A Novel Energy-Efficient Approach Based on Clustering Using Grey Prediction in WSNs for IoT Infrastructures." *IEEE Internet of Things Journal* (2024).

[7] Yadav, Santosh Kumar, and Rakesh Kumar. "Scalable energy optimization of resources for mobile cloud computing using sensor enabled cluster-based system." *Wireless Networks* (2024): 1-26.

[8] Chellathurai Amirthabai, Subasini, Udita Malhotra, Saravanan Thapasimuthu Rajeswari, and Sudhahar Thachankurichy Natesan. "Healthcare security in cloud-based wireless sensor networks: Botnet attack detection via autoencoder-aided goal-based artificial intelligent agent." *Concurrency and Computation: Practice and Experience*: e8152.

[9] Alawi, Mahmoud A., Rashid A. Saeed, and Aisha A. Hassan. "Cluster-based multi-hop vehicular communication with multi-metric optimization." In *2012 international conference on computer and communication engineering (ICCCE)*, pp. 22-27. IEEE, 2012.

[10] Islam, Shayla, Aisha-Hassan Abdalla Hashim, Mohamed Hadi Habaebi, and Mohammad Kamrul Hasan. "Design and implementation of a multihoming-based scheme to support mobility management in NEMO." *Wireless Personal Communications* 95 (2017): 457-473.

[11] Hassan, Mona Bakri, Rashid A. Saeed, Othman Khalifa, Elmustafa Sayed Ali, Rania A. Mokhtar, and Aisha A. Hashim. "Green machine learning for green cloud energy efficiency." In *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, pp. 288-294. IEEE, 2022.

[12] Badawi, Ahmed SA, Nurul Fadzlin Hasbullaha, Y. Yusoff, Sheroz Khan, Aisha Hashim, Alhareth Zyoud, and Mohammed Elamassie. "Evaluation of wind power for electrical energy generation in the mediterranean coast of Palestine for 14 years." *International Journal of Electrical and Computer Engineering* (2019).

[13] Elagib, Sara B., Atahur Rahman Najeeb, Aisha H. Hashim, and Rashidah F. Olanrewaju. "Big data analysis solutions using MapReduce framework." In *2014 International Conference on Computer and Communication Engineering*, pp. 127-130. IEEE, 2014.

[14] Khalifa, Othman O., Adil Roubleh, Abdelrahim Esgiar, Maha Abdelhaq, Raed Alsaqour, Aisha Abdalla, Elmustafa Sayed Ali, and Rashid Saeed. "An IoT-platform-based deep learning system for human behavior recognition in smart city monitoring using the Berkeley MHAD datasets." *Systems* 10, no. 5 (2022): 177.

[15] Hashim, M. M., Mustafa Sabah Taha, Azana Hafizah Mohd Aman, Aisha Hassan Abdalla Hashim, Mohd Shafry Mohd Rahim, and Shayla Islam. "Securing medical data transmission systems based on integrating algorithm of encryption and steganography." In *2019 7th International Conference on Mechatronics Engineering (ICOM)*, pp. 1-6. IEEE, 2019.

https://doi.org/10.31436/iiumej.v26i2.3419

NUMERICAL ANALYSIS OF PARTIAL DISCHARGE BEHAVIOUR UNDER DC STRESS WITH VOLTAGE DISTURBANCE AT DIFFERENT VOID SIZES

NUR SHAHIDA MIDI*, MUHAMMAD ALIF SULAIMAN

Electrical and Computer Engineering Department, International Islamic University Malaysia, Kuala Lumpur, Malaysia

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

(Received: 23 August 2024; Accepted: 24 April 2025; Published online: 15 May 2025)

ABSTRACT: Partial discharge (PD) is commonly related to electrical insulator degradation. It occurs in a high electric field environment, especially in high-voltage systems. It can lead to the electrical breakdown of insulators. Knowledge of the characteristics of PD allows for testing and monitoring of insulation properties in power system equipment. PD can be observed in both AC and DC power. However, research on AC-PD is much more mature than DC-PD due to the established pattern of AC's changing magnitude and polarity characteristics. This work uses finite element analysis to study PD activity under DC stress with voltage disturbance at different void sizes, with AC harmonic as the focused disturbance. The characteristics are evaluated based on the electric potential distribution, electric field distribution, electrical charges, and repetition rate. As the void radius increases from 1 mm to 4 mm, the inception voltage decreases by approximately 65%, while the PD repetition rate increases by 58%. In addition, a comparison between DC-PD and AC harmonic disturbance and AC-PD is performed to analyze their differences. The simulation result shows that the repetition rate for DC-PD with AC harmonics is approximately 75% lower than that of AC-PD, indicating fewer PD events in the case of DC-PD with AC harmonics. The conducted simulation provides valuable insights and guidance for the formation of DC-PD testing, and consequently, a validated method can be approved to monitor insulating material condition under DC power.

ABSTRAK: Penyahcasan separa (partial discharge, PD) merupakan fenomena biasa yang berkait rapat dengan degradasi penebat elektrik. Ia berlaku dalam persekitaran medan elektrik tinggi, khususnya dalam sistem voltan tinggi, dan boleh menyebabkan kerosakan elektrik pada penebat. Pengetahuan mengenai ciri-ciri PD membolehkan ujian dan pemantauan sifat penebat dalam peralatan sistem kuasa dijalankan. PD boleh diperhatikan dalam kedua-dua kuasa arus ulang-alik (AC) dan arus terus (DC). Namun, kajian mengenai PD dalam AC adalah lebih matang berbanding dalam DC disebabkan oleh corak perubahan magnitud dan polariti AC yang telah mapan. Kajian ini meneliti aktiviti PD di bawah tekanan DC dengan gangguan voltan pada saiz kekosongan yang berbeza menggunakan analisis elemen terhingga, dengan gangguan harmonik AC sebagai tumpuan utama. Ciri-ciri yang dinilai termasuk taburan potensi elektrik, taburan medan elektrik, cas elektrik, dan kadar pengulangan. Apabila jejari kekosongan meningkat dari 1 mm kepada 4 mm, voltan permulaan menurun sebanyak kira-kira 65%, manakala kadar pengulangan PD meningkat sebanyak 58%. Selain itu, perbandingan antara PD-DC dengan gangguan harmonik AC dan PD-AC turut dilakukan bagi menganalisis perbezaannya. Hasil simulasi menunjukkan bahawa kadar pengulangan bagi PD-DC dengan gangguan harmonik AC adalah kira-kira 75% lebih rendah berbanding PD-AC, menandakan bilangan kejadian PD yang lebih sedikit dalam kes PD-DC dengan gangguan harmonik. Simulasi yang dijalankan ini memberikan pandangan yang bernilai dan

panduan berguna bagi pembentukan ujian PD-DC, dan seterusnya, kaedah yang disahkan boleh diluluskan untuk memantau keadaan bahan penebat di bawah kuasa DC.

KEYWORDS: *Partial discharge, DC, AC harmonics, void, FEA*

1. INTRODUCTION

In electrical power systems, partial discharge (PD) is the first indication that the insulation properties of an electrical apparatus are deteriorating. The failure of the insulation system can eventually cause the breakdown of the apparatus, jeopardizing the reliability of the transmission and distribution system. PD is an electrical breakdown that occurs in insulating materials caused by operating stresses [1] such as an overhigh electric field [2]. Under normal circumstances, the insulators contain defects such as voids or air-filled gaps that promote the activity of PD. However, PDs do not cause an abrupt breakdown of the insulation property. When subjected to long and harsh operations, insulating properties start to degrade, causing an increase in PD activity, which eventually could lead to total breakdown [3], [4].

PD detection and measurement are known as a standard method to evaluate the quality of the insulation system [5]; where the knowledge of the PDs' characteristics is crucial for their detection and measurement. Various techniques have been developed for PD measurement, analysis, and recognition, particularly for partial discharge under AC stress (AC-PD). AC voltage operation in terms of its polarity and direction changes is naturally made suitable for monitoring PDs. The pure AC output from transformer conversion allows simplicity in measuring the discharge value without any disturbances.

The critical parameters for measurement include partial discharge inception voltage (PDIV) [6] apparent charge, repetition rate, and the Phase-Resolved Partial Discharge (PRPD) pattern [7]. These parameters are highly influenced by the types of insulating materials, voltage stress, temperature, humidity, impurities, and defects (usually referred to as voids) geometry [4], [6], [8]. The knowledge of the fundamental characteristics of AC-PD, combined with information about the parameters, allows for monitoring the insulation properties of the power system apparatus. Current research is also integrating PD simulation modelling and machine learning [9], [10] in facilitating this method of insulation monitoring.

On the other hand, recent developments in high voltage transmission and distribution systems show an increasing interest in high voltage direct current (HVDC) systems. Similar to the traditional HVAC systems, the insulation systems for HVDC also face the same partial discharge (DC-PD) problem. Unlike AC-PD, which is already well characterized, PRPD is well established as a method of analysis; the fundamental characteristics of DC-PD are still not well known. DC power behaves differently from AC power in terms of its electrical characteristics, such as the magnitude, type, and phase. These differences in characteristics are enough to influence the PD behaviours between AC and DC systems. However, the testing and monitoring procedures of DC-PD are still unclear because the study is still being conducted, and researchers are studying and developing new methods for a definitive solution.

This paper aims to contribute to this study by simulating the PD activity under DC stress with voltage disturbance at different void sizes. In relation to DC power, the output transmission voltage in an HVDC system contains voltage disturbances such as traces of AC harmonic voltage or DC voltage ripple. Thus, it is essential to investigate the effect of voltage disturbance on the characteristics of DC-PD. Here, finite element method (FEM) analysis will be used to analyze the DC-PD characteristic for different void size in cross-linked polyethylene (XLPE) material under DC stress superimposed with AC harmonics.

https://doi.org/10.31436/iiumej.v26i2.3419

2. METHODOLOGY

This chapter consists of several sections. Section 2.1 covers the model's geometry and the simulation parameters. Section 2.2 shows the governing equation of PD modelling using the FEA method. Fig. 1 shows a summary of the methodology.

2.1. Model Configuration

The model for this simulation uses a 2D axisymmetric dimension, allowing for only half of the geometry to be drawn, with the remainder completed through 3D revolution or 2D mirroring. Fig. 2 shows the model following the CIGRE Method-II (CM-II) Type Metal-Air Gap-Metal Electrode system [11]; comprises an insulator material sandwiched by two electrodes, and a cylindrical-shaped air gap is created in the middle of the insulator. The electrodes' pole is 3mm in radius and 20mm in height, while the electrode plates are 20mm in radius and 2mm in height. The insulator material is 25mm in radius and 2mm in height. The void radius varies as a parameter, taking 1 mm, 2 mm, 3 mm, and 4 mm values.

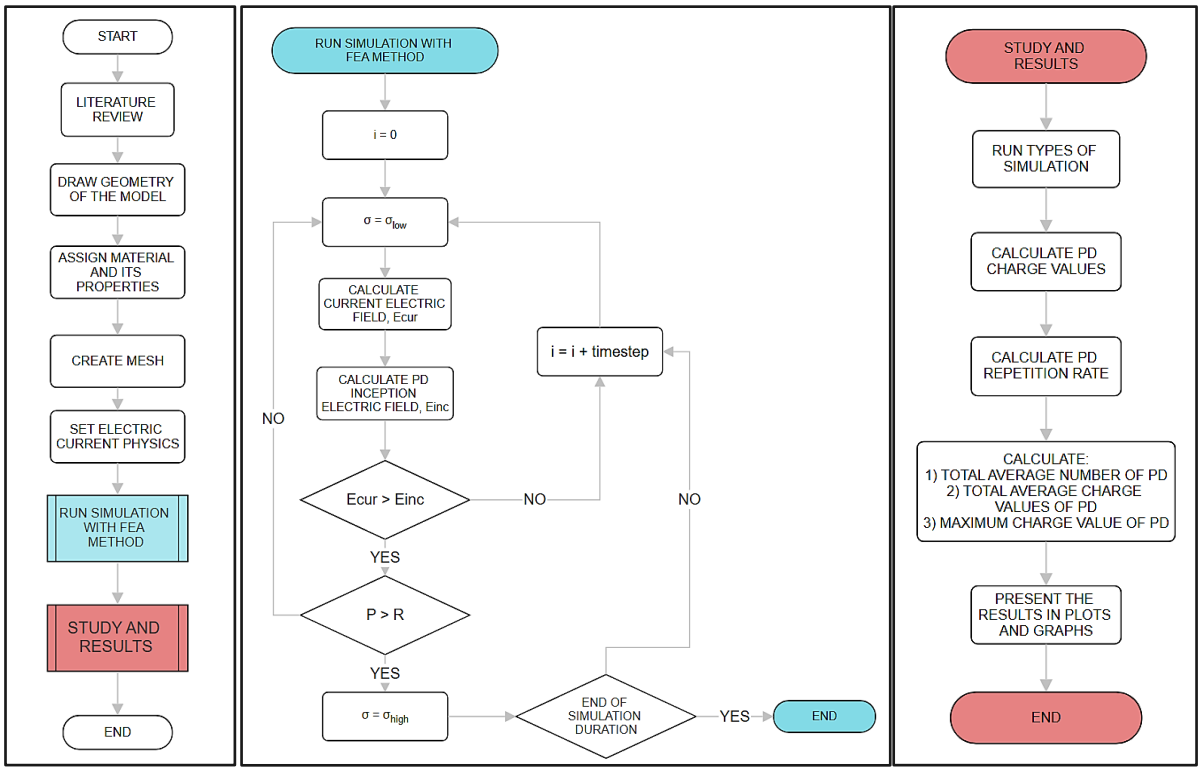


Figure 1. Flowchart of the study.

https://doi.org/10.31436/iiumej.v26i2.3419

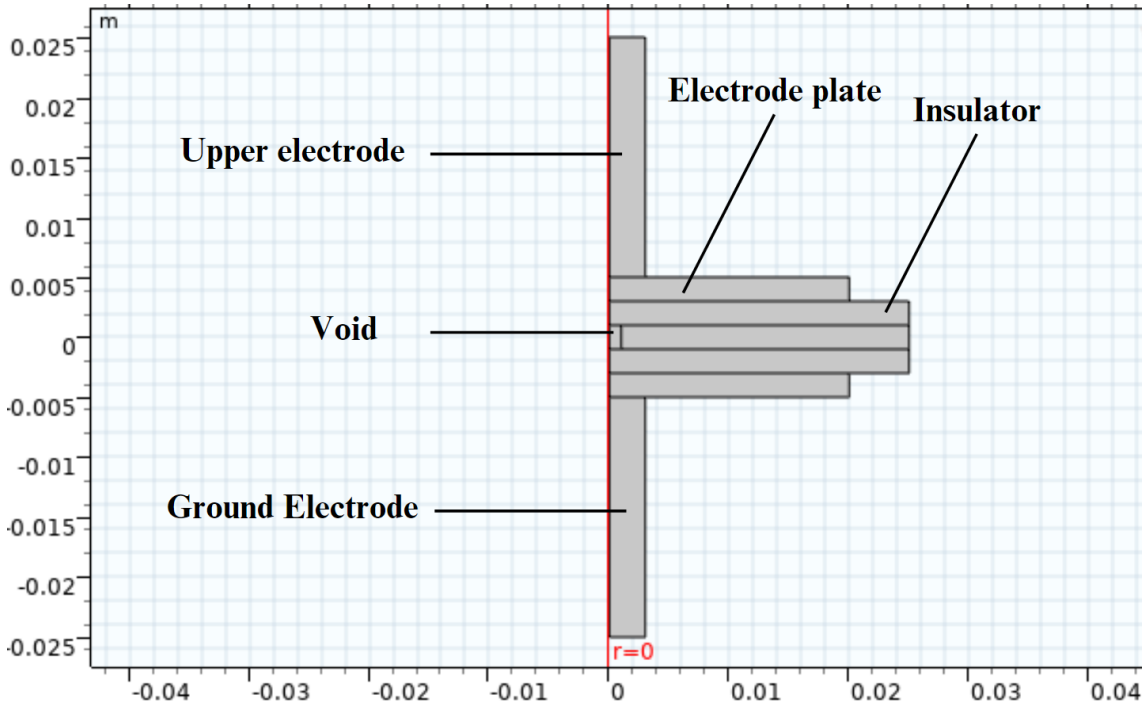


Figure 2. Geometry of the model

Each part of the geometry is assigned materials used experimentally to simulate the partial discharge (PD) model. The electrodes are made of copper metal, suitable for electrical conductivity. The material of the insulator cable is cross-linked polyethylene (XLPE) since this material is often utilized in HVDC systems where the cross-linking by-products, impurities, voids, and gaps can easily cause space charge accumulation under HVDC stress, consequently leading to PD. On the other hand, the material of the void is air [12], with significantly lower dielectric strength than XLPE, a standard material used for PD simulation. Properties of the materials are configured to fulfill simulation requirements, namely the relative permittivity and electrical conductivity, summarized in Table 1. The relative permittivity is the ratio between the absolute permittivity of the insulator and the absolute permittivity of vacuum. At the same time, the electrical conductivity determines the material's ability to conduct electric current. Next, the mesh of the model is created to establish the numerical computation process. Triangular meshing is used for this model, and the mesh element size is set to fine to allow accurate measurement, but this reduces computational resources. Fig. 3 represents the built mesh on the model.

Table 1. Material properties

Materials	Relative permittivity	Electrical conductivity [S/m]
Copper	1	5.998×10^{7}
XLPE	2.3	1×10^{-13}
Air	1	1×10^{-15} (Initial)

https://doi.org/10.31436/iiumej.v26i2.3419

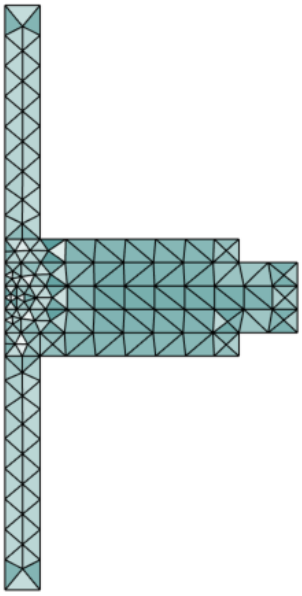
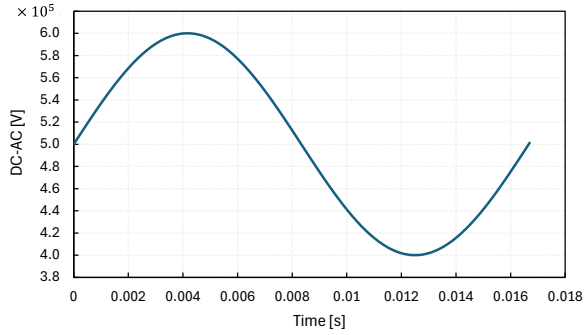
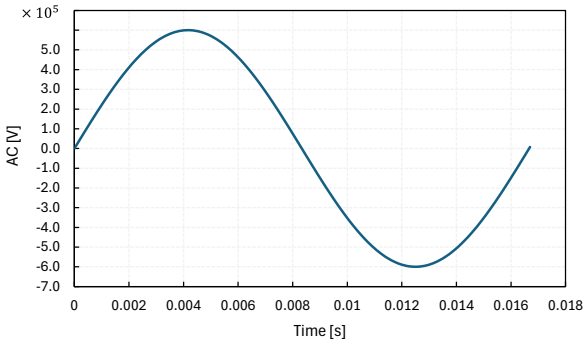


Figure 3. Mesh of the model

Two types of input voltage are defined for the simulations: a) DC input voltage superimposed with AC harmonics and b) AC input voltage. The DC + AC harmonics voltage contains 500kV (DC) + 100kV (AC), while the AC input voltage has a magnitude of 600kV, as shown in Fig. 4.



(a). The 500kV (DC) + 60Hz 100kV (AC Harmonic)



(b). The 60 Hz 600kV (AC)

Figure 4. Input voltages

2.2. Governing Equations

First, the upper electrode is determined as the terminal to receive the input voltage, while the ground electrode becomes the ground terminal. The assumptions for this analysis are as follows:

a. The geometry, material properties, and boundary conditions are symmetric about a central axis.

b. The current continuity equation is satisfied throughout the entire domain.

c. The electrical insulation is perfect, where no current flows through the insulated boundaries or regions.

The electric potential and electric field distribution are calculated based on the governing equations as follows [13]:

https://doi.org/10.31436/iiumej.v26i2.3419

$$\nabla . \mathrm{D} = \rho \tag{1}$$

$$\sigma \mathrm{E} = J \tag{2}$$

$$\nabla . J = \frac{\partial \rho}{\partial t} \tag{3}$$

$$\mathrm{D} = \varepsilon E = -\varepsilon \nabla \mathrm{U} \tag{4}$$

$$\mathrm{J} = \sigma \mathrm{E} = -\sigma \nabla \mathrm{U} \tag{5}$$

$$-\nabla(\sigma \nabla V + \varepsilon \nabla \left(\frac{\mathrm{dv}}{\mathrm{dt}}\right) = 0 \tag{6}$$

Eq. (1) is Gauss' Law, which states that the total electric flux is equal to the total electric charge across a closed surface. D is the electric flux density, ρ is the volume charge density and ∇ is the gradient operator. Eq. (2) is from the derivation of Ohm's Law, and it takes into account the electric field, E, current density, J, and conductivity, σ. The next three are further derivations of the first two, and these equations are translated into Eq. (6).

The first FEA condition of PD occurrence is when the current electric field value, E_{cur}, inside the void is higher than the minimum inception electric field value, E_{inc} [14], [15]. This is known as the streamer inception criterion, where the electrons gain energy, perform further ionization in molecules and atoms, and cause an avalanche of excited electrons. The formula for E_{inc} is represented as follows:

$$E_{inc} = (E_1/\mathrm{p})_{cr}\, \mathrm{p}\left(1 + \frac{B}{(2\mathrm{pa})^n}\right) \tag{7}$$

where p is the air pressure inside the void, a is the void radius and $(E_1/\mathrm{p})_{cr}$, B and n is parameters that are involved in gas ionizations. The constant values for these parameters are $(E_1/\mathrm{p})_{cr}$ = 24.2 $\mathrm{VPa^{-1}m^{-1}}$, B = 8.6 $\mathrm{Pa^{1/2}m^{1/2}}$ and n = 0.5. After a period of discharge occurs, the electric field inside the void starts to drop because there is a difference of strength between the electric field left by the charged particles and the supplied voltage. When the value falls below an extinction value (E_{ext}), the streamer discharge cannot be fulfilled. The formula for E_{ext} is represented as follows:

$$E_{ext} = \mathrm{p}\gamma(\mathrm{E}/\mathrm{p})_{cr} \tag{8}$$

where γ depends on the voltage polarity and is usually obtained from experiments [13].

The second condition is the electron generation rate and the probability of electron availability. Eq. (9) is used to assess the availability of electrons for PD to occur, where N_{es} is a constant, U_{cav} is the voltage across void, U_{inc} is the PD inception voltage across the void, and Δt is the time step of the simulation. Then, in Eq. (10), P is calculated to determine the probability of generating free electrons at the time step of the simulation [16]. Subsequently, P is compared with R, where R is a random number between 0 and 1 (0 < R < 1).

At $E_{cur} > E_{inc}$ and P > R, PD occurrences are dynamically simulated by switching the void's state from non-conducting to conducting (i.e., increasing the conductivity σ_{cav} from σ_{low} to σ_{max}), as shown in Eq. (11). This causes the field inside the void to increase again, and the discharges are repeated. Table 2 summarizes the constant parameters implemented in this simulation.

$$\dot{N}_{tot}(\mathrm{t}) = N_{es} \exp(|U_{cav}/\ U_{inc}|) \tag{9}$$

$$\mathrm{P} = \dot{N}_{tot}(\mathrm{t})\Delta t \tag{10}$$

$$\sigma_{cav} = \begin{cases} \sigma_{max} & E_{CUR} \geq E_{INC}\ \&\ P > R \\ \sigma_{low} & E_{CUR} \leq E_{EXT} \end{cases} \tag{11}$$

https://doi.org/10.31436/iiumej.v26i2.3419

Table 2. Constant parameters of the simulation [13], [16], [17]

Symbols	Parameters	Values
F	Frequency of the input voltages	60 Hz
T	Period of the input voltages	0.0167 s
σ_{max}	Max conductivity of the void	$5x10^{-4}$ S/m
σ_{low}	Min conductivity of the void	$1x10^{-15}$ S/m
P	Void pressure	101325 Pa
γ	Streamer propagation factor	0.35
$(E/p)_{cr}$	Gas ionization constant	24.2 V. $Pa^{-1}.m^{-1}$
B	Gas ionization parameter, air	8.6 $Pa^{1/2}m^{1/2}$
n	Gas ionization constant, air	0.5
N_{es}	Initial electron generation rate	650 1/s
Δt	Simulation time step	-

PD charge magnitudes and repetition rate are standard parameters used to evaluate PD, and they can be calculated using Eq. (12) and Eq. (13), respectively. It is important to note that the electric field values are taken from the simulation, where only the electric field inception and extinction are considered. Meanwhile, the charge magnitudes and repetition rate are considered in the first condition, and the electron probability is calculated through the electron generation rate formula.

$$q_{real} = \int_{tinc}^{text} \int_{Scav,surf} \vec{J}(t).d\vec{S}dt \tag{12}$$

$$PD\ repetition\ rate = \begin{cases} 1 & \begin{matrix} E_{CUR} \geq E_{INC}\ \&\ P>R \\ E_{CUR} \leq E_{EXT} \end{matrix} \\ 0 \end{cases} \tag{13}$$

In Eq. (12), the current density values $\vec{J}(t)$ is integrated over the $S_{cav,surf}$ which is the void surface, obtaining the current within the void. Then, the values are integrated again during the interval of PD inception and extinction times. When PD does not occur, the time intervals are flipped and integrated oppositely. The value is recorded at each time step of the simulation. In Eq. (13), the PD repetition rate is counted when the FEA conditions are fulfilled. Only one count is taken for each PD event, which is increased for the next PD event.

3. RESULT AND DISCUSSION

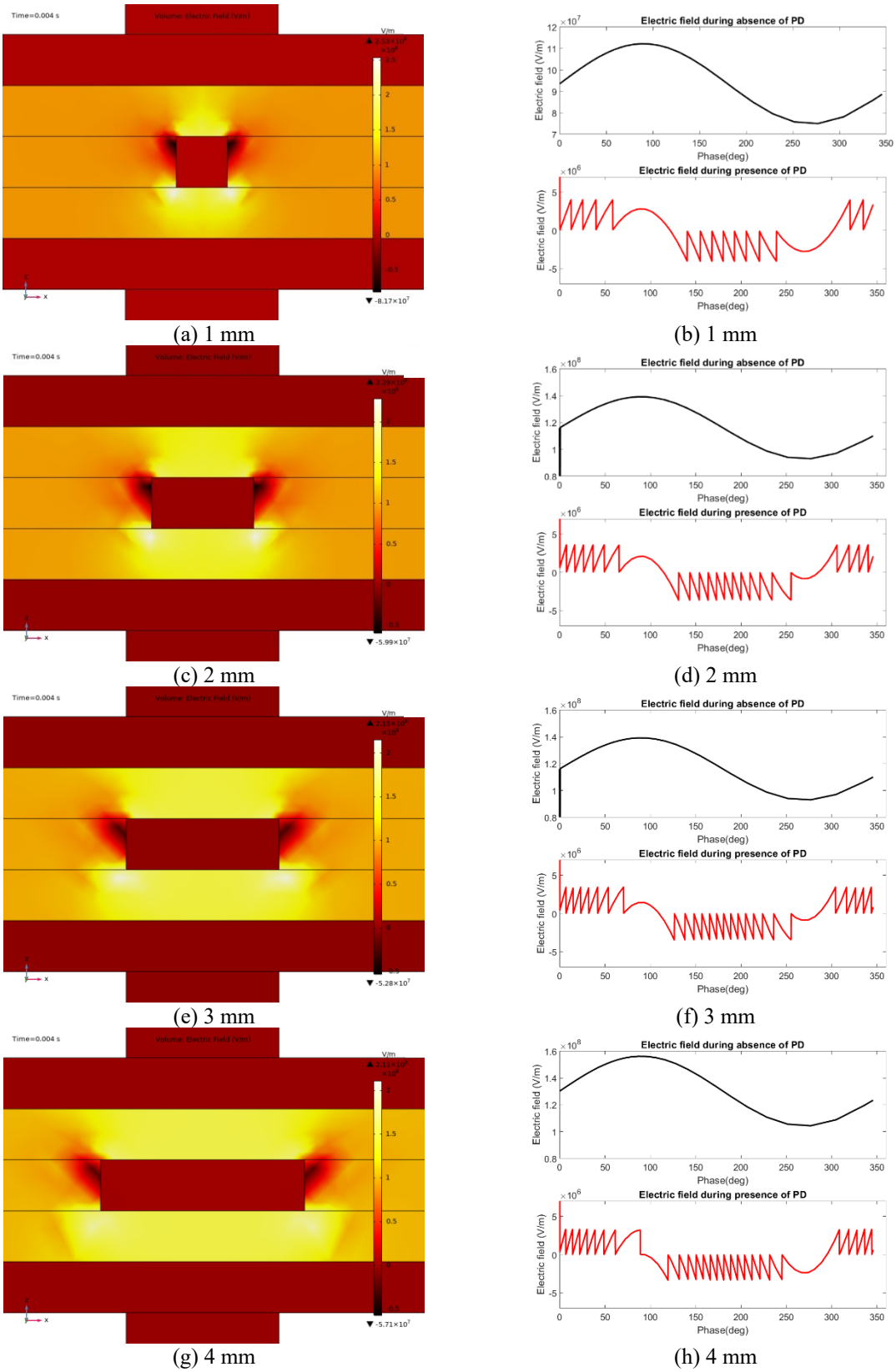
This section covers the results of the DC-PD simulation with AC harmonic disturbance at different void sizes. Section 3.1 shows the electric field distribution of the model and line graphs during the absence and presence of PD. Section 3.2 shows the PD charge magnitudes in a scatter graph and repetition rate in a 2D bar graph. Section 3.3 compares the DC-PD simulation with the AC harmonic disturbance and the pure AC-PD simulation.

3.1. Electric Field Distribution at Different Void Sizes

The influence of AC harmonics disturbance on DC-PD is determined by measuring the electric field. The occurrence of PD is heavily dependent on the electric field values. The plots and values are recorded after simulating one cycle. Fig. 5(a), (b), (c), and (d) indicate that the electric field distribution across all void sizes is uniform, and the voids have lower values than the input voltage. Fig. 5(b), (d), (f), and (h) show the changing electric field values throughout the cycle period. When the electric field PD inception condition is fulfilled, the conductivity changes to a high value. After that, the electric field value drops instantly, fulfilling the PD extinction condition. The conductivity then changes back to a low value, and the electric field value is in accordance with the electric field during the absence of PD. Fig. 5(h) has the highest changing conductivity values because it has the lowest E_{inc} and PD inception voltage, U_{inc}

https://doi.org/10.31436/iiumej.v26i2.3419

value. Thus, the PD inception condition is fulfilled faster in the 4mm void radius simulation than the lower void radius simulations. Table 3 shows the E_{inc} and U_{inc} values for each void size simulation.



(a) 1 mm (b) 1 mm

(c) 2 mm (d) 2 mm

(e) 3 mm (f) 3 mm

(g) 4 mm (h) 4 mm

Figure 5. Electric field surface distribution and the electric field for different void sizes

https://doi.org/10.31436/iiumej.v26i2.3419

Table 3. The PD electric field and amplitude inception values

Void radius [mm]	E_{inc} [V/m]	U_{inc} [V]
1	3.93×10^{6}	8499.6
2	3.5×10^{6}	3519.2
3	3.31×10^{6}	3130.6
4	3.19×10^{6}	2984.5

3.2. Evaluation of Electrical Charges and Repetition Rate at Different Void Sizes

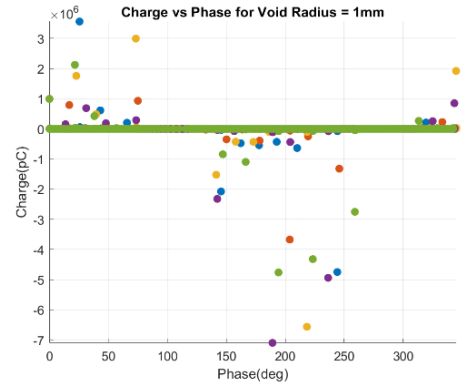
The charge magnitude and repetition rate across the void under DC-PD + AC harmonics are calculated for the different void sizes. The value and number of these charges indicate the occurrence of PDs when the conditions for PD are fulfilled. The PD inception and electron generation rate are considered when evaluating the charges. When these conditions are fulfilled, the electrons gain enough energy to move quickly due to the increased electric field strength. This increases the electrical current and, in turn, increases the electrical charges inside the void. The plots and values are recorded after simulating five cycles.

Fig. 6 shows the charge magnitudes and the repetition rate at different void sizes. Fig. 6(e) and Fig. 6(g) show the highest number of PD charges at a high range of charge values. Considering the electron generation rate condition, the increase in void sizes allows the charges to reach very high values. The 3mm and 4mm void size for the PD repetition rate shows the highest number of PDs. Table 4 shows the number of PDs, total charge values, and maximum charge values for each void radius simulation per cycle. The results show that all the parameters increase in value when the void radius increases. As the radius of the void increases, the effective discharge area and the number of charged particles in the discharge channel grow, leading to a higher discharge quantity and increased discharge repetition rate.

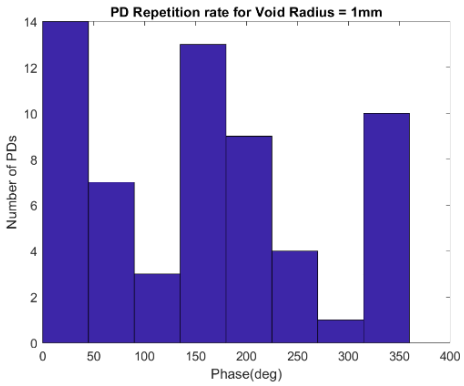
Table 4. Evaluation of PD parameters

Void radius [mm]	Average PD repetition rate per cycle	Average PD charges value per cycle [pC]	Maximum PD charge [pC]
1	13.2	1.3416×10^{3}	5.763×10^{6}
2	16.8	2.7851×10^{3}	4.9488×10^{6}
3	18.2	3.93×10^{3}	1.1040×10^{7}
4	20.8	6.0024×10^{3}	1.13862×10^{7}

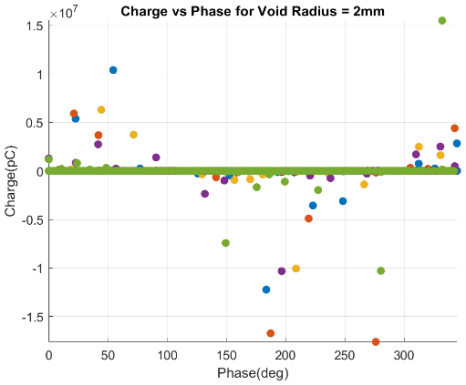
https://doi.org/10.31436/iiumej.v26i2.3419



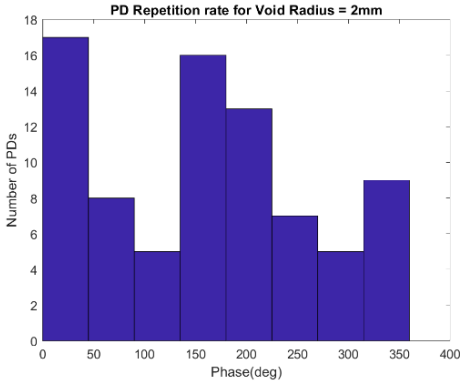
(a) 1 mm



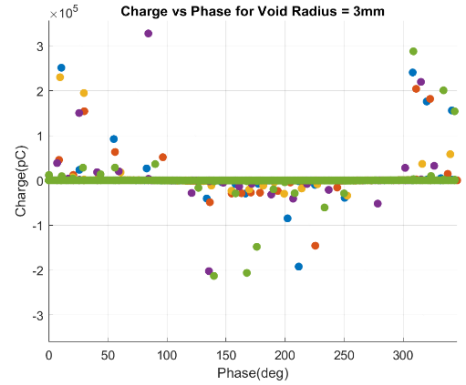
(b) 1 mm



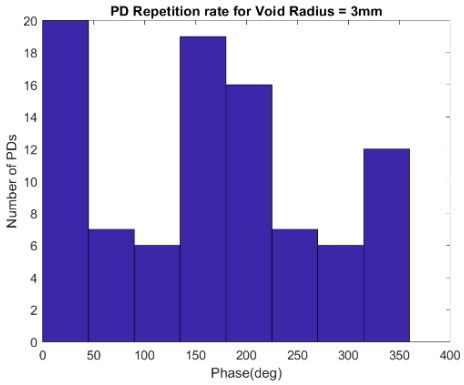
(c) 2 mm



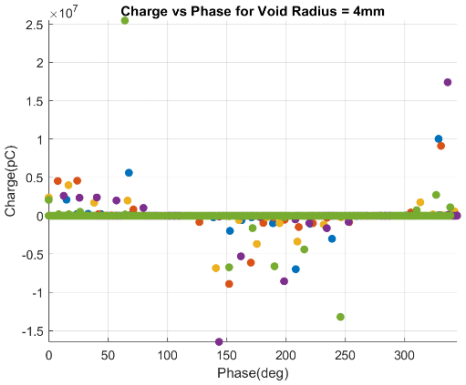
(d) 2 mm



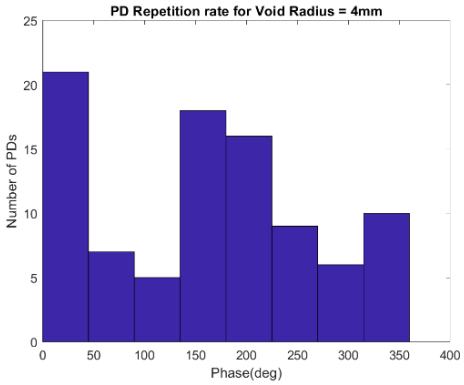
(e) 3 mm



(f) 3 mm



(g) 4 mm



(h) 4 mm

Figure 6. Charge vs phase and PD repetition rate for different void sizes

https://doi.org/10.31436/iiumej.v26i2.3419

3.3. Comparison of Electrical Charges and Repetition Rate between AC-PD and DC-PD + AC harmonics

After evaluating the simulation characteristics of DC-PD with AC harmonics, the PD behavior between this simulation and the AC-PD simulation is compared to see the differences between these configurations. The void radius is kept constant at 4mm for these comparisons. Fig. 7 displays the PD charges values and repetition rate graph between AC-PD and DC-PD + AC harmonics.

Comparing Fig. 7(a) and Fig. 7(c), the simulations of AC-PD contain a higher range of charge compared to DC-PD + AC harmonics. The comparison of PD repetition rate in Fig. 7(b) and Fig. 7(d) is also significant, where the number of PDs can reach up to 70 counts in AC-PD, while for DC-PD + AC harmonics, it can only reach up to 19 counts. AC-PD possesses a significantly higher average PD repetition rate per cycle than the other simulations. Table 5 shows the recorded PD parameters for the two simulations. This observation might be related to the difference in the input voltages shown in Fig. 5. It can be seen from the figure that amplitudes are different despite the almost similar peak value, indicating a difference in the magnitude of the voltage. AC's input voltage is higher than the DC+AC harmonics, suggesting a higher electric field in the case of AC-PD. However, the average and maximum PD charges in the DC-PD+AC harmonics simulation are not that different from those in the AC-PD simulation. Even though the number of PDs is lower in DC-PD compared to AC-PD, the AC harmonic component of the combined voltage can cause higher PD charge values due to the increase in electric field when the AC values change alternately.

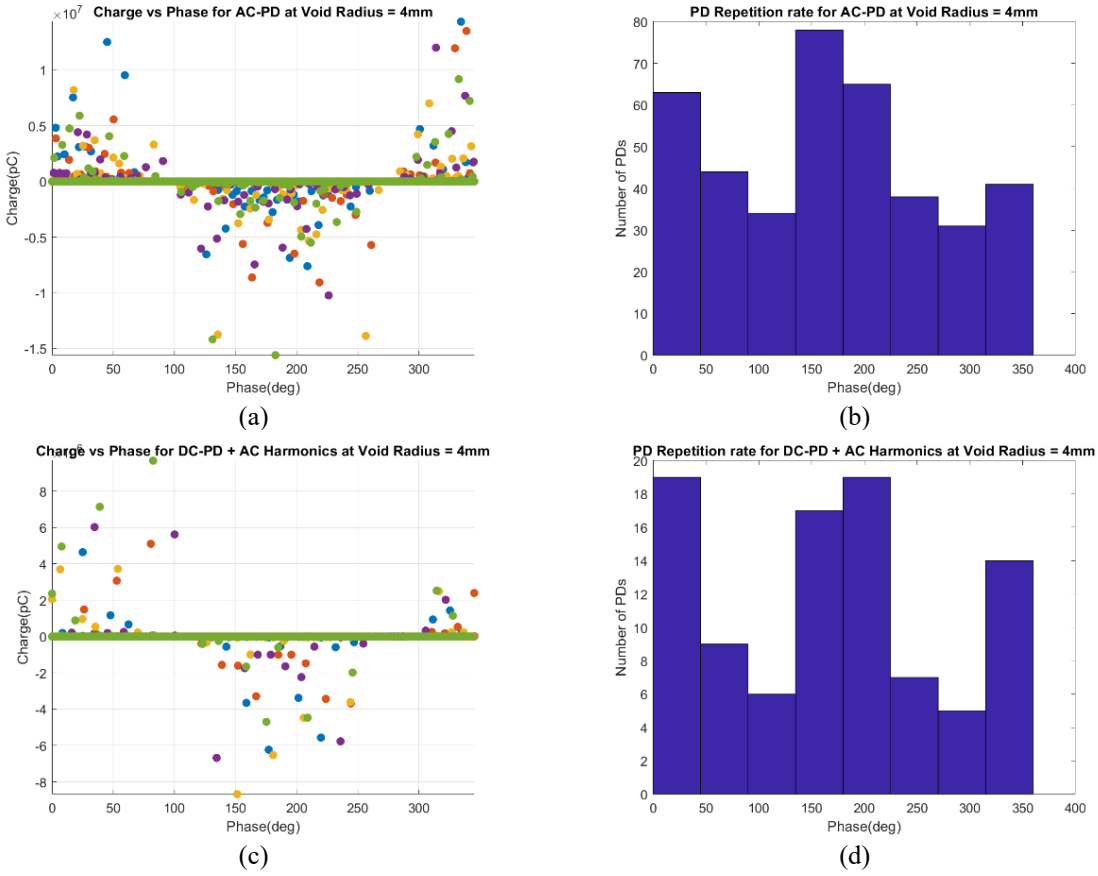


Figure 7. The charge and PD repetition rate at 4mm radius void: (a)(b) AC-PD and (c)(d) DC-PD + AC harmonics

Table 5. Results evaluation of PD parameters between AC-PD and DC-PD with AC harmonics

Simulation type	Average PD repetition rate per cycle	Average PD charges value per cycle [pC]	Maximum PD charge [pC]
AC-PD	78.8	3.1083×10^{6}	1.4345×10^{7}
DC-PD + AC harmonics	19.2	1.8881×10^{6}	9.6837×10^{6}

4. CONCLUSION

PDs are temporary electrical connections that occur between the surface of the insulator and the conductor in high electric fields and voltage systems. Monitoring the condition of the insulator is vital to recognizing the defects. Compared to AC-PD testing, monitoring PD under DC power is not well-developed. Thus, research and study are conducted to unveil new methodologies to improve the monitoring of DC-PD. For simulation, FEA modelling is discovered as one of the methods to model the PD behavior using conditions such as electric field inception, electron generation rate, and the availability of electrons.

The simulation of DC-PD with AC harmonic disturbance has been conducted under different void sizes. The electric field at the void origin is plotted, and the results show that the values change in accordance with the electric field PD inception and extinction conditions at each void size. The inception voltage decreases by approximately 65% with the increase in the radius from 1mm to 4mm. After that, the evaluation of PD charges and repetition rate is recorded. As the void size increases, the charge values and number of PDs also increase, with the maximum charge PD charge increasing by 95% and the average PD repetition rate per cycle increasing by 58%.

A comparison of AC-PD and DC-PD with AC harmonic is performed to view the performance of the testing PDs in these two simulations. It is concluded that AC-PD shows higher values compared to the DC-PD + AC harmonic simulation, where the repetition rate for DC-PD with AC harmonics is approximately 75% lower than that of AC-PD, indicating fewer PD events in the case of DC-PD with AC harmonics.

The simulation of DC-PD with AC harmonics across all simulations provides promising results for observing PDs. At the very least, PDs appear in DC-PD + AC harmonic simulation because the number of PD occurrences is very low in pure DC-PD testing. To conclude, the simulation is successful as the occurrence of PDs is obtained through the combination of DC voltage + AC harmonic disturbance. The simulation provides valuable insights and guidance for developing DC-PD testing and monitoring techniques for assessing the condition of insulating materials. Further research under varying environmental conditions, such as temperature fluctuations and thermal stress, can enhance the application of these methods. Additionally, the DC-PD behavior of various materials used in HVDC facilities, including emerging insulator materials like High-Performance Thermoplastic Elastomer (HPTE), should be further explored. This will help further expand the applicability and understanding of DC-PD behavior in insulating materials.

ACKNOWLEDGEMENT

This is supported by the Ministry of Higher Education (MoHE) Malaysia through the Research Management Centre, International Islamic University Malaysia (IIUM) under the Fundamental Research Grant Scheme FRGS (Ref: FRGS/1/2023/TK07/UIAM/02/5– Project ID: FRGS23-320-0929)

REFERENCES

[1] Seri P *et al.* (2019) Partial discharge inception voltage in DC insulation systems: A comparison with AC voltage supply. 2019 IEEE Electrical Insulation Conference, EIC 2019. 176–179. https://doi: 10.1109/EIC43217.2019.9046630

[2] Zhang X *et al*. (2021) Review on detection and analysis of partial discharge along power cables. https://doi: 10.3390/en14227692

[3] Borghei M, Ghassemi M, Kordi B, Gill P, Oliver D. (2021) A finite element analysis model for internal partial discharges in an air-filled, cylindrical cavity inside solid dielectric. 2021 Electrical Insulation Conference, EIC 2021. 260–263. https://doi: 10.1109/EIC49891.2021.9612268

[4] Du Y, Li Y. (2022) Discharge characteristics of epoxy resin partial discharge under different harmonic voltages. 2022 9th International Forum on Electrical Engineering and Automation, IFEEA 2022. 728–731. https://ddoi: 10.1109/IFEEA57288.2022.10037812

[5] Hussain GA, Hassan W, Mahmood F, Shafiq M, Rehman H, Kay JA. (2023) Review on partial discharge diagnostic techniques for high voltage equipment in power systems. IEEE Access. 11. 51382–51394. https://doi: 10.1109/ACCESS.2023.3279355

[6] Sahlen F, Takala M. (2023) Factors influencing the partial discharge inception voltage in Type I insulation systems. 2023 IEEE Electrical Insulation Conference, EIC 2023. 1-4. https://doi: 10.1109/EIC55835.2023.10177313

[7] Mohamed Asseri MA, Hezri Fazalul Rahiman M, Halim Abdul Majid MH. (2022) Partial discharge monitoring and analysis using phase resolve partial discharge pattern. 2022 IEEE 10th Conference on Systems, Process and Control, ICSPC 2022 – Proceedings. 231–235. https://doi: 10.1109/ICSPC55597.2022.10001739

[8] Montanari GC, Shafiq M, Chen Z. (2024) A discussion on the dependence of partial discharge inception voltage on supply voltage waveform: sinusoidal and modulated AC. 2024 IEEE Electrical Insulation Conference, EIC 2024. 392–396. https://doi: 10.1109/EIC58847.2024.10579408

[9] Ding B *et al.*(2021) Pattern recognition of partial discharge based on deep learning. The 16th IET International Conference on AC and DC Power Transmission (ACDC 2020). 1187–1191. https://doi: 10.1049/ICP.2020.0223

[10] Banjare HK, Sahoo R, Karmakar S. (2022) Study and analysis of various partial discharge signals classification using machine learning application. Proceedings of 2022 6th International Conference on Condition Assessment Techniques in Electrical Systems, CATCON 2022. 52–56. https://doi: 10.1109/CATCON56237.2022.10077703

[11] Sakakibara T, Kawashima T, Hozumi N, Murakami Y. (2023) Measurement of spark discharge inception voltage using CIGRE Method-II Type metal-air gap-metal electrode system. Proceedings of the International Symposium on Electrical Insulating Materials. 285–287. https://doi: 10.23919/ISEIM60444.2023.10329272

[12] Lagrotteria G, Pietribiasi D, Marelli M. (2019) HVDC Cables - The technology boost. 2019 AEIT HVDC International Conference, AEIT HVDC 2019. https://doi: 10.1109/AEIT-HVDC.2019.8740645

[13] Musa U *et al.* (2023) FEA-based simulation of accelerated ageing in a power cable due to sustained partial discharge activities in a spherical cavity. Arab J Sci Eng. 48(11). 15029–15043. https://doi: 10.1007/S13369-023-07967-7/TABLES/4

[14] Borghei M, Ghassemi M, Rodriguez-Serna JM, Albarracin-Sanchez R. (2021) A finite element analysis and an improved induced discharge concept for partial discharge modeling. IEEE Transactions on Power Delivery. 36(4). 2570–2581. https://doi: 10.1109/TPWRD.2020.2991589

[15] Rodríguez-Serna JM, Albarracín-Sánchez R, Dong M, Ren M. (2020) Computer simulation of partial discharges in voids inside epoxy resins using three-capacitance and analytical models. Polymers 2020. 12(1). 77. https://doi: 10.3390/POLYM12010077

[16] Haiba AS, Elfaraskoury AA, Elkoshairy AD, Halawa MM. (2022) Modeling and simulation of partial discharge measurement for defected solid dielectrics. Journal of Measurement Science and Applications (JMSA). 2(1). 43–53. https://doi: 10.21608/JMSA.2022.219865

[17] Saghafi M, Ghassemi M, Lehr J, Borghei M, Kordi B, Oliver D. (2022) A finite element analysis model for internal partial discharges under DC voltage. 2022 IEEE International Power Modulator and High Voltage Conference, IPMHVC 2022. 8–11. https://doi: 10.1109/IPMHVC51093.2022.10099411

https://doi.org/10.31436/iiumej.v26i2.3429

DEVELOPMENT OF A MOBILE APPLICATION FOR DYSLEXIC-FRIENDLY LEARNING MATERIALS

DAYANGKU NUR AMELIA ASRI, HUDA ADIBAH MOHD RAMLI *, NORAZLINA SAIDIN

Department of Electrical and Computer Engineering, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia

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

(Received: 3 September 2024; Accepted: 2 May 2025; Published online: 15 May 2025)

ABSTRACT: Dyslexia, a prevalent learning disorder affecting approximately 10% to 15% of primary school children in Malaysia, presents a significant educational hurdle. Traditional interventions often rely on in-person sessions, imposing additional demands on parents. Existing technology-based platforms are partially dyslexic-friendly with limitations such as poor user interfaces, limited content, lack of progression to advanced learning, and attempts to address multiple skills simultaneously. This paper proposes an enhanced dyslexic-friendly mobile application (app) to address these issues. The app employs the Orton-Gillingham approach to ensure a continuous and progressive learning environment. Moreover, it also uses the Structured Literacy approach by dividing learning into three modules that reinforce basic phonology, syllable knowledge, and meaning association. This app is developed using several software components: Visual Studio Code, Android Studio, Flutter, and Adobe XD. Besides, the app includes games that serve as diagnostic tools to monitor improvement, identify weaknesses, and measure progress. By catering to diverse learning styles and needs, this app offers a more accessible and supportive learning experience through an innovative mobile tool that promotes inclusivity, independence, and continuous skill development.

ABSTRAK: Disleksia, gangguan pembelajaran lazim yang menjejaskan kira-kira 10% hingga 15% kanak-kanak sekolah rendah di Malaysia, memberikan cabaran pendidikan yang ketara. Intervensi tradisional sering bergantung pada sesi bersemuka, menyebabkan tuntutan tambahan kepada ibu bapa. Platform berasaskan teknologi sedia ada sebahagiannya mesra disleksia dengan batasan seperti antara muka pengguna yang lemah, kandungan terhad, kekurangan kemajuan kepada pembelajaran lanjutan dan percubaan untuk menangani pelbagai kemahiran secara serentak. Untuk menangani isu ini, aplikasi mudah alih mesra disleksia yang dipertingkatkan dicadangkan dalam kertas kerja ini. Aplikasi ini menggunakan pendekatan Orton-Gillingham untuk memastikan persekitaran pembelajaran yang berterusan dan progresif. Selain itu, ia juga menggunakan pendekatan Literasi Berstruktur dengan membahagikan pembelajaran kepada tiga modul yang mengukuhkan fonologi asas, pengetahuan suku kata, dan perkaitan makna. Aplikasi ini dibangunkan menggunakan beberapa komponen perisian yang merangkumi Kod Visual Studio, Android Studio, Flutter dan Adobe XD. Selain itu, aplikasi tersebut menggunakan permainan yang berfungsi sebagai alat diagnostik untuk memantau peningkatan, mengenal pasti kelemahan dan mengukur kemajuan. Dengan memenuhi gaya dan keperluan pembelajaran yang pelbagai, aplikasi ini menawarkan pengalaman pembelajaran yang lebih mudah diakses dan menyokong melalui alat mudah alih yang inovatif yang menggalakkan keterangkuman, kebebasan dan pembangunan kemahiran berterusan.

KEYWORDS: *dyslexia, mobile application, learning materials, Orton-Gillingham, Structured Literacy*

https://doi.org/10.31436/iiumej.v26i2.3429

1. INTRODUCTION

Learning disabilities have become a significant educational and developmental challenge, affecting a substantial portion of the population. A comprehensive report by the Department of Social Welfare revealed an astonishing 236,972 registered cases of individuals grappling with these disabilities [1]. Additionally, a 2023 report from the Ministry of Education discloses that there are up to 97,494 cases of learning disorders in Malaysian schools, with 17,271 (18%) identified as dyslexic [2]. Dyslexia, classified as a neurodevelopmental disorder, affects an individual's ability to grasp basic literacy-related skills, such as spelling and reading [3]. This distinctive impact makes dyslexia easily identifiable, especially in children as they begin their formal education. Individuals displaying symptoms of dyslexia frequently exhibit significant delays compared to their peers, particularly in reading. Regrettably, if left untreated, this can result in diminished self-esteem and reduced confidence [4].

Table 1. Summary of Existing Mobile Applications

Mobile application	**Features**	**Limitations**
Dyslexia Baca [5]	• A Malay-language app for early dyslexic readers. • Introduces confusing letters, supports alphabet recognition with multisensory methods, and features rewards for achievements.	Needs better interfaces, offers overly simple levels, focuses narrowly, and lacks continuous learning progression.
Easylexia [6]	A level-based number quiz game with a user-friendly interface and structured learning.	Overemphasis on multiple skills may confuse.
Alexza [7]	A user-friendly app helping dyslexic users overcome reading difficulties in real-life situations.	• Not focused on learning. • Ideal for those with basic reading skills.
Funologo [8]	An Android-based virtual reality app that enhances phonological awareness in dyslexic children using the Orton-Gillingham approach.	Unspecified information on how the learning modules work.
Dys-I-Can [9]	• An app with modules for dyslexic children aged 3 to 12. • Offers a relaxing environment and rewards throughout their learning journey.	Can confuse module transitions
The Hope [10]	• A stage-by-stage app that monitors dyslexic individuals' progress. • Uses the Orton-Gillingham approach to enhance learning benefits.	Requires extra equipment for handwriting recognition assessment and parental guidance.
The Cure [11]	An alternative for individuals with dyslexia and dysgraphia is to practice reading and writing independently with a user-friendly app and diverse practice words.	• No variation in level progression. • Not very engaging for children.
Arunalu [12]	A multisensory mobile app in Sinhala with interactive interfaces and a reward system.	• Only in Sinhala. • No clear distinction between screening and intervention stages.

Given the significant number of dyslexic individuals and the challenges they face, specific educational and support interventions such as treatment sessions and special classes within schools are essential. While these intervention strategies have been employed for a more effective learning environment for dyslexic individuals, these mainly rely on in-person sessions, imposing additional strain on parents already managing demanding schedules. Consequently, there is an urgent need for more accessible and dyslexia-friendly educational

https://doi.org/10.31436/iiumej.v26i2.3429

interventions. Though numerous technology-based learning platforms, such as mobile applications (apps) for dyslexia, have emerged, many of these apps struggle with issues like poor user interfaces and content limitations. Response often features distracting colors and unclear designs. Likewise, many apps focus solely on early learning phases, such as alphabet recognition, without a smooth transition to more advanced learning. Moreover, attempts to address multiple skills simultaneously may lead to confusion and burnout, not fulfilling the need for balanced content complexity and user engagement. These limitations are summarized in Table 1.

Considering these limitations, this paper presents the development of a mobile application tailored to the specific needs of dyslexic individuals. Designed for smartphones and tablets, the app offers support anytime, anywhere to promote continuous learning. It features a user-friendly interface to enhance clarity, a structured progression from foundational to more advanced and complex materials, and targeted learning modules to prevent cognitive overload. The app aims to provide accessible, personalized tools for effective education, incorporating interactive features and feedback to boost engagement, support self-directed learning, and foster the development of essential educational skills.

The structure of the following sections is as follows: Section 2 offers an overview of dyslexia, Section 3 outlines the methodology of this study, Section 4 presents and analyzes the results, and Section 5 summarizes the paper.

2. OVERVIEW OF DYSLEXIA

The Special Education Division within the Ministry of Education Malaysia characterizes dyslexic individuals as those with intellectual abilities equal to or surpassing those of their peers of the same age, but who confront substantial challenges in spelling, reading, and writing [3]. Therefore, dyslexia is generally associated with a reading disorder or reading disability. Common characteristics presented by dyslexic individuals include deficits in phonological processing and letter reversal. Phonological processing, essential for understanding written and spoken language, includes phonological awareness [13], memory, and access to lexical storage components. Dyslexic individuals repeatedly struggle with these components, impacting their reading fluency and comprehension and making it harder for them to decode and understand text. Letter reversal is another common deficit in dyslexic children. They have difficulty recognizing between similarly shaped letters [14]. For example, they may write /b/ as /d/, /p/ as /q/, or /m/ as /w/. Furthermore, they might reverse the sequence of syllables in words, such as writing “load” as “laod”. Considering the challenges dyslexic individuals face, it is necessary to adopt specific teaching approaches instead of traditional classroom settings. These approaches should be tailored to target specific areas of difficulty to enhance learning outcomes.

This paper studies two well-known teaching approaches, namely Structured Literacy and Orton-Gillingham. These approaches are detailed in the following sub-sections.

2.1. Structured Literacy

Structured Literacy is a teaching approach developed by the International Dyslexia Association (IDA). It offers a comprehensive approach to literacy education by focusing on six key pillars: phonology, phonics, syllable knowledge, morphology, syntax, and semantics. Phonology refers to the rules governing how sounds are pronounced in a specific language. Knowledge of phonology enables individuals to recognize and manipulate letter sounds, identify rhyming words, and segment sounds within words. The sound-symbol correspondence

https://doi.org/10.31436/iiumej.v26i2.3429

(or phonics), involves the skill of associating phonemes (sounds) with graphemes (letters). For example, associating the sound /ch/ with the letter "ch."

Syllable knowledge enhances the ability of dyslexic individuals to analyze the comprehension and pronunciation of words, while morphology focuses on the structure of words, including prefixes, suffixes, and roots. Syntax, conversely, concerns how sentences are structured and how words are arranged to convey meaning. Finally, semantics is the study of the meanings of words, symbols, and word units, intending to help individuals understand written and oral texts. In conclusion, this approach introduces reading in small, sequential steps, beginning with sound recognition and progressing to meaning and word association. This method effectively strengthens the foundations of reading and writing, particularly for individuals with dyslexia, by providing them with essential literacy skills.

2.2. Orton-Gillingham

The Orton-Gillingham is an approach designed for one-on-one instruction [15]. Developed in the early 1930s by Anna Gillingham and Dr. Samuel Torrey Orton, it provides a systematic, sequential, and cumulative learning process, with lessons reviewed repeatedly to reinforce understanding and focus. Instruction begins with simple concepts and progressively advances to more complex material. The approach also incorporates multisensory structured language (MSL), which engages visual, auditory, tactile, and motor skills across listening, speaking, reading, and writing activities. This multisensory technique allows students to associate sounds with what they see, hear, and feel, using their eyes, ears, mouths, and hands. Lessons are tailored based on continuous assessments to meet each student's specific needs. In conclusion, the Orton-Gillingham method offers a comprehensive and individualized approach to support individuals with learning disabilities, helping them better understand and retain educational material. Table 2 presents a comparison of the two teaching approaches.

Table 2. Comparison of Structured Literacy and Orton-Gillingham Approaches

	Structured Literacy	**Orton-Gilligham**
Aim	This approach introduces the fundamentals of Literacy to strengthen students' foundational reading and writing skills.	This approach promotes a systematic, sequential, and cumulative learning environment to help students maintain focus and better understand the learning content.
Characteristics	• Consists of six pillars, namely phonology, sound-symbol correspondence (phonics), syllable knowledge, morphology, syntax, and semantics. • Breaks reading into smaller, sequential steps.	• Applies prerequisite concepts to ensure progressive improvement. • Encourages engagement through MSL. • Tailors teaching plans according to individuals' strengths and weaknesses.

3. METHODOLOGY

This section describes the methodology used in developing the mobile app, encompassing the flowchart, the software components utilized, and the user interface design. The details of each are described next.

3.1. Flowchart

The proposed mobile app incorporates Structured Literacy and Orton-Gillingham approaches to create a dyslexic-friendly learning environment. The flowchart of the proposed

https://doi.org/10.31436/iiumej.v26i2.3429

mobile app is illustrated in Fig. 1. It can be observed in the figure that the app encompasses three learning modules aligned with the Orton-Gillingham approach in ensuring progressive and cumulative learning processes. Each module is based on the pillars of Structured Literacy, where Module 1 teaches basic phonology, Module 2 teaches syllable knowledge, and Module 3 teaches meaning association. The app will be available in Malay to boost Literacy for dyslexic individuals in Malaysia. Primarily targeted at users aged 7 to 9 years, the app helps address early signs of dyslexia, such as difficulty learning. It is also suitable for kindergarten-aged children who can use it to begin their learning journey. Designed to improve literacy skills, the app helps dyslexic users spell and read simple sentences.

The app begins each session by randomly selecting a word when users click "Start." This word will be used across Module 1, Module 2, and Module 3, ensuring a consistent and progressive learning experience. Module 1 introduces users to sounds associated with each letter, such as teaching the sound /ch/ for the letter "k." In Module 2, users progress to combining the letters learned in Module 1 into syllables, understanding their corresponding sounds. For instance, if Module 1 teaches the sounds for "a" and "k," Module 2 combines these to introduce the syllable "ka." Module 3 applies learned syllables to construct complete words. This structured progression ensures a systematic approach to phonological learning, supporting gradual and continuous improvement in users' abilities to understand and use sounds and syllables to form words.

After completing each learning module, users take an evaluation test presented as an interactive game. This game incorporates Rapid Automatized Naming (RAN) to assess the speed of verbal material retrieval by measuring how quickly users identify letters, syllables, and words on screen [16]. While the game does not use speech input, it simulates RAN through visual recognition tasks. Dyslexic individuals often score lower on RAN tasks, affecting their reading fluency and comprehension, which can hinder academic progress. Integrating RAN into the monitoring process helps educators and parents identify areas where users may need additional support, promoting continuous growth and academic achievement. Test scores decide whether users can progress to the next module.

The accuracy of answers and response times are recorded to calculate the overall score. A congratulatory page appears if a user's score is above the required minimum. Thereafter, the user can proceed to the next module. Otherwise, the user will be redirected back to the current module. This evaluation continues throughout the modules, generating a cumulative progress report that allows parents to monitor their children's progress over time.

3.2. Software Components

The development process of the proposed app utilizes several software components involving Visual Studio Code, Android Studio, Flutter, and Adobe XD. Visual Studio Code is a widely used code editor that supports multiple programming languages. Additionally, it includes essential tools like debugging and Git integration. Android Studio is a powerful platform for coding, debugging, and testing. Besides, it offers an emulator that allows testing of applications on various Android devices. Flutter empowers the development of local applications for mobile, web, and desktop platforms from a single codebase using Dart. Finally, Adobe XD is a component used to create interactive and visually appealing designs for mobile applications. Its design process seamlessly adapts to diverse mobile sizes, perfecting the user experience. These software components are summarized in Table 2.

https://doi.org/10.31436/iiumej.v26i2.3429

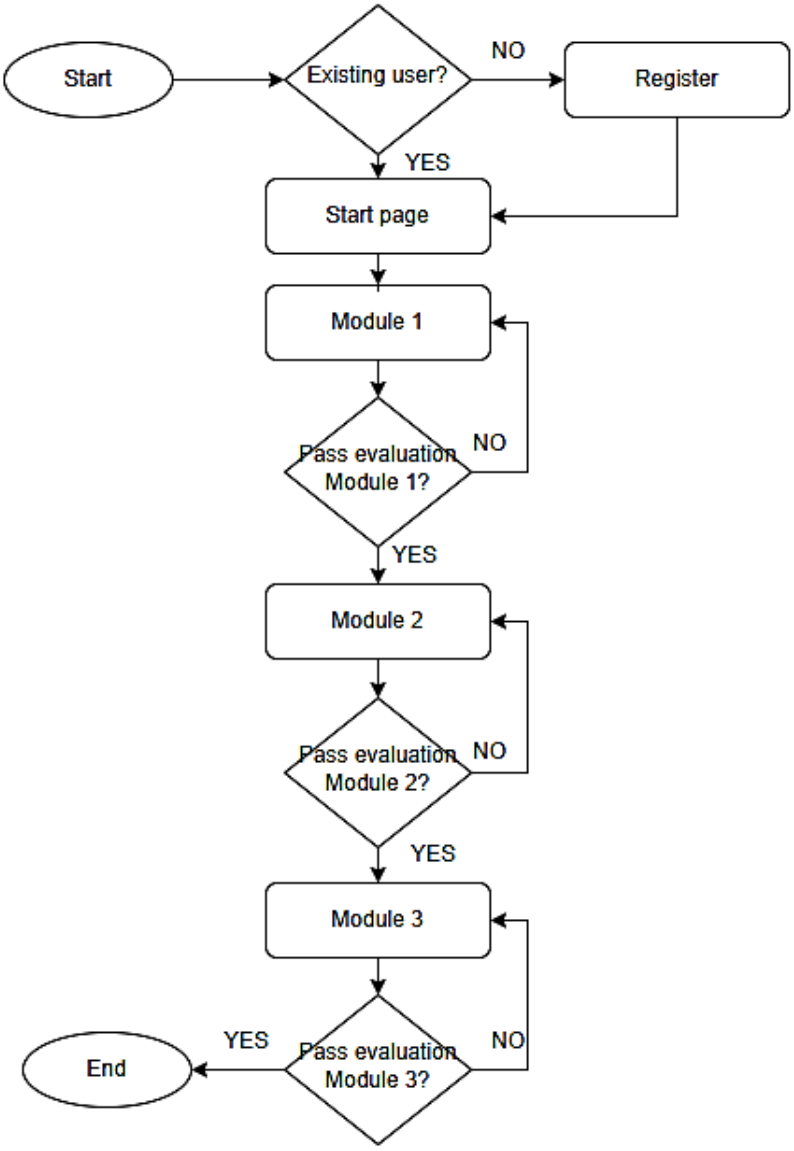


Figure 1. Flowchart of the proposed mobile app

3.3. User Interface Design

The user interface design carefully considers font and color choices to support individuals with dyslexia. Dyslexic readers often struggle when letters are closely spaced or when colors are distracting [17]. As greater spacing has been shown to improve reading speed by reducing visual crowding, the app increases letter spacing (2.0 spacing) and widens the space between words. Helvetica font was chosen because it reduces fixation durations, which are a key measure of reading fluency, and has been associated with faster reading speeds and high user preference among individuals with dyslexia [18]. Furthermore, the app incorporates colored overlays and adjusts text presentation to favor lower brightness and contrast, following recommendations from the British Dyslexia Association. Studies show that highly contrasted colors are less effective for individuals with dyslexia, while cream-black combinations and pastel colors improve readability [19]. Personalized color adjustments are important, as color preferences can vary among dyslexic users.

https://doi.org/10.31436/iiumej.v26i2.3429

In addition to the font and color considerations, other factors listed in Table 3 were integrated into the user interface design to accommodate individuals with dyslexia. These user interface designs are grounded in established guidelines and recommendations reported in prior research, ensuring that the developed app follows validated practices rather than independently developed criteria.

Table 3. User Interface Guideline Summary [17]

Components	**Descriptions**
Fonts	Use large font size: 18 to 26 points.
	Use consistent large spacing within words and larger on between words.
	Suggested font: Courier, Helvetica, Verdana and San Serif font type.
	Use lowercase for ease of reading.
Color	Use dark color text on a light background.
	The background should not be too bright.
Navigation	Easy navigation.
Consistency	The same theme/design will be used throughout the entire application design.
Interaction	Simple instruction in text and voiceover.
	Avoid using too much flashing and moving text.
	Tapping of buttons for selection to avoid many steps and fine motor movement for fingers.
	Simple Click.
Game Type	Drag and drop.

4. RESULTS AND DISCUSSIONS

As previously discussed in Section 3, the proposed mobile app encompasses three progressive learning modules, aligning with the Orton-Gillingham approach. A user must register to use the app for the first time, which must only be done once. This registration ensures that user activity is recorded, allowing for the monitoring of progress and improvements. After registration, the user will be directed to Module 1, Module 2, or Module 3, depending on their achievements. Each module is described next.

4.1. Module 1

After logging in, users can click the Start button, prompting the app to select a word from a predefined list, identify its characters, and begin the learning module. For instance, if the chosen word is “Kucing,” users will be prompted to learn each character in “Kucing,” including “K,” “U,” “C,” “I,” “N,” and “G.” Each character is displayed in both uppercase and lowercase formats. This module integrates key components of the Structured Literacy approach, such as phonology and phonics. As the user clicks the forward arrow icon, the application audibly pronounces each letter, enhancing the learning experience. Fig. 2 shows the user interface and the flow of Module 1.

https://doi.org/10.31436/iiumej.v26i2.3429

Figure 2. Module 1 (Learning Module)

Upon completing Module 1, users move on to a Drag-and-Drop game (Fig. 3) to test their comprehension. In this figure, each box contains a character. The upper boxes contain the characters 'K', 'U', 'C', 'I', 'N', and 'G', respectively, while the lower boxes contain the characters 'I', 'G', and 'N'. The user must drag each lower box and drop it onto its correct matching box above. For example, the lower box labeled 'I' must be dropped onto the upper box labeled 'I'. Incorrect placements/drops trigger sound effects to signal errors, and each drag action plays the corresponding character's sound. Performance is scored based on accuracy. If users achieve 75% accuracy or higher, a congratulatory message as shown in Fig. 4 appears, and they can move on to the next module. Otherwise, they can either retry the game or revisit the learning session. This game uses the Orton-Gillingham method, with repeated lessons to reinforce understanding and gradual progression to more complex content. The time taken to complete the game is also recorded.

Figure 3. Module 1 (Drag & Drop Game)

https://doi.org/10.31436/iiumej.v26i2.3429

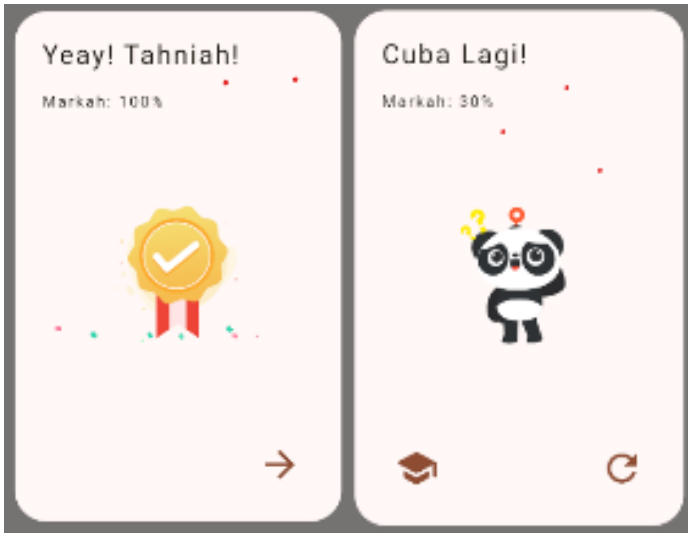


Figure 4. Congratulatory and Retry Message

4.2. Module 2

In Module 2, users progress to the next foundational pillar of Literacy as outlined by the Structured Literacy approach, syllable knowledge. Users are now introduced to syllables based on the knowledge gained in Module 1, where words were studied and learned as individual characters. For instance, a word previously introduced, such as "Kereta," is segmented into its constituent syllables: "Ke," "re," and "ta." Like Module 1, Module 2 includes an audio feature that reads each syllable aloud, enhancing phonological awareness and supporting a more comprehensive approach to literacy development. The user interface and the flow of Module 2 are illustrated in Fig. 5.

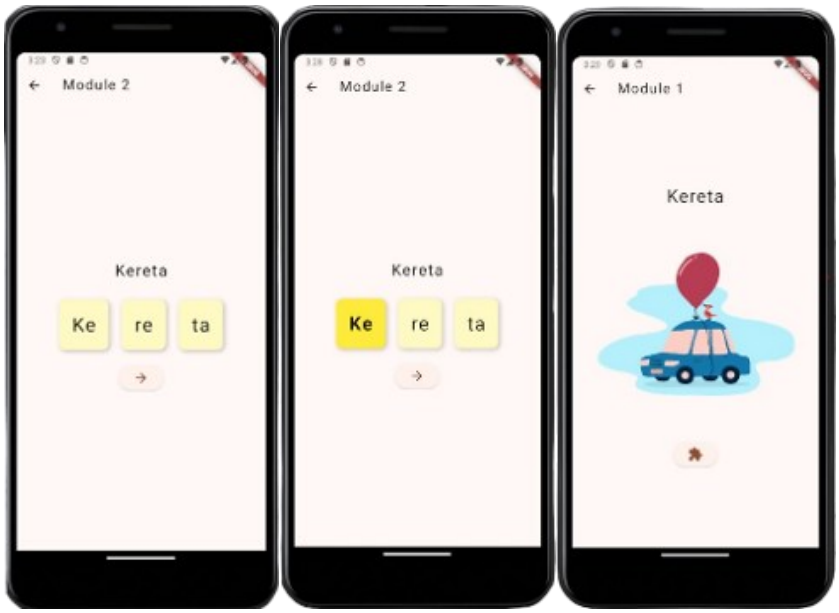


Figure 5. Module 2 (Learning Module)

After successfully completing Module 2, users engage in a puzzle game where they arrange syllables to form complete words, as shown in Fig. 6. This game format enhances syllable recognition and word formation skills. A blue box placed incorrectly in a grey box counts as an incorrect attempt, while a correctly placed blue box that turns the grey box yellow

https://doi.org/10.31436/iiumej.v26i2.3429

is considered correct. Accuracy is calculated similarly to Module 1, with a congratulatory message displayed for successful performance.

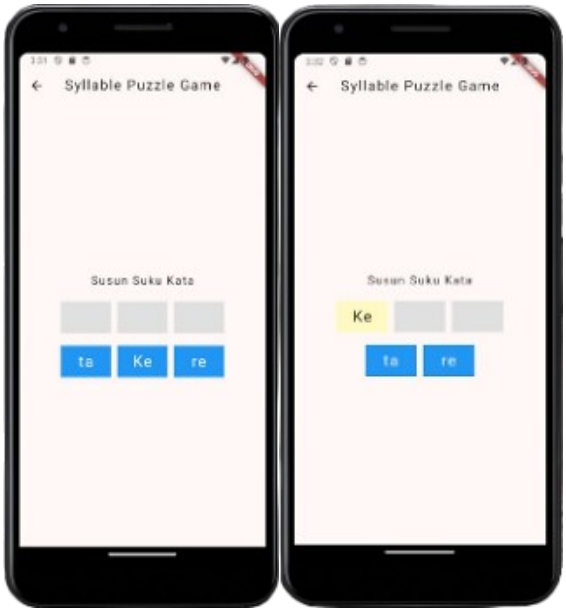


Figure 6. Module 2 (Syllable Puzzle Game)

4.3. Module 3

Following mastering syllables, users progress to Module 3, which focuses on meaning association. To activate this module, users must interact with the app at least twice to familiarize themselves with two different words. This module features two interactive games designed to reinforce meaning association and assess understanding. The first game is a matching game where users pair images with corresponding words. A selected box highlights, and if the correct image is chosen, the box turns yellow to indicate a correct match, as shown in Fig. 7.

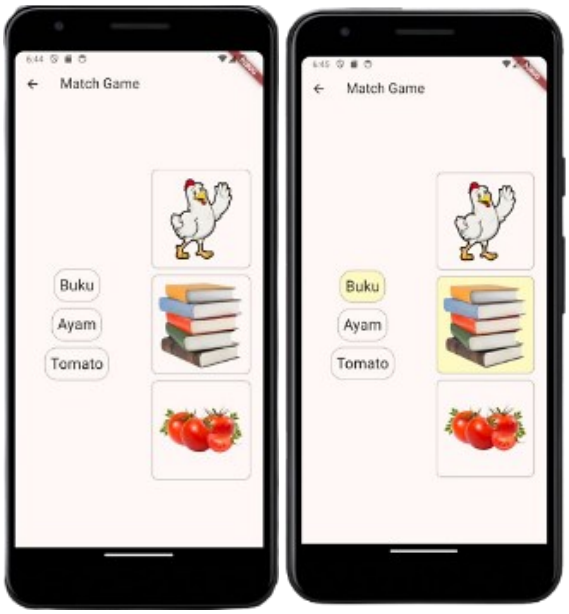


Figure 7. Module 3 (Match Game)

https://doi.org/10.31436/iiumej.v26i2.3429

The second game in this module is a classic card memory game where players match pairs of cards with words and images. This game is crucial for enhancing phonological memory, which involves coding information in a sound-based system for temporary storage. Since limitations in phonological memory can impact reading comprehension, engaging in this game helps users practice and strengthen their phonological memory, improving their ability to store and recall sound information effectively.

Each game serves as a diagnostic tool to monitor improvement and identify weaknesses, with accuracy and completion time carefully recorded as critical measures of progress in phonological memory and comprehension skills. To avoid distractions, accuracy results are not displayed on the user interface. Still, they are accessible within the app, as shown in Fig. 9 and Fig. 10. A study comparing word reading accuracy found significant differences between typically developing children and those with dyslexia, with typically developing children achieving a mean accuracy of 91.43%, while dyslexic children averaged 72.47%. These findings underscore the challenges dyslexic children face in decoding and recognizing words. Given these results, the app's 75% accuracy requirement is well justified. This standard applies to all students, including those with dyslexia, who must achieve or exceed this accuracy level to advance to the next module. Setting the threshold slightly above the average accuracy observed among dyslexic students (72%) implies a significant improvement in their reading abilities. It indicates meaningful progress in overcoming the specific challenges associated with dyslexia, potentially facilitated by effective interventions or strategies aimed at enhancing word recognition and overall reading fluency [20].

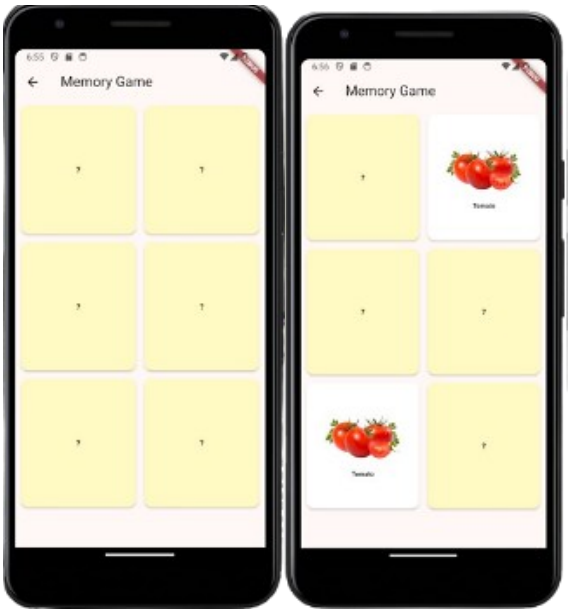


Figure 8. Module 3 (Memory Game)

```
I/flutter ( 8356): User spelled out the word: Kucing
I/flutter ( 8356): Correct Attempts: 6
I/flutter ( 8356): Incorrect Attempts: 1
I/flutter ( 8356): Total Attempts: 7
I/flutter ( 8356): The accuracy: 85.71%
I/flutter ( 8356): Time taken to finish test: 13 seconds
I/flutter ( 8356): User will proceed to next module, Module 2: Syllable Knowledge
```

Figure 9. Accuracy above 75%

https://doi.org/10.31436/iiumej.v26i2.3429

```
I/flutter ( 8356): User spelled out the word: Kucing
I/flutter ( 8356): Correct Attempts: 6
I/flutter ( 8356): Incorrect Attempts: 3
I/flutter ( 8356): Total Attempts: 9
I/flutter ( 8356): The accuracy: 66.67%
I/flutter ( 8356): Time taken to finish test: 20 seconds
I/flutter ( 8356): User needs to reattempt Module 1
```

Figure 10. Accuracy below 75%

5. CONCLUSIONS

This paper proposes a dyslexic-friendly mobile learning application designed to address the educational challenges of dyslexic children. It integrates the Orton-Gillingham and Structured Literacy approaches to provide progressive and cumulative learning through modules focused on phonology, syllable knowledge, and meaning association. The app was developed in Malay to improve literacy outcomes for dyslexic students in Malaysia and contribute to a more inclusive educational landscape. The app features an enhanced dyslexic-friendly interface with accessible fonts, color configurations, and interactive games that monitor user progress by identifying weaknesses and improvements.

Despite its promising features, the app has several limitations, particularly related to the learning materials and its effectiveness in delivering a dyslexic-friendly user experience. Although the materials were based on a syllabus for dyslexic individuals, further clarification and validation from teaching professionals are needed to ensure the content is appropriately tailored. Additionally, the app's usability, including average completion time, scores across modules, and clarity of instructions and content, has not yet been fully evaluated with both dyslexic and non-dyslexic users. Future work should involve thorough usability and acceptance testing with both dyslexic and non-dyslexic learners to gather feedback and subsequently refine the app's effectiveness. A thorough review of the learning content should also be conducted with input from teaching professionals to strengthen its educational foundation. Planned enhancements will also include improving the text-to-speech functionality, refining syllable segmentation algorithms for Malay, and creating a more interactive score reporting interface for parents. Eventually, the app's performance should be compared with other available tools to assess its capability and effectiveness in supporting dyslexic-friendly learning.

REFERENCES

[1] K. M. Jabatan, "Laporan Statistik 2022," 2023. [Online]. Available: https://www.jkm.gov.my/jkm/uploads/files/Bahagian Kawalan Standard/Laporan Statistik 2022.pdf

[2] Ministry of Education Malaysia, *Buku Data Pendidikan Khas Tahun 2023*. 2023. Accessed: Aug. 25, 2024. [Online]. Available: https://www.collegesidekick.com/study-docs/3227194

[3] M. A. Dzulkifli, "Children with Learning Disabilities in Malaysia," *IntechOpen-Intellectual Learn. Disabil. - Inclusiveness Contemp. Teach. Environ.*, pp. 1–12, Mar. 2024,

[4] A. Osman, A. Che Ahmad, and W. A. J. Wan Yahaya, "A Glimpse of Malaysian Teachers' Reading Intervention Strategies for Children with Dyslexi," *J. Comput. Res. Innov.*, vol. 1, pp. 116–126, 2016, Accessed: Aug. 23, 2024. [Online]. Available: https://www.researchgate.net/publication/309857154_A_Glimpse_of_Malaysian_Teachers%27_Reading_Intervention_Strategies_for_Children_with_Dyslexi

[5] S. M. Daud and H. Abas, "'Dyslexia baca' mobile app - The learning ecosystem for dyslexic

children," *Proc. - 2013 Int. Conf. Adv. Comput. Sci. Appl. Technol. ACSAT 2013*, pp. 412–416, 2013,

[6] R. Skiada, E. Soroniati, A. Gardeli, and D. Zissis, "EasyLexia: A mobile application for children with learning difficulties," *Procedia Comput. Sci.*, vol. 27, pp. 218–228, 2014,

[7] S. Rajapakse, D. Polwattage, U. Guruge, I. Jayathilaka, T. Edirisinghe, and S. Thelijjagoda, "ALEXZA: A Mobile Application for Dyslexics Utilizing Artificial Intelligence and Machine Learning Concepts," *2018 3rd Int. Conf. Inf. Technol. Res. ICITR 2018*, Dec. 2018,

[8] R. E. P. MacUlada, A. R. Caballero, C. G. Villarin, and E. M. Albina, "FUNologo: An Android-based Mobile Virtual Reality Assisted Learning with Speech Recognition Using Diamond-Square Algorithm for Children with Phonological Dyslexia," *2023 8th Int. Conf. Bus. Ind. Res. ICBIR 2023 - Proc.*, pp. 403–408, 2023,

[9] I. Sarah, K. Soundarya, S. Tamil Thendral, R. Dhanalakshmi, and T. Deenadayalan, "DYS-I-CAN: An Aid for the Dyslexic to improve the skills using Mobile Application," *2020 Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2020*, Jul. 2020,

[10] S. Thelijjagoda, M. Chandrasiri, D. Hewathudalla, P. Ranasinghe, and I. Wickramanayake, "The hope: An interactive mobile solution to overcome the writing, reading and speaking weaknesses of dyslexia," *14th Int. Conf. Comput. Sci. Educ. ICCSE 2019*, pp. 808–813, Aug. 2019,

[11] I. Avishka, K. Kumarawadu, A. Kudagama, M. Weerathunga, and S. Thelijjagoda, "Mobile App to Support People with Dyslexia and Dysgraphia," *2018 IEEE 9th Int. Conf. Inf. Autom. Sustain. ICIAfS 2018*, Dec. 2018,

[12] L. Sandathara, S. Tissera, R. Sathsarani, H. Hapuarachchi, and S. Thelijjagoda, "Arunalu: Learning ecosystem to overcome sinhala reading weakness due to dyslexia," *ICAC 2020 - 2nd Int. Conf. Adv. Comput. Proc.*, pp. 416–421, Dec. 2020,

[13] A. B. L. dos Anjos, A. L. de A. Barbosa, and C. A. S. Azoni, "Phonological processing in students with developmental dyslexia, ADHD and intellectual disability," *Rev. CEFAC*, vol. 21, no. 5, 2019,

[14] V. Subramaniam and N. S. A. Nasir, "Multisensory Therapy in Letter Reversal of Dyslexic Pupils," *Univers. J. Educ. Res.*, vol. 8, no. 12, pp. 7118–7130, Dec. 2020,

[15] J. Peavler and T. Rooney, *Orton Gillingham, 'Join the journey in support of reading', Teacher Manual*. M.A. Rooney Foundation, 2019.

[16] M. W. E. Norton, "Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities," *Annu. Rev. Psychol.*, 2012.

[17] M. 'Azizi C. Sulaiman and A. Bain, "User Interface Guidelines for Dyslexic Game-Based Learning on Selected Usability Test Method," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.4, pp. 439–445, 2019.

[18] L. Rello and R. Baeza-Yates, "Good fonts for dyslexia," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*, 2013, no. May.

[19] F. T. Yoliando, "A Comparative Study of Dyslexia Style Guides in Improving Readability for People With Dyslexia," vol. 502, no. Imdes, pp. 32–37, 2020,

[20] P. B. da Silva *et al.*, "Rapid Automatized Naming and Explicit Phonological Processing in Children With Developmental Dyslexia: A Study With Portuguese-Speaking Children in Brazil," *Front. Psychol.*, vol. 11, May 2020,

https://doi.org/10.31436/iiumej.v26i2.3463

DESIGN AND OPTIMIZATION OF A FLEXIBLE ANTENNA FOR ISM BAND WEARABLE DEVICES VIA INSET SLOT INTEGRATION AND PARASITIC ELEMENTS

AIMAN HAKIMI RAHIMI, SARAH YASMIN MOHAMAD*, NORUN FARIHAH ABDUL MALEK, FARAH NADIA MOHD ISA, AHMAD ZAMANI JUSOH, AISHA HASSAN ABDALLA HASHIM

Microwave, Communication and Information Systems Engineering (MCISE) Research Group, Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

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

(Received: 4 October 2024; Accepted: 13 January 2025; Published online: 15 May 2025)

ABSTRACT: This paper aims to analyze and optimize the electrical properties of a flexible antenna to maximize its potential. The antenna's performance is enhanced by incorporating inset slots on both the patch and ground plane, as well as utilizing parasitic elements to improve the resonant frequency and achieve a wider bandwidth. The flexible antenna is designed to resonate within the Industrial, Scientific, and Medical (ISM) frequency band of 5.725–5.875 GHz, using Rogers RO4003C with a dielectric constant of 3.55 and dimensions of 20 × 25 × 0.2032 mm³. The proposed antenna demonstrates promising results, with S_{11} of -37.05 dB at 5.78 GHz, and a bandwidth of 158 MHz, ranging from 5.739–5.897 GHz. The usable ISM bandwidth is 136 MHz, representing 90.67% of the ISM frequency band. Theoretical, simulation, and experimental analyses confirm that embedding slots on the patch or ground plane, along with the use of parasitic elements, significantly enhances the antenna's resonant frequency, bandwidth, and efficiency. Because of its flexibility, improved resonant frequency, and wide bandwidth, this antenna has potential uses in the ISM frequency band to enable faster and more efficient data transmission.

ABSTRAK: Kajian ini bertujuan bagi menganalisa dan mengoptimum sifat elektrik antena fleksibel bagi memaksimum potensinya. Ciri-ciri antena dipertingkatkan lagi dengan penggunaan slot sisipan pada tampalan dan satah tanah, dan menggunakan elemen parasitik bagi meningkatkan frekuensi resonan dan mencapai lebar jalur yang lebih luas. Antena fleksibel ini direka untuk frekuensi Industri, Saintifik dan Perubatan (ISM) iaitu pada frekuensi 5.725–5.875 GHz menggunakan Rogers RO4003C dengan pemalar dielektrik 3.55 dan dimensi 20 × 25 × 0.2032 mm³. Antena yang dicadangkan ini menunjukkan hasil yang baik dengan S_{11} -37.05 dB pada frekuensi 5.78 GHz, dan jalur lebar 158 MHz, dalam julat 5.739–5.897 GHz. Jalur lebar ISM mencapai 136 MHz iaitu bersamaan 90.67% jalur frekuensi ISM. Analisis teori, simulasi dan eksperimen membuktikan bahawa penggunaan slot pada tampalan atau di atas satah tanah, bersama elemen parasitik dapat membantu meningkatkan frekuensi resonan, lebar jalur dan kecekapan antena. Oleh kerana fleksibilitinya, frekuensi resonan meningkat dan jalur lebar bertambah luas. Antena ini berpotensi bagi penggunaan jalur frekuensi ISM dengan penghantaran data yang lebih pantas dan cekap.

KEYWORDS: *Flexible microstrip antenna, inset slots, parasitic elements, Industrial, Scientific, and Medical (ISM) band.*

https://doi.org/10.31436/iiumej.v26i2.3463

1. INTRODUCTION

Modern communication networks have seen substantial advancements in system development, driven by technological integration and innovation. Among the critical components in communication technology, antennas stand out for their reliability. However, traditional antennas, with their rigid and fragile structures, often lack the durability and flexibility required for specific applications [1,2]. In contrast, flexible antennas, which rely heavily on the substrate material, are crucial to the functionality of contemporary communication systems. A variety of materials have been researched, tested, and employed in the development of flexible antennas, including rubber [3], polydimethylsiloxane (PDMS) [4–5], polyethylene terephthalate (PET), thermoplastic polyurethane, and elastomers (TPU and TPE) [6], and Rogers RO4003C [7]. These materials provide flexible antennas with excellent elasticity, stable electrical properties, and the added benefits of being lightweight and low profile, making them more adaptable compared to traditional antennas mounted on rigid substrates. For wearable applications, the antenna must be physically flexible, capable of conforming to the body, especially since it involves integrating with the body on uneven surfaces, as shown in Figure 1 [8].

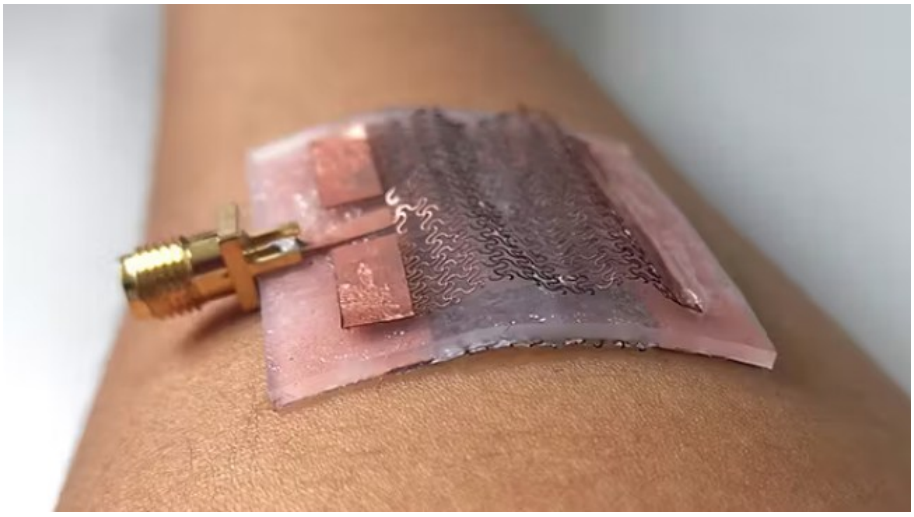


Figure 1. On-body flexible antenna [8].

According to [9], wearable antennas are highly versatile, capable of achieving multiband designs and characteristics suited for various applications, including the ISM band, C-band, Bluetooth, Worldwide Interoperability for Microwave Access (WiMAX), 5G smartphones, radio altimeters, and satellite TV. While several flexible antennas have been developed, many do not operate within the ISM band, lack multiband functionality, or do not provide wideband performance.

Wireless applications greatly benefit from antennas that offer both flexibility and broad bandwidth. A wider bandwidth is particularly advantageous for wireless networking, allowing for better utilization of the designated frequency spectrum. Over the past decade, the adoption of broadband technology has grown substantially due to its numerous advantages, including enhanced communication security, higher data transfer rates, improved power efficiency, reduced signal interference, optimized spectrum utilization, and simplified circuitry.

Several techniques have been used to enhance the bandwidth of microstrip antennas, including the use of inset slots on the patch [10–11], the introduction of parasitic elements [12], and modifications to the ground plane [13]. In [10], the authors employed slots and an inset

feed to improve impedance matching, gain, and return loss for Wireless Body Area Network (WBAN) applications. Antenna design plays a crucial role in WBAN, as comfort and reliability directly influence antenna efficiency. Polyimide was selected as the substrate due to its flexibility and low dielectric constant, which minimizes surface wave losses, making it ideal for this application. Proper feeding techniques were used to control impedance matching and return loss, which in turn affect various antenna performance parameters.

In [12], U-shaped slotted antennas have gained popularity in antenna design due to their ability to improve performance, particularly in terms of gain. The U-slot extends the current path, allowing for current perturbation that generates circular polarization. The inclusion of parasitic elements, positioned beside the central patch, further enhances the antenna's bandwidth by acting as additional radiators. The authors also integrated parasitic elements in a single-layer patch, along with slots, to boost bandwidth, gain, and efficiency for applications such as Wireless Gigabit (WiGig) and Wireless Personal Area Network (WPAN). Additionally, modifications to the ground plane have been implemented to enhance impedance bandwidth [13]. Ground plane alterations, such as inserting slots and using metamaterials, have been widely accepted by researchers as they significantly contribute to generating multiband frequencies.

Moreover, the laminating manufacturing technique employed for the single-patch and Multiple Input Multiple Output (MIMO) textile antenna in [14] enhances the reproducibility of textile antennas. The MIMO system, which utilizes multiple transmitters and receivers, improves performance, reliability, and coverage. This system is designed for fifth-generation (5G) on-body applications, further advancing IoT development and extending 5G frequencies to cover the ISM band at 5.8 GHz. Traditional manufacturing processes often introduce inaccuracies, such as frequency shifts, that can negatively impact antenna performance.

All these studies demonstrate that adding slots improves both resonant frequency and return loss, while parasitic elements contribute to broader bandwidth. Ground plane modifications increase parasitic capacitance in the fringing field due to a defective ground structure, thereby enhancing the coupling between the ground plane and the patch, ultimately resulting in increased bandwidth.

In this paper, we present methods to enhance the bandwidth and performance of a flexible microstrip antenna operating within the ISM frequency band range of 5.725–5.875 GHz. The antenna design includes a coplanar waveguide (CPW) technique, parasitic elements, and inset slots on both the patch and ground plane. The inset slots on the patch are expected to improve resonant frequency and return loss by extending the surface current paths of the resonant modes, thus lowering the corresponding resonant frequencies. Moreover, including parasitic elements enhances antenna efficiency and broadens the bandwidth. The design methodology and computations, based on fundamental equations, are detailed in Section 2. Section 3 presents an analysis of the simulation and experimental results, focusing on key performance metrics such as return loss, bandwidth, directivity, efficiency, VSWR, and radiation pattern. The optimized antenna, utilizing the flexible Rogers RO4003C substrate, demonstrated a return loss of -37.05 dB and a wide bandwidth of 158 MHz (5.739–5.897 GHz). Within the 5.725–5.875 GHz ISM frequency band, the antenna achieved a bandwidth of 136 MHz. We also present the antenna's performance under 45° bending conditions, with both simulated and measured results demonstrating acceptable performance. Despite frequency shifts, the antenna maintains reliable functionality thanks to its wide bandwidth.

https://doi.org/10.31436/iiumej.v26i2.3463

2. METHODOLOGY

2.1. Antenna Equations

The proposed antenna was designed, simulated, and analyzed using Computer Simulation Technology Microwave Studio (CST MWS) software. Based on calculations derived from the fundamental equation for a basic rectangular patch antenna, a rectangular patch antenna targeting an operating frequency of 5.78 GHz (the center of the 5.725–5.875 GHz ISM band) was developed.

Three key factors were considered to determine the patch dimensions and ensure proper impedance matching between the radiating patch and the transmission line: the selected material's resonant frequency, dielectric thickness, and dielectric constant. For this design, the Rogers RO4003C substrate was chosen for its flexibility and favorable characteristics, including a dielectric constant of 3.55, a thickness of 0.2032 mm, and a dielectric strength of 780 V/mil. The specified values of the operating frequency (f_o), substrate thickness (h) and substrate relative permittivity of the dielectric constant (ε_r) are applied accordingly to define the computational value of patch width (W_p) and patch length (L_p). The patch width (W_p) is calculated as in Eq. (1).

$$W_p = \frac{C}{2f_o\sqrt{\frac{(\varepsilon_r+1)}{2}}} \tag{1}$$

Where c is the speed of light, 3×10^8 m/s. The effective dielectric constant (ε_{eff}) is calculated using Eq. (2), which is based on the substrate height, dielectric constant, and calculated patch width in Eq. (1).

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\left[1 + 12\frac{h}{W_p}\right]^{-\frac{1}{2}} \tag{2}$$

The patch length (L_p) is calculated using Eq. (3), which involves the difference between effective length (L_{eff}) and length extension (ΔL), calculated using Eqs. (4) and (5) respectively.

$$L_p = L_{eff} - 2\Delta L \tag{3}$$

$$L_{eff} = \frac{C}{2f_o\sqrt{\varepsilon_{eff}}} \tag{4}$$

$$\Delta L = 0.412h\frac{(\varepsilon_{eff}+0.3)\left(\frac{W_p}{h}+0.264\right)}{(\varepsilon_{eff}-0.258)\left(\frac{W_p}{h}+0.8\right)} \tag{5}$$

The feedline width (W_f) is determined using macros in CST features, which calculate the analytical line impedance to achieve a suitable width for good impedance matching of 50Ω between the feedline and the antenna radiating patch. Moreover, the feedline width can be calculated using Eq. (6), where Z_o represents the input impedance. Securing impedance matching is essential to ensure that power can be supplied without having losses, leading to improved accuracy in antenna performance.

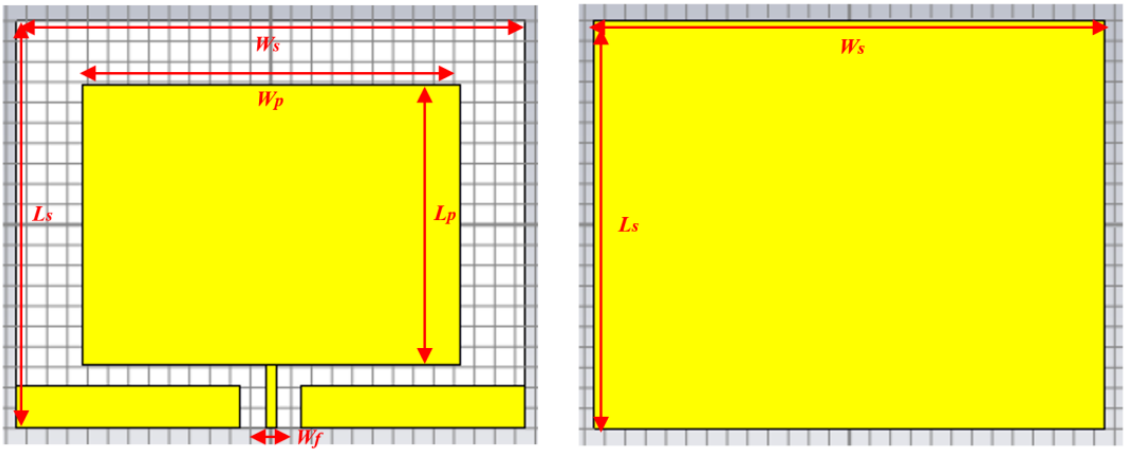
$$W_f = \frac{7.48h}{e^{\left(Z_o\sqrt{\frac{(\varepsilon_r+1.41)}{87}}\right)}} - 1.25t \tag{6}$$

Applying these equations and optimizing the dimensions obtained a patch width of 18.53 mm and a length of 13.81 mm.

https://doi.org/10.31436/iiumej.v26i2.3463

2.2. Optimization and Finalized Dimensions

Figure 2 illustrates the antenna dimensions before optimization, derived from the equations in Section 2.1. Several modifications were applied to the initial patch design to achieve the desired resonant frequency of 5.78 GHz and enhance overall performance. These included the introduction of an inset feedline (Figure 3a), the addition of slots on both the patch (Figure 3b–c) and ground plane (Figure 3d), the integration of two parasitic elements (Figure 3e), and the incorporation of cutting edges on the patch antenna (Figure 3f).



(a) Top View (b) Bottom View

Figure 2. Initial antenna dimensions before optimization.

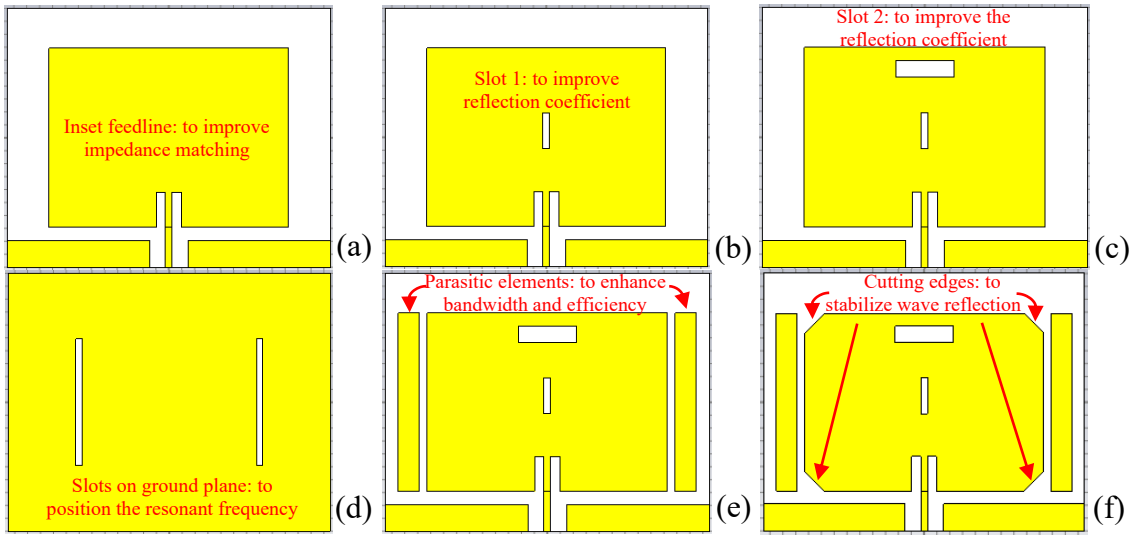


Figure 3. Optimization stages of the proposed antenna: (a) inset feedline, (b) slot 1 on patch, (c) slot 2 on patch, (d) slots on ground plane, (e) parasitic elements, and (f) cutting edges.

To address the limited bandwidth of a standard rectangular patch antenna, the feedline is designed with a coplanar waveguide (CPW), incorporating ground plane elements. The process begins with introducing an inset feedline, as shown in Figure 3a, to ensure impedance matching between the radiating patch and the feedline. Subsequently, slots are introduced to the patch and the ground plane to improve reflection coefficient performance and enhance overall efficiency, as depicted in Figure 3 b–d. Previous studies [9] have demonstrated that introducing a slot into the antenna structure can extend the surface current paths of the resonant modes,

https://doi.org/10.31436/iiumej.v26i2.3463

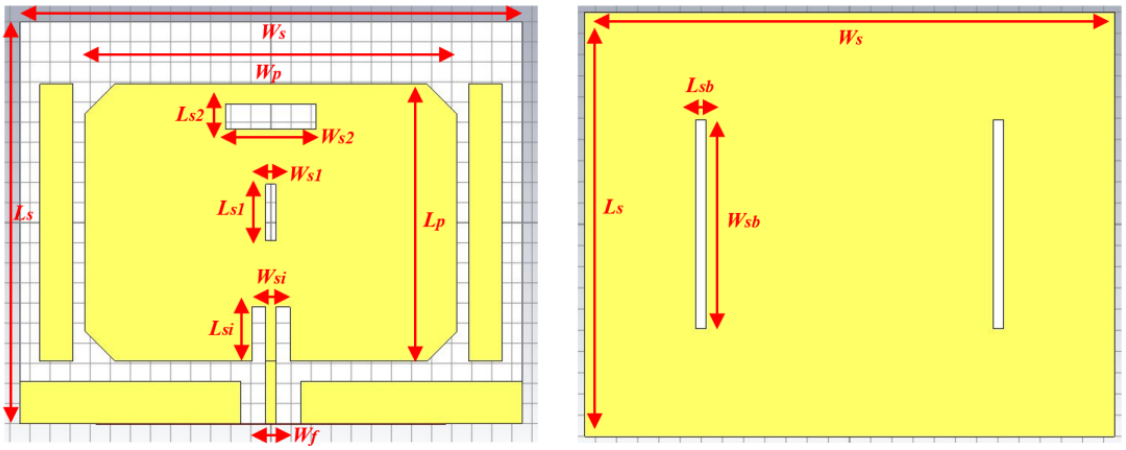
thereby lowering the corresponding resonant frequencies and improving both resonance and return loss. In implementing the CPW technique, a partial ground plane near the feedline creates gaps of 1 mm and 1.25 mm between the radiating patch and the feedline, respectively.

Additionally, two parasitic elements are positioned on the left and right sides of the radiating patch, as shown in Figure 3e, with gaps of 1 mm and 0.6 mm from the patch edge and the radiating patch, respectively. This parasitic element technique enhances bandwidth and improves antenna efficiency, as demonstrated in [13]. The cutting-edge design of the patch, depicted in Figure 3f, further stabilizes the antenna by minimizing wave reflection.

The final design was implemented on a 20 × 25 × 0.2032 mm^3 Rogers RO4003C substrate, as shown in Figure 4. The dimensions of the proposed antenna are detailed in Table 1, which includes parameters such as substrate width (W_s), substrate length (L_s), substrate thickness (h), patch width (W_p), patch length (L_p), patch thickness (t), feedline width (W_f), multiple slots consisting of inset slot (si), slot 1 ($s1$), slot 2 ($s2$) and back slot (sb).

Table 1. Finalized dimensions of the proposed antenna as in Figure 4.

Parameter	mm	Parameter	mm	Parameter	mm	Parameter	mm
W_s	25	W_p	18.53	W_{si}	0.7	W_{s2}	4.5076
L_s	20	L_p	13.81	L_{si}	2.7	L_{s2}	1.25
h	0.2032	t	0.035	W_{s1}	0.5076	W_{sb}	0.5
W_f	0.5076			L_{s1}	2.81	L_{sb}	9.81



(a) Top View (b) Bottom View

Figure 4. Finalized antenna dimensions after optimization with inset slots, parasitic elements, and cutting edges.

3. RESULTS AND DISCUSSION

The antenna's performance was evaluated in terms of return loss, bandwidth, directivity, efficiency, VSWR, and radiation pattern. The simulated return loss and bandwidth are presented in Figure 5. After optimization, the antenna achieved a resonant frequency at the center of the ISM band, with a return loss of -37.05 dB at 5.78 GHz (red curve), a significant improvement over the pre-optimization value of only -2.74 dB (green curve), which was inadequate for proper operation. Adding inset slots on both the patch and the ground plane (Figure 3 b–d) greatly enhanced the antenna's performance, precisely controlling the targeted frequency band. These slots contributed to improved bandwidth and efficiency, while

https://doi.org/10.31436/iiumej.v26i2.3463

extending the slot length on the ground plane shifted the resonant frequency to a lower range, effectively reducing the antenna size.

Furthermore, the antenna resonated within the usable frequency range of 5.739–5.897 GHz, meeting the -10 dB threshold and resulting in a bandwidth of 158 MHz, as depicted in Figure 5. This bandwidth effectively covers 136 MHz, corresponding to 90.67% of the 5.725–5.875 GHz ISM band. In the design, the parasitic elements positioned beside the central patch (Figure 3e) are electrically isolated from it but are inductively coupled, facilitating energy transfer to and from the central patch. This inductive coupling influences the antenna's impedance, generating an additional resonant frequency close to the original. While the resonant frequency produced by the parasitic elements is weaker than the primary frequency, the key objective of this technique is to broaden the antenna's bandwidth. This method has proven effective in significantly enhancing the overall bandwidth.

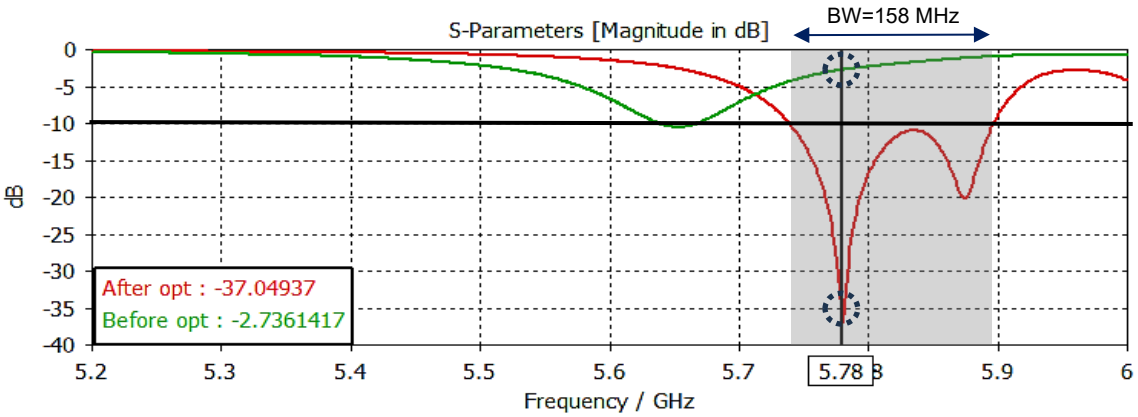
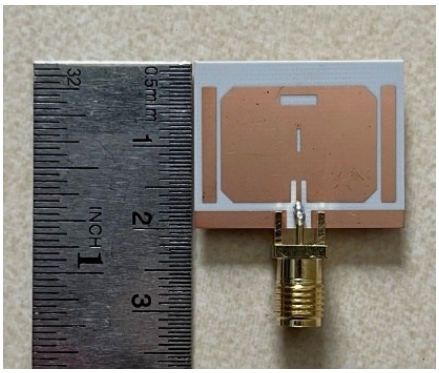
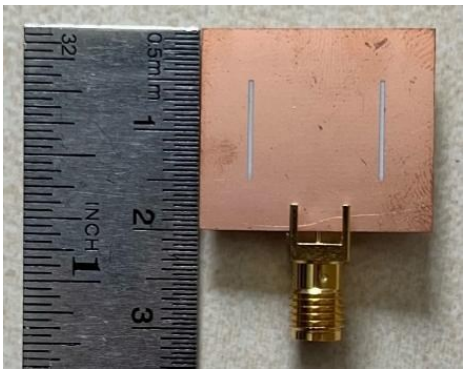


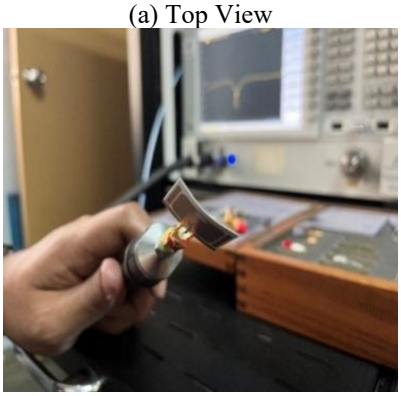
Figure 5. Simulated return loss S_{11} and bandwidth of the proposed antenna at 5.78 GHz before and after optimization.



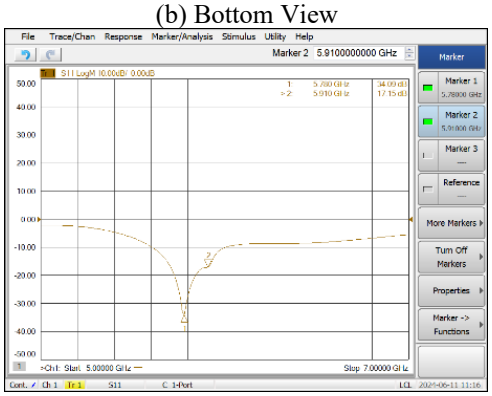
(a) Top View



(b) Bottom View



(c) Antenna Under Test using VNA



(d) Measured Return Loss

Figure 6. Fabricated antenna.

https://doi.org/10.31436/iiumej.v26i2.3463

The prototype displayed in Figure 6a–b showcases the fabricated antenna, which has overall dimensions of 20 × 25 × 0.2032 mm³. The antenna's performance was evaluated using a Vector Network Analyzer (VNA) to obtain the actual return loss results, as illustrated in Figure 6c–d.

Figure 7 compares the simulated (red curves) and measured (blue curves) return loss of the microstrip patch antenna used in this study, showing strong agreement between the two sets of results. The antenna showed a measured return loss of -35.24 dB at 5.78 GHz, with a usable frequency range of 5.6–6.0 GHz. The return loss measurements closely matched the simulations, showing two resonant frequencies in both datasets. The flexibility and electrical properties of the chosen substrate helped achieve a better measured return loss, resulting in a bandwidth of 400 MHz, over twice that of the simulated return loss. Therefore, the antenna effectively covered a frequency range of 150 MHz, corresponding to 100% of the 5.725–5.875 GHz ISM frequency band.

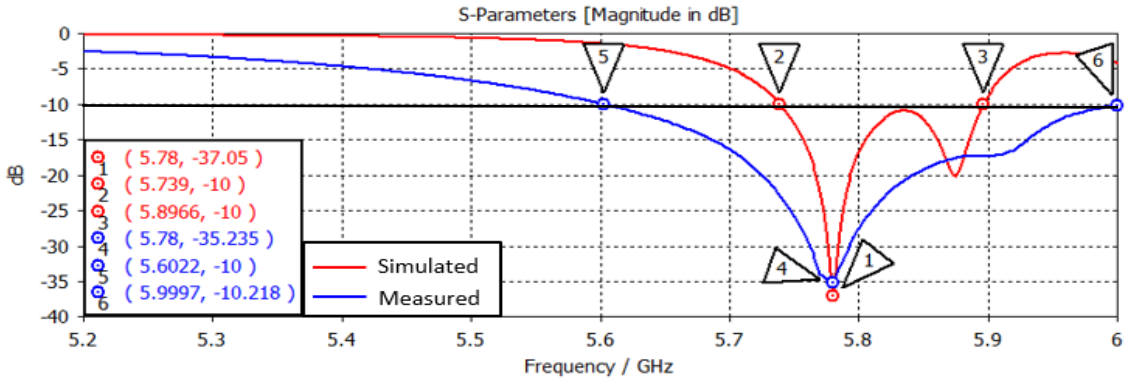


Figure 7. Simulated and measured return loss (S_{11}) under flat conditions at 5.78 GHz.

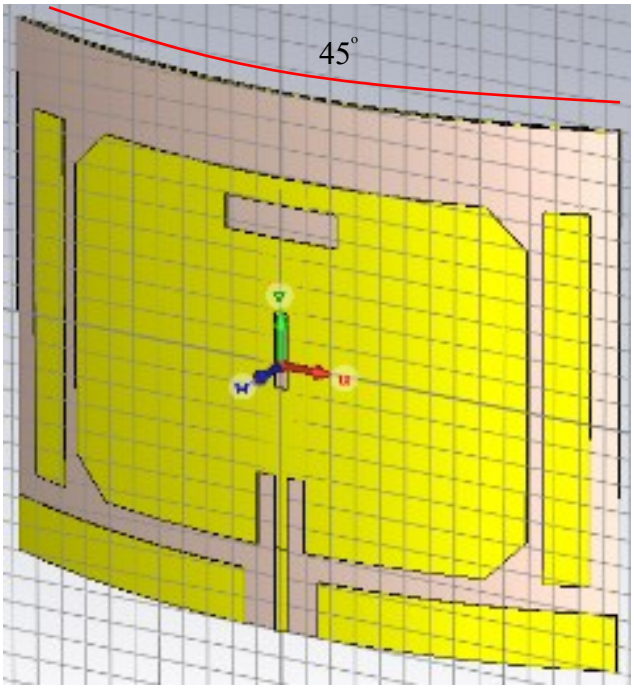


Figure 8. Antenna on 45° bend curvature in CST.

https://doi.org/10.31436/iiumej.v26i2.3463

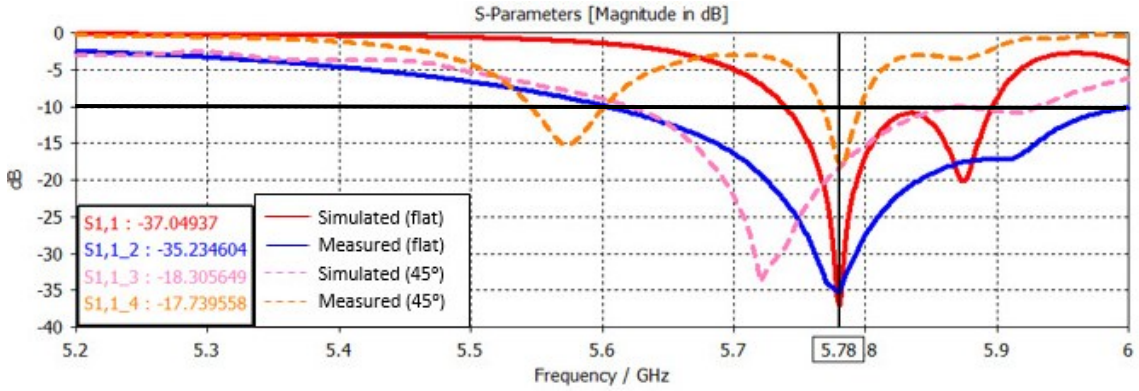
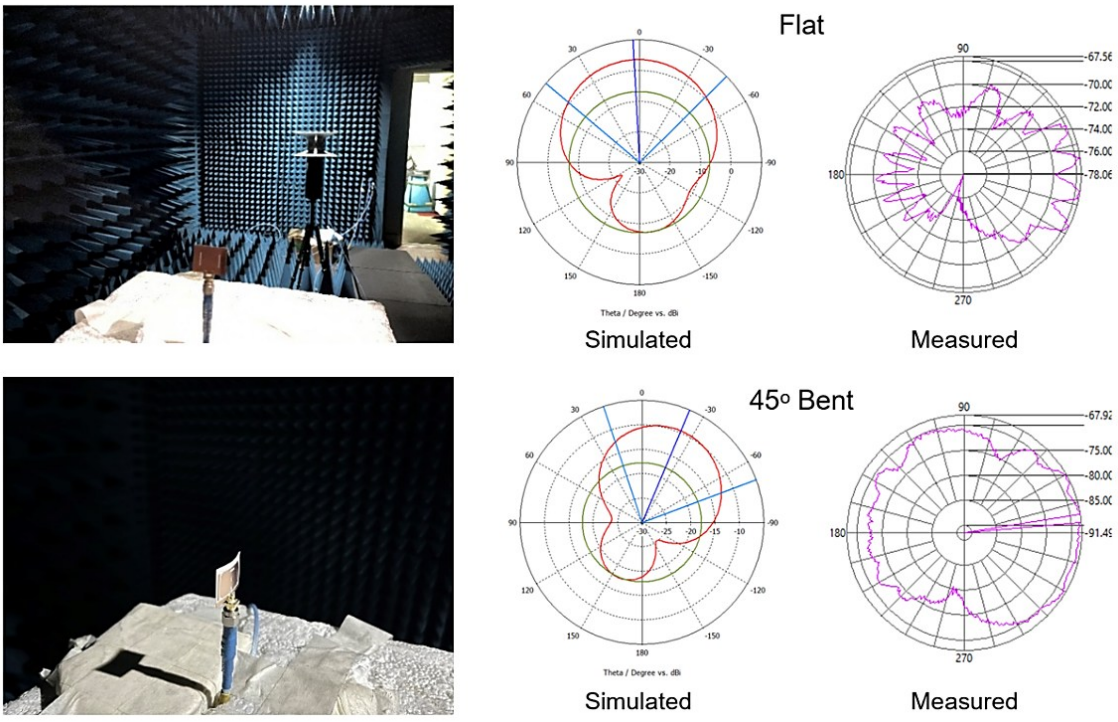


Figure 9. Simulated and measured return loss of the antenna under flat conditions and 45° bending state at 5.78 GHz.

In addition to analyzing the antenna in a flat condition, the bending of the flexible antenna was also simulated and examined. Demonstrating the antenna's flexibility is essential for wearable applications to ensure that the antenna can work efficiently in both flat and bent conditions. The microstrip patch antenna was simulated at a bending curvature of 45° (Figure 8). Bending significantly affects the flexible antenna's return loss, leading to signal strength variations due to impedance mismatches with the feedline and the influence of surface waves. As a result, signal strength increases while the resonant frequencies shift to lower values (Figure 9). These findings indicate a notable improvement in signal strength and underscore impedance mismatch as the most significantly impacted factor in this context.

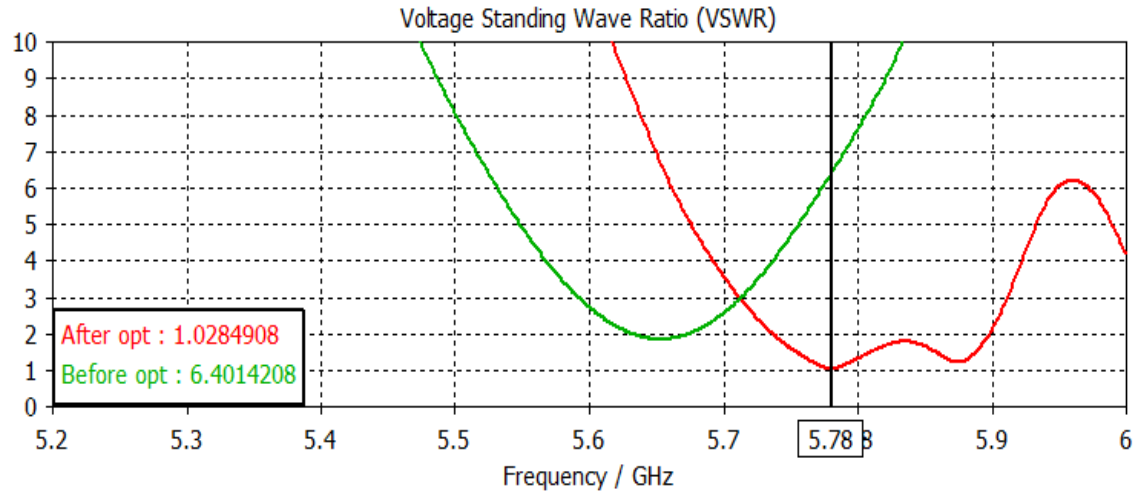


(a) (b)

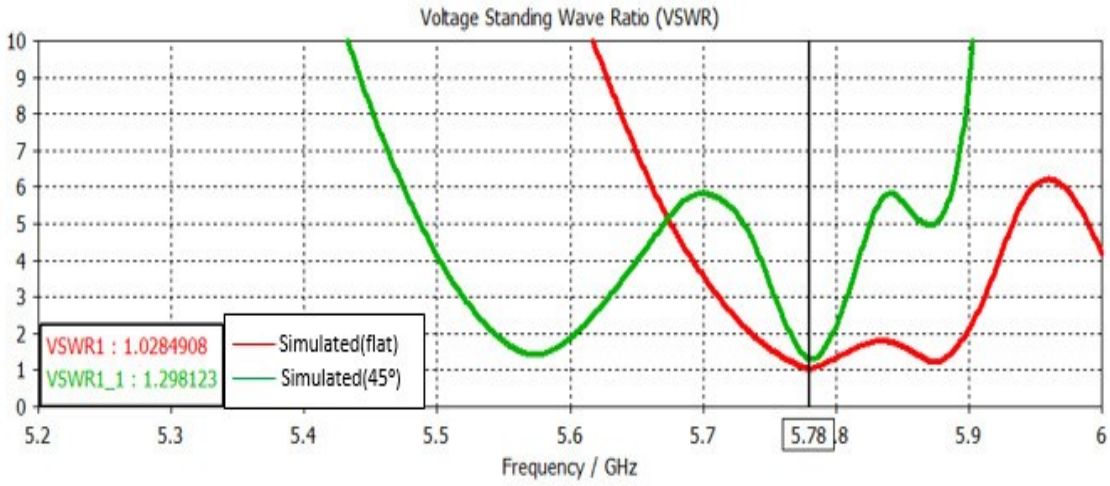
Figure 10. (a) Antenna measurement in anechoic chamber under flat and 45° bent state, (b) simulated and measured antenna pattern at 5.78 GHz under flat and 45° bent state

https://doi.org/10.31436/iiumej.v26i2.3463

The measurement setup for the antenna within an anechoic chamber, along with the polar pattern of the antenna, is presented in Figure 10a and Figure 10b, respectively. The microstrip patch antenna was placed on a custom 45° carved Styrofoam to justify the angle and ensure it can be precisely adjusted. Then, the bending is done carefully by tilting the antenna to the desired 45° angle using a mechanical tool. It is essential to ensure that the bending does not damage the antenna structure, especially on the substrate. Once the antenna is bent, it is secured in the bending position for a significant amount before it reverts to its original state. This procedure effectively allows the antenna to be measured in the anechoic chamber to achieve the highest efficiency possible. The antenna demonstrates a directional radiation pattern, which enables effective operation within the specified frequency range. The advanced patch design facilitates circular polarization with an axial ratio of 2.54, allowing the antenna to receive power from any incident wave angle consistently. Practical circularly polarized antennas typically have values of 1 to 3 in many real-world scenarios. The closer the axial ratio is to 1, the better the circular polarization. When slots are introduced on the ground plane (Figure 3d), the electric field influences the radiation pattern, behaving as a magnetic dipole. The E-plane radiation pattern experiences some changes due to diffraction effects resulting from the edges of the finite-sized ground plane.



(a) Before and after optimization



(b) Under flat and bending state of 45°

Figure 11. Simulated VSWR of the proposed antenna at 5.78 GHz.

https://doi.org/10.31436/iiumej.v26i2.3463

The radiation patterns of the antenna exhibit similar characteristics under both simulated and measured conditions in a flat configuration. However, the measured pattern under a 45° bend condition shows slight deviations from the simulated pattern due to the effects on the surface current wave. The variability in the measured patterns is likely caused by cable attenuation, which can lead to unavoidable signal distortion. Nonetheless, these observations are acceptable as they reflect the realistic conditions considered in this analysis. The antenna demonstrates a directivity of 6.26 dBi at 5.78 GHz, resulting in an antenna efficiency of 55.65%. This performance is considered satisfactory for a single radiating patch antenna. Under bending conditions, the directivity decreased to 4.29 dBi with corresponding efficiencies of 52.86% at 45° bending. However, the overall performance could be enhanced in future work by applying various modification principles related to the microstrip patch antenna design.

The Voltage Standing Wave Ratio (VSWR) fundamentally measures the degree of mismatch between an antenna and its connecting feedline. After optimization with slots and parasitic elements, the proposed antenna recorded an acceptable VSWR of 1.03, while the high VSWR of 6.44 observed prior to optimization is not ideal for patch antennas, as illustrated in Figure 11a. The effects of bending on VSWR are also depicted in Figure 11b, which shows that the VSWR remains stable even after bending, with a VSWR of 1.3 (45° bent).

Table 2 summarizes the performance of the flexible microstrip patch antenna in terms of S_{11} (dB), usable and ISM bandwidth (MHz), directivity (dBi), efficiency (%), and VSWR, before and after optimization. At 5.78 GHz, the antenna exhibits excellent performance with a simulated and measured return loss of -37.05 dB and -35.24 dB, respectively. The antenna exhibits a usable bandwidth of 158 MHz (5.739–5.897 GHz), and an ISM bandwidth of 136 MHz. The antenna also produced a directivity of 6.26 dBi, an efficiency of 55.65%, and a VSWR of 1.03.

Table 2. Antenna performances in terms of S_{11}, usable bandwidth, ISM bandwidth, directivity, efficiency, and VSWR before and after optimization in flat and 45° bend conditions at 5.78 GHz.

Parameters	Flat Before OPT	Flat After OPT	Bent (45°) After OPT	Value Improvement Flat Before vs. After OPT
S_{11} (dB)	-2.74	-37.05 (sim) -35.24 (mea)	-18.31 (sim) -17.74 (mea)	34.31 dB
Usable Bandwidth (MHz)	0	158 (sim) 400 (mea)	260 (sim) 29 (mea)	158 MHz
ISM Bandwidth (MHz)	0	136 (sim) 150 (mea)	150 (sim) 29 (mea)	136 MHz
Directivity (dBi)	6.39	6.26	4.29	-0.13 dBi
Efficiency (%)	1.75	55.65	52.86	53.9%
VSWR	6.40	1.03	1.30	5.37

*OPT = optimized, sim = simulation, mea = measurement.

Following a 45° bend, the antenna continues to perform effectively, with an S_{11} value of -18.31 dB (measured -17.74 dB). It is well known that bending can impact antenna performance, particularly by causing frequency shifts, underscoring the importance of a wide bandwidth to maintain functionality at the target frequency. While directivity and efficiency are slightly reduced to 4.29 dBi and 52.86%, respectively, these values remain within acceptable limits. Additionally, the VSWR remains below 2, with a 1.30 value indicating stable performance.

Table 3 compares the proposed antenna in this work with other ISM antennas available in the literature, operating within the frequency range of approximately 5.725–5.875 GHz. This includes a variety of different flexible materials such as FR4 [15–16], Rogers RT5880 [17–19], jeans [20], and polyimide [21]. Our antenna exhibits commendable performance within the ISM frequency band with enhanced S_{11}, bandwidth, and VSWR while maintaining reliable functionality under bending conditions. These results imply that the suggested antenna would be a good option for various ISM band applications that require a flexible structure and a wide bandwidth.

Table 3. Comparison of the proposed antenna with other latest flexible ISM antennas found in the literature [15–21] in terms of dielectric constant, flexibility, S_{11}, usable bandwidth, VSWR, bending study, and experimental analysis.

Ref	Substrate	Freq (GHz)	$\mathcal{E}_r$	Flexibility	S_{11} (dB)	BW (MHz)	VSWR	Bending study	Fabricated & measured
[15]	FR4	5.80	4.40	No	-18.32	150	1.5	No	NR
[16]	FR4	5.80	4.40	No	-20.59	315	NR	No	Yes
[17]	RT5880	5.80	2.20	No	-13.75	NR	NR	No	No
[18]	RT5880	5.80	2.20	No	-39.30	NR	1.6	No	No
[19]	RT5880	5.80	2.20	No	-13.90	NR	1.01	No	No
[20]	Jeans	5.80	NR	Yes	-31.72	NR	1.73	No	No
[21]	Polyimide	5.70	3.50	Yes	-25.00	150	NR	Yes	Yes
Proposed antenna	**RO4003C**	**5.78**	**3.55**	**Yes**	**-37.05**	**158**	**1.03**	**Yes**	**Yes**

*NR = not reported, Freq = frequency, $\mathcal{E}_r$ = dielectric constant, BW = bandwidth.

4. CONCLUSIONS

A flexible microstrip patch antenna with a coplanar waveguide (CPW) has been proposed for the 5.725–5.875 GHz ISM frequency band. The antenna is designed to resonate at 5.78 GHz on 8 mil (0.2032 mm) RO4003C material, with optimized dimensions of 20 × 25 × 0.2032 mm³. It was simulated using CST Microwave Studio (CST MWS) software and measured with a Vector Network Analyzer (VNA) to assess return loss, and with an anechoic chamber to assess the radiation pattern. Incorporating inset slots on the patch, feedline, and ground plane, along with the addition of parasitic elements adjacent to the central patch, has led to significant improvements in the antenna's performance, particularly in enhanced resonant frequency and wider bandwidth. The inset slots improve the reflection coefficient, while the parasitic element technique amplifies bandwidth and increases antenna efficiency. The achieved usable bandwidth is 158 MHz and an ISM bandwidth of 136 MHz, corresponding to 90.67% of the 5.725–5.875 GHz ISM frequency band. Additionally, the return loss performance of the antenna has improved remarkably from -2.74 dB to -37.05 dB, exhibiting acceptable directivity compared to conventional methods for designing standard patch antennas. Apart from that, the antenna was also analyzed under 45° bending, demonstrating the flexibility and satisfactory performance of the proposed antenna, which can benefit wearable applications. This antenna has potential applications in the ISM frequency band, such as in smartwatch technology and wireless body-area network (WBAN) applications, to facilitate faster and more efficient data transmission due to its flexibility, enhanced resonant frequency, and large bandwidth.

ACKNOWLEDGEMENTS

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/1/2023/TK07/UIAM/02/1.

REFERENCES

[1] Kirtania SG, Elger AW, Hasan MR, Wisniewska A, Sekhar K, Karacolak T, Sekhar PK. (2020). Flexible Antennas: A Review. Micromachine, vol. 11, p. 847. doi: 10.3390/mi11090847.

[2] Al-Haddad MASM, Jamel N, Nordin AN. (2021). Flexible Antenna: A Review of Design, Materials, Fabrication, and Applications. Journal of Physics, vol. 1878, p. 012068. doi: 10.1088/1742-6596/1878/1/012068.

[3] Ruslan AA, Mohamad SY, Malek NFA, Yusoff SH, Ibrahim SN, Isa FNM. (2020). Design of Flexible Microstrip Patch Antenna using Rubber Substrate for Brain Tumor Detection. IEEE Student Conference on Research and Development (SCOReD), pp. 1-5. doi: 10.1109/SCOReD50371.2020.9250994.

[4] Sharma PK, Chung JY. (2023). Evaluation of Polydimethylsiloxane (PDMS) as a Substrate for the Realization of Flexible/Wearable Antennas and Sensors. Micromachines, vol. 14, p. 735. https://doi.org/10.3390/mi14040735

[5] Bakar AA, Hasnan F, Razali AR, Rahim AFA, Osman MS, Ali T, Radzali R. (2019). Polydimethylsiloxane as a Potential Antenna Substrate. Acta Physica Polonica Series A, pp. 938-941. doi: 10.12693/APhysPolA.135.938.

[6] Li J, Jiang Y, Zhao X. (2019). Circularly Polarized Wearable Antenna Based on NinjaFlex-Embedded Conductive Fabric. International Journal of Antennas and Propagation, vol. 2019, pp. 1-8. doi: 10.1155/2019/3059480.

[7] Koohestani M, Azadi-Tinat N, Skrivervik AK. (2023). Compact Slit-Loaded ACS-Fed Monopole Antenna for Bluetooth and UWB Systems With WLAN Band-Stop Capability. IEEE Access, vol. 11, pp. 7540-7550. doi: 10.1109/ACCESS.2023.3238577.

[8] Wearable Antenna for Well-Being. Available: https://bte-jkt.telkomuniversity.ac.id/wearable-antenna-for-well-being-iot-untuk-hidup-yang-lebih-baik/

[9] Kaur H, Chawla P. (2022). Design and Performance Analysis of Wearable Antenna for ISM Band Applications. International Journal of Electronics, vol. 110, pp. 986-1005. doi: 10.1080/00207.2022.2068199.

[10] John AS, Murugan K. (2022). Design and Simulation of a Slotted Inset Feed Patch Antenna for Wireless Body Area Network Applications. International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1-6. doi: 10.1109/ICKECS56523.2022.10060299.

[11] Noor MSNM, Mohamad SY, Malek NFA, Isa FNM, Rahimi AH, Ramli HA. (2024). A Dual-Band Slotted Antenna at 2.4/5.8GHz for ISM Band Applications. IEEE Symposium on Wireless Technology & Application (ISWTA), pp. 67-71. doi: 10.1109/ISWTA62130.2024.10651642.

[12] Ibrahim MS. (2019). Low-Cost, Circularly Polarized and Wideband U-Slot Microstrip Patch Antenna with Parasitic Element for WiGig and WPAN applications. 13th European Conference on Antennas and Propagation (EuCAP), pp. 1-4.

[13] Rahimi AH, Mohamad SY, Malek NFA, Islam MR, Midi NS, Shuhaimi NI. (2024). A Flexible Wideband Microstrip Antenna on TPU Substrate Using Inset Slot Feed and Partial Ground Plane. IEEE Symposium on Wireless Technology & Application (ISWTA), pp. 29-33. doi: 10.1109/ISWTA62130.2024.10651864.

[14] Loss C, Silveira TM, Pinho P, Salvado R, de Carvalho NB. (2020). Design and Analysis of the Reproducibility of Wearable Textile Antennas. 12th International Symposium on Communication Systems, Network and Digital Signal Processing (CSNDSP), pp. 1-5. doi: 10.1109/CSNDSP49049.2020.9249634.

[15] Taqdeer MM, Amjad QM, Zahid M, Amin Y. (2023). 2×2 Hexagonal-Shaped Antenna Array for 5.8 GHz ISM Band Applications. 7th International Multi-Topic ICT Conference (IMTIC), pp. 1-4. doi: 10.1109/IMTIC58887.2023.10178485.

[16] Sawant VG, Kadam P, Mangrulkar V, Gawade H, Patil Y. (2023). 5.8GHz ISM Band Low Cost Button Antenna for Smart Wearable Application. 6th International Conference on Advances in Science and Technology (ICAST), pp. 638-642. doi: 10.1109/ICAST59062.2023.10454967.

[17] Gupta K, Jain K, Singh P. (2021). Analysis and Design of Circular Microstrip Patch Antenna At 5.8 GHz. International Journal of Computer Science and Information Technologies, vol. 5, pp. 3895-3898. doi: 10.1088/1742-6596/1804/1/012200.

[18] Valjibhai GJ, Bhatia D. (2023). Design of 4×1 Microstrip Patch Antenna Array for 5.8 GHz ISM Band Applications. International Conference on Communication and Electronics System Design. *https://doi.org/10.1117/12.2012464.*

[19] Salelkar SS, Kerkar P. (2019). 5.8 GHz Semi Slotted Patch Antennas for ISM Band Applications. International Research Journal of Engineering and Technology (IRJET), vol. 6(4), pp. 3791-3794.

[20] Kumar VR, Anushree T, Aarthi S, Sanju A, Aishwariya VU, Kumar VP. (2021). Design and Analysis of Wearable Textile Jean Antenna for ISM Band Applications. Turkish Online Journal of Qualitative Inquiry (TOJQI), vol. 12(5), pp. 2442-2447.

[21] Saeidi T, Mahmood SN, Alani S, Ali SM, Ismail I, Alhawari ARH. (2020). Triple-Band Transparent Flexible Antenna for ISM Band and 5G Applications. IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 1-6. doi: 10.1109/BlackSeaCom48709.2020.9235009.

https://doi.org/10.31436/iiumej.v26i2.3488

COMPARATIVE ANALYSIS OF VISION TRANSFORMERS AND CNN MODELS FOR DRIVER FATIGUE CLASSIFICATION

FADHLAN HAFIZHELMI KAMARU ZAMAN*, NG KOK MUN, SYAHRUL AFZAL CHE ABDULLAH

Vehicle Intelligence and Telematics Lab, Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor

**Corresponding author: fadhlan@uitm.edu.my*

(Received: 30 October 2024; Accepted: 28 January 2025; Published online: 15 May 2025)

ABSTRACT: This study provides a comprehensive evaluation of Convolutional Neural Network (CNN) and Vision Transformer (ViT) models for driver fatigue classification, a critical issue in road safety. Using a custom driving behavior dataset, state-of-the-art CNN and ViT architectures, including VGG16, EfficientNet, MobileNet, Inception, DenseNet, ResNet, ViT, and Swin Transformer, were analyzed in this study to determine the best model for practical driver fatigue monitoring systems. Performance metrics such as accuracy, F1-score, training time, inference time, and frames per second (fps) were assessed across different hardware platforms, including a high-performance workstation, Raspberry Pi 5, and a desktop with a Graphic Processing Unit (GPU). Results demonstrate that CNN models, particularly VGG16, achieve the best balance between accuracy and efficiency, with an F1-score of 0.97 and 77.00 fps on a desktop. On the other hand, Swin V2S outperforms all models in terms of accuracy, achieving an F1-score of 0.99 and 61.18 fps on a GPU, although it exhibits limited efficiency on embedded systems. This study significantly contributes by providing practical recommendations for selecting models based on performance needs and hardware constraints, highlighting the suitability of ViTs for high-computation environments. The findings support the development of more efficient driver fatigue monitoring systems, offering practical implications for enhancing road safety and reducing traffic accidents.

ABSTRAK: Kajian ini merupakan penilaian komprehensif terhadap model Konvolusi Rangkaian Neural (CNN) dan Transformer Penglihatan (ViT) bagi pengelasan keletihan pemandu, iaitu satu isu kritikal dalam keselamatan jalan raya. Menggunakan set data tingkah laku pemanduan tersuai, seni bina terkini CNN dan ViT, termasuk VGG16, EfficientNet, MobileNet, Inception, DenseNet, ResNet, ViT dan Transformer Swin dianalisa dalam kajian ini bagi menentukan model terbaik bagi sistem pemantauan keletihan pemandu yang praktikal. Metrik prestasi seperti ketepatan, skor F1, masa latihan, masa inferens, dan bingkai sesaat (fps) telah dinilai merentasi pelbagai platfom perkakasan, termasuk stesen kerja berprestasi tinggi, Raspberry Pi 5, dan komputer meja dengan Unit Pemprosesan Grafik (GPU). Dapatan kajian menunjukkan bahawa model CNN, khususnya VGG16, mencapai keseimbangan terbaik antara ketepatan dan kecekapan, dengan skor F1 sebanyak 0.97 dan 77.00 fps pada komputer meja. Sebaliknya, Swin V2S mengatasi semua model dari segi ketepatan, mencapai skor F1 sebanyak 0.99 dan 61.18 fps pada GPU, walaupun menunjukkan kecekapan yang terhad pada sistem terbenam. Kajian ini memberikan sumbangan yang signifikan dengan menyediakan cadangan praktikal bagi pemilihan model berdasarkan keperluan prestasi dan kekangan perkakasan, serta menonjolkan kesesuaian ViT bagi persekitaran berkomputasi tinggi. Penemuan ini menyokong pembangunan sistem pemantauan keletihan pemandu yang lebih cekap, dengan implikasi praktikal bagi meningkatkan keselamatan jalan raya dan mengurangkan kemalangan.

https://doi.org/10.31436/iiumej.v26i2.3488

KEYWORDS: *Deep learning, CNN, vision transformer, driving behavior, embedded system, Raspberry Pi*

1. INTRODUCTION

Traffic accidents have claimed the lives of an estimated 1.35 million people annually, or 3,700 people per day. Road accident victims, their families, and entire nations suffer significant financial losses in the event of traffic accidents. In Malaysia, the government has lost at least RM3.12 million for every life, according to MIROS's 2018 Value of Statistical Life (VSOL). On average, Malaysia has 18 fatal road accidents every day, making it a serious public health issue for the nation [1].

Improving vehicle driving safety has become a priority in academia and the automotive industry to lower the risk of vehicle accidents. Numerous elements affect vehicle driving safety, but one of the most important is the driver's health and condition. The driver's emotions, psychology, and physiology all impact this condition. The two most important negative indicators of the driver's state are fatigue and distraction. According to studies, 36% of highway fatalities are caused by exhaustion and distraction [2]. There is a negative impact of fatigue and sleep-related safety risks on driving performance [3].

As a result of the rapid advancement of technologies like machine vision and deep learning [4], it is now a popular research area to use images and videos to identify driver fatigue and attention states to reduce the risk of vehicle accidents. One of the deep learning algorithms that is commonly used for image and video recognition is the Convolutional Neural Network (CNN) [5], [6]. It uses convolutional layers to automatically and adaptively learn the spatial hierarchies of features in grid-like data, such as images and photographs. Convolutional, pooling, and fully connected layers are some of the layers that CNNs use to identify patterns and structures in visual input. Another popular deep learning architecture intended for image classification is called a Vision Transformer (ViT) [7], [8]. ViTs employ transformer models, which were first created for natural language processing, as opposed to Convolutional Neural Networks (CNNs), which employ convolutional layers. To capture global context and dependencies, they split an image into fixed-size patches, insert them into sequences, and process them using self-attention techniques. Because of this, Vision Transformer can perform image classification tasks with excellent accuracy.

There are various state-of-the-art architectures for CNN, such as ResNet, VGGNet, DenseNet, InceptionNet, MobileNet, and EfficientNet. At the same time, Vision Transformer has popular architectures such as ViT and Swin Transformer. In numerous previous studies, CNN and ViT have been used recently to classify driving behaviors. Poon et al. developed a non-contact driving behaviour detection system for the improvement of driving safety using a YOLO-based CNN architecture [9] whereas ResNet50 has been shown to produce a very good performance in driving behavior classification for night driving [10]. On the other hand, Lian et al. proposed Stargazer, a straightforward yet powerful action temporal localization framework that utilises rich temporal aspects about human behavioural data. Stargazer is an efficient ViT-based system that has produced good performance and shown the efficacy of our model on the Naturalistic Driver Action Recognition of the AI City Challenge 2022 [11].

Several studies have benchmarked CNNs over ViT to evaluate their performance in many classification tasks, including the recognition of Japanese Sign Language [12], detection of bacterial strains [13], detection of distracted drivers [14], wheat disease classification [15], and COVID-19 classification [16]. These studies found CNNs outperforming ViTs only in COVID-19 classification, while ViTs consistently excelled in other classification tasks. However, the

computational complexity of ViTs in real-world applications remains unexamined, and their comparative performance with CNNs in classifying driver fatigue behaviors is still unclear. It is also not yet clear whether ViT or CNN will perform better against the other in classifying fatigue driving behaviours. Thus, this paper provides a comprehensive overview and performance analysis of several state-of-the-art CNN architectures and Vision Transformer models for classifying driver fatigue behaviours, such as nodding and yawning. We utilize our driving behavior dataset for benchmarking. The architectures are compared using standard evaluation metrics, including classification accuracy, precision, recall, F1 scores, and model training and inference time on a desktop with a GPU and a Raspberry Pi 5, followed by a detailed analysis and discussion.

The paper is structured as follows: Section II provides an overview of CNNs and Vision Transformers and the models this study considers. Section III reviews previous approaches for detecting distracted driving behaviours while Section IV details the experimental procedures, evaluation metrics, and dataset used. Section V presents our experimental results and analysis. Finally, Section VI offers conclusions.

2. OVERVIEW OF CNN AND ViT ARCHITECTURES

CNNs and ViTs, two well-known deep learning architectures utilised in image classification problems, are examined in this section. Effective feature extraction and visual data classification are made possible by these models' unique structures and processing methods. CNNs are excellent at identifying spatial relationships in images because of their layered approach of convolutional and pooling operations. In contrast, ViTs use transformer-based self-attention techniques, which were first created for natural language processing, to capture global context and long-range dependencies in an image.

2.1. Convolutional Neural Network Architectures

CNNs are deep learning models designed to analyze images and videos. Convolutional layers, activation functions, pooling layers, fully connected layers, and an output layer are among the basic layers that make up its architecture. The convolutional layers apply filters to the input image to extract features like edges, textures, and patterns. The model can learn intricate patterns thanks to the non-linearity introduced by activation functions like ReLU. By downsampling the feature maps, pooling layers minimise computational cost and spatial dimensions while avoiding overfitting. The output layer usually employs a softmax function to calculate probabilities for each class. In contrast, fully connected layers, positioned at the network's end, convert the high-level characteristics into classification results.

A CNN uses several convolutional and pooling layers to extract hierarchical features for image classification, starting with low-level edges and textures and working up to high-level shapes and objects. By modifying the weights of its filters and fully connected layers, the model is trained on a labeled dataset and uses backpropagation to minimize the error between its predictions and actual labels. Once trained, the CNN recognises learnt patterns to classify new images accurately. CNNs are an invaluable tool in computer vision applications because of their ability to effectively analyse visual data and recognise spatial dependencies, which makes them highly successful for image classification. In this work, ResNet [17], EfficientNet [18], Inception [19], VGGNet[20], DenseNet [21], and MobileNet [22] are selected as the CNN architectures to benchmark against Vision Transformer.

- **ResNet (Residual Network)**: ResNet tackles the issue of vanishing gradients in deep networks by introducing the idea of residual learning. To enable the network to learn

https://doi.org/10.31436/iiumej.v26i2.3488

identity mappings, it uses shortcut connections that bypass one or more levels. By promoting gradient flow during backpropagation, this architecture makes it possible to build incredibly deep networks like ResNet-50 and ResNet-101, which achieve great performance.

- **EfficientNet**: EfficientNet is intended to provide excellent accuracy using fewer parameters and lower processing expenses. It uses compound scaling to balance network depth, width, and resolution to optimize speed. Neural Architecture Search (NAS) was used to create the baseline model, EfficientNet-B0, to identify an efficient structure. Due to its Mobile Inverted Bottleneck Convolution (MBConv) blocks and Swish activation function, EfficientNet achieves state-of-the-art accuracy in image classification tasks while being computationally efficient.
- **Inception**: Using the Inception architecture, several convolutional filters of varying widths are applied within a single layer using Inception modules. This lessens the computational load and enables the network to capture various spatial features. These modules are stacked in the design to provide reliable performance and effective learning. Through significant optimisations, Inception-v3 and subsequent versions further increase accuracy and efficiency.
- **VGGNet**: The Visual Geometry Group (VGG) created VGGNet, renowned for its depth and simplicity. Small (3×3) convolutional filters are used throughout the network, and additional layers are stacked to boost depth, as in VGG-16 and VGG-19. This method increases classification performance, but because there are more parameters, it requires more memory and processing power.
- **MobileNet**: MobileNet prioritises efficiency and low computing costs in its design of mobile and embedded vision applications. It uses depth-wise separable convolutions, drastically reducing the number of parameters and calculations by dividing normal convolution into depth-wise and pointwise convolutions. MobileNet is appropriate for resource-constrained environments, with variations like MobileNetV2 and MobileNetV3, which further optimise performance and efficiency.

2.2. Vision Transformer Architectures

Vision Transformer is a new deep learning model designed for image classification that utilizes the transformer architecture originally developed for language processing. To maintain spatial information, it divides an image into fixed-size patches, embeds them into vectors, and subsequently incorporates positional encodings. The output is then utilized by the classification head to predict image classifications. ViTs excel in picture classification tasks, particularly with large datasets, due to their ability to capture global context and long-range dependencies. To benchmark CNNs against Vision Transformer, ViT [8] and Swin Transformer [23] architectures and their variants are selected.

- **Vision Transformer (ViT)**: The Vision Transformer (ViT) is a groundbreaking architecture that employs the transformer model for image classification. It flattens and linearly embeds the fixed-size patches it creates from an input image. To preserve spatial information, these patch embeddings are augmented with positional encodings. A conventional transformer encoder receives the embedded patches and employs multi-head self-attention to capture relationships throughout the image. ViT has demonstrated competitive performance on extensive image classification tasks and is excellent at capturing global context.

https://doi.org/10.31436/iiumej.v26i2.3488

- **Swin Transformer**: The Swin Transformer (Shifted Window Transformer) is an improved vision transformer architecture that aims to increase efficiency and scalability. It presents a hierarchical structure that uses a shifted window method to handle images. This approach splits the image into non-overlapping windows to reduce computational complexity and calculates self-attention within each window. The windows are moved between layers to facilitate cross-window connections and allow the model to gather local and global information efficiently. Swin Transformer is particularly effective for various vision applications, including segmentation, object detection, and image classification.

3. RELATED WORKS

3.1. CNN-Based Behavior Classification Approaches

Li et al. introduced the Deep Multichannel Network Model (DMNM) framework [24], which processes real driving data divided into external, internal, and habit dimensions using fully connected layers and CNNs. The model achieves high classification accuracy (95%) by filtering correlations between factors and integrating multidimensional data. DMNM includes a comprehensive analysis and effective correlation elimination. However, it also presents challenges like data processing and model training complexity, potential overfitting due to high feature dimensionality, and significant computational resource requirements. Multiple CNN models have also been used to detect driving behaviors under different lighting conditions. The model uses MobileNetV2, GoogleNet, InceptionV3, and ResNet-50 to analyze various driving patterns, distinguishing between day and night driving behaviors. The data for training the models includes video and sensor inputs collected from actual driving scenarios [25].

A 3D CNN called DriftNet has also been used in detecting aggressive driving behaviors, particularly car drifting, from video data [26]. The implementation employs DenseNet architecture to effectively learn spatial and temporal features from traffic videos. The model was trained on a custom dataset of car drifting clips collected from YouTube, achieving a validation accuracy of 77.5%. Among its advantages, DriftNet demonstrates high accuracy and effectiveness in learning complex features, utilizes transfer learning to enhance performance on limited data, and achieves superior validation accuracy compared to other models tested.

Qu et al. developed HAR-Net, a deep learning model to detect dangerous driving behaviors such as eating, drinking, smoking, and phone use [27]. The model processes optical flow, RGB, and RGBD data separately, integrating them through spatial-temporal fusion. It combines ResNet-50 with an hourglass network and incorporates an attention mechanism, achieving a mean average precision (mAP) of 98.84% on the constructed dataset [27]. Octave-Like Convolutional Neural Network (OLCMNet) [28] has been proposed for detecting driver distraction, where its main implementation focuses on incorporating a high-frequency branch and two low-frequency branches in the OLCM block, enhancing the network's ability to capture diverse feature information. The results demonstrate that the OLCMNet achieves better overall accuracy than other networks, particularly on the StateFarm and LDDB datasets. The advantages of the OLCMNet include its superior accuracy and effective feature capture. At the same time, its disadvantages are the increased computational cost and the need for careful tuning of hyperparameters to balance accuracy and speed.

3.2. Vision Transformer-Based Behavior Classification Approaches

Liang et al. introduced a system that uses a transformer-based approach to detect driver actions from in-vehicle video footage [11]. Stargazer leverages an improved multi-scale vision

https://doi.org/10.31436/iiumej.v26i2.3488

transformer (MViT) network to learn hierarchical representations of driver actions. It employs a sliding-window classification strategy for accurate temporal localization. The work shows that re-training on large-scale video action datasets and multi-crop data augmentation enhances training efficiency and model robustness. The advantages of this method include high accuracy in detecting and localizing driver actions and effective temporal feature capture. However, the system has high computational requirements and faces challenges in real-time deployment on vehicle onboard devices.

Similarly, a hierarchical vision transformer with shifted windows has been proposed to detect distracted driving behaviors [29]. This model leverages the Swin Transformer architecture, fine-tuned with pre-trained weights on the ImageNet dataset, to classify distracted drivers accurately. The implementation includes a detailed analysis of the model's performance on the AUC Distracted Driver Dataset, achieving a classification accuracy of 95.72%. This approach achieved high accuracy in distracted driver detection and has efficient computation due to limited self-attention calculations within local windows.

Besides, ViT and CNN have been used in a hybrid approach that combines the strengths of both methods to detect driver distraction [30]. The proposed model, FPT, integrates the Twins Transformer framework with redesigned residual embedding and lightweight group convolution modules, reducing computational complexity and improving feature extraction capabilities. This hybrid approach enhances the model's ability to capture local and global features, which is crucial for accurately identifying distracted driving behaviors. The implementation includes a cross-entropy loss function with label smoothing to enhance model learning and prevent overfitting. The model's effectiveness is demonstrated on two large-scale driver distraction datasets, achieving high accuracy and stability. However, the complexity of the transformer architecture and the need for substantial computational resources are noted as disadvantages, requiring further optimization for practical deployment.

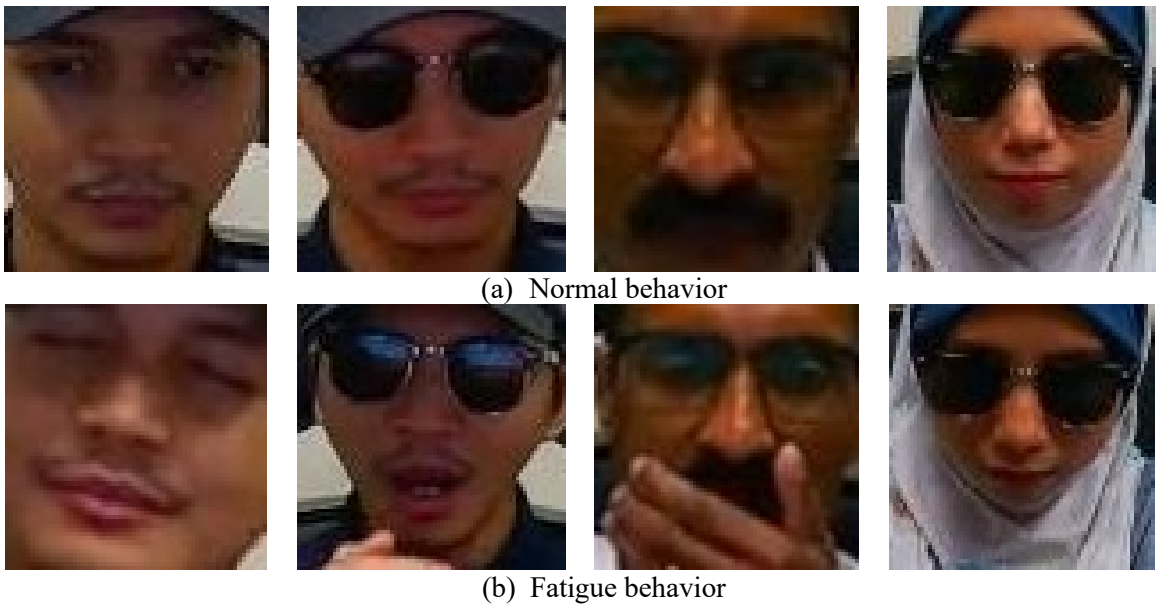
4. EXPERIMENTS

This section provides details of the data preparation and experimental setup used to evaluate CNNs and ViTs. Evaluation is conducted during model development and deployment to assess the models' performance.

4.1. Data Preparation

To enable the system to classify fatigue behaviors, the model was trained using two classes of driving behaviors: normal and fatigue. Figure 1 presents examples from each behavior class. This targeted classification approach allows the model to focus on specific behaviors, enhancing its learning efficiency. The datasets, encompassing normal, yawning, and drowsiness behaviors, were collected in the VITAL Laboratory, Universiti Teknologi MARA, where 44 participants engaged in a driving simulation designed to mimic real-life conditions. Each participant simulated scenarios aligned with the predefined categories of normal driving, yawning, and drowsiness. An essential consideration during data collection was ensuring data diversity. Given that the models are intended for real-time applications, diverse training data is critical to enable the model to adapt to various real-world scenarios. The diversity was achieved by including different individuals as drivers and varying the accessories worn, such as hats and spectacles.

https://doi.org/10.31436/iiumej.v26i2.3488



(a) Normal behavior

(b) Fatigue behavior

Figure 1. Example of images used in the dataset, showing the behaviors (a) normal behavior and (b) fatigue behavior.

Before training and validation, the data underwent pre-processing steps, including face cropping, normalization, and augmentation techniques to ensure a more robust model could be obtained. The image augmentations employed here are image scaling, random zoom, width and height shift, horizontal flip, and random brightness. CNN models are trained using images of 100x100 dimensions, whereas Vision Transformer models are trained using 224×224 images as required by the models. A manual annotation method accurately labeled the processed images into normal, yawning, and drowsiness classes. This approach ensured that the model was provided with high-quality training data for recognizing fatigue-related driving behaviors. In total, 43,034 images were used for training and validation, whereas 3,599 images were used for testing, with distinct individuals (non-overlapping) for each dataset to ensure unbiased evaluation. During training, all layers in all pre-trained models are unfrozen and optimized.

4.2. Experimental Setup

This work uses three separate computing platforms to evaluate the models under varying hardware configurations. The primary platform for training is a high-performance workstation equipped with an Intel® Core™ i7-14700KF 3.40 GHz processor, 32 GB of memory, and an NVIDIA RTX 4070 12GB GPU. This setup ensures efficient and accelerated training of deep learning models, capable of handling large datasets and complex computations essential for developing CNN and ViT models. For testing, a Raspberry Pi 5 features an ARM Cortex-A76 (ARM v8) 2.4 GHz processor, 8 GB of memory, and a VideoCore VII (800 MHz) GPU. This platform represents low-cost, embedded computing environments, allowing us to assess the models' performance on resource-constrained devices. Evaluating the Raspberry Pi 5 is critical for real-world applications in automotive systems, where such embedded systems are prevalent. Additionally, a desktop setup is used with an Intel® Core™ i7-13700 2.10 GHz processor, 16 GB of memory, and an NVIDIA RTX 4060 8GB GPU for further testing. This desktop configuration provides a middle ground in computing power between the workstation and the Raspberry Pi. It offers a more accessible and typical consumer-grade environment to assess the models' performance, which is relevant for broader applications. The specifications for each computing platform are given in Table 1.

https://doi.org/10.31436/iiumej.v26i2.3488

Table 1. Computing platforms used for training and testing

Specs.	**Computing Platforms**		
	Workstation (for model training)	***Raspberry Pi 5 (for model testing)***	***Desktop with GPU (for model testing)***
Processor	Intel® Core™ i7-14700KF 3.40 GHz	ARM Cortex-A76 (ARM v8) 2.4GHz	Intel® Core™ i7-13700 2.10 GHz
Memory	32 GB	8 GB	16GB
GPU	NVIDIA RTX 4070 12GB	VideoCore VII (800 MHz)	NVIDIA RTX 4060 8GB

During the training session, the time required to train the models is collected to provide valuable insights into their efficiency, practicality, and suitability for different applications. After that, the accuracy, precision, recall, and F1-score are evaluated along with the average inference time for all models. The equations for the accuracy, precision, recall, and F1-score are given in Eq. (1), Eq. (2), Eq. (3), and Eq. (4), respectively. Subsequently, the models are deployed on the Raspberry Pi 5 and a desktop computer with a GPU to evaluate their performance in classifying distracted driving behaviors. The models are run as part of the intelligent behavior detection program shown in Figure 2, which consists of face detection and eye detection, whereby the performance is evaluated in terms of frames per second (fps). The pseudocode of the intelligent behavior detection program, encompassing the functions and procedures required to assess the CNN and ViT models, is provided in Algorithm 1.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

where True positive (TP) is a correct classification of the fatigue behavior, True negative (TN) is a correct classification of the normal behavior, False positive (FP) is an incorrect classification of the normal behavior as fatigue behavior, and False negative (FN) is an incorrect classification of the fatigue behavior as normal behavior.

Figure 2. Visualization of the outputs from the intelligent behavior detection program used to evaluate the model performance.

Algorithm 1. The pseudocode for the intelligent behavior detection program

```
Start program

1: Import necessary AI, computer vision, and sensor interfacing libraries.
2: Define auxiliary functions:
    - 'eye_close_detector': Detects if the driver's eyes are closed based on eye aspect ratio (EAR).
    - 'post_process_behavior': Processes behavior status (e.g., distracted, drowsy) based on thresholds.
    - 'behavior_detector': Identifies driver behavior using a pre-trained model.
    - 'detect_face': Detects face and extracts facial region for further analysis.
    - 'get_frames': Extracts RGB and grayscale frames for processing.
    - 'load_model_and_labels': Loads AI model and label data for driver behavior classification.
3: Initialize video capture and other counters.
    - Load AI model using 'load_model_and_labels'
4: Start main loop to process video frames:
    - Capture a video frame and apply overlay text.
    - Get full, cropped, and grayscale frames using 'get_frames'.
5: Every 'face_detector_interval' frames:
    - Detect face using `detect_face`.
    - Detect eye status using 'eye_close_detector'.
6: Every 'behavior_detector_interval' seconds:
    - If a face is detected, classify driver behavior using 'behavior_detector'.
    - Process behavior status with 'post_process_behavior'.
7: Display information on the video frame, including fps, behavior, eye status, and behavior history.

End program
```

The overall experiment workflow adopted in this work to benchmark the CNNs and ViTs in classifying driver fatigue behaviors is given in Figure 3. The pre-trained models that are compared in the experiments are Vision Transformer models: ViT B16, ViT B32, ViT L16, ViT L32, Swin B, Swin S, Swin T, Swin V2B, Swin V2S, Swin V2T, and CNN models: VGG 16, VGG 19, ResNet-50, MobileNetV2, InceptionV3, MobileNetV3, DenseNet-121, EfficientNetB0, EfficientNetV2 B0, and ResNet-152. According to Figure 3, CNN and ViT pre-trained models are trained using only the training data from the dataset. After model training is completed, each model is tested using the test data, and the performance is evaluated using accuracy, precision, recall, F1-score, training time, and inference time metrics. Subsequently, each model is deployed into the intelligent behavior detection program on a desktop with a GPU and a Raspberry Pi 5. Here, the performance evaluated is the frames per second (fps).

5. RESULTS AND DISCUSSIONS

This section presents the results of the experiment, in which the performance of the CNN and Vision Transformer models is analyzed in terms of accuracy, precision, recall, and F1-score, as well as training time, inference time, and computation requirements on the desktop with GPU and Raspberry Pi 5, measured in frames per second.

5.1. Overall Accuracy and Performance

The results tabulated in Table 2 show the test performance of CNN and Vision Transformer models evaluated in this work. Vision Transformer models are trained using different numbers of embeddings according to the architecture requirements defined in Table

https://doi.org/10.31436/iiumej.v26i2.3488

2. The batch size and epochs are also set to match the models' memory requirements and provide the best test accuracy.

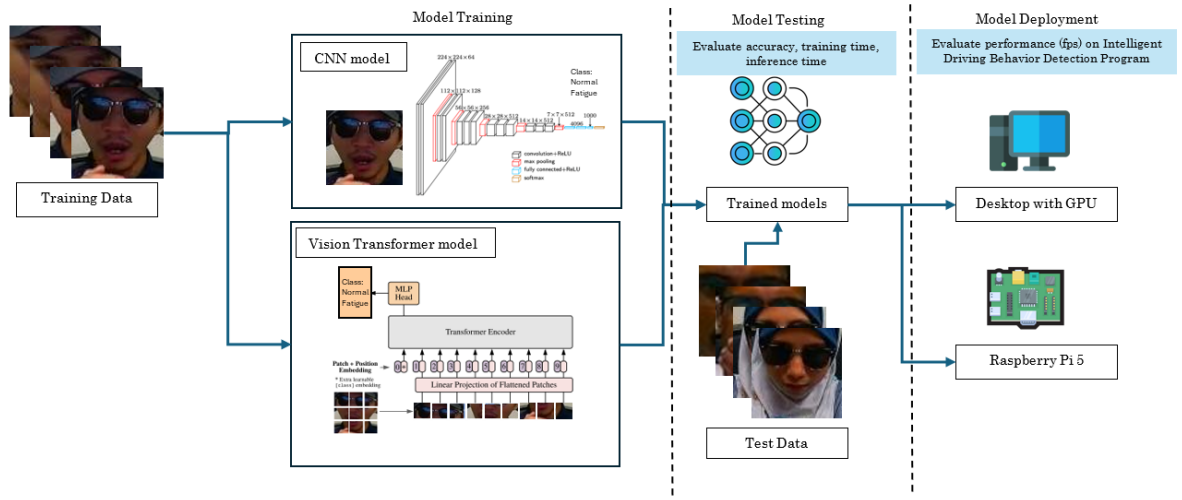


Figure 3. The overall experiment workflow adopted in this work is to benchmark the performance of CNN and ViT models in driver fatigue classification

Table 2. Overall performance comparisons

Models	Test Performance				# params	# emb.	batch size	epochs	RPi 5 (fps)	Desktop (fps)
	Acc.	Prec.	Recall	F1						
ViT B16	0.9686	0.9459	0.9674	0.9561	85,801,730	768	32	10	7.73	64.86
ViT B32	0.8413	0.7843	0.8700	0.8062	87,458,306	768	16	10	18.67	73.22
ViT L16	0.8032	0.7518	0.8398	0.7669	303,305,730	1024	16	10	0.73	51.27
ViT L32	0.9177	0.8664	0.9434	0.8942	306,537,450	1024	16	10	0.89	66.6
Swin B	0.9705	0.9451	0.9757	0.9592	86,745,274	1024	16	10	1.32	58.35
Swin S	0.9716	0.9490	0.9734	0.9604	48,838,796	768	16	10	9.64	58.17
Swin T	0.9472	0.9073	0.9584	0.9290	28,291,428	768	32	10	18.1	69.06
Swin V2B	0.9619	0.9305	0.9693	0.9478	86,907,898	1024	16	10	1.27	58.82
Swin V2S	0.9897	0.9915	0.9789	0.9850	48,969,980	768	16	10	4.05	61.18
Swin V2T	0.9505	0.9123	0.9610	0.9332	27,584,108	768	32	10	14.29	64.38
VGG16	0.9808	0.9793	0.9653	0.9721	14,715,714	-	256	20	37.81	77.52
VGG19	0.9763	0.9816	0.9506	0.9650	20,025,410	-	256	20	36.75	73.71
ResNet-50	0.9566	0.9321	0.9462	0.9389	23,591,810	-	256	20	36.5	65.03
MobileNet V2	0.9519	0.9553	0.9051	0.9272	2,260,546	-	256	20	40.19	80.13
Inception V3	0.9691	0.9696	0.9411	0.9544	21,806,882	-	256	20	40.68	69.76
MobileNet V3	0.9605	0.9717	0.9155	0.9400	2,998,274	-	256	20	40.11	79.69
DenseNet-121	0.9713	0.9779	0.9399	0.9573	7,039,554	-	256	20	38.81	67.72
EfficientNet B0	0.9686	0.9669	0.9421	0.9538	4,052,133	-	256	20	42.91	69.61
EfficientNet V2B0	0.9605	0.9722	0.9150	0.9399	5,921,874	-	256	20	42.25	68.59
ResNet-152	0.9416	0.9046	0.9382	0.9198	58,375,042	-	32	20	27.82	61.31

According to Table 2, among the CNN models, MobileNetV2 and MobileNetV3 demonstrate strong performance across all metrics, with particularly high accuracy and F1-scores. MobileNet V3 slightly outperforms MobileNet V2 in precision and recall. EfficientNet B0 and EfficientNet V2B0 also perform well, maintaining high accuracy and F1-scores. DenseNet121 exhibits robust performance similar to EfficientNet models, with balanced metrics. VGG16 and VGG19 show excellent performance, especially in precision and recall, resulting in high F1-scores. VGG16 delivers the best F1-score among all tested CNN models at 0.9721. ResNet-50 exhibits strong performance with balanced metrics, slightly lower than

https://doi.org/10.31436/iiumej.v26i2.3488

InceptionV3 and VGG models, but still robust. For the ViT models, Swin T, Swin S, Swin VSS, and Swin V2B perform well, particularly Swin V2S, which shows high precision and recall, translating to a high F1-score. ViT B16 shows competitive performance with high accuracy and F1-score, but slightly lower precision and recall than Swin models. ViT B32, ViT L16, and ViT L32 show varying performance, with ViT L32 lagging in recall and F1-score but maintaining decent accuracy and precision. Swin V2S gives the best F1-score among Vision Transformer models at 0.9850, and surpasses VGG16, which is the best-performing CNN model.

Based on Table 2, comparing the two types of models, CNN models like VGG16, VGG19, and Inception V3 exhibit high accuracy. ViT models such as Swin S, Swin V2S, and ViT B16 also show high accuracy but with some variability among variants. Regarding precision, ViT models, particularly Swin V2S, show higher precision than many CNN models. CNN models generally exhibit strong precision, with VGG and Inception models leading. Regarding recall, ViT models such as Swin V2S, Swin S, and ViT B16 exhibit higher recall, indicating better performance in identifying all relevant instances. CNN models like VGG19 and InceptionV3 also demonstrate high recall.

5.2. Model Training and Model Inference Performance

As shown in Table 2, ViT models also tend to have a much higher number of parameters. ViT L32 has over 303 million parameters, significantly higher than the 2.9 million parameters in MobileNetV2. This higher parameter count can contribute to the increased training time and potentially higher model complexity, impacting the inference time. Thus, every model's training and inference time is measured and shown in Figure 4 and Figure 5, respectively.

According to Figure 4, ViT models generally require significantly longer training times than CNN models. For instance, ViT L16 has the longest training time at 17087.90 seconds, whereas MobileNetV2 has the shortest at 1682.62 seconds. Even CNN ResNet-152's longest model training time, at 2698.13 seconds, is almost two times shorter than the fastest model training of Vision Transformer ViT B32, at 4688.91 seconds. This indicates that Vision Transformers are more computationally intensive and require more computing resources for training.

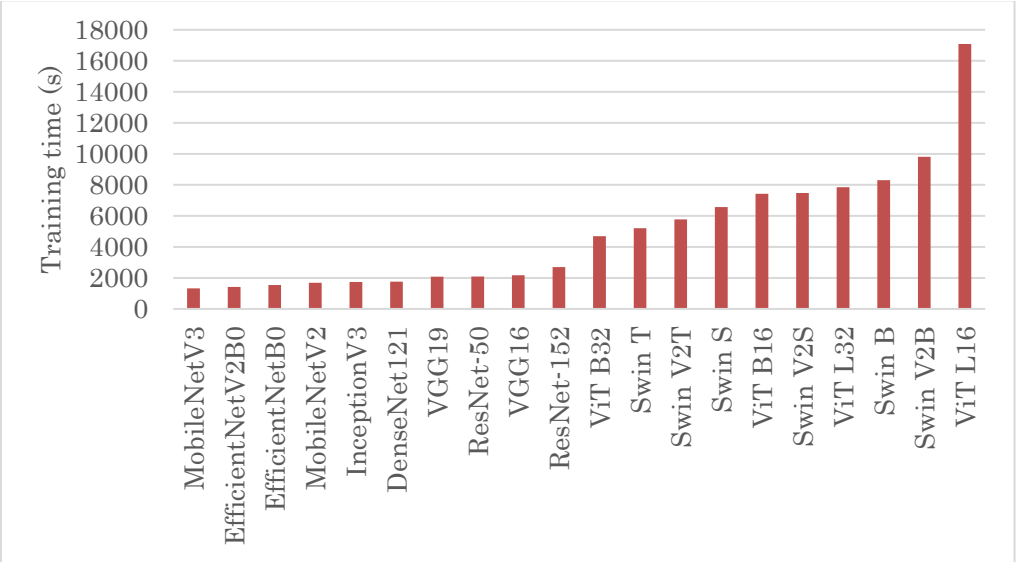


Figure 4. The training time required for each model, run on the workstation

https://doi.org/10.31436/iiumej.v26i2.3488

Furthermore, the inference times for these models are measured. The time taken to process the image and apply the models is considered in the measurement of inference time. Surprisingly, based on Figure 5, Vision Transformers' inference times are shorter than CNNs. For example, ViT L32 has an average inference time of 0.0055 seconds, whereas MobileNet V2 has an inference time of 0.0731 seconds, even though the ViT L32 model has higher complexity. To ensure that the inference performance on the powerful workstation used for training is translated into the actual application implementation, each model is tested in an intelligent driving behavior detection program, as mentioned earlier. This comparison is crucial for understanding how these models perform in different hardware environments, particularly for real-time applications like driving behavior monitoring systems.

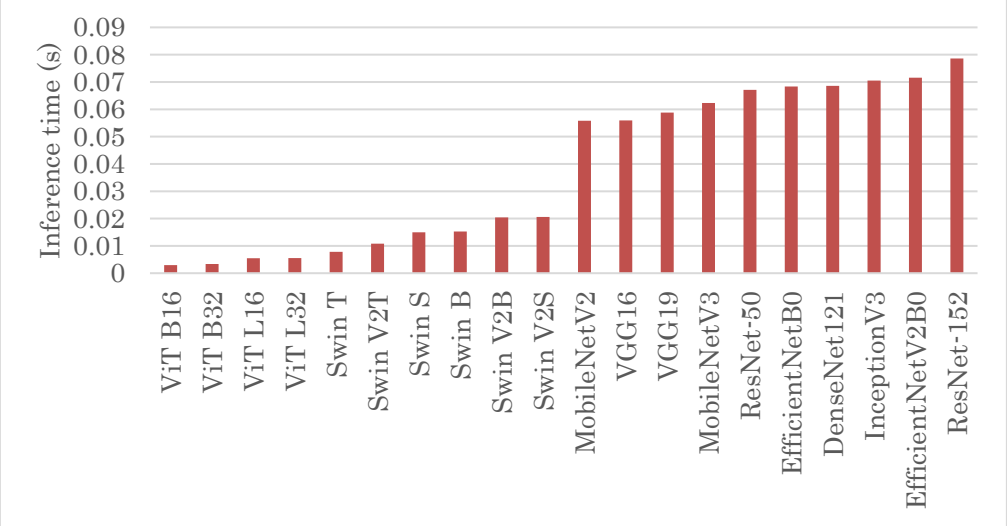
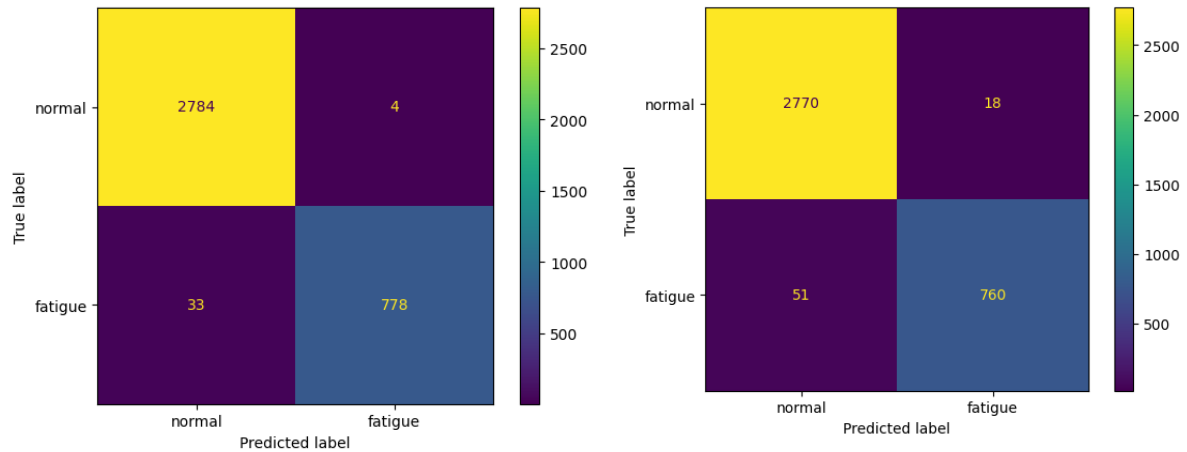


Figure 5. The inference time required for each model, run on the workstation

5.3. Swin V2S vs. VGG-16 Classification Results Analysis

Based on the results tabulated in Table 2, the top two best CNN and ViT models according to the F1 Score are the VGG-16 and Swin V2S models, with 0.9721 and 0.9850 F1 Scores, respectively. In this section, the results of the two models are compared through an analysis of the confusion matrix, as presented in Figure 6.



(a) Swin V2S (b) VGG-16

Figure 6. The confusion matrix from the test dataset classification using the Top-2 classifiers: (a) Swin V2S and (b) VGG-16

https://doi.org/10.31436/iiumej.v26i2.3488

From Figure 6, Swin V2S demonstrates high accuracy, particularly for the "normal" class, with a low number of false positives (4) compared to VGG-16 (18). Its ability to correctly identify both classes, especially "fatigue," is evident, with only a small number of fatigue cases misclassified as normal cases (33) compared to (51) by the VGG-16 model. The VGG-16 model shows a slightly higher rate of errors in both false positives and false negatives compared to Swin V2S. Specifically, it has slightly more difficulty differentiating between "normal" and "fatigue" states, leading to greater misclassification in both cases. It clearly shows that the Swin V2S model consistently outperforms VGG-16 in distinguishing between normal and fatigue driving conditions, showing fewer errors across both classes. Figure 7 shows several examples of misclassifications made by the models.

Swin V2S performs superiorly over VGG16 in driver fatigue classification due to its hierarchical architecture with shifted windows, which enables efficient capture of both local details and global context. This ability to integrate fine-grained features with broader patterns is enhanced by its self-attention mechanism within non-overlapping windows, making distinguishing between fatigue and normal states robust. While VGG16 offers a simpler sequential architecture with small filters, it cannot handle long-range dependencies of the relationship between spatially distant regions of an image, a strength of Swin V2S.

According to Figure 7, most misclassifications are due to the apparent similarity of some fatigue behavior images with normal behavior images. This similarity challenges the model by having shared visual characteristics and low inter-class variability. Additionally, since the data collection requires the participants to imitate fatigue behavior, the imitations may not be as realistic as the actual behaviors, i.e., when the drivers are in a fatigue or drowsy condition while driving. This has contributed to several misclassifications by the models.

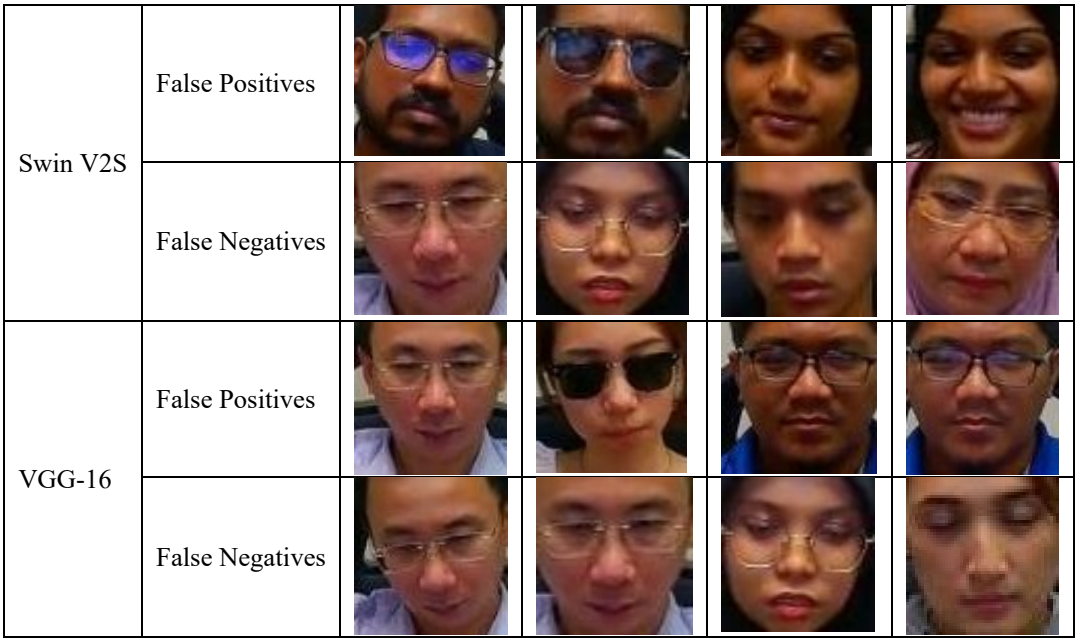


Figure 7. Examples of several misclassifications by the Swin V2S and VGG-16 models. False Positives refer to the incorrect classification of normal behavior as fatigue behavior, whereas False Negatives refer to the incorrect classification of fatigue behavior as normal behavior

https://doi.org/10.31436/iiumej.v26i2.3488

5.4. Model Performance in an Intelligent Driving Behavior Detection Program

Each of the CNN and ViT models trained earlier is loaded and deployed in the intelligent driving behavior detection program to evaluate the inference speed of each model on different computing platforms. The platforms used in the experiment are a Raspberry Pi 5 and a desktop computer with a GPU. The inference performance is shown in Figure 8.

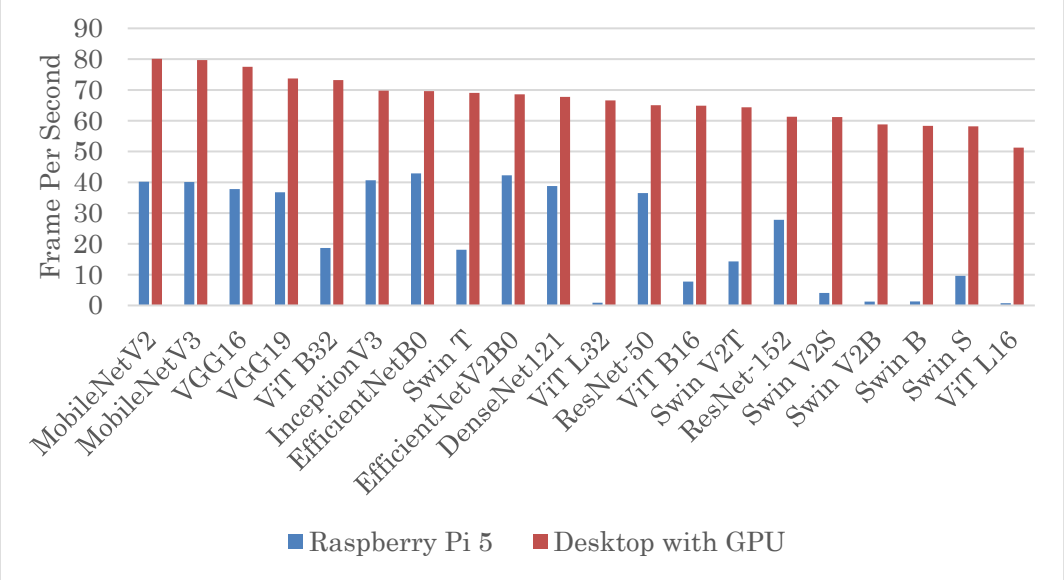


Figure 8. The comparison of processing time measured in fps for each model implemented in the intelligent driving behavior detection program.

According to Figure 8, the inference performance on the desktop with a GPU shows a clear advantage across all models, achieving significantly higher fps than the Raspberry Pi 5. MobileNet models demonstrate exceptional performance on the desktop, reaching close to 80 fps, indicating their efficiency and suitability for high-speed inference tasks. Other CNN models, such as VGG16, VGG19, and InceptionV3, also perform faster on the desktop, maintaining fps values around 70 and above. ResNet models achieve moderate fps values, around 60, indicating good performance but slightly less efficiency than the more streamlined models.

The inference performance on the Raspberry Pi 5 is notably lower, reflecting the constraints of running complex models on a less powerful, embedded platform. EfficientNetB0 and V2B0 still perform well, achieving more than 30 fps, which is respectable for such a device. Other CNN models perform relatively well, maintaining fps values of more than 25 fps. However, Vision Transformer models like ViT B16, ViT B32, and larger ViT variants such as ViT L16 and L32 exhibit very low fps on the Raspberry Pi 5, often below 10 fps, highlighting the significant computational demands of these models that are not well-suited for low-power embedded devices.

On the other hand, Vision Transformer models achieved almost equivalent fps performance to CNN models on a desktop with a GPU. With a GPU for computation, the fastest Vision Transformer model (ViT B32) can achieve higher fps than CNN InceptionV3. Even the slowest Vision Transformer model, ViT L16, achieved more than 50 fps, indicating that Vision Transformer models, while offering higher accuracy and better handling of complex patterns, demonstrate excellent fps in environments with high computational capabilities, such as desktops with powerful GPUs. To further illustrate the performance comparisons of CNN and Vision Transformer models in actual applications, the comparisons are given in Figs. 9 and 10.

https://doi.org/10.31436/iiumej.v26i2.3488

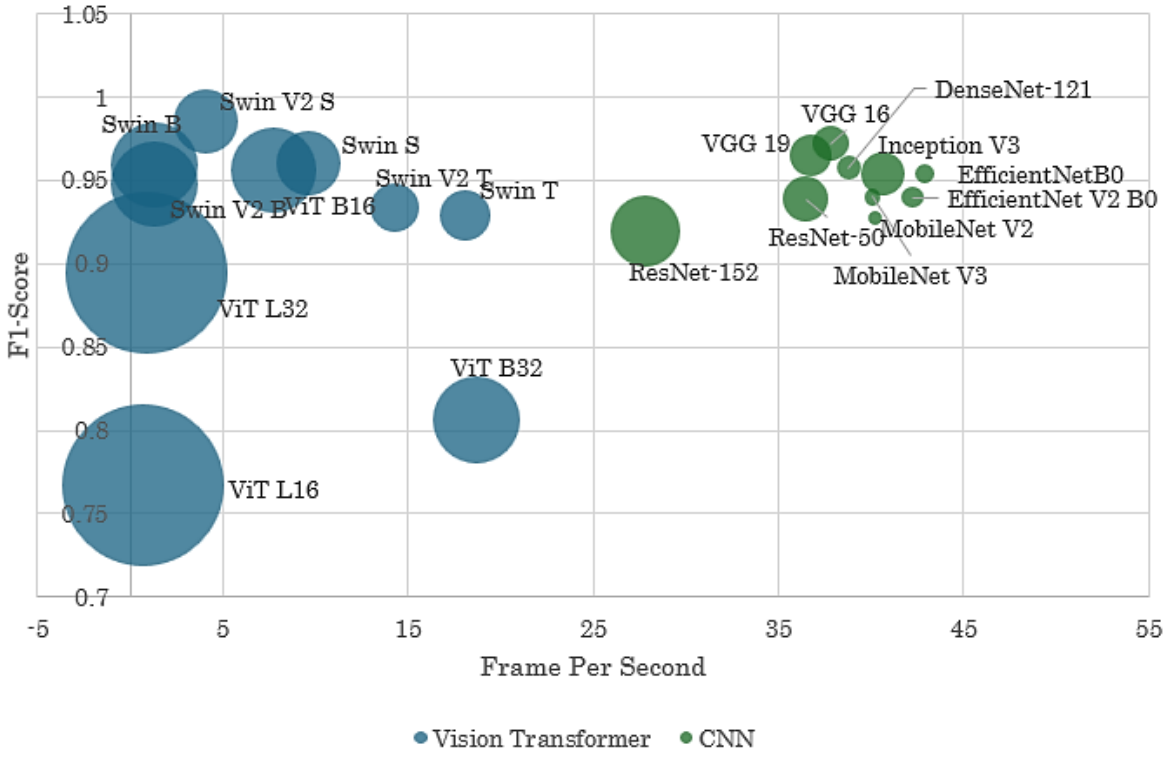


Figure 9. Performance comparison in terms of F1-score, fps, and number of parameters (defined by the bubble size) for each model run in the intelligent digital video recorder program running on Raspberry Pi 5.

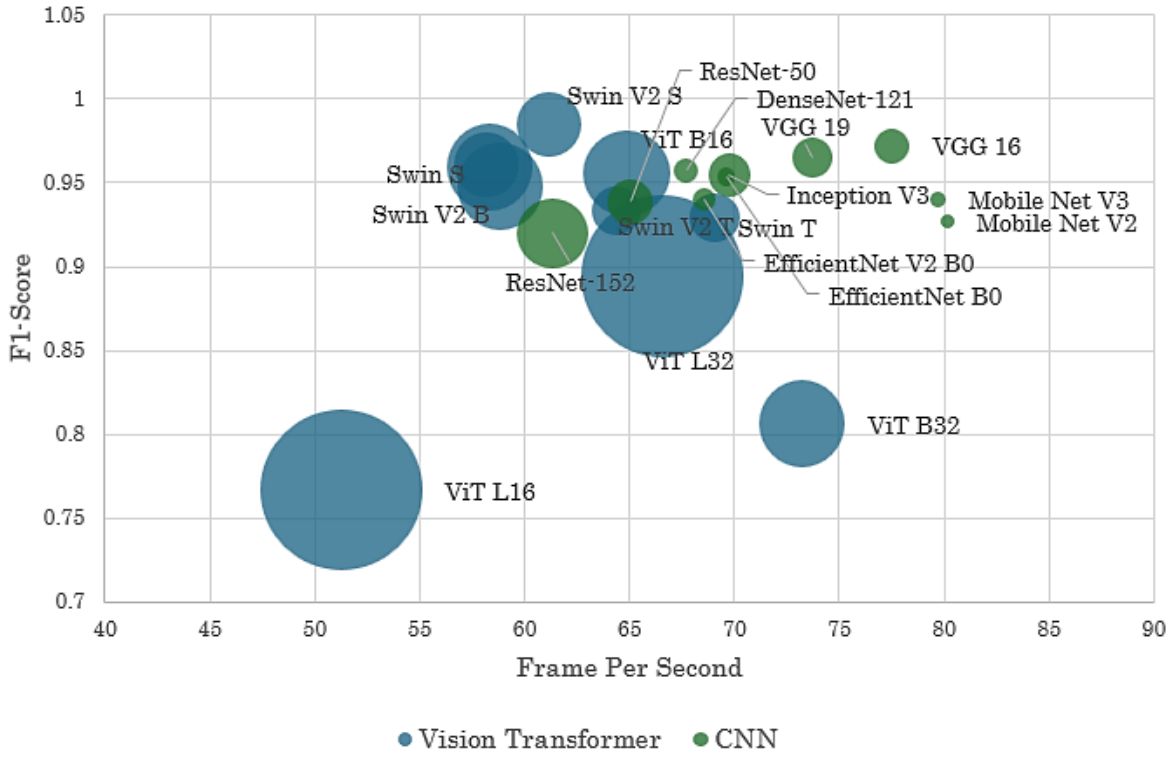


Figure 10. Performance comparison in terms of F1-score, fps, and number of parameters (defined by the bubble size) for each model run in the intelligent digital video recorder program running on the desktop with a GPU.

Based on Figure 9, all CNN models can deliver more than 25 fps performance on the Raspberry Pi 5. However, only ViT B32 and Swin V2T can provide acceptable fps on Raspberry Pi 5 at around 14-18 fps. The Vision Transformer model with the best F1-Score, Swin V2S, can only deliver 4 fps on the Raspberry Pi 5. Meanwhile, according to Figure 10, all models benefit from the enhanced computational power on the desktop with a GPU, but CNNs still slightly outperform Vision Transformer in inference speed.

VGG16 and VGG19 achieve more than 70 fps with high F1-scores of 0.97 and 0.95, respectively, indicating their efficiency and suitability for high-speed, real-time applications. MobileNetV2 and MobileNetV3 perform exceptionally well, maintaining fps around 80 with good F1-scores. In contrast, Vision Transformer models like ViT B16, ViT B32, and Swin variants achieve lower fps on GPU but still more than 45 fps, though Swin V2S stands out among Vision Transformers with relatively better performance, with 61 fps and 0.99 F1-score.

To summarize the findings, VGG16 is recommended as the best CNN model due to its superior balance of speed and accuracy across different hardware environments, making it ideal for both embedded and high-performance applications. Swin V2S is recommended as the best Vision Transformer model for its high accuracy and reasonable inference speed, especially in high-performance environments, despite its lower fps than CNN models. For deployment on computation machines with a GPU, Swin V2S will deliver the best performance and excellent computation speed, while VGG16 should be used if deployment is targeted at embedded platforms.

6. CONCLUSION

This paper investigates and compares CNN models against Vision Transformer (ViT) models in classifying fatigue driving behaviors. Both model types were evaluated on their accuracy and performance during training and inference on desktop and embedded platforms. CNN models consistently deliver high accuracy with balanced precision and recall, with VGGs, InceptionV3, and MobileNets showing exceptional performance across all metrics. ViT models, particularly the Swin variants, excel in precision and recall but require longer training and computational resources. CNN models like VGGs, InceptionV3, and MobileNets are reliable for balanced performance with high accuracy and efficient processing. In contrast, ViT models, especially Swin variants, are suitable for applications demanding higher precision and recall, provided robust computational resources are available. VGG16, with a 0.97 F1-score and 77 fps on GPU and 37.81 fps on Raspberry Pi 5, is the top CNN model for its balance of speed and accuracy. Swin V2S, achieving 0.99 F1-score and 61.18 fps on GPU, is the best ViT model, recommended for high-performance settings despite lower fps than CNNs. ViT models, though slower on embedded platforms, can surpass CNNs in accuracy in environments with powerful GPUs. Future work should explore hybrid models combining CNN and ViT strengths to enhance driving behavior classification further. Real-time testing on larger, more diverse datasets and advanced data augmentation and temporal context integration through RNNs or LSTMs could improve model robustness and accuracy. Optimizing models for deployment on low-power embedded devices while maintaining high accuracy is crucial for practical, real-world automotive applications.

ACKNOWLEDGEMENT

The authors extend their gratitude to the Ministry of Higher Education for their support through the Fundamental Research Grant Scheme FRGS/1/2023/TK07/UITM/02/23 and to the College of Engineering at Universiti Teknologi MARA for their assistance in this project.

REFERENCES

[1] Ministry of Transport Malaysia, "Malaysia Road Fatalities Index," Ministry of Transport Malaysia Official Portal. Accessed: Jun. 13, 2024. [Online]. Available: https://www.mot.gov.my/en/land/safety/malaysia-road-fatalities-index

[2] A. Fernández, "Facial attributes recognition using computer vision to detect drowsiness and distraction in drivers," Electronic Letters on Computer Vision and Image Analysis, vol. 16, no. 2, pp. 25–28, 2017, doi: 10.5565/rev/elcvia.1134.

[3] B. Shandhana Rashmi and S. Marisamynathan, "Factors affecting truck driver behavior on a road safety context: A critical systematic review of the evidence," Oct. 01, 2023, KeAi Communications Co. doi: 10.1016/j.jtte.2023.04.006.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[5] L. Cun et al., "Handwritten Digit Recognition with a Back-Propagation Network."

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

[7] A. Vaswani et al., "Attention Is All You Need," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[8] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.11929

[9] Y. S. Poon et al., "Driver Distracted Behavior Detection Technology with YOLO-Based Deep Learning Networks," in ISPCE-ASIA 2021 - IEEE International Symposium on Product Compliance Engineering-Asia, Proceeding, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ISPCE-ASIA53453.2021.9652435.

[10] M. F. Ishak, F. H. K. Zaman, N. K. Mun, S. A. C. Abdullah, and A. K. Makhtar, "Improving night driving behavior recognition with ResNet50," Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 3, pp. 1974–1988, 2024, doi: 10.11591/ijeecs.v33.i3.pp1974-1988.

[11] J. Liang, H. Zhu, E. Zhang, and J. Zhang, "Stargazer: A Transformer-based Driver Action Detection System for Intelligent Transportation," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society, 2022, pp. 3159–3166. doi: 10.1109/CVPRW56347.2022.00356.

[12] T. Kondo, S. Narumi, Z. He, D. Shin, and Y. Kang, "A Performance Comparison of Japanese Sign Language Recognition with ViT and CNN Using Angular Features," Applied Sciences (Switzerland), vol. 14, no. 8, Apr. 2024, doi: 10.3390/app14083228.

[13] H. E. Kim, M. E. Maros, T. Miethke, M. Kittel, F. Siegel, and T. Ganslandt, "Lightweight Visual Transformers Outperform Convolutional Neural Networks for Gram-Stained Image Classification: An Empirical Study," Biomedicines, vol. 11, no. 5, May 2023, doi: 10.3390/biomedicines11051333.

[14] H. V. Koay, J. H. Chuah, and C. O. Chow, "Convolutional Neural Network or Vision Transformer? Benchmarking Various Machine Learning Models for Distracted Driver Detection," in IEEE Region 10 Annual International Conference, Proceedings/TENCON, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 417–422. doi: 10.1109/TENCON54134.2021.9707341.

[15] S. Z. Ul Abidin, H. M. Lashari, and R. F. Ahmad, "ViT vs CNN: A Comparative Study of Wheat Disease Classification for Custom Data," in Proceedings - 2023 International Conference on Frontiers of Information Technology, FIT 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 274–279. doi: 10.1109/FIT60620.2023.00057.

[16] M. M. Sufian, E. G. Moung, J. A. Dargham, F. Yahya, and S. Omatu, "Pre-trained Deep Learning Models for COVID19 Classification: CNNs vs. Vision Transformer," in 4th IEEE International

Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/IICAIET55139.2022.9936852.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H

[18] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[19] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1409.1556S

[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Aug. 2016, [Online]. Available: http://arxiv.org/abs/1608.06993

[22] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2017arXiv170404861H

[23] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," arXiv e-prints, p. arXiv:2103.14030, Mar. 2021, doi: 10.48550/arXiv.2103.14030.

[24] D. Li, Y. Wang, and W. Xu, "A Deep Multichannel Network Model for Driving Behavior Risk Classification," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 1, pp. 1204–1219, Jan. 2023, doi: 10.1109/TITS.2022.3201378.

[25] M. F. Bin Ishak, F. H. K. Zaman, N. K. Mun, and A. K. Makhtar, "Day Driving and Night Driving Behavior Detection Using Deep Learning Models," in International Conference on ICT Convergence, IEEE Computer Society, 2023, pp. 463–468. doi: 10.1109/ICoICT58202.2023.10262745.

[26] A. Noor, B. Benjdira, A. Ammar, and A. Koubaa, "DriftNet: Aggressive Driving Behaviour Detection using 3D Convolutional Neural Networks," in Proceedings - 2020 1st International Conference of Smart Systems and Emerging Technologies, SMART-TECH 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 214–219. doi: 10.1109/SMART-TECH49988.2020.00056.

[27] Z. Qu, L. Cui, and X. Yang, "HAR-Net: An Hourglass Attention ResNet Network for Dangerous Driving Behavior Detection," Electronics (Switzerland), vol. 13, no. 6, Mar. 2024, doi: 10.3390/electronics13061019.

[28] P. Li et al., "Driver Distraction Detection Using Octave-Like Convolutional Neural Network," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 8823–8833, Jul. 2022, doi: 10.1109/TITS.2021.3086411.

[29] H. Vin Koay, J. Huang Chuah, and C. O. Chow, "Shifted-Window Hierarchical Vision Transformer for Distracted Driver Detection," in TENSYMP 2021 - 2021 IEEE Region 10 Symposium, Institute of Electrical and Electronics Engineers Inc., Aug. 2021. doi: 10.1109/TENSYMP52854.2021.9550995.

[30] H. Wang et al., "FPT: Fine-Grained Detection of Driver Distraction Based on the Feature Pyramid Vision Transformer," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 2, pp. 1594–1608, Feb. 2023, doi: 10.1109/TITS.2022.3219676.

https://doi.org/10.31436/iiumej.v26i2.3494

MODIFICATION OF GREY RELATIONAL ANALYSIS FOR DYNAMIC CRITERIA WEIGHTING IN DECISION-MAKING SYSTEMS

MUHAMMAD NAJIB DWI SATRIA[1], ERLIYAN REDY SUSANTO[1], SETIAWANSYAH[2*], SUFIATUL MARYANA[3], PRITASARI PALUPININGSIH[4]

[1]*Department of Information System, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia*
[2]*Department of Informatics, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia*
[3]*Department of Artificial Intelligence and Robotics, Universitas Pakuan, Bogor, Indonesia*
[4]*Department of Information System, Institut Teknologi Perusahaan Listrik Negara, Tanggerang, Indonesia*

**Corresponding author: setiawansyah@teknokrat.ac.id*

(Received: 2 November 2024; Accepted: 21 April 2025; Published online: 15 May 2025)

ABSTRACT: Grey relational analysis (GRA) is a grey system theory method used to solve multi-criteria decision problems with incomplete or uncertain data. The GRA analyzes the level of closeness or relationship between several alternatives based on a series of criteria. One of the limitations in using the GRA method is the weight of the criteria, which is often fixed or subjective. In many GRA applications, the criterion weights are set based on expert considerations or decision-maker preferences, which can be highly subjective and influenced by individual biases. Grey relational analysis change data driven (GRA-C) method emphasizes the increased effectiveness and flexibility of this method in performance appraisal for multi-criteria decision-making. GRA-C allows for more precise adjustments according to the importance of each criterion, leading to more accurate and relevant evaluation results. By modifying the weights, the GRA-C becomes more flexible and can be adapted to different contexts and specific decision-making needs, so that it can be applied in various industry sectors. These modifications help reduce bias due to improper weight allocation, resulting in more objective performance assessments. The results of the modified GRA-C can provide better insights for decision-makers, supporting a more effective and informed decision-making process. The comparison with the Spearman correlation shows that the GRA-C method has a very strong degree of conformity in producing alternative rankings, with a correlation value 1. This indicates that these methods provide similar results, making them reliable for consistent decision-making.

ABSTRAK: Analisis Perhubungan Kelabu (Grey Relational Analysis, GRA) merupakan satu kaedah dalam teori sistem kelabu yang digunakan untuk menyelesaikan masalah keputusan berbilang kriteria (multi-criteria decision-making) yang melibatkan data tidak lengkap atau tidak pasti. GRA menganalisis tahap keterkaitan atau hubungan antara beberapa alternatif berdasarkan satu siri kriteria. Salah satu kekangan dalam penggunaan kaedah GRA ialah pemberat kriteria yang sering kali bersifat tetap atau subjektif. Dalam banyak aplikasi GRA, pemberat kriteria ditentukan berdasarkan pertimbangan pakar atau keutamaan pembuat keputusan, yang boleh menjadi sangat subjektif dan dipengaruhi oleh bias individu. Kaedah Grey Relational Analysis Change Data Driven (GRA-C) menekankan keberkesanan dan fleksibiliti yang lebih tinggi dalam penilaian prestasi bagi sistem keputusan berbilang kriteria. GRA-C membolehkan pelarasan yang lebih tepat mengikut kepentingan setiap kriteria, yang membawa kepada keputusan penilaian yang lebih tepat dan relevan. Dengan pengubahsuaian pemberat, GRA-C menjadi lebih fleksibel dan boleh disesuaikan dengan pelbagai konteks serta keperluan khusus dalam membuat keputusan, membolehkannya diaplikasikan dalam

https://doi.org/10.31436/iiumej.v26i2.3494

pelbagai sektor industri. Pengubahsuaian ini membantu mengurangkan bias akibat pengagihan pemberat yang tidak sesuai, sekali gus menghasilkan penilaian prestasi yang lebih objektif. Hasil daripada GRA-C yang telah diubah suai dapat memberikan pandangan yang lebih baik kepada pembuat keputusan, seterusnya menyokong proses membuat keputusan yang lebih berkesan dan berasaskan maklumat. Perbandingan dengan korelasi Spearman menunjukkan bahawa kaedah GRA-C mempunyai tahap kesesuaian yang sangat tinggi dalam menghasilkan kedudukan alternatif, dengan nilai korelasi sebanyak 1. Ini menunjukkan bahawa kaedah-kaedah tersebut memberikan hasil yang serupa dan boleh dipercayai untuk proses membuat keputusan yang konsisten.

KEYWORDS: *Comparison, Decision, GRA-C, Modification.*

1. INTRODUCTION

Grey relational analysis (GRA) is a method in grey system theory that is used to solve multi-criteria decision problems with incomplete or uncertain data [1]. The GRA analyzes the level of closeness or relationship between several alternatives based on a series of criteria. In the GRA, the relationship between different data is calculated using the concept of gray relational grade, which represents the relative proximity of each alternative to the ideal condition [2]. The general application of GRA in multi-criteria systems is very broad and varied, due to its ability to handle problems with incomplete or uncertain data [3]. GRA is often used for performance evaluation in various fields, such as employee performance assessment, supplier selection, and product or service selection. In the manufacturing industry, GRA is applied to optimize production processes and evaluate product quality. In project management, GRA selects the best projects based on several criteria such as cost, time, and quality. In addition, GRA is also used in environmental research to analyze environmental performance and choose the best strategy in natural resource management [4]. The advantage of GRA in a multi-criteria system is its flexibility in assessing various alternatives based on several complex and diverse criteria [5]. One of the main advantages of GRAs is their ability to handle problems with incomplete or uncertain information, which often arises in multi-criteria decision-making. This method does not require complete or perfect data to perform the analysis, so it is suitable for applications when the available data is limited, inconsistent, or uncertain. GRA reduces complexity by focusing on patterns of relationships between data, rather than demanding absolute precision [6]. Despite the lack of information, GRA can use the gray relational grade to determine the degree of proximity between the evaluated alternatives to the ideal solution. This makes GRA a flexible and reliable tool in a variety of applications such as performance evaluation, risk analysis, and decision-making in environments full of uncertainty or incomplete information [7].

One limitation in using the GRA method is the weight of the criteria, which is often fixed or subjective [8]. In many GRA applications, the criterion weights are set based on expert considerations or decision-maker preferences, which can be highly subjective and influenced by individual biases. This weight determination is also less adaptive to changes in different contexts or decision scenarios, so it can reduce the accuracy of the analysis results if the weights do not reflect the actual conditions. In addition, unchanging weights often do not consider the dynamics in the relationship between criteria, for example, when criteria become more or less critical depending on the specific situation. These limitations can cause the results of decisions to be less than optimal or not follow real conditions, especially in complex and dynamic environments. The main limitation in using GRA is the often fixed or subjective weight of the criteria, resulting in a lack of flexibility in the analysis. The importance of adjusting the weight of the requirements is crucial so that the analysis can be more relevant to the desired goal [9].

With weights that can be adjusted objectively or dynamically, GRA can capture changes in priorities between criteria, improve the accuracy of analysis results, and provide more effective solutions according to the contextual needs of the decisions at hand.

Problems that arise due to weights that are not adjusted to the reality of complex and varied decision-making can result in several negative impacts [10]. The resulting decision may not reflect actual conditions or needs, as the weight of the fixed criteria cannot capture the dynamics and complexity of the situation. Inaccuracies in weights can lead to dissatisfaction among stakeholders, as decisions taken may not be in line with their expectations or priorities [11]. This can reduce trust in the decision-making system used. The use of irrelevant weights can result in neglecting important criteria that should be considered in the decision-making process, potentially resulting in fatal errors in alternative selection. If the weight is not adjusted to the change, the decisions taken may become ineffective and unresponsive to the challenges that arise [12]. As such, it is important to develop methods that allow for objective and dynamic weight adjustments to make decision-making results more accurate and relevant. The main problem in this study is the lack of ability of conventional RA methods to accommodate the dynamics of criterion weights in decision-making. In many situations, the weight of the criteria is often considered fixed, although the relevance of each criterion may change based on context or changes in the data. For example, when selecting suppliers, criteria such as price may have a higher weight in times of economic crisis, but product quality can be a top priority in normal times. The GRA's inability to capture these changes can lead to less accurate analysis results, potentially resulting in suboptimal decisions. In addition, standard GRA methods tend to be sensitive to large-scale or multi-dimensional data, as they lack an adequate mechanism to balance the influence between criteria based on their relevance dynamically. This problem is exacerbated when data is complex, heterogeneous, or frequently updated. The reliance on fixed weights can also reduce the flexibility of these methods in responding to evolving decision needs, especially in highly competitive or rapidly changing environments, such as the business, technology, or crisis management sectors.

The development of a more flexible and adaptive GRA model for applications in multi-criteria performance assessment in this study uses change data-driven. The change-data-driven approach allows for real-time adjustment of criterion weights based on the analysis of the latest data [13], so that models can respond quickly to changing needs or evolving situations. Thus, decision-making becomes more precise and relevant, resulting in more accurate and effective performance evaluations. This approach also increases the flexibility of the model, allowing for better adaptation to frequent changes in operational and business contexts [14]. Increasing the accuracy and relevance of assessment results in the GRA can be achieved through a more dynamic adjustment of the criteria weight. By applying a responsive approach to change data, criterion weights can be adjusted in real-time based on recent data analysis on performance and environmental conditions. This means that when situations or priorities change, the weight of the criteria will reflect new needs and conditions, resulting in a more accurate and relevant analysis. This research contributes to enriching GRA-based decision-making methodologies by developing more effective and efficient models through dynamic and responsive adjustment of criterion weights. By integrating a change-driven data-driven approach, this study provides a framework that allows real-time adjustment of criterion weights based on changes in conditions and evaluation needs [15]. This not only improves the accuracy of the assessment results, but also ensures that the analysis remains relevant in an ever-changing context [16].

The main problem to be solved in this study is the limitation of the GRA method in handling the weight of criteria that change dynamically according to the conditions or context of the decision. In many cases, the weight of the criteria is often considered fixed, thus ignoring

the dynamic nature of factors that affect the decision-making process, such as changing priorities, market situations, or environmental contexts. This can lead to less accurate and relevant analysis results. By modifying the GRA to integrate dynamic criteria weighting, this study aims to improve the flexibility and accuracy of the decision support system, so that it is more adaptive to changing conditions and produces more optimal decisions. The study aims to develop and improve the flexibility of the GRA method by modifying the weight parameters in performance assessment to provide a more accurate and relevant evaluation in the context of multi-criteria decision-making. This modification aims to overcome the limitations of traditional GRA methods that often use fixed weights by offering a more dynamic and responsive approach to the relative importance of each criterion. Thus, this research is expected to provide deeper insights for decision-makers in various industry sectors and improve the quality and effectiveness of the decision-making process.

2. RELATED WORKS

The GRA method has been widely used in various fields to solve multi-criteria decision problems. With its wide range of applications and innovations, GRA remains an interesting and relevant research topic in decision support systems. The following is related to research conducted using the GRA method.

Research from Andika (2024) shows that combining GRA and Rank Order Centroid (ROC) in determining supervisor promotions can provide a more objective and comprehensive approach. The GRA evaluates supervisors' performance based on several criteria that are not completely clear or cannot be measured with certainty. Meanwhile, ROC is used to determine the relative weight of each criterion by giving higher priority to the aspects that decision-makers consider most important. By combining these two methods, supervisor performance assessment can be carried out by considering the relationship between existing criteria, as well as giving more proportional weight to the most relevant criteria for promotion, so that the decisions taken are more transparent and fair [17].

Research from Gao (2024) shows that combining GRA and Entropy in optimizing laser coating parameters for steel presents an effective approach in improving coating quality. GRA is used to evaluate the performance of various laser coating parameters such as laser power, scanning speed, and feed rate based on several quality criteria, such as coating thickness, hardness, and wear resistance. Meanwhile, the Entropy method helps to determine the objective weight for each criterion based on the variation of the data obtained from the experiment, so that the parameters significantly influencing the coating quality get a higher priority. By combining these two methods, the optimization process becomes more directional and data-driven, allowing for the selection of optimal coating parameters to achieve the best results in steel applications [18].

Research from Lu (2024) The combination of GRA and Criteria Importance Through Intercriteria Correlation (CRITIC) in selecting agricultural machinery provides a smarter, data-driven approach to support decision-making. GRA measures the performance of various alternative agricultural machinery based on several important criteria. On the other hand, it is used to objectively calculate the weight of the criteria by considering the variation and correlation between them, so that the more important and unrelated criteria get greater weight. By combining GRA and CRITIC, decision-making becomes more comprehensive, taking into account both the relative performance of each machine and the importance of each criterion objectively, which ultimately allows the selection of agricultural machinery that best suits the needs and operational conditions [19].

https://doi.org/10.31436/iiumej.v26i2.3494

Research from Arshad (2024) suggests that combining GRA and the Pivot Pairwise Relative Criteria Importance Assessment (PIPRECIA) method in selecting warehouse heads offers an effective and structured approach. GRA is used to evaluate the performance of prospective warehouse heads based on several criteria. Meanwhile, the PIPRECIA method helps to determine the weight of the criteria subjectively by involving repeated assessments from experts who consider the relative importance of each criterion. Through this stage, the weight of the criteria can be obtained more accurately according to the specific priorities and needs of the organization. By combining GRA for performance analysis and PIPRECIA for criteria weighting, the warehouse head selection process becomes more transparent and accurate, ensuring that the best candidates are selected based on a thorough and relevant evaluation [20].

The results of comparing the combination of the GRA method with various weighting methods, such as ROC, Entropy, CRITIC, and PIPRECIA, show how different approaches can be used for optimization and decision selection based on relevant criteria.

3. MATERIALS AND METHODS

The GRA method has been widely used in various fields to solve multi-criteria decision problems. With its wide range of applications and innovations, GRA remains an interesting and relevant research topic in decision support systems.

3.1. Case Study and Dataset

In this study using a case study on the selection of the best division head, the analysis was carried out using a decision support system approach that involved several performance criteria, such as leadership (C-01) it is a person's ability to motivate, direct, and influence team or organization members to achieve a common goal, managerial ability (C-02) this ability includes skills in managing company resources, including human, financial, time, and material, to achieve organizational targets efficiently, task completion (C-03) this a person's ability to complete assigned tasks on time and according to the expected standards. This includes work discipline, punctuality, reliability in meeting deadlines, and the ability to overcome obstacles and challenges in the execution of tasks, innovation (C-04), the ability to think creatively and create new solutions to existing challenges. These criteria include the development of new ideas, the ability to improve existing processes or products, and the courage to take risks in exploring untested approaches, and collaboration between teams (C-05). The ability to work with other team members effectively within and between divisions. Individuals who are strong in collaboration can communicate, share information, and work with various parties to achieve common goals. Each candidate is assessed using historical data on their performance and the results of interviews and assessments from colleagues and direct supervisors. By applying this approach, companies not only rely on intuition or experience alone but also obtain results based on comprehensive quantitative and qualitative analysis. Thus, selecting the best division head can be done objectively, reducing bias, increasing credibility and fairness in decision-making, and ensuring that the chosen candidates have optimal potential to lead the division towards better performance.

The dataset used contains performance appraisal data from some division head candidates, where each candidate is evaluated by the company's internal appraisal panel, which includes senior managers and related team members. The assessment is based on a predetermined value scale for each criterion, and the weight of the criteria is set to reflect the company's strategic priorities. This dataset allows testing various multi-criteria decision-making methods to

https://doi.org/10.31436/iiumej.v26i2.3494

determine the most qualified candidates based on the overall rating obtained. Table 1 is the dataset used in this study.

Table 1. Dataset assessment

Division Name	C1	C2	C3	C4	C5
Production Division	8	9	9	6	8
R&D Division	7	7	8	10	7
Quality Division	6	8	9	5	7
Logistics and Supply Chain Division	7	9	8	7	9
Maintenance Division	6	7	9	6	6
Finance Division	8	9	8	5	7
Human Resources Division	9	8	7	6	9
Purchasing Division	7	8	8	6	7
IT Division	6	7	8	8	8

Assessment data for each division is based on five criteria: leadership, managerial ability, task completion, innovation, and team collaboration. Scores are given on a scale of 1-10, where 1 is the lowest and 10 is the highest.

3.2. Modification of Grey Relational Analysis

The modification of the GRA aims to improve the accuracy and flexibility of the GRA method in the multi-criteria decision-making process. One of the most common modifications is to integrate advanced weighting techniques that are objective-based to better reflect the importance of each criterion. Other modifications focus on adjusting the gray relational coefficient to handle high-dimensional data more effectively or applying normalization techniques to minimize outlier influence and ensure consistency of results. These improvements aim to make GRA more reliable and versatile, enabling it to tackle more complex decision-making scenarios with higher precision.

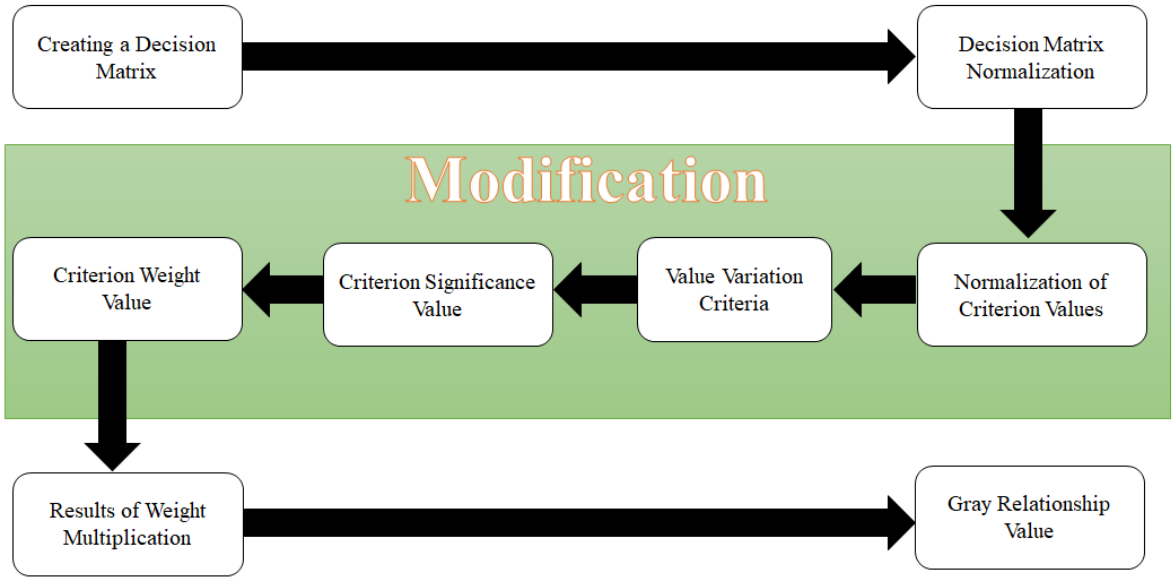


Figure 1. Framework GRA-C

https://doi.org/10.31436/iiumej.v26i2.3494

The modification of GRA with a change data-driven approach called grey relational analysis change data-driven (GRA-C) focuses on adjusting the GRA method to be more responsive to dynamic data changes in the decision-making process. In this approach, the GRA algorithm is modified to accommodate real-time or periodic data updates, so the analysis results can continue to reflect current conditions. The weighting of the criteria and the relational coefficient are changed based on significant changes in data patterns, which allows for more adaptive analysis. As such, these modified GRA-C can handle scenarios with ever-evolving data, such as market changes or technological developments, providing more relevant and accurate results in real-time. Figure 1 is the GRA-C framework conducted in this study.

The GRA-C framework is designed to improve the effectiveness of grey relational analysis by handling dynamic data in multi-criteria decision-making. Within this framework, there are five main modification processes. First is the normalization of criterion values, which are changed to the same scale to ensure equivalence in the analysis. Second, the criterion average score is used to calculate the middle score of each criterion, providing an overview of the candidate's performance. Third, the criterion variation value measures how far the criterion value spreads, aiding in identifying performance stability. Fourth, the significance value of the criteria assesses the relative contribution of each criterion to the final decision, so that the more critical criteria can be recognized. Fifth, the criterion weighting value determines the priority of the criteria based on the calculated significance value, providing a solid basis for the final calculation in the GRA. By integrating these five processes, GRA-C can provide more accurate and relevant analysis results in an ever-changing situation.

Creating a Decision Matrix: This stage involves collecting data from alternatives that will be evaluated based on predetermined criteria. The data is organized in the form of a decision matrix, where rows represent alternatives [21], and columns represent evaluation criteria. Each cell in the matrix contains a value that indicates how well each alternative meets those criteria. Eq. (1) creates a decision matrix.

$$X = \begin{bmatrix} x_{11} & x_{21} & x_{2n} \\ x_{12} & x_{22} & x_{2n} \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \quad (1)$$

Decision matrix normalization: At this stage, the values in the decision matrix are normalized to convert them into the same scale [22]. Normalization aims to eliminate differences in units and scales between criteria, so that all criteria can be directly compared. Decision matrix normalization uses Eq. (2).

$$X_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Normalization of criterion values: After the decision matrix is normalized, the criterion values for each alternative are further processed to ensure that the values obtained reflect consistent performance [23]. Normalization of these criteria is important to provide a more accurate picture of how each alternative performs in the same context. Normalization of criterion values uses Eq. (3).

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} {x_{ij}}^2}} \quad (3)$$

The average value of the criteria results from a calculation that describes the level of performance or the general score of a group based on certain criteria. This modification stage is carried out in this study. The average value of the criteria is important because it provides an

https://doi.org/10.31436/iiumej.v26i2.3494

overview of how the group as a whole performs against those criteria[24], which can help in more objective and balanced decision-making. The average value of the criteria in Eq. (4).

$$N_i = \frac{1}{m}\sum_{i=1}^{m} r_{ij} \tag{4}$$

Value variation criteria: At this stage, the variation value is calculated for each criterion to measure how much the spread of the existing value is at a modification stage carried out in this study. The value of variation provides insight into the consistency of alternative performance under certain criteria [25], aiding in identifying highly variable criteria that can influence decisions. For the value variation criteria, use Eq. (5).

$$\emptyset_j = \sum_{i=1}^{m}(r_{ij} - N_i)^2 \tag{5}$$

Criterion significance value: This stage measures the relative significance of each criterion in the overall context of this analysis, which is the modification stage carried out in this study. The significance value is calculated by taking into account the average value and variation [26] to determine which criteria have a greater impact on the final decision. The criterion significance value uses Eq. (6).

$$\varphi_j = 1 - \emptyset_j \tag{6}$$

Criterion weight value: Once the significance value is determined, this step involves determining the weight for each of these criteria, and a modification stage is carried out in this study. The criterion weights reflect the relative priority of each criterion based on its significance and are used to make the appropriate contribution in the final calculation [27]. The criterion weight value uses Eq. (7).

$$w_j = \frac{\varphi_j}{\sum_{j=1}^{n}\varphi_j} \tag{7}$$

Results of weight multiplication: At this stage, the normalized criterion value is multiplied by the predetermined weight. The result of this multiplication contributes from each criterion to the overall value of each alternative, thus allowing for a more precise comparison between the alternatives being evaluated [28]. The results of weight multiplication use Eq. (8).

$$V_{ij} = w_j \times x_{ij} \tag{8}$$

Gray relationship value: Finally, the gray relationship value is calculated by integrating the results of multiplying the weights and the criterion value. This value reflects the relationship between the alternatives and the criteria, indicating how well each alternative meets all the criteria that have been set [29]. These end results are used to rank alternatives based on their overall performance. Grey relationship value uses Eq. (9).

$$GRG_i = \frac{1}{n}\sum_{j=1}^{n} V_{ij} \tag{9}$$

4. RESULT AND DISCUSSION

Modifying the weight parameters in the GRA for performance assessment in multi-criteria decision-making provides greater flexibility than traditional methods. By changing the weight of the criteria, the GRA can better adjust to specific preferences or different priorities in the evaluation, so that the results are more accurate and relevant to the specific context. This increased flexibility allows decision-makers to consider the importance of each criterion more dynamically, providing more precise solutions in complex and diverse situations. It also expands the application of GRA in various fields, from performance appraisal to selection of the best alternatives.

https://doi.org/10.31436/iiumej.v26i2.3494

GRA-C modification is a development of the GRA method that focuses on dynamic changes in data in multi-criteria decision-making. This approach takes advantage of constantly changing or updated data, so that it can adjust analysis and decision results more responsively to changes in information or conditions. GRA-C allows for more accurate performance evaluations or decision alternatives because it considers trends and changes in the data. GRA-C becomes more adaptive and relevant in contexts requiring rapid response to changes, such as fast-moving business environments or technologies, using this dynamic, data-driven approach.

4.1. Implementation of GRA-C in Case Studies

The implementation of GRA-C in the case study of the performance assessment of division heads can be applied to evaluate the performance of each division based on various criteria that are important in achieving organizational goals. Using GRA-C, the analysis is carried out by taking into account dynamic changes in the performance data of division heads from time to time. GRA-C will calculate the level of gray relationship (gray relational grade) between the performance of the division head and the desired ideal target. Changes in the weight of criteria and up-to-date performance data allow for a more adaptive and accurate assessment of the performance of division heads, which may be subject to change due to internal or external conditions.

The decision matrix is the first stage in GRA-C, a representation of data in tables used to assist decision-making. Normalization of the decision matrix uses Eq. (10).

$$X = \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \\ x_{13} & x_{23} & x_{33} & x_{43} & x_{53} \\ x_{14} & x_{24} & x_{34} & x_{44} & x_{54} \\ x_{15} & x_{25} & x_{35} & x_{45} & x_{55} \\ x_{16} & x_{26} & x_{36} & x_{46} & x_{56} \\ x_{17} & x_{27} & x_{37} & x_{47} & x_{57} \\ x_{18} & x_{28} & x_{38} & x_{48} & x_{58} \\ x_{19} & x_{29} & x_{39} & x_{49} & x_{59} \end{bmatrix} = \begin{bmatrix} 8 & 9 & 9 & 6 & 8 \\ 7 & 7 & 8 & 10 & 7 \\ 6 & 8 & 9 & 5 & 7 \\ 7 & 9 & 8 & 7 & 9 \\ 6 & 7 & 9 & 6 & 6 \\ 8 & 9 & 8 & 5 & 7 \\ 9 & 8 & 7 & 6 & 9 \\ 7 & 8 & 8 & 6 & 7 \\ 6 & 7 & 8 & 8 & 8 \end{bmatrix} \quad (10)$$

The normalization of the decision matrix is the second stage in the GRA-C to convert it to the same scale calculated using Eq. (11).

$$X_{11} = \frac{x_{11} - x_{11,19}}{x_{11,19} - x_{11,19}} = \frac{8-6}{9-6} = \frac{2}{3} = 0.667 \quad (11)$$

Table 2 is the calculation result of the matrix normalization assessment of each alternative for each criterion.

Table 2. Normalization matrix

Division Name	C1	C2	C3	C4	C5
Production Division	0.667	1	1	0.2	0.667
R&D Division	0.333	0	0.5	1	0.333
Quality Division	0	0.5	1	0	0.333
Logistics and Supply Chain Division	0.333	1	0.5	0.4	1
Maintenance Division	0	0	1	0.2	0
Finance Division	0.667	1	0.5	0	0.333
Human Resources Division	1	0.5	0	0.2	1
Purchasing Division	0.333	0.5	0.5	0.2	0.333
IT Division	0	0	0.5	0.6	0.667

The criterion of value variation is the third stage in the GRA-C to measure how much the spread of existing values is calculated using Eq. (12).

$$r_{11} = \frac{x_{11}}{\sqrt{\sum_{i=1}^{m} x_{11,19}}} = \frac{8}{\sqrt{8^2+7^2+6^2+7^2+6^2+8^2+9^2+7^2+6^2}} = \frac{8}{\sqrt{464}} = 0.371 \quad (12)$$

Table 3 is the result of the calculation of the normalization of the criteria of each alternative for each criterion.

Table 3. Normalization of the criteria

Division Name	C1	C2	C3	C4	C5
Production Division	0.371	0.373	0.364	0.297	0.350
R&D Division	0.325	0.290	0.323	0.496	0.306
Quality Division	0.279	0.332	0.364	0.248	0.306
Logistics and Supply Chain Division	0.325	0.373	0.323	0.347	0.394
Maintenance Division	0.279	0.290	0.364	0.297	0.263
Finance Division	0.371	0.373	0.323	0.248	0.306
Human Resources Division	0.418	0.332	0.283	0.297	0.394
Purchasing Division	0.325	0.332	0.323	0.297	0.306
IT Division	0.279	0.290	0.323	0.397	0.350

The average value of the criteria is the fourth stage in the GRA-C, which helps provide an overview of how the group as a whole performs against certain criteria calculated using Eq. (13).

$$N_1 = \frac{1}{9}\sum_{i=1}^{m} r_{11,19} = 0.111 * 2.971 = 0.3301 \quad (13)$$

Table 4 is the result of calculating the average value for each criterion.

Table 4. Average value of the criteria

Criteria	C1	C2	C3	C4	C5
Average value	0.3301	0.3316	0.3324	0.3249	0.3307

The value of the variation criterion is the fifth stage in the GRA-C, helping with criteria with high variability and influencing the decision calculated using Eq. (14).

$$\emptyset_1 = \sum_{i=1}^{m}(r_{11,19} - N_1)^2 = 0.019157 \quad (14)$$

Table 5 is the result of calculating the variation value for each criterion.

Table 5. Variation value of the criteria

Criteria	C1	C2	C3	C4	C5
Variation value	0.019157	0.010309	0.005810	0.049686	0.015751

The significance value of the criterion is the sixth stage in the GRA-C to help determine which criterion has the greatest impact on the final decision, calculated using Eq. (6), i.e., $\varphi_1 = 1 - \emptyset_1 = 1 - 0.019157 = 0.980843$. Table 6 is the result of the calculation of the significance value of the criterion.

https://doi.org/10.31436/iiumej.v26i2.3494

Table 6. Significance value of the criteria

Criteria	C1	C2	C3	C4	C5
Significance value	0.980843	0.989691	0.994190	0.950314	0.984249

The criterion weight value is the seventh stage in the GRA-C to reflect the relative priority of each criterion based on its significance, calculated using Eq. (7), i.e., $w_1 = \frac{\varphi_1}{\sum_{j=1}^{n} \varphi_1} = \frac{0.980843}{4.899286} = 0.2002$. Table 7 is the result of the calculation of the weight value of the criterion.

Table 7. Weight value of the criteria

Criteria	C1	C2	C3	C4	C5
Weight value	0.2002	0.202	0.2029	0.194	0.2009

The result of weight multiplication is the eighth stage in the GRA-C to contribute from each criterion to the overall value of each alternative calculated using Eq. (8), i.e., $V_{11} = w_1 * x_{11} = 0.2002 * 0.667 = 0.13347$. Table 8 is the result of calculating the value of weight multiplication for each alternative of the existing criteria.

Table 8. Calculating the value of weight multiplication

Division Name	C1	C2	C3	C4	C5
Production Division	0.13347	0.20201	0.20293	0.03879	0.13393
R&D Division	0.06673	0	0.10146	0.19397	0.06697
Quality Division	0	0.10100	0.20293	0	0.06697
Logistics and Supply Chain Division	0.06673	0.20201	0.10146	0.07759	0.20090
Maintenance Division	0	0	0.20293	0.03879	0
Finance Division	0.13347	0.20201	0.10146	0	0.06697
Human Resources Division	0.20020	0.10100	0	0.03879	0.20090
Purchasing Division	0.06673	0.10100	0.10146	0.03879	0.06697
IT Division	0	0	0.10146	0.11638	0.13393

The gray relationship value is the eighth, ninth, or last stage in the GRA-C. This value reflects the relationship between the alternative and the criterion, which indicates how well each alternative meets all the predetermined criteria calculated using Eq. (9), i.e., $GRG_1 = \frac{1}{5}\sum_{j=1}^{n} V_{11,51} = 0.13347 + 0.20201 + 0.20293 + 0.03879 + 0.13393 = 0.71112$. Table 9 is the result of calculating the grey relationship value for each existing alternative.

The ranking of division heads' performance assessments involves sorting or arranging objects, individuals, or alternatives based on certain criteria to determine the best priority or performance. The ranking results provide an overview of the best to worst alternatives, which helps decision-makers choose the option that best suits the division head's performance assessment. The ranking results are shown in Figure 2.

https://doi.org/10.31436/iiumej.v26i2.3494

Table 9. Calculation of the grey relationship value

Division Name	Grey Value
Production Division	0.71112
R&D Division	0.42913
Quality Division	0.37089
Logistics and Supply Chain Division	0.64869
Maintenance Division	0.24172
Finance Division	0.50390
Human Resources Division	0.54090
Purchasing Division	0.37496
IT Division	0.35178

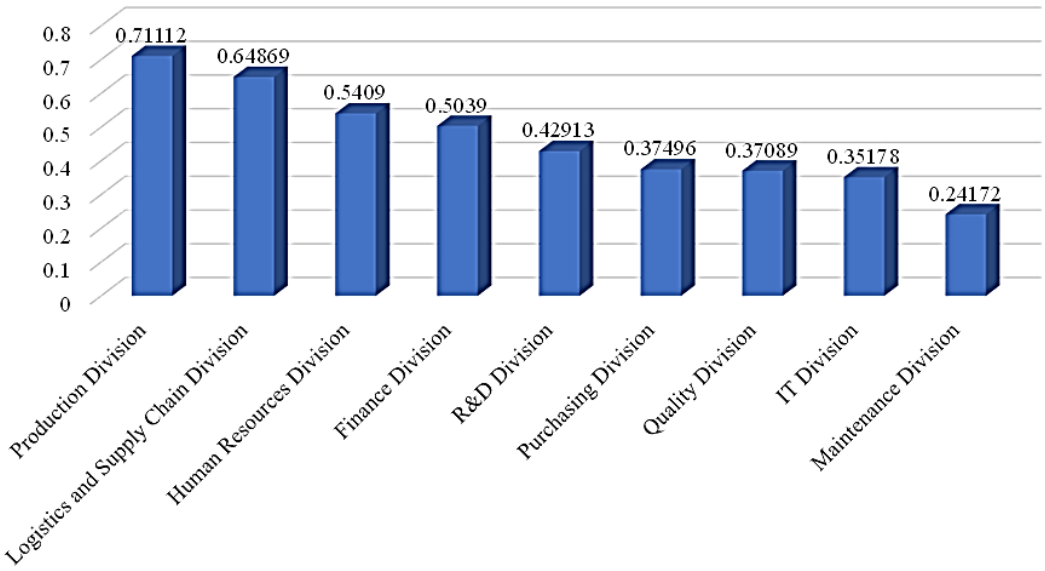


Figure 2. Results of the ranking of the performance assessment of division heads

The results of the division head's performance assessment ranking, Production Division, ranked highest with a score of 0.71112, indicating that this division head performs superior to other divisions. Followed by the Logistics and Supply Chain Division with a score of 0.64869 and the Human Resources Division in third place with a score of 0.5409. The Finance Division is ranked fourth with a score of 0.5039, slightly higher than the R&D Division with a score of 0.42913. The Purchasing Division, Quality Division, and IT Division recorded scores of 0.37496, 0.37089, and 0.35178, respectively. The Maintenance Division was in last position with a score of 0.24172, indicating the lowest performance among all the divisions assessed.

4.2. Discussion

In the context of multi-criteria decision-making, the GRA method has been widely used because of its ability to analyze the relationship between various alternatives based on established criteria. However, one of the main challenges in implementing GRA is its limited flexibility, especially when setting parameter weights that can affect the final result. Therefore, modifications to the weighting parameters can increase the flexibility of the GRA, allowing this method to be more adaptive in handling the complexity of performance assessment problems. Modifying weights in GRA enables this method to be more responsive to differences in significance between criteria used in performance assessment. In the case of decision-making involving many criteria, not all criteria have the same weight or level of importance.

https://doi.org/10.31436/iiumej.v26i2.3494

Therefore, by introducing customized weighting parameters, GRA can prioritize more relevant criteria more precisely, increasing accuracy in ranking results. This increased flexibility also allows decision-makers to tailor the analysis to specific needs, including considering various scenarios and dynamic business or operational environment changes.

Grey Relational Analysis Change Data Driven (GRA-C) is a modified GRA method designed to handle dynamic changes in assessment data. This method introduces a data-driven approach, where changes in the dataset directly affect the weighting process and the relationships between alternatives. By accounting for real-time data changes, GRA-C enables more responsive analysis of fluctuations in criteria or dynamic environmental conditions, such as changes in customer preferences, performance, or market conditions. The GRA-C method leverages a data-change-based algorithm to adapt weighting and increase flexibility in multi-criteria evaluations. For example, suppose there is a significant change in relevant criteria, such as a decrease in the performance of a division or an improvement in the performance of an alternative. In that case, the GRA-C automatically adjusts the relationship between the alternatives based on the new data. This method is particularly effective in applications requiring continuous adjustment and real-time assessment, such as project management, performance appraisal, or supplier selection in highly competitive environments. GRA-C offers a more dynamic solution than conventional GRA methods, as it can model changes directly in a multi-criteria decision-making system, resulting in more relevant and accurate decisions as data continues to change.

The comparison between GRA-C and GRA combined with the criterion weighting method offers an interesting perspective on flexibility and adaptability in multi-criteria decision-making. GRA-C has the advantage of responsiveness to real-time data changes, where criteria weighting and relationships between alternatives can be adjusted automatically based on data dynamics. On the other hand, GRA combined with criterion weighting methods, such as ROC [17], Entropy [18], CRITIC [19], and PIPRECIA [20], prioritizes a more objective or user-preference-based initial weighting before the analysis. However, these methods tend to be static, so they cannot easily adjust weights when there are changes in the data or environment. Table 10 compares the GRA-C and the GRA methods combined with the criterion weighting methods.

Table 10. The result of a comparison of the GRA-C method and the GRA method

Division Name	Original Rank	ROC Rank	Entropy Rank	CRITIC Rank	PIPRECIA Rank	GRA-C Rank
Production Division	1	1	4	1	1	1
Logistics and Supply Chain Division	2	4	2	2	4	2
Human Resources Division	3	3	3	4	2	3
Finance Division	4	2	6	3	3	4
R&D Division	5	6	1	5	5	5
Purchasing Division	6	5	7	7	6	6
Quality Division	7	7	8	6	7	7
IT Division	8	9	5	8	8	8
Maintenance Division	9	8	9	9	9	9

The results of the comparison of Original, ROC, Entropy, CRITIC, PIPRECIA, and GRA-C methods show a variety of approaches in weighting criteria and performance evaluation. The

https://doi.org/10.31436/iiumej.v26i2.3494

Original Method is the result of the ranking obtained from the company. ROC gives weight based on a simple order of priority, but does not consider the relationship between criteria. Entropy offers a more objective weighting by considering the variation of data between criteria, which is suitable for situations where quantitative data is the basis for assessment. CRITIC combines objectivity by looking at correlations between criteria, resulting in a more balanced weight, and considering the influence of factors. PIPRECIA provides an approach based on repeated assessments from experts, suitable for decisions that require a subjective but more systematic professional assessment. Finally, GRA-C integrates performance between alternatives with criterion weighting, resulting in a comprehensive approach to decision-making involving various factors.

The comparison results using Spearman's correlation between the Original, ROC, Entropy, CRITIC, PIPRECIA, and GRA-C methods provide an overview of the extent to which the alternative rankings of each method correlate with each other. Spearman correlation measures the strength of the monotonic relationship between two ranks, so the higher the correlation value, the more similar the ratings produced by the two methods. Figure 3 is the result of Spearman's correlation of the existing ranking.

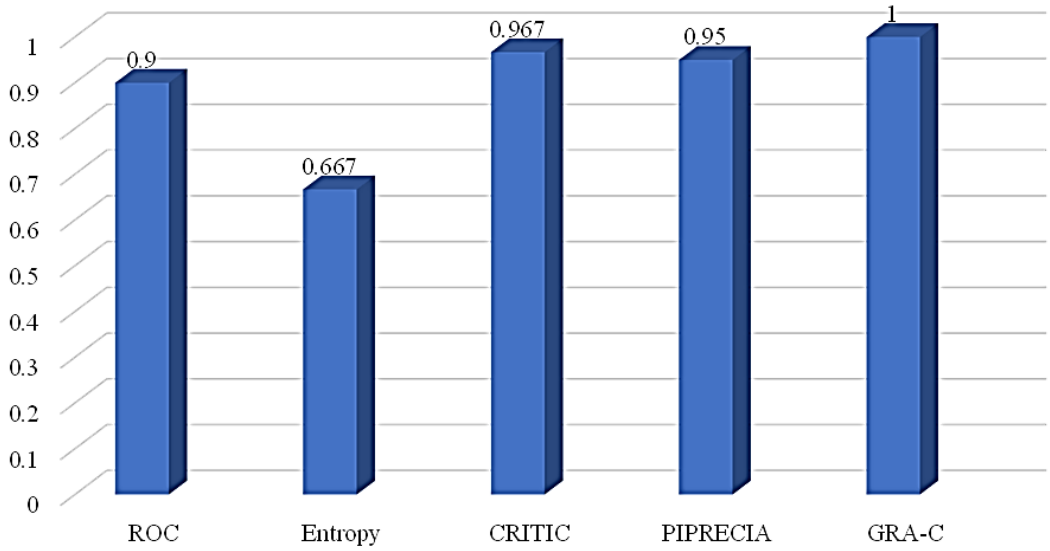


Figure 3. The results of the comparison used Spearman's correlation.

The comparison results displayed in the graph using Spearman correlation show the degree of alignment between the different methods compared. The Spearman correlation value varied from 0.667 to 1, indicating the degree of correlation between these methods in generating alternative rankings. ROC has a correlation value of 0.9, indicating a relatively strong alignment with the ranking of the calculation results. Entropy showed the lowest correlation, 0.667, indicating that this method had a more significant difference in alternative rankings than other methods. CRITIC has a high correlation of 0.967, indicating an almost perfect fit with ratings from different methods, especially those based on objective data. PIPRECIA also showed a strong correlation of 0.95, indicating that this method provides a fairly consistent rating with other methods used. GRA-C achieves the highest correlation value, 1, indicating that the rating generated by GRA-C is entirely consistent with the reference rating or other methods.

The conclusion of the comparison with the Spearman correlation shows that the GRA-C method has an extreme degree of conformity in producing alternative rankings, with a

correlation value of 1 each. This indicates that these methods provide similar results, making them reliable for consistent decision-making.

The limitations of the proposed research related to the GRA method need to be considered to provide a clear context regarding the limitations of this methodology. One of the main limitations is the reliance of GRA on fixed criterion weights, which do not consider the dynamics or changes in the relevance of the criteria based on a particular context or situation. This makes this method less flexible when changing data or in rapid decision-making environments. In addition, GRAs tend to be sensitive to the scale of the data and the normalizations used. Extreme scale differences between criteria can influence specific criteria to become too dominant, thus affecting the analysis results. Improper normalization can also result in gray relationships that do not reflect the proper relationship between alternatives and criteria. The GRA method also has limitations in handling large and complex datasets, especially those that involve many criteria or alternatives. This can increase computing time and decrease the efficiency of the analysis. In addition, GRA is less efficient in handling subjective or qualitative data, as the measurement and quantification process for these kinds of criteria can be complex and prone to bias. Interpreting results from GRA often requires a deep understanding of these methods, which can be challenging for non-technical users or stakeholders who do not have a technical background. Discussing these limitations is essential to provide insight into the scope and application limits of the GRA method and guide further development to address existing shortcomings.

5. CONCLUSION

The GRA-C method emphasizes the increased effectiveness and flexibility of this method in performance appraisal for multi-criteria decision-making. GRA-C allows for more precise adjustments according to the importance of each criterion, leading to more accurate and relevant evaluation results. By modifying the weights, the GRA-C becomes more flexible and can be adapted to different contexts and specific decision-making needs, so that it can be applied in various industry sectors. These modifications help reduce bias due to improper weight allocation, resulting in more objective performance assessments. The results of the modified GRA-C can provide better insights for decision-makers, supporting a more effective and informed decision-making process. The comparison with the Spearman correlation shows that the GRA-C method has extreme conformity in producing alternative rankings, with a correlation value of 1 each. This indicates that these methods provide similar results, making them reliable for consistent decision-making. The comparison results using Spearman's correlation show the degree of alignment between the different methods being compared. The Spearman correlation value varies between 0.667 and 1, showing the degree of correlation between the methods producing alternative rankings. ROC has a correlation value of 0.9, which shows a fairly strong alignment with the ranking of the calculation results. Entropy shows the lowest correlation of 0.667, which shows that this method has a more significant difference in alternative rankings than other methods. CRITIC has a high correlation of 0.967, which indicates almost perfect agreement with the rankings of different methods, especially those based on objective data. PIPRECIA also shows a strong correlation of 0.95, which shows that this method provides a ranking consistent with other methods. GRA-C obtains the highest correlation value of 1, which shows that the ranking produced by GRA-C is entirely consistent with the reference ranking or other methods. Future work in this study can be focused on developing and testing a more flexible GRA-C analysis model with modification of weight parameters for performance assessment in multi-criteria decision-making. Furthermore, research can investigate integrating these methods with other analytical techniques, such as

boundary-based decision analysis or multi-objective programming methods, to provide a more comprehensive view of performance appraisal.

ACKNOWLEDGEMENT

This research was funded by Universitas Teknokrat Indonesia, Hibah Publikasi Terindeks Scopus (HiPuTS) grant number 022/UTI/LPPMI/E.1.1/VIII/2023.

REFERENCES

[1] N. Xiaoyan, W. Ying, W. Zhenduo, and S. Zhiguo, "Link-16 anti-jamming performance evaluation based on grey relational analysis and cloud model," *J. Syst. Eng. Electron.*, pp. 1–11, 2024, doi: 10.23919/JSEE.2023.000120.

[2] K. Mausam, A. Pare, S. K. Ghosh, and A. K. Tiwari, "Thermal performance analysis of hybrid-nanofluid based flat plate collector using Grey relational analysis (GRA): An approach for sustainable energy harvesting," *Therm. Sci. Eng. Prog.*, vol. 37, p. 101609, 2023.

[3] T. Škrinjarić, "Dynamic portfolio optimization based on grey relational analysis approach," *Expert Syst. Appl.*, vol. 147, p. 113207, 2020.

[4] M. Gerus-Gościewska and D. Gościewski, "Grey relational analysis (gra) as an effective method of research into social preferences in urban space planning," *Land*, vol. 11, no. 1, p. 102, 2022.

[5] S. Chatterjee and S. Chakraborty, "Optimization of friction stir welding processes using multi-attributive border approximation area comparison (MABAC) method in neutrosophic fuzzy environment," *Int. J. Interact. Des. Manuf.*, vol. 17, no. 4, pp. 1979–1994, Aug. 2023, doi: 10.1007/s12008-023-01308-6.

[6] B. S. Nithyananda, G. V Naveen Prakash, N. Ankegowda, K. B. Vinay, and A. Anand, "Optimization of Performance and Emission Responses of Common Rail Direct Injection Engine by Taguchi-Grey Relational Analysis Technique," in *RAiSE-2023*, Jan. 2024, vol. 59, no. 1, p. 140. doi: 10.3390/engproc2023059140.

[7] X. Wang *et al.*, "Integration of the grey relational analysis with machine learning for sucrose anaerobic hydrogen production prediction," *Int. J. Hydrogen Energy*, vol. 68, pp. 388–397, 2024.

[8] M. O. Esangbedo, J. Xue, S. Bai, and C. O. Esangbedo, "Relaxed Rank Order Centroid Weighting MCDM Method With Improved Grey Relational Analysis for Subcontractor Selection: Photothermal Power Station Construction," *IEEE Trans. Eng. Manag.*, 2022, doi: 10.1109/TEM.2022.3204629.

[9] K.-H. Chang, "Integrating Subjective–Objective Weights Consideration and a Combined Compromise Solution Method for Handling Supplier Selection Issues," *Systems*, vol. 11, no. 2, p. 74, 2023.

[10] A. Le Gouais *et al.*, "Understanding how to create healthier places: A qualitative study exploring the complex system of urban development decision-making," *Health Place*, vol. 81, p. 103023, May 2023, doi: 10.1016/j.healthplace.2023.103023.

[11] H. Sulistiani, Setiawansyah, P. Palupiningsih, F. Hamidy, P. L. Sari, and Y. Khairunnisa, "Employee Performance Evaluation Using Multi-Attribute Utility Theory (MAUT) with PIPRECIA-S Weighting: A Case Study in Education Institution," in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, 2023, pp. 369–373. doi: 10.1109/ICIMCIS60089.2023.10349017.

[12] H. Sulistiani, S. Setiawansyah, A. F. O. Pasaribu, P. Palupiningsih, K. Anwar, and V. H. Saputra, "New TOPSIS: Modification of the TOPSIS Method for Objective Determination of Weighting," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 5, pp. 991–1003, Oct. 2024, doi: 10.22266/ijies2024.1031.74.

[13] S. Dhruva, R. Krishankumar, E. K. Zavadskas, K. S. Ravichandran, and A. H. Gandomi, "Selection of Suitable Cloud Vendors for Health Centre: A Personalized Decision Framework with Fermatean Fuzzy Set, LOPCOW, and CoCoSo," *Informatica*, vol. 35, no. 1, pp. 65–98, Nov. 2024, doi: 10.15388/23-INFOR537.

[14] Setiawansyah, A. A. Aldino, P. Palupiningsih, G. F. Laxmi, E. D. Mega, and I. Septiana,

"Determining Best Graduates Using TOPSIS with Surrogate Weighting Procedures Approach," in *2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology (IConNECT)*, 2023, pp. 60–64. doi: 10.1109/IConNECT56593.2023.10327119.

[15] J. Więckowski, B. Kizielewicz, B. Paradowski, A. Shekhovtsov, and W. Sałabun, "Application of Multi-Criteria Decision Analysis to Identify Global and Local Importance Weights of Decision Criteria," *Int. J. Inf. Technol. Decis. Mak.*, vol. 22, no. 06, pp. 1867–1892, Nov. 2023, doi: 10.1142/S0219622022500948.

[16] M. Deveci, D. Pamucar, I. Gokasar, M. Köppen, B. B. Gupta, and T. Daim, "Evaluation of Metaverse traffic safety implementations using fuzzy Einstein based logarithmic methodology of additive weights and TOPSIS method," *Technol. Forecast. Soc. Change*, vol. 194, p. 122681, Sep. 2023, doi: 10.1016/j.techfore.2023.122681.

[17] R. Andika, "Kombinasi Grey Relational Analysis (GRA) dan ROC Dalam Penentuan Promosi Jabatan Supervisor," *Chain J. Comput. Technol. Comput. Eng. Informatics*, vol. 2, no. 1, pp. 37–44, 2024, doi: 10.58602/chain.v2i1.94.

[18] S. Gao *et al.*, "Optimization of Laser Cladding Parameters for High-Entropy Alloy-Reinforced 316L Stainless-Steel via Grey Relational Analysis," *Coatings*, vol. 14, no. 9, p. 1103, Sep. 2024, doi: 10.3390/coatings14091103.

[19] H. Lu, Y. Zhao, X. Zhou, and Z. Wei, "Selection of Agricultural Machinery Based on Improved CRITIC-Entropy Weight and GRA-TOPSIS Method," *Processes*, vol. 10, no. 2, p. 266, Jan. 2022, doi: 10.3390/pr10020266.

[20] M. W. Arshad, S. Setiawansyah, and S. Sintaro, "Comparative Analysis of the Combination of MOORA and GRA with PIPRECIA Weighting in the Selection of Warehouse Heads," *BEES Bull. Electr. Electron. Eng.*, vol. 4, no. 3, pp. 112–122, Mar. 2024, doi: 10.47065/bees.v4i3.4922.

[21] N. Moldybayeva *et al.*, "Decision Matrix in an Autonomous Power System for Agro-Industrial Complexes with Renewable Energy Sources," *Energies*, vol. 17, no. 17, p. 4357, Aug. 2024, doi: 10.3390/en17174357.

[22] M. Baydaş, M. Yılmaz, Ž. Jović, Ž. Stević, S. E. G. Özuyar, and A. Özçil, "A comprehensive MCDM assessment for economic data: success analysis of maximum normalization, CODAS, and fuzzy approaches," *Financ. Innov.*, vol. 10, no. 1, p. 105, Mar. 2024, doi: 10.1186/s40854-023-00588-x.

[23] I. Z. Mukhametzyanov, "Elimination of the Domains' Displacement of the Normalized Values in MCDM Tasks: The IZ-Method," *Int. J. Inf. Technol. Decis. Mak.*, vol. 23, no. 01, pp. 289–326, Jan. 2024, doi: 10.1142/S0219622023500037.

[24] A. Puška, A. Štilić, D. Pamučar, D. Božanić, and M. Nedeljković, "Introducing a Novel multi-criteria Ranking of Alternatives with Weights of Criterion (RAWEC) model," *MethodsX*, vol. 12, p. 102628, Jun. 2024, doi: 10.1016/j.mex.2024.102628.

[25] H. U. Khan *et al.*, "Multi-criteria decision-making methods for the evaluation of the social internet of things for the potential of defining human behaviors," *Comput. Human Behav.*, vol. 157, p. 108230, Aug. 2024, doi: 10.1016/j.chb.2024.108230.

[26] I. Naz *et al.*, "Integrated Geospatial and Geostatistical Multi-Criteria Evaluation of Urban Groundwater Quality Using Water Quality Indices," *Water*, vol. 16, no. 17, p. 2549, Sep. 2024, doi: 10.3390/w16172549.

[27] H. Aljaghoub *et al.*, "Comparative analysis of various oxygen production techniques using multi-criteria decision-making methods," *Int. J. Thermofluids*, vol. 17, p. 100261, Feb. 2023, doi: 10.1016/j.ijft.2022.100261.

[28] Marković, Stajić, Stević, Mitrović, Novarlić, and Radojičić, "A Novel Integrated Subjective-Objective MCDM Model for Alternative Ranking in Order to Achieve Business Excellence and Sustainability," *Symmetry (Basel).*, vol. 12, no. 1, p. 164, Jan. 2020, doi: 10.3390/sym12010164.

[29] J. Wei, C. Huang, S. Li, J. Liu, L. Huang, and Y. Liang, "Stress Analysis and Gray Relation Optimization of QFP Solder Joints Under Torsional and Bending Loading," in *2024 25th International Conference on Electronic Packaging Technology (ICEPT)*, Aug. 2024, pp. 1–6. doi: 10.1109/ICEPT63120.2024.10668549.

https://doi.org/10.31436/iiumej.v26i2.3653

LATENCY PERFORMANCE EVALUATION OF LEO STARLINK AND SES-12 GEO HTS NETWORK UNDER TROPICAL RAINFALL CONDITIONS

HAZRUL HAFIZ ABDUL SHUKUR, YASSER ASRUL AHMAD*, MUHAMMAD SHARIR FATHULLAH MOHD YUNUS, KHAIRAYU BADRON

Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University of Malaysia, Kuala Lumpur, Malaysia

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

(Received: 17 March 2025; Accepted: 22 April 2025; Published online: 15 May 2025)

ABSTRACT: High Throughput Satellites (HTS) in geostationary Earth orbit (GEO) have been rapidly launched to meet the growing demand for high-speed data. However, the latency of HTS remains the same as that of conventional GEO satellites due to the characteristics of their orbit. Recently, Starlink HTS satellites in low Earth orbit (LEO) mega constellations have been operationalized globally, providing low-latency internet services compared to GEO HTS. Despite their high-speed benefits, Ku-band HTS systems are highly susceptible to rain-induced signal attenuation, particularly in regions with heavy rainfall, such as the tropics and equatorial countries. This phenomenon weakens the radio frequency signals and impacts real-time latency in the communication link. This research aims to determine the latency effect of the HTS satellite in LEO and GEO and evaluate its performance under heavy rainfall conditions. This study utilises real satellite link services from SES-12 GEO HTS and LEO Starlink for performance assessment. Continuous latency measurements are recorded over six months to analyse Ku-band performance in a heavy rainfall region. The results indicate that extreme rainfall in the tropical region significantly affects GEO satellite links, causing prolonged signal degradation due to fixed ground stations. In contrast, Starlink's LEO network experiences less impact from rain fade, as it dynamically switches between multiple satellites. The results show latency for the GEO terminal link ranges from 600 milliseconds (ms) to 3000 milliseconds (ms), whereas latency for the LEO Starlink terminal ranges from 20 milliseconds (ms) to 100 milliseconds (ms). Starlink provides higher satellite link availability at 99.6% onsite compared to 94% for the tropical region's SES-12 GEO HTS satellite services.

ABSTRAK: Satelit Jalur Lebar Berkapasiti Tinggi (HTS) dalam orbit geostasioner (GEO) berkembang pesat bagi memenuhi permintaan data berkelajuan tinggi. Namun, latensi HTS GEO kekal tinggi seperti satelit GEO tradisional. Sebaliknya, HTS dalam orbit bumi rendah (LEO) seperti Starlink menawarkan internet berlatensi rendah secara global.Walaupun berkelajuan tinggi, sistem HTS menggunakan jalur Ku terdedah kepada pelemahan isyarat hujan, terutamanya di kawasan tropika. Fenomena ini bukan sahaja melemahkan isyarat frekuensi radio tetapi juga menjejaskan latensi komunikasi masa nyata. Penyelidikan ini bertujuan untuk menentukan kesan latensi sistem satelit HTS di orbit LEO dan GEO serta menilai prestasinya dalam keadaan hujan lebat. Pengukuran latensi berterusan direkodkan selama enam bulan bagi menganalisis prestasi jalur Ku di kawasan beriklim hujan lebat.Kajian ini menilai kesan latensi HTS LEO dan GEO dalam keadaan hujan lebat menggunakan perkhidmatan satelit SES-12 GEO HTS dan LEO Starlink di Makmal Komunikasi Satelit IIUM. Keputusan menunjukkan hujan lebat memberi kesan besar kepada sambungan GEO kerana stesen bumi tetap, manakala rangkaian LEO Starlink kurang terjejas

https://doi.org/10.31436/iiumej.v26i2.3653

kerana boleh beralih antara satelit. Kelewatan GEO ialah 600ms–3000ms, manakala Starlink hanya 20ms–100ms. Starlink juga lebih stabil dengan 99.6% ketersediaan berbanding 94% bagi SES-12 GEO HTS di kawasan tropika menyebabkan degradasi isyarat berpanjangan.

KEYWORDS: *High Throughput Satellite, Ku Band, Starlink Satellite Latency, Tropical Regions, Rain Attenuation.*

1. INTRODUCTION

A high-throughput Satellite (HTS) is a type of satellite designed to handle large data communication links, offering high data throughput capabilities. Each spot beam reuses the frequencies in multiple transponders or carriers, which a single HTS can provide more than 100 Gbps of capacity, which is more than 100 times the capacity offered by conventional Ku-Band satellites [1].

A geostationary satellite (GSO) is an Earth-orbiting satellite at the same rate as the Earth's rotation, located approximately 36,000 km from the Earth's surface. This shows that the satellite appears stationary at a fixed point in the sky [2]. The first Ku GEO HTS launched was Thaicom in 2005, also known as IPSTAR, with 45 Gbps global capacity, and the launch of ViaSat 2 in 2017 with over 300 Gbps [3]. Furthermore, Geostationary satellite lifetimes are now extending to 15 years, and traffic demand changes are occurring rapidly during their operational period, affecting service areas and applications [4].

While for LEO, two major announcements were made in April 2019 where Amazon is planning an (HTS) constellation for broadband services with over 3,000 satellites in Low Earth Orbit (LEO) known as Project Kuiper in the year 2020 and subsequently SpaceX announced its HTS constellation known as Project Starlink and is going to launch nearly 1,600 satellites [5].

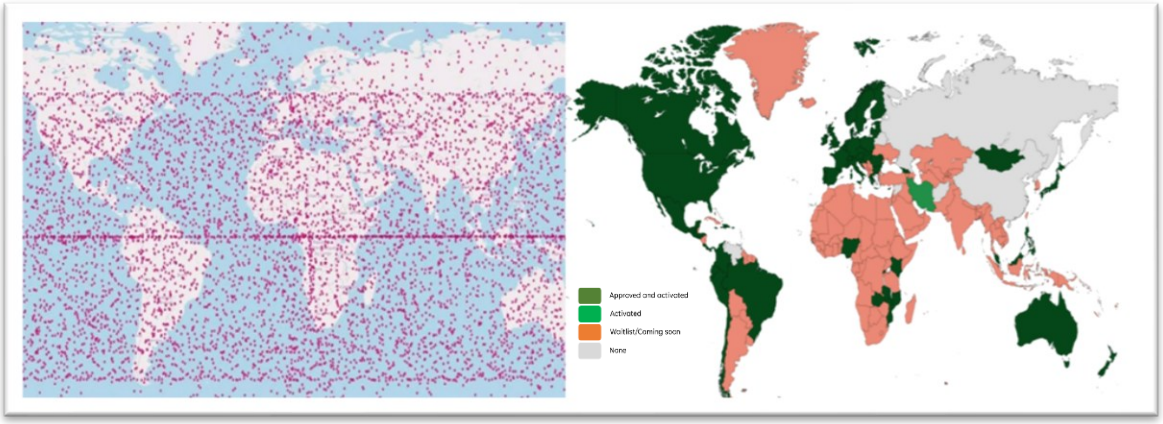


Figure 1. Starlink Satellite operation and distribution map[10]

Placing an HTS in LEO offers advantages such as lower latency, improved signal propagation, and enhanced data transmission speeds due to the satellite's proximity to Earth[6]. The satellite industry has experienced a rapid transformation to meet the ever-growing demand for high data throughput. Recently, Starlink has emerged as a prominent LEO satellite network service and entered the Malaysian market in 2023 [7]. SpaceX has received Class Assignment (CA) approval on a non-interference and non-protection basis to operate Starlink terminals using the Ku-band in Malaysia, with its gateway operating in the Ka-band at frequencies ranging from 12 to 40 GHz [8]. GEO HTS satellite services also utilise these frequency ranges,

https://doi.org/10.31436/iiumej.v26i2.3653

and they are particularly susceptible to rain attenuation [9]. Fig. 1 shows the Starlink satellite distribution in September 2024 from the Non-GEO Constellations Analysis Toolkit**,** with more than 6,000 active Starlink satellites [10]. The green code represents countries that have approved Starlink to operate, while the orange code indicates countries where SpaceX is obtaining approval**.**

Indirectly, those satellite signal link performances in Ku Band satellites operating beyond 10 GHz are subject to propagation loss, where rain attenuation is highly variable and, depending on the liquid water content, will impose a significant impact on signal link attenuation [11]. This propagation effect causes the receiving terminal signal level to be attenuated, causing the satellite link to suffer outages during rain [12].

Radio wave propagation is the process by which electromagnetic waves are transmitted from one point to another through the Earth's atmosphere or space. It exhibits reflection, refraction, diffraction, absorption, polarization, and scattering [13]. Atmospheric losses occur in the Earth's atmosphere due to energy absorption by the many atmospheric conditions, some of which are from adverse weather conditions. Weather-related losses are called atmospheric attenuation and absorption losses [14].

Annual precipitation data show that Peninsular Malaysia experiences the highest average annual rainfall compared to the neighbouring countries within the Asia Pacific region [15]. Malaysia experiences the highest rainfall rates in the world, frequently exceeding 200 mm/hr, attributed to its equatorial climate, which is characterized by high humidity and is further influenced by tropical monsoon patterns [16]. The climate results in a consistent rainfall distribution throughout the year, with heavy downpours often occurring in the afternoon due to the intense heat and moisture build-up during the day [17]. Two distinct monsoon seasons further shape the country's rainfall patterns. The Southwest Monsoon, which spans from late May to September, generally brings relatively lighter and less consistent rainfall. In contrast, the Northeast Monsoon, occurring between November and March, is associated with heavier and more sustained rainfall [16]. This monsoon primarily affects the eastern regions of Peninsular Malaysia, where heavy showers are known to occur in bursts, lasting an average of about an hour.

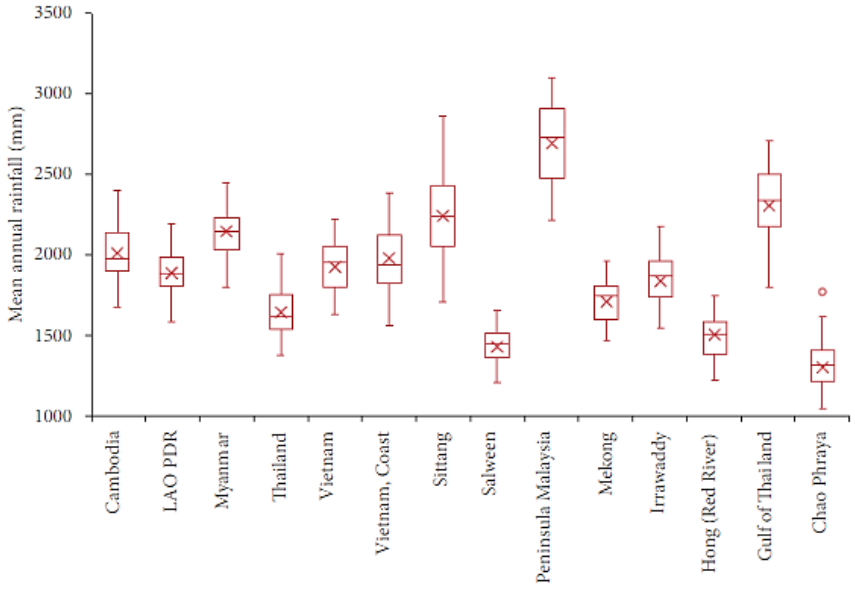


Figure 2. East and South Asia Precipitation Map [18]

https://doi.org/10.31436/iiumej.v26i2.3653

Understanding rainfall patterns, specifically in tropical regions, is crucial for developing robust strategies to mitigate the effects of the intensity and frequency of rain, ensuring resilience in critical infrastructure and services in the face of this natural climatic phenomenon. However, research on Low Earth Orbit (LEO) satellite constellations remains limited, particularly in addressing their performance analysis in tropical regions. Most existing studies on rainfall effects have only focused on Geostationary Earth Orbit (GEO) satellites, largely due to their historical dominance in satellite communication services over the past decades.

Latency is often called round-trip time (RTT), the time it takes to receive and return data between two nodes. GEO satellites, positioned at approximately 36,000 km, experience inherently higher latency as little as 480 ms round-trip due to their extended transmission path, making them more susceptible to prolonged degradation during heavy rain events [19], [20]. In contrast, Starlink satellites orbiting in low Earth orbit are positioned much closer to Earth, at an altitude of approximately 550 km. Starlink network benefits from the reduced signal path loss, adaptive beamforming, and dynamic satellite handovers, which may mitigate rain-related latency fluctuations [19], [21].

This research aims to determine, compare, and evaluate the latency effect under heavy rain conditions in the equatorial region using the cumulative distribution function (CDF) for LEO HTS constellations versus GEO HTS, which provides a probabilistic view of exceedance times for different latency thresholds. The method uses real satellite link paid internet services from the Starlink terminal and the SES 12 GEO HTS very small aperture satellite (VSAT) terminal installed onsite at the International Islamic University Malaysia satellite communication lab to measure continuous latency measurements for 6 months. This setup measures latency in Ku-band link from both LEO and GEO HTS exposed under a heavy rainfall region. Table 1 shows a key comparison summary for both HTS GEO & LEO Starlink [8], [22], [23], [24].

Table 1. Comparison between SES12 and Starlink specifications

Satellite Name	Type	Frequency band	Satellite Count	Lifespan	Nominal Latency	Data Throughput
SES-12	Geostationary (GEO) HTS - Specific wide beam covers the entire Asia-Pacific region	Ku-band (VSAT Terminal) & Ka-band (Gateway)	1	15 years	500-700 ms	Download of 50-100 Mbps & Upload of 10-20 Mbps
Starlink	Low Earth Orbit (LEO) mega constellation - Global coverage focusing on underserved and remote areas.	Ku-band (Starlink Terminal) & Ka-band (Gateway)	> 3000	5 to 7 years	20-40 ms	Download of 100-250 Mbps & Upload of 20-40 Mbps

2. METHODOLOGY

This research methodology is structured into two key phases: (1) Onsite Measurement and Data Collection, and (2) Onsite Performance Analysis. Each phase focuses on distinct aspects to achieve the research objectives.

A critical aspect of the applied method is developing and implementing an automated monitoring system for LEO Starlink and SES-12 GEO HTS VSAT terminals to measure latency and ensure proper data storage accurately. This automated monitoring system enables continuous, consistent, and reliable data collection for six months, facilitating the analysis of latency trends across varying weather conditions.

https://doi.org/10.31436/iiumej.v26i2.3653

2.1. Onsite measurement and data collection

LEO Starlink and SES-12 GEO HTS VSAT terminals are co-located and were installed at the satellite communication lab at the International Islamic University Malaysia (IIUM). This phase comprises setting up and testing satellite link services using Ku Band transmission from two different satellite systems: LEO Starlink and conventional SES-12 GEO HTS, a Geostationary Earth Orbit (GEO) satellite. A real-time gravity tipping bucket rain gauge is integrated into this setup to measure rainfall intensity accurately. It records data at one-minute intervals, providing high temporal resolution for detailed analysis. With a measurement resolution of 0.28 mm, the system ensures precise detection of even light rainfall events. Subsequently, satellite link latency was captured at 10 ms intervals and monitored continuously, serving as a primary parameter for understanding the behaviour of satellite communication links under tropical weather conditions. By implementing this automated monitoring approach, the methodology provides a comprehensive dataset for further analysis.

The LEO Starlink terminal is connected via the LEO Starlink Satellite network, as shown in Fig. 1, which is marked as an approved and activated region on the map. This LEO Starlink terminal is physically set up on a pole using a custom extended mount shown in Fig. 3. It performs an automatic alignment to locate and connect with Starlink satellites. The terminal will connect to the Starlink network and provide internet access. The Starlink terminal stays in a fixed position while tracking Starlink satellites using its advanced phased array antenna as shown in Fig. 3. Utilising a phased array antenna for electronically steered beam control, the Starlink terminal tracks satellites by its integrated a GPS receiver for its location with Starlink satellites' ephemeris data to predict satellite positions. Compared to a VSAT parabolic antenna, a phased array antenna scans faster due to electronic beam steering, while the parabolic antenna depends on the initial mechanical setup of its azimuth and elevation [25].

However, information regarding the azimuth and elevation angles of the Starlink terminal is not publicly available through its standard product dashboard. Therefore, the Starlink azimuth and elevation angles are obtained separately using the open-source Starlink-Grafana-Dashboard tool developed through outsourced scripting [26]. This tool was specifically created to capture the Starlink terminal's azimuth and elevation angles and to enable continuous monitoring of these parameters. Based on this setup for 6 months of observation, the azimuth and elevation angles measured in Kuala Lumpur, Malaysia, are approximately 4° and 76.5°, respectively.

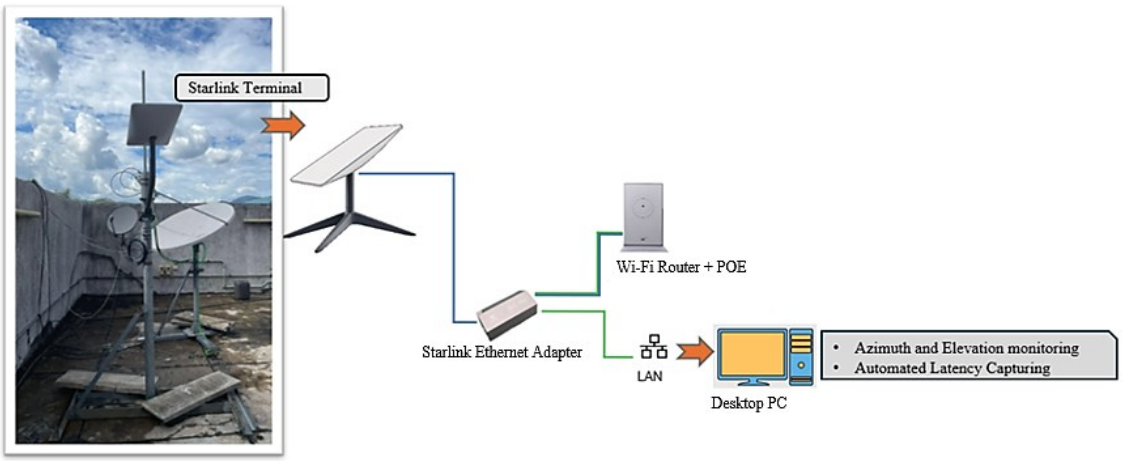


Figure 3. Onsite experimental arrangement for Starlink

https://doi.org/10.31436/iiumej.v26i2.3653

Fig. 4 shows the SES-12 GEO HTS Ku-band terminal setup. The VSAT terminal is connected to the SES-12 Satellite, which is positioned in a Geostationary orbit at 95° East with a fixed azimuth of 244.35° and elevation of 81.20° [27].

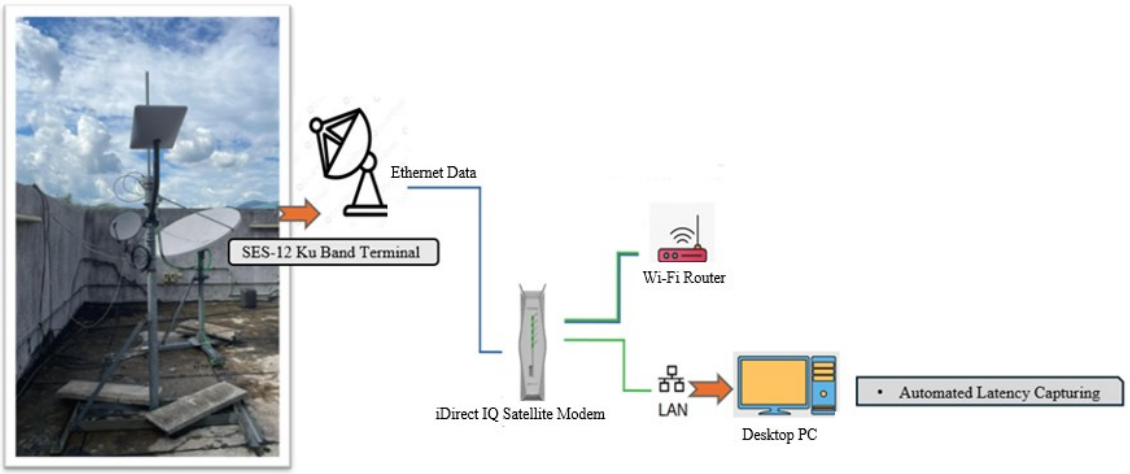


Figure 4. Onsite experimental arrangement for SES-12 GEO HTS

As shown in Fig. 5, this orbital slot allows it to provide coverage across the Asia-Pacific, specifically covering the Malaysian region in providing various communication services.

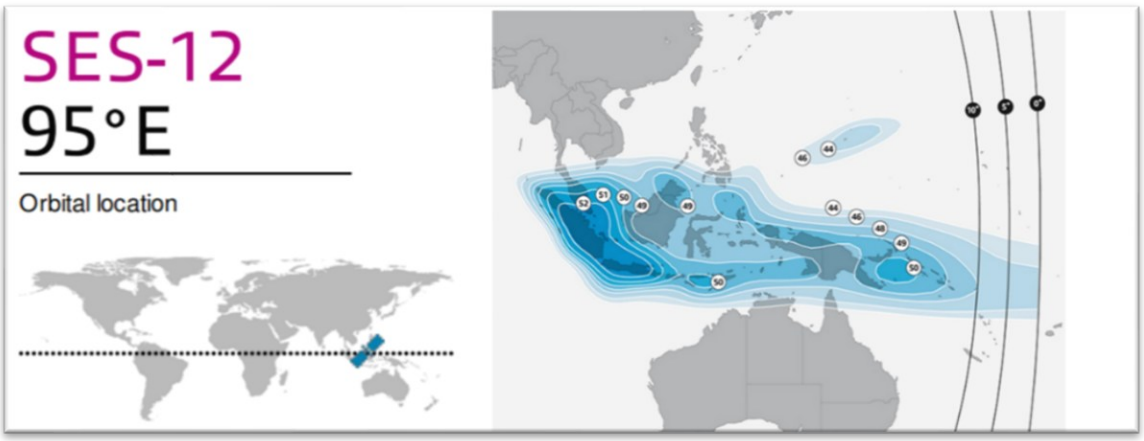


Figure 5. SES-12 GEO HTS satellite with Malaysian footprints [22]

A batch script was developed during this stage to automate the measurement of satellite link latency. The script was developed to ensure continuous operation, logging real-time data by sending ping requests to Google's DNS server (8.8.8.8) for both satellite links (SES-12 GEO HTS and LEO Starlink). An automated script runs continuously, measuring and recording latency results every 10 seconds. These results are saved in a file named ping_log.txt and are also displayed in real-time on the desktop PC, allowing users to monitor the data as it updates. Fig. 6 shows a simplified block diagram of the automated script process, where results are stored in Google Cloud Storage in a structured format for future analysis.

https://doi.org/10.31436/iiumej.v26i2.3653

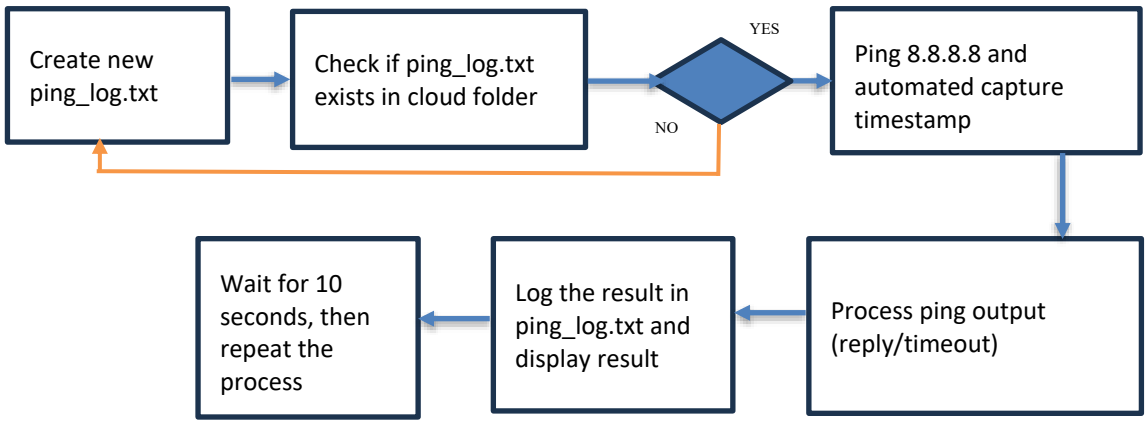


Figure 6. Flow Chart for Automated Latency Script

2.2. Data Analysis

This section focuses on the performance evaluation methodology of the new Low Earth Orbit (Starlink) HTS technology compared to the satellite communication HTS technology of Geostationary Earth Orbit (GEO). Empirical data collected were evaluated with respect to their critical parameters, such as latency and rainfall rate impact, resulting in their communication signal link reliability.

The study starts with data preprocessing, which involves cleaning and organizing the collected data, including rainfall intensity and satellite link delay (latency). Next, an analysis is done to find patterns between the rainfall and latency over time, and to create Cumulative Distribution Functions (CDFs) for both rainfall rates and latency for the LEO Starlink and GEO SES-12 satellite systems. Finally, the performance of both satellite links is evaluated by comparing their latency and connection timeouts during different rain conditions. The CDFs are used to calculate useful statistics that help compare how each satellite system handles heavy rain and how reliable they are overall. Given the tropical region's high rainfall intensity and frequency, the empirical observations provide insights into these technologies' operational limits and resilience under adverse conditions.

These three phases provide a comprehensive performance analysis, highlighting the strengths and limitations of SES-12 GEO HTS and LEO Starlink satellite technologies, specifically in tropical environments. Additionally, this analysis establishes a foundation for future research and the development of region-specific recommendations for satellite communication in high-rainfall areas.

3. EFFECTS OF RAIN ON LATENCY

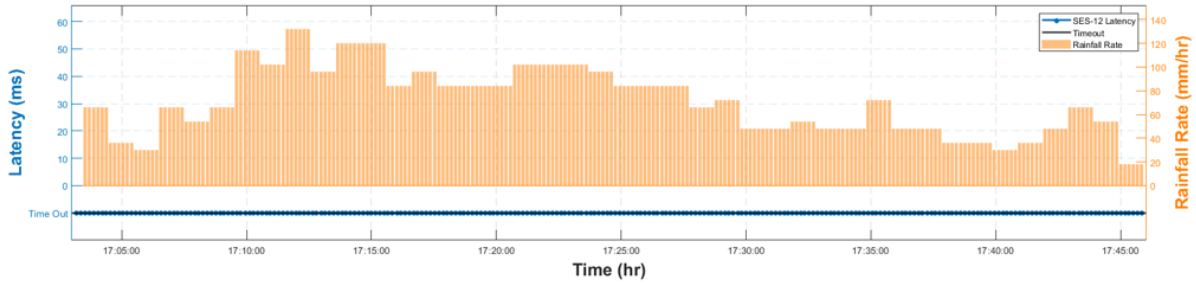
This section analyzes the performance of the SES-12 GEO HTS and LEO Starlink satellite services during rain events, focusing on how precipitation affects signal latency and discussing the experimental findings in detail. This section also discusses the correlation between rainfall rate and latency based on experimental data collected under various weather conditions. The study aims to evaluate the impact of rainfall intensity on the Starlink satellite internet system's latency, focusing on signal degradation caused by rain fade.

3.1. Rain conditions

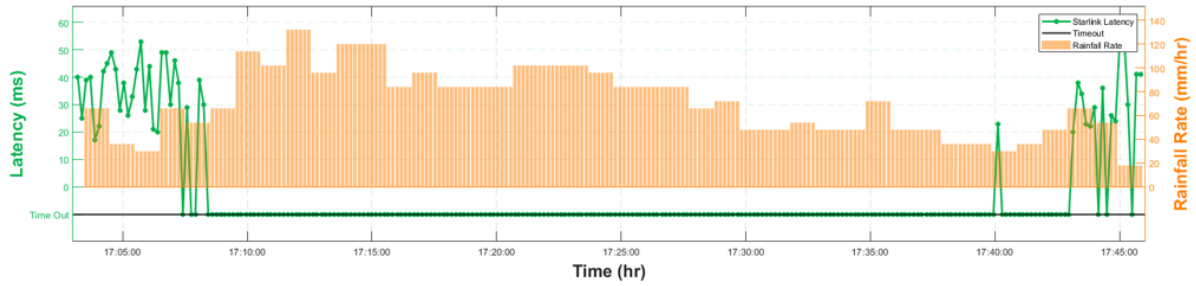
The graph consists of two subplots, each analyzing the latency performance of SES-12 GEO HTS and LEO Starlink under various rainfall conditions on August 22, 2024. Fig. 7(a)

https://doi.org/10.31436/iiumej.v26i2.3653

and 7(b) highlight the significant impact of heavy rainfall on the latency performance of Starlink and the SES-12 GEO HTS satellite services. High rainfall rate caused the SES-12 GEO HTS VSAT terminal to become fully disconnected at Fig.7(a). Rainfall rates consistently exceeded 100 mm/hr, peaking at 120 mm/hr, causing disconnections and latency spikes above 60ms, occasionally exceeding 100ms from the LEO Starlink terminal. This is attributed to raindrops' scattering and absorption of satellite signals, highlighting the vulnerability of LEO satellite links to rain fade (Fig.7 (b)).

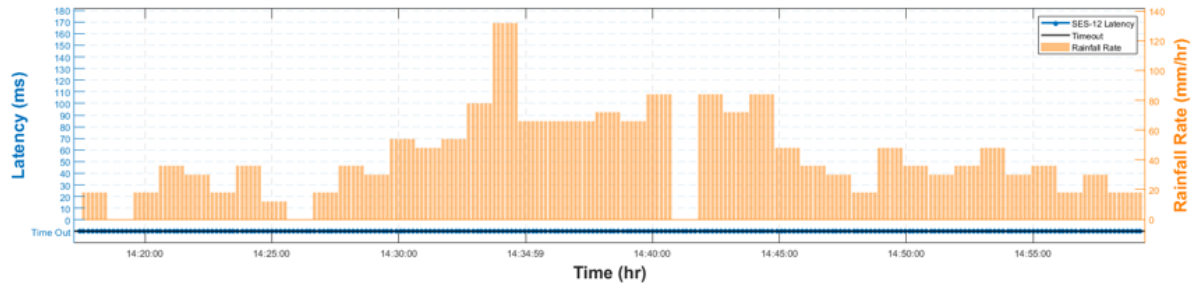


(a) SES-12 GEO HTS

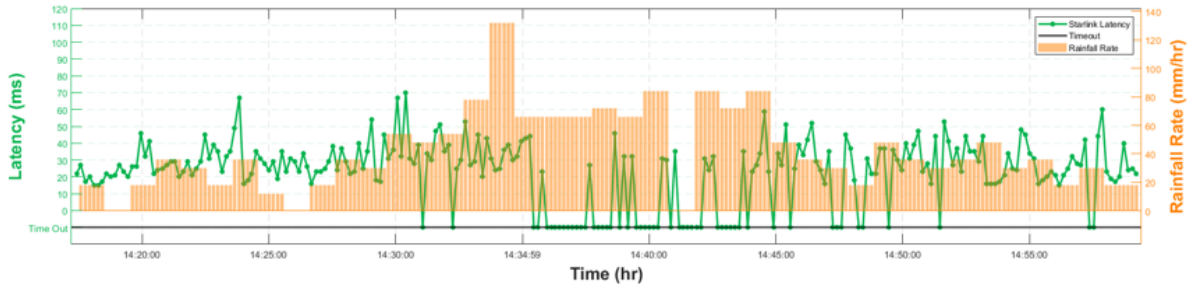


(b) LEO Starlink

Figure 7. A Rain Event occurred on 22 August 2024



(a) SES-12 GEO HTS



(b) LEO Starlink

Figure 8. A Rain Event occurred on 23 August 2024

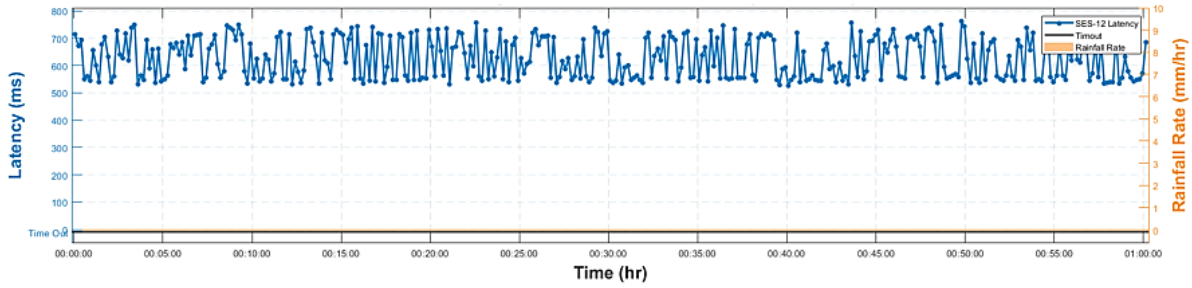
Fig.8(a) and 8(b) illustrate an August 23, 2024, event featuring prolonged medium to heavy rainfall with multiple peaks exceeding 100 mm/hr for SES-12 GEO HTS and LEO Starlink VSAT Terminal. During this period, as shown in Fig. 8(a), the SES-12 GEO HTS VSAT

https://doi.org/10.31436/iiumej.v26i2.3653

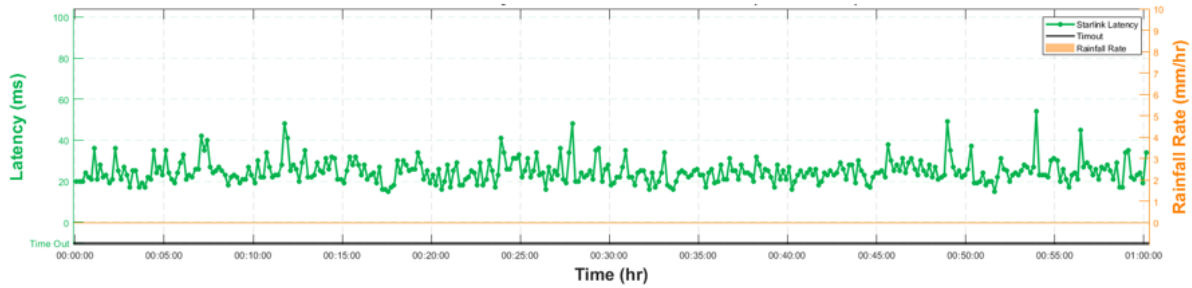
terminal still suffered a total loss of connectivity. The LEO Starlink satellite system experienced only a few disconnects during this period, as shown in Fig. 8(b). In summary, multiple rainfall rate conditions and rain events demonstrate a strong correlation between rainfall intensity and higher latency, highlighting the susceptibility of high-frequency satellite systems to adverse weather. These findings also confirm the vulnerability of Ku-band frequencies to high-intensity rainfall, even in a LEO satellite system mega-constellation closer to Earth. Rain fade causes signal degradation and higher latency. However, the LEO Starlink terminal maintained link availability at certain rainfall rates below 60 mm/hr. In contrast, the SES-12 GEO satellite system experienced complete disconnection.

3.2. Clear sky

The analysis in Fig. 9(b) highlights the stable performance of the Starlink satellite system during no rain events, where rainfall rates consistently remained near 0 mm/hr, indicating clear weather conditions. During these periods, latency for the LEO Starlink terminal showed minimal variability, remaining within a predictable range between 25ms and 35ms. This stable performance reflects the system's ability to operate optimally without atmospheric disturbances, with no signal attenuation or interference caused by rain fade. Compared to the GEO satellite services, which range between 550ms and 750ms in Fig. 9(a), Starlink demonstrates greater latency stability under clear skies, showcasing its robustness in favourable weather conditions.



(a) SES-12 GEO HTS



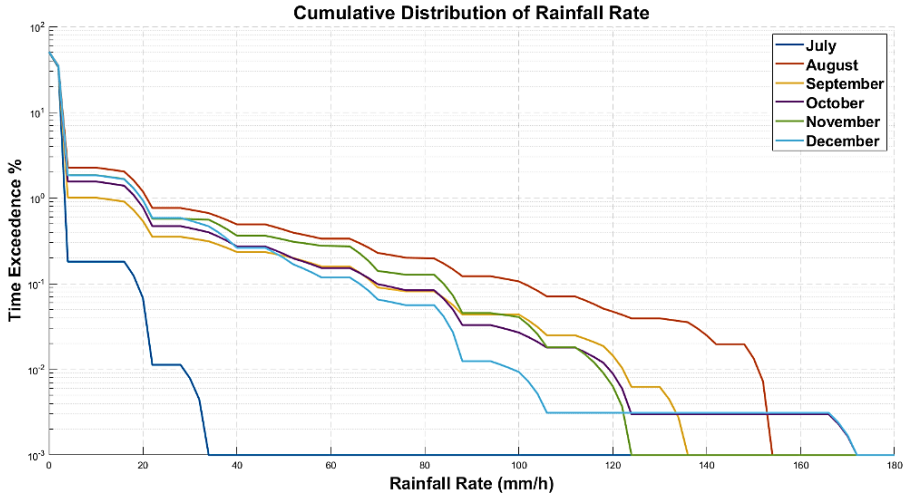
(b) LEO Starlink

Figure 9. A clear sky event occurred on 5 December 2024

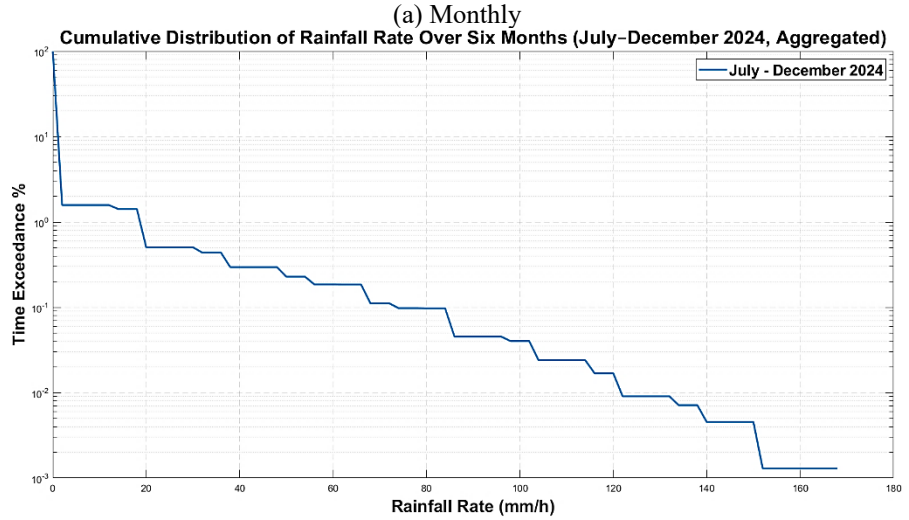
3.3. Cumulative distribution function (CDF) analysis of rain rate

The CDF analysis of rainfall, illustrated in Fig. 10(a) and 10(b), provides a detailed statistical evaluation of the distribution and frequency of rainfall rates across six months (July to December 2024) and accumulated. This analysis examines the percentage of time that specific rainfall rates are exceeded, offering insights into the occurrence and intensity of precipitation during the data collection period. Each month's rainfall distribution is represented by a distinct line, with the x-axis indicating rainfall rates (in mm/hr) and the y-axis showing the percentage of time exceedance.

https://doi.org/10.31436/iiumej.v26i2.3653



(a) Monthly



(b) Aggregated

Figure 10. CDF Analysis of Rainfall Rate

The cumulative distribution function (CDF) of the rainfall rate in Fig. 10(a) highlights seasonal variations in rainfall intensity. July experiences mostly light to moderate rainfall, with minimal heavy rain. From August to November, the probability of moderate-to-heavy rainfall increases, likely due to the monsoon season, as seen in the closely aligned curves. December records the highest extreme rainfall events, exceeding 160 mm/h, but these are rare. Most of the time, rainfall remains in the light-to-moderate range. Overall, September to November shows the most frequent heavy rain, while December, despite occasional extremes, has a lower time exceedance for intense rainfall. These seasonal variations directly impact Starlink's performance, particularly in the Ku-band range, as increased rainfall can cause signal attenuation, leading to higher latency, reduced data rates, and potential service disruptions during heavy rain periods. Though infrequent, heavy rainfall events above 100 mm/hr significantly degrade performance, leading to substantial signal attenuation and increased latency. Moderate rainfall, ranging between 40 and 100 mm/hr, also introduces signal degradation, but connectivity remains mostly intact. In contrast, light rainfall below 40 mm/hr has minimal impact, allowing for stable network performance.

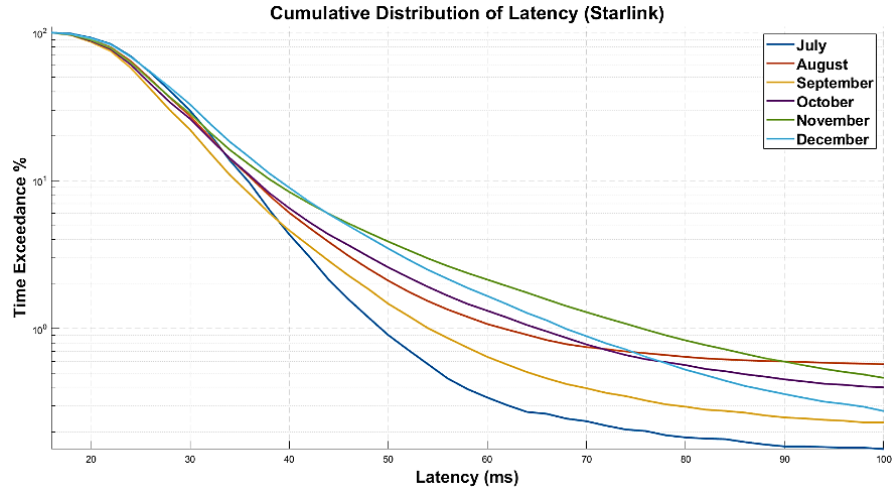
The data collected over six months is not to be directly comparable to ITU-R standards, which are based on long-term averages. However, since the data was gathered during the monsoon season, when rainfall is most intense, it still offers valuable insight. Therefore, the

https://doi.org/10.31436/iiumej.v26i2.3653

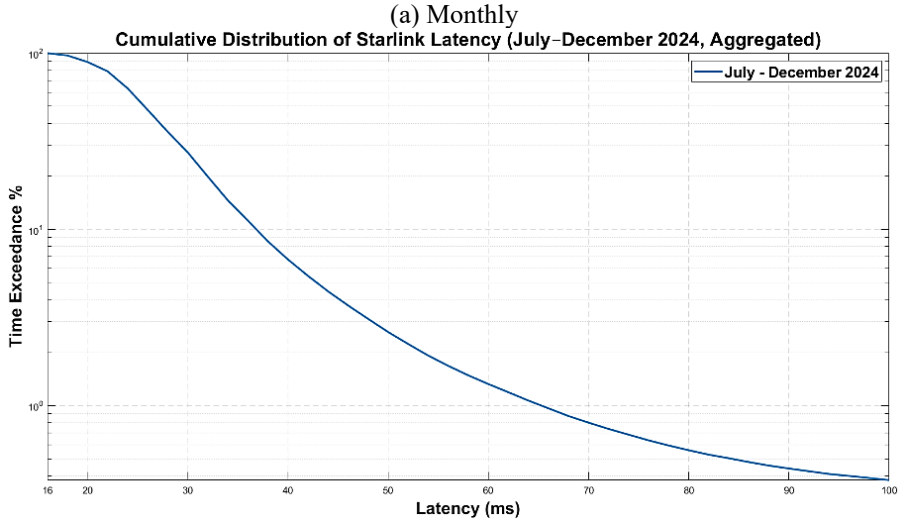
result shown estimated $R_{0.01}$ value from Fig. 10(b), ranging between 110–130 mm/h, is slightly higher than the value shown by the ITU-R P.837-7 model, which generalizes Malaysia's 0.01% exceedance rainfall rate as 100 mm/h [23]. This suggests that the measured data may reflect localised variations in rainfall intensity that exceed the ITU-R's generalised value. The study emphasises the importance of using tools like CDF to quantify the likelihood of various rainfall intensities and their correlation with signal degradation. Integrating CDF insights with latency data enables a comprehensive understanding of the relationship between rainfall intensity and Starlink system performance.

3.4. Cumulative distribution function (CDF) analysis of latency

The Cumulative Distribution Function (CDF) analysis of latency, as depicted in Fig. 11(a) and (b), provides a comprehensive evaluation of the percentage of time that specific latency thresholds were exceeded over a six-month period (July to December 2024). This analysis is crucial for assessing the variability and consistency of the Starlink satellite internet system's latency performance under different conditions. By examining the CDF curves, trends in latency distribution across multiple months can be identified, highlighting periods of increased or reduced network delays.



(a) Monthly



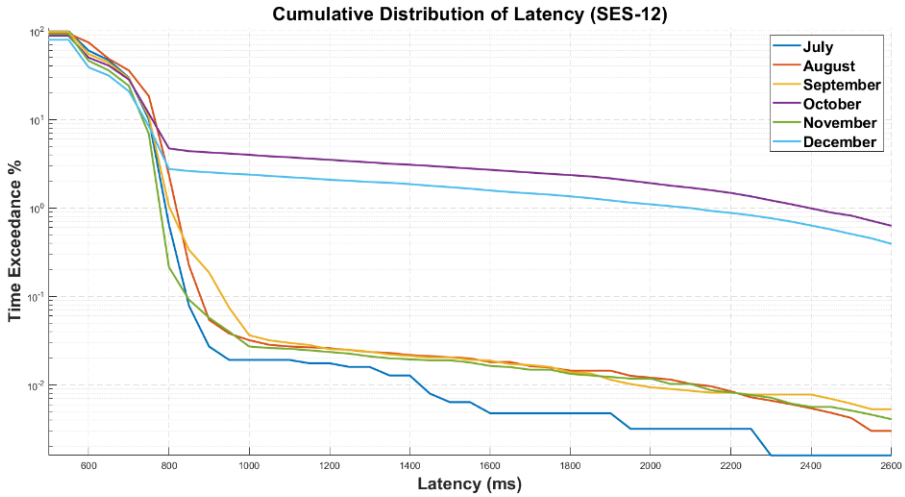
(b) Aggregated

Figure 11. CDF Analysis of Latency for LEO Starlink

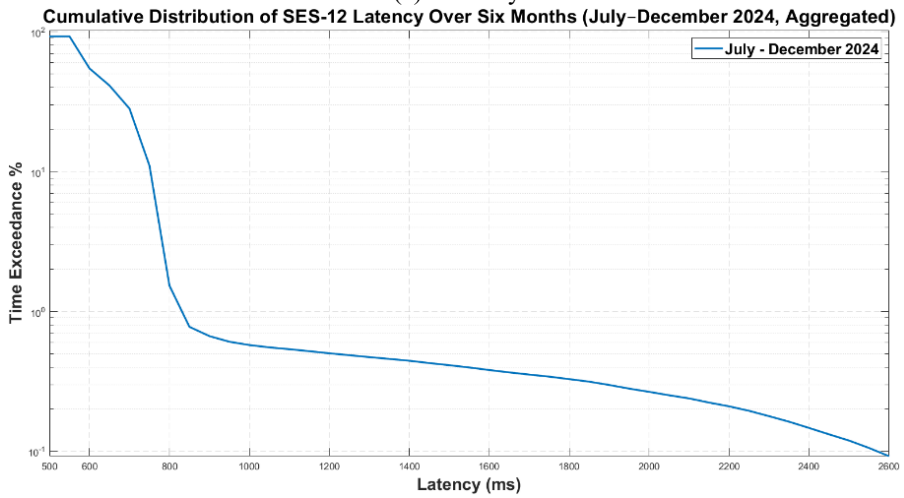
https://doi.org/10.31436/iiumej.v26i2.3653

Furthermore, this analysis helps to establish a correlation between latency fluctuations and environmental factors, specifically rainfall intensity. The CDF analysis highlights in Fig. 11(a) monthly trends in Starlink's latency performance, showing that latency averaged at 20ms about 90% of the time between July and August. From September to November, latency performance increases closer to 40ms at 90% of the time. These changes in latency performance align with higher rainfall intensities observed in the same period, where rain fade likely caused significant signal attenuation and increased latency.

December recorded the highest latency, with latency surpassing other months. Overall, the aggregated data show in Fig. 11(b) that average latency remained below 40ms for 90% of the time, highlighting the reliability of the Starlink system under varying weather conditions. Understanding these variations is essential for optimising Starlink's quality of service (QoS), predicting network behaviour, and improving system reliability for real-time applications.



(a) Monthly



(b) Aggregated

Figure 12. CDF Analysis of Latency for SES-12 GEO HTS

Meanwhile, for SES-12 GEO HTS, as shown in Fig. 12(a) and 12(b), the latency averages around 600ms for approximately 90% of the time across the observed six-month period. The cumulative distribution function (CDF) curves in Fig. 12(a) indicate that the latency remains below 1,000ms during most months, suggesting relatively stable performance under normal conditions. However, higher latency values exceeding 1,500ms are occasionally recorded, particularly in October and December, where the curves gradually deteriorate beyond this

https://doi.org/10.31436/iiumej.v26i2.3653

threshold. Notably, in December, the latency remains persistently high, with a significant portion exceeding 2,000 ms, likely due to the extensive rainy season causing network congestion and signal degradation in the GEO satellite link.

Overall, the findings established a clear correlation between rainfall intensity and latency performance, emphasising the vulnerability of high-frequency satellite signals to rain fade. The insights gained from this chapter provide a strong foundation for future research to enhance the resilience and reliability of satellite internet systems in varying environmental conditions.

3.5. Performance comparison

Fig. 13 shows a performance comparison graph between the SES-12 GEO HTS and LEO Starlink satellite systems, analyzing the timeout occurrences and site availability from July to December. The timeout percentage, represented by bars, indicates that Starlink in blue bars consistently experiences minimal timeouts across all months. In contrast, the SES-12 GEO HTS VSAT terminal in red bars shows significantly higher timeout occurrences, particularly in August, November, and December.

Onsite availability performance, represented by solid and dashed lines, shows that Starlink, in the solid blue line, maintains above 99% availability throughout the period, highlighting its reliability. In contrast, in the dashed red line, the SES-12 GEO HTS VSAT terminal exhibits a downward trend in its site availability, with a notable decline in December, when it drops below 90% due to heavy rain events.

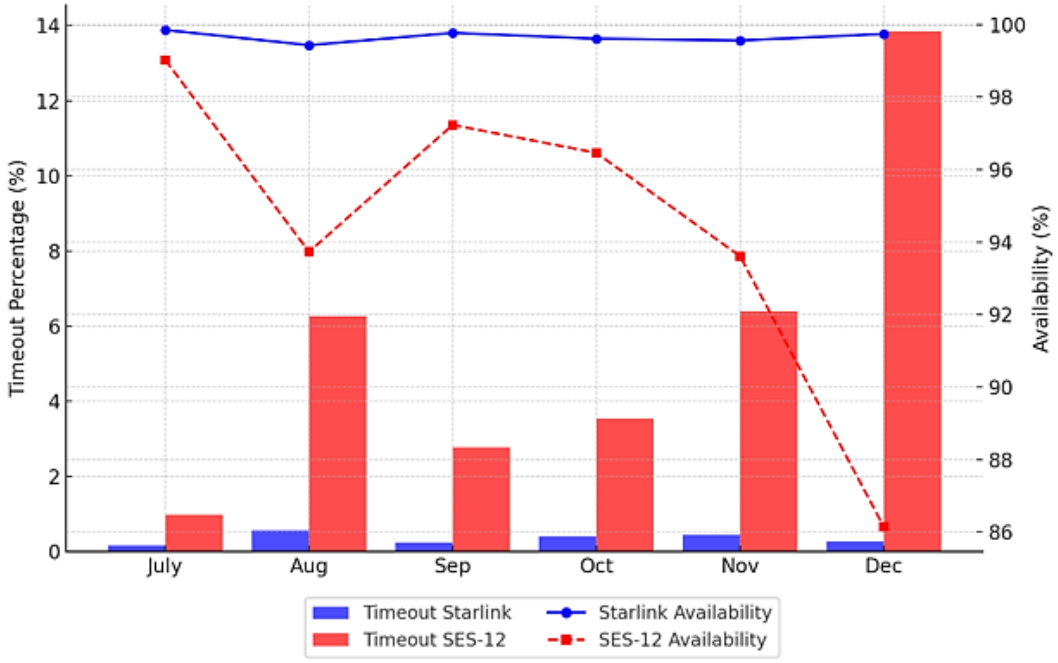


Figure 13. Monthly timeout and site availability

Table 2 summarizes the overall performance analysis, indicating that LEO Starlink satellite services maintain superior stability and availability despite using similar Ku-bands. In contrast, SES-12 GEO HTS Satellite services experience higher disruptions, especially during extreme weather. The LEO Starlink terminal shows temporary latency spikes but recovers quickly, whereas the SES-12 GEO HTS VSAT terminal suffers from persistent signal degradation due to rain events. Additionally, the LEO Starlink terminal has much lower latency, captured at 20ms to 100ms, compared to the SES-12 GEO HTS VSAT terminal, which

is 550ms to 1000ms, further emphasising its advantage in responsiveness. The mean site availability for the LEO Starlink terminal remains close to 100%. In comparison, the SES-12 GEO HTS VSAT terminal fluctuates significantly to 94% for 6 months of the analysis, highlighting its vulnerability to network disruptions in this tropical region.

Table 2. Summary of performance

Parameters	LEO Starlink terminal	SES-12 GEO HTS VSAT terminal
Measured Latency	~20-100ms	~550 -1000ms
Effect of Rain	Temporary spikes, but recovers fast	Persistent degradation
Site Availability	99.63% (6 months average)	93.99% (6 months average)

4. CONCLUSION

This study analyzed the impact of rainfall on Starlink's signal latency compared to the GEO satellite link signals. Over six months, heavy rainfall exceeding 120 mm/hr was recorded. This extreme weather condition resulted in significant signal degradation for both GEO and LEO satellite services, highlighting the susceptibility of satellite communications to intense precipitation. The findings emphasize the need for advanced mitigation strategies to enhance signal reliability during severe weather conditions. Latency performance evaluation shows that Starlink demonstrates superior satellite link availability, achieving 99.6% uptime onsite compared to 94% for SES-12 GEO HTS satellite services in the tropical region. This higher availability highlights Starlink's resilience and reliability, particularly in areas where consistent connectivity is critical. The difference in uptime suggests that LEO Starlink's satellite network is less susceptible to weather-related disruptions and signal latency issues that can impact traditional geostationary (GEO) satellite services like SES-12. Further analysis is required to develop a latency model that accounts for the effects of rain in tropical regions, particularly for the LEO Starlink satellite network.

ACKNOWLEDGEMENT

This project's funding comes from the International Islamic University Malaysia (IIUM) sponsored research project's grant number SPI24-224-0224.

REFERENCES

[1] O. Vidal, B. Roy, S. Dimitrov, R. Barrios, D. Giggenbach, and A. Le Kernec, "Next generation high throughput satellite system based on optical feeder links," *https://elib.dlr.de/*, 2015.

[2] F. Geng, D. B. Gomez, Y. Guan, and J. H. Saleh, "Monte-Carlo value analysis of high-throughput satellites: Value levers, tradeoffs, and implications for operators and investors," *PLoS One*, vol. 14, no. 9, 2019, doi: 10.1371/journal.pone.0222133.

[3] A. García Zaballos, W. Wagner, and R. Brazil David, *The Provision of Satellite Broadband Services in Latin America and the Caribbean*. 2016. doi: 10.18235/0012642.

[4] S. E. Hobbs, J. P. Sanchez, and J. Kingston, "Extended lifetime Laplace plane Geo SAR mission design," in *IET Conference Publications*, 2015. doi: 10.1049/cp.2015.1489.

[5] Jason Rainbow, "Amazon gets key FCC approval for more than 3,000 LEO broadband satellites," *Spacenews*, Feb. 08, 2023.

[6] S. Agnelli, P. Feltz, P. F. Griffiths, and D. Roth, "Satellite's role in the penetration of broadband connectivity within the European Union," in *2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop, ASMS/SPSC 2014*, 2014. doi: 10.1109/ASMS-SPSC.2014.6934560.

[7] Reuters, "Malaysia issues license to Musk's Starlink to bring internet services to remote areas," Reuters. Accessed: Mar. 09, 2025. [Online]. Available: https://www.reuters.com/technology/malaysia-issues-license-musks-starlink-bring-internet-services-remote-areas-2023-07-20/

[8] Alexander Wong, "MCMC says co-existence study not done for Starlink before green light given to operate," *Malay Mail*, Kuala Lumpur, Sep. 08, 2023.

[9] Y. Guan, F. Geng, and J. H. Saleh, "Review of high throughput satellites: Market disruptions, affordability-Throughput map, and the cost per bit/second decision tree," 2019. doi: 10.1109/MAES.2019.2916506.

[10] Christopher Baugh, "SpaceX Starlink and Amazon Kuiper look set to dominate the satellite connectivity market," https://www.analysysmason.com/research/content/articles/starlink-kuiper-connectivity-nsi040/. Accessed: Sep. 07, 2024. [Online]. Available: https://www.analysysmason.com/research/content/articles/starlink-kuiper-connectivity-nsi040/

[11] "ITU-R P.839-4," *Rain height model for prediction methods* , 2013.

[12] J. S. Mandeep and S. I. S. Hassan, "Slant-path rain attenuation predictions in tropical regions," *J Atmos Sol Terr Phys*, vol. 68, no. 8, 2006, doi: 10.1016/j.jastp.2006.01.009.

[13] Ajayi Taiwo Seun, "Mobile Satellite Communications," Blekinge Institute of Technology, 2007.

[14] "ITU-R P618.14," *Propagation data and prediction methods required for the design of Earth-space telecommunication syste*, vol. 8, 2023.

[15] M. Edirisinghe, N. Alahacoon, M. Ranagalage, and Y. Murayama, "Long-Term Rainfall Variability and Trends for Climate Risk Management in the Summer Monsoon Region of Southeast Asia," *Advances in Meteorology*, vol. 2023, pp. 1–12, Oct. 2023, doi: 10.1155/2023/2693008.

[16] N. H. H. Khamis, J. Din, and T. A. Rahman, "Rainfall rate from meteorological radar data for microwave applications in Malaysia," in *2005 13th IEEE International Conference on Networks jointly held with the 2005 7th IEEE Malaysia International Conference on Communications, Proceedings*, 2005. doi: 10.1109/ICON.2005.1635660.

[17] I. Shayea, T. Abd Rahman, M. Hadri Azmi, and M. R. Islam, "Real Measurement Study for Rain Rate and Rain Attenuation Conducted over 26 GHz Microwave 5G Link System in Malaysia," *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2810855.

[18] UN Office for the Coordination of Humanitarian Affairs (OCHA), "Asia Pacific Regional Reference Map: Annual Precipitation." Accessed: Mar. 09, 2025. [Online]. Available: https://reliefweb.int/map/world/asia-pacific-regional-reference-map-annual-precipitation

[19] J. Conde, G. Martínez, P. Reviriego, and J. A. Hernández, "Round Trip Times (RTTs): Comparing Terrestrial and LEO Satellite Networks," in *2024 27th Conference on Innovation in Clouds, Internet and Networks (ICIN)*, IEEE, Mar. 2024, pp. 42–46. doi: 10.1109/ICIN60470.2024.10494421.

[20] M. Beyaz, "Satellite Communications with 5G, B5G, and 6G: Challenges and Prospects," *International Journal of Communications, Network and System Sciences*, vol. 17, no. 03, pp. 31–49, 2024, doi: 10.4236/ijcns.2024.173003.

[21] Starlink, "Satellite Technology." Accessed: Feb. 22, 2025. [Online]. Available: https://www.starlink.com/technology

[22] "SES-12 at 95 degree East," https://www.ses.com/our-coverage#/explore/satellite/365.

[23] Y. A. Ahmad, A. F. Ismail, M. N. A. Abdul 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: 10.31436/iiumej.v25i2.3279.

https://doi.org/10.31436/iiumej. v26i2.3518

THEORETICAL STUDY OF ELLIPTIC DRUM OF VERTICAL SPINDLE COTTON PICKER

SHAVKAT RAVUTOV[1], ABDUMALIK RAKHIMOV[2*], GAFURJAN IBRAGIMOV[3,4,5]

[1]Machine building, Tashkent State Technical University, Tashkent, Uzbekistan,
[2]Department of Science in Engineering. International Islamic University Malaysia, Malaysia,
[3]Dynamical systems and their applications, V.I. Romanovskiy Institute of Mathematics, Uzbekistan Academy of Sciences, Tashkent, Uzbekistan,
[4]Department of Econometrics, Tashkent State University of Economics, Tashkent, Uzbekistan,
[5]Department of International Scientific Journals and Rankings, Alfraganus University, Tashkent, Uzbekistan.

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

(Received: 22 November 2024; Accepted: 26 March 2025; Published online: 15 May 2025)

ABSTRACT: The article focuses on the structural optimization of the planetary mechanism of the spindle drum of a vertical-spindle cotton-picking machine in Uzbekistan and discusses the problems of synthesizing a new vertical-spindle drum that significantly enhances the efficiency of using spindles and increases the productivity of the machine. Utilizing the interaction characteristics of drum spindles with cotton plants and ensuring the effective functioning of spindle friction drives, it is theoretically justified that the trajectory of spindle relative motion has an elliptical shape. Special attention is given to the shape and dimensions of the directional path, ensuring the elliptical movement of the spindles. An analytical expression is obtained, and a convenient calculation algorithm is developed to automate the investigation of the influence of the dimensions of drum mechanism parts on the directional path shape and its dimensions. The theoretical results allow for preliminary conclusions regarding the dynamic processes occurring during drum operation and provide recommendations for selecting the directional path shape and dimensions.

ABSTRAK: Artikel ini memberi tumpuan kepada pengoptimuman struktur mekanisme planet bagi dram gelendong mesin pemetik kapas gelendong-menegak di Uzbekistan, serta membincangkan permasalahan dalam mensintesis dram gelendong menegak baharu yang dapat meningkatkan kecekapan penggunaan gelendong dan produktiviti mesin secara signifikan. Dengan mengambil kira ciri interaksi antara gelendong dram dan pokok kapas serta memastikan keberkesanan operasi pemacu geseran gelendong, didapati secara teori bahawa trajektori pergerakan relatif gelendong berbentuk elips. Perhatian khusus diberikan kepada bentuk dan dimensi arah laluan yang menjamin pergerakan elips gelendong. Satu ungkapan analitik telah diperoleh dan algoritma pengiraan yang sesuai telah dibangunkan bagi tujuan automasi penyelidikan terhadap pengaruh dimensi komponen mekanisme dram terhadap bentuk dan dimensi laluan arah. Dapatan teori yang diperoleh membolehkan rumusan awal dibuat berkaitan proses dinamik yang berlaku semasa operasi dram dan seterusnya memberikan cadangan dalam pemilihan bentuk serta dimensi laluan arah.

KEYWORDS: *Cotton picker, vertical spindle, planetary mechanism, directional path.*

1. INTRODUCTION

About 80% of the world's cotton is produced in India, China, the United States, Pakistan, Uzbekistan, Turkey, Brazil, Greece, and Egypt. The successful application of machinery in cotton harvesting operations depends on selecting appropriate mechanisms for the working

https://doi.org/10.31436/iiumej. v26i2.3518

parts of cotton pickers as shown in Fig.1. In addition, it is always necessary to address the analysis of the system, including geometric, kinematic, and dynamic aspects, and continuous improvement of its characteristics. The activities and research conducted in this regard are extensive and diverse in the field of application.

We will dwell on several studies conducted on mechanisms closely related to the vertical-spindle cotton-picking machine in terms of their structural configuration and working principles. Note that the two-shaft module is widely used in light industrial technological machines. The work is devoted to elucidating the principles of distributing friction forces in a dual-shaft module. Its working shaft covers are made of materials with different hardness, which are widely employed in various industrial sectors. In addition, in [1], the advantages of an asymmetric two-shaft module are substantiated.

In [2], the issue of utilizing toothed-lever differential mechanisms in technological machines was discussed. In this technological process, the distance between the axes of the working shafts changes with a roller. Such mechanisms are widely used in agricultural machinery and light industrial technological equipment. In addition [2], information about the principle of operation of the new device developed by the authors is presented. The possibility of reducing the variability of the kinematic characteristics of side-by-side shafts by using the device was also justified.

Figure 1. Vertical spindle Cotton Picker.

The work [3] studied a comparative analysis of two types of gear-lever transmission mechanisms in agro-technological machines with drum working bodies. For this, the centroid method was used to derive an analytical expression for the dependence of the kinematic chain parameters of the mechanism. Based on this, the kinematic parameters of the toothed chain were obtained, and computational experiments were conducted. The obtained results established that their transmission ratios also change with the change of the distance between the axes of the leading and driven shafts. The work [4] is devoted to the mathematical modeling of the shape of contact curves of two-roller technological module rollers. The proposed contact curve equation allows the technological module's theoretical research under different operating conditions.

The following works dedicated to the improvement and increasing reliability of the spindle drums of the vertical-spindle cotton picking machine are also noteworthy. One of the factors that negatively affects the completeness of picking in the working process of a vertical spindle cotton-picking machine is the bending of cotton plants along the direction of movement of the

machine. The bending of the cotton into the working slot causes the cotton bolls to be blocked by the branches from the spindles, limiting the activity of the spindles. To solve this problem and increase the machine's working speed by 20-25%, the authors of the work [5, 6] proposed a mechanism for inserting the cotton into the working slot for the cotton picking machine. They also attempted to justify its parameters theoretically. However, the experiments on this apparatus show that its reliability is much reduced because of the complexity of the drum construction.

The spindle drives of the vertical spindle cotton picking machine are reversible friction drives. Instability in the spindle's speed modes negatively affects the quality of cotton picking. The authors of the work [7] proposed a method of experimental determination of the actual speed regimes of spindles and gave recommendations for improving the drive. The work [8] experimentally studied the mechanism of friction and the deteriorating processes occurring in the friction drives of the spindles of the vertical-spindle cotton picker to increase the durability of drive belts. It also proposed covering the working surfaces of drive rollers with nanocrystals using the ultrasonic method. Experimental studies have shown that the durability of drive belts increases due to the reduction of the roughness of the working surfaces.

The results of the works [9, 10] dedicated to the theoretical study of the kinematic modes of the elliptical drums of the cotton picking machine with a vertical spindle show that this drum construction is promising. They also indicate that research work should be continued in this direction. In particular, the work considers the issue of synthesizing a structurally developed mechanism that increases the contact zone of spindles with cotton. This mechanism allows for the optimal values of their kinematic modes to be chosen based on the characteristics of the technological process of harvesting and the results of theoretical research.

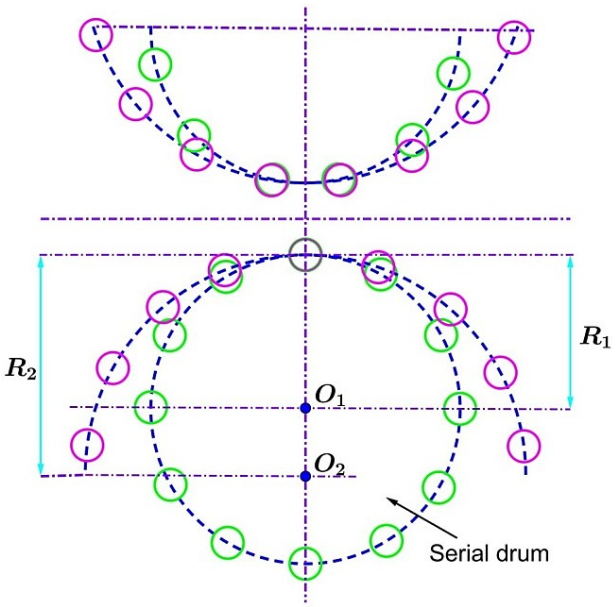


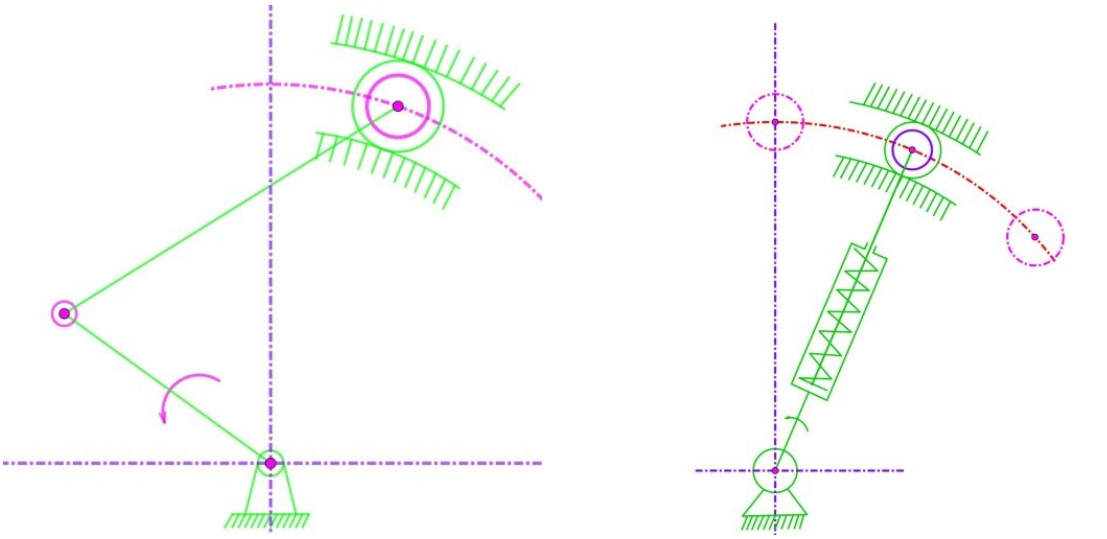
Figure 2. Using a large diameter drum to increase the contact area of the vertical spindle drum spindles with the cotton.

According to the requirements of the technological process of cotton picking, the distance between the centers of the spindles on the drum of vertical spindle cotton picking machines should not be less than 60 mm [11]. This factor limits the number of spindles in a drum being mass-produced. To increase the number of spindles involved in the cotton harvest at one time, it will be necessary to increase the diameter of the drums. However, the metal and energy

https://doi.org/10.31436/iiumej. v26i2.3518

consumption of the cotton picking machine is high, and it is difficult to fit the machine equipped with large diameter drums between the rows of cotton as shown in Fig. 2.

As a solution to this problem, in the 1980s in Uzbekistan, the idea emerged to increase the number of spindles simultaneously involved in harvesting and expand the zone of direct contact of the working bodies with cotton. This was achieved by structurally developing the mechanism of the spindle drum of the vertical spindle cotton machine [12-14]. The idea is to change the spindles' trajectory relative to the drum apparatus from a circle to an ellipse. Fig.3 shows the structural development options of the serial drum mechanism.



(a) replacement to a crank-slide mechanism (b) replacement to a telescopic mechanism

Figure 3. Options for replacing the drum mechanism with a lever-crank planetary mechanism.

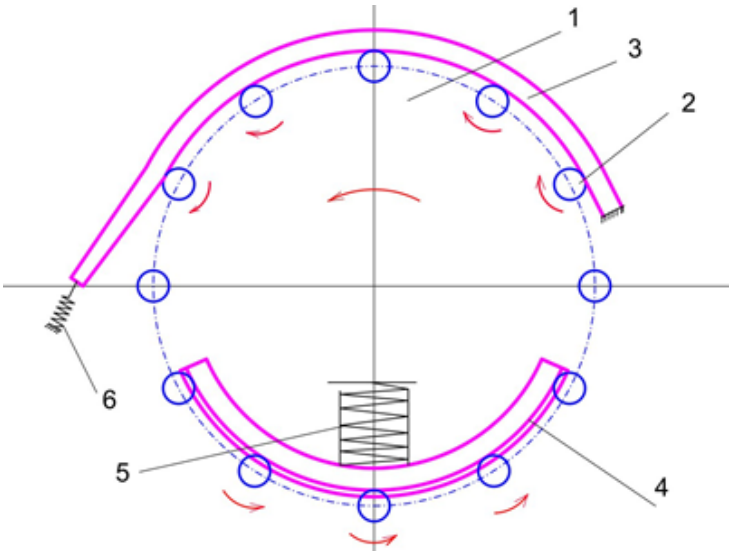


Figure 4. Series vertical spindle drum scheme: 1 is the drum; 2 is the spindle; 3 is drive belt; 4 is a reversible pad; 5 and 6 are springs.

The preparation and testing of both variants showed that option (a) was more suitable [12, 14]. There are some problems with the preparation and its use as a cotton picking drum for option (b). This can be explained, firstly, by the fact that the number of excess connections in the mechanism increases and requires very high precision in its preparation. Secondly, using a

telescopic mechanism at the bottom of the drum will not fit. With this in mind, we choose option (a).

Before considering the drum based on the mechanism with the developed structure, we will briefly review the design and technological features of the currently used vertical spindle drum. In a serial drum, the spindles are positioned so the lower and upper discs can rotate around their axes. They are the planetary mechanisms' satellites, which move in reverse rotation around their axes during the process (Fig. 4).

In the proposed design, an additional link is added to the serial drum mechanism, which is converted into a planetary-claw-clamp mechanism (Fig.5). This allows the spindles to move closer to the rows of cotton.

Further, the paper is organized as follows: Section 2 presents the equation for the directing ditch and discusses some features of the mechanism in designing an elliptical drum for the distinct values of parameters. In Section 3, we present the main conclusions.

2. SYNTHESIZING THE DIRECTING DITCH

According to the structure, the mechanism can be conditionally considered as a curved-sliding mechanism in which the slide moves in a curved line. Let $OA = r$ be a crank, $AB = l$ be a rod, and C be a slider (Fig. 5).

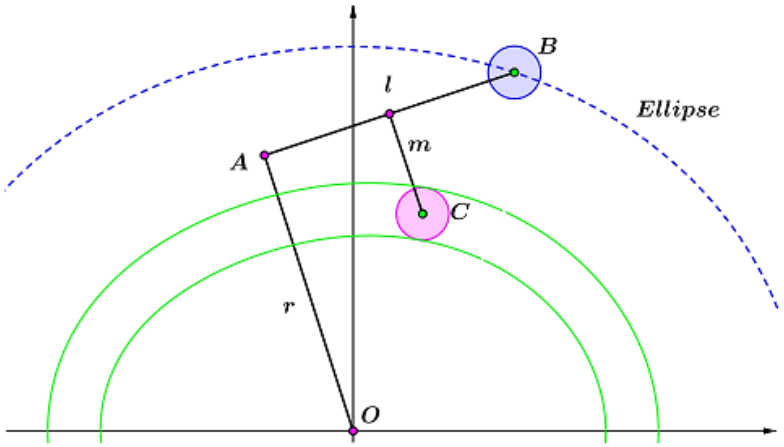


Figure 5. Additional link insertion scheme for serial drum mechanism.

In practice, this mechanism allows the desired trajectory of the spindles to be obtained. However, following the requirements of cotton picking technology using vertical spindles, it is recommended that the line forming the trajectory of the spindles be smooth, closed, and convex. In addition, to ensure the satisfactory operation of the friction drives of the spindles, the same conditions should be met [14]. As such, the ellipse was the most suitable [12]. Because of this reason, the proposed drum is called an elliptical drum.

By the condition, the point B must move along the ellipse when the point A of the crank r moves along a circle in the direction shown in Fig. 5. The results of the structural analysis of this mechanism show that when the point B moves along the ellipse, the point C, which is at a distance m from the axis of the connecting rod l, moves along a line different from the ellipse. We determine the analytical expression of the line drawn through the point C.

From the point of view of construction, the mechanism can be considered one whose roller moves along a curve (Fig. 5). We can find the spindles' trajectories for such mechanisms.

https://doi.org/10.31436/iiumej. v26i2.3518

Let the point A moves with an angular speed equal to l on the circle of radius r centered at the point O ($OA = r$) (Fig. 5) in a counterclockwise direction. We assume that the center of the spindle B with $AB = l$ moves along a specified ellipse. This ellipse is the curve passing through the center of the directing ditch. However, it is important to facilitate the service of spindles (B) to moving rollers (C) and to prevent roller fractures that may occur due to the increased distance from the boundary of the roller to its support. Therefore, it is expedient to install the directional roller (Fig. 6, point C) and the directional ditch close to the crank plane.

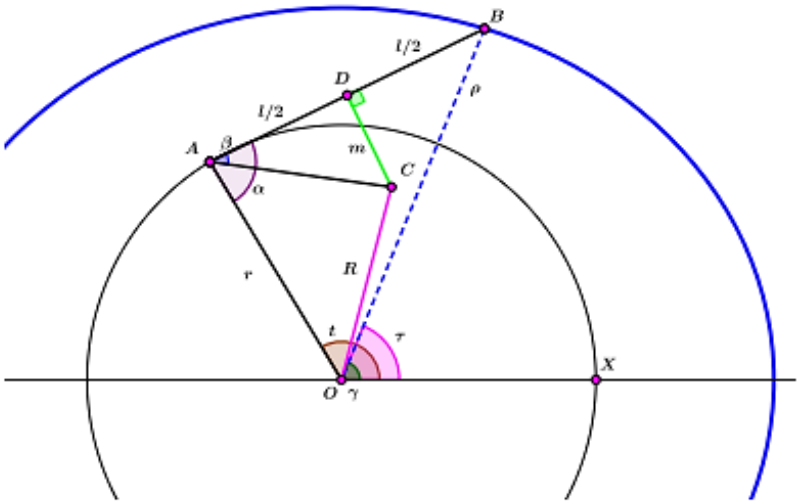


Figure 6. Scheme for determining the shape of the directing ditch.

The results of the structural analysis of this mechanism show that when the point B moves along the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,\ a > b \tag{1}$$

where a and b are given numbers, and they are the lengths of half of the major and minor axes of the ellipse, respectively, the point C moves along a curve that differs from the ellipse. Let us find an analytic formula that describes this curve, that is, the trajectory of the point (C).

For constructional reasons, the point C is at a distance DC=m from the connecting rod AB along its perpendicular bisector (Fig. 6), that is, $AB \perp DC$ and $AD = DB$. The length of the connecting rod AB is l.

Denote

$$\angle BOX = \tau,\ \angle AOX = t,\ \angle DAC = \beta,\ \angle COX = \gamma,\ OB = \rho,\ OC = R \tag{2}$$

where the points O and X are in a horizontal line. We assume that the angle τ is given, and we find the coordinates of the point C. We have, the coordinates of the point B,

$$x_B = \rho Cos\tau,\ y_B = \rho Sin\tau \tag{3}$$

Since the point B moves along the ellipse (1), substituting x_B and y_B in (3), into the equation (1) we have

$$\frac{\rho^2 Cos^2\tau}{a^2} + \frac{\rho^2 Sin^2\tau}{b^2} = 1, \tag{4}$$

where $\rho = \frac{ab}{(a^2 Sin^2\tau + b^2 Cos^2\tau)}$, which is the equation of the ellipse, Eq. (1), in polar coordinates.

https://doi.org/10.31436/iiumej. v26i2.3518

Since $\angle AOB = t - \tau$ and $AB = l$, we obtain from the triangle AOB by the law of cosines that

$$l^2 = r^2 + \rho^2 - 2r\rho Cos(t - \tau). \tag{5}$$

This implies that $t = \tau + arcCos\frac{r^2+\rho^2-l^2}{2r\rho}$. Next, we find the angle $\angle OAB = \alpha$. To this end, we find from the triangle AOB by the law of cosines that

$$\alpha = arcCos\frac{r^2+l^2-\rho^2}{2rl} \tag{6}$$

We obtain from the right triangle ACD that

$$\beta = arctan\frac{2m}{l} \tag{7}$$

Applying the law of cosines to the triangle AOC gives

$$R^2 = OA^2 + AC^2 - 2OA \cdot AC \cdot Cos\,(\alpha - \beta) \tag{8}$$

where $OA = r$ and $AC = \frac{1}{2}\sqrt{l^2 + 4m^2}$. Therefore, using Eq. (8),

$$R = (r^2 + \frac{l^2}{4} + m^2 - r\sqrt{l^2 + 4m^2}\,Cos\,(\alpha - \beta))^{1/2} \tag{9}$$

Next, to determine $< COX = \gamma$ we consider the triangle AOC. Since $< AOC = t - \gamma$, by the law of cosines, we obtain $cos(t - \gamma) = \frac{R^2+r^2-AC}{2rR}$. And, hence,

$$\gamma = t - arccos\frac{R^2+r^2-AC^2}{2rR} = t - arccos\frac{4R^2+4r^2-l^2-4m^2}{8Rr} \tag{10}$$

Finally, we find the coordinates of the point C as shown in Equation (11)

$$x_C(\tau) = R(\tau)Cos\gamma(\tau), \qquad y_C(\tau) = R(\tau)Sin\gamma(\tau) \tag{11}$$

We discuss why x_C and y_C depend on τ in these formulas. By (4) ρ is expressed in terms of τ, that is, $\rho = \rho(\tau)$. Equation (5) shows that time t depends on τ and ρ , and so by the (4) parameter t can be expressed only in terms of τ, that is, $t = t(\tau)$. The equation (6) allows us to conclude that α depends on τsince ρ is expressed in terms of τ. Equation (7) shows that β is constant, since r and l are given numbers. By equation (9) R depends on α, and so R can be expressed in terms of τ, that is, $R = R(\tau)$. Finally, by equation (10) γ can be expressed in terms of τ, that is, $\gamma = \gamma(\tau)$.

$$\rho = \rho(\tau), \quad t = t(\tau), \quad \alpha = \alpha(\tau), \quad R = R(\tau), \quad \gamma = \gamma(\tau) \tag{12}$$

In reality, τ varies depending on t. But here we are using the inverse relationship, that is, t varies dependently on τ. This is because we want to obtain only the trajectory of the point C.

Note that if τ varies depending on t, then the radius OA rotates in anti clockwise direction with unit angular speed, and the angular speed of OC. In general, it is not constant. Conversely, if t varies dependently on τ, then the radius OC rotates in anti clockwise direction with unit angular speed, and the angular speed of OA. In general, it is not constant. To find the trajectory of the point C, both options work and therefore, we can use the option where t varies depending on τ.

https://doi.org/10.31436/iiumej. v26i2.3518

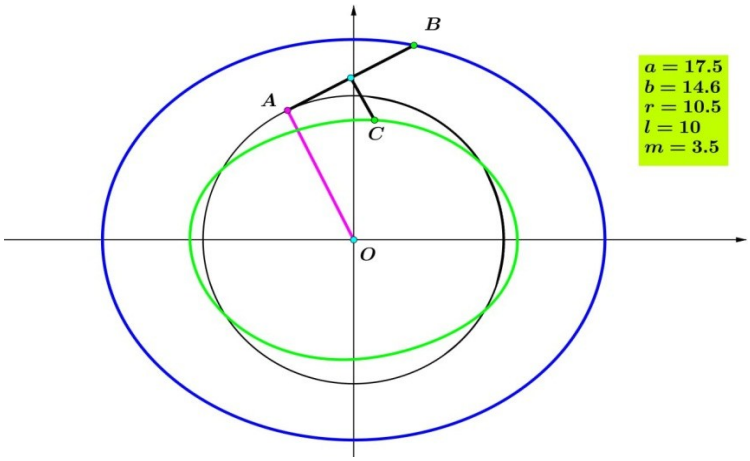


Figure 7. Directing the ditch when $a = 17.5,\ b = 14.6, r = 10.5, l = 10, m = 3.5$.

The elliptic drums furnished by the cotton picking mechanism must move along the cotton plant rows, and this circumstance is reflected in traditional circular drums. To guarantee this condition in a new elliptic drum, we make the half axis of the ellipse equal to the radius of the traditional circular drum, that is, $b = R$ =14.6 cm. Note that the diameter of the circular drum for the current cotton picking machine is $D = 2R$ = 29.2 cm. For the various values of the parameter a from the interval (14.6 cm, 20 cm), we study the trajectory of the point C.

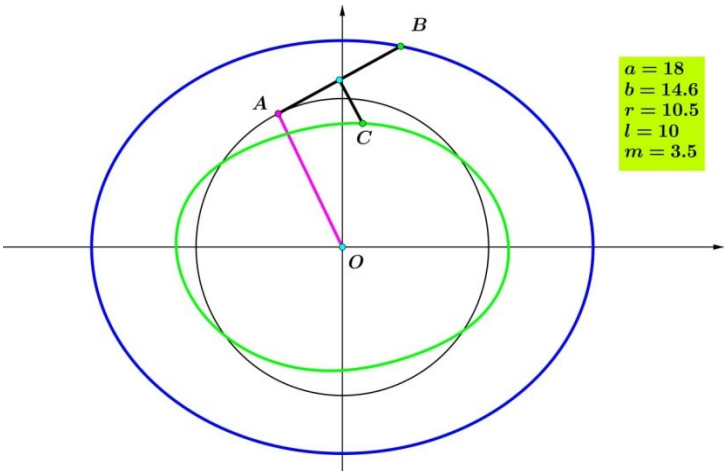


Figure 8. Directing the ditch when $a = 18,\ b = 14.6, r = 10.5, l = 10, m = 3.5$.

From the point of view of construction, we choose the length of the connecting rod AB of the elliptic drum to be equal to l=10 cm, the radius of the crank OA to be equal to r =10.5 cm, and the distance of the roller C from the connecting rod to be equal to m =3.5 cm. The numerical experiment was conducted for the distinct values of the semi-major axis a of the elliptical drum. The trajectories of the point C are illustrated in Fig.7 for a=17.5 cm, in Fig.8 for a =18 cm, and in Fig.9 for a=19 cm.

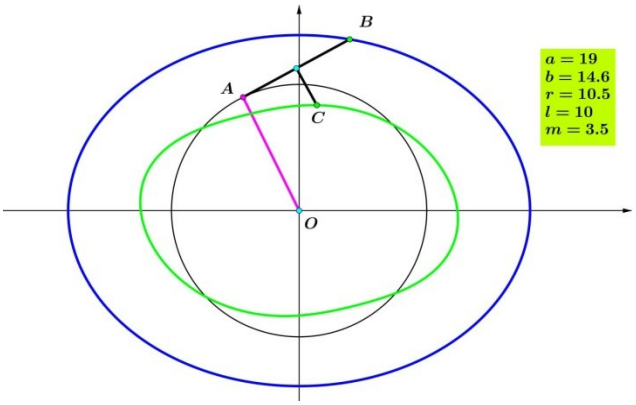


Figure 9. Directing the ditch when $a = 19,\ b = 14.6, r = 10.5, l = 10, m = 3.5$.

In the mechanism of the drum, the arc of trajectory of the spindle, where the spindle contacts the cotton plant, can be enlarged by increasing the major axis of the ellipse a. We can see, however, that if we increase the major axis of the ellipse, the shape of the trajectory of the point C gets to differ from the ellipse. Also, the difference between the curvatures of the spindle's trajectories gets larger in the picking and passage zones. This may cause an additional vibration and complicate the drum's operation. Therefore, it is important to consider the mechanism's features when designing an elliptical drum.

3. CONCLUSION

In the present paper, we have theoretically studied the elliptic drum of a vertical spindle cotton picker. The study allows us to draw the following important conclusions.

1. Through the structural development of the mechanism of the cotton picking drum with a vertical spindle, it is possible to develop a new (elliptic) drum that increases both the contact zone of the spindles with the cotton and the number of spindles participating in picking.
2. An effective analytical expression describing the change of the shape and dimensions of the guide ditch was obtained for the structurally developed mechanism of the elliptical drum, depending on the dimensions of other links of the mechanism.
3. Mathematical software allows for the automation of calculations and the visual observation of the size and shape changes that occur in the mechanism guide ditch when the geometric dimensions of the mechanism links change. It is an effective tool at the drum design stage.
4. Calculation experiments using the obtained analytical expression show that the value of the semi-major axis of the ellipse in the elliptical drum increases. In addition, the trajectory of the point C (guide ditch) deviates from the shape of an ellipse (quasi-ellipse), and the differences in the guide ditch curvature radii in the places corresponding to the movement trajectories in the pick-up and transition zones of the spindles get larger.
5. The formation of differences in the radii of curvature of the guide ditch of the new mechanism, synthesized based on the structural development of the drum mechanism, can cause dynamic movements in the mechanism. This means studying its kinematic and dynamic properties is necessary to use the mechanism in technological machines.
6. The research results are essential for using elliptical drums in vertical spindle cotton picking machines.

REFERENCES

[1] Khurramov SR, Abdukarimov A, Khalturaev FS, Kurbanova FZ. (2020) Modeling of friction forces in an asymmetric two-roll module. *ModTech 2020*, IOP Publishing, *IOP Conf. Series: Materials Science and Engineering*, 916, 012051, 1-8. https://doi.org/10.1088/1757-899X/916/1/012051

[2] Bahadirov GA, Sultanov TZ, Abdukarimov A. (2020) Kinematic analysis of tooth-lever differential transmission mechanisms. *First International Conference on Energetics, Civil and Agricultural Engineering 2020*, Tashkent, Uzbekistan. *IOP Conf. Series: Earth and Environmental Science*, 614, 012102. https://doi.org/10.1088/1755-1315/614/1/012101

[3] Bahadirov GA, Sultanov TZ, Abdukarimov A. (2020) Comparative analysis of two gear-lever differential inter-roller transmission mechanisms. *First International Conference on Energetics, Civil and Agricultural Engineering 2020*, Tashkent, Uzbekistan. *IOP Conf. Series: Earth and Environmental Science*, 614, 012101. https://doi.org/10.1088/1755-1315/614/1/012101

[4] Khurramov SR, Abdukarimov A, Khalturaev FS, Kurbanova FZ. (2021) Modeling the shape of the roll contact curves in two-roll modules. *ICEMP 2021*, *Journal of Physics: Conference Series*, 1789, 012008. https://doi.org/10.1088/1742-6596/1789/1/012008

[5] Turanov Kh, Abdazimov A, Shaumarov M, Siddikov Sh. (2021) Type analysis of a multiloop coulisse mechanism of a cotton harvester. *In: Murgul V, Pukhkal V, editors. International Scientific Conference Energy Management of Municipal Facilities and Sustainable Energy Technologies (EMMFT 2019)*. *Advances in Intelligent Systems and Computing*, Springer, Cham, 1258:290–305. https://doi.org/10.1007/978-3-030-57450-5_27

[6] Turanov K, Abdazimov A, Shaumarova M, Siddikov S. (2021) Mathematical modeling of a multiloop coulisse mechanism of a vertical spindle cotton harvester. *In: Murgul V, Pukhkal V, editors. International Scientific Conference Energy Management of Municipal Facilities and Sustainable Energy Technologies (EMMFT 2019)*. *Advances in Intelligent Systems and Computing*, Springer, Cham, 1258:306–321. https://doi.org/10.1007/978-3-030-57450-5_28

[7] Uljayev E, Ravutov ST, Ubaydullayev UM. (2020) Remote control device to control the contact uniformity of the brush strippers on the spindle's surface of the cotton picking apparatus. *IOP Conference Series: Earth and Environmental Science*, 614, 012139. https://doi.org/10.1088/1755-1315/614/1/012139

[8] Amanov A, Sembiring JPBA, Amanov T. (2019) Experimental investigation on friction and wear behavior of the vertical spindle and V-belt of a cotton picker. *Materials*, 12(5), 773. https://doi.org/10.3390/ma12050773

[9] Bahadirov G, Ravutov S, Abdukarimov A, Toshmatov E. (2021) Development of the methods of kinematic analysis of elliptic drum of vertical-spindle cotton harvester. *IOP Conference Series: Materials Science and Engineering*, 1030, 012160. https://doi.org/10.1088/1757-899X/1030/1/012160

[10] Ravutov ST, Rizayev AA, Rajapbaev UA. (2022) On the issue of increasing the efficiency of the spindles of a vertical-spindle cotton picker. *IOP Conf. Series: Earth and Environmental Science*, 1112, 012043. https://doi.org/10.1088/1755-1315/1112/1/012043

[11] Rizaev AA. (2017) Research and creation of working bodies of a cotton picker with high efficiency. *Tashkent: Fan*.

[12] Khudayberdiev R, Turapov A. (1989) On the use of ellipsocycloid in the design of promising cotton pickers. *Doc. Academy of Sciences of the Uzbek SSR*, 1989(8):18–21.

[13] Karimov KA. (1986) Planetary friction mechanisms with variable carrier length. *Monograph, Tashkent: Fan*, 107 p.

[14] Turapov AT, Usmanova BK, Ravutov ST, Ergashev A. (1988) The spindle drum of the cotton picking machine. *Academy of Sciences of USSR*, No. 1419582. *Innoventions, Inventions*, No. 32.

https://doi.org/10.31436/iiumej.v26i2.3509

UTILIZATION OF MSW-DERIVED BIOCHAR FOR PRODUCTION OF IRON THROUGH SOLID-SOLID REACTION

HADI PURWANTO [1], ALYA NAILI ROZHAN [2*], DANIAL FAIZ ZULAINI [2], MOHD HANAFI ANI [2]

[1]Department of Engineering Management, University of International Semen Indonesia, Gresik, West Java, 61122 Indonesia
[2]Department of Manufacturing and Materials Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia

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

(Received: 21 November 2024; Accepted: 15 January 2025; Published online: 15 May 2025)

ABSTRACT: Every year, the increasing human population has contributed to the abundance of Municipal Solid Waste (MSW) at designated landfills, especially in Malaysia. On the other hand, in the iron and steel industry, a high demand for steel production has caused an increase in CO2 production, which would, unfortunately, contribute to global warming. Therefore, sustainable solutions are needed to avoid this environmental crisis from worsening. This study employed an effective approach, utilizing low-grade iron ore from Pahang to produce iron by using MSW-derived biochar as an alternative to coal as a reductant and energy source. This project aims to study the mechanism of the solid-solid reduction reaction of iron ore with carbon. MSW was collected and prepared before the slow pyrolysis to obtain the MSW-derived biochar. Iron ore was characterized by using X-Ray Fluorescence (XRF) for elemental composition determination and X-Ray Diffraction (XRD) to observe the presence of mineral phases. In this research, the solid-solid reaction was conducted for the reduction reaction at 900°C with residence times of 1 hour, 3 hours, and 5 hours. From the research that has been conducted, temperature and residence time were found to influence the elemental composition, reduction degree, and mineral phase transformation of iron.

ABSTRAK: Setiap tahun, peningkatan populasi manusia telah menyumbang kepada kelimpahan Sisa Pepejal Perbandaran (MSW) di tapak pelupusan yang ditetapkan, terutamanya di Malaysia. Manakala dalam industri besi, permintaan yang tinggi terhadap pengeluaran keluli telah menyebabkan peningkatan pengeluaran karbon dioksida, yang malangnya menyumbang kepada pemanasan global. Oleh itu, penyelesaian yang mampan diperlukan untuk mengelakkan krisis alam sekitar ini menjadi lebih buruk. Dalam kajian ini, pendekatan yang berkesan digunakan dengan memanfaatkan bijih besi gred rendah untuk pembuatan besi dengan menggunakan arang yang dihasilkan daripada MSW sebagai alternatif kepada arang batu sebagai bahan agen penurunan serta sumber tenaga. Tujuan projek ini adalah untuk mengkaji mekanisma tindak balas dalam pembuatan besi. MSW telah dikumpulkan dan disediakan sebelum menjalani proses pirolisis untuk mendapatkan arang yang dihasilkan daripada MSW. Bijih besi dianalisa menggunakan Fluoresens X-Ray (XRF) untuk penentuan komposisi unsur dan Penghabluran X-Ray (XRD) untuk memperoleh komposisi fasa mineral. Dalam penyelidikan ini, eksperimen dijalankan untuk tindak balas penurunan pada suhu 900°C dengan masa selama 1 jam, 3 jam dan 5 jam. Berdasarkan kajian yang telah dijalankan, suhu dan masa didapati mempengaruhi komposisi unsur serta tahap penurunan besi.

KEY WORDS: *Municipal Solid Waste, Reduction of Iron, Biochar, Biomass Ironmaking.*

https://doi.org/10.31436/iiumej.v26i2.3509

1. INTRODUCTION

Municipal solid waste (MSW) is mostly generated from homes, businesses, offices, schools, and other places regularly. It covers all types of non-industrial solid waste, including garbage, street cleanings, decaying organic matter, and medical waste. MSW comprises organic food waste, paper, plastics, metal, and glass. Today, most developing countries use landfills as their final disposal method. This is because landfills are considered the cheapest, easiest, and most cost-effective method of waste disposal. The rising amount of generated MSW is a major cause of environmental problems such as recurring land and water pollution. MSW poses a significant global challenge, with an annual production that exceeds 2 billion tons, which is forecasted to double by the end of the century [1]. As the population grows in Malaysia, an increasing amount of MSW is reportedly deposited at landfills yearly. Foods, papers, and plastics were discovered to be the major composition of Malaysian MSW, which is expected to increase 3.3% annually [2].

However, since the significant components of MSW are found to be organic matter, this makes it a viable candidate to be utilized as a source for bioenergy generation [3]. Pyrolysis is one of the most prominent processes for producing energy from these organic materials. This prominent thermochemical process is conducted at temperatures between 300 and 700°C to thermally convert organic matter into biochar and bio-oil, which can be later used as alternative energy sources [4]. Generally, the temperature for producing biochar is up to 700°C, which would take 20 to 60 minutes. As the temperature rises, the quantity of acidic functional groups, particularly carboxylic functional groups, decreases and causes the development of basic functional groups. In other words, the structure of biochar will change to become graphite carbon at higher pyrolysis temperatures. The function of this biochar is comparable to that of normal coal, which provides an alternative method to coal utilization in iron production [5].

Carbon emissions, mainly from iron mining and steel production, are also increasing due to high market demand in developing countries. As a result, the world is at risk of a worsening global warming crisis. This environmental crisis's risk will affect human health and all life on land and in water. Therefore, prevention is always essential to sustaining this world and all humanity. To overcome these problems, one of the solutions is to utilize other renewable resources, such as MSW, as an alternative to coke.

On the other hand, in Malaysia, iron mines are still operating to extract iron ore. However, Malaysian iron ores are regarded as low-grade iron ores because of the 56 to 59% iron content and a relatively high combined water content. These low-grade iron ores do not fulfill the specifications to be used directly in the blast furnace as a feedstock for pig iron production without pre-reduction treatment [6]. According to the World Bureau of Metal Statistics, in 2021, Malaysia's iron ore production volume was approximately 4.98 million metric tons [7].

Scrutinizing the issues from both perspectives, MSW is a good candidate for biochar production before being utilized as a reductant for the reduction reaction of iron ore. Research on the reduction reactions of iron ore using MSW-derived biochar by solid-state responses is minimal. Therefore, this study was conducted to observe the outcomes.

2. METHODOLOGY

MSW was collected from households, and in general, the main components of this MSW in Malaysia are organic waste, plastic waste, and paper waste. The ratios were determined from the proportions of each specific component and were based on 1kg of hematite, Fe_2O_3. To obtain MSW-derived biochar that would be used as a reductant for iron ore reduction reaction,

MSW underwent a slow pyrolysis process at 350°C for 2 hours. Based on the chemical reactions, the minimum carbon required to reduce iron ore was calculated, and thus, the ratio for sample preparation could be determined.

For sample preparation, MSW was prepared, with the organic materials, plastics, and papers weighing 60, 20, and 20g, respectively. The samples were oven dried at 105°C for 24 hours to completely remove contact water. Subsequently, MSW was subjected to a heating process from room temperature to 350°C with a 10°C/minute heating rate. This slow thermal decomposition was conducted in an inert atmosphere for 2 hours to produce MSW-derived biochar.

On the other hand, goethite, a low-grade iron ore, was selected as the raw sample for this study. The physical form of this ore was reddish-brown in appearance, consisting of hard clay and hard rock. For sample preparation, iron ore was oven dried at 105°C for 24 hours to remove moisture before being ball milled for 1 hour to break the ore into small particles. The dough was compacted with a single shaft compressor at 1000 psi for 5 minutes. It was then placed in a cylindrical iron mold with water as the binding agent. Each sample weighed approximately 18g with a 2 cm diameter and 2.5 cm thickness. After pelletizing was completed, all pellets were left at room temperature for approximately 20 hours, then the samples were placed in an oven at 105°C for 24 hours to remove moisture inside the pellets.

For iron ore reduction experiments, an electric tubular furnace was used to examine the solid-solid reaction between iron ore bodies and MSW biochar in an argon atmosphere. Iron ore samples were mixed with biochar particles before being placed in enclosed steel baskets in an electric furnace. For the reduction process, samples were prepared according to iron ore to MSW char ratios, which were 8:2 and 8:4. These samples were subjected to experiments with different residence times: one hour, three hours, and five hours. For sample characterization, X-Ray fluorescence (XRF) and X-Ray diffraction (XRD) were utilized to determine the raw iron ore's chemical composition and mineral phases, respectively. The reduction degree was also calculated to observe the reduction reaction of iron.

3. RESULTS AND DISCUSSION

The elemental composition of the raw iron ore collected is shown in Table 1. The primary composition of the ore was iron, which is 59.1 wt%. Fig. 1 shows the thermal characteristic curve for this iron ore. At the temperature range of 220 to 350°C, there was a significant weight loss of 3.65%, where an endothermic peak appeared, indicating heat absorption. This was due to the removal of combined water in the form of hydroxyl, which often exists in limonitic ore containing the goethite mineral. This process can also be seen in the endothermic peak on the DTA curve. Heat was absorbed at 220 to 370°C for the iron ore sample, where the weight loss ranged from 1.50% to 3.75%, indicating the occurrence of combined water removal at this temperature range.

Table 1. Elemental compositions of Malaysian iron ore (wt%)

Locality	T-Fe	Si	Al	S	Mn	V	Ti	Ca	T-Fe	Si
3	59.1	0.8	1.3	0.2	0.4	0.5	6.2	0.3	59.1	0.8

https://doi.org/10.31436/iiumej.v26i2.3509

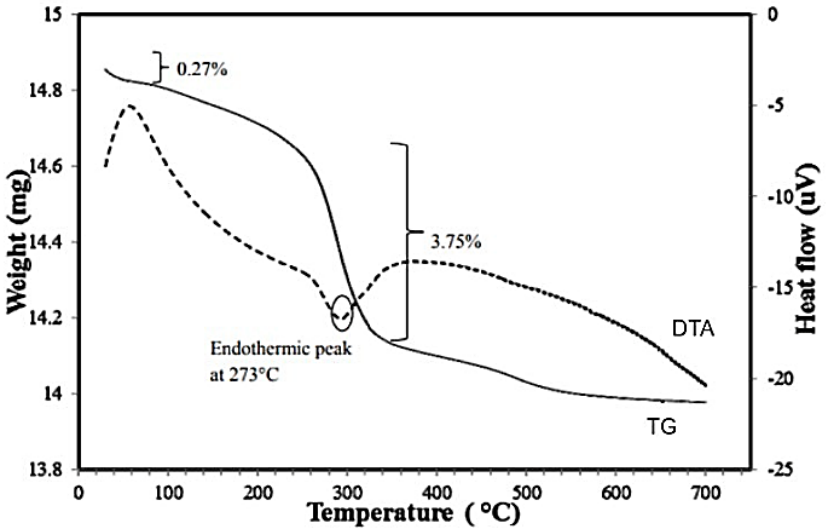


Figure 1. Raw iron ore TG-DTA profile

Fig. 2 shows the raw iron ore's XRD pattern. The primary minerals observed were goethite, an iron oxide hydrate, with the predominant hematite phase. The XRD profile also indicated some quartz as the main impurity.

The proximate and ultimate analyses of the MSW are shown in Table 2 below. The analysis shows that the contents of volatile matter, fixed carbon, and ash were 69.35, 9.00, and 7.05 mass% of MSW, respectively. Furthermore, the elemental analysis of MSW indicates that it contained 45.00, 6.11, 0.40, 0.12, and 48.37 mass% of carbon, hydrogen, nitrogen, sulfur, and oxygen (by difference), respectively. Table 3 shows that during the pyrolysis process of MSW, 100.0g of the raw material was heated without oxygen, leaving behind an approximate 60% char yield, where 40.5g was lost during the process.

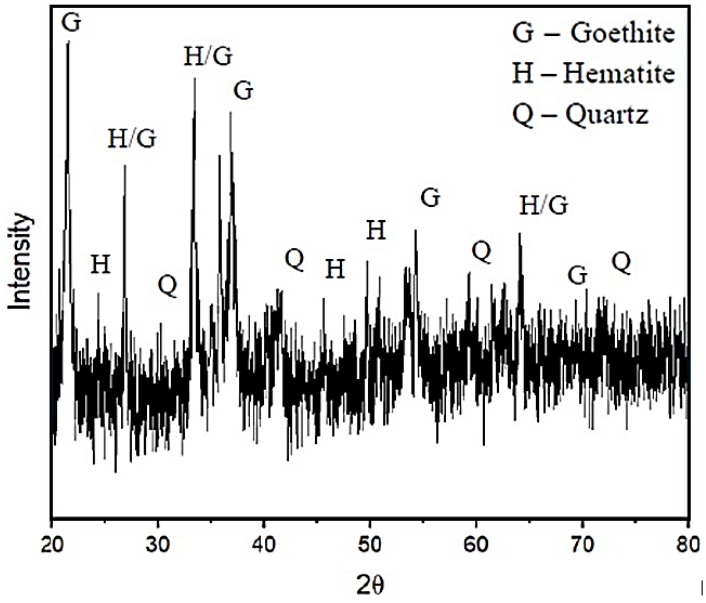


Figure 2. XRD analysis of raw iron ore.

https://doi.org/10.31436/iiumej.v26i2.3509

Table 2. Proximate and Ultimate analysis of MSW.

Analysis	Proximate Analysis			Ultimate Analysis				
Elemental Mass (%)	Volatile Matter	Fixed Carbon	Ash	C	H	N	O	S
MSW	69.35	9.00	7.05	45.00	6.11	0.40	48.37	0.12

Table 3. Weight differences of MSW after pyrolysis.

Sample	Weight before (g)	Weight after (g)	Weight loss (g)
MSW	100.0	59.5	40.5

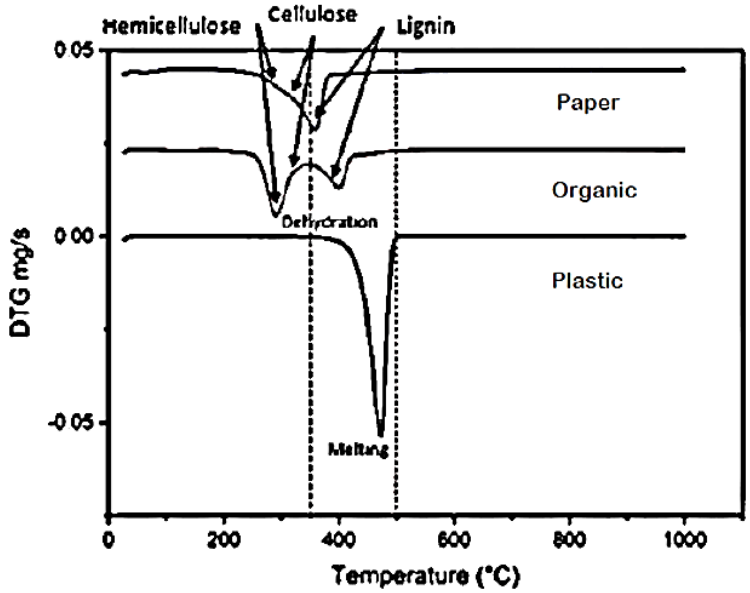


Figure 3. DTG pattern of MSW by component.

Fig. 3 depicts the DTG curve for the MSW sample by component. As the temperature increased to 300 °C, the weight of MSW steadily decreased as cellulose, hemicellulose, and lignin decomposed. The weight of the samples does appear to significantly decline over 600°C, indicating that the decomposition of hydrocarbons has completed. This demonstrates that certain elements, such as CO, H2, CO2, and hydrocarbons, were released. The DTA curve shows apparent peaks or valleys between 300 and 500 °C, representing reactive devolatilization in the materials [8]. The pyrolysis process in this experiment was conducted at 350°C so that during the reaction, there would be some diffusion of gaseous compounds, which would help reduce iron [9]. Fig. 4 shows the weight loss of the iron pellets after each sample was reduced with different reduction times within the range from one hour to five hours, with different iron ore to MSW char ratios, to study the effectiveness of the reduction process. The weight loss of pellets significantly increased with time, and the highest weight loss was obtained when twice the mass of MSW char is used to reduce iron ore at a temperature of 900°C. As the reduction temperature increases to 900°C, the rate of iron ore reduction also increases [10]. At this temperature, MSW char was gasified into CO and H2, which act as a reduction agent for the reduction reaction of iron oxide in the ore. Excess carbon and hydrogen in the mixture would also help to proceed with the reduction reaction further, as shown in equations 1 to 6. Therefore, higher carbon and hydrogen contents in the biomass promote more oxygen removal from iron species, which causes greater weight loss of iron ore pellets.

Reducing gases are formed during the reduction process through the reaction of chemicals between iron oxide and carbonaceous materials. The Boudouard reaction (Eq. 7) and the water

https://doi.org/10.31436/iiumej.v26i2.3509

gas reaction (Eq. 8) produced syngas, which would react with iron oxide to convert iron ore to metallic iron. In the lower oxide and iron direction, simultaneous and continuous reduction processes occur from hematite to magnetite, from magnetite to wustite, and from wustite to metallic iron. The following sequence summarizes the main reduction processes of iron from hematite to iron (Fe):

Hematite to Magnetite

$$3Fe_3O_4 + CO \rightarrow 2\ Fe_3O_4 + CO_2 \quad (1)$$

$$3Fe_3O_4 + H_2 \rightarrow 2Fe_4O_3 + H_2O \quad (2)$$

Magnetite to Wustite

$$Fe_4O_3 + 4CO \rightarrow FeO + 4CO_2 \quad (3)$$

$$Fe_4O_3 + 4H_2 \rightarrow FeO + 4H_2O \quad (4)$$

Wustite to Iron

$$FeO + CO \rightarrow Fe + CO_2 \quad (5)$$

$$FeO + H2 \rightarrow Fe + H_2O \quad (6)$$

Boudouard reaction

$$CO_2 + C \rightarrow CO \quad (7)$$

Water-Gas reaction

$$H_2O + C \rightarrow H_2 + CO \quad (8)$$

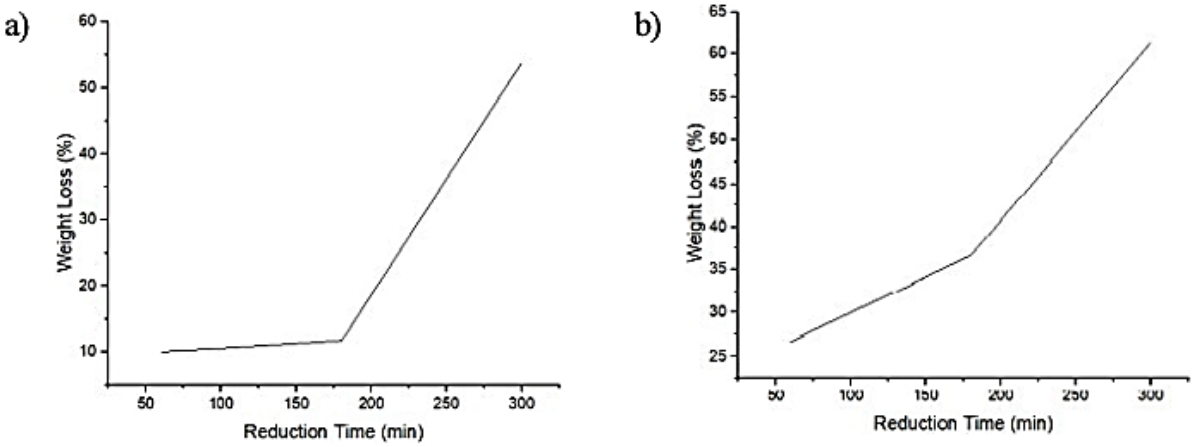
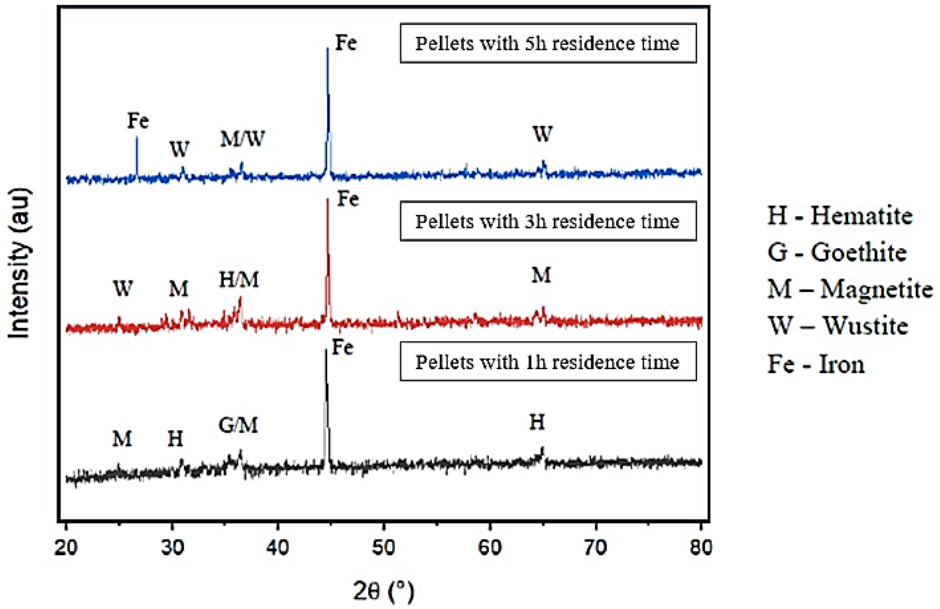


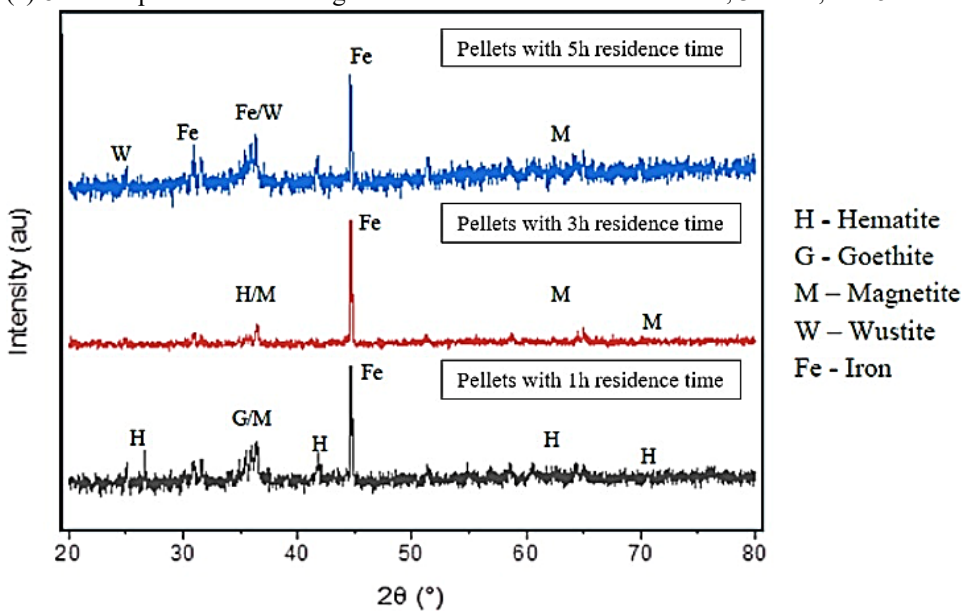
Figure 4. Effect of time on weight loss during iron reduction at 900°C for: (a) 8:2 ratio (b) 8:4 ratio pellets.

As iron ore was reduced, the characterization using XRD was conducted to observe the transformation phase of iron ore according to the reduction time. A longer reduction time allows hydrocarbons and gases to diffuse and thus reduce the iron ore. Samples with 8:2 ratios were exposed to reduction reaction with a holding time of 1 hour (Sample 1), 3 hours (Sample 3), and 5 hours (Sample 5). Samples with an 8:4 ratio were also exposed to the same residence time; those were 1 hour, 3 hours, and 5 hours for Sample 2, Sample 4, and Sample 6, respectively.

https://doi.org/10.31436/iiumej.v26i2.3509



(a) 8:2 ratio pellets for reducing iron with residence times of 1 hour, 3 hours, and 5 hours.



(b) 8:4 ratio pellets for reducing iron with residence times of 1 hour, 3 hours, and 5 hours.

Figure 5. XRD Pattern

Both reduction reactions for samples with 8:2 and 8:4 ratios indicate the presence of iron, as the minimum temperature to obtain iron is 900°C, as shown in Fig. 5 (a) and Fig. 5 (b) below. One hour is sufficient to reduce hematite to iron, but only a small number of ores can be reduced directly to iron. Samples 1 to 4 show the presence of magnetite, hematite, and some iron-Fe. Samples 5 and 6 illustrate the presence of mainly wustite and iron with little magnetite. From this trend, it can be deduced that by increasing time and adding more biochar, there is a possibility for iron ore to be reduced further.

The degree of reduction was calculated depending on the mass loss of the iron ore. A similar pattern can be observed in the mass change of the pellets after the reduction process. The degree of reduction increases with a reduction time of up to five hours. Maximum reduction has been achieved when a 16:2 ratio of the weight of the MSW was used. As the reduction continues, the biomass gasifies to produce more reducing gas in the chamber.

https://doi.org/10.31436/iiumej.v26i2.3509

Diffusion of CO and H2 through the pellets resulted from the gasification of MSW, which was facilitated by the reduction reaction of iron oxide.

As the weight loss percent of both ratios of iron to MSW shows an increasing pattern with time, the percentage of Fe content is also increasing, as shown in Table 4. The 8:2 ratio has an increment of 23.9% in total, starting from one hour of reduction time up to five hours, while the 8:4 ratio shows an increment of 27% of Fe produced. This indicates that the weight loss due to the char element diffusion would lead to reduced iron. The mass of MSW biochar could also affect the percentage of Fe increment.

This is because, as the mass of MSW increases, the iron oxides can be expected to fully reduce into iron when exposed to a longer residence time. A short residence time would limit the degree of reduction reaction. From the weight loss data of iron ore, ore pellet, and MSW, the reduction degree was calculated, and it was found that as the residence time increased, the percentage of Fe increment also increased, showing that MSW can be used as a reducing agent for iron ores.

Table 4. Percent of Fe increment with time.

Sample	Holding Time (hour)	Fe Content (%)
8:2 Ratio	(Raw)	59.1
	1	68.6
	3	76.2
	5	83.0
8:4 Ratio	(Raw)	59.1
	1	80.9
	3	83.4
	5	86.7

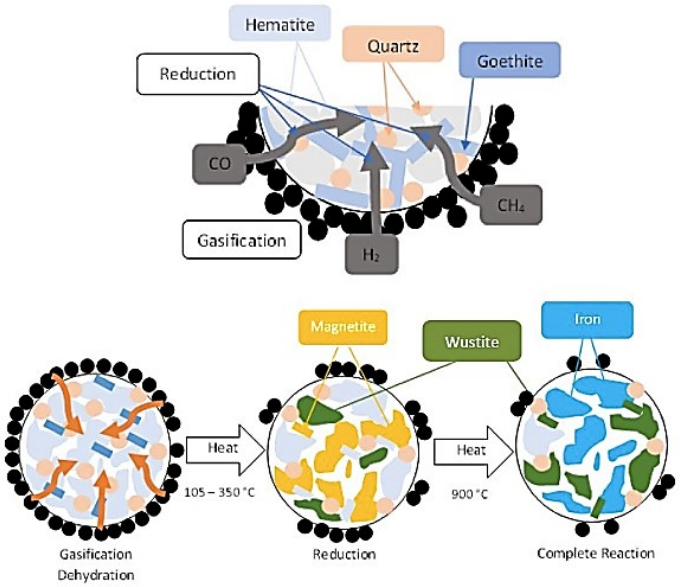


Figure 6. Solid-solid reaction mechanism of MSW biochar with iron ore.

Fig. 6 illustrates the reduction mechanisms in the reduction process of iron pellets with MSW biochar as reducing agent via solid-solid reaction. As the pellets were heated with bedded char around them, the temperature inside the pellet increased due to heat transfer from the outside to the inside of the pellet. In this experiment, combined water removal from iron ore occurred when heated to a low temperature by releasing the hydroxyl chain, creating many

voids and generating cracks on the iron ore surface with a large specific surface area. Gases produced from MSW char gasification also diffused into the pellets of biochar-iron ore mixture because of the formation and volume expansion of syngas occurring at this location. Therefore, the hydrocarbons and gases produced a large amount of heat, spreading throughout the bodies of the pellet. Biomass gasification produces syngas such as CO and H_2 that can be utilized as reducing agents in reducing iron ore. As these gases are being produced and diffused in the inner area of the pellets, they also diffuse outwards to reduce the iron oxide species in the outer part of the pellets. For the iron reduction, iron ore was first reduced by H_2 and hydrocarbon gases generated from MSW biochar, because reduction by H_2 is a faster process than that of other gases. However, the reaction process is very complicated because many elemental reactions can take place in this process. During MSW biochar gasification, syngas and hydrocarbon gases were produced as a potential reduction agent for iron oxide, and the decomposition of hydrocarbon gases into carbon to inhabit the formed pores on the surface of iron oxide. On the other hand, gases reacted together along with ore oxides, reducing Hematite to Magnetite, Magnetite to Wustite, and Wustite to metallic iron phase.

4. CONCLUSION

Low-grade iron ore with 59.1% of Fe was used as the raw material for this study. MSW was selected to be converted into biochar, a carbon material that acts as the reductant and/or energy source in reducing iron. A sample with an 8:2 ratio of iron to MSW showed an increasing weight loss pattern. A sample with an 8:4 ratio presented the same pattern but with more weight loss because of the extra carbon particles. As the holding time increased, the degree of reduction also increased.

After the experiment, the XRD results of each sample indicated the presence of different iron species such as magnetite, wustite, and iron-Fe. During the heating process, MSW-derived biochar produced reductant gases that would assist in reducing iron from goethite to iron-Fe. By doubling the mass of the MSW-derived biochar, the reaction towards iron ore particles is highly active, with an increasing percentage of total iron in the sample, and more weight loss, which indicates the removal of oxygen atoms. This study promotes the reduction of MSW at landfills, which has become a troubling issue in Malaysia year by year. MSW can be a new alternative source of energy for metal smelting.

REFERENCES

[1] Ali HM, Tyagi VK, Kazmi AA, Ojha CSP. (2024) Biomethane Production from Organic Fraction of Municipal Solid Waste: Clean and Green Energy Generation under Circular Economy Framework. International Journal of Hydrogen Energy, 95:1080-1094.

[2] Yong ZJ, Bashir MJK, Ng CA, Sethupathi S, Lim JW, Show PL. (2019) Sustainable Waste- to-Energy Development in Malaysia: Appraisal of Environmental, Financial, and Public Issues Related with Energy Recovery from Municipal solid Waste. Processes, 7(10):676.

[3] Traven, L. (2023) Sustainable Energy Generation from Municipal Soild Waste: A Brief Overview of Existing Technologies. Case Studies in Chemical and Environmental Engineering, 8:100491.

[4] Khamis SS, Purwanto H, Salleh HM, Rozhan AN, Rahman MA, Warimani M, Alam N. (2024) Novel Energy Recovery from an Integrated Municipal solid Waste and Leachate Treatment System. Waste Disposal & Sustainable Energy, 6:53-61.

[5] Rozhan AN, Ani MH, Salleh HM, Akiyama T, Purwanto H. (2015) Development of Carbon-Infiltrated Biochar from Oil Palm Empty Fruit Bunch. ISIJ International, 55(2):436-440.

[6] Yunus NA, Ani MH, Salleh HM, Rashid RZA, Akiyama T, Purwanto H. (2013) Reduction of Iron Ore/Empty Fruit Bunch Char Briquette Composite. ISIJ International, 53(10):1749–1755.

[7] Siddharta A (2024) Production Volume of Iron Ore Malaysia 2012-2021. Available: https://www.statista.com/statistics/1132484/malaysia-iron-ore-production/

[8] Siregar SRH, Nursani D, Setyawan MIB, Surjosatyo A. (2021) Study of Kinetic and Thermal Decomposition from Municipal Solid Waste Pellets using Model-Fitting Methods. IOP Conference Series: Materials Science and Engineering, 1173(2021):012033.

[9] Purwanto H, Salleh HM, Rozhan AN, Mohamad AS, Zakiyuddin A. (2018) Phase Change of Iron Ore Reduction Process using EFB as Reducing Agent at 900-1200C. IOP Conference Series: Materials Science and Engineering, 342(1):012054.

[10] Rashid RZA, Salleh HM, Ani MH, Yunus NA, Akiyama T, Purwanto H. (2014) Reduction of Low Grade Iron Ore Pellet using Palm Kernel Shell. Renewable Energy, 63:617-623.

https://doi.org/10.31436/iiumej.v26i2.3556

ENHANCEMENT OF THE PHOTOCATALYTIC ACTIVITY OF MIL (53) METAL-ORGANIC FRAMEWORKS THROUGH THE ADDITION OF THE REDUCED GRAPHENE OXIDE FOR IMPROVING DEGRADATION OF ORGANIC DYE POLLUTANTS IN WATER TREATMENT APPLICATIONS

MUTTAQIN MUTTAQIN[1*], FANDRY HOSEA JABY[1], RIS KEVIN BRAMASTA[1], NONA MERRY MERPATI MITAN[1], YOSE FACHMI BUYS[2]

[1]*Department of Chemistry, Universitas Pertamina, Jakarta, Indonesia*
[2]*Department of Mechanical Engineering, Universitas Pertamina, Jakarta, Indonesia*

**Corresponding author: muttaqin@universitaspertamina.ac.id*

(Received: 29 December 2024; Accepted: 17 February 2025; Published online: 15 May 2025)

ABSTRACT: A simple nanocomposite consisting of MIL-53(Al) and reduced graphene oxide (rGO), denoted as MIL-53(Al)/rGO, was synthesized as a photocatalyst driven by sunlight and UV light to study the decomposition of methylene orange and methylene blue in aqueous solution. The MIL-53(Al)/rGO ultrafine particles were produced by an in situ method using the solvothermal technique. The nanocomposite was made with two different amounts of rGO, 2.5% and 5% by weight. Various tests, including XRD, N_2 adsorption-desorption isotherms, SEM, SEM-EDS, UV-Vis DRS (Diffuse Reflectance Spectroscopy), and FTIR, were performed on all photocatalyst variations to analyse their properties. Results from SEM and EDS showed the creation of small MIL-53(Al) particles measuring 10-20 μm and rGO spread evenly on the MIL-53(Al) surface, particularly in the 2.5% rGO sample. The photocatalytic effectiveness of the MIL-53(Al)/rGO nanocomposites was tested for degrading organic dyes (MO and MB) in water under both sunlight and UV light for 60- and 120-minute durations. The 2.5% rGO photocatalyst showed the highest performance, removing over 96% and 98% of the dyes after one hour of sunlight exposure for MB and MO, respectively. This demonstrates that the combined effect of MIL-53(Al) and rGO composite can be seen as an effective photocatalyst for breaking down reactive dyes, such as MO and MB, in water treatment applications.

ABSTRAK: Kajian ini adalah berkaitan nanokomposit sederhana daripada MIL-53(Al) dan grafit oksida yang tereduksi (rGO), atau MIL-53(Al)/rGO, berjaya disintesis sebagai fotopemangkin oleh cahaya matahari dan cahaya UV bagi mengkaji penguraian metil jingga (MO) dan metilena biru (MB) dalam larutan akueus. Zarah ultrahalus MIL-53(Al)/rGO dihasilkan melalui kaedah in situ menggunakan teknik solvotermal. Nanokomposit dibuat dengan dua jumlah rGO berat berbeza, 2.5 wt% dan 5 wt%. Pelbagai ujian termasuk XRD, N_2 penyerapan-nyahserapan isoterma (BET), SEM, SEM-EDS, UV-Vis DRS, dan FTIR telah dilakukan pada semua variasi fotopemangkin bagi mengkaji sifatnya. Dapatan kajian dari SEM dan EDS menunjukkan penciptaan zarah kecil MIL-53(Al) berukuran 10-20 μm dan rGO tersebar secara rata pada permukaan MIL-53(Al), terutamanya dalam sampel rGO 2.5%. Keberkesanan fotopemangkin nanokomposit MIL-53(Al)/rGO telah diuji bagi mengurai pewarna organik dalam air, di bawah kedua-dua cahaya matahari dan cahaya UV selama tempoh 60 dan 120 minit. Fotopemangkin rGO 2.5% menunjukkan prestasi tertinggi, dengan penyingkiran lebih dari 96% MB dan 98% MO, selepas pendedahan cahaya matahari selama satu jam. Ini menunjukkan, kesan gabungan komposit MIL-53(Al) dan rGO, boleh dilihat

sebagai fotopemangkin yang berkesan bagi memecahkan pewarna reaktif, seperti MO dan MB, dalam aplikasi rawatan air.

KEYWORDS: *Metal organic frameworks, Photocatalyst, Reduced graphene oxide, Organic Dye Pollutants, Degradation.*

1. INTRODUCTION

In the last decade, freshwater pollution by organic dyes has become a problem of concern for governments in every country. Industrial wastewater containing organic dyes is the most difficult waste to eliminate due to its high level of toxicity, difficult degradation process, high chromaticity content, and difficulty in removing the color [1]. Several methods have been introduced to remove organic dyes, such as physicochemical processes like adsorption, coagulation, and filtration. However, they include high costs, low efficiency, and the generation of large volumes of sludge that need further treatment [2][3]. Therefore, finding alternative technologies for removing toxic organic dyes in wastewater is essential. Advanced oxidation processes (AOPs) have emerged as a highly effective alternative for converting organic materials into simple mineral products [4][5]. Among AOPs, photocatalysis is seen as the most promising solution for sustainable energy conversion [6]. Researchers have focused on photocatalysis for its non-selectivity, efficiency in removing pollutants, and toxicity reduction. However, the complete degradation of organic dyes is time-consuming and less effective in some studies. Therefore, developing new and improved photocatalysts is necessary to address these issues. Materials that can be used as photocatalysts are semiconductors [7][8]. Semiconductors have proven to be promising materials as photocatalysts for degrading liquid organic waste [9]. In addition, photocatalysts have several advantages, such as low energy consumption, mild reaction conditions, reduction of secondary pollution, and so on, which have attracted wide attention domestically and internationally [10][11]. However, some semiconductor materials have disadvantages, such as being unresponsive to visible light (around 43% of the solar spectrum) and rapid recombination between electrons and holes [12][13]. Therefore, to overcome these limitations, the development of new photocatalysts with high efficiency, long-term stability, and the ability to harness visible light is necessary[14][15].

One of the materials that are included as semiconductors and receiving wide attention for application in the field of photocatalysis is metal-organic frameworks (MOFs) [16][17]. MOFs have a band gap that is not too large [18][19]. MOFs have the potential as semiconductor materials because they can transfer charge from ligands to metals (LMCT) or transfer charge from metals to ligands (MLCT). With this charge transfer ability, MOFs have photocatalytic capabilities to reduce hazardous materials [20]. One type of MOF with an interesting structure and widely studied is Hong Kong University of Science and Technology-1 (HKUST-1), a Cu-based MOF with benzene dicarboxylate organic ligands [21]. By adding graphene oxide (GO) to the HKUST-1 structure, the composition of these two materials has very good photocatalytic performance on organic dyes [22]. Furthermore, GO is a graphene-like material that has been functionalized with different oxygen-containing groups such as hydroxyl, carboxylic, carbonyl, and epoxide groups. Studies have shown that there are chemical bonds formed between MOF and substrates in composite materials [21]. Research indicates that the oxygen groups of GO are linked to the metal centres of the MOF in MOF/GO composites, leading to an enhanced ability to adsorb toxic gases such as NH_3, H_2S, and NO_2 [23][24]. But the use of nanocomposites combining MOF MIL-53 with Aluminium (Al) metal ion cluster and graphene-based materials like reduced graphene oxide (rGO) as highly effective photocatalysts has not been extensively studied until this point.

https://doi.org/10.31436/iiumej.v26i2.3556

In this study, we will perform and report the degradation process of organic dyes, methylene orange (MO) and methylene blue (MB), contaminating water using MOF MIL-53(Al) nanocomposite with rGO as co-doping. The photocatalyst synthesis method uses solvothermal techniques, while rGO is produced using the modified Marcano method [25]. MIL-53(Al) bandgap measurement shows great potential in photocatalysis because it has a band gap between 3.5 – 4.5 eV. The formation of nanocomposites between MOFs and rGO is expected to widen the range of absorption light and slow down the electron-hole recombination process of semiconductor materials based on MOFs.

2. METHODOLOGY

This section contains the materials we used, the synthesis method and process of the target materials, and various characterizations employed to determine the physical and chemical properties of the synthesized materials. The final part describes the testing methods used to evaluate the synthesized materials as photocatalysts.

2.1. Materials

The materials needed for this procedure are aluminium nitrate nonahydrate ($Al(NO_3)3{\cdot}9H_2O$, Merck KGaA), terephthalic acid or benzene-1,6-dicarboxylic acid (H_2BDC, 98%, Sigma Aldrich), graphite as a carbon source for GO/rGO synthesis, concentrated sulfuric acid (H_2SO_4, Mallinckrodt), phosphoric acid (H_3PO_4, Merck KGaA), 30% hydrogen peroxide (H_2O_2, Merck KGaA), potassium permanganate ($KMnO_4$, Sigma Aldrich 98%), N,N'-dimethylformamide (DMF, Merck KGaA), methanol (CH3OH, Merck KGaA), and distilled water. All materials are used directly without any purification steps.

2.2. Synthesis of Reduced Graphene Oxide (rGO)

The rGO synthesis followed previous work [25]. The process begins by preparing a 9:1 mixture of H_2SO_4/ H_3PO_4 (23 ml) and adding 1g of graphite powder. Afterward, 3g of $KMnO_4$ is slowly added to the solution, and the mixture temperature is maintained below 10°C during the addition process. The mixture is stirred for 40 minutes at 50°C and then diluted with deionized water. 30% hydrogen peroxide (H_2O_2) enhances the oxidation process. Adding H_2O_2 to the solution changes its colour to greenish yellow with bubbles indicating a high oxidation level. The solution is filtered and washed until the pH is near 7. The washing and filtration process of the rGO suspension was carried out by following the following steps: The rGO powder was dispersed in excess distilled water and stirred for 10 minutes using a magnetic stirrer at 300 rpm to neutralize the residual acid attached to the rGO. The rGO solid was then separated from the suspension solvent using a centrifuge, set at 5000 rpm for 10 minutes. This suspension formation and centrifugation process was repeated several times with excess distilled water until the pH of the rGO suspension approached 7. Finally, the product, rGO, is recovered and heated to 60°C overnight.

2.3. Synthesize nanocomposite Mil-53 (Al)/rGO

Synthesis of MIL-53 (Al)/rGO nanocomposites started by preparing the metal source from aluminium nitrate salt and the ligand source from terephthalic acid in a 1:1 molar ratio. Both materials are mixed and homogenized in DMF for 60 minutes under standard conditions. Afterwards, the rGO suspension was added in an amount of 2.5 to 5 % wt. of the total weight of the metal salt and ligand used and homogenized using a magnetic stirrer at around 400 rpm for 60 minutes. Then, the suspension is in a Durant bottle and is heated in an oven at 120°C for 72 hours to trigger the growth of MIL-53 (Al)/rGO crystals. After heating, the solid MIL-53

(Al)/rGO catalyst was recovered using a centrifugation technique at 5000 rpm for 10 minutes, then washed twice with methanol. The wet MIL-53 (Al)/rGO catalyst was dried at 80°C for 24 hours in the oven. Next, synthesized MIL-53 (Al)/GO is ground and sieved using a 200-mesh sieve. Finally, the catalyst was activated using a furnace at 160°C for 5 hours to produce dry nanocomposite MIL-53 (Al)/rGO.

2.4. Characterization

The samples were characterized using various analytical techniques. The infrared spectra were recorded using a NICOLET iS5 FT-IR spectrophotometer with ATR windows. Powder X-ray diffraction was performed with a Rigaku Miniflex 600 with a detector 1D−DteX/Ultra, Kβ filter. The XRD instrument uses a Cu-Kα radiation source. Furthermore, the morphological surface of the catalyst was examined using a Phenom Pro-X SEM. Finally, a Micrometrics TriStar II gas sorption and porosimeter device determined specific surface areas and porosity.

2.5. Photocatalytic Test

The photocatalyst testing of all photocatalyst materials was conducted by mixing 12 mg of the material with 20 mL of Methylene Blue (MB) or 20 mL of Methylene Orange (MO) at a concentration of 20 ppm in a reactor vessel. The mixture was stirred at 400 rpm using a magnetic stirrer for 60 and 120 minutes. The experimental setup for testing the photocatalyst material uses a glass reactor with the equipment arrangement as shown in Fig. 1. The photocatalyst performance tests were conducted under two types of light: UV light and sunlight. For UV light, the test was performed using a UV-A lamp with a wavelength range of 320-400 nm. For sunlight testing, the reactor containing the test solution was placed near a window to expose it to sunlight throughout the testing period. The cooling jacket was also filled with water, and the temperature was maintained at 30°C. This temperature was chosen as it is a moderate temperature typical for Jakarta and its surroundings.

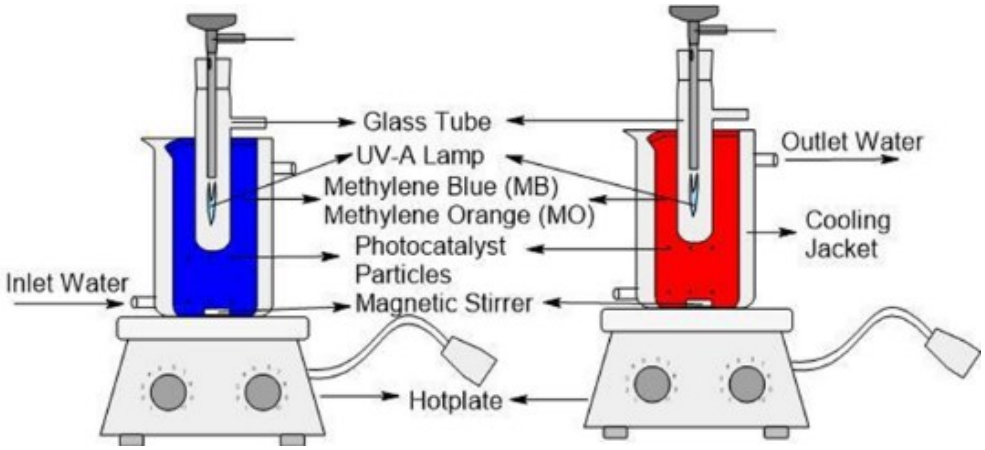


Figure 1. Experimental setup for organic dyes degradation using photocatalyst materials.

3. RESULTS AND DISCUSSION

3.1. FTIR SPECTRA

The interaction of aluminum (Al) ions and organic linker, H_2BDC, in the solvent, triggers the formation of covalent coordination bonds between metal ions and oxygen atoms derived from bidentate terephthalic acid [26]. As presented in Fig. 2, the wave numbers at 1417, 1511, and 1603 cm^{-1} correspond to a carbonyl (C=O) group and CO stretching in the C-OH

carboxylic group. Meanwhile, the wave numbers at 1001, 755, and 593 cm^{-1} indicate the C=O double bond in terephthalic acid, which interacts with metal ions and forms a single metal-O bond [27]. From FTIR data, we qualitatively assume that the target materials had been synthesized successfully.

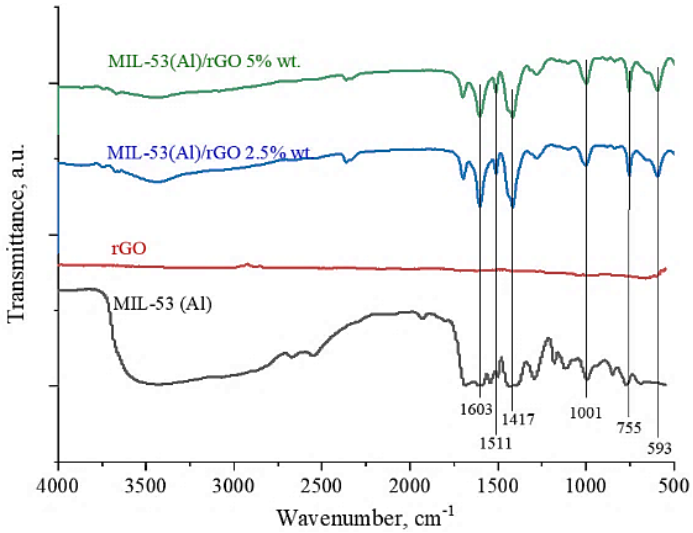


Figure 2. FTIR spectrums of MIL-53 (Al), rGO, MIL-53(Al)/rGO 2.5% wt., MIL-53(Al)/rGO 5% wt.

3.2. PHOTOCATALYST BANDGAP

The bandgap measurement of the photocatalytic materials was conducted using a UV-Vis DRS spectroscopy instrument. The sample reflects electromagnetic waves in the UV-Visible wavelength range in this measurement. Using the Tauc plot proposed by J. Tauc [28], the bandgap of a material can be calculated. The bandgap values for all photocatalyst variants are presented in the figure below.

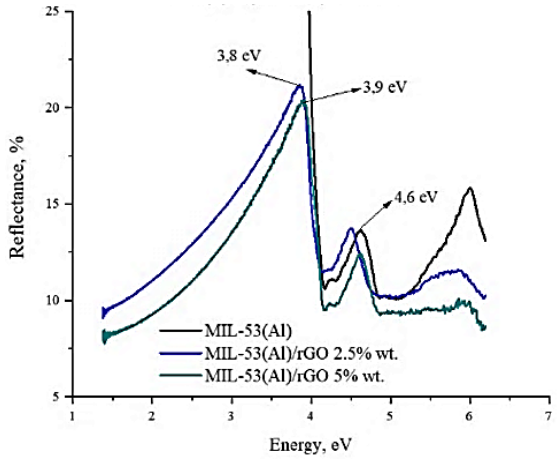


Figure 3. Bandgap measurement of photocatalyst materials.

Fig. 3 shows that the addition of 2.5% wt. rGO is effective in reducing the bandgap of the photocatalytic material. However, increasing the rGO ratio to 5% wt. does not further decrease the bandgap value. Instead, it remains the same or even slightly increases compared to the bandgap of the photocatalytic material without rGO. This phenomenon appears to be influenced by the decrease in the photocatalytic material's crystallinity and the increase in the size of the photocatalyst due to the increasing coverage by rGO.

https://doi.org/10.31436/iiumej.v26i2.3556

3.3. X-Rays Diffraction

XRD patterns of GO, rGO, and MIL-53(Al) and MIL-53(Al)/rGO composites are shown in Fig. 4. In the XRD pattern of the GO sample, the sharp intense peak at 2θ of 10° is assigned to the plane of [002] belonging to the GO layer [24]. In the XRD pattern of rGO, the sharp peak no longer exists but is replaced by a diffractogram representing material with amorphous properties. Interestingly, the diffractogram peak of the rGO sample appears at 2θ of 23°, a value similar to the 2θ peak of graphite [29]. In the XRD pattern of MIL-53 (Al), the strong peaks appear at 2θ of 9.13°, 10.05°, 15.36°, 18.26°, 20.91°, 25.45°, and 32.6° are several peaks characteristic of the MIL-53 (Al) phase, which match the peaks reported in the previous report [30]. In the XRD pattern of MIL-53(Al)/rGO, some diffraction peaks of MIL-53(Al) are still clearly visible, but these peaks appear at 2θ values that are either higher or lower. Furthermore, the amorphous profile, which belongs to rGO, does not appear due to mixing with the crystal structure of MIL-53(Al). This result indicates that MIL-53(Al) crystals may be well dispersed within the interlayers of GO/rGO sheets.

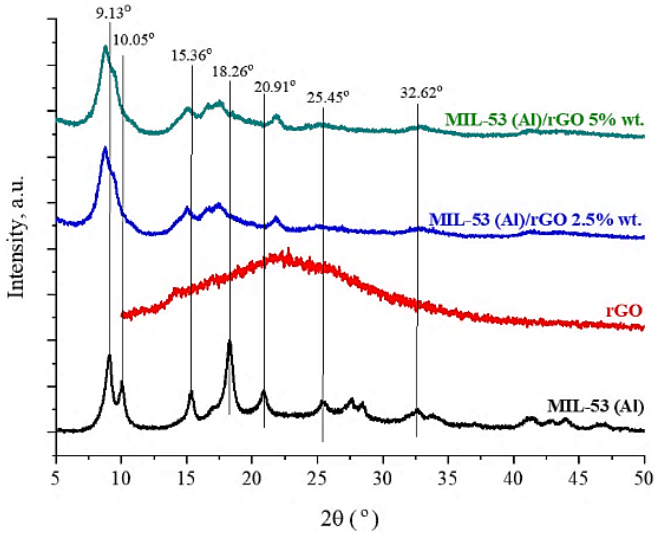


Figure 4. XRD patterns of: MIL-53 (Al), rGO, MIL-53(Al)/rGO 2.5% wt., MIL-53(Al)/rGO 5% wt.

3.4. SEM and SEM-EDS Analysis

The photocatalytic process is influenced by the physical characteristics of the catalyst, such as its active surface area and porosity. Smaller catalyst sizes result in a higher surface-to-mass ratio, reducing the distance electrons need to travel to the solid/solution interface and increasing efficiency. Therefore, our objective was to produce semiconductor nanocrystals to enhance light absorption. Scanning electron microscopy (SEM) images in Fig. 5 illustrate the morphology of MIL-53(Al), MIL-53(Al)/rGO 2.5% wt., MIL-53 (Al)/rGO 5% wt. From XRD testing results, the addition of rGO, both at 2.5% w.t and 5% w.t, did not significantly affect the crystal structure of MIL-53 (Al) because they have similar diffractogram profiles. This XRD result is supported by SEM profiles for MIL-53(Al)/rGO composites for all variants. The general SEM image of MIL-53(Al) appears as sub-micrometer particles with bar and flat cube surfaces and sharp or rounded edges [31]. The presence of GO/rGO surrounding the crystal structure of MIL-53(Al) did not alter its crystal structure dramatically, as seen in the following image.

https://doi.org/10.31436/iiumej.v26i2.3556

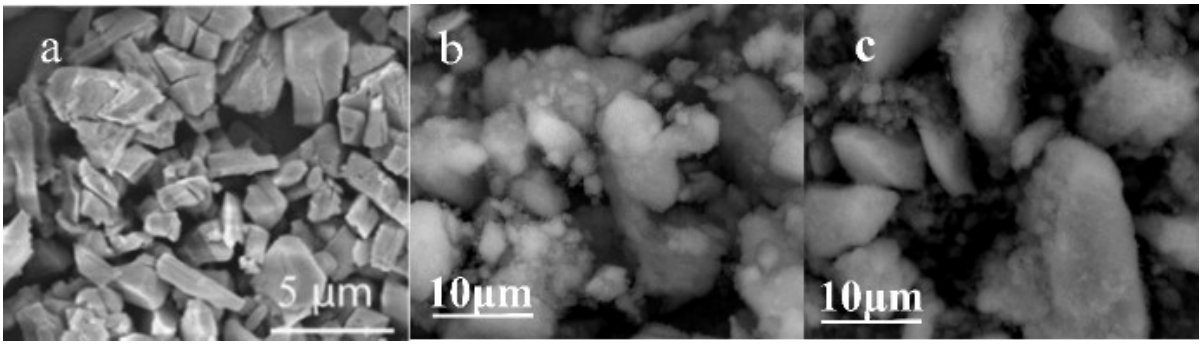
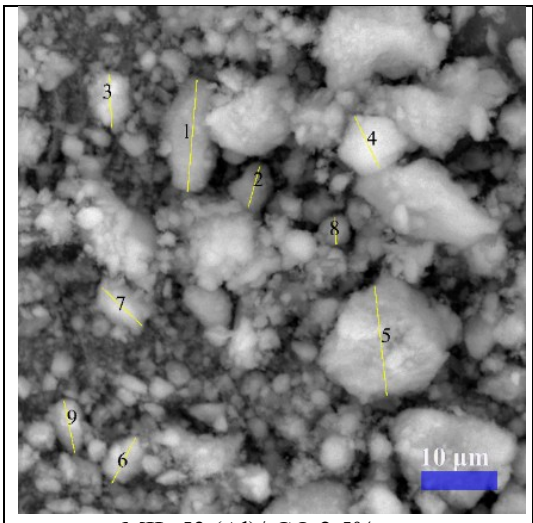


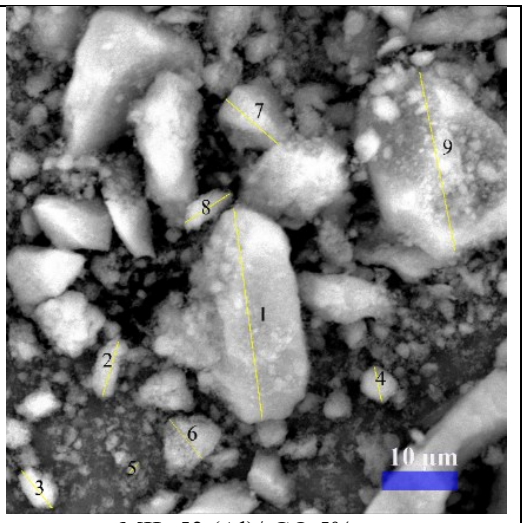
Figure 5. SEM profile for (a) MIL-53 (Al) [30], (b) MIL-53 (Al)/rGO 2.5% wt. (c) MIL-53/rGO 5% wt.

Fig. 5 shows that the addition of rGO ratio to the photocatalyst material leads to the coverage of the MIL-53 (Al) surface, resulting in larger crystallite sizes compared to the composite material with a lower rGO content and without rGO. As the crystallite size increases, the surface area decreases as shown in Table 1. The crystallite size measurements, presented in Table 1, show that the amount of rGO strongly influences the crystallite size added to the photocatalyst composition. As the rGO ratio increases, most MIL-53(Al) MOF surface interacts with rGO sheets. Since rGO still contains many oxygen-based functional groups, the interaction between crystallites becomes stronger due to hydrogen interaction, leading to an increase in crystallite size.

Table 1. The effect of rGO ratio on the crystallite size of the photocatalyst-composite material.



MIL-53 (Al)/rGO 2.5% wt.



MIL-53 (Al)/rGO 5% wt.

No Particle	**Length of Particle (µm)**	**No Particle**	**Length of Particle (µm)**
1	10.15	1	28.32
2	4.63	2	8.04
3	5.92	3	6.05
4	6.27	4	4.74
5	13.47	5	1.95
6	4.49	6	9.22
7	5.52	7	9.51
8	3.03	8	7.46
9	7.29	9	25.08
Average	6.75	*Average*	11.15

https://doi.org/10.31436/iiumej.v26i2.3556

SEM-EDS analysis was performed to investigate the chemical composition of MIL-53(Al)/rGO. The EDX spectra (Fig. 6) of MIL-53(Al)/rGO showed that the amount of aluminium decreases with the increase in the rGO weight ratio. In Fig. 6, for MIL-53 (Al)/rGO 2.5% wt., the atomic percentage of Al is 10.04% when the rGO ratio is increased to 5% wt. (MIL-53 (Al)/rGO 5% wt.). The atomic percentage of Al decreases to 5.43%, nearly halving. The EDX spectra results qualitatively match the composition of the two variations of the photocatalyst composite materials well.

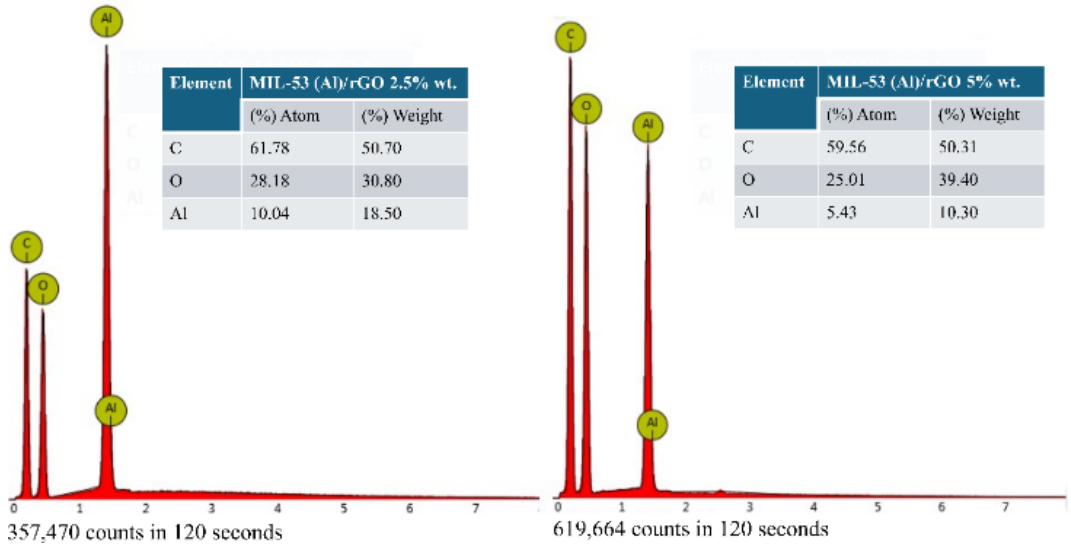


Figure 6. Elemental composition for each photocatalyst material

3.5. Nitrogen Isotherm adsorption

Fig. 7 shows that the isothermal curve of all photocatalyst samples shows type III based on the IUPAC classification.

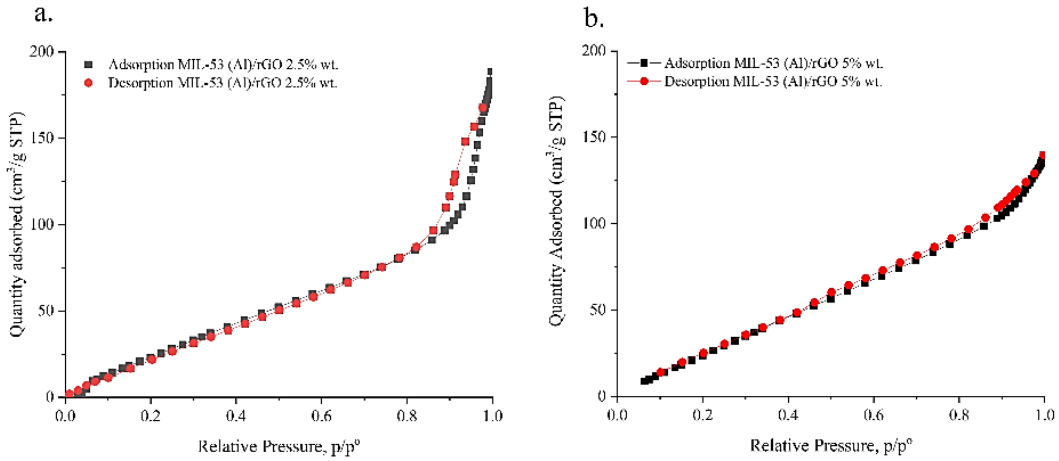


Figure 7. N_2 adsorption-desorption for synthesized MIL-53 (Al)/rGO. (a) 2.5% wt. rGO, (b) 5% wt. rGO.

The type-III adsorption isotherm, like the type-II isotherm, shows a general increase in adsorption with increasing pressure or concentration. As shown in Fig. 7, this rapid rise continues until the pressure is equal to the saturation pressure. At this point, adsorption is typically stable after relative pressure larger than 0.8, suggesting that the surface is saturated by adsorbate [32].

https://doi.org/10.31436/iiumej.v26i2.3556

3.6. Photocatalyst Performances

To evaluate the performance of the synthesized photocatalytic materials, tests were conducted to degrade two organic dyes, namely methylene blue (MB) 20 ppm and methyl orange (MO) 20 ppm, under UV and sunlight irradiation. The results of these tests are shown in Fig. 8.

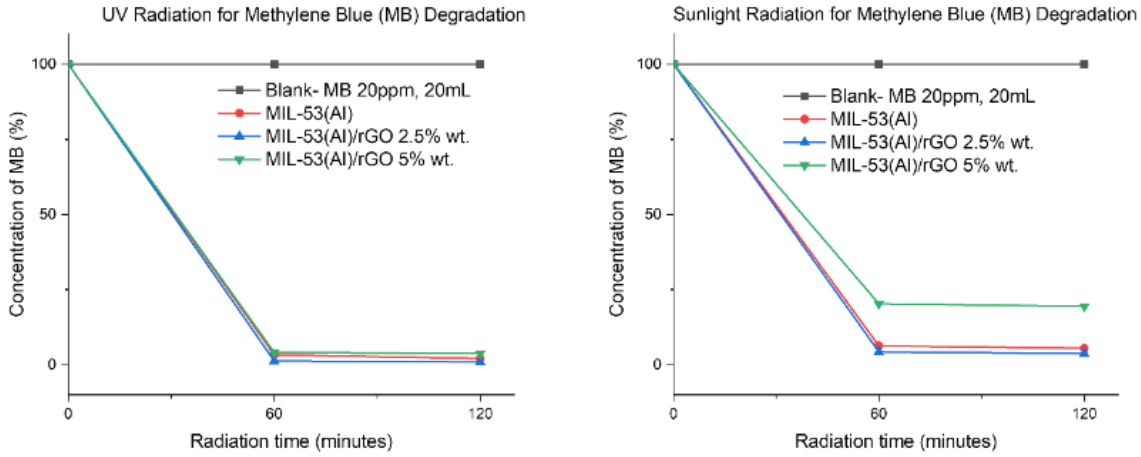


Figure 8. Photocatalyst performance under UV and Sunlight irradiation in 20 ppm MB solution.

Fig. 8 shows the performance of each photocatalyst material in degrading a 20 ppm MB test solution. Generally, all photocatalyst materials can degrade the test solution by more than 75% during the testing period. The MIL-53(Al)/rGO 2.5% wt. composition exhibits the best performance in degrading the test solution. Under sunlight, this composition can degrade the test solution by up to 97%, while under UV light, the test solution can be degraded by up to 99%. These results show that the MIL-53(Al) photocatalyst yields higher degradation of the test solution under sunlight than the MIL-53(Al)/rGO 5% wt. material shows slightly lower performance than the MIL-53(Al)/rGO 2.5% wt. composition. The poor photocatalytic performance of MIL-53(Al)/rGO 5% wt. under visible light, compared to other photocatalyst materials, is likely attributed to the larger crystallite size of this material resulting from the highest rGO ratio of 5%. As is well known, larger crystallite sizes can lead to a reduction in the number of active sites on the catalyst surfaces [33], facilitate faster recombination of electrons and holes [34], widen the bandgap [35], and enhance the stability of the photocatalyst material [36]. Ultimately, the increased crystallite size in the photocatalyst reduces its performance in degrading organic dye molecules.

A similar trend was observed for the methylene orange (MO) test solution. The photocatalyst with 2.5% wt rGO exhibited the best performance in degrading the 20 ppm MO solution within 60 and 120 minutes, under UV light and sunlight. The photocatalytic performance in the MO solution is presented in Fig. 9.

https://doi.org/10.31436/iiumej.v26i2.3556

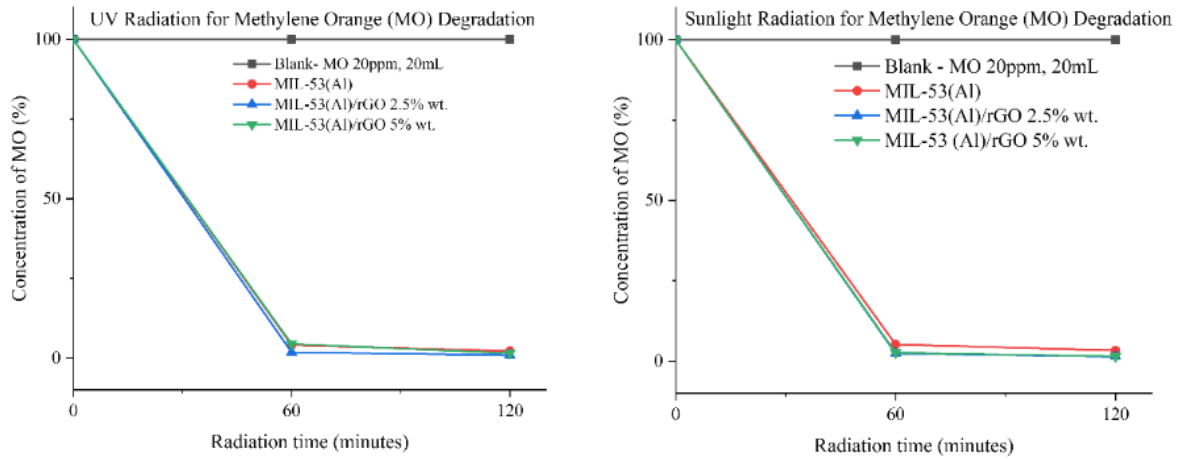


Figure 9. Photocatalyst performance under UV and Sunlight irradiation in 20 ppm MO solution.

3.7. General Mechanism of MIL-53(Al)/rGO Photocatalyst.

Methylene blue and methylene orange are synthetic dyes commonly used in various applications, including in the textile and pharmaceutical industries. Its decomposition under UV and sunlight irradiation in the presence of a photocatalyst involves photochemical reactions that lead to the breakdown of the dye molecules into smaller, non-toxic components. This process can be enhanced by using photocatalysts or other semiconductors, which facilitate photodegradation [24][37].

Here is an explanation of the mechanism of methylene blue and methylene orange decomposition using photocatalysts under UV and sunlight irradiation. UV-Irradiation: When UV light (typically in the range of 200-400 nm) is irradiated onto the photocatalyst, MIL-53 (Al), it excites the electrons (e^-) in the photocatalyst's valence band to the conduction band due to its relatively small bandgap. This creates electron-hole pairs, where the photogenerated electron (e^-) in the conduction band and hole (h^+) in the valence band are highly reactive. On the other hand, Sunlight Irradiation contains a broader spectrum, including visible light. To enhance the photocatalysis under sunlight, doping MIL-53(Al) with substances like rGO can allow the catalyst to absorb visible light and become activated under sunlight [9].

The electron-hole pairs generated in the photocatalyst due to UV or sunlight irradiation can migrate to the surface of the photocatalyst. The electron in the conduction band can participate in reducing species (such as oxygen molecules or water), while the hole in the valence band can oxidize other species (such as water or hydroxyl ions) [9]. Electron in the conduction band reacts with dissolved oxygen (O_2) in the solution to form superoxide anions ($O_2^{\bullet-}$), or they can reduce oxygen into hydrogen peroxide (H_2O_2)[38]. This leads to the formation of various reactive oxygen species (ROS), such as hydroxyl radicals (•OH) [38], [39], which are highly reactive and play a crucial role in the degradation of MB and MO.

$$O_2 + e^- \rightarrow O_2^{\bullet-} \text{ (superoxide anion)} \quad (1)$$

$$O_2^{\bullet-} + H_2O \rightarrow H_2O_2 + OH^- \text{ (hydrogen peroxide and hydroxyl ion)} \quad (2)$$

On the other side, the hole in the Valence Band reacts with water (H_2O) or hydroxide ions (OH^-) on the surface of the photocatalyst to generate hydroxyl radicals (•OH), which are strong oxidants capable of attacking and degrading MB molecules.

$$H_2O + h^+ \rightarrow \bullet OH + H^+ \text{ (formation of hydroxyl radicals)} \quad (3)$$

https://doi.org/10.31436/iiumej.v26i2.3556

$$OH^- + h^+ \rightarrow \bullet OH \text{ (formation of hydroxyl radicals)} \quad (4)$$

The reactive oxygen species (ROS), such as hydroxyl radicals (•OH), superoxide anions ($O_2\bullet^-$), and hydrogen peroxide (H_2O_2), are highly reactive and can attack the methylene blue and methylene orange molecules. The degradation of MB and MO primarily involves the following steps. The ROS oxidizes the aromatic structure of MB and MO, breaking the chemical bonds in the dye and leading to the formation of smaller, less complex molecules, eventually mineralizing to carbon dioxide (CO_2) and water (H_2O)[39]. The next step is demethylation and decolorization. The methyl groups on the MB and MO molecules may be oxidized or broken down, leading to a loss of color. The photodegradation process of MO and MB is presented schematically in Fig. 10.

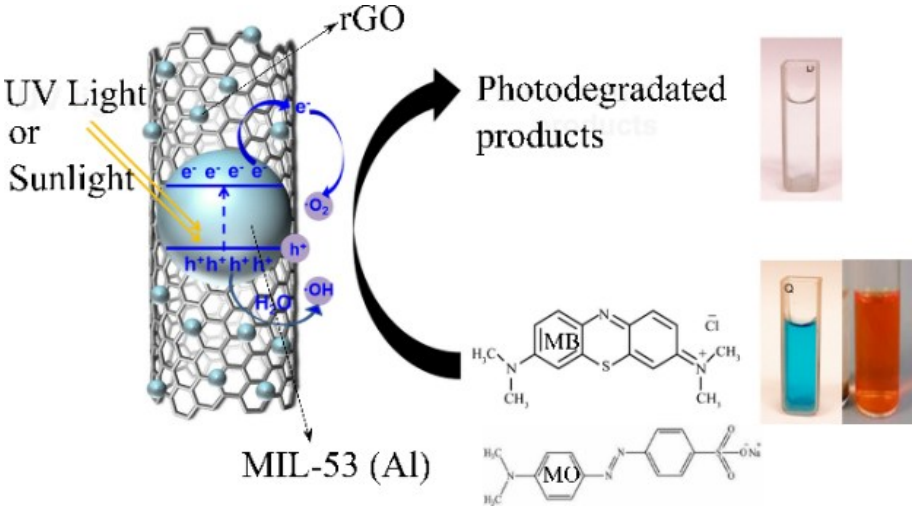


Figure 10. The process of photodegradation of MO and MB using MIL-53 (Al)/rGO

The breakdown of the chromophore (the part of the molecule responsible for its color) leads to the decolorization of MB and MO. From our experimental data, the process of dye decomposition is more efficient under UV light because of the high energy available to activate the photocatalyst. Under sunlight, the process can still occur, but it may be slower or require modified photocatalysts that can absorb visible light. Furthermore, factors like temperature and pH can improve this mechanism, and the presence of additional substances like co-catalysts or sacrificial agents can enhance the efficiency of MB and MO degradation.

4. CONCLUSION

The photocatalyst materials were successfully synthesized via the solvothermal method, producing three variants: MIL-53(Al), MIL-53(Al)/rGO (2.5 wt.%), and MIL-53(Al)/rGO (5 wt.%). FTIR analysis confirmed their composition, with characteristic absorption peaks at 1417 cm^{-1}, 1511 cm^{-1}, and 1603 cm^{-1}, indicating the presence of carbonyl and carboxyl functional groups. The IR spectra at wave numbers 1001, 755, and 593 cm^{-1} indicate the vibrational interaction between metal ions and oxygen bonds. XRD characterization confirmed the crystalline structure of the materials. MIL-53(Al) showed distinct diffraction peaks consistent with reported data, while rGO exhibited an amorphous peak at $2\theta = 23°$, characteristic of graphite. Additionally, combining MIL-53(Al) with rGO caused a shift in diffraction peaks at 2θ and peak broadening due to the amorphous nature of rGO. However, both materials retained their characteristic peaks, confirming successful synthesis. The incorporation of rGO

influenced the photocatalyst's morphology, with higher rGO content nearly doubling its size, decreasing surface area, and affecting dye degradation efficiency. N_2 adsorption-desorption data revealed that larger particles had reduced gas adsorption capacity, which impacted their photocatalytic activity. Additionally, introducing a small amount of rGO reduced the bandgap from ~4 eV to 3.8 eV, though further rGO addition had no significant effect. Photocatalytic tests showed that MIL-53(Al)/rGO (2.5 wt.%) achieved the highest degradation efficiency, breaking down MB and MO dyes almost completely (~100%) within 120 minutes under both UV and sunlight irradiation.

ACKNOWLEDGEMENT

This work was supported by the Early-Career Researcher (Penelitian Dosen Pemula) Grant Scheme 2024, from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, contract number 105/E5/PG.02.00.PL/2024.

REFERENCES

[1] L. Liu, Z. Chen, J. Zhang, D. Shan, Y. Wu, L. Bai, B. Wang. Treatment of industrial dye wastewater and pharmaceutical residue wastewater by advanced oxidation processes and its combination with nanocatalysts: A review. *J. Water Process Eng.* 2021; 42(22): 102122. doi: 10.1016/j.jwpe.2021.102122.

[2] S. Jorfi, S. Pourfadakari, and B. Kakavandi. A new approach in sono-photocatalytic degradation of recalcitrant textile wastewater using MgO@Zeolite nanostructure under UVA irradiation. Chem. Eng. J. 2018 Jul; 343: 95–107. doi: 10.1016/j.cej.2018.02.067.

[3] M. S. Jami, M. Mel, A. R. M. Ariff, Q. M. Abdulazeez. Investigation of bioflocculant as dewatering aid in sludge treatment. IIUM Eng. J. 2018; 19(1): 15–23. doi: 10.31436/iiumej.v19i1.735.

[4] J. Wang, S. Wang. Reactive species in advanced oxidation processes: Formation, identification and reaction mechanism. Chem. Eng. J.2020; 401:126158. doi: 10.1016/j.cej.2020.126158.

[5] G. I. Lupu, C. Orbeci, L. Bobirică, C. Bobirică, L. F. Pascu. Key Principles of Advanced Oxidation Processes: A Systematic Analysis of Current and Future Perspectives of the Removal of Antibiotics from Wastewater. Catalysts. 2023; 13(9). doi: 10.3390/catal13091280.

[6] S. A. Younis, E. E. Kwon, M. Qasim, K-H. Kim, T. Kim, D. K, X. Dou, I. Ali. Metal-organic framework as a photocatalyst: Progress in modulation strategies and environmental/energy applications. Progress in Energy and Combustion Science. 2020 Nov. 01;81. doi: 10.1016/j.pecs.2020.100870.

[7] F. Zhang, X. Wang, H. Liu, C. Liu, Y. Wan, Y. Long, Z. Cai. Recent advances and applications of semiconductor photocatalytic technology. Appl. Sci. 2019; 9(12). doi: 10.3390/app9122489.

[8] M. Nemiwal, V. Subbaramaiah, T. C. Zhang, D. Kumar. Recent advances in visible-light-driven carbon dioxide reduction by metal-organic frameworks. Sci. Total Environ. 2021; 762: 144101. doi: 10.1016/j.scitotenv.2020.144101.

[9] Y. Chen, B. Y. Zhai, Y. N. Liang, Y. Li. Hybrid photocatalysts using semiconductor/MOF/graphene oxide for superior photodegradation of organic pollutants under visible light. Mater. Sci. Semicond. Process. 2020 Mar;107. doi: 10.1016/j.mssp.2019.104838.

[10] M. A. Hassaan, M. A. El-Nemr, M. R. Elkatory, S. Ragab, V. C. Niculescu, A. El Nemr. Principles of Photocatalysts and Their Different Applications: A Review. 2023; 381(6). Springer International Publishing.

[11] N. Mohammed, P. Palaniandy, F. Shaik, H. Mewada. Experimental and Computational Analysis

for Optimization of Seawater Biodegradability Using Photo Catalysis. IIUM Eng. J. 2023;24(2):11–33. doi: 10.31436/iiumej.v24i2.2650.

[12] N. Sun, X. Si, L. He, J. Zhang, Y. Sun. Strategies for enhancing the photocatalytic activity of semiconductors. Int. J. Hydrogen Energy. 2024 Mar; 58: 1249–1265. doi: 10.1016/J.IJHYDENE.2024.01.319.

[13] C. B. Anucha, I. Altin, E. Bacaksiz, V. N. Stathopoulos. Titanium dioxide (TiO_2)-based photocatalyst materials activity enhancement for contaminants of emerging concern (CECs) degradation: In the light of modification strategies. Chem. Eng. J. Adv. 2022 Feb; 10:100262. doi: 10.1016/j.ceja.2022.100262.

[14] V. García-Salcido, P. Mercado-Oliva, J. L. Guzmán-Mar, B. I. Kharisov, L. Hinojosa-Reyes. MOF-based composites for visible-light-driven heterogeneous photocatalysis: Synthesis, characterization and environmental application studies. J. Solid State Chem.2022 Mar; 307: 122801. doi: 10.1016/J.JSSC.2021.122801.

[15] Y. Bai, S. Zhang, S. Feng, M. Zhu, S. Ma. The first ternary Nd-MOF/GO/Fe3O4nanocomposite exhibiting an excellent photocatalytic performance for dye degradation. Dalt. Trans. 2020 Aug; 49(31): 10745–10754. doi: 10.1039/d0dt01648a.

[16] E. Özcan, Z. Mermer, Y. Zorlu. Metal-organic frameworks as photocatalysts in energetic and environmental applications. Turkish J. Chem. 2023;47(5): 1018–1052. doi: 10.55730/1300-0527.3592.

[17] W. Ma, L. Yu, P. Kang, Z. Chu, Y. Li. Modifications and Applications of Metal-Organic-Framework-Based Materials for Photocatalysis. Molecules. 2024; 29(24). doi: 10.3390/molecules29245834.

[18] A. Sugie, K. Nakano, K. Tajima, I. Osaka, H. Yoshida. Dependence of Exciton Binding Energy on Bandgap of Organic Semiconductors. J. Phys. Chem. Lett. 2023;14(50):11412–11420. doi: 10.1021/acs.jpclett.3c02863.

[19] P. H. M. Andrade, C. Volkringer, T. Loiseau, A. Tejeda, M. Hureau, A. Moissette. Band gap analysis in MOF materials: Distinguishing direct and indirect transitions using UV–vis spectroscopy. Appl. Mater. Today. 2024 Jan; 37: 102094. doi: 10.1016/j.apmt.2024.102094.

[20] S. Javanbakht, M. Pooresmaeil, H. Hashemi, H. Namazi. Carboxymethylcellulose capsulated Cu-based metal-organic framework-drug nanohybrid as a pH-sensitive nanocomposite for ibuprofen oral delivery. Int. J. Biol. Macromol. 2018 Nov;119: 588–596. doi: 10.1016/j.ijbiomac.2018.07.181.

[21] M. Muttaqin, F. I. Hidayat. A modest method of synthesis Cu- based metal-organic frameworks using benzene dicarbocxylate as a ligan for promising candidate of flue gas CO_2 adsorption. J. Nat. 2021 Oct;21(3): 128–134. doi: 10.24815/jn.v21i3.20035.

[22] J. Zhang, C. Su, X. Xie, P. Liu, M. E. Huq. Enhanced visible light photocatalytic degradation of dyes in aqueous solution activated by HKUST-1: performance and mechanism. RSC Adv. 2020;10(61):37028–37034. doi: 10.1039/d0ra05275b.

[23] Y. Chen, B. Y. Zhai, Y. N. Liang, Y. Li. Hybrid photocatalysts using semiconductor/MOF/graphene oxide for superior photodegradation of organic pollutants under visible light. Mater. Sci. Semicond. Process. 2020 Mar;107: 104838. doi: 10.1016/J.MSSP.2019.104838.

[24] H. T. Vu, L. T. Tran, G. H. Le, Q. K. Nguyen, T. M. Vu, T. A. Vu. Synthesis and application of novel Fe-MIL-53/GO nanocomposite for photocatalytic degradation of reactive dye from aqueous solution. Vietnam J. Chem. 2019 Dec;57(6): 681–685. doi: 10.1002/vjch.201900055.

[25] M. Husnah, H. A. Fakhri, F. Rohman, A. H. Aimon, F. Iskandar. A modified Marcano method for improving electrical properties of reduced graphene oxide (rGO). Mater. Res. Express. 2017

Jun;4(6). doi: 10.1088/2053-1591/aa707f.

[26] C. G. Carson, K. Hardcastle, J. Schwartz, X. Liu, C. Hoffmann, R. A. Gerhardt, R. Tannenbaum. Synthesis and structure characterization of copper terephthalate metal-organic frameworks. Eur. J. Inorg. Chem. 2029;(16): 2338–2343. doi: 10.1002/ejic.200801224.

[27] C. G. Carson, G. Brunnello, S. G. Lee, S. S. Jang, R. A. Gerhardt, R. Tannenbaum. Structure solution from powder diffraction of copper 1,4- benzenedicarboxylate. Eur. J. Inorg. Chem. 2014;(12): 2140–2145. doi: 10.1002/ejic.201301543.

[28] P. Makuła, M. Pacia, W. Macyk. How To Correctly Determine the Band Gap Energy of Modified Semiconductor Photocatalysts Based on UV-Vis Spectra. J. Phys. Chem. Lett. 2018; 9(23): 6814–6817. doi: 10.1021/acs.jpclett.8b02892.

[29] S. Mundinamani. Large Area, Multilayer Graphene Films as a Flexible Electronic Material. ACS Omega. 2020 Jul;5(28): 17479–17485. doi: 10.1021/acsomega.0c01982.

[30] T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille, G. Ferey. A Rationale for the Large Breathing of the Porous Aluminum Terephthalate (MIL-53) Upon Hydration. Chem. - A Eur. J. 2004 Apr;10(6): 1373–1382. doi: 10.1002/chem.200305413.

[31] X. Qian, B Yadian, R. Wu, Y Long, K. Zhou, B. Zhu, Y. Huang. Structure stability of metal-organic framework MIL-53 (Al) in aqueous solutions. Int. J. Hydrogen Energy. 2013 Dec;38(36): 16710–16715. doi: 10.1016/j.ijhydene.2013.07.054.

[32] M. Burhan, M. W. Shahzad, K. C. Ng. Energy distribution function based universal adsorption isotherm model for all types of isotherm. Int. J. Low-Carbon Technol. 2018;13(3):292–297. doi: 10.1093/IJLCT/CTY031.

[33] J. Gorimbo, R. Chikati, P. Khangale, I. N. Beas, L. L. Mguni, D. Nkazi. Debunking the impact of crystallite/particle size in cobalt-based Fischer-Tropsch synthesis. Chemical Engineering Communications. 2024;211(8):1262–1287. Taylor and Francis Ltd. doi: 10.1080/00986445.2024.2341263.

[34] R. Chuliá-Jordán, E. J. Juarez-Perez. Short Photoluminescence Lifetimes Linked to Crystallite Dimensions, Connectivity, and Perovskite Crystal Phases. J. Phys. Chem. C. 2022 Feb;126(7): 3466–3474. doi: 10.1021/acs.jpcc.1c08867.

[35] B. M. John, S. W. Mugo, J. M. Ngaruiya. Dependence of Optical Band Gap on Crystallite Size of TiO_2 Thin Films Prepared Using Sol Gel Process. European Journal of Material Science.2021; 8(1): 1-12. doi: 10.2139/ssrn.3819728.

[36] A. B. D. Nandiyanto, R. Zaen, R. Oktiani. Correlation between crystallite size and photocatalytic performance of micrometer-sized monoclinic WO3 particles. Arab. J. Chem. 2020 Jan;13(1): 1283–1296. doi: 10.1016/j.arabjc.2017.10.010.

[37] A. Fawzy, H. Mahanna, M. Mossad. Effective photocatalytic degradation of amoxicillin using MIL-53(Al)/ZnO composite. Environ. Sci. Pollut. Res. 2022 Sep;29(45): 68532–68546. doi: 10.1007/s11356-022-20527-0.

[38] B. Wang, L. Shen, J. Xu, Z. Yang, Y. Chen, S. Chen, B. Li, C. Chen, H. Lin. Graphene oxide-assisted dispersion and assembly of photocatalytic self-cleaning MOF membrane for enhanced water purification. Sep. Purif. Technol. 2025 Apr;356:129928. doi: 10.1016/J.SEPPUR.2024.129928.

[39] Z. U. Zango, K. Jumbri, N. S. Sambudi, A. Ramli, N. H. H. Abu Bakar, B. Saad, M. N. H. Rozani, H. A. Isiyaka, A. H. Jagaba, O. Aldaghri, A. Sulieman. A critical review on metal-organic frameworks and their composites as advanced materials for adsorption and photocatalytic degradation of emerging organic pollutants from wastewater. Polymers. 2020 Nov;12(11):1–42. MDPI AG. doi: 10.3390/polym12112648.

https://doi.org/10.31436/iiumej.v26i2.3575

AN INTEGRATION METHOD FOR SUPPLIER SELECTION: A PRACTICAL STUDY IN THE INDONESIAN CONCRETE COMPANY

AGUS RISTONO*

Department of Industrial Engineering, University of Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia

**Corresponding author: agus.ristono@upnyk.ac.id*

(Received: 13 January 2025; Accepted: 10 March 2025; Published online: 15 May 2025)

ABSTRACT: The construction of toll roads in Indonesia has massively increased the need for concrete. Therefore, concrete manufacturing companies must select new raw material suppliers for concrete. This selection is based on several criteria, including qualitative and quantitative criteria. This study proposes a new methodology for supplier selection. The model captures decision-makers' dynamics in ranking suppliers by integrating quantitative and qualitative criteria. The first stage in the proposed model is determining the criteria using the Fuzzy Delphi Method (FDM). The second stage is weighting the criteria using the Fuzzy Step-wise Assessment Ratio (F-SWARA) II and the Method Based on Removal Effects of Criteria (MEREC). The last stage is supplier assessment using Fuzzy Measurement of Alternatives and Ranking According to Compromise Solution (F-MARCOS). One of Indonesia's largest concrete manufacturing companies was used to validate the proposed model. The results showed that the proposed model is valid and more effective than some previous methods. The company in the case study experienced a 5% reduction in production costs in the later stages of the toll project, following several months of implementation, indicating the effectiveness of the proposed methodology.

ABSTRAK: Pembinaan jalan tol di Indonesia telah meningkatkan keperluan konkrit secara besar-besaran. Oleh itu, syarikat pembuatan konkrit mesti memilih pembekal bahan mentah baru konkrit. Pemilihan ini berdasarkan beberapa kriteria, termasuk kriteria kualitatif dan kuantitatif. Kajian ini mencadangkan metodologi baru untuk pemilihan pembekal. Model ini menfokuskan pembuat keputusan dinamik dalam kedudukan pembekal dengan menyepadu kedua-dua kriteria kuantitatif dan kualitatif. Peringkat pertama dalam model yang dicadangkan ini adalah menentukan kriteria menggunakan Kaedah Rawak Delphi (FDM). Peringkat kedua adalah menimbang kriteria menggunakan Nisbah Pentaksiran Rawak Langkah-Bijak (F-SWARA) II dan kaedah berdasarkan Kriteria Kesan Pengasingan (MEREC). Peringkat terakhir merupakan penilaian pembekal menggunakan Solusi Kompromi berdasarkan Pengukuran Rawak Alternatif dan Kedudukan (F-MARCOS). Salah satu syarikat pembuatan konkrit terbesar di Indonesia telah digunakan bagi mengesahkan model yang dicadangkan. Dapatan kajian menunjukkan bahawa model yang dicadangkan ini adalah sah dan lebih berkesan berbanding beberapa kaedah sebelumnya. Syarikat dalam kes kajian mengalami pengurangan 5% dalam kos pengeluaran pada peringkat akhir projek tol, iaitu selepas beberapa bulan pelaksanaan, menunjukkan keberkesanan metodologi yang dicadangkan.

KEYWORDS: *Fuzzy Delphi Method, F-SWARA II, MEREC, F-MARCOS, Supplier selection*

https://doi.org/10.31436/iiumej.v26i2.3575

1. INTRODUCTION

The Ministry of Public Works (PU) Indonesia plans to construct freeways or toll roads along 17,865.43 kilometers (km) between 2025 and 2040. Meanwhile, Java is the island most connected to toll roads, with an operating toll road length of 1,632.63 km. However, this length is still far from the total planned toll road length, which is 4,688.94 km. Due to this policy implemented by the Indonesian government, concrete will be required. PT Wika Beton, one of Indonesia's leading concrete manufacturers, will enhance its production capacity to prepare for the anticipated rise in demand for concrete. As a result, acquiring materials from additional raw material providers is essential.

PT Wika Beton selects its suppliers not for the short term but to establish long-term partners. This is mainly because raw materials are an essential component and play a crucial role in the development of our organization. In the long term, suppliers are required to ensure the availability of essential materials. Consequently, it is imperative to identify the appropriate supplier. Suppliers are indispensable to the organization's supply chain. Supply chain management significantly influences enterprise performance and profitability. Consequently, it is imperative to cultivate positive relationships with suppliers and preserve a robust supply chain.

Supplier selection can be done using Multi-Criteria Decision Making (MCDM). There are manyCDM, such as Analytical Hierarchy Process (AHP), Analytic Network Process (ANP), Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE), VIšekriterijumska Optimizacija I Kompromisno Rešenje (VIKOR), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Elimination et Choix Traduisant la Réalité/Elimination and Choice Expressing Reality (ELECTRE), Multi-Objective Optimization based on Ratio Analysis (MOORA), Complex Proportional Assessment (COPRAS), and the Measurement of Alternatives and Ranking according to the Compromise Solution (MARCOS). Recent studies show the latest developments in supplier selection methods, including four strategies [1]. They are (1) utilizing fuzzy logic, which is suitable for the dynamic and data-rich environment of Industry 4.0; (2) using mixed MCDM strategies to handle the complex problems of supplier selection in Industry 4.0; (3) using hybrid fuzzy MCDM techniques, which offer a more thorough and complex decision-making process [1].

The MARCOS method demonstrates high stability in ranking options than others [2]. It can identify the optimal solution regardless of the number of alternatives or the weighting method employed [3]. The criteria in MARCOS consist of qualitative and quantitative criteria, so both need to be weighted. Qualitative criteria can only be assessed using the decision maker's perception, while quantitative criteria can be evaluated based on historical supplier data. The primary limitation of subjective weighting is its ineffectiveness as the number of criteria increases, including additional criteria, which can lead to decreased accuracy in these preferences [4]. The combined weighting approach reduces the potential bias associated with a singular subjective or objective weight [4].

Throughout its evolution, MARCOS has found extensive utility in addressing issues related to supplier selection. Badi and Pamucar [5] used the MARCOS system to choose the best supplier and the Grey theory to establish the criteria weights in steel companies. Chattopadhyay et al. [6] proposed the D number theory to incorporate MARCOS into the supplier selection process in the iron and steel sector. Taş et al. [7] employed F-MARCOS and Spherical Fuzzy Stepwise Weight Assessment Ratio Analysis (SF-SWARA) to identify environmentally conscious vendors for textile businesses. To help textile businesses choose ecologically friendly vendors, Tus and Aytac-Adali [8] presented a strategy that merges F-

SWARA and F-MARCOS. Abdulla et al. (2023) suggested integrating MARCOS with machine learning for supplier selection. Y. Wang et al. [9] present two methods for selecting sustainable food suppliers: PF-CRITIC, which uses Pythagorean fuzzy measurement of alternatives and ranking to compromise, and PF-MARCOS, which uses weights to determine the criterion. The position of the proposed method in supplier selection research utilizing MARCOS is illustrated in Table 1.

Table 1. Supplier selection research using MARCOS

Ref	Criteria selection	Criteria weighting		Alternative selection		Case
		Objective method	Subjective method	Original MARCOS	Extended MARCOS	
[5]	-	-	Grey method	MARCOS	-	Steel-making company
[6]	-	-	-	-	D-MARCOS	Steel and iron industry
[7]	-	-	SF-SWARA	-	Fuzzy MARCOS	Green suppliers in textile companies
[8]	-	-	F-SWARA	-	Fuzzy MARCOS	Green suppliers in textile companies
[10]	-	-	-	MARCOS	-	General
[11]	-	-	Machine learning	MARCOS	-	Oil and gas sector
[9]	-	PF-CRITIC	-	-	PF-MARCOS	Sustainable food suppliers
Proposed method	Fuzzy Delphi	MEREC	F-SWARA II	-	Fuzzy MARCOS	Concrete manufacturing company

Previous research on supplier selection utilizing MARCOS only partially used subjective or objective weighting. Such as Badi and Pamucar [5], Taş et al. [7], Tus and Aytac-Adali [8], and Abdulla et al. [11] used only subjective criteria for weighting. Whereas Y. Wang et al. [9] used only objective weighting. However, not all criteria can be weighed objectively or subjectively. It depends on its characteristics. So, it is prone to bias [4]. In addition, the criteria in previous studies already exist. The process of determining these criteria is essential in selecting suppliers [4], The proposed method then employs Fuzzy Delphi for criteria selection and integrates objective and subjective criteria weighting.

2. METHODS

The proposed method has three stages. The first stage is determining the supplier selection criteria using FDM. The second stage is weighting criteria using F SWARA II and MEREC. The last stage is supplier assessment. Some potential criteria are selected using FDM to produce the chosen criteria. The chosen criteria consist of qualitative criteria and quantitative criteria. Qualitative criteria are weighted using F SWARA II, while quantitative criteria are weighted using MEREC. Supplier performance assessment using F MARCOS based on weighted criteria. The criteria are based on the weighting of F SWARA II and MEREC.

The selection of criteria uses FDM because FDM can reduce subjectivity in decision-making based on expert opinion, produce more objective agreement in the early stages of research or selection of essential factors, and does not require many iterations compared to classical Delphi [12]. The objective criteria weighting uses MEREC because MEREC is a more

efficient, precise, and accurate objective weighting than other methods [3]. The MEREC is recommended for application because of its high accuracy [3]. The subjective criteria weighting uses SWARA because SWARA is more straightforward than other subjective methods, and experts can participate more spontaneously [13]. Supplier assessment uses MARCOS because it is more stable than other MCDM [2] and can always find the best one, no matter how many options are considered or how the weights are distributed [3].

2.1. Fuzzy Delphi Method

Compared to other expert-based methods (such as AHP or fuzzy AHP), the FDM can obtain expert consensus to determine essential factors or variables in a study [14]. AHP or fuzzy AHP can only choose the priority weight between criteria and alternatives in the multi-criteria decision-making process [1]. The FDM's steps are:

- **Step 1**: Expert panel. This document includes the division head and all procurement, building, and commissioning staff members. They have been employed by comparable organizations for over 15 years, which is why their responses are credible. Ten specialists satisfy these procurement, building, and commissioning prerequisites. This investigation exclusively employed the cases of PT. Wika Beton. All of the specialists selected are experts in supplier selection at PT. Wika Beton, which is why they are the population used in FDM.
- **Step 2**: List of the potential criteria. The most widely used criteria in the literature are price, reject, delivery [6], and service [8]. Price and reject are critical criteria because they directly affect the company's profit. The delivery and service criteria affect production lead time and its impact on increasing penalty costs. This will reduce the company's profit. In the FGD with decision makers at the largest concrete company in Indonesia, additional criteria were obtained: communication and flexibility. These two criteria have an impact on company trust. The selection of suppliers in concrete manufacturing is not only temporary but also for the long term.
- **Step 3**: The experts measure each criterion using linguistic variables (LV). The $\mu_{\tilde{n}}$ ranges from 1 to 0 in the LV, from Very Suitable to Very Unsuitable, as seen in Table 2. A score is represented as a TFN $\tilde{n}_i^k = (n_{1i}^k, n_{2i}^k, n_{3i}^k)$ [14].
- **Step 4.a**: Determining threshold value (Th_i) using Eq. (1) [14]. The minimum threshold on each criterion i is 0.2 [15].

$$Th_i = \left(\frac{1}{K}\right)\sum_{k=1}^{K}\sqrt{\frac{1}{3}\left[(n_{i1}^k - a_{i1})^2 + (n_{i2}^k - a_{i2})^2 + (n_{i3}^k - a_{i3})^2\right]} \quad (1)$$

- **Step 4.b**: Testing the experts' consensus using Eq. (2) [14]. The minimum level of experts' consensus is 75% for every criterion [15].

$$EA_i = \left\{\frac{\sum_{k=0}^{K}\left(c_i^k\right)\begin{cases} c_i^k=1, if\ d \le 0.2 \\ 0, \quad otherwise \end{cases}}{K}\right\} \times 100\% \quad (2)$$

- **Step 4.c**: Defuzzification. The aggregate TFN is calculated using Eqs. (3) [14]. Then the calculation of the α-cut using Eq. (4) [14] which this value must be more than 0.5 [15].

$$\tilde{n}_i = \left(\min\left(n_{i1}^k\right), \left[\prod_{k=1}^{K} n_{i2}^k\right]^{\frac{1}{K}}, \max\left(n_{i3}^k\right)\right) \quad (3)$$

$$\alpha_{cut} = n_{i1} + \frac{(n_{i3} - n_{i1}) + (n_{i2} - n_{i1})}{3} \quad (4)$$

The diagram of the FDM can be seen in Figure 1. After forming the expert panel, the second step is collecting data from experts. The questionnaire is given to the expert panel. The answers are converted to fuzzy numbers. After that, the fuzzy average calculation is carried out. The third step is defuzzification (conversion to crisp value) to obtain a precise value. The fourth step is determining the Expert consensus, namely, the minimum threshold on each criterion i is 0.2; the minimum level of experts' consensus is 75% for each criterion; and the α-cut value must be more than 0.5. If the factors do not reach consensus, reiterate with a revised questionnaire. Final Decision Factors that reach consensus are included in the analysis. After iteration, factors that still do not meet consensus can be removed or reviewed.

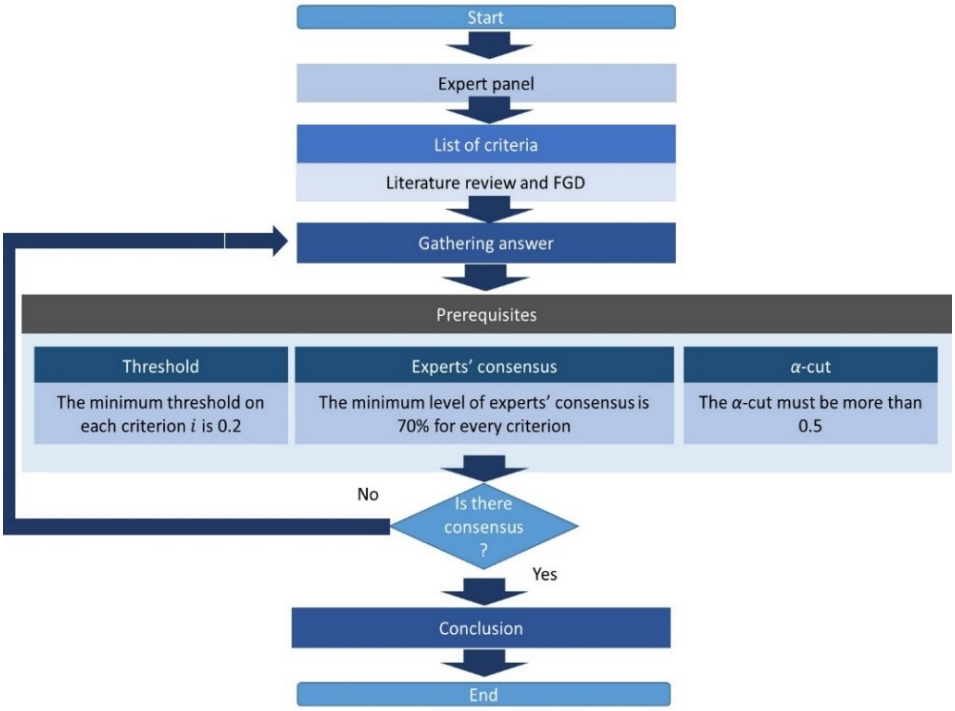


Figure 1. Flow chart of the FDM [12].

Table 2. Scale of fuzzy numbers in the FDM

No	5-Point LV	Most reasonable (n_2)	Pessimistic value (n_1)	Optimistic value (n_3)
1	Very Suitable	0.9	0.8	1
2	Suitable	0.7	0.6	0.8
3	Acceptable	0.5	0.4	0.6
4	Unsuitable	0.3	0.2	0.4
5	Very Unsuitable	0.1	0	0.2

2.2. Fuzzy SWARA II

SWARA II is an improved version of SWARA, as presented by Keshavarz-Ghorabaee [16]. However, specific adjustments to the methodology make SWARA II more attainable and practical for decision-makers. This study proposed integrating fuzzy theory with the SWARA II to enhance its intrinsic ambiguity and account for data imprecision and uncertainty. The steps of the Fuzzy SWARA II are as follows and can be seen in Figure 2.

- **Step 1**: Determination of the rank of the j^{th} criterion using the symbol t_j $(t_j \epsilon\{1, 2, 3, \dots, n\})$.
- **Step 2**: Ask the decision-maker to specify each criterion's relative preference $(\widetilde{RP})$ by comparing it $([t_j]^{th})$ with the following criterion $([t_j + 1]^{th})$ in the prioritized list from Step 1. This study employs linguistic variables (LVs) and associated values, as illustrated in Table 3.
- **Step 3**: Determine each criterion's fuzzy number of the preference degree $(\widetilde{PD})$ using Eq. (5).

$$\widetilde{PD}_{[t_j]} = u\left[\tilde{P}_{[t_j]}\right] = \left[\frac{\tilde{P}_{[t_j]}}{10}\right]^2 \tag{5}$$

- **Step 4**: Calculate the fuzzy number of the relative weighting coefficients using Eq. (6). The coefficients are established based on the arrangement of each criterion within the ordered list and the corresponding fuzzy number values of $\widetilde{PD}$. Starting from the nth criterion, the following equation is utilized for calculation.

$$\tilde{V}_{[t_j-1]} = \left(1 + \widetilde{PD}_{[t_j-1]}\right).\tilde{V}_{[t_j]} \tag{6}$$

- **Step 5**: The relative weights of the evaluation criteria are determined using Eq. (7).

$$\widetilde{W}_j^s = \frac{\tilde{V}_{[t_j]}}{\tilde{V}_{[t_1]} \oplus \tilde{V}_{[t_2]} \oplus \dots \tilde{V}_{[t_n]}} \tag{7}$$

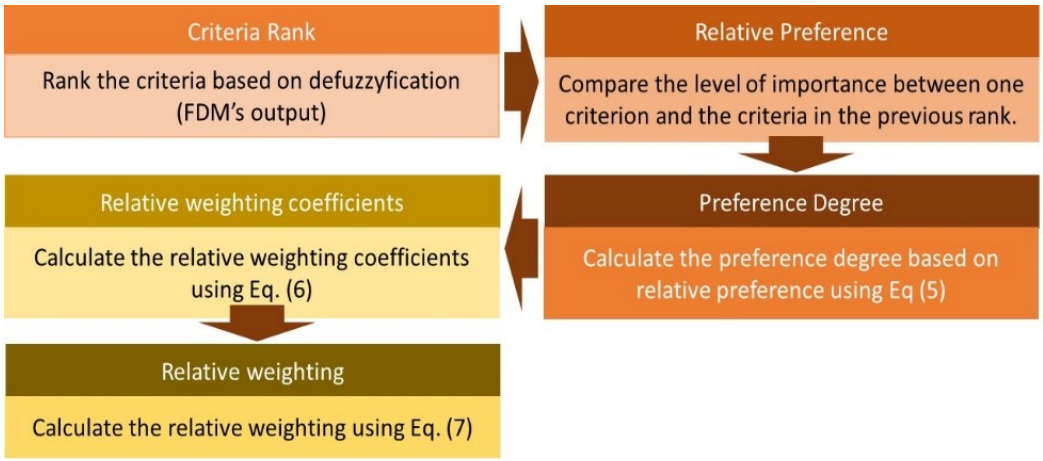


Figure 2. Step of the F SWARA II.

2.3. MEREC

The MEREC stages in the proposed method can be seen in Çelebi Demirarslan et al. [17]. The MEREC calculates weights based on the contribution of each criterion to the total difference between alternatives [17]. In contrast, the CRITIC is based on the correlation between criteria [18], and Entropy is based on the probability distribution of data in the criteria [19]. Therefore, the CRITIC is ineffective if there is no correlation between the criteria [18], and conversely, Entropy is useless if there is a correlation between criteria [19]. However, the MEREC is unaffected by the presence or absence of correlation between criteria [17]. The MEREC is not very sensitive to data scale because it is based on variation between criteria [17]. At the same time, the CRITIC is less flexible for data with different scales, especially if

the data is not normalized first [18]. The entropy is sensitive to un-normalized data, requiring more careful preprocessing [19]. Although the Entropy is easy to implement and calculate, it is less accurate if the data has a uniform distribution pattern, because the weights can become insignificant [19]. At the same time, the MEREC is not affected by such patterns.

2.4. Fuzzy MARCOS

The proposed method's F-MARCOS stages can be seen in Stanković et al. [20]. Table 3 illustrates the transformation of the linguistic variable into a fuzzy initial decision matrix.

Table 3. Scale of fuzzy numbers

No	9-Point LV in the F MARCOS [20]	Relative Preference (n_1, n_2, n_3)	9-Point LV in the F SWARA II	Relative Preference (n_1, n_2, n_3)
1	Extremely poor (EP)	(1, 1, 1)	Very very low (VVL)	(1, 1, 2)
2	Very poor (VP)	(1, 1, 3)	Very low (VL)	(1, 2, 3)
3	Poor (P)	(1, 3, 3)	Low (L)	(2, 3, 4)
4	Medium poor (MP)	(3, 3, 5)	Medium-low (ML)	(3, 4, 5)
5	Medium (M)	(3, 5, 5)	Medium (M)	(4, 5, 6)
6	Medium good (MG)	(5, 5, 7)	Medium-high (MH)	(5, 6, 7)
7	Good (G)	(5, 7, 7)	High (H)	(6, 7, 8)
8	Excellent (VG)	(7, 7, 9)	Very high (VH)	(7, 8, 9)
9	Extremely good (EG)	(7, 9, 9)	Very very high (VVH)	(8, 9, 9)

Table 4. Linguistic variable of the criteria assessment

		Criteria						
		C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	D_1	Very Suitable	Very Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
2	D_2	Very Suitable	Very Suitable	Suitable	Suitable	Suitable	Very Suitable	Acceptable
3	D_3	Suitable	Very Suitable	Very Suitable	Very Suitable	Very Suitable	Suitable	Suitable
4	D_4	Very Suitable	Suitable	Suitable	Suitable	Suitable	Acceptable	Very Suitable
5	D_5	Very Suitable	Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
6	D_6	Suitable	Very Suitable	Suitable	Acceptable	Suitable	Very Suitable	Very Suitable
7	D_7	Very Suitable	Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
8	D_8	Very Suitable	Suitable	Suitable	Suitable	Acceptable	Very Suitable	Suitable
9	D_9	Very Suitable	Very Suitable	Acceptable	Very Suitable	Acceptable	Suitable	Very Suitable
10	D_{10}	Very Suitable	Very Suitable	Acceptable	Suitable	Very Suitable	Suitable	Very Suitable

https://doi.org/10.31436/iiumej.v26i2.3575

Table 5. Fuzzy number of the criteria assessment linguistic variable

		Criteria						
		C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	D_1	Very Suitable	Very Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
2	D_2	Very Suitable	Very Suitable	Suitable	Suitable	Suitable	Very Suitable	Acceptable
3	D_3	Suitable	Very Suitable	Very Suitable	Very Suitable	Very Suitable	Suitable	Suitable
4	D_4	Very Suitable	Suitable	Suitable	Suitable	Suitable	Acceptable	Very Suitable
5	D_5	Very Suitable	Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
6	D_6	Suitable	Very Suitable	Suitable	Acceptable	Suitable	Very Suitable	Very Suitable
7	D_7	Very Suitable	Suitable	Very Suitable	Very Suitable	Acceptable	Suitable	Suitable
8	D_8	Very Suitable	Suitable	Suitable	Suitable	Acceptable	Very Suitable	Suitable
9	D_9	Very Suitable	Very Suitable	Acceptable	Very Suitable	Acceptable	Suitable	Very Suitable
10	D_{10}	Very Suitable	Very Suitable	Acceptable	Suitable	Very Suitable	Suitable	Very Suitable

3. RESULTS

3.1. Criteria selection

Seven evaluation criteria were established to identify the optimal supplier for PT Wika Beton: reject (c_1), delivery (c_2), price (c_3), communication (c_4), complaint procedure (c_5), service (c_6), and flexibility (c_7). Table 4 presents the assessment criteria decision-makers utilize in their expert evaluations—step two. The experts' responses on the 5-point linguistic scale were converted into a fuzzy number consisting of three specific values: n_1, n_2, and n_3.

Table 6. The threshold

		Criteria						
		C_1	C_2	C_3	C_4	C_5	C_6	C_7
1	D_1	0.04	0.04	0.04	0.04	0.36	0.16	0.16
2	D_2	0.04	0.04	0.16	0.16	0.16	0.04	0.36
3	D_3	0.16	0.04	0.04	0.04	0.04	0.16	0.16
4	D_4	0.04	0.16	0.16	0.16	0.16	0.36	0.04
5	D_5	0.04	0.16	0.04	0.04	0.36	0.16	0.16
6	D_6	0.16	0.04	0.16	0.36	0.16	0.04	0.04
7	D_7	0.04	0.16	0.04	0.04	0.36	0.16	0.16
8	D_8	0.04	0.16	0.16	0.16	0.36	0.04	0.16
9	D_9	0.04	0.04	0.36	0.04	0.36	0.16	0.04
10	D_{10}	0.04	0.04	0.36	0.16	0.04	0.16	0.04
$\sum_{k=1}^{K} d(\tilde{n}_i^k, \tilde{a}_i)$		0.64	0.88	1.52	1.20	2.36	1.44	1.32
	Th_i	0.06	0.09	0.15	0.12	0.24	0.14	0.13

Table 6 presents the results of this step. The criterion c_5 (Thc_5) has a threshold value over 0.2, and the expert agreement is under 70%, leading to its removal. It transforms the computed

average fuzzy rating score for each criterion *i* into its corresponding crisp values [14] utilizing Eq. (4).

Table 7. The expert's agreement value

		Triangular Fuzzy Numbers		Condition of Defuzzification Process		
	Criteria	Threshold Value	Percentage of Experts' Agreement	Fuzzy Number (A)	Ranking	Experts Consensus
1	C_1	0.06	100%	0.819	1	Accepted
2	C_2	0.09	100%	0.805	2	Accepted
3	C_3	0.15	80%	0.708	6	Accepted
4	C_4	0.12	90%	0.722	3	Accepted
5	C_5	0.24	50%	0.674	7	Not Accepted
6	C_6	0.24	90%	0.710	5	Accepted
7	C_7	0.13	90%	0.716	4	Accepted

Table 8. The results of F SWARA II

No	Criteria	Fuzzy number		BnFP	$[t_j]$	$\widetilde{RP}$	$\tilde{P}_{[t_j]}$			$\widetilde{PD}_{[t_j]}$			$\tilde{V}_{[t_j]}$			$\widetilde{W}_j^s$
							n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3	
1	c_4	n_{41}	0.400	0.722	1	ML	3.00	4.00	5.00	0.09	0.16	0.25	1.36	1.58	1.86	0.31
		n_{42}	0.767													0.40
		n_{43}	1.000													0.52
2	c_6	n_{61}	0.400	0.716	2	MH	5.00	6.00	7.00	0.25	0.36	0.49	1.25	1.36	1.49	0.29
		n_{62}	0.748													0.35
		n_{63}	1.000													0.41
3	c_7	n_{71}	0.400	0.710	3	VL	1.00	2.00	3.00	0.01	0.04	0.09	1.00	1.00	1.00	0.23
		n_{72}	0.730													0.25
		n_{73}	1.000													0.28

3.2. Criteria weighting

Three subjective criteria exist: communication (c_4), service (c_6), and flexibility (c_7). Table 8 is obtained using the F-SWARA II steps. The fuzzy weights for the subjective criteria are c_4 = (0.33; 0.40; 0.33), c_6 = (0.33; 0.35; 0.33), and c_7 = (0.33; 0.25; 0.33). Three objective criteria exist: reject (c_1), delivery (c_2), price (c_3). In contrast to subjective data, objective data are clear and precise. The objective data includes minimum, average, and maximum values—the application of objective criteria weighting through MEREC. The initial phase of MEREC involves the creation of a decision matrix. The second and third steps involve normalizing the decision matrix and assessing the overall performance of suppliers. The fourth step consists of calculating by removing each criterion. The fifth step is calculating the outcome effect of removing the j^{th} criterion. The final step involves establishing the criteria's weights**.** The MEREC method, as described by Çelebi Demirarslan et al. [17], is employed to carry out these steps, resulting in Table 9. According to Table 10, the fuzzy weights for the objective criteria are c_1 = (0.290; 0.330; 0.340), c_2 = (0.230; 0.232; 0.349), and c_3 = (0.361; 0.428; 0.440).

https://doi.org/10.31436/iiumej.v26i2.3575

Table 9. The results of MEREC

Supplier		D_{ij}^a			N_{ij}^a			S_{ij}^a	$S_{ij}'^a$			E_j^a		
		C_1	C_2	C_3	C_1	C_2	C_3		C_1	C_2	C_3	C_1	C_2	C_3
1	PT.ISBS	0.07	4	60	0.714	0.750	0.667	0.295	0.208	0.221	0.189	0.087	0.074	0.106
2	PT.MS	0.05	3	50	1.000	1.000	0.800	0.072	0.072	0.072	0.000	0.000	0.000	0.072
3	PT.CBS	0.065	4	60	0.769	0.750	0.667	0.276	0.208	0.201	0.168	0.069	0.076	0.108
4	PT.JCT	0.06	3	40	0.833	1.000	1.000	0.059	0.000	0.059	0.059	0.059	0.000	0.000
	Weight(w_j^a)											0.330	0.230	0.440
Supplier		D_{ij}^b			N_{ij}^b			S_{ij}^b	$S_{ij}'^b$			E_j^b		
		C_1	C_2	C_3	C_1	C_2	C_3		C_1	C_2	C_3	C_1	C_2	C_3
1	PT.ISBS	0.09	4.4	80	0.778	0.909	0.750	0.192	0.120	0.165	0.109	0.072	0.027	0.082
2	PT.MS	0.07	4.2	70	1.000	0.952	0.857	0.065	0.065	0.050	0.016	0.000	0.015	0.049
3	PT.CBS	0.085	5.2	80	0.824	0.769	0.750	0.222	0.168	0.149	0.142	0.053	0.073	0.080
4	PT.JCT	0.08	4	60	0.875	1.000	1.000	0.044	0.000	0.044	0.044	0.044	0.000	0.000
	Weight(w_j^b)											0.340	0.232	0.428
Supplier		D_{ij}^c			N_{ij}^c			S_{ij}^c	$S_{ij}'^c$			E_j^c		
		C_1	C_2	C_3	C_1	C_2	C_3		C_1	C_2	C_3	C_1	C_2	C_3
1	PT.ISBS	0.11	6	100	0.818	0.833	0.800	0.184	0.127	0.132	0.120	0.057	0.052	0.064
2	PT.MS	0.09	5	90	1.000	1.000	0.889	0.039	0.039	0.039	0.000	0.000	0.000	0.039
3	PT.CBS	0.105	6	100	0.857	0.833	0.800	0.171	0.127	0.118	0.106	0.044	0.053	0.065
4	PT.JCT	0.1	6	80	0.900	0.833	1.000	0.092	0.059	0.035	0.092	0.033	0.057	0.000
	Weight(w_j^c)											0.290	0.349	0.361

3.3. Supplier Evaluation

Five decision-makers evaluated the suppliers as experts following the predetermined criteria; the results are summarized in Table 10. The geometric mean of the fuzzy initial decision matrix is employed to construct the initial matrix, as illustrated in Table 11. The subsequent phase entails the development of an extended initial fuzzy matrix (see Table 12). The extension involves the identification of the fuzzy ideal and fuzzy anti-ideal solutions. The third stage consists of normalizing the data to guarantee its awareness. The results of the normalization procedure are summarized in Table 13.

Following the normalization of the initial matrix, the aggregated values are calculated in the fourth step utilizing the weighting coefficients derived from the preceding phase. The fourth step's result is in Table 14. The calculation of the fuzzy utility degree takes place in the fifth step. Identifying the ideal and anti-ideal solutions was crucial for executing this step. Anti-ideal values represent the minimum acceptable level of a specific criterion, while the perfect solution indicates the maximum standard of that criterion. The utility degrees were determined by aggregating the values of each supplier in conjunction with the ideal and anti-ideal solutions. The sixth step of the F-MARCOS method involves defining the utility function of the suppliers. The utility function for the ideal and anti-ideal solutions was computed to determine this function. The final value of the suppliers was determined by incorporating these values, which established their ranking.

Table 10. Linguistic variable and fuzzy parameter of supplier assessment

Expert		PT.ISBS			PT.MS			PT.CBS			PT.JCT		
		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
1	Ex_1	MG	M	EG	MG	MP	EG	M	VP	VG	G	M	VP
2	Ex_2	M	MP	EG	VG	G	G	MP	P	G	MG	VG	MP
3	Ex_3	M	MG	EG	P	MP	MG	MP	MG	EG	M	MG	G
4	Ex_4	M G	P	VP	VP	EG	EG	M	G	P	G	EG	P
5	Ex_5	EG	MP	EG	M	EG	P	VG	MG	M	MP	VG	M

Expert		PT.ISBS			PT.MS			PT.CBS			PT.JCT		
		C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
1	Ex_1	(5,5,7)	(3,5,5)	(7,9,9)	(5,5,7)	(3,3,5)	(7,9,9)	(3,5,5)	(1,1,3)	(7,7,9)	(5,7,7)	(3,5,5)	(1,1,3)
2	Ex_2	(3,5,5)	(3,3,5)	(7,9,9)	(7,7,9)	(5,7,7)	(5,7,7)	(3,3,5)	(1,3,3)	(5,7,7)	(5,5,7)	(7,7,9)	(3,3,5)
3	Ex_3	(3,5,5)	(5,5,7)	(7,9,9)	(1,3,3)	(3,3,5)	(5,5,7)	(3,3,5)	(5,5,7)	(7,9,9)	(3,5,5)	(5,5,7)	(5,7,7)
4	Ex_4	(5,5,7)	(1,3,3)	(1,1,3)	(1,1,3)	(7,9,9)	(7,9,9)	(3,5,5)	(5,7,7)	(1,3,3)	(5,7,7)	(7,9,9)	(1,3,3)
5	Ex_5	(7,9,9)	(3,3,5)	(7,9,9)	(3,5,5)	(7,9,9)	(1,3,3)	(7,7,9)	(5,5,7)	(3,5,5)	(3,3,9)	(7,7,9)	(3,5,5)

Expert		PT.ISBS			PT.MS			PT.CBS			PT.JCT		
		C_4	C_6	C_7	C_4	C_6	C_7	C_4	C_6	C_7	C_4	C_6	C_7
1	Ex_1	M	MG	EG	P	MP	MG	M	M	EG	M	MG	G
2	Ex_2	M G	P	VP	VP	EG	EG	M	G	P	VG	VG	VP
3	Ex_3	EG	MP	EG	M	EG	P	VG	MG	M	MP	VG	M
4	Ex_4	M	MP	EG	EG	MG	MG	MP	P	G	MG	VG	MP
5	Ex_5	G	MP	VG	MG	MP	EG	M	VP	VG	G	M	VP

Expert		PT.ISBS			PT.MS			PT.CBS			PT.JCT		
		C_4	C_6	C_7	C_4	C_6	C_7	C_4	C_6	C_7	C_4	C_6	C_7
1	Ex_1	(3,5,5)	(5,5,7)	(7,9,9)	(1,3,3)	(3,3,5)	(5,5,7)	(3,5,5)	(3,5,5)	(7,9,9)	(3,5,5)	(5,5,7)	(5,7,7)
2	Ex_2	(5,5,7)	(1,3,3)	(1,1,3)	(1,1,3)	(7,9,9)	(7,9,9)	(3,5,5)	(5,7,7)	(1,3,3)	(7,7,9)	(7,7,9)	(1,1,3)
3	Ex_3	(7,9,9)	(3,3,5)	(7,9,9)	(3,5,5)	(7,9,9)	(1,3,3)	(7,7,9)	(5,5,7)	(3,5,5)	(3,3,9)	(7,7,9)	(3,5,5)
4	Ex_4	(3,5,5)	(3,3,5)	(7,9,9)	(7,9,9)	(5,5,7)	(5,5,7)	(3,3,5)	(1,3,3)	(5,7,7)	(5,5,7)	(7,7,9)	(3,3,5)
5	Ex_5	(5,7,7)	(3,3,5)	(7,7,9)	(5,5,7)	(3,3,5)	(7,9,9)	(3,5,5)	(1,1,3)	(7,7,9)	(5,7,7)	(3,5,5)	(1,1,3)

Table 11. Initial matrix of the F-MARCOS

Supplier	C_1			C_2			C_3		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
PT.ISBS	4.36	5.62	6.43	2.67	3.68	4.83	4.74	5.80	7.22
PT.MS	2.54	3.50	4.90	4.66	5.52	6.77	4.15	6.11	6.53
PT.CBS	3.55	4.36	5.62	2.63	3.50	4.99	3.74	5.81	6.11
PT.JCT	4.08	5.16	6.12	5.52	6.43	7.61	2.14	3.16	4.36
Supplier	C_4			C_6			C_7		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
PT.ISBS	4.36	6.02	6.43	2.67	3.32	4.83	4.74	5.52	7.22
PT.MS	2.54	3.68	4.90	4.66	5.16	6.77	4.15	5.71	6.53
PT.CBS	3.55	4.83	5.62	2.37	3.50	4.66	3.74	5.81	6.11
PT.JCT	4.36	5.16	6.43	5.52	6.12	7.61	2.14	2.54	4.36

https://doi.org/10.31436/iiumej.v26i2.3575

Table 12. Extended initial matrix of the F-MARCOS

Supplier	C_1			C_2			C_3		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	2.54	3.50	4.90	2.63	3.50	4.83	2.14	3.16	4.36
PT.ISBS	4.36	5.62	6.43	2.67	3.68	4.83	4.74	5.80	7.22
PT.MS	2.54	3.50	4.90	4.66	5.52	6.77	4.15	6.11	6.53
PT.CBS	3.55	4.36	5.62	2.63	3.50	4.99	3.74	5.81	6.11
PT.JCT	4.08	5.16	6.12	5.52	6.43	7.61	2.14	3.16	4.36
A_{ID}	4.36	5.62	6.43	5.52	6.43	7.61	4.74	6.11	7.22
Supplier	C_4			C_6			C_7		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	2.54	3.68	4.90	2.37	3.32	4.66	2.14	2.54	4.36
PT.ISBS	4.36	6.02	6.43	2.67	3.32	4.83	4.74	5.52	7.22
PT.MS	2.54	3.68	4.90	4.66	5.16	6.77	4.15	5.71	6.53
PT.CBS	3.55	4.83	5.62	2.37	3.50	4.66	3.74	5.81	6.11
PT.JCT	4.36	5.16	6.43	5.52	6.12	7.61	2.14	2.54	4.36
A_{ID}	4.36	6.02	6.43	5.52	6.12	7.61	4.74	5.81	7.22

Table 13. Normalized extended initial matrix of the F-MARCOS

Supplier	C_1			C_2			C_3		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	0.39	0.45	0.58	0.35	0.41	0.48	0.30	0.35	0.45
PT.ISBS	0.39	0.45	0.58	0.54	0.71	0.98	0.30	0.37	0.45
PT.MS	0.52	0.72	1.00	0.39	0.48	0.56	0.33	0.35	0.52
PT.CBS	0.45	0.58	0.71	0.53	0.75	1.00	0.35	0.37	0.57
PT.JCT	0.41	0.49	0.62	0.35	0.41	0.48	0.49	0.68	1.00
A_{ID}	0.52	0.72	1.00	0.54	0.75	1.00	0.49	0.68	1.00
Supplier	C_4			C_6			C_7		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	0.39	0.57	0.76	0.31	0.44	0.61	0.30	0.35	0.60
PT.ISBS	0.68	0.93	1.00	0.35	0.44	0.63	0.66	0.76	1.00
PT.MS	0.39	0.57	0.76	0.61	0.68	0.89	0.57	0.79	0.90
PT.CBS	0.55	0.75	0.87	0.31	0.46	0.61	0.52	0.80	0.85
PT.JCT	0.68	0.80	1.00	0.73	0.80	1.00	0.30	0.35	0.60
A_{ID}	0.68	0.93	1.00	0.73	0.80	1.00	0.66	0.80	1.00

Table 14. Weighted by the normalized extended initial matrix

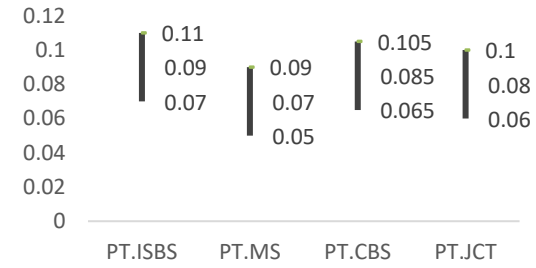
Supplier	C_1			C_2			C_3		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	0.39	0.45	0.58	0.35	0.41	0.48	0.30	0.35	0.45
PT.ISBS	0.39	0.45	0.58	0.54	0.71	0.98	0.30	0.37	0.45
PT.MS	0.52	0.72	1.00	0.39	0.48	0.56	0.33	0.35	0.52
PT.CBS	0.45	0.58	0.71	0.53	0.75	1.00	0.35	0.37	0.57
PT.JCT	0.41	0.49	0.62	0.35	0.41	0.48	0.49	0.68	1.00
A_{ID}	0.52	0.72	1.00	0.54	0.75	1.00	0.49	0.68	1.00
Supplier	C_4			C_6			C_7		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
A_{AI}	0.39	0.57	0.76	0.31	0.44	0.61	0.30	0.35	0.60
PT.ISBS	0.68	0.93	1.00	0.35	0.44	0.63	0.66	0.76	1.00
PT.MS	0.39	0.57	0.76	0.61	0.68	0.89	0.57	0.79	0.90
PT.CBS	0.55	0.75	0.87	0.31	0.46	0.61	0.52	0.80	0.85
PT.JCT	0.68	0.80	1.00	0.73	0.80	1.00	0.30	0.35	0.60
A_{ID}	0.68	0.93	1.00	0.73	0.80	1.00	0.66	0.80	1.00

Table 15. Final value of suppliers

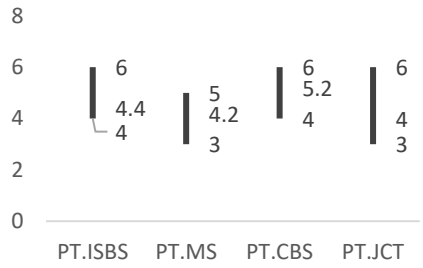
Supplier	K_i^-			K_i^+			$f(K_i^-)$		
	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
PT.ISBS	0.589	1.381	3.085	0.347	0.764	1.753	0.235	0.551	1.230
PT.MS	0.574	1.349	3.050	0.338	0.746	1.733	0.229	0.538	1.216
PT.CBS	0.553	1.376	3.062	0.326	0.761	1.740	0.220	0.549	1.221
PT.JCT	0.629	1.432	3.288	0.371	0.792	1.868	0.251	0.571	1.311

Supplier	$f(K_i^+)$			$K_i^-{}_{crisp}$	$K_i^+{}_{crisp}$	$f(K_i^-)_c$	$f(K_i^+)_c$	$f(K_i\ )_{cris}$	Rank
	n_1	n_2	n_3						
PT.ISBS	0.138	0.305	0.699	1.533	0.859	0.611	0.343	0.81941	2
PT.MS	0.135	0.297	0.691	1.503	0.843	0.599	0.336	0.7878	4
PT.CBS	0.130	0.303	0.693	1.520	0.852	0.606	0.340	0.80528	3
PT.JCT	0.148	0.316	0.745	1.608	0.901	0.641	0.359	0.90114	1

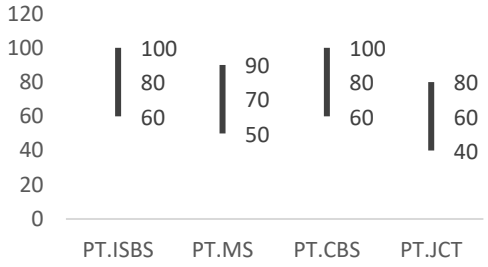
https://doi.org/10.31436/iiumej.v26i2.3575



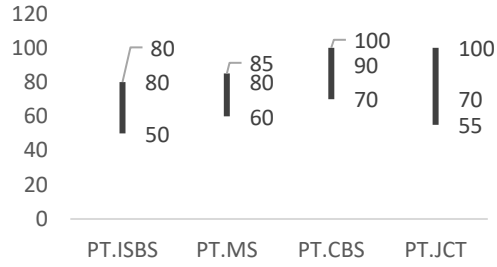
(a) Reject



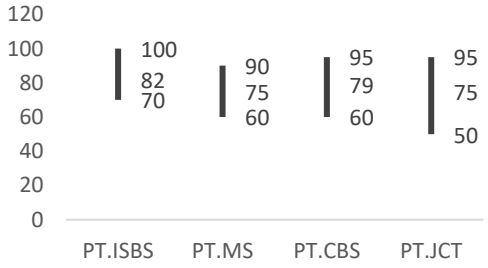
(b) Delivery



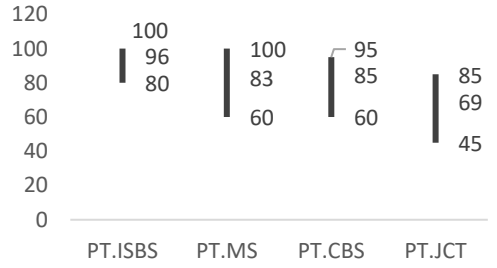
(c) Price



(d) Communication



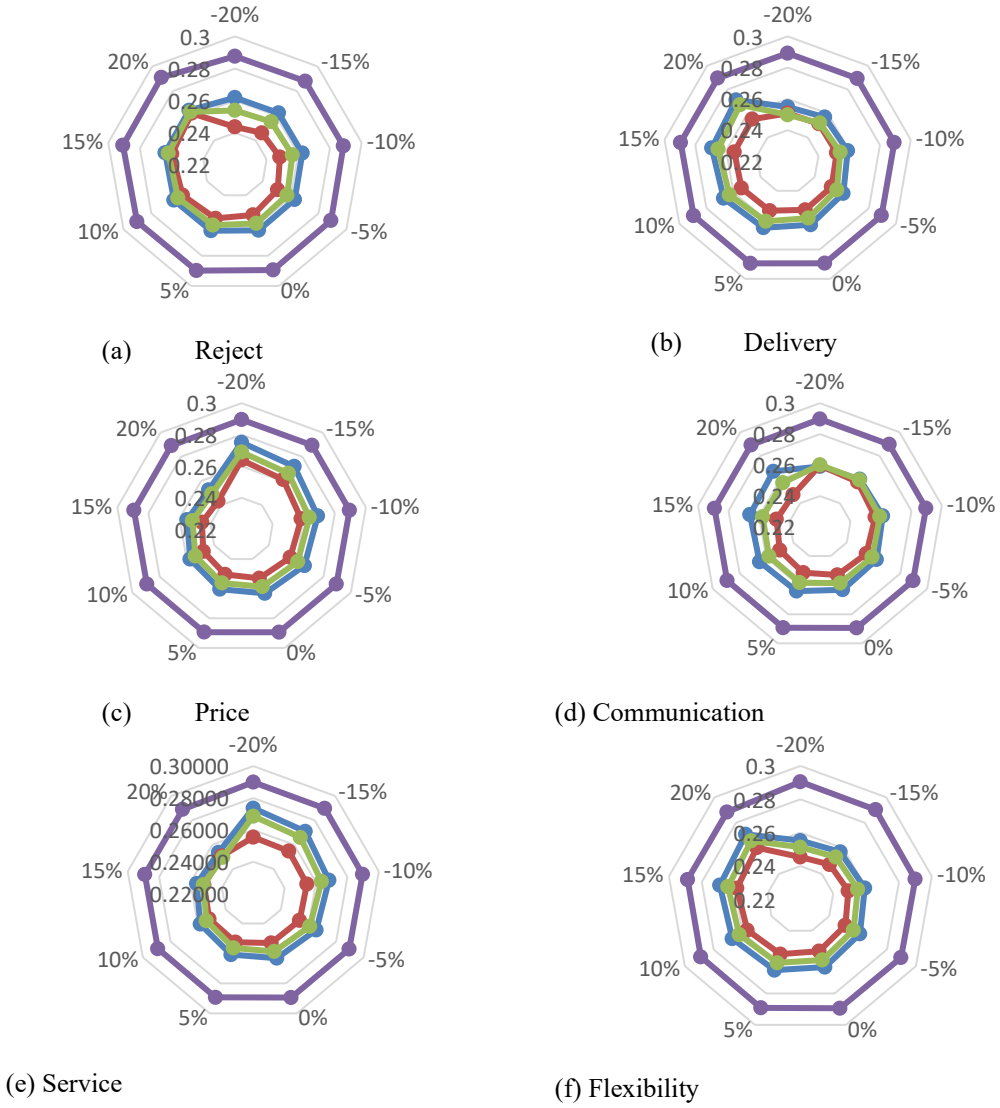
(e) Service



(f) Flexibility

Figure 3. Performance of the suppliers.

https://doi.org/10.31436/iiumej.v26i2.3575



(a) Reject (b) Delivery

(c) Price (d) Communication

(e) Service (f) Flexibility

Figure 4. Sensitivity analysis.

3.2. Discussion

The ranking of supplier evaluation for PT. Wika Beton is PT. JCT, PT. ISBS, PT. CBS, and PT. MS. The weight of the criteria and supplier performance determines this evaluation. The crisp values after normalization for the weights of reject, delivery, price, communication, service, and flexibility are 0.159, 0.134, 0.204, 0.204, 0.174, and 0.126, respectively. If associated with supplier performance, as seen in Figures 3(c) and 3(d), PT. JCT has the lowest price compared to others; likewise, for the performance of the communication sector, PT. JCT has a relatively higher performance value than others. At the same time, the weight of the price and communication criteria is the largest, making PT. JCT is the best supplier candidate. The weight of the reject and delivery criteria is only fourth and fifth, so their contribution is not the most significant. However, based on Figures 3(a) and 3(b), PT JCT's reject and delivery performance is excellent. In general, PT.JCT performs well in all criteria except flexibility, with the lowest weight of the flexibility criteria. Even the performance of the price criteria is the best, with the highest weight of the price criteria. So, the proposed method is the best one for providing the best alternative for PT. Wika Beton is determining suppliers to increase its production.

Moreover, managers should emphasize considerations beyond cost, which have been overshadowed due to the product's critical role in the organization's survival and elevated demand. The production process should function efficiently by emphasizing the reject and delivery criteria. Therefore, besides the lowest price, the supplier exhibiting the quickest delivery time and lowest rejection should be prioritized and regarded as the optimal choice, regardless of their performance in other domains.

A sensitivity analysis is needed to validate the proposed method, with changes for each weight range from -5% to 20%. Based on Figure 4, the proposed method is not sensitive to changes in the criteria weight because the supplier ranking does not change. So, the proposed method is valid. Based on Figures 4.b and 4.d, changes in the delivery and communication criteria weight will start to cause changes in the second and third supplier rankings if they are below 20%. Meanwhile, changes in the weight of the reject and service criteria will start to cause changes in the second and third supplier rankings if they exceed 20% (see Figures 4(a) and 4(e)). Based on Figures 4.c and 4.f, regardless of the change in the weight of the price and flexibility criteria, it does not affect the supplier ranking.

Table 16. Supplier rank for each hybrid MARCOS method

Supplier	MARCOS (Stević et al. [2], Paradowski and Szyjewski [10])		F SWARA MARCOS (Tadić et al. [21])		F-MARCOS (Ecer and Pamucar [22])		F-SWARA F-MARCOS (Tus and Aytac-Adali [8])		Proposed method	
	Utility function	Rank	Utility function	Rank	Utility function	Rank	Utility function	Rank	Utility function	Rank
PT.ISBS	0.7504	2	0.7524	2	0.8070	1	0.9654	3	0.8194	2
PT.MS	0.7435	3	0.7491	3	0.7782	3	0.9488	2	0.7878	4
PT.CBS	0.7869	1	0.7904	1	0.8057	2	0.9774	1	0.8053	3
PT.JCT	0.7202	4	0.7205	4	0.7756	4	0.9294	4	0.9011	1

Table 17. Borda count for each performance's selected supplier

Criteria		MARCOS (Stević et al. [2], Paradowski and Szyjewski [10]) - PT.CBS	F SWARA MARCOS (Tadić et al. [21]) - PT.CBS	F-MARCOS (Ecer and Pamucar [22]) - PT.ISBS	F-SWARA F-MARCOS (Tus and Aytac-Adali [8]) - PT.CBS	Proposed method - PT.JCT
	Performance	0.065	0.065	0.07	0.065	0.06
C_1	Rank	2	2	3	2	1
	Borda	3	3	1	3	5
	Performance	4	4	4	4	3
C_2	Rank	2	2	2	2	1
	Borda	2.5	2.5	2.5	2.5	5
	Performance	60	60	60	60	40
C_3	Rank	2	2	2	2	1
	Borda	2.5	2.5	2.5	2.5	5
	Performance	100	100	80	100	100
C_4	Rank	1	1	2	1	1
	Borda	3.5	3.5	1	3.5	3.5
	Performance	60	60	70	60	50
C_6	Rank	2	2	1	2	3
	Borda	3	3	5	3	1
	Performance	60	60	80	60	45
C_7	Rank	2	2	1	2	3
	Borda	3	3	5	3	1
	Sum	17.5	17.5	17	17.5	20.5

Table 16 compares the supplier's utility function and the rank of the methods. It shows that the supplier rank for each technique is different. Meanwhile, supplier selection aims to select the supplier with the best performance in each criterion. Often, no single supplier performs best across all of these criteria. So, MCDM is needed to solve this problem. The proposed method is compared with the previous methods to determine its effectiveness. This test uses the Borda count on the performance of selected suppliers for each method. The Borda count generates the overall ranking of suppliers by combining inconsistent results from different evaluation models [23]. The Borda rule suits multi-person decision-making when exploring multiple options [23]. In other words, the Borda technique assigns ranks to alternatives based on the assumption that the higher an option's position on the list, the higher its rating. The winner is the option with the highest calculated score, where each alternative is awarded a score beginning with 1 for the least favorable answer, 2 for the second worst, 3 for the third worst, and so on. The Borda score for each option is calculated by weighting all scores [23]Figure 2 shows each supplier's performance for each criterion. Table 17 shows the test results. Based on Table 18, the supplier of the proposed method has the highest Borda value. Thus, the proposed method is more effective than the previous method because it can select the supplier with the best performance.

4. CONCLUSION

The optimal raw material supplier for the Indonesian concrete industry was determined by material quality, competitive pricing, prompt delivery, communication, service, and flexibility. PT JCT emerged as the optimal choice, exhibiting a utility value of 0.9011. PT Wika Beton places significant emphasis on the quality of raw materials and prompt delivery. The company strategically selects raw materials at competitive prices, aligning with its cost-cutting initiatives in the supply chain to ensure customer satisfaction and maintain high service quality.

The decision to select PT JCT as the optimal supplier is justified and sound. Cost is traditionally viewed as a critical factor, often leading to immediate rejection, while other aspects are evaluated with varying degrees of importance. In Indonesia, concrete products play a crucial role in infrastructure development, and managers are responsible for ensuring the operational efficiency of their production. The sustainability of the Indonesian concrete industry is contingent upon the selection of suppliers. Selecting an appropriate supplier enhances a concrete manufacturing company's downstream operation by decreasing production costs and improving customer satisfaction.

The company in the case study experienced a 5% reduction in production costs in the later stages of the project, following several months of implementation, indicating the effectiveness of the proposed methodology. The management of the case companies is fully committed to executing the supplier ranking established in this study and allocating orders according to that ranking. This commitment demonstrates their confidence in the proposed methodology's effectiveness.

The future research: (1) consider applications in other industries; (2) the use of Deep Reinforcement Learning (DRL) in optimal supplier selection in dynamic environments; (3) the use of Machine Learning models, such as Random Forest, or Neural Networks to predict supplier performance based on historical data; (4) the use of AI to automatically determine criteria weights; (5) the use of Fuzzy Logic accompanied by AI to handle uncertainty in decision-maker preferences.

ACKNOWLEDGEMENT

The author would like to thank the management of PT Wika Beton for their technical support.

REFERENCES

[1] Sahoo, S. K. Goswami, S. S., Halder, R. (2024). Supplier Selection in the Age of Industry 4.0: A Review on MCDM Applications and Trends. Decision Making Advances, 2(1): 32–47, doi: 10.31181/dma21202420.

[2] Stević, Ž., Pamučar, D., Puška, A., Chatterjee, P. (2020). Sustainable supplier selection in healthcare industries using a new MCDM method: Measurement of alternatives and ranking according to Compromise solution (MARCOS). Computers and Industrial Engineering, 140: 106231, doi: 10.1016/j.cie.2019.106231.

[3] Trung, D. D. (2024). Assessing the Impact of Criterion Weights on the Ranking of the Top Ten Universities in Vietnam. Engineering, Technology and Applied Science Research, 14(4): 14899–14903, doi: 10.48084/etasr.7607.

[4] Worku, V (2025). Formwork material selection and optimization by a comprehensive integrated subjective–objective criteria weighting MCDM model. Discover Materials, 5(2): 2025, doi: 10.1007/s43939-024-00162-x.

[5] Badi, I., Pamucar, D. (2020). Supplier selection for steelmaking company by using combined grey-marcos methods. Decision Making: Applications in Management and Engineering, 3(2): 37–47, doi: 10.31181/dmame2003037b.

[6] Chattopadhyay, R., Chakraborty, S., Chakraborty, S. (2020). An Integrated D-Marcos Method for Supplier. Decision Making: Applications in Management and Engineering, 3(2): 49–69.

[7] Taş, M. A., Çakır, E., Ulukan, Z. (2021). Spherical fuzzy SWARA-MARCOS approach for green supplier selection. 3C Tecnología_Glosas de innovación aplicadas a la pyme, Edición Es(special issue 7): 115–133, doi: 10.17993/3ctecno.2021.specialissue7.115-133.

[8] Tus, A., Aytac Adali, E. (2022). Green Supplier Selection Based on the Combination of Fuzzy SWARA (SWARA-F) and Fuzzy MARCOS (MARCOS-F) Methods. Gazi University Journal of Science, 35(4): 1535–1554, doi: 10.35378/gujs.978997.

[9] Wang, Y., Wang, W., Wang, Z., Deveci, M., Roy, S. K., Kadry, S. (2024). Selection of sustainable food suppliers using the Pythagorean fuzzy CRITIC-MARCOS method. Information Sciences, 664: 120326, doi: 10.1016/j.ins.2024.120326.

[10] Paradowski, B., Szyjewski, Z. (2022). Comparative analyses of multi-criteria methods in supplier selection problem. Procedia Computer Science, 207: 4593–4602, doi: 10.1016/j.procs.2022.09.523.

[11] Abdulla, A., Baryannis, G., Badi, I. (2023). An integrated machine learning and MARCOS method for supplier evaluation and selection. Decision Analytics Journal, 9: 100342, doi: 10.1016/j.dajour.2023.100342.

[12] Naghipour, M. S. , Rahim, Z. A., Iqbal, M. S. (2024). A 5G competency model based on the fuzzy Delphi method. Journal of Infrastructure, Policy and Development, 8(10): 6788, doi: 10.24294/jipd.v8i10.6788.

[13] Ahmed, O. S., Al-Gahtani, K. S., Altuwaim, A. (2025). Cost–Benefit Framework for Selecting a Highway Project Using the SWARA Approach. Buildings, 15(3): 439, doi: 10.3390/buildings15030439.

[14] Aris, N., Dayana, N., Abd, B., Hanani, N., Nidzam, M. (2025). Determining Design Thinking Elements in Chemistry Classroom Teaching Strategies : A Fuzzy Delphi Method. Eclética Química, 50: e–1566, doi: 10.26850/1678-4618.eq.v50.2025.e1566.

[15] Erdiani, N., Hashim, H., Sulaiman, N. A. (2024). Application of Fuzzy Delphi Method to Identify the Construct for Designing and Developing the Multimodal Learning Framework for Writing Skills in ESL Context. International Journal of Communication Networks and Information Security, 16(S1): 1085-1097.

[16] Keshavarz-Ghorabaee, M. (2021). Assessment of distribution center locations using a multi-expert subjective–objective decision-making approach. Scientific Reports, 11(1): 1–20, doi:

10.1038/s41598-021-98698-y.
[17] Çelebi Demirarslan, P., Sönmez Çakır, F., Akansel, I. (2024). Ranking the quality of life indexes by years in Asian countries using multi-criteria decision-making methods. Asia-Pacific Journal of Regional Science, 8(3): 911-942, doi: 10.1007/s41685-024-00350-w.
[18] Abdi, A. P.., Damci, A., Kirca, O., Turkoglu, H., Arditi, D., Demirkesen, S., Korkmaz, M., Arslan, A. E. (2024). A Spatial Decision-Support System for Wind Farm Site Selection in Djibouti. Sustainability, 16(22): 9635, doi: 10.3390/su16229635.
[19] Grzywacz, N. M. (2025). Perceptual Complexity as Normalized Shannon Entropy. Entropy, 27(2): 166.
[20] Stanković, M., Stevi´, Ž., Das, D. K., Suboti´, m., Pamuˇcar, D. (2020). A New Fuzzy MARCOS Method for Road Tra ffi c Risk Analysis. Mathematics, 8: 457.
[21] Tadić, S., Kilibarda, M., Kovač, M., Zečević, S. (2021). the Assessment of Intermodal Transport in Countries of the Danube Region. International Journal for Traffic and Transport Engineering, 11(3): 375–391, doi: 10.7708/ijtte2021.11(3).03.
[22] Ecer, F., Pamucar, D. (2021). MARCOS technique under an intuitionistic fuzzy environment for determining the COVID-19 pandemic performance of insurance companies in terms of healthcare services. Applied Soft Computing, 104: 107199, doi: 10.1016/j.asoc.2021.107199.
[23] Oufella, S. (2024). Hybrid use of Borda count and PROMETHEE method for maintenance strategy selection problem. Foundations of Computing and Decision Sciences, 49(2): 139–160, doi: 10.2478/fcds-2024-0009.

https://doi.org/10.31436/iiumej.v26i2.3577

VOLTAGE-INDUCED VOID FORMATION IN HIGH-TEMPERATURE OXIDE SCALES OF BOILER TUBES

MUHAMMAD RAFIQ HAIKAL ROSDIN[1], SYED NOH SYED ABU BAKAR[2], ABD MALEK ABDUL HAMID[1], AHMAD ABDUL MUN'IM ISMAIL[1], MOHD HANAFI ANI[1*]

[1]Department of Manufacturing and Materials Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia
[2]Department of Mechanical Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia

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

(Received: 15 January 2025; Accepted: 8 April 2025; Published online: 15 May 2025)

ABSTRACT: Corrosion monitoring remains a significant challenge at high temperatures. Understanding the varying factors in high-temperature cathodic protection is crucial for developing mitigation strategies and predictive maintenance. This study assesses how cathodic protection influences oxidation in T91 alloys at elevated temperatures by evaluating the effects of exposure duration and voltage-induced void development in the oxide layer. It is hypothesized that polarizing the sample affects the diffusivity of cations and anions in the oxide scale, which is the rate-determining step of the oxidation process. This study measured the number of voids directly on T91 alloys exposed at 823K under various induced voltages. T91 alloy was externally induced with voltages of 0V, 50V, and 300V for 43.2 ks, 259.2 ks, and 432 ks at 923 K in air (P_{O_2}= 0.21 atm = 2.1×10^4 Pa). The presence of oxide layers was analysed using X-Ray Diffraction (XRD), and the void formed was inspected using Scanning Electron Microscopy (SEM). XRD results reveal that Fe_2O_3, Fe_3O_4, $FeCr_2O_3$, and Cr_2O_3 peaks were formed on all samples. The parabolic rate constant, K_p, was calculated as 3.83 × 10^{-14} m^2/s, 2.17 × 10^{-14} m^2/s, and 9.25 × 10^{-14} m^2/s, respectively, verifying that the reaction occurred by solid-state diffusion. Changes in K_p at different induced voltages are clear evidence that the diffusivity was altered by external electrical potential. It was observed that the overall void formation decreased by 17%. Inducing voltage onto T91 alloy affects the ionic diffusivity. It changes the void formation, suggesting it may promote the diffusivity of more inert species, such as Cr, to form a protective layer at the early oxidation stage.

ABSTRAK: Pemantauan hakisan kekal sebagai cabaran utama pada suhu tinggi. Memahami pelbagai faktor perlindungan katodik pada suhu tinggi adalah penting untuk membangunkan strategi pengurangan dan ramalan penyelenggaraan. Kajian ini menilai perlindungan katodik mempengaruhi pengoksidaan dalam aloi T91 pada suhu tinggi dengan melihat kesan tempoh pendedahan dan pembangunan rongga yang disebabkan oleh potensi elektrik luaran pada lapisan oksida. Pemolaran sampel mempengaruhi keberaliran kation dan anion dalam oksida, yang menentukan kadar dalam proses pengoksidaan. Melalui kajian ini, jumlah ruang kosong diukur secara langsung pada aloi T91 yang didedahkan pada suhu 823K di bawah pelbagai voltan teraruh. Aloi T91 dikenakan voltan luaran sebanyak 0V, 50V, dan 300V bagi tempoh 43.2 ks, 259.2 ks, dan 432 ks pada suhu 923K dalam udara (P_{O_2} = 0.21 atm = 2.1 × 10^4 Pa). Kehadiran lapisan oksida dianalisa menggunakan Pembelauan Sinar-X (XRD), dan ruang kosong yang terbentuk diperiksa menggunakan Mikroskop Elektron Imbasan (SEM). Dapatan XRD menunjukkan bahawa puncak Fe_2O_3, Fe_3O_4, $FeCr_2O_3$, dan Cr_2O_3 terbentuk pada semua sampel. Pemalar kadar parabola, K_p, dikira masing-masing sebanyak 3.83 × 10^{-14} m²/s, 2.17 × 10^{-14} m²/s, dan 9.25 × 10^{-14} m²/s, mengesahkan bahawa tindak balas yang berlaku adalah

penyebaran keadaan pepejal. Perubahan dalam K_p pada voltan teraruh berbeza membuktikan bahawa keberaliran telah diubah oleh potensi elektrik luaran. Hasil kajian mendapati bahawa pembentukan ruang kosong secara keseluruhan berkurangan sebanyak 17%. Proses penguraian voltan pada aloi T91 mempengaruhi keberaliran ionik dan mengubah pembentukan ruang kosong, mencadangkan bahawa ia mungkin digunakan bagi mempromosi keberaliran spesies yang lebih lengai seperti Cr bagi membentuk lapisan pelindung pada peringkat awal pengoksidaan.

KEYWORDS: *high temperature oxidation, T91 Alloy, void formation, induced voltage*

1. INTRODUCTION

The most economical method of producing hydrogen is through steam methane reforming (SMR), which accounts for 95% of global hydrogen production with an efficiency of 85% [1]. SMR involves reacting pure methane with superheated steam in the presence of a nickel catalyst to produce syngas. The steam is generated at 923K using a heat recovery steam generator (HRSG). Seamless ferritic alloy tubes, specifically T91 alloy (ASME SA-213 grade T91), are the common boiler tubes used in the HRSG system. These tubes are exposed to high temperatures, leading to oxidation in the presence of oxygen, which causes fouling and spalling of oxide scales. Although these tubes are designed to last 100,000 service hours, studies by Li et al. [2] show degradation starting as early as 4,000 hours, raising safety and operational concerns.

Research done by Maruyama et al. [3] had linked void formation in oxide scales during high-temperature oxidation to divergence of ionic flux. In a related study, Maruyama et al. [4] also examined void formation in growing oxide scales with Schottky defects and p-type conduction, finding that voids preferentially form near the metal/scale interface, their volume increases parabolically, and their volume fraction remains time-independent. Using the Nernst-Planck equation, Schlögl and Helffreich [5] theorized changes in ionic diffusivity under electrical potential. Further, Jin et al. [6] had analysed the influence of applied electric fields on carrier distribution during high-temperature corrosion, concluding that positive electric fields promote cation vacancy migration toward the metal interface, thereby altering the growth rate of corrosion film. However, the impact of electric fields on high-temperature corrosion remains underexplored. This study investigates how applied voltages (0V, 50V, and 300V) at 923K in normal air affect void formation in T91 alloy, aiming to understand the relationship between induced electrical potential and high-temperature corrosion.

2. EXPERIMENTAL METHODS

2.1. Sample Preparations

T91 alloy was cut into smaller pieces before being further cut into an average dimension of 10.9 mm × 6.4 mm × 4.2 mm using a precision saw machine. Then, it was ground using a grinding machine with various abrasive paper grades (P600 – P2000). A polishing machine was used to polish the ground sample using up to 0.3 micron-sized diamond particles.

2.2. Voltage-Induced Experiment

Figure 1 shows the schematic diagram for the experimental setup. A platinum wire of 1 metre length was used as the working and counter electrode. Platinum was chosen due to its high resistivity towards corrosion, high electrical conductivity, and low chemical reactions at high temperatures. Then, the T91 alloy sample was spot welded onto the working electrode.

https://doi.org/10.31436/iiumej.v26i2.3577

This was done to ensure the electrodes remained firmly in place during the experiment. An R-Type thermocouple was placed in the isothermal zone, 5mm from the sample, to measure precise temperature in the zone. Then, heat was supplied by the furnace.

The platinum electrodes were then connected to the Cockroft-Walton circuit, also known as a voltage multiplier. The amount of voltage supplied to the sample was 0V, 50V, and 300V. The experiment was conducted at 1 atm (2.1×10^4 Pa of oxygen) with standard dry air composition. For 20 minutes, the temperature was progressively raised to 923K to ensure stability inside the heated space. It was then left alone for another 15 minutes to check that the provided voltage is stable. Table 1 summarizes the parameters used for the experiment.

Table 1. Parameters used for the experiment.

Time (ks)	Voltage Supplied (V)
43.2	0
	50
	300
259.2	0
	50
	300
432	0
	50
	300

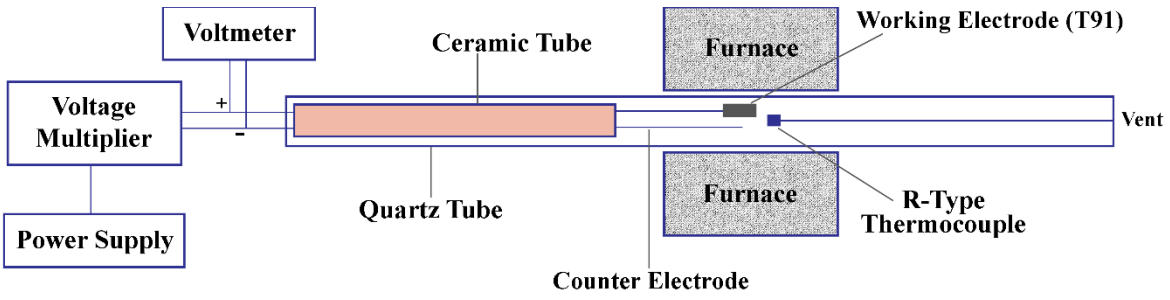


Figure 1. Schematic diagram for the Voltage Induced Experiment.

The selection of applied voltages was based on their expected impact on oxidation behaviour. 0V was a control condition, representing natural oxidation without external electric influence. 50V was chosen as a moderate voltage level, simulating conditions where minor electric fields could affect ionic mobility and void formation. 300V was selected to represent high-field conditions and evaluate whether the oxidation kinetics would still follow the parabolic rate constant (K_p) behaviour, which is typically observed in high-temperature oxidation. Deviations from the parabolic trend would indicate the presence of alternative mechanisms influencing oxidation under high-voltage conditions.

The exposure times of 43.2 ks (12 hours), 259.2 ks (72 hours), and 432 ks (120 hours) were selected to capture different oxidation stages. 43.2 ks represents the early-stage oxidation behaviour and initial void formation. 259.2 ks provides insight into progressive oxide scale growth and void evaluation. 432 ks enables observation of long-term oxidation effects, including steady-state oxide growth and potential saturation of void formation. However, common high-temperature corrosion studies often involve exposure periods longer than suggested durations, extending beyond 432 ks.

https://doi.org/10.31436/iiumej.v26i2.3577

2.3. Phase Characterization

Oxide phases were analysed using XRD with divergence slit of 2.5 degree. The diffraction angle was set from 20° to 80°. Cross-sectional images generated by SEM were used to determine the oxide thickness, void formation, and parabolic rate constant.

The methodology of calculating volume fraction of voids was adopted from Kaderi et al. [7], wherein uniformly sized grids were superimposed onto the SEM images of oxide scale cross-sections. The area fraction of voids was calculated by assessing the proportion of void areas relative to the total oxide scale area within these grids, which was a standard approach in such analyses. To ensure measurement accuracy, multiple images were analysed for each sample to account for variability and enhance statistical stability. This method provides a reasonable estimation of void volume fractions.

3. RESULTS AND DISCUSSIONS

3.1. XRD Analysis

Figure 2 shows XRD patterns of all sample surfaces at respective conditions. The XRD analysis of the raw sample shows intensity peaks at diffraction angles of 44.8° and 65°, which confirms the presence of α-Fe in the BCC structure [8]. Furthermore, there are also noticeable peaks of Fe_2O_3 (hematite), Fe_3O_4 (magnetite), $FeCr_2O_3$ (spinel), and Cr_2O_3 (chromium), confirming that major phases have been identified.

Meanwhile, samples subjected to 300V for 43.2 ks and all samples exposed for 259.2 ks and 432 ks began to show a less intense peak of magnetite due to their variations in crystal structure and atom arrangement. Nasrazadani and Raman [9] stated that magnetite is a mixed valence iron oxide containing both Fe^{2+} and Fe^{3+} ions in an inverse spinel structure, while hematite has a corundum structure with only Fe^{3+} ions. Fe^{2+} ions in magnetite cause more disorder in the crystal structure, causing magnetite to have weaker diffraction peaks than hematite. The dense formation of hematite can also affect the XRD in detecting the magnetite phase [10].

The peaks of chromia started to increase from the sample exposed at 50V for 43.2 ks until 300V for 432 ks. However, the peak was less intense than magnetite and hematite. The presence of multiple phases can lead to a reduction in diffraction peaks. Other than that, the chromium content of T91 alloy was quite low at 9%, which can also lead to reduced intensity.

The XRD results demonstrate that increasing applied voltage influences the phase composition of the oxide scale, particularly by altering the formation and stability of magnetite and hematite. The observed reduction in magnetite intensity at higher voltage suggests potential changes in oxidation kinetics and defect structures, which will be discussed further in this paper. The gradual increase in chromium formation with voltage indicates how electric fields may promote selective oxidation, further affecting the alloy's resistance to corrosion.

https://doi.org/10.31436/iiumej.v26i2.3577

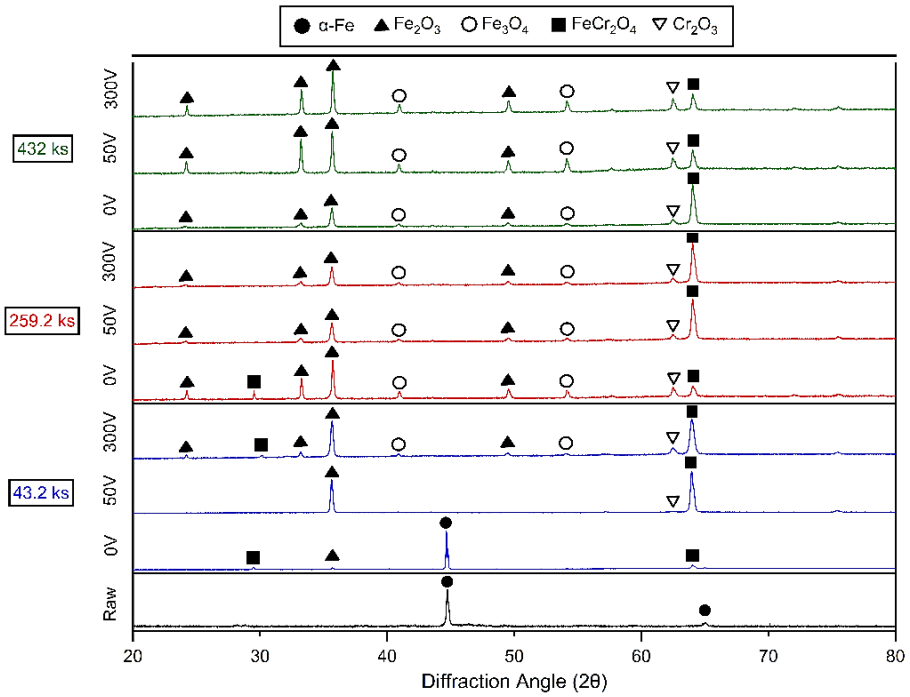


Figure 2. XRD patterns of all samples exposed at 650°C after 43.2, 259.2, and 432 ks of exposure with the proposed voltage supplied. The figure also includes the XRD pattern of the raw sample.

3.2. SEM Analysis

The morphology of the sample at the proposed timeframe was summarized in Figure 3. A red dashed line indicates the interphase boundary between the alloy, inner scale, and outer scale. For ease of discussion, the inner scale will be referred to as magnetite, while the outer scale is hematite, and the term will be used interchangeably in the discussion. Based on Figure 3, all samples show the presence of oxide layers at various supplied external voltages. It was also observed that the voids' shape was irregular, and some aligned in a lath-like pattern at a specific oxide thickness. A clear example of voids and lath-like patterns can be seen in Figure 4.

Voids in the inner and outer oxide scale are formed due to the ionic flux. A study by Ueda et. al. [11] stated that the partial pressure of oxygen can affect the void formation. The higher the oxygen partial pressure, the higher the oxygen chemical potential; hence, more voids are formed. Even though the oxygen partial pressure of this experiment is high (2.1×10^4 Pa), there is a presence of voids in the outer scale, and only a few appear in the inner scale. In addition, Taniguchi [12] proposed that the oxide expanded under a compressive force if the volume of oxide formed by the metal exceeded unity. If the ratio, however, were less than one, the resulting oxide would expand under tensile stress. As a result, a porous oxide coating develops. The oxide layer was either island-like or layer-by-layer when subjected to high temperatures [13].

https://doi.org/10.31436/iiumej.v26i2.3577

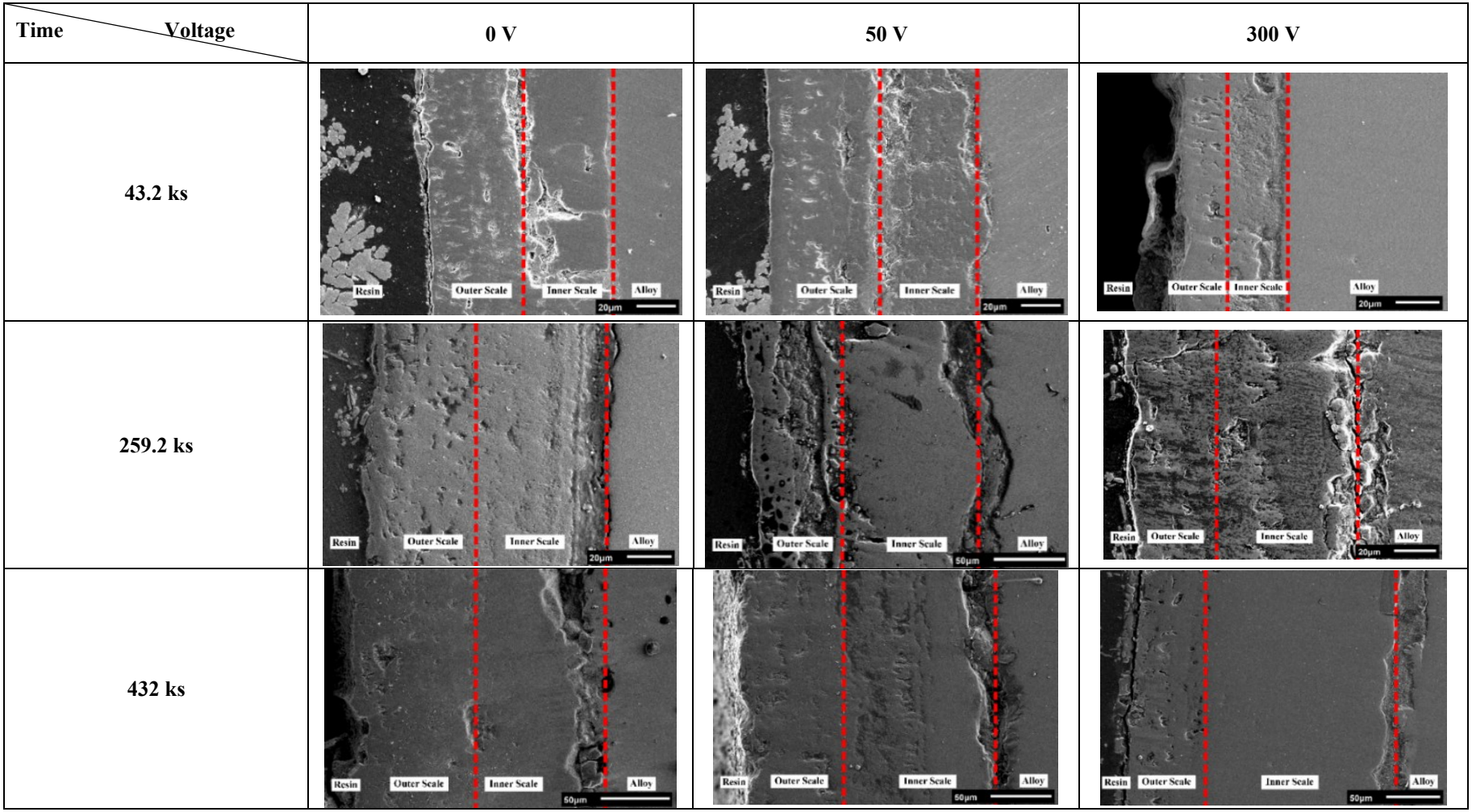


Figure 3. Cross sectional micrograph of T91 alloy through various timeframe and voltage.
Red dashed line indicates the interphase boundary.

https://doi.org/10.31436/iiumej.v26i2.3577

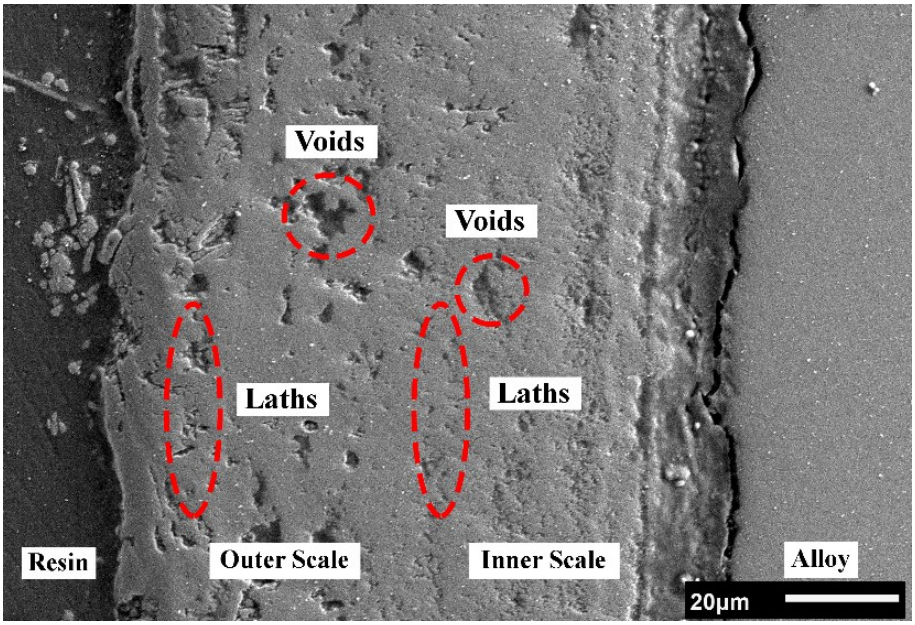


Figure 4. Formation of voids and laths from one of the cross sections exposed to the proposed condition. The image was taken from a sample exposed for 259.2 ks at 0V.

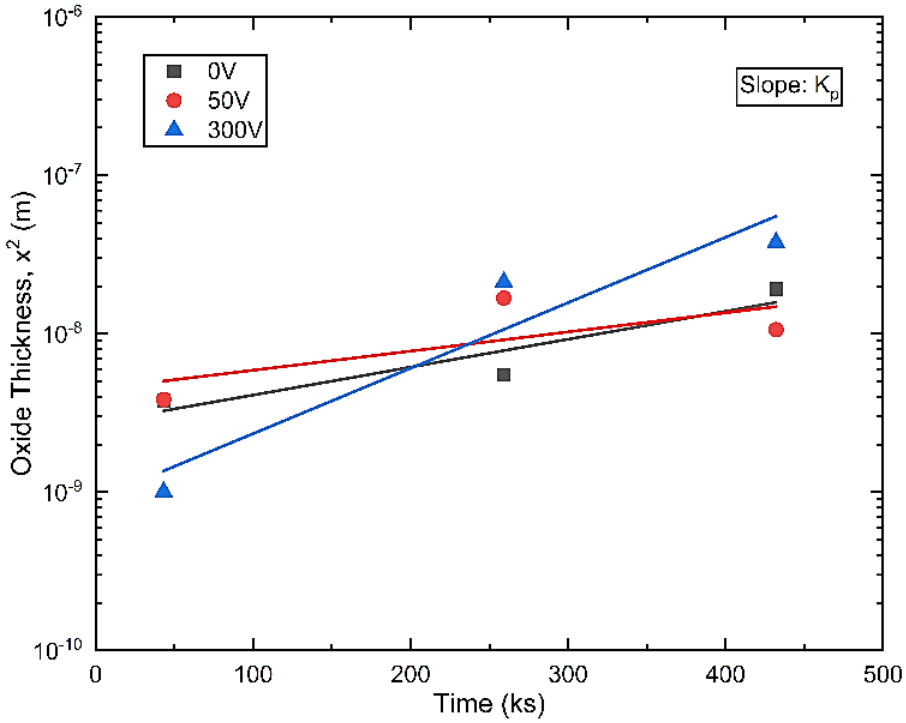


Figure 5. Overall oxide thickness vs Time.

Figure 5 shows the parabolic plots of the overall oxide thickness from the perspective of the weight gain of the sample. It shows a sharp increase at the beginning phase and then slowly gains weight up to 432 ks. The weight gain for the sample exposed to 0V does not follow the parabolic pattern, especially at 259.2 ks of exposure. This was due to oxide spallation, which causes it to become fragile during the final weight measurement. Overall, it indicates that almost all oxide scales formed obey the parabolic rate law. From the cross-section of the sample, the thickness of oxide scales can be determined to calculate the parabolic rate constant,

https://doi.org/10.31436/iiumej.v26i2.3577

K_p. The K_p values for 0V, 50V and 300V were 3.83×10^{-14} m^2/s, 2.17×10^{-14} m^2/s and 9.25×10^{-14} m^2/s respectively.

3.3. Volume Fraction of Voids in Oxide Scale

Figure 6 shows the graphical representation of the volume fraction of voids for the sample's overall phase. The image illustrates that the volume percentage of voids and the governing equations are in good agreement. It also shows that as exposure time increases, void formation decreases. Overall, the formation of voids has decreased by 17%.

Similarly, as the voltage supplied increased, void formation on the oxide layers also decreased. Generally, oxygen ions from the air will diffuse into the sample when exposed to a high-temperature environment. Iron and chromium ions within the sample will also diffuse to the surface. The difference in flux between iron, chromium, and oxygen ions triggers a faster reconfiguration for some of them, causing slower-moving ions to have difficulty moving and redistributing themselves. Thus, it will lead to the accumulation of voids. However, if external driving forces are present, for this case, externally supplied voltage, the flux increased as well. Hence, the diffusion process will speed up, causing void formations to be reduced.

For T91 alloy, chromium reacts with oxygen and will form a passive layer called chromia, which prevents further oxidation [14]. Chromia has a much slower diffusion rate than other oxide layers. Thus, some of the iron ions managed to react with oxygen to form an oxide layer [15]. By inducing voltage onto the sample, the flux of chromia can be accelerated, allowing it to reach the surface quicker and participate in oxide layer formation, enhancing both the passive layer and resistance against void formation and corrosion. Figure 7 shows the overall void percentage of T91 alloy decreases linearly with time.

A plot of overall void percentage vs voltage is shown in Figure 8. The correlation indicates that void percentage decreases exponentially as voltage increases. Conversely, Figure 8 also illustrates that increasing the voltage causes the void percentage to plateau, eventually reaching a limit — indicating that beyond a certain point, applying higher voltage does not significantly affect the void percentage.

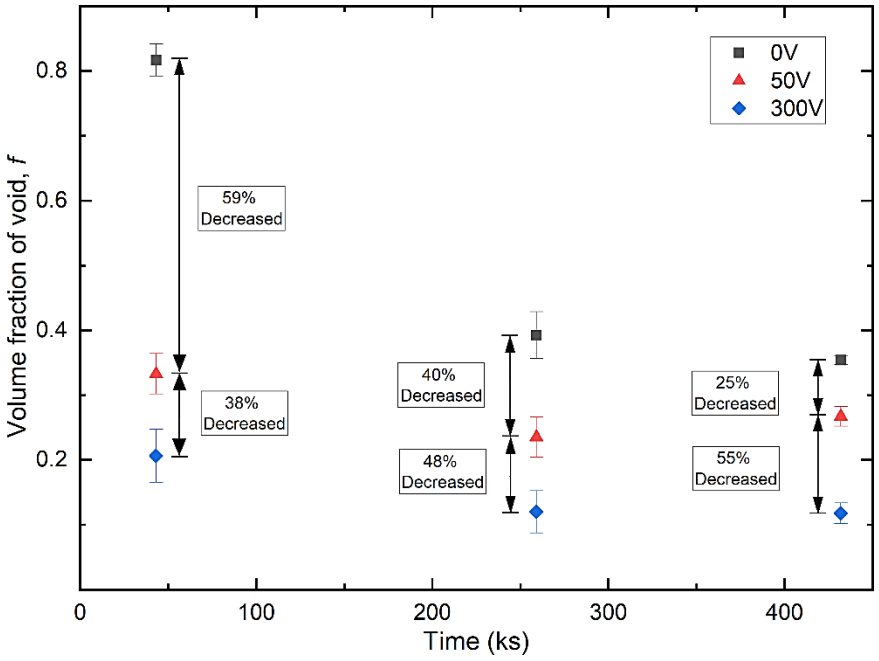


Figure 6. Volume fraction of void for overall phase of the sample.

https://doi.org/10.31436/iiumej.v26i2.3577

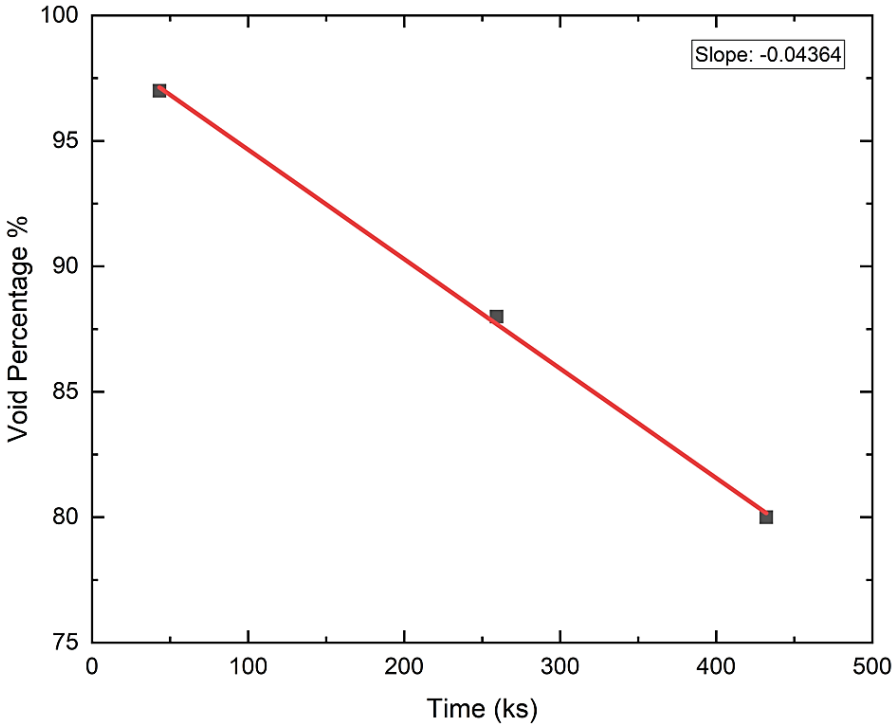


Figure 7. Overall void percentage vs Time.

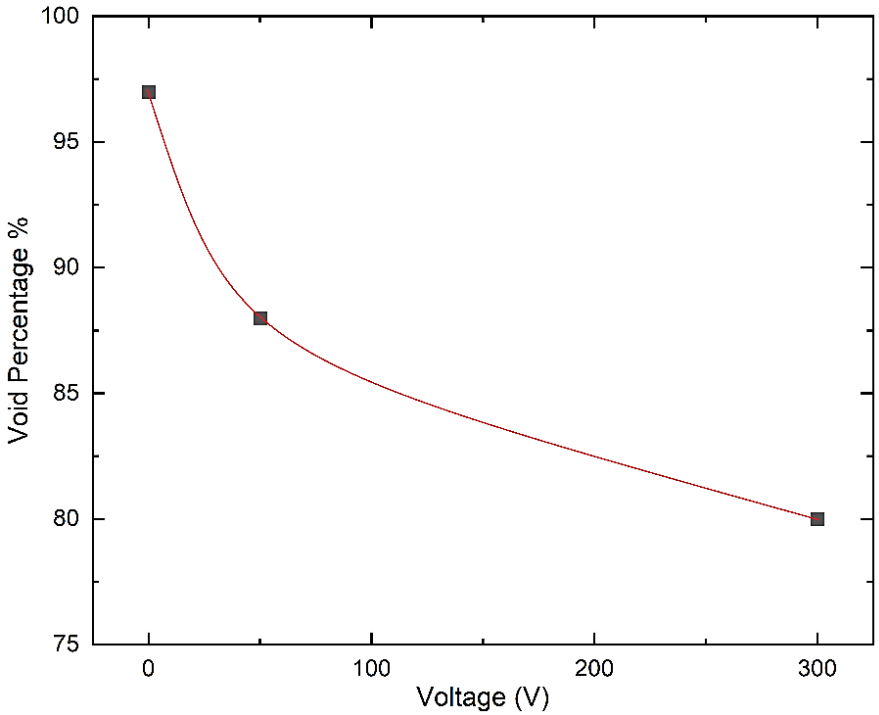


Figure 8. Overall void percentage vs Voltage.

Based on the graphs obtained in Figures 6, 7, and 8, a more detailed analysis of the void percentage in hematite and magnetite layers can be conducted, as illustrated in Figures 9 and 10. These figures clearly show a distinction: a higher volume of voids is formed in hematite compared to magnetite. This phenomenon can be explained by several factors related to these two oxide layers' crystal structure and properties.

https://doi.org/10.31436/iiumej.v26i2.3577

Magnetite has a spinel structure and consists of both Fe^{2+} and Fe^{3+} ions. This structure allows more oxygen vacancies and defects within the oxide lattice [16]. Hence, magnetite tends to exhibit a higher probability of void formation because it can facilitate the movement of iron ions. In contrast, hematite has an alpha-corundum crystal structure and contains only Fe^{3+} ions. Hematite is more compact and less prone to the formation of defects and voids [17]. This is not the case when an external voltage has been supplied. The volume fraction of voids for both layers appeared to decrease as a higher voltage was supplied. Plus, the exponential pattern on the volume fraction of void in magnetite phase appears steeper than that of hematite.

At exposed temperature, magnetite demonstrates an intriguing interplay between its magnetic and electrical properties, being a ferrimagnetic material that inherently exhibits both. Due to its mixed-valence compound, it creates a charged imbalance in the crystal lattice, resulting in the presence of electrical conductivity. As noted by Radoń et al. [18], rising temperatures significantly enhance the electrical conductivity of magnetite due to increased thermal energy, which facilitates a higher concentration of charge carriers. When combined with an applied external voltage, the resulting induced current accelerates diffusion within the magnetite layer, thereby reducing void formation.

On the contrary, hematite behaves as paramagnetic at temperatures above the Neel temperature (TN ≈ 955K), where the magnetic moments become disordered, resulting in no long-range magnetic ordering [19]. Therefore, hematite is considered an insulator or weakly conducting material as well as having a wide bandgap, which restricts the flow of electrical current [20]. Due to the mentioned hematite properties, an increase in voltage supplied shows a less significant steep slope, as shown in Figures 9 and 10. Moreover, when the voltage was set to 300V, their volume fraction of void shows an almost linear pattern (Figure 10). Overall, the effect of inducing different amounts of voltage onto T91 alloy shows that as the voltage supplied increases, it speeds up the diffusion rate of magnetite, causing less formation of voids, and vice versa for hematite.

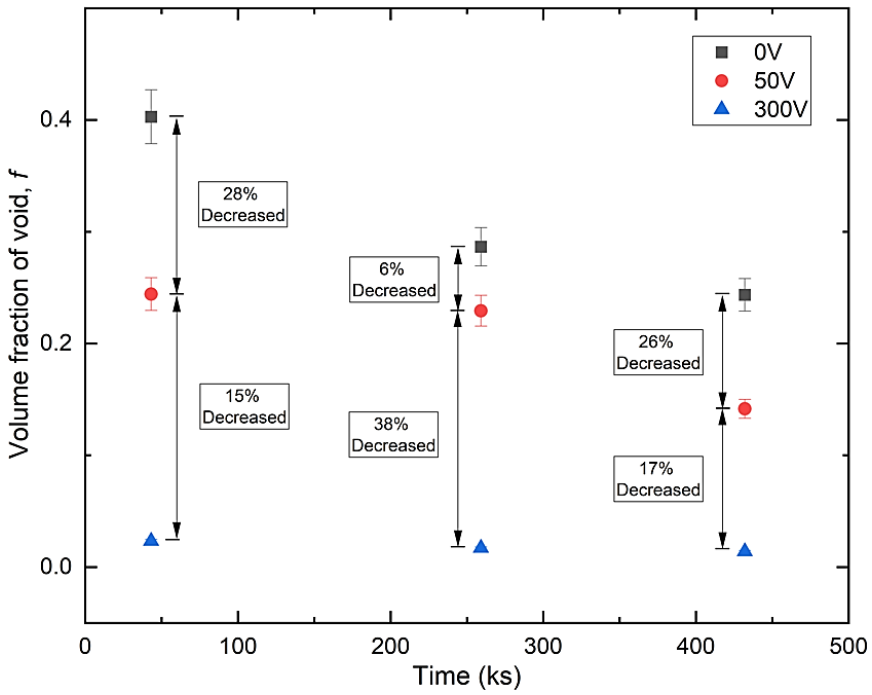


Figure 9. Void Percentage for all samples according to the hematite layer.

https://doi.org/10.31436/iiumej.v26i2.3577

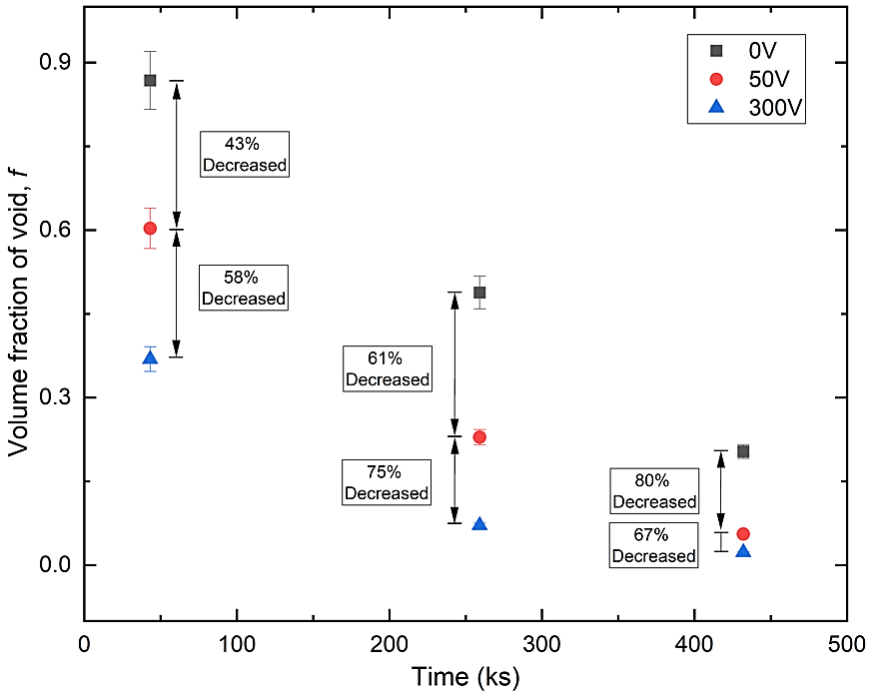


Figure 10. Void Percentage for all samples according to the magnetite layer.

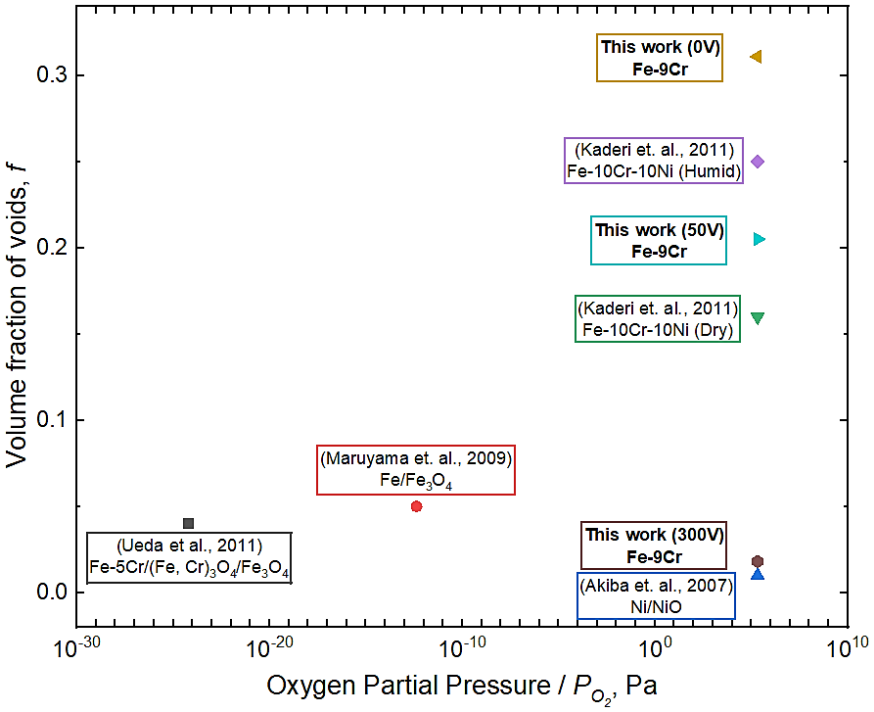


Figure 11. Comparison of volume fraction of voids on the outermost scale with other works.

Figure 11 shows a summary of work done on the volume fraction of voids on the outermost scale of the sample with respect to oxygen partial pressure, P_{O_2}. A study on void formation at the magnetite scale conducted by Maruyama et al. [3] and Ueda et al. [18] showed a very small range of voids formed in the range of 0.05 to 0.06. Note that these two studies were conducted in a controlled environment. In contrast, a study on void formation of NiO scale by Akiba et al. [21] and void on Fe-Cr-Ni alloy by Kaderi et al. [7] was done in normal air and is comparable

with this work. Note that the volume fraction of void on the outer scale is much smaller when voltage is supplied.

4. CONCLUSION

T91 alloy has been exposed to a high-temperature corrosion environment at 923 K. The samples were subjected to impressed voltages of 0V, 50V, and 300V at 43.2 ks, 259.2 ks, and 432 ks. XRD analysis confirmed the formation of oxide layers in all conditions, with phase identification indicating the presence of hematite, magnetite, and chromium. SEM cross-sectional analysis revealed the development of voids within oxide scales, with void formation quantified at both the overall scale and within individual phases.

Oxidation kinetics, assessed through oxide thickness measurements, followed the parabolic law, verifying that solid-state diffusion governed the reaction. Overall K_p value for samples exposed at 0V, 50V, 100V and 300V were 3.83×10^{-14} m^2/s, 2.17×10^{-14} m^2/s and 9.25×10^{-14} m^2/s respectively. This suggested that 50V suppressed oxidation rates while higher voltages (300V) accelerated diffusion kinetics, leading to increased oxidation rates. The enhancement of oxidation at 300V can be attributed to increased mobility of oxygen and metal cations under the influence of the external electric field, which facilitates faster oxide scale growth. SEM images also revealed that voids formed in all samples.

The void percentage for the overall, hematite, and magnetite layers has been calculated. The result shows that overall void formation decreased by 17% as exposure time increases, and a smaller number of voids was observed on magnetite phases compared to hematite. This is due to magnetite's higher electrical conductivity, which enhances diffusion and reduces void formation under applied voltage, and vice versa for hematite. At 300V, void formation in hematite stabilized while magnetite showed a steeper decline. These findings confirm that applied voltage influences oxidation behaviour by accelerating diffusion in magnetite, reducing void formation, while having a lesser effect on hematite.

ACKNOWLEDGEMENT

The author would like to acknowledge that no external funding sources were utilized in the conduct of his research. This work was carried out independently without financial support from organizations or institutions.

REFERENCES

[1] S. G. Subraveti, S. Roussanaly, R. Anantharaman, L. Riboldi, and A. Rajendran, "Techno-economic assessment of optimised vacuum swing adsorption for post-combustion CO_2 capture from steam-methane reformer flue gas," *Sep Purif Technol*, vol. 256, Feb. 2021, doi: 10.1016/j.seppur.2020.117832.

[2] Y. Li, J. Du, L. Li, K. Gao, X. Pang, and A. A. Volinsky, "Mechanical properties and phases evolution in T91 steel during long-term high-temperature exposure," *Eng. Failure Analysis*, vol. 111, Apr. 2020, doi: 10.1016/j.engfailanal.2020.104451.

[3] T. Maruyama, M. Ueda, and K. Kawamura, "Void formation in the growing scale induced by the divergence of the diffusive ionic flux in high temperature oxidation of metals," *Defect Diffus. Forum*, Trans Tech Publications Ltd, 2009, pp. 1–13. doi: 10.4028/www.scientific.net/DDF.289-292.1.

[4] T. Maruyama, K. Akiba, M. Ueda, and K. Kawamura, "Void formation in growing oxide scales with Schottky defects and p-type conduction," *Mater. Sci. Forum*, vol. 595–598, pp. 1039–1046, 2008, doi: 10.4028/www.scientific.net/MSF.595-598.1039.

[5] R. Schlögl and F. Helfferich, “Comment on the significance of diffusion potentials in ion exchange kinetics,” *J Chem. Phys.*, vol. 26, no. 1, pp. 5–7, 1957, doi: 10.1063/1.1743264.

[6] Y. Jin, Y. Zhang, and Y. Liu, “Corrosion reaction kinetics and high-temperature corrosion testing of contact element strips in ultra-high voltage bushing based on the phase-field method,” *IET Gener. Transm. Distrib.*, vol. 16, no. 2, pp. 249–260, 2022, doi: 10.1049/gtd2.12487.

[7] A. Kaderi, A. Zaki, M. Zainal, H. Ani, and R. Othman, “Observation on Void Formed in Oxide Scale of Fe-Cr-Ni Alloy at 1073K in Dry and Humid Environments,” *IIUM Eng. J.*, vol. 12, no. 5, pp. 69–78, 2011, doi: 10.31436/iiumej.v12i5.235.

[8] M. H. B. Ani, T. Kodama, M. Ueda, K. Kawamura, and T. Maruyama, "The effect of water vapor on high temperature oxidation of Fe-Cr alloys at 1073 K," *Mater. Trans.*, vol. 50, no. 11, pp. 2656–2663, 2009, doi: 10.2320/matertrans.M2009212.

[9] S. Nasrazadani and A. Raman, "The application of infrared spectroscopy to the study of rust systems—II. Study of cation deficiency in magnetite (Fe_3O_4) produced during its transformation to maghemite (γ-Fe_2O_3) and hematite (α-Fe_2O_3)," *Corrosion Sci.*, vol. 34, no. 8, pp. 1355–1365, 1993, doi: 10.1016/0010-938X(93)90092-U.

[10] T. Simmonds and P. C. Hayes, “Isothermal Oxidation of Magnetite to Hematite in Air and Cyclic Reduction/Oxidation Under Carbon Looping Combustion Conditions,” *Metall. Mater. Trans. E*, vol. 4, no. 2–4, pp. 114–122, Dec. 2017, doi: 10.1007/s40553-017-0111-7.

[11] M. Ueda, K. Kawamura, and T. Maruyama, "Void formation in magnetite scale formed on iron at 823 K—Elucidation by chemical potential distribution," *Mater. Sci. Forum*, vol. 523, pp. 37–44, 2006, doi: 10.4028/www.scientific.net/MSF.522-523.37.

[12] S. Taniguchi, “Stresses developed during the oxidation of metals and alloys,” *Trans. Iron Steel Inst. Jpn.*, vol. 25, pp. 3–13, 1985, doi: 10.2355/isijinternational1966.25.3.

[13] A. Fluri, D. Pergolesi, V. Roddatis, A. Wokaun, and T. Lippert, "In situ stress observation in oxide films and how tensile stress influences oxygen ion conduction," *Nat. Commun.*, vol. 7, no. 1, p. 10692, 2016, doi: 10.1038/ncomms10692.

[14] R. D. Armstrong, “Electrochemical Dissolution,” *Encyclopedia of Materials: Science and Technology*. Elsevier, pp. 2521–2525, 2001. doi: 10.1016/B0-08-043152-6/00456-3.

[15] T. Jonsson (2007) Microscopy of High Temperature Oxidation of Iron and Some Stainless Steels PhD thesis. Chalmers University of Technology. Retrieved from https://research.chalmers.se/en/publication/40412.

[16] R. Arras, B. Warot-Fonrose, and L. Calmels, “Electronic structure near cationic defects in magnetite,” *J. Phys. Condens. Matter*, vol. 25, no. 25, Jun. 2013, doi: 10.1088/0953-8984/25/25/256002.

[17] K. R. Tolod, S. Hernández, E. A. Quadrelli, and N. Russo, “Visible light-driven catalysts for water oxidation: Towards solar fuel biorefineries,” in *Studies in Surface Science and Catalysis*, vol. 178, S. Albonetti, S. Perathoner, and E. A. Quadrelli, Eds., Elsevier, 2019, pp. 65–84, doi: 10.1016/B978-0-444-64127-4.00004-5.

[18] A. Radoń et al., “Influence of magnetite nanoparticles shape and spontaneous surface oxidation on the electron transport mechanism,” *Materials*, vol. 14, no. 18, Sep. 2021, doi: 10.3390/ma14185241.

[19] D. Varshney and A. Yogi, “Structural and Electrical conductivity of Mn doped Hematite (α-Fe_2O_3) phase,” *J. Mol. Struct*, vol. 995, no. 1–3, pp. 157–162, May 2011, doi: 10.1016/j.molstruc.2011.04.011.

[20] J. Engel and H. L. Tuller, “The electrical conductivity of thin film donor doped hematite: From insulator to semiconductor by defect modulation,” *Phys. Chem. Chem. Phys.*, vol. 16, no. 23, pp. 11374–11380, Jun. 2014, doi: 10.1039/c4cp01144a.

[21] K. Akiba, M. Ueda, K. Kawamura, and T. Maruyama, “Quantitative prediction of voids formation in a growing nickel oxide scale at 1371 K,” *Mater. Trans.*, vol. 48, pp. 2753–2761, Oct. 2007, doi: 10.2320/matertrans.MER2007122.

https://doi.org/10.31436/iiumej.v26i2.3413

REGENERATIVE BRAKING SYSTEM (RBS) MOSFET SWITCHING-BASED DRIVE CYCLE FOR AN ELECTRIC MOTORCYCLE

NURUL MUTHMAINNAH MOHD NOOR[1], SITI FAUZIAH TOHA[1*], MUHAMMAD ABDULLAH[2], SALMIAH AHMAD[3], MD ATAUR RAHMAN[4]

[1]*Department of Mechatronics, International Islamic University Malaysia, Kuala Lumpur, Malaysia*
[2]*Department of Mechanical and Aerospace Engineering, Kulliyyah of Engineering, International Islamic University, Gombak 53100, Malaysia*
[3]*College of Engineering and Technology, University of Doha for Science and Technology, Doha, 24449, Qatar*
[4]*GameAbove College of Engineering and Technology, Sill Hall, Eastern Michigan University, Ypsilanti, Michigan, USA 48197*

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

(Received: 21 August 2024; Accepted: 20 March 2025; Published online: 15 May 2025)

ABSTRACT: A regenerative braking system is an advanced technology applicable to transportation, particularly electric vehicles. The purpose of incorporating regenerative braking is to recover energy during braking and deceleration, which can be stored in the battery. This paper aims to study the operation of the regenerative braking system based on an urban drive cycle. This study selects the US60 and NEDC drive cycles as inputs to evaluate future powertrain systems and vehicle concepts. The output torque is calculated longitudinally based on the vehicle dynamic equation to determine whether the torque is negative or positive. When the torque is negative, regenerative braking applies, and the state of charge (SoC) of the battery increases. The concept of regenerative braking is that this system uses four MOSFETs as switches. As a result, at the 50% level of SoC, the first regeneration improved performance by 12.22%, whereas the second showed a smaller gain of 5.96%. Similarly, at the 80% level of SoC, the first regeneration yielded a 12.55% increase, while the second achieved only a 6.19% improvement. The rise in SoC for both levels demonstrates that energy can be recovered when implementing regenerative braking. Therefore, the results obtained from the MATLAB simulation will be used for future studies in implementing a regenerative braking control strategy.

ABSTRAK: Sistem brek jana semula adalah teknologi canggih yang digunakan untuk pengangkutan, terutamanya kenderaan elektrik. Tujuan menggabungkan brek jana semula adalah bagi memulihkan tenaga semasa brek dan nyahpecutan, yang boleh disimpan dalam bateri. Kajian ini bertujuan bagi mengkaji operasi sistem brek jana semula berdasarkan kitaran pacuan bandar. Dalam kajian ini, kitaran pemacu US60 dan NEDC dipilih sebagai input bagi menilai sistem *powertrain* dan konsep kenderaan masa hadapan. Tork keluaran dikira berdasarkan persamaan dinamik membujur kenderaan bagi menentukan tork negatif atau positif. Apabila tork negatif, brek jana semula terpakai, dan keadaan cas (SoC) bateri meningkat. Konsep brek sistem jana semula ini menggunakan empat MOSFET sebagai suis. Hasilnya, pada tahap 50% SoC, penjanaan semula pertama meningkatkan prestasi sebanyak 12.22%, manakala tahap kedua menunjukkan kenaikan lebih kecil iaitu 5.96%. Begitu juga, pada tahap 80% SoC, penjanaan semula pertama menghasilkan peningkatan 12.55%, manakala yang kedua hanya mencapai peningkatan 6.19%. Peningkatan SoC bagi kedua-dua tahap menunjukkan bahawa tenaga boleh dipulihkan bagi melaksanakan brek jana semula.

https://doi.org/10.31436/iiumej.v26i2.3413

Oleh itu, dapatan kajian yang diperoleh dari simulasi MATLAB akan digunakan untuk kajian masa hadapan dalam melaksanakan strategi kawalan brek jana semula.

KEYWORDS: *Regenerative, Braking, Electric Motorcycle, Drive Cycle, and Recovery Energy.*

1. INTRODUCTION

In this era, technological advancements are paralleled by energy consumption, leading to an energy shortage crisis. This forces researchers to find solutions in new energy generation. In Malaysia, fossil fuels contribute to about 90% of energy production, especially in Peninsular Malaysia, which includes greenhouse gas (GHG) emissions [1]. To address this issue, the Malaysian government has introduced the National Energy Transition Roadmap (NETR) to achieve net-zero carbon emissions by 2050 with renewable energy shares targeting an impressive 40% by 2040 [2]. One significant effort to reduce emissions is the adoption of electric vehicles.

Moreover, Malaysia also aims to reduce GHG emissions through the increased use of renewable energy (RE), as outlined in the 11th Malaysia Plan (2016–2020). This plan continues for the 12th Malaysia (2021–2025), and the main agenda still focuses on low-carbon development, resource efficiency, and the preservation and sustainability of natural resources. However, a major challenge is that industries are the most significant contributors to carbon emissions and environmental impacts. Nevertheless, the government is also committed to the Paris Agreement of 2015 and the decarbonization agenda; the energy transition from gasoline cars to electric vehicles is becoming more imminent [3].

Selected Targets	2018	Low Carbon Nation Aspiration 2040
1. Percentage of urban public transport modal share	20%	50%
2. Percentage of electric vehicle (EV) share	<1%	38%
3. Alternative fuel standard for heavy transport	B5	B30
4. Percentage of Liquefied Natural Gas (LNG) as alternative fuel for marine transport	0%	25%
5. Percentage of industrial and commercial energy efficiency savings	<1%	11%
6. Percentage of residential energy efficiency savings	<1%	10%
7. Total installed capacity of RE	7,597 MW	18,431 MW
8. Percentage of coal in installed capacity	31.4%	18.6%
9. Percentage of RE in TPES	7.2%	17%

Legend: Energy security, Energy affordability, Environmental sustainability

Figure 1. Selected targets on Low Carbon Nation Aspiration 2040 compared to 2018 [3].

Through the National Energy Policy 2021– 2024 (NEP 2040), the Minister of Environment and Water (KASA) introduced several strategies to take advantage of the energy sector for

https://doi.org/10.31436/iiumej.v26i2.3413

socio-economic development, including the Low Carbon Nation Aspiration 2024 [4]. As illustrated in Figure 1, the percentage of electric car use will increase from less than 1% to 38% of the total industry volume (TIV) by 2040. This demonstrates the government's commitment to reducing carbon emissions by building 10,000 charging facilities, comprising 9,000 alternating current (AC) units and 1,000 direct current (DC) units, by 2050.

In Malaysia, the use of electric motorcycles has been increasing since the government announced a rebate of up to RM2,400 in Budget 2024. This incentive scheme encourages electric motorcycle adoption among individuals earning less than RM120,000 annually or less than RM10,000 monthly [5]. However, some modern electric motorcycles still have certain flaws, such as high purchase costs, 'range anxiety' among riders, and a lack of charging stations. The main issue behind these problems is the battery. To overcome this issue, regenerative braking technology can recover and store energy in the energy storage pack.

This paper is divided into five sections. The first section discusses Malaysia's energy policy. Section 2 covers related literature reviews on regenerative braking systems. Section 3 describes the methodology and concept of electric motorcycles, introducing the mathematical equation for the longitudinal direction. Section 4 presents the findings and discussion. Finally, Section 5 provides the conclusion, summarizing the findings and offering recommendations for future work.

2. LITERATURE REVIEW

Electric motorcycles fall under the category of electric vehicles, which are particularly useful in urban areas for riders to move efficiently from one place to another, especially during congested city driving conditions. Electric motorcycles' size is sleeker than other electric vehicles, allowing users to navigate traffic congestion more easily. An electric motor drives battery-powered electric motorcycles. Additionally, the electric drive and control system are key components connected to the motor drive, power supply, and speed control of the motor [6]. Unlike traditional motorcycles that propel themselves using internal combustion engines (ICEs), electric motorcycles are distinguished by their electric devices, control systems, and mechanical systems, such as power drives.

The power supply from the battery sent to the drive motor converts electrical energy into mechanical energy through a transmission system or by directly driving the wheel. Lead-acid batteries are the most widely used power source in electric vehicles [7]. However, due to their lower specific energy, slower charging speed, and shorter lifespan, they are gradually being replaced by other types of batteries, such as lithium-ion, as electric vehicle technology advances [8]. Meanwhile, the driving device generates torque, which are translated and converted into ground forces through the wheel, propelling the electric motorcycle forward.

Brakes are another critical vehicle component, responsible for slowing down and stopping the vehicle. In electric motorcycles, most braking systems are electromagnetic, which can control the driving motor circuit to generate power. During deceleration and braking, the energy produced can be converted into current to charge the battery, making it reusable [9]. This concept is known as regenerative braking.

2.1. Regenerative Braking System (RBS)

Many motorcyclists use their motorcycles for long journeys, leading to concerns about battery capacity and frequent recharging to maintain the battery's state of charge (SoC) until they reach their destination [10]. However, modifying an electric motorcycle to increase battery capacity is challenging, costly, and cumbersome. Finding a solution to this problem is crucial,

https://doi.org/10.31436/iiumej.v26i2.3413

as electric motorcycles have the potential to revolutionize urban commuting by reducing carbon emissions and traffic congestion and providing an efficient alternative to traditional internal combustion engine-powered vehicles [11].

Regenerative braking is a technology designed to extend the travel range of electric vehicles. When the vehicle slows down, kinetic energy is converted into electrical energy, preventing energy wastage [12]. This process utilizes the motorcycle's motion to generate electricity stored in the battery. It assists in slowing down the motorcycle while recharging the battery, enhancing overall energy efficiency. Depending on the vehicle's requirements, the stored energy can be utilized immediately or later [13]. Figure 2 illustrates the energy flows between traditional and electric vehicles. In traditional vehicles, kinetic energy is converted into heat and wasted. However, with regenerative braking in electric vehicles, kinetic energy is captured and stored in a storage device for future use.

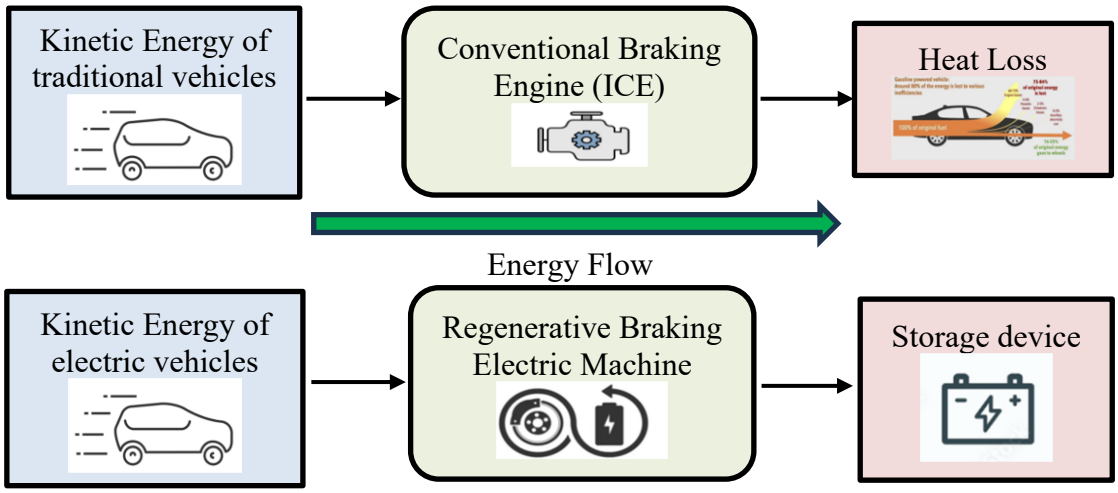


Figure 2. Energy flows for both traditional and electric vehicles.

Moreover, regenerative braking can capture 10% to 30% of kinetic energy, depending on the system. During aggressive braking, the friction brake mainly slows down the motorcycle, with regenerative braking offering additional support. However, under normal braking conditions, the regenerative system can recover up to half of the energy [14]. Advanced regenerative braking systems can recover 70% to 80% of energy using sophisticated control strategies, allowing for more precise and responsive braking. When the brakes are applied quickly, the energy recovery rate is lower than with moderate braking [15].

Regenerative braking is the only technique to recharge the batteries of electric vehicles without requiring any additional connection between the electric motor and the drive wheels. According to Vasiljevic et. al. [16], this type of brake can increase the travel range by up to 30%, depending on the type of vehicle, terrain, temperature, and environment. It allows batteries to recharge during downhill movement and deceleration and further store it in the onboard battery for later use [17]. The advantages of the regenerative braking system include extending the vehicle's range, reducing fuel consumption for hybrid cars, and decreasing maintenance for disc and brake pads of friction brakes [18].

On the other hand, electric vehicles without regenerative braking may not benefit from energy recovery capabilities, but they could offer other advantages. Depending on the vehicle was implemented, these might include simpler system designs, lower maintenance requirements, and reduced overall system complexity [19].

https://doi.org/10.31436/iiumej.v26i2.3413

2.2. Regenerative Braking Implementation using MOSFET

Electric motors are propulsive motors and act as generators to recharge the vehicle's battery during braking. Motor controllers can be adjusted to enable regenerative braking. The motor can be modelled as a series circuit consisting of an inductor and a voltage generator. The resistor and inductor represent the resistance and inductance of the motor's electromagnets. The voltage generator signifies the motor's spinning voltage, known as back electromagnetic force (EMF). The back EMF voltage is proportional to the motor's RPM. Figure 3 shows the connection between the inductor and the voltage source.

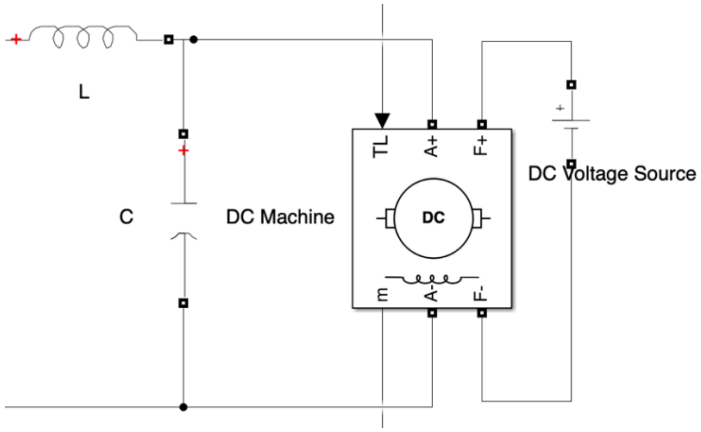


Figure 3. The circuit connection of the inductor and DC motor

When the motor rotates, it generates the back EMF voltage proportional to its rotational speed. During acceleration, the current decreases, and the motor's speed stabilises the back EMF to produce torque [20]. This torque is enough to overcome friction and the motor's mechanical load. However, if an external force drives the motor, such as the vehicle going downhill, the rotation can cause the back EMF to match battery voltage, and current will flow from the motor to the battery, creating a regenerative braking situation. Equation 1 describes the relationship between back EMF and torque.

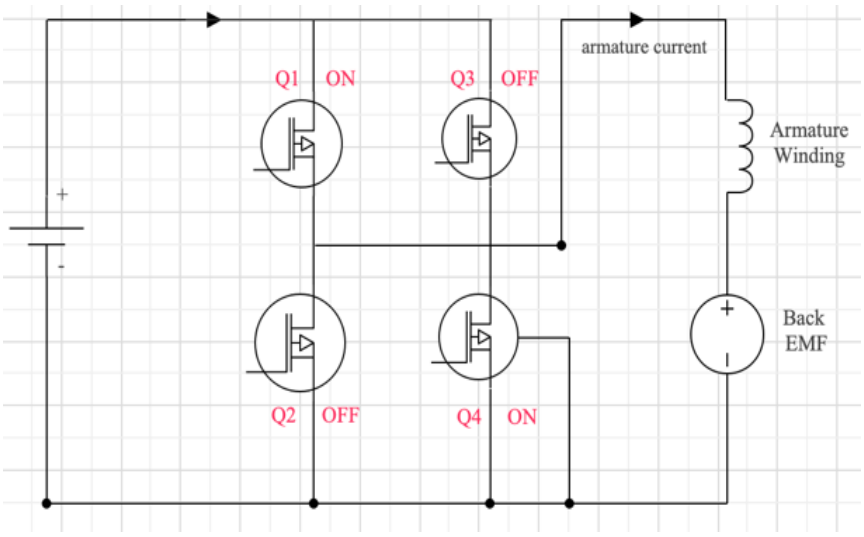
$$e_a(t) = \frac{1}{K_t}\left(R_a T_m(t) + L_a \frac{dT_m(t)}{dt}\right) + K_b V_b(t) \quad (1)$$

where T_m is the torque developed by the motor, V_b is the back EMF, R_a is the armature resistance, L_a is the armature inductance, K_t is the motor torque constant, and K_b is the back EMF constant.

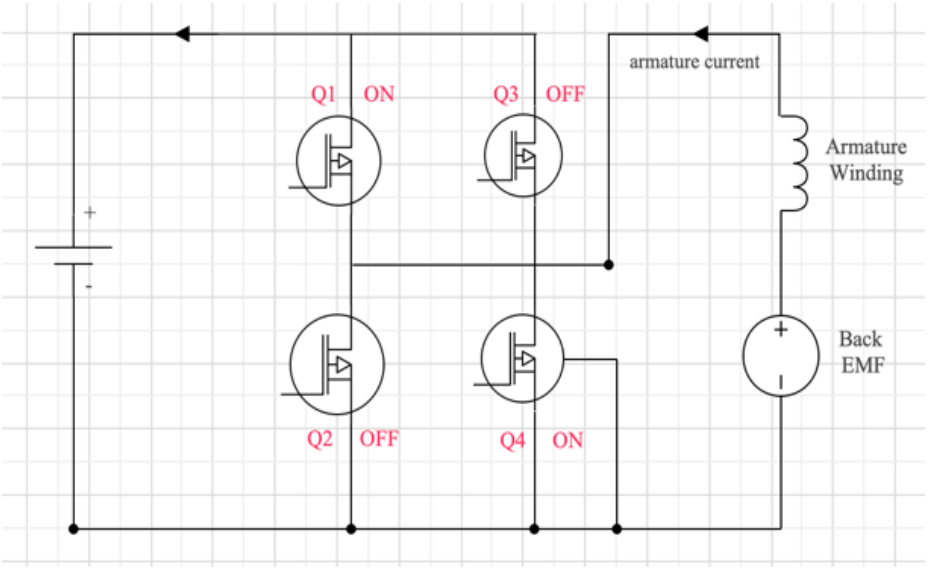
In regenerative braking systems of electric motorcycles, energy recovery can be achieved using an inverter and a DC machine, such as a brushless DC motor (BLDC) [21], a permanent magnet synchronous DC motor (PMSM) [22], and a DC motor. MOSFETs are widely used in power devices, such as DC–DC and DC–AC converters, due to their ability to switch rapidly between ON and OFF states by applying voltage to the gate. MOSFETs can operate in both motoring and generating modes. Meanwhile, the battery pack is also an essential part of the regenerative braking system, providing power during motoring mode and serving as an energy storage device during regenerative mode [24].

The MOSFET circuit consists of diodes that store energy back into the battery pack by conducting in the forward direction. This circuit is known as an inverter circuit, as it inverts the DC voltage supplied by the battery using a controlled switching sequence of the switches Q1 to Q4. In monitoring mode, a controller and gate driver integrated circuit (IC) drive the motor during the initial state and acceleration. A BLDC motor offers power-saving advantages

over other motor drives, high efficiency, and excellent controllability [25]. The circuit diagram operation for the electric motorcycle can be divided into two stages: a) motoring mode and b) regenerative mode, as shown in Figure 4.



(a) Driving mode



(b) Regenerative mode

Figure 4. The equivalent circuit of a BLDC circuit and an inverter

During motoring mode, the high-side switches Q1 and Q4 operate in pulse width modulation (PWM), while the low-side switches Q2 and Q3 operate in standard high/low switching mode. PWM allows for control of the motor's torque, with current flowing from the positive to the negative terminal of the voltage source. Varying voltage can be achieved by turning the power to the motor ON and OFF using MOSFETs arranged in a half-bridge (unidirectional) or full bridge (bidirectional). Combined with the motor's inductance, this switching effectively makes the controller behave like an adjustable voltage source, proportional to the switch's ON/OFF duty cycle; however, in braking energy regeneration for electric motorcycles, the electric machine functions as a generator, inducing a back EMF. This back EMF is usually lower than the battery voltage. The induced back EMF must be boosted

https://doi.org/10.31436/iiumej.v26i2.3413

to charge the battery using a DC-DC converter. In regenerative mode, every switch from Q1 to Q4 operates in PWM switching mode. Due to the continuous switching of the PWM signal, a distinction must be made between the ON and OFF PWM conditions. Compared to motoring mode, in regenerative braking mode, the current flows from the positive terminal of the back EMF to the negative terminal.

3. METHODOLOGY

Each electric motorcycle has a few main parts, including an electric motor, a chassis, a transmission, and two wheels. The sizes and features of these components depend on the type of motorcycle. There are five categories of motorcycles: standard bikes, cruisers, touring bikes, sport bikes, and dirt bikes. Motorcycles are quite complex machines with many finely tuned mechanical parts. Figure 5 shows the electric motorcycle structure for the regenerative braking system.

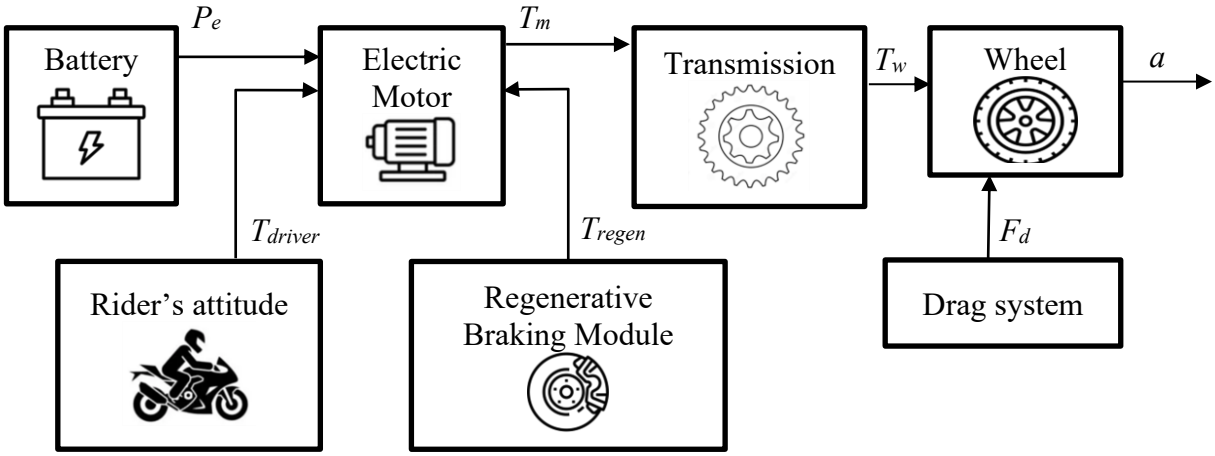


Figure 5. Structure of the motorcycle model.

The electric motorcycle model will capture the longitudinal dynamics, braking forces, energy, and battery state of charge (SoC). The battery simulates the consumed electric power, $P_{e,}$ to produce the torque on the electric motor, T_m. This motor torque is then used to simulate the single gear in the transmission, considering the gear ratio to generate the wheel torque, T_w. The longitudinal vehicle dynamics help with the vehicle's longitudinal deceleration. The model takes rider input and the regenerative braking module as inputs for the motorcycle model.

3.1. Electric Motorcycle Modeling

The motorcycle is designed to move primarily in one direction, influenced by all the forces acting along this direction. We only consider single-dimensional longitudinal dynamics. Figure 6 shows the force at work on a motorcycle moving up an inclined road at an angle of θ.

https://doi.org/10.31436/iiumej.v26i2.3413

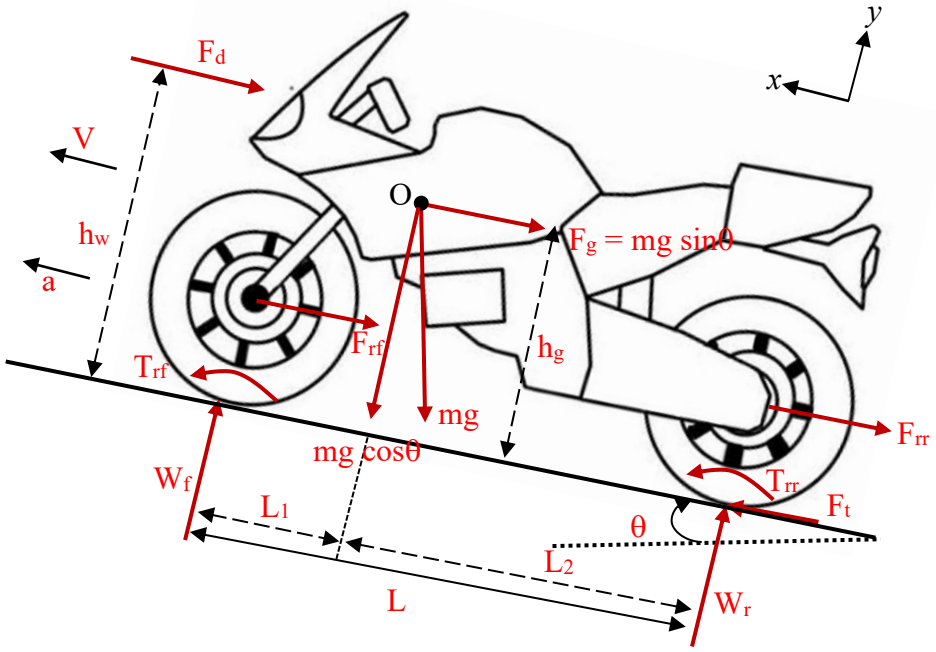


Figure 6. Forces acting on a motorcycle in the longitudinal direction [26].

The traction force, F_t, is created by the friction between the drive wheel's tire and the road surface, propelling the motorcycle forward. This force comes from the engine's torque and is transferred through the transmission to the drive wheel. As the motorcycle moves, it faces resistance from tire rolling, air drag, and gravity, especially when going up or down a hill. According to Newton's second law, acceleration and deceleration can be calculated by finding the difference between the traction force and the sum of these resistance forces.

$$\frac{dV}{dt} = \frac{F_t - \sum F_f}{m} \tag{2}$$

where V is the speed of the motorcycle along the longitudinal x-direction, F_t is the total tractive force, $\sum F_f$ is the total resistance force, m is the motorcycle's total weight.

In the longitudinal direction, the major external forces on a two-axle motorcycle are the rolling resistance of the front and rear tires, the air resistance force, the uphill or downhill resistance force, and the tractive force. Then, the mathematical modeling of motorcycle motion in the longitudinal direction can be expressed by

$$m\frac{d}{dt}V = F_t - (F_{rf} + F_{rr} + F_d + F_g) \tag{3}$$

The first term on the right side is the total traction force, while the second term is the resistance.

$$\left.\begin{aligned} F_r &= mgf_r \cos\theta \\ F_d &= \tfrac{1}{2}\rho A_f C_d (V - V_w)^2 \\ F_g &= mg \sin\theta \end{aligned}\right\} \tag{4}$$

https://doi.org/10.31436/iiumej.v26i2.3413

Equation (4) represents the resistance of the motorcycle, which hinders its motion, encompassing the rolling resistance of the tires, Fr, aerodynamic drag, Fd, and grading resistance, Fg. Moreover, Cd signifies the aerodynamic drag coefficient, which is contingent on the motorcycle's shape, and Vw denotes the wind speed in the motorcycle's direction. The frontal area, A_f is usually 70-90% of the area calculated from the motorcycle's width and height with riders. The aerodynamic drag coefficient, C_d, typically ranges from 0.5 to 0.7 and g represents gravity, which is 9.8 m/s².

3.2. Modeling the Regenerative Braking System

The regenerative braking system considers various factors to recuperate energy and store it in the battery pack. The regenerative braking force produces torque for the system. While ascending, the regenerative force, Fregen, diminishes because of the augmented force, Fload, from the incline. Conversely, the regenerative force escalates while descending as the load force interacts with the downward slope.

$$\left.\begin{aligned} F_{regen} &= F_{br} - F_{fric} \\ F_b &= F_{load} - F_a \\ ma &= F_{br} - F_{load} \end{aligned}\right\} \quad (5)$$

where F_{br} is the motorcycle braking force, F_{load} is the road load force, which includes aerodynamic and rolling resistance, and F_a is the inertial force. However, F_{br} £ 0 is the deceleration force from the deceleration device, and F_{fric} £ 0 is the frictional force from the hydraulic mechanical braking system.

Energy and power are key factors in energy management systems. Therefore, several formula formulations can be used to analyze energy recovery as well as energy consumption for the system.

$$\left.\begin{aligned} E_c &= \frac{E_b}{d} \\ E_b &= E_t + E_{load} + E_{loss} - E_{regen} \\ E_b &= \int_{traction} P_{b_{out}} dt + \int_{regen} P_{b_{in}} dt \\ P_{b_{out}} &= \frac{V}{\eta_m}\left(m\frac{dV}{dt} + mgf_r + \frac{1}{2}\rho C_d A_f V^2\right) \\ P_{b_{in}} &= \frac{\alpha V}{\eta_m}\left(m\frac{dV}{dt} + mgf_r + \frac{1}{2}\rho C_d A_f V^2\right) \end{aligned}\right\} \quad (6)$$

Equation (6) is about energy consumption and energy. Ec is the energy consumed, Eb is the total energy from the battery, and d is the distance travelled. Et is the total energy required to move the motorcycle, Eload is the energy for accessories on the motorcycle, Eloss is the energy lost due to inefficiencies, and Eregen is the energy generated during regenerative braking. So, the net energy consumption is the sum of the power output at the battery terminal and the regenerative braking power, with a negative sign. When the battery energy consumption matches the total energy in the battery, the resistance power, and the effectiveness of regenerative braking. Pbout is the battery power output, and Pbin is the regenerative braking power output with the regenerative braking factor, a, which ranges from 0 to 1. Finally, the total power required for the electric motorcycle can be calculated using equation (7).

$$P_{req} = F_t V \quad (7)$$

https://doi.org/10.31436/iiumej.v26i2.3413

3.3. MATLAB Simulation Modeling

In this part, a vehicle dynamic modelling approach is designed using MATLAB/Simulink. The proposed modelling is based on mathematical modelling in the longitudinal direction. The detailed MATLAB/Simulink model of the electric motorcycle is shown in Figure 7. It consists of three primary blocks: forces acting on the motorcycle, power, and energy. The input is the speed from the drive cycle, and the outputs are distance travelled, energy, power, and torque. This simulation allows each functional block to be easily upgraded and improved. Additionally, the model is developed using equations (3-6), which can be solved directly within the model. This simulation aims to understand the concept of electric motorcycles before implementing any control strategies in the regenerative braking system. Meanwhile, Table 1 shows the motorcycle parameters used in the MATLAB simulation based on the retrofitted Honda CBR250RR [10].

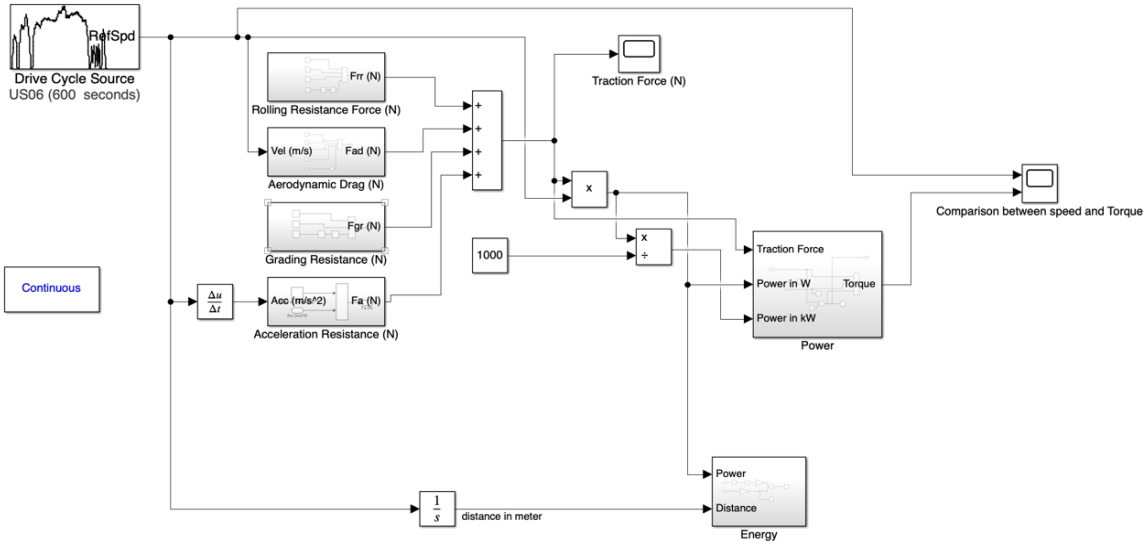


Figure 7. MATLAB/Simulink for electric motorcycle model.

Table 1. Motorcycle dynamic parameters

Parameter	Symbol	Value
Mass of vehicle	M	180 kg
Acceleration due to gravity	G	9.81 m/s^2
Rolling resistance coefficient (on road)	$_{fr}$	0.013
Grading angle	θ	0 ^{o}
Density of air	ρ	1.202 kg/m^3
Radius of tire	r_w	0.3149 m
Vehicle frontal area	A_f	0.6 m^2
Aerodynamic drag coefficient	C_d	0.7
Gear ratio	GR	0.192054

Figure 8 shows the model for the regenerative braking system using four MOSFETs and a DC machine. The load torque input is applied to the motor in this MATLAB simulation. The model includes a 50 HP, 240 V, 1750 RPM machine with a series inductor and capacitor, connected to a lithium-ion battery. For a battery, it will represent the voltage, current, and state

https://doi.org/10.31436/iiumej.v26i2.3413

of charge (SoC). The MOSFETs act as the switch. When S1 and S4 are ON, the electric motorcycle can either move forward or brake, depending on acceleration and deceleration. During deceleration, the torque becomes negative, indicating regenerative braking, increasing the battery's charge state. The battery's state of charge will start to increase if it is at 80% or below.

The input torque, TL, is the torque applied to the shaft. The flow of the armature determines the motor's mode. The electromechanical torque, T_e, is proportional to the armature current, I_a.

$$T_e = K_T I_a \tag{8}$$

K_T is the torque constant. The motor operates in motor mode if the input torque, TL, is greater than 0 (TL > 0). The system operates in generator mode if TL is less than 0 (TL < 0).

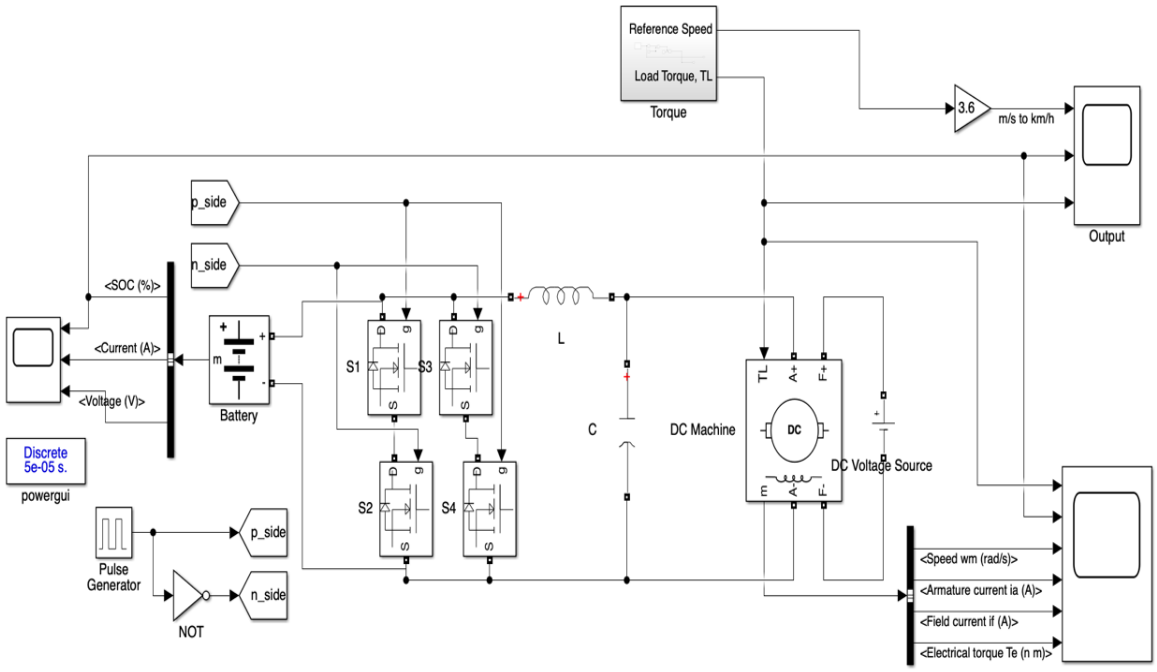


Figure 8. MATLAB/Simulink for regenerative braking of the electric motorcycle model.

In the MATLAB simulation, the four MOSFET switches (S1–S4) operate in pulse width modulation (PWM) mode. Pulse width modulation (PWM) is used as the switching mode to control the torque developed by the motor. The PWM signal continuously switches the MOSFETs ON and OFF, distinguishing between the ON and OFF PWM conditions. When the switching mode is changed to the energy-regenerative mode, a braking signal from the rider is sent to the controller. The PMW switching mode continuously switches the switches ON and OFF. In the ON PWM condition, switches S1 and S4 are turned ON, supplying voltage from the battery to energize winding L. To achieve this, switches S2 and S3 are turned OFF, changing to the OFF PWM condition, thereby supplying energy back to the battery.

https://doi.org/10.31436/iiumej.v26i2.3413

4. SIMULATION RESULTS AND DISCUSSIONS

This paper aims to study and implement the regenerative braking system that recovers energy wasted during braking and stores it back in the battery for further driving. The new model of regenerative braking in the electric motorcycle is presented in a MATLAB simulation. This study considers two drive cycles: US60 and NEDC. The drive cycles are represented by graphs that plot vehicle speed against time. They are used to evaluate the performance of the motor, transmission, and emissions control system. The US60 drive cycle represents aggressive, high-speed, and/or high-acceleration driving behavior, rapid speed fluctuations, and driving behavior following startup. Meanwhile, the NEDC drive cycle, known as the New European Driving Cycle, represents the typical European car usage. It consists of four repeated urban driving cycles lasting until 800 s and an extra-urban driving cycle.

Figure 9 shows the drive cycle of US60 and the torque graph obtained from the motorcycle modelling. From the figure, the US60 drive cycle features high speed and quick acceleration, covering it in 10 minutes. It averages 77 km/h and reaches a top speed of 130 km/h, with four stops included. The cycle represents a 12.8 km route with an average speed of 78 km/h, with a duration of 596 seconds. For the torque, negative torque occurs when the speed is decreasing in the last 10 minutes. This negative torque helps determine the regenerative braking system.

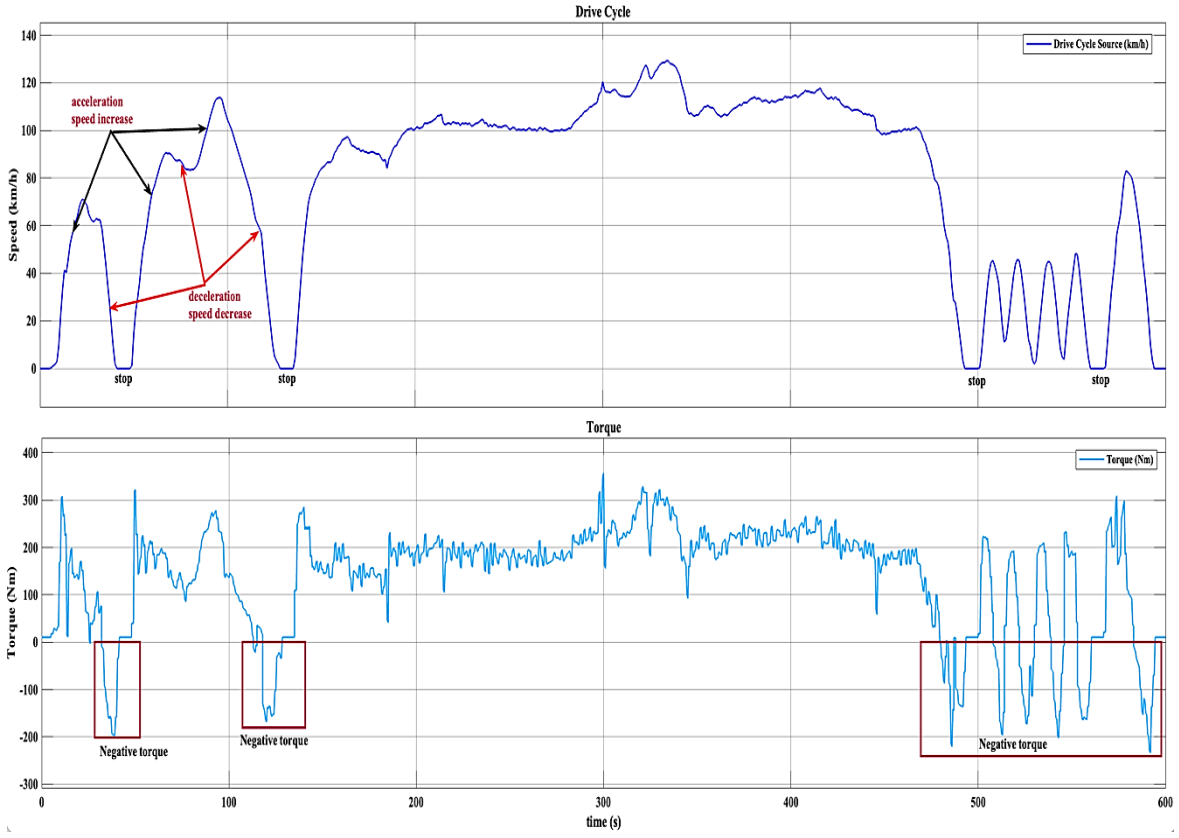


Figure 9. US60 drive cycle and produced torque.

Figure 10 shows the NEDC drive cycle in kilometers per hour for the first graph, while the second graph illustrates the torque over 1180 seconds. These graphs are generated using the vehicle modeling equation. The complete NEDC consists of four repeated urban cycles followed by an extra-urban cycle to simulate highway driving, reaching a maximum 120 km/h speed. This cycle is commonly used for testing electric motorcycle energy consumption and regenerative braking performance testing. The speed pattern for the extra-urban cycle

https://doi.org/10.31436/iiumej.v26i2.3413

fluctuates and increases until it reaches 120 km/h, then decreases. The negative torque for the extra-urban cycle occurs only once before the motorcycle stops, allowing maximum energy recovery optimization in this phase.

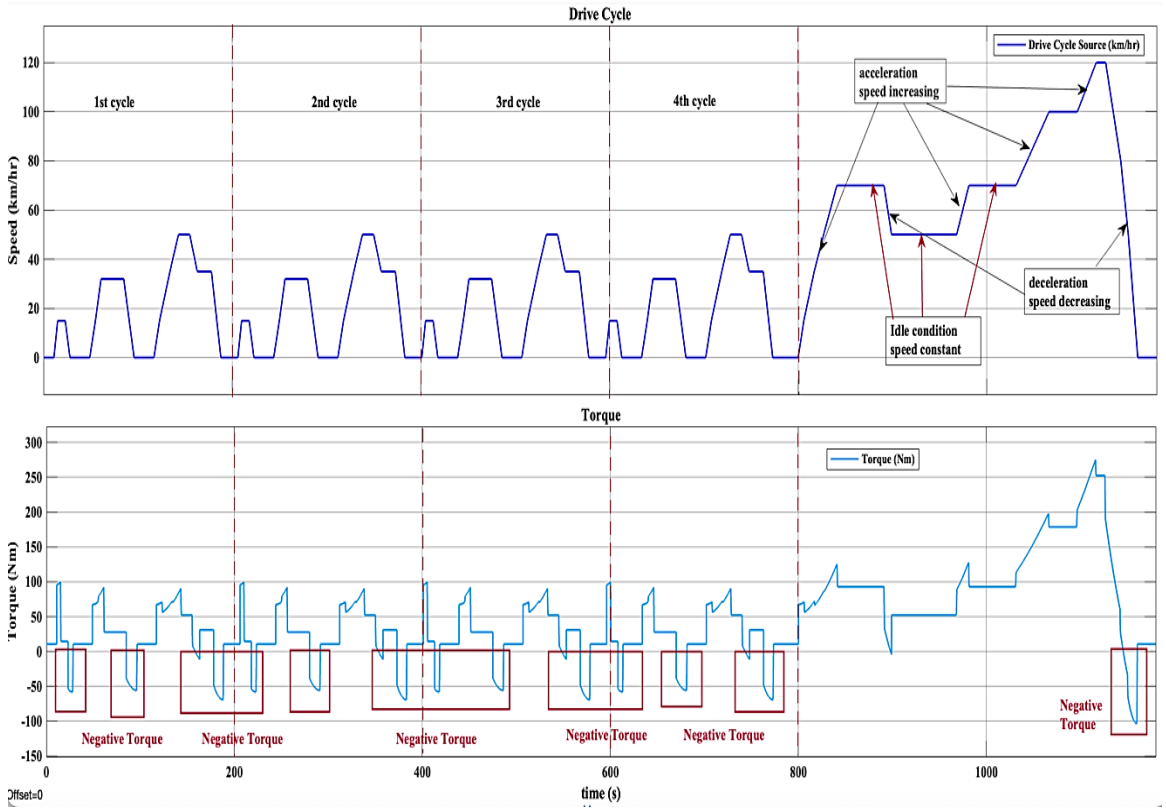
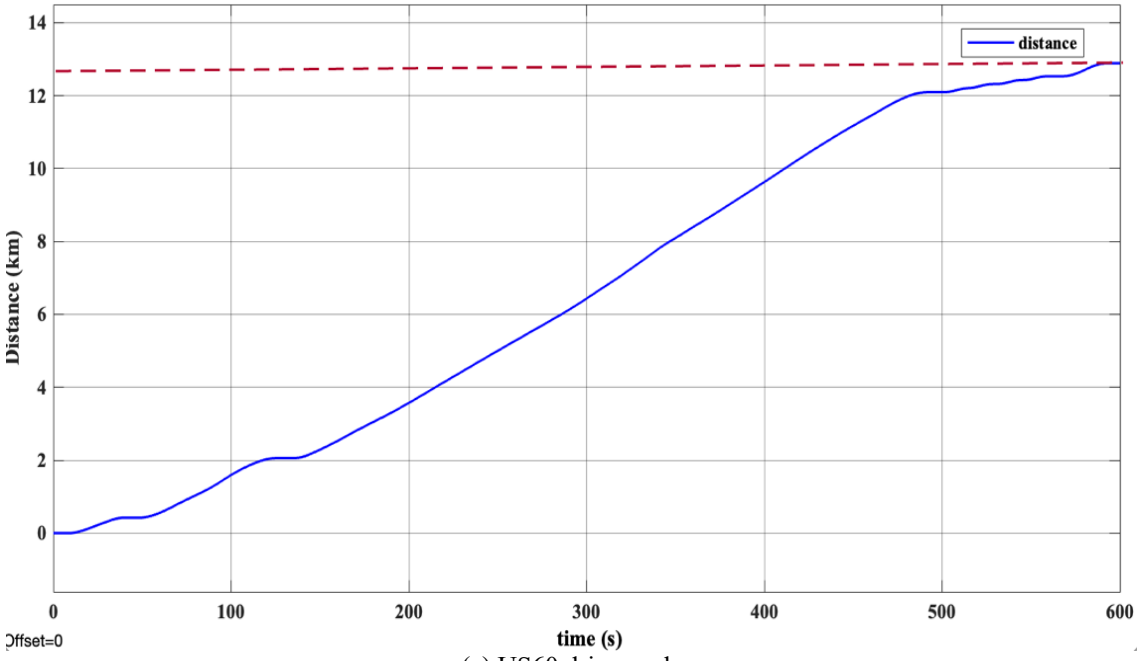


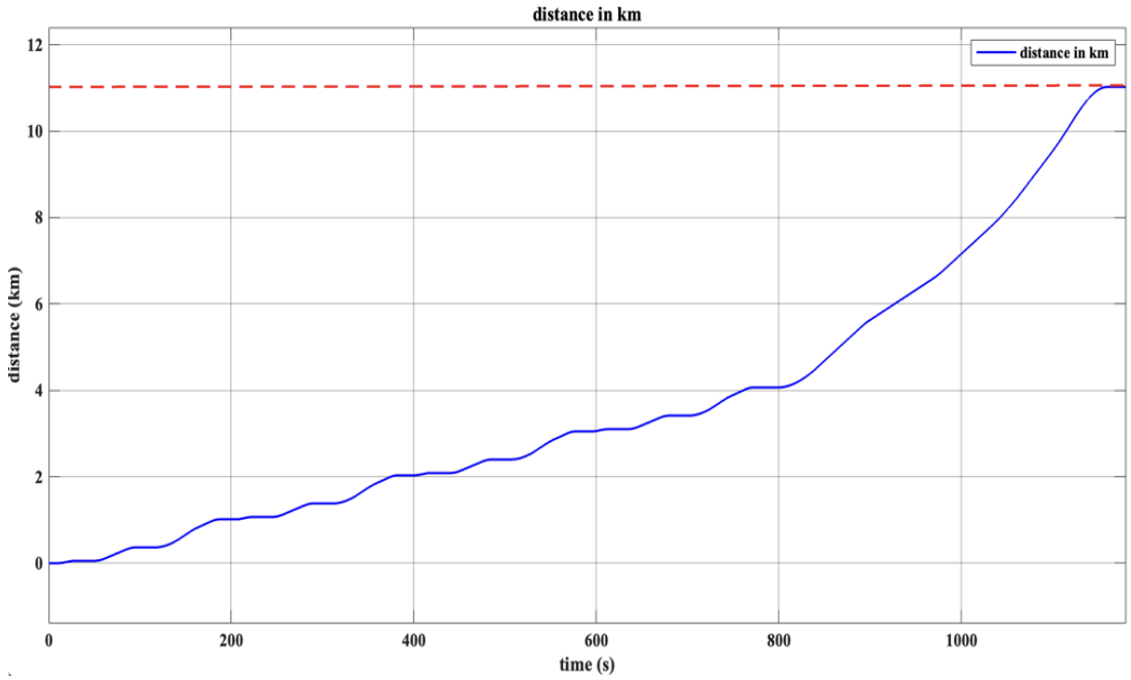
Figure 10. NEDC drive cycle and produced torque.

In the MATLAB simulation, the energy consumed can be estimated by integrating the total power and dividing it by the total distance covered by the electric motorcycle in kilometers. The total distance for the US60 drive cycle in 600 seconds is approximately 13 km, while for the NEDC drive cycle in 1180 seconds, the distance travelled is 11 km, as shown in Figure 11. Meanwhile, Figure 12 illustrates the power consumption for the US60 and NEDC drive cycles. The total power consumed for both cycles is 120 Wh/km and 61 Wh/km, respectively. Once energy consumption is determined, the energy recovery through regenerative braking can be calculated.

https://doi.org/10.31436/iiumej.v26i2.3413



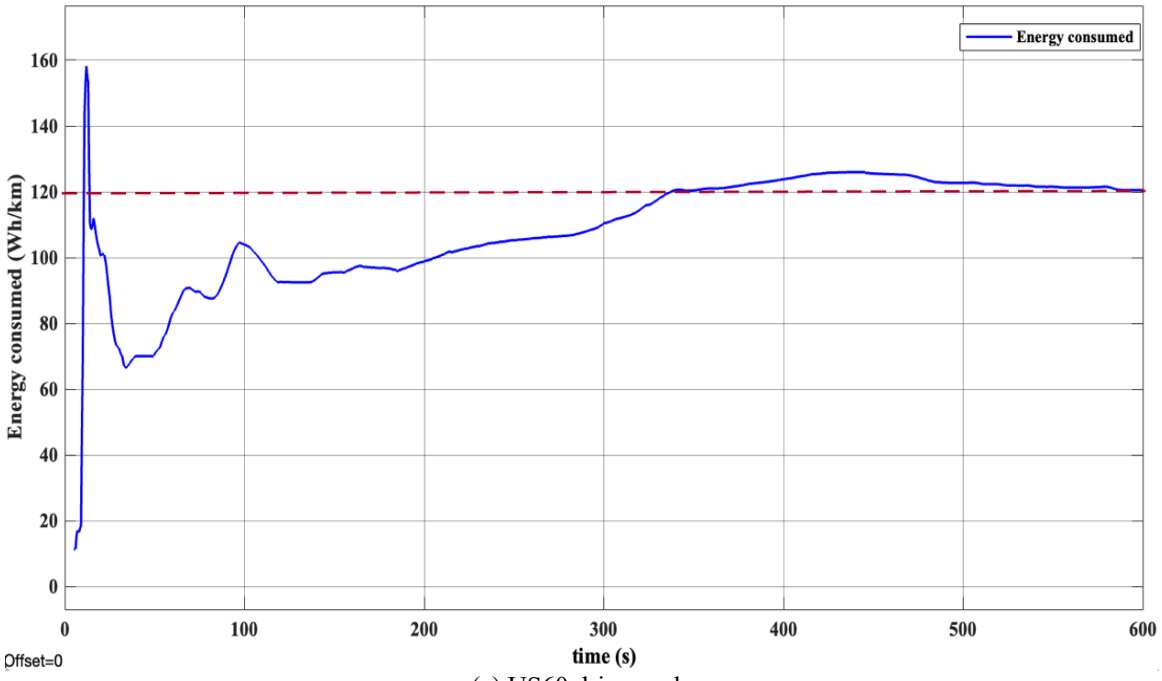
(a) US60 drive cycle



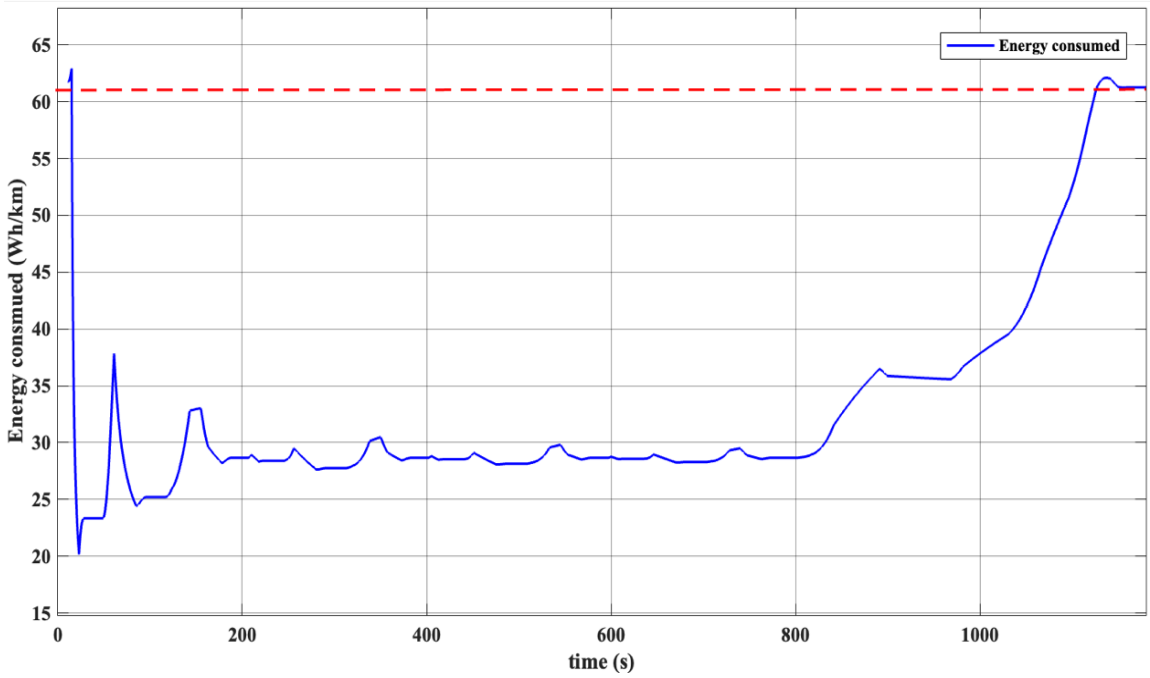
(b) NEDC drive cycle

Figure 11. Total distance for US60 and NEDC drive cycles.

https://doi.org/10.31436/iiumej.v26i2.3413



(a) US60 drive cycle



(b) NEDC drive cycle

Figure 12. Total energy consumed for US60 and NEDC drive cycles.

The concept of a regenerative braking system is studied in the second implementation using MATLAB simulation. As mentioned earlier, negative torque indicates the occurrence of regenerative braking. During regenerative braking, the state of charge (SoC) of the battery increases, essentially recharging the battery. This recharging occurs when negative torque is applied, causing the battery to charge as the speed decreases. In this study, the concept of regenerative braking is applied specifically to the US60 cycle. Two SoC levels, 80% and 50%, are chosen to evaluate the effectiveness of braking within the system. In this system, the input torque, TL, is provided by a lithium-ion battery and four MOSFETs connected to a DC

https://doi.org/10.31436/iiumej.v26i2.3413

machine. Figure 13 displays the speed graphs based on the drive cycle, input torque, and state of charge over 250 seconds.

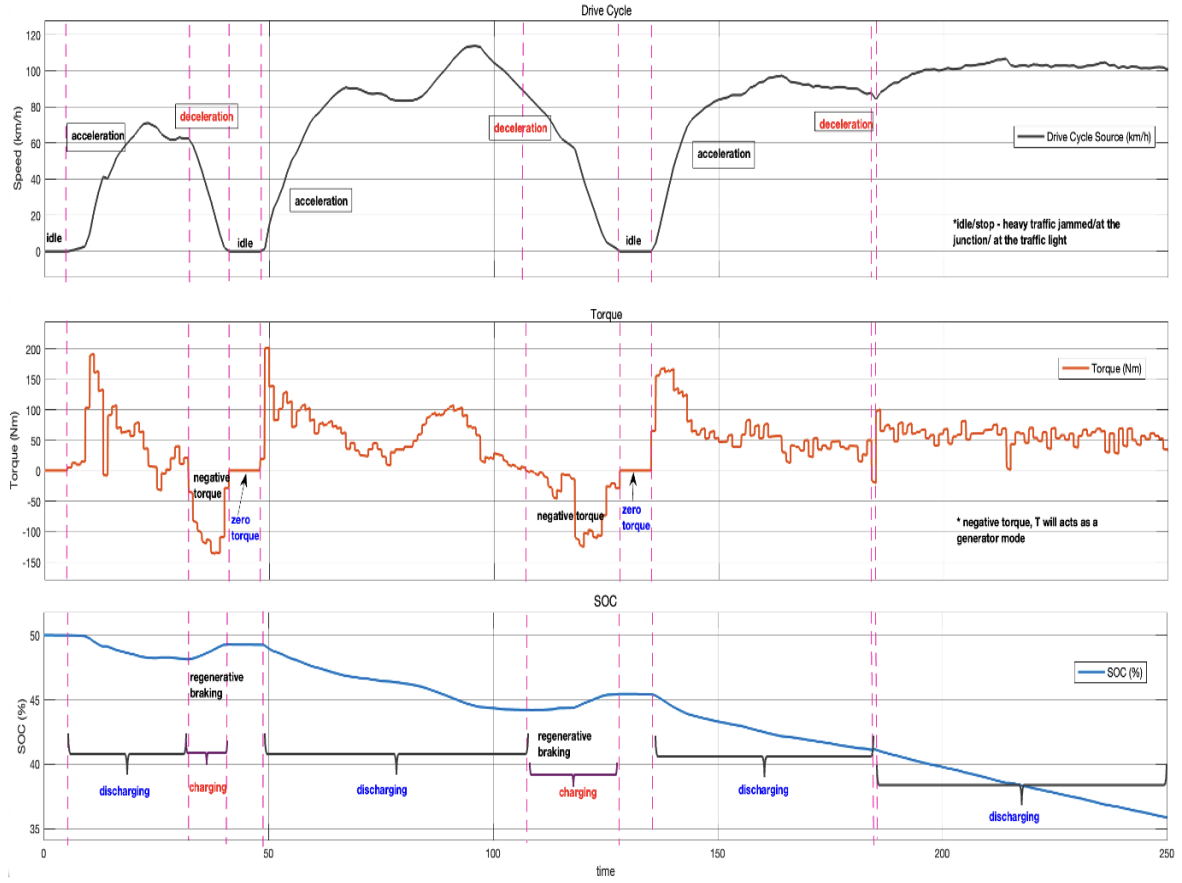


Figure 13. The graphs of drive cycle, torque, and SoC for US60.

In regenerative braking mode, the performance study is conducted separately for a specific level of SOC and speed using a US60 duty cycle. Figure 14 illustrates the details of regenerative braking for two levels of SoC: 80% and 50%. These graphs demonstrate that as the speed decelerates, the torque decreases. While the battery charges, the SoC graph increases, indicating the system is in a regenerative braking state. In both torque graphs, regenerative braking occurs twice between t = 0 s and t = 250 s when the torque is negative.

The increase in SoC (%) during braking, as shown in the graph in Figure 14, can be calculated using the following equation:

$$\text{SoC}\ (\%) = \frac{\Delta Y}{\Delta t} = \frac{Y_2 - Y_1}{t_f - t_i} \times 100\% \tag{9}$$

where Y_2 and Y_1 are the final and initial values on the *y-axis,* and t_f and t_i refer to the initial time and final time.

https://doi.org/10.31436/iiumej.v26i2.3413

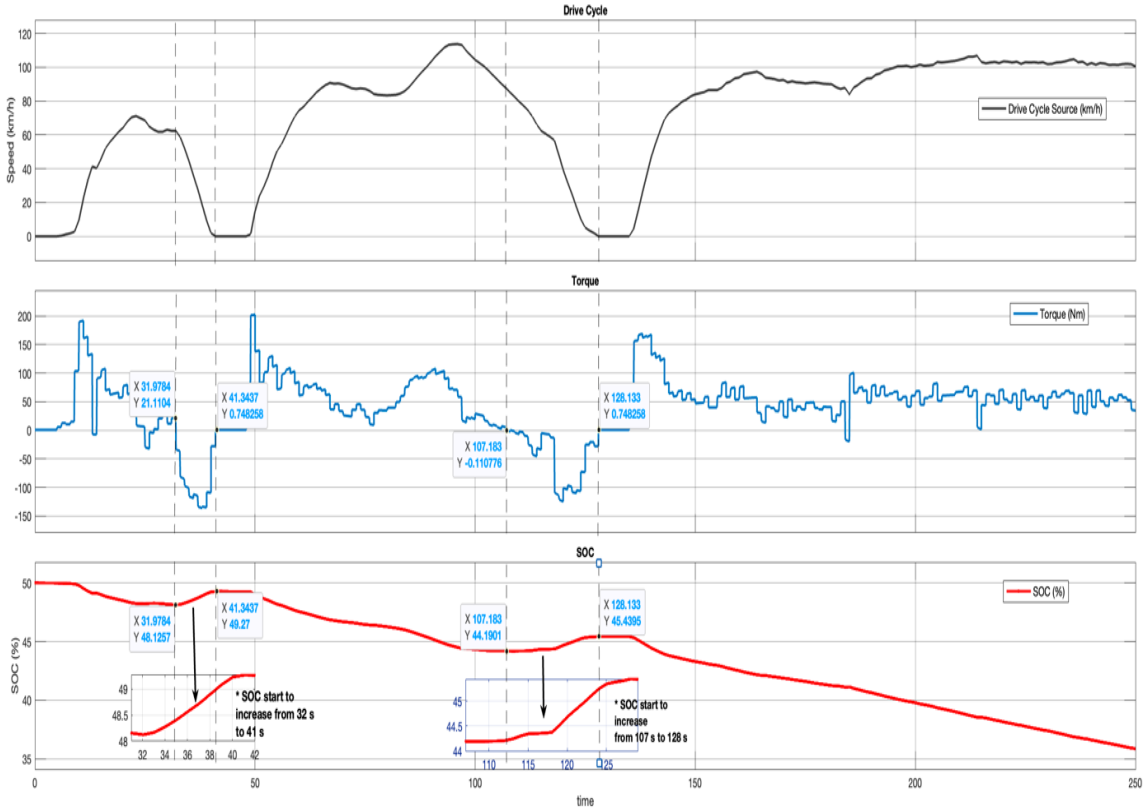


Figure 14. The details graph for regenerative braking occurs between 0 to 250 s for US60.

The results indicate that over a longer duration, the SoC increases slightly more than over a shorter duration. At a 50% SoC level, the first regeneration resulted in a 12.22% increase, while the second regeneration increased by only 5.96%. At an 80% SoC level, the first regeneration achieved a 12.55% increase, while the second regeneration resulted in a 6.19% increase. The summarized results that were obtained are recorded in Table 2. The objective remains to optimize energy recovery and minimize wastage or loss.

Table 2. The percentage of SoC increases during regenerative braking

Level of State of Charge (%)	1st regen	2nd regen
50	12.22	5.96
80	12.55	6.19

The results obtained for both SoC levels show that the percentage increase during each regeneration is similar. Therefore, optimizing energy recovery in both regenerations is feasible. To maximize energy recovery in electric motorcycles, regenerative braking control strategies such as fuzzy logic control (FLC), neural network (NN), adaptive control, and model predictive control (MPC) can be employed. Additionally, implementing these control strategies can enhance the range traveled and reduce rider anxiety.

5. CONCLUSION

In conclusion, regenerative braking is an effective technique for optimizing energy recovery in electric motorcycles. The results from the simulation indicate that using regenerative braking increases the SoC, enabling the generated energy to be stored in the battery for future use. The longer braking periods allow for maximum energy recovery.

Therefore, regenerative braking in electric vehicles represents another technological approach to effective energy management. Additionally, energy recovery through regenerative braking can reduce the overall cost of electric motorcycles and decrease the need for regular maintenance compared to traditional motorcycles.

Once the fundamentals of regenerative braking in electric motorcycles are understood, where the electric motor functions as a generator, future studies should focus on maximising energy recovery during deceleration or braking using braking strategy control, the system can optimize energy usage by reducing consumption while increasing the required energy through the use of various control strategy methods. These control strategies effectively enhance energy recovery. With the implementation of a regenerative braking system controlled by methods like Model Predictive Control (MPC), energy recovery could potentially increase by at least 20% compared to the current systems.

ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude and appreciation to the Ministry of Higher Education Malaysia for research grant FRGS1/2023/TK02/UIAM/01/1 Optimal Deceleration Formulation Based on Reinforcement Deep Learning for Energy Generation From Regenerative Braking Mode of an Electric Motorcycle, as well as the support provided by Department of Mechatronics, Kulliyyah of Engineering, International Islamic University and Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh for their support in completion this research project.

REFERENCES

[1] National Energy Transition Roadmap. Retrieved from: https://www.ekonomi.gov.my/sites/default/files/2023-09/National%20Energy%20Transition%20Roadmap_0.pdf

[2] Laporan Tahunan 2020. Retrieved from:https://www.parlimen.gov.my/ipms/eps/2021-12-16/ST.200.2021%20-%20ST%20200.2021.pdf

[3] National Energy Policy, 2022 – 2024. Retrieved from: https://www.ekonomi.gov.my/sites/default/files/2022-09/National%20Energy%20Policy_2022_2040.pdf

[4] Dua pelan hala tuju peralihan tenaga akan dilancar tahun ini. (26 Jun 2023). Retrieved from: https://www.mida.gov.my/ms/berita-mida/dua-pelan-hala-tuju-peralihan-tenaga-akan-dilancar-tahun-ini/

[5] Chapree C. (13 October 2023) Budget 2024: Electric motorcycle incentive scheme to provide rebates of up to RM2,400. Retrieved from: https://soyacincau.com/2023/10/13/budget-2024-electric-motorcycle-rebate/

[6] Silaghi H, Gamcova M, Silaghi A, Spoială V, Silaghi A, Spoială D. (2018) Intelligent Control of Electrical Drive System Used for Electric Vehicles. The Scientific Bulletin of Electrical Engineering Faculty,18(1) 5-10.

[7] Muslimin S, Nawawi Z, Suprapto BY, Dewi T. (2022) Comparison of Batteries Used in Electrical Vehicles (A Review). Atlantis Highlights in Engineering, volume 9, pp 421 – 425.

[8] Ga Bui V, Minh Tu Bui T, Tuan Hoang A, Nižetić, S, Sakthivel R, Nam Tran V, Hung Bui V, Engel D, Hadiyanto H. (2021) Energy storage onboard zero-emission two-wheelers: Challenges and technical solutions. Sustain. Energy Technol. Assess, 47, 101435.

[9] Ahmed TH, Mehmet FO. (2020) An overview of regenerative braking systems. Journal of Energy Storage, Volume 52, Part C.

[10] Hanifah RA, Toha SF, Hanif NHHM, Kamisan NA. (2019). Electric Motorcycle Modeling for Speed Tracking and Range Travelled Estimation. IEEE Access, volume 7, pp 26821–26829.

[11] Rego N, Castro R. (2023) Regenerative Braking Applied to a Student Team's Electric Racing Motorcycle Prototype: A Theoretical Study. Applied Sciences, 13(6):3784.

[12] Balasubramanian B, Leyland A. (2017) Development of regeneration braking model for electric vehicle range improvement. 2017 IEEE Transportation Electrification Conference (ITEC-India), 1 – 5.

[13] Paredes MGSP, Pomilio JA. (2018) Control System to Regenerative and Anti-Lock Braking for Electric Vehicles. 2018 IEEE Transportation Electrification Conference and Expo (ITEC), Long Beach, CA, USA, 2018, pp. 135 – 139.

[14] Szumska EM, Jurecki R. (2022) The Analysis of Energy Recovered during the Braking of an Electric Vehicle in Different Driving Conditions. Energies, 15(24):9369.

[15] Muzir NAQ, Mojumder MRH, Hasanuzzaman M, Selvaraj J. (2022) Challenges of Electric Vehicles and Their Prospects in Malaysia: A Comprehensive Review. Sustainability, 14(14):8320.

[16] Vasiljevic S, Aleksandrovic B, Maslac M. (2022) Regenerative braking on electric vehicles: working principles and benefits application. IOP Con. Series. Materials Science and Engineering, 1271 (2022) 012025.

[17] Gao Y, Lin H, Yi F, Zhou X, Qi L, Li Y. (2022) A Car-Following Model with the Acceleration Generalized Force Coupled with External Resistance and the Temporal-Spatial Distribution of Battery Decline. Batteries, 8(12):257.

[18] Pavan KP. (2023) Design of Regenerative Braking System. International Journal For Multidisciplinary Research, Volume 5, Issue 4.

[19] Alanazi F. (2023) Electric Vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation. Applied Sciences, 13(10):6016.

[20] Nama T, Mondal P, Tripathy P, Adda R, and A. K. Gogoi AK. (2022) Design, Modeling and Hardware Implementation of Regenerative Braking for Electric Two-Wheelers for Hilly Roads. IEEE Access, vol. 10, pp. 130602-130618.

[21] Mamur H, Candan AK. (2020) Detailed simulation of regenerative braking of BLDC motor for electric vehicles. Bilge International Journal of Science and Technology Research, 4 (2), 63 – 72.

[22] Qian G, Chengxing L, Na Z, Hechao Z, Huilue J, Zhaowen Z, Fengli Z. (2018) Regenerative braking system of PM synchronous motor. AIP Conf. Proc, 1955 (1): 040111.

[23] Nuñez-Dorantes J, Martinez-Sibaja A, Rodriguez-Jarauin J, Aquavo-Alquicira J. (2022) Comparison of energy recovery in the regenerative braking system in an electric vehicles using MOSFET's and the IRAMY inverter module. DYNA, 97(6), 625 – 630.

[24] Anurag B, Adinath K, Mredhul KP, Jaydeep A. (2019) Design and Fabrication of Regenerative Braking in EV. VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 2, Article No. 2 PP 1-6.

[25] Minh Dinh B. (2023) Torque Performance Analysis of BLDC Motor for Electric Motorcycle by Halbach and Skewing Rotor. In 13th Annual International Conference on Industrial Engineering and Operations Management.

[26] How to Draw a Motorcycle. Available: https://www.wikihow.com/Draw-a-Motorcycle/

https://doi.org/10.31436/iiumej.v26i2.3448

SMART IOT ENERGY OPTIMISATION AND LOCALISATION MONITORING FOR E-BIKE SHARING

MAWADA AHMED MOHAMED[1], SITI FAUZIAH TOHA[1]*, MD ATAUR RAHMAN[2], MOH. KHAIRUDIN[3]

[1] *Department of Mechatronics Engineering, International Islamic University Malaysia, Malaysia*
[2] *Eastern Michigan University, Sill Hall, GameAbove College of Engineering & Technology, Ypsilanti, Michigan, USA 48197*
[3] *The Doctoral Program on Engineering Science, Department of Electrical Engineering, Universitas Negeri Yogyakarta, Yogyakarta 55281, Indonesia*

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

(Received: 31 October 2023; Accepted: 22 April 2025; Published online: 10 May 2025)

ABSTRACT: E-bike sharing has emerged as a sustainable and convenient mode of transportation, offering lightweight, energy-efficient mobility solutions. However, existing systems face challenges such as limited input parameters for modeling, leading to inefficiencies in energy optimization algorithms and power assist mechanisms. A significant concern is the rapid depletion of batteries, which reduces the availability of e-bikes, increases operational costs for fleet managers, and impacts user satisfaction. To address these challenges, this project developed a Smart IoT Energy Optimization and Localization Monitoring System that integrates multi-sensor data, IoT connectivity, and advanced data analytics to monitor real-time usage patterns, battery levels, and the location of e-bikes. The methodology involved integrating sensors to collect key data, implementing connectivity for real-time monitoring, and developing an energy optimization algorithm to prolong battery life, improving the efficiency of the e-bike sharing system. The results demonstrated a 15% improvement in energy efficiency, which increased battery state-of-charge (SOC) and extended operational range. Additionally, the system enabled better fleet management by ensuring optimal energy usage and the availability of e-bikes for users. This approach aligns seamlessly with the Sustainable Development Goals (SDGs) by promoting eco-friendly transportation and enhancing user accessibility. The integration of IoT technology has proven effective in overcoming the limitations of traditional systems, offering a scalable and efficient solution for modern urban mobility.

ABSTRAK: Perkongsian e-basikal telah muncul sebagai kaedah pengangkutan yang lestari dan mudah, menawarkan penyelesaian mobiliti yang ringan dan cekap tenaga. Walau bagaimanapun, sistem sedia ada menghadapi cabaran seperti parameter input yang terhad untuk pemodelan, yang menyebabkan ketidakcekapan dalam algoritma pengoptimuman tenaga dan mekanisme bantuan kuasa. Masalah utama adalah penurunan bateri yang cepat, yang mengurangkan ketersediaan e-basikal, meningkatkan kos operasi untuk pengurus armada, dan memberi kesan kepada kepuasan pengguna. Untuk mengatasi cabaran ini, projek ini membangunkan Sistem Pemantauan Pengoptimuman Tenaga dan Lokalisasi IoT Pintar yang mengintegrasikan data multi-sensor, sambungan IoT, dan analitik data lanjutan untuk memantau corak penggunaan masa nyata, tahap bateri, dan lokasi e-basikal. Metodologi ini melibatkan pengintegrasian sensor untuk mengumpulkan data penting, pelaksanaan sambungan untuk pemantauan masa nyata, dan pembangunan algoritma pengoptimuman tenaga untuk memanjangkan hayat bateri, dengan itu meningkatkan kecekapan sistem perkongsian e-basikal. Hasil kajian menunjukkan peningkatan kecekapan tenaga sebanyak 15%, yang meningkatkan status pengecasan bateri (SOC) dan memanjangkan jarak operasi. Selain itu, sistem ini membolehkan pengurusan armada yang lebih baik dengan memastikan

https://doi.org/10.31436/iiumej.v26i2.3448

penggunaan tenaga dan ketersediaan e-basikal yang optimum untuk pengguna. Pendekatan ini selaras sepenuhnya dengan matlamat pembangunan mampan (SDG) dengan mempromosikan pengangkutan mesra alam dan meningkatkan aksesibiliti pengguna. Integrasi teknologi IoT terbukti berkesan dalam mengatasi kelemahan sistem tradisional, menawarkan penyelesaian berskala dan cekap untuk mobiliti bandar moden.

KEYWORDS: *E-bike, E-bike sharing systems, Energy optimization, Localization monitoring, Smart cities.*

1. INTRODUCTION

Electric vehicles are a superior option for mobility for several reasons. They often feature simpler engines, which make them easier to service, and are lighter, which makes them more economical, resilient, and energy efficient. They also have better options for mobility. Their only emissions come from the facility that produces electricity. However, the adoption of this technology, which has been around since the creation of the first Ford, is incredibly slow. In response to rising traffic, longer and more intense "smog" seasons, and a lack of parking spaces, cities are striving to urge their citizens to use other modes of transportation, such as the bus, rail, or bicycle. Bicycles are the most practical solution and have the finest health benefits for both the commuter and the city. However, the scarcity and restrictions of e-biking systems, as well as the difficulty of reaching the bicycle itself in a dangerous region, are what make using e-bikes difficult today [1].

Most recent studies primarily focus on electric vehicles rather than e-bikes, and they emphasize the hardware platform of those vehicles, including components such as powertrains and charging systems. However, similar research dedicated to e-bike platforms remains limited. In a related study [2], researchers explored the performance of a dynamic wireless charging system for electric vehicles (EVs) with characterization of different ferrite core geometries. These were to improve power efficiency and overcome issues like air gaps and misalignment between the primary coils, which generally cause inefficiency of energy transfer. The study revealed that air gaps and misalignment are the primary sources of lower power transfer efficiency. At the same time, high battery cost is still a major challenge affecting EV deployments, combined with limited driving range and deterrence to mass adoption caused by the virtual lack of static charging stations for consumers. While this study examined dynamic wireless charging for EVs, it highlights critical principles applicable to energy management with electric vehicles that can also be exploited in solving efficiency-related challenges [3].

Numerous cities, nations, and regions have implemented various regulations over the past ten years to encourage sustainable mobility. Bicycles are widely regarded as a cheap, healthy, and ecologically beneficial means of transportation [4]. Implementing bike-sharing programs (BSS) has been associated with substantial financial gains, improved public health, and time savings. As a result, many cities and local governments support and finance BSSs to promote cycling. However, writers call for more research on the topics because the effectiveness and effects of these new services are not fully understood. Another study done by [5] many cities worldwide introduced new, shared mobility modes, such as dockless electric scooters (e-scooters) and electric bikes (e-bikes). Researchers investigated the potential causes of users' distaste for e-bike battery systems, which deters people from using these systems more frequently. They also discussed the challenge of integrating electric bike sharing systems (BSS) with cargo-friendly solutions.

https://doi.org/10.31436/iiumej.v26i2.3448

Since they are powered, e-bikes provide unique possibilities for tracking usage and comprehending how they interact with urban environments, which may be helpful for both e-bikers and conventional cyclists. It is essential to have a thorough grasp of how e-bikes are utilized in specific geographical and cultural contexts to comprehend and communicate their potential benefits for sustainable transportation and beyond [6]. Electric bike sharing aims to improve the efficiency and comfort of the current public transportation system. Electric bicycles (e-bikes) have become increasingly popular in China since the 2000s, and over the past ten years, they have also gained favor in Asia. Electric bicycles are typically thought to have more range, speed, and overall improved performance than normal bikes. Bicycling is, however, hindered by difficult geography, long travel distances, high temperatures, poor air quality, and other circumstances requiring physical effort. Therefore, "e-bikeshare" offers the ability to make bikeshare more appealing to people who may not have previously considered it a possibility. By removing some of the obstacles to cycling, electrically assisted bikes are faster, enabling longer trips, and are easier to ride over mountainous terrain [7]. It was discovered that perceived simplicity and functionality substantially predicted the desire to use regular BSSs. The e-bike delivers a competitive travel speed compared to rush-hour driving and public transportation.

According to a study done by [8], a significant obstacle to the widespread use of e-bikes has been recognized as the actual and perceived risks associated with cycling in cities and sharing the road with other motorized vehicles. It was determined that the cyclist was in danger from overtaking automobiles, passing pedestrians, road junctions, and parked vehicles in his or her path. Additionally, they demonstrated that collisions with pedestrians, light vehicles, and other bicycles were the most frequent. Moreover, it was noted that heavy traffic, fast traffic, icy, snowy, and debris-filled roads, as well as the perception of risk from other motorized vehicles, were the main obstacles to cycling adaptation. A type of motor vehicle, longitudinal road lines dividing slow traffic or bike lanes, and the stability of a cyclist in controlling the wheel angle and variation in their speed were proposed as solutions to these problems [9]. This investigation used an e-bike called a "Vektron S10" with four power support levels. The bike had three GoPro cameras for front, left, and rear views, five ultrasonic sensors, GPS, an IMU, a potentiometer, and a datalogger. The impact of road-related factors, such as the kind of road, slope, construction, pavement, type of intersections, etc., was examined.

People across the world are paying more and more attention to the Internet of Things. This concept's core tenet is the dominant presence around users of various things or "smart objects," such as Radio-Frequency Identification (RFID) tags, sensors, actuators, mobile phones, etc., that can communicate with one another and work with their neighbors to accomplish shared objectives thanks to special addressing schemes. Thus, the IoT (Internet of Things) and the increased ability to use less expensive, portable pollution sensors have changed the paradigm of particulate measurement [10]. These advancements have made it possible to collect real-time data that can be used in various scenarios. This new sensor family opens new possibilities for applications that directly involve citizens in monitoring air pollution, from which came the idea of "Canarin II" [11]. An electric bike that is linked to the user's smartphone to assist data gathering and sharing.

Its operational characteristics are analyzed based on the effects of essential parameters such as rider mass, wind speed, and slope. Dynamic equations mimic this bicycle's operation under specific operating circumstances. The fundamentals require a power flow controller for e-bikes to control the energy transfer from the battery to the motor. The motor produces electrical power functions in coordination with the user's mechanical power generated when pedaling, putting both in action. The basic scheme of an e-bike proposed in this paper can be

https://doi.org/10.31436/iiumej.v26i2.3448

seen in Fig. 1. Suitable power is chosen for the motor of the electric bicycle based on the required power determined by simulation. A previous study investigated the operating parameters of the electric bike in two scenarios: human power alone and human power assisted by an electric motor, to support the simulation study [12]. Data is gathered using LabVIEW programming. To verify the accuracy of the simulations that are being presented, experimental and simulation findings were compared.

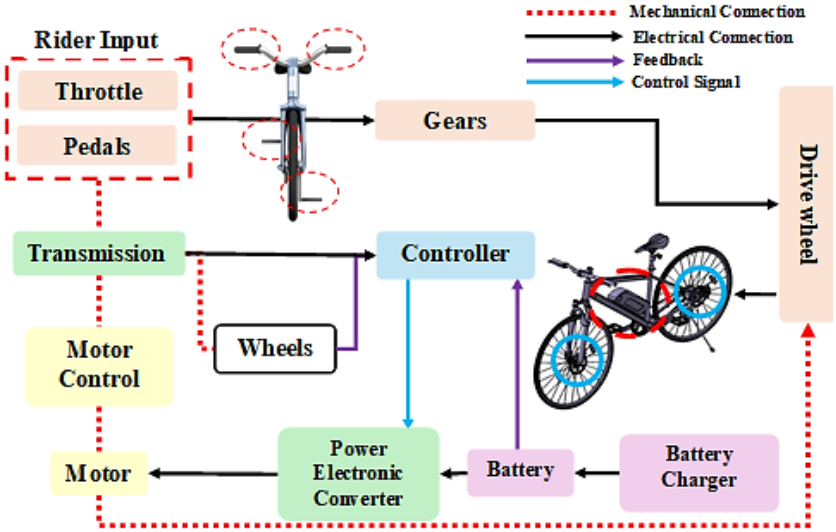


Figure 1. Basic scheme of an e-bike

In 2019, we witnessed the launch of two types of personal light-electronic vehicle sharing systems in the Gdansk, Gdynia, and Sopot metropolitan area (also known as "Tricity"). The first was a public bike sharing system (BSS) called MEVO, which was supported by the metropolitan authorities and had a fleet of 1224 e-bikes [3], [13]. The second category included e-scooter sharing programs offered by three private, rival businesses. No solutions at the time supported freight transfer and were integrated with micro-mobility services. The single BSS introduced in Tricity, the only major city in Europe without traditional bikes, was an electric bike. This gave them a rare opportunity to contrast the actions of users of shared e-bikes and e-scooters in the same setting. Bike ownership, a lack of adequate infrastructure for bike sharing and cycling, adverse weather, rugged terrain, and safety concerns on the road are some of the issues arising from these restrictions. The failure of MEVO can be attributed to several factors. Despite having experience in BSS management, the system's operator did not foresee issues with a somewhat large system that was half free-floating and electric. Implementation and upkeep costs, such as battery charging and fleet rebalancing, surpassed the financial capacity set mainly by the local authorities.

According to [14] it was determined the importance of 23 elements when designing an e-scooter: vehicle size, trunk capacity, seat design, riding range between battery changes, maximum speed, hill-climbing capability, horsepower, handle design, brake function, light brightness, time indicator, tire pressure indicator, constant speed, trip odometer, single trip odometer, battery capacity, revolutions per minute, digital speedometer, average speed, and average electric consumption were design considerations. Electric vehicles powered by batteries still have a lot of problems regarding those from which smart e-bike monitoring system (SEMS) has been developed, as shown in Fig. 2. However, maximizing the driving range of electric vehicles and extending the battery life; addressing the energy needs of electric vehicles both in the short- and long-terms; and increasing regenerative braking energy are the most crucial ones [15].

https://doi.org/10.31436/iiumej.v26i2.3448

Figure 2. SEMS was developed to collect real-time usage and sensor data, combining open-source software and open hardware (Kiefer & Behrendt, 2016) [15]

Recent studies come in the form of an annotated dataset and an application study that made use of Artificial Neural Networks (ANN) to implement a driving behavior control system [16], [17], [18]. This work aimed to compare signal processing methods applied to automatically trigger an ANN-based control system based on sensor data collected in real-time regarding driving behavior parameters related to throttle and braking. And steering influencing a gasoline-powered average vehicle. This system delivered driving comfort by modifying the air-fuel ratio (AFR), enabling fuel economy on par with eco-driving mode and maximum engine power in sporty driving mode. While the findings are limited to gasoline engines and did not quantify factors such as the hill grade effect, they indicate increasing importance in having data feedback suited well for fine-tuning vehicle dynamics.

Similarly, our work targets a Smart IoT Energy Optimization and Localization Monitoring system for e-bike sharing. Though our system is not an ANN, it shares the idea of exploiting real-time sensor data to improve energy efficiency with the prior one. This is part of our larger effort to optimize energy management and promote the scalability and sustainability performance of e-bike sharing systems by tracking different parameters relevant to e-bike handover and battery utilization.

The remainder of this paper provides a detailed method for our E-bike sharing system, "I-bike," and a detailed description and assessment of the energy optimization for power consumption and localization monitoring mapping.

2. METHODOLOGY

The methodology for the smart IoT energy optimization and localization system in e-bike sharing is designed to ensure that the development process for this system is effectively done and whole. This method shall integrate both hardware and software, providing coherence from an early design phase to the final implementation of this system. It commences with a careful evaluation of the project specifications, requirements, and bicycle data route information to ensure that every element tallies to enhance energy efficiency and track the location of e-bikes. The NodeMCU ESP8266 will be used as the core controller of the hardware system; it serves as the central communication point to which multiple sensors and other components, including a GPS module and energy monitoring device, are attached. Energy optimization in an e-bike-sharing system can be defined as minimizing the energy consumption of electric bikes while

https://doi.org/10.31436/iiumej.v26i2.3448

maximizing their availability for use. That could be achieved by enabling hardware-software solutions, which could use data from IoT devices and sensors. This section shows the software and hardware design of the energy monitoring system. Moreover, localization monitoring within an e-bike sharing system involves accurately tracking real-time locations with complete accuracy and oversight. The development of both software and hardware for the localization monitoring system is focused on robust communication module development, efficient data collection strategies, and the utilization of IoT devices and sensors to enable accurate tracking and monitoring. If followed correctly, this systematic approach ensures the system fulfills its main goals while establishing a foundation for further developments toward better energy efficiency and user satisfaction.

2.1. Developing the System

In terms of software, the algorithms developed will optimize energy usage by predicting and adjusting consumption patterns based on the bike's operational parameters and usage patterns. Location tracking is implemented via GPS data synchronized with the IoT cloud for real-time monitoring and data analysis. This IoT platform will ensure that data transfer between the bike and the cloud, including the user interface, happens well and smoothly to attain its intended purpose of accurate tracking and energy optimization. The most critical stage of this development process involves integration and testing, where hardware and software components are integrated to check for compatibility and performance. Various scenarios were simulated to investigate and evaluate how the system responds to various situations, like low battery levels or inconsistent GPS signals. This phase also ensures the system is scalable and integrates perfectly well with the bike-sharing networks and platforms. Throughout the entire process, MATLAB is pivotal, particularly at the analysis and optimization stages. The software is used to formulate the algorithms, simulate systems modeling, and design control systems. The real-time feedback of the MATLAB toolset gives, within the development phase, a scope for data-driven adjustments to be made in the hardware and software parts of the system. In the final stages, the system deploys in a controlled environment to test its performance in real-world cases. The thorough and detailed documentation of all steps taken enables future developments, and the data gained from testing helps to increase scalability and compatibility. Hence, this integration process shall offer a smart, energy-efficient, and user-friendly system to offer a better experience in e-bike sharing systems.

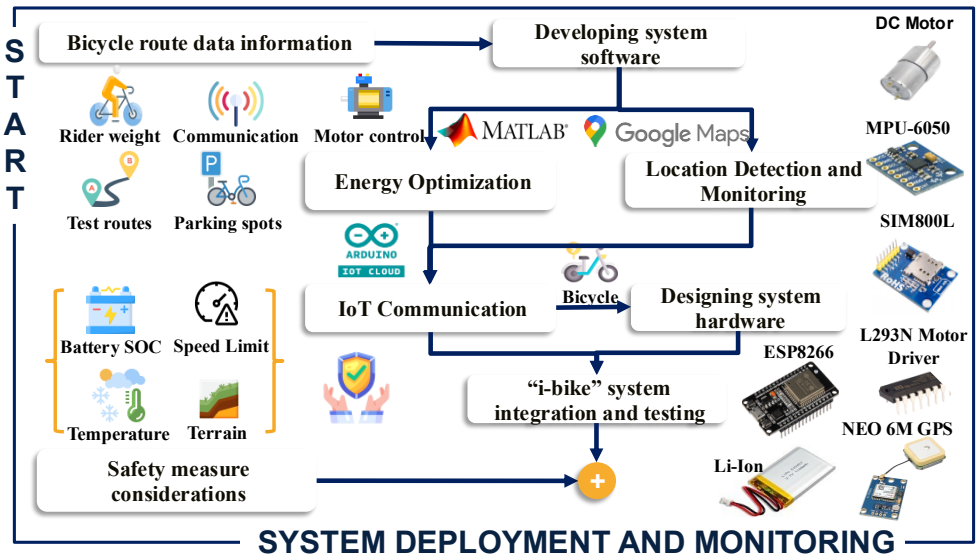


Figure 3. I-bike system monitoring block diagram

https://doi.org/10.31436/iiumej.v26i2.3448

Fig. 3 illustrates the methodology for developing and deploying a smart IoT energy optimization and localization system for e-bike sharing, whereby it starts with collecting bicycle route data from sensors and GPS to locate and track the movement and speed of the bicycle on the travel route. The software for the system is designed to handle and manage real-time communication of this data between the e-bike, the cloud, and user interfaces. Operational data is analyzed to optimize energy consumption and utilization, while the GPS-based e-bike detection and monitoring system enables the precise tracking of locations. The IoT communication stage integrates all components, facilitating seamless data transfer and ensuring uninterrupted data transmission. The hardware circuit setup consists of key elements such as NodeMCU ESP8266 (central communication hub), DC motor, Li-Ion battery, MPU-6050 (gyroscope/accelerometer) for motion sensing, SIM800L GSM module for mobile communication, L293N motor driver for motors, and NEO 6M GPS for tracking location. These components have been selected to meet the system requirements of durability, low power usage, and real-time communication needs. Additionally, increasing rider safety is considered by implementing safety measures, such as monitoring environmental conditions. The system is then deployed in a real-world environment, monitored continuously to ensure seamless functionality and readiness for future expansion.

2.2. E-Bike Sharing System Workflow

An e-bike sharing system usually consists of users locating bikes using a mobile app, signing up to access the bike, and organizing their trip. The alarm system, a vital part of the intelligent IoT energy optimization and localization system for e-bike sharing, includes various important hardware and software components. The hardware components consist of motion sensors that identify unauthorized movements or attempts to tamper. These sensors link to a microcontroller NodeMCU, which handles the sensor data and activates the alarm (buzzer) when needed. Furthermore, the system could incorporate LED indicators or blinking lights for visual notifications. The software component needs programming of the NodeMCU to trigger planned responses like sounding an alarm and sending alerts in response to unauthorized access. Moreover, the system can connect with a central control panel or dashboard to control the alarm system's condition. By integrating these parts, the alarm system increases security and discourages theft in the e-bike sharing system. It includes projected battery level, energy efficiency choices, and route tracking. Users can ride a bike to their destination or drop it off at a different station. An E-Bike Sharing System Workflow is further illustrated in Fig. 4.

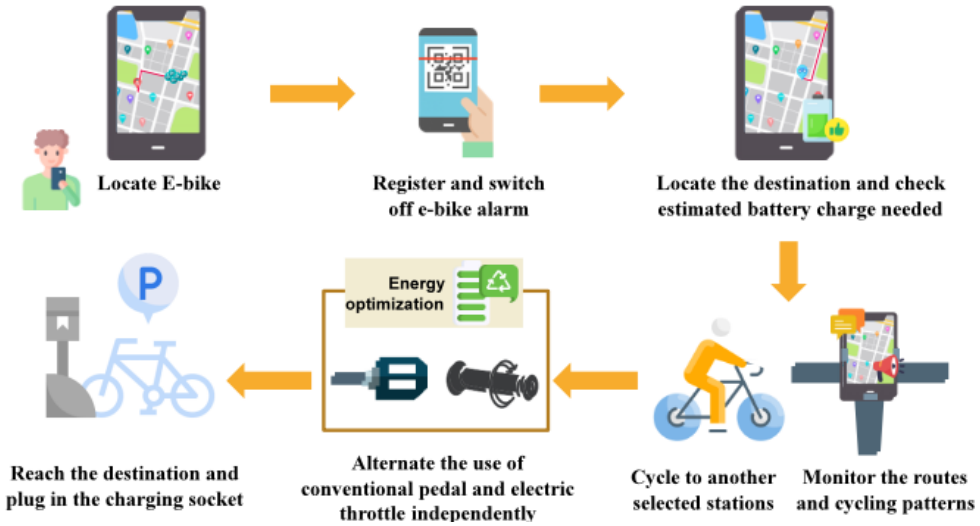


Figure 4. I-Bike Sharing System Workflow

https://doi.org/10.31436/iiumej.v26i2.3448

2.3. IoT Integration for E-Bike Sharing System

An IoT-based energy optimization system can be seamlessly integrated with a localization monitoring system to enhance the overall efficiency and user experience of an e-bike sharing service. By leveraging data collected from sensors and IoT devices, the system can dynamically adjust the e-bike's performance based on real-time factors such as location, battery level, and rider behavior. To achieve this integration, a robust cloud platform like Arduino IoT Cloud is essential for collecting, processing, and storing large volumes of data generated by the e-bikes. Developing a user-friendly mobile application enables end-users to interact with the system and access real-time information. An overall IoT communication diagram is shown in Fig. 5. Frameworks like React Native, Flutter, or Ionic can be employed to create efficient and intuitive mobile interfaces. To ensure the security and privacy of data transmitted between the e-bikes, IoT gateway, and cloud platform, cryptographic libraries such as OpenSSL or Mbed TLS can be implemented. These libraries provide robust security measures to protect sensitive information and prevent unauthorized access.

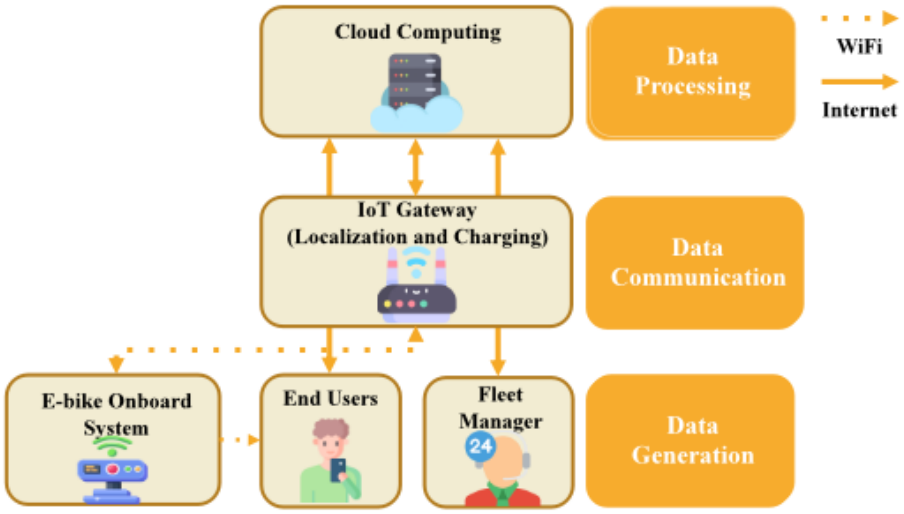


Figure 5. IoT communication diagram

2.4. IoT Connection Flowchart for E-Bike Sharing System

The flowchart shown in Fig. 6 depicts the basic process to be performed by an IoT-enabled e-bike-sharing system for efficient tasks such as battery management, precise localization, and seamless user experience. The working process of the IoT-based e-bike sharing system starts when the user enters the system's dashboard by logging in. In case of a successful login, it triggers the alarm system to be turned off. Once the alarm system has been turned off, the e-bike gets unlocked. The power monitoring panel is checked to assess the charge level in the batteries. If the battery level falls below a predefined threshold level X=20%, it notifies the user to start the bike charging process. Once the battery is sufficiently charged, the system checks the localization monitoring dashboard to know the e-bike destination. Then it continuously monitors battery consumption, adjusting the speed of the motor accordingly. If the gyroscope detects lower acceleration due to climbing uphill, it increases the battery consumption and raises the speed of the motor to maintain optimal speed.

https://doi.org/10.31436/iiumej.v26i2.3448

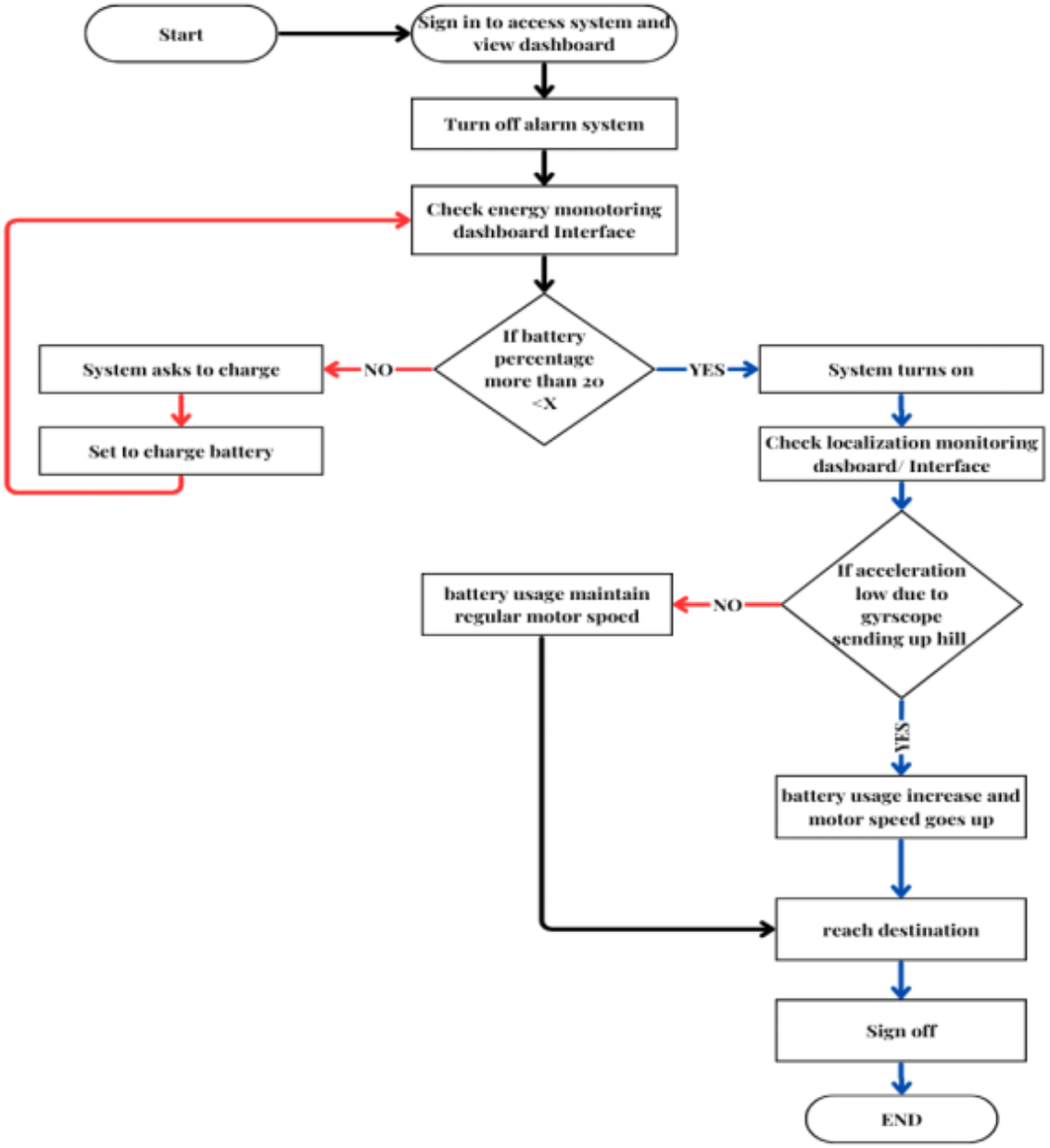


Figure 6. IoT Connection Flowchart for E-Bike Sharing System

2.5. Bicycle Dynamics

The motion of a bicycle follows Newton's second law, which is described by:

$$F_P - (F_R - F_S + F_W) = M \frac{d^2y}{dt^2} \quad (1)$$

Where, F_R is the rolling resistance force, F_P is the propulsion force, F_S is slope resistance force, F_W is the wind resistance force, M is the total mass of the bicycle $M_b$and rider M_R, and x is distance (m). This study uses a DC motor installed in the rear wheel to support riders during pedaling. The dynamics of the DC motor are described by Eq. (2) and Eq. (3):

$$L_a.\frac{di_a}{dt} + i_a(t).R_a + K_b\,\omega_m = U_a \quad (2)$$

$$J.\frac{d\omega_m}{dt} + T_l + B_1\,\omega_m = K_b\,i_a(t) \quad (3)$$

https://doi.org/10.31436/iiumej.v26i2.3448

where, i_a is the armature current, R_a is the armature resistance, U_a is the terminal voltage of the DC motor, L_a is the armature inductance, K_b is the back emf constant, J is torque of inertia, B_1 is the viscous friction coefficient, T_l is the load torque, ω_m is the speed motor.

2.6. Bicycle Performance

The power consumed to push a bicycle and rider is primarily due to overcoming air resistance, slope, and friction, as described by Eq. (4) and Eq. (5):

$$P_{total} = P_{drag} + P_{hill} + P_{friction} \quad (4)$$

$$P_{drag} = \frac{C_d \,.\, D \,.A}{2} . (v_g + v_w)^2 . v_g \quad (5)$$

where C_d is the drag coefficient,D is the density of air, A is the frontal area, v_g is ground speed, v_w is wind speed. The power to overcome the slope P_{hill} is calculated using Eq. (6):

$$P_{hill} = 9.81 \,.\, G \,.\, v_g \,.\, M \quad (6)$$

where G is the coefficient of slope. M is the total mass of the bicycle and rider. The power to overcome the friction $P_{friction}$ is calculated using Eq. (7):

$$P_{friction} = 9.81 \,.\, R_c \,.\, v_g \,.\, M \quad (7)$$

where R_c is the coefficient of rolling resistance.

The motion of the e-bike is governed by Newton's second law, considering forces such as rolling resistance, slope resistance, and wind resistance. The power consumed by the e-bike is calculated using Eq. (8):

$$P_{total} = P_{drag} + P_{hill} + P_{friction} \quad (8)$$

where P_{drag} is the power to overcome air resistance, P_{hill} is the power to overcome slope, and $P_{friction}$ is the power to overcome rolling resistance. These calculations are integrated into the energy optimization algorithm to adjust motor output dynamically. For instance, on steep slopes, the system increases motor torque to compensate for increased resistance, while on flat terrains, it reduces power output to conserve energy. This approach ensures efficient energy usage and contributes to the overall improvement in electrical efficiency.

2.7. Gyroscope-Based Energy Optimization Strategies

The IoT-based smart energy optimization and localization system in e-bike sharing involves many advanced hardware-software integrations to help improve the general riding experience and increase energy efficiency. Gyroscopes, integrated with other sensors such as accelerometers, play a crucial role. These updated movement and orientation data are available in real time, thus offering accurate power management tuning. For instance, when the system notices any shifts in speed or direction, it can optimize the motor output accordingly. Such instantaneous adjustments would be critical in keeping energy efficiency uphill or around sharp turns. In the e-bike system, NodeMCU provides the board that controls all functionalities of several other components integrated into the system for proper functionality. The GPS module communicates with NodeMCU through a serial port and will provide the current location in real time. This will be important in tracking the position of an e-bike and mapping the routes taken. The Wi-Fi module enables the NodeMCU to send location, speed, and battery status data to any remote server or cloud services for continuous monitoring and analysis. Also, integrating an OLED 0.96" I2C display gives the rider all the essential current data: speed and battery SOC (State of Charge), so he is always kept informed about his journey.

https://doi.org/10.31436/iiumej.v26i2.3448

The circuit in Fig. 7 has two major parts: the Localization Monitoring System and the Energy Optimization System. In the Localization Monitoring System, the SIM800L GSM module sends all its data to a remote cloud service to track the real-time position. More precisely, it does geographical tracking using the NEO 6M GPS, and for object detection or tampering, it uses an HC-SR04 ultrasonic sensor. Once tampering is detected, a buzzer triggers an alert. Energy Optimization System targets power management and motion control. In the NodeMCU ESP8266, the core coordination among sensors, communication modules, and cloud services is done. Motor speed and torque are managed to optimize energy consumption during operation using an L293N motor driver. This, along with the MPU-6050 sensor that monitors the orientation and movement of the bicycle, allows the system to make real-time adjustments to improve performance while being energy-efficient. The power supply is guaranteed through a rechargeable Li-Ion battery, the consumption of which is monitored very carefully to extend its SOC while sustaining the performance of the system. All these components combine into one robust, energy-efficient system to enhance riders' experience with real-time data displays, optimize energy usage, and ensure safety and efficiency. This system improves not only the aspect of localization and communication but also optimizes energy consumption to enhance performance and sustainability for better e-bike sharing.

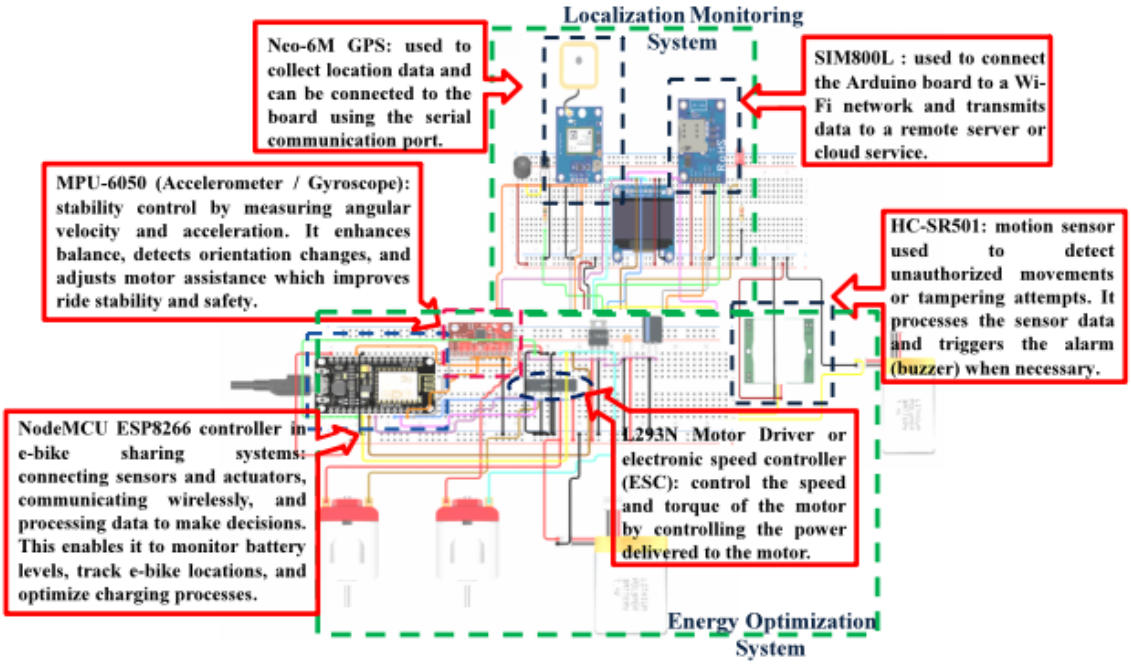


Figure 7. Developed a circuit for energy optimization and localization monitoring systems

2.8. Utilizing MATLAB And Arduino Cloud / IoT for Enhanced Data Processing And Control In E-Bike System Development

Most aspects of the e-bike system were developed and implemented using the powerful and versatile framework of MATLAB/Simulink, specifically Simscape Electrical. Because of its rich toolkit and features regarding data processing, algorithm development, and system modeling, we could analyze and process sensor data efficiently from different components, like the gyroscope of the e-bike and other sensors. All the libraries for processing and analyzing signals and data helped us extract meaningful information from raw sensor data, enabling correct tracking and control of the motion of the e-bike. Its intuitive graphical user interface and simulation capability allowed us to model and fine-tune parameters of the control system

for optimal performance and stability of the e-bike. Furthermore, MATLAB is compatible with external hardware and different communication protocols, facilitating seamless integration of the e-bike electronic system for real-time data acquisition and control. Arduino Cloud and IoT enhanced the ability to connect with outside devices, elevating the system's performance, adaptability, and user satisfaction.

The system optimizes battery usage through real-time monitoring and adaptive energy management. Sensors such as the MPU-6050 gyroscope and GPS module provide data on terrain, rider behavior, and speed. This information is fed into an energy optimization algorithm that adjusts motor output to minimize energy consumption. For instance, during uphill climbs, the system increases motor assistance, while on flat terrains, it reduces power output to conserve energy. Additionally, the system employs predictive maintenance by monitoring battery voltage and charge levels, ensuring optimal performance.

In addition to MATLAB, a mobile application and a dashboard user interface were made using Arduino Cloud and Arduino IoT. These platforms make it easy to create a responsive interface that enables users to monitor key e-bike metrics such as location, speed, and battery status. In this regard, Arduino IoT has ensured seamless connectivity of an e-bike with the cloud over real-time data transmission and its control on mobile devices.

3. RESULTS AND DISCUSSION

We start by describing the performance metrics adopted for measuring the system's performance concerning energy efficiency, localization accuracy, and user satisfaction. The obtained results are then presented in detail, together with discussions. The analysis will be conducted from multiple dimensions: effectiveness for energy optimization, localization system accuracy, and influence of user behavior on energy consumption. Furthermore, the segment delves into incorporating gyroscope data for energy optimization, highlighting the improved energy efficiency. The results are carefully examined and critically analyzed for their implications on scalability, sustainability, and overall accomplishment of the e-bike sharing program.

3.1. Integration of Gyroscope and Feedback System for Enhanced Monitoring

First of all, the information from the gyroscope can be used to optimize the operation of the motor. In this case, angular movement and orientation are controlled all the time, which enables the system to make proper adjustments in the operation of the motor by adjusting the motor's power output accordingly. For example, in uphill or heavy terrain conditions, the system is informed of how much assistance the motor has to give. Optimized motor performance decreases the rider's fatigue and enhances comfort during riding. The input from the gyroscope would contribute to optimizing energy and prolonging battery SOC.

Tracing the motion and orientation of an e-bike precisely would intelligently adjust the motor's power consumption according to the nature of the ground and riding conditions. For example, when the gyroscope detects downhill or flat surfaces—downhill being the maximum condition where there is less need for a motor assistance system—it reduces the power output, thus conserving battery energy. In particular, energy optimization enables an increase in e-bike range by enhancing the overall efficiency of battery use. We implemented algorithms in MATLAB that processed and analyzed data gathered from several sensors. In turn, the gyroscope provides information about the angular motion and orientation of the e-bike, which is essential for accurate tracking and control. We will also integrate a feedback control system into the MATLAB model. This will continuously monitor and adjust several parameters, like

https://doi.org/10.31436/iiumej.v26i2.3448

motor power output, based on real-time sensor data. The e-bike uses a feedback control system for efficient and optimal operation under various riding conditions. Incorporating these sensor inputs into the operational model developed in Fig. 8 has greatly enhanced the stability, control, and energy efficiency of the e-bike. The gyroscope was incorporated with a feedback control system to improve these results with more precise and responsive control.

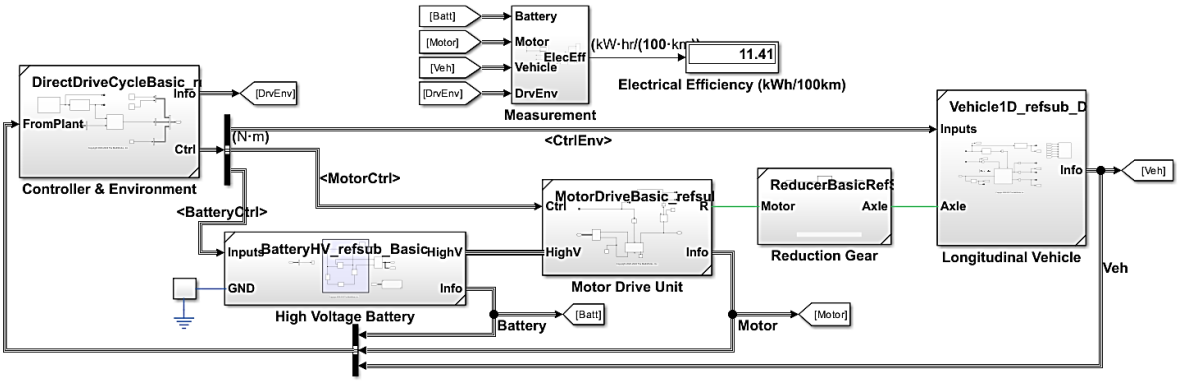


Figure 8. E-bike MATLAB/Simulink new control system with feedback

3.2. Efficiency Calculation and Gyroscope Integration

Calculating electrical efficiency within an electrical drive system, such as a three-phase AC multi-drive, requires detailed behavioral simulation of components like a motor, controller, and inverter. In this respect, Simscape Electrical is one of the best tools for modeling and designing the system to appropriately estimate electrical losses, which is required to calculate efficiency. First, a detailed three-phase permanent synchronous motor drive (PMSM) system, including elements such as the motor, controller, and inverter, is set up. Efficiency will be calculated by setting some main operating conditions, including speed and torque. Significant operating factors influence how the motor will consume and dissipate electrical power. Electrical losses can be measured when running the system under the set operating conditions.

Once the simulations are done, the API (Application Programming Interface) called upon is the Get Power Loss Summary API to extract the electrical losses from the system. The electrical efficiency is calculated by comparing the input power at the system input to its output power, where the difference is the electrical losses. It is a model for calculation efficiency under different conditions. The general rule is that electrical efficiency in such systems can vary from 11% to 15%, as shown previously in Fig. 8, depending on factors such as operating load and system parameters. The integration of extra sensors, such as the gyroscope, increases the completeness of the model in all respects. The system dynamically changes its motor output based on real-time data from the gyroscope on the orientation and movement of the bike. For example, when the gyroscope detects angle changes in the bike while making climbs or sharp turns, the system can adjust motor torque and speed to maintain optimal performance while avoiding unnecessary energy consumption. This would, in return, work to enhance the energy efficiency of the whole e-bike system.

Besides, the results can be used in higher-scale models, such as the simplified PMSM drive block, where the estimated electrical losses would serve as parameters for more efficient simulations. Employing models with detailed loss information, the system will give a more accurate prediction of energy performance in various conditions and improve power efficiency while enhancing the rider's experience. The Simscape Electrical-enabled workflow captures the system's complexities, allowing efficient solutions for energy optimization.

https://doi.org/10.31436/iiumej.v26i2.3448

From the three graphs plotted in Fig. 9, very useful information can be acquired about energy consumption by the vehicle, performance by the motor, and the efficiency in general while it operates. The graph shows three time plots of an electric vehicle's performance: the top graph shows the speed of the vehicle, which rapidly accelerates from zero to a maximum of 40 MPH and then levels off; the middle graph shows battery voltage decreases gradually as the vehicle is in use but with a slight dip around 150 seconds. The graph below depicts the development of motor temperature, constantly growing during the operation of the motor, and then leveling off at around 150 seconds.

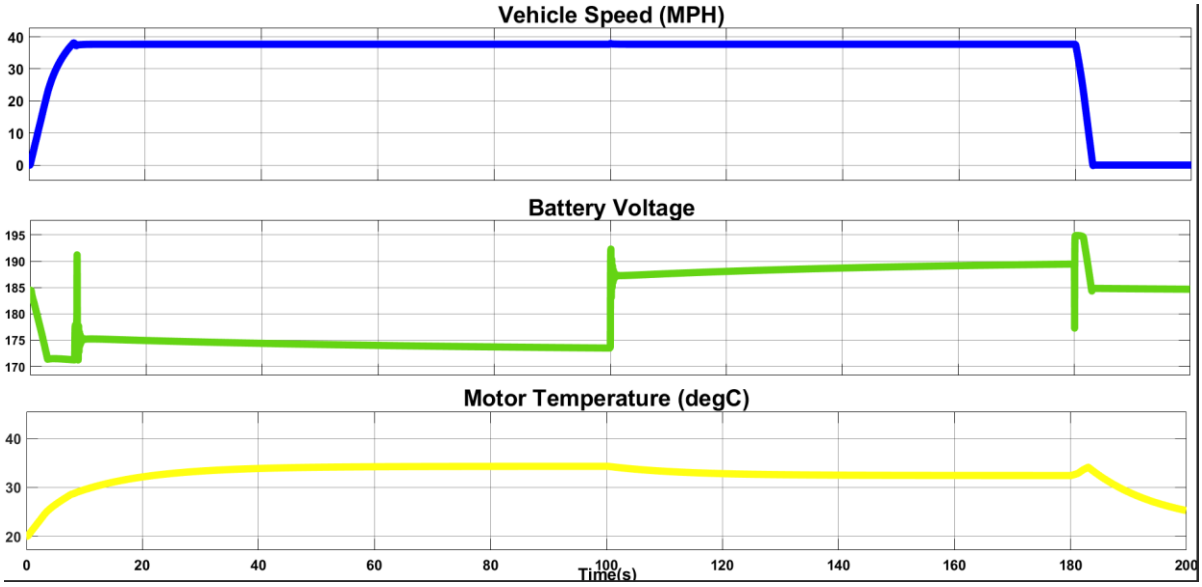


Figure 9. Performance Analysis of e-bike

The graph in Fig. 10 illustrates the mechanical and electrical power consumption of an e-bike as a function of time. Mechanical power output is represented in red and wildly oscillates due to changes in both load and speed. Electrical power input, in blue, has a similar shape but with deviation due to a combination of factors, including the battery voltage and motor efficiency. Energy efficiency can be achieved through the analysis of the relationship that exists between mechanical and electrical power. For example, suppose the mechanical power output is high. In that case, it will also be expected that the electrical power input is equally high because of the increased load the motor has to drive. However, if the electrical power input is way higher than the mechanical power output, energy is lost through some inefficiencies in the system. The differential value of mechanical and electrical power should be minimized to improve the vehicle's general efficiency. It can be done by adopting specific strategies, like optimization in the motor control algorithm, reduction in the electrical losses within the system, and enhancement in the efficiency of energy storing and converting components.

https://doi.org/10.31436/iiumej.v26i2.3448

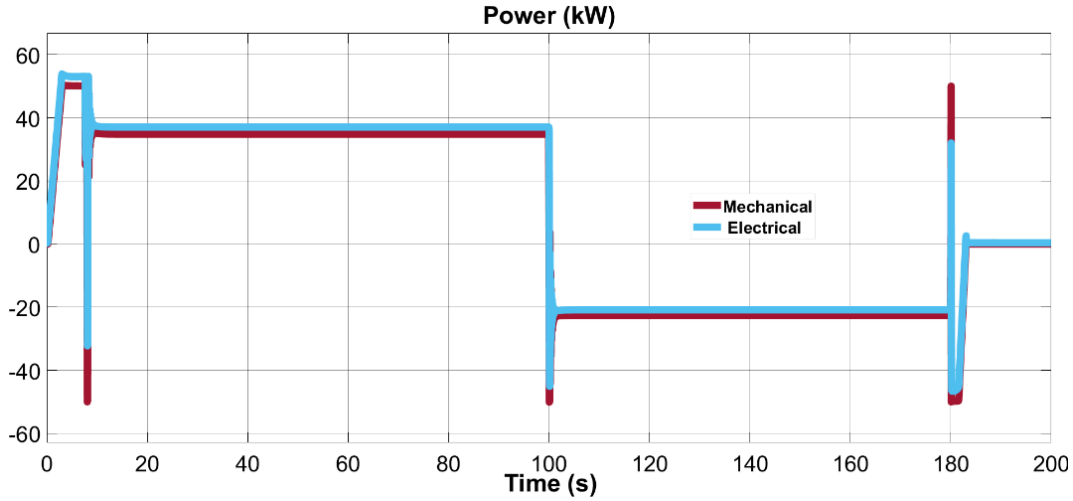


Figure 10. Power Consumption Analysis of the e-bike

Electrical efficiency was improved by integrating real-time sensor data with advanced motor control algorithms. The system dynamically adjusts motor output based on terrain, rider behavior, and battery status. The electrical efficiency is calculated using Eq. (9):

$$Elec_{Eff} = \frac{BattPwr}{VehSpd} \times 100 \tag{9}$$

where $BattPwr$ is the battery power in kWh, and $VehSpd$ is the vehicle speed in km/h. This adaptive approach minimizes energy wastage and maximizes energy conversion efficiency. Experimental results confirmed the efficiency gain by comparing energy consumption before and after optimization under identical riding conditions. Fig. 11 shows the MATLAB simulation subsystem used to calculate battery power ($BattPwr$), vehicle speed ($VehSpd$), and electrical efficiency ($Elec_{Eff}$). The subsystem integrates data from the battery and motor to compute energy consumption and optimize efficiency in real-time.

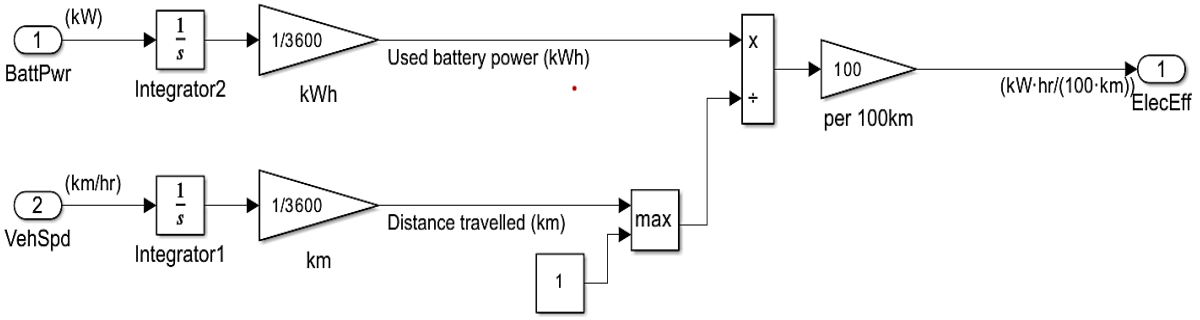


Figure 11. MATLAB simulation subsystem for electrical efficiency

The following graphs in Fig. 12 represent the battery charge and electrical efficiency of an e-bike with respect to time. The graph of battery charge shows that the charge progressively deteriorates as the usage of the electric vehicle increases, but it tends to increase around 1500 seconds. The electrical efficiency graph represents oscillations, although the trend is generally rising with the increase in the time axis. This improvement might be due to various reasons related to tuning the motor control algorithms or because the system has been adapted to a driving condition. The highs and lows of fluctuations in the graph depict how much load was applied to the motor and, consequently, the energy conversion efficiency. These graphs, in conjunction with system simulation results, show the location of optimizations that could be

https://doi.org/10.31436/iiumej.v26i2.3448

made by enhancing motor control algorithms or employing better energy management techniques. This could be a data-driven approach to improving electric vehicles' performance and energy efficiency for extended range while reducing their environmental impact.

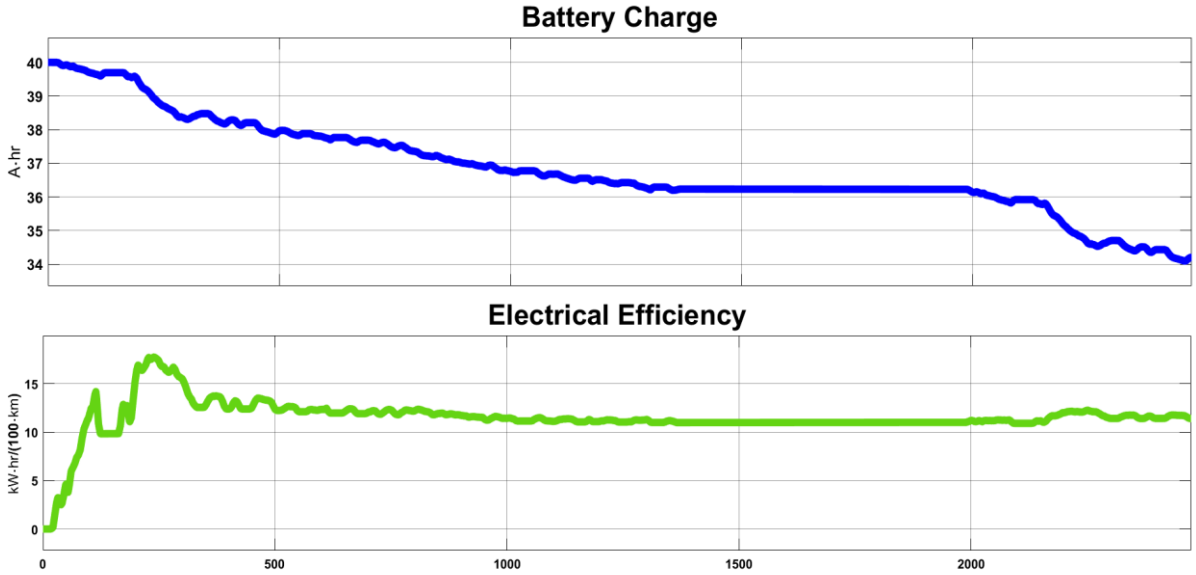


Figure 12. Battery Charge and Electrical Efficiency Analysis

3.3. Prototype Setup

The developed electrical efficiency of the e-bike system is achieved through an optimized control algorithm that intelligently manages power delivery to the motor. The system adapts to various factors, including speed, load, and terrain, for minimal energy dissipation while ensuring the maximum possible conversion of electrical energy into mechanical propulsion. Sensors like the MPU6050 accelerometer/gyro, GPS, current, voltage, and charge allow it to gain real-time operating data and make better decisions. It will precisely enable power output control, making terrain-based adjustments and performing predictive maintenance to ensure the e-bike operates most efficiently. The energy sensor calculates the battery's remaining life and regulates performance for sustained battery management. It also consists of a remote control system that connects to the user's mobile application for conditions and battery locations, among others.

The smart IoT energy optimization and localization system prototype integrated on e-bikes supports the implementation of innovative technology. The hardware will include a carefully chosen array of components comprising a microcontroller board, a GPS module, a gyroscope, and other sensors. These components have been mounted in a case on the collected bicycles, as shown in Fig. 13, offering stability and functionality during operation.

https://doi.org/10.31436/iiumej.v26i2.3448

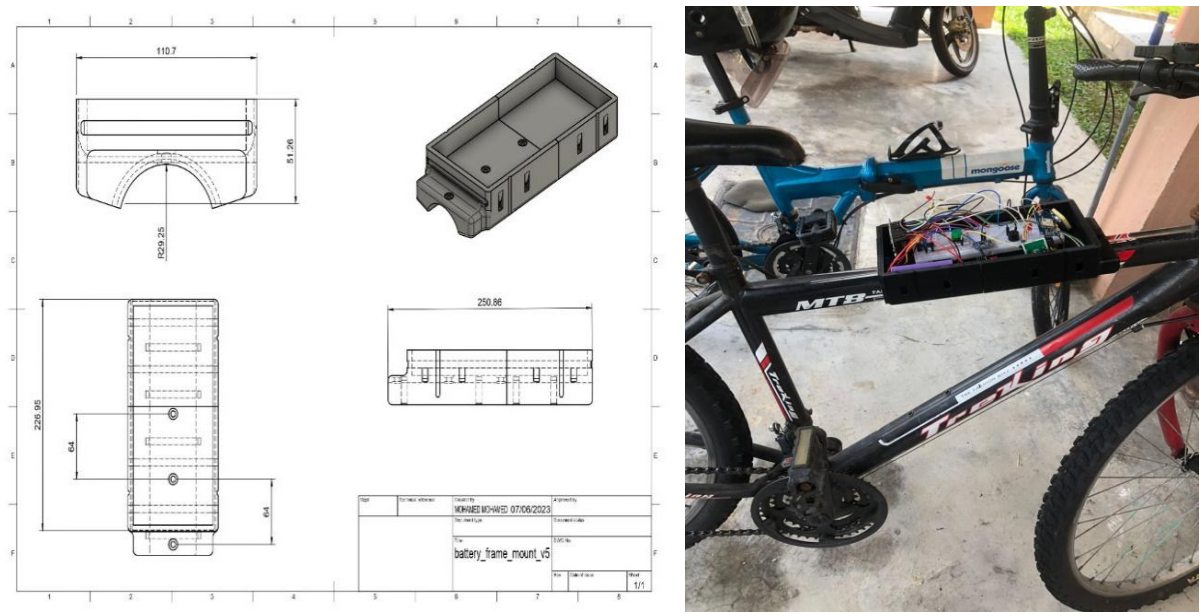


Figure 13. 3D case prototype

It has been elaborately undertaken to realize the best connectivity with optimum placement of hardware for the detailed design and integration of the prototype. As an integrated result, a system that combines the latest technology and the practicality of bicycles is developed. Fig. 14 shows images of the prototype in action and the appearance of the integrated components and bicycles. These views help better present the prototype and how it would look in real life within the e-bike-sharing system.

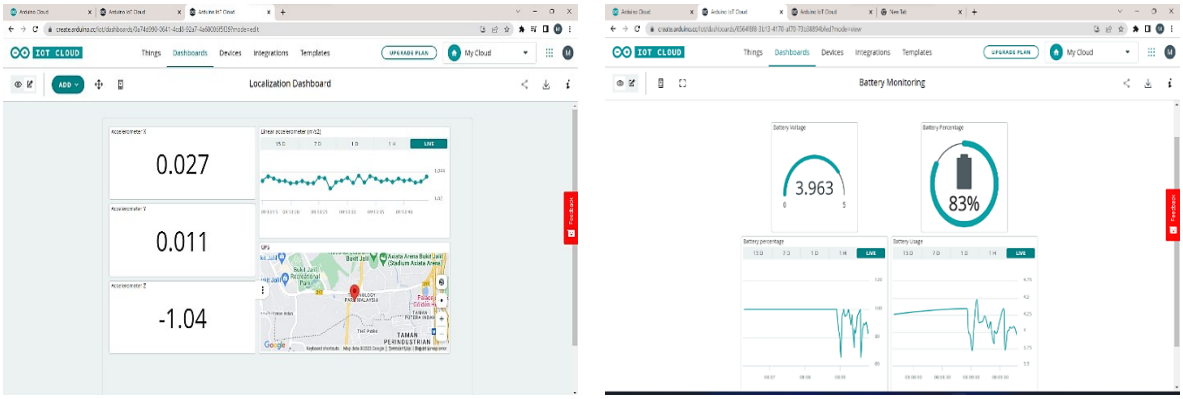
Figure 14. Prototype Integration - Showcasing the smart IoT system seamlessly integrated onto a bicycle

3.4. User Dashboard / Monitoring Interface

The images below, shown in Fig. 15, present a user interface developed on Arduino IoT Cloud that allows the opportunity to show two types of dashboards designed for e-bike localization and battery monitoring. The Localization Dashboard (a) presents real-time data from various sensors: three-axis acceleration values (X, Y, Z) and a GPS map with the current position of the e-bike. This gives users insights into the bike's position and movement dynamics. The Battery Monitoring Dashboard (b) presents key battery metrics such as voltage and charge percentage. The visual gauges display the battery's current voltage (3.963V) and charge level (83%).

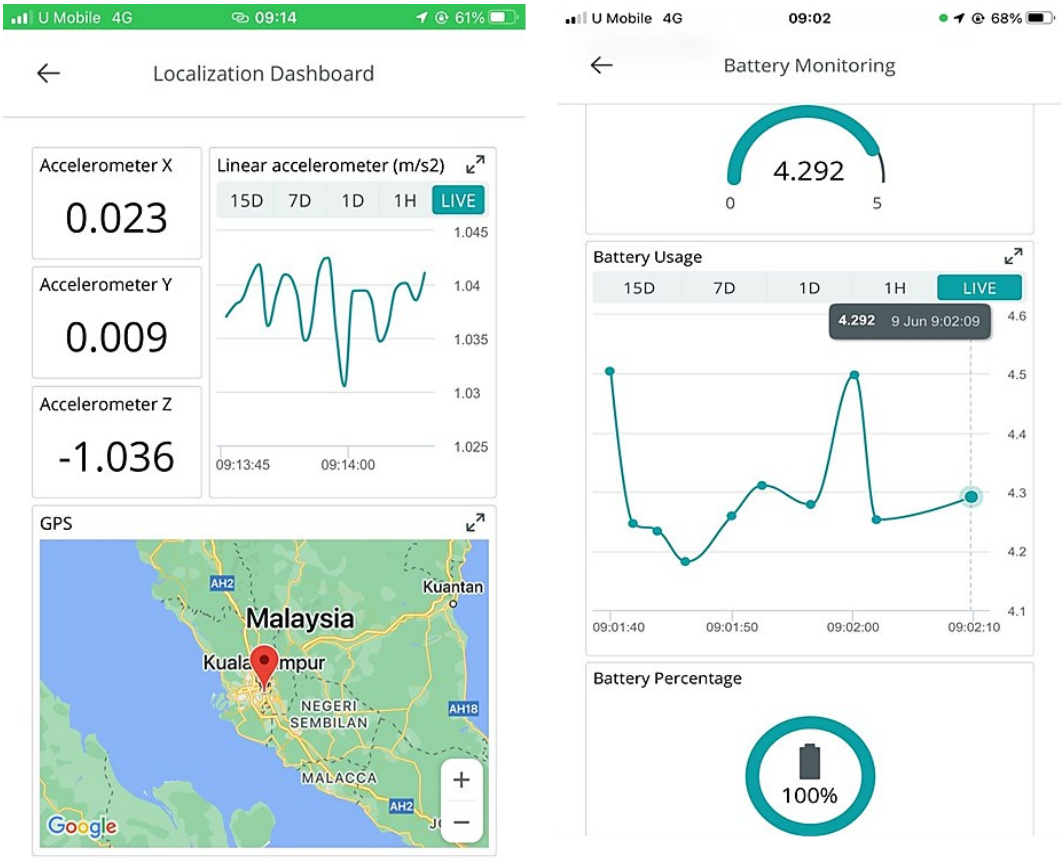
https://doi.org/10.31436/iiumej.v26i2.3448

The two graphs plot the battery percentage versus time and voltage over time in detail to track battery status and performance trends in the e-bike. It integrates sensor data with real-time monitoring for effective energy management and enhances usability through mobile-friendly interfacing. Accordingly, Fig. 16 reports the interface of the mobile application in different riding conditions, directly mirroring the localization and battery monitoring dashboards from the user's smartphone.



(a) Localization - Acceleration (b) Battery SOC Monitoring

Figure 15. Desktop Dashboard



(a) Localization - Acceleration (b) Battery SOC Monitoring

Figure 16. Mobile App Dashboard

4. CONCLUSION

In conclusion, the aims mentioned in the design of the communication module strategy for location and speed data gathering, energy profile information evaluation with travel distance optimization, and IoT-based energy optimization system development for e-bike sharing localization monitoring were met by this project.

We implemented and strategized a bright approach in embedding a communication module that would share real-time data on the location and speed of the e-bikes through careful planning and implementation. This was a good backbone for the subsequent energy profile evaluation. We have analyzed energy profiles, optimized the travel distance to find areas of improvement, and implemented intelligent algorithms for optimizing energy consumption for an e-bike sharing system. The optimization ensured efficiency in using the resources and extending battery life for overall sustainability and operational efficiency. Thus, IoT-based energy optimization of e-bike sharing began to be developed for localization monitoring, marking a significant milestone. The advanced technologies of GPS, gyroscope, and other sensors were merged into the e-bikes to perform real-time monitoring and control. Hence, it provided accurate localization, which permitted effective management and optimization of the bike fleet. This would, in turn, help achieve the project's outcomes regarding SDG 11: Sustainable Transportation and SDG 7: Affordable Clean Energy, while promoting innovation (SDG 9) for a greener future.

The project effectively achieved its goals by creating a communication module plan, enhancing energy profiles, and creating an energy optimization system based on the Internet of Things. The results support the progress of energy-saving e-bike sharing systems, encouraging eco-friendly transportation, and improving overall user satisfaction. The successful implementation of the project showcases how IoT technologies have the potential to enhance energy usage and efficiency in shared transportation systems.

ACKNOWLEDGEMENT

This work was supported in part by the International Islamic University Malaysia under Research Grant “Optimal Deceleration Formulation Based on Reinforcement Deep Learning for Energy Generation from Regenerative Braking Mode of an Electric Motorcycle” FRGS/1/2023/TK02/UIAM/01/1, and in part by IIUM under the IIUM ENGINEERING MERIT SCHOLARSHIP 2023.

REFERENCES

[1] V. Albuquerque, M. S. Dias, and F. Bacao, “Machine learning approaches to bike-sharing systems: A systematic literature review,” *ISPRS Int J Geoinf*, vol. 10, no. 2, Feb. 2021, doi: 10.3390/ijgi10020062.

[2] S. H. Yusoff, A. A. Abdullah, N. N. Nanda, and A. S. A. Badawi, “PERFORMANCE ANALYSIS ON DYNAMIC WIRELESS CHARGING FOR ELECTRIC VEHICLE USING FERRITE CORE,” *IIUM Engineering Journal*, vol. 23, no. 1, pp. 46–59, 2022, doi: 10.31436/IIUMEJ.V23I1.1663.

[3] T. Bieliński, A. Kwapisz, and A. Ważna, “Electric bike-sharing services mode substitution for driving, public transit, and cycling,” *Transp Res D Transp Environ*, vol. 96, Jul. 2021, doi: 10.1016/j.trd.2021.102883.

[4] L. Stilo, D. Segura-Velandia, H. Lugo, P. P. Conway, and A. A. West, “Electric bicycles, next generation low carbon transport systems: A survey,” *Transp Res Interdiscip Perspect*, vol. 10, Jun. 2021, doi: 10.1016/j.trip.2021.100347.

[5] T. Bieliński and A. Ważna, "Electric scooter sharing and bike sharing user behaviour and characteristics," *Sustainability (Switzerland)*, vol. 12, no. 22, pp. 1–13, Nov. 2020, doi: 10.3390/su12229640.

[6] Shen, W., & Li, J. (2018). Design and Implementation of Parking Monitoring System of Bike Sharing In the Electronic Barrier. *Proceedings of the 2018 Joint International Advanced Engineering and Technology Research Conference (JIAET 2018)*. 2018 Joint International Advanced Engineering and Technology Research Conference (JIAET 2018), Xi'an, China. https://doi.org/10.2991/jiaet-18.2018.6

[7] N. B. Hung, J. Sung, K. Kim, and O. Lim, "A Simulation and Experimental Study of Operating Characteristics of an Electric Bicycle," in *Energy Procedia*, Elsevier Ltd, 2017, pp. 2512–2517. doi: 10.1016/j.egypro.2017.03.723.

[8] C. H. Hsu, B. S. Liu, S. L. Jhuang, and Y. C. Li, "The study of ergonomic evaluation and critical design factors for electric scooter," in *2016 International Conference on Applied System Innovation, IEEE ICASI 2016*, Institute of Electrical and Electronics Engineers Inc., Aug. 2016. doi: 10.1109/ICASI.2016.7539571.

[9] S. Pejhan, M. Agelin-Chaab, M. Yusuf, and D. Eng, "Analysis of ebike dynamics and cyclists' anxiety levels and interactions with road vehicles that influence safety," *Accid Anal Prev*, vol. 159, Sep. 2021, doi: 10.1016/j.aap.2021.106272.

[10] S. I. . Ao, Len. Gelman, D. W. L. . Hukins, Andrew. Hunter, and Alexander. Korsunsky, *World Congress on Engineering : WCE 2015 : 1-3 July, 2015, Imperial College London, London, U.K.* Newswood Limited : International Association of Engineers, 2015.

[11] *Aguiari, D., Delnevo, G., Monti, L., Ghini, V., Mirri, S., Salomoni, P., Pau, G., Im, M., Tse, R., Ekpanyapong, M., & Battistini, R. (2018). Canarin II: Designing a smart e-bike eco-system. 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), 1–6. https://doi.org/10.1109/CCNC.2018.8319221*

[12] Seria, S., Quintero, V., Espinoza, P. A., Pérez, A., Jaramillo, F., Benavides, M., & Orchard, M. (2017). Energy Management of Electric Bicycles Given a Traveling Elevation Profile. *Annual Conference of the PHM Society*, *9*(1). https://doi.org/10.36001/phmconf.2017.v9i1.2390

[13] D. Makurat and K. Wojewódzka-Król, "PROBLEMS OF IMPLEMENTING THE MEVO METROPOLITAN BICYCLE SYSTEM." [Online]. Available: https://doi.org/10.26881.etil.2019.81.11

[14] W. Deleenheer, L. Jáneš, and A. Jayakumar, "DEVELOPMENT OF AN ELECTRIC BICYCLE FOR A SHARING SYSTEM IN PRAGUE," *Acta Polytech CTU Proc*, vol. 12, p. 24, Dec. 2017, doi: 10.14311/app.2017.12.0024.

[15] C. Kiefer and F. Behrendt, "Smart e-bike monitoring system: Real-time open source and open hardware GPS assistance and sensor data for electrically-assisted bicycles," *IET Intelligent Transport Systems*, vol. 10, no. 2, pp. 79–88, Mar. 2016, doi: 10.1049/iet-its.2014.0251.

[16] A. Triwiyatno, S. Munahar, M. Munadi, and J. D. Setiawan, "APPLICATION OF DRIVING BEHAVIOR CONTROL SYSTEM USING ARTIFICIAL NEURAL NETWORK TO IMPROVE DRIVING COMFORT BY ADJUSTING AIR-TO-FUEL RATIO," *IIUM Engineering Journal*, vol. 24, no. 2, pp. 337–353, 2023, doi: 10.31436/iiumej.v24i2.2781.

[17] G. A. Aranda-Corral, M. A. Rodríguez, I. F. de Viana, and M. I. G. Arenas, "Genetic hybrid optimization of a real bike sharing system," *Mathematics*, vol. 9, no. 18, Sep. 2021, doi: 10.3390/math9182227.

[18] M. Giliberto, F. Arena, and G. Pau, "A fuzzy-based solution for optimized management of energy consumption in e-bikes," *J Wirel Mob Netw Ubiquitous Comput Dependable Appl*, vol. 10, no. 3, pp. 45–64, Sep. 2019, doi: 10.22667/JOWUA.2019.09.30.045.

[19] Al-Ali, A. R., Zualkernan, I., & Aloul, F. (2010). A Mobile GPRS-Sensors Array for Air Pollution Monitoring. *IEEE Sensors Journal*, *10*(10), 1666–1671. https://doi.org/10.1109/JSEN.2010.2045890

[20] MathWorks. (2023). *MATLAB (R2023a)*. The MathWorks, Inc. https://www.mathworks.com