IIUM ENGINEERING JOURNAL

Volume 27 Number 1

January 2026

IIUM Press

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

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

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

IIUM ENGINEERING JOURNAL

Volume 27, Issue 1, January 2026

https:/doi.org/10.31436/iiumej.v27i1

Copyright Notice

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

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

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

The IIUM Engineering Journal follows the open access policy.

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

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

IIUM ENGINEERING JOURNAL EDITORS

CHIEF EDITOR
Ahmad Faris Ismail, IIUM, Malaysia

TECHNICAL EDITOR
Khairul Azami Sidek, IIUM, Malaysia

EXECUTIVE EDITOR
Teddy Surya Gunawan, IIUM, Malaysia

ASSOCIATE EDITOR
Nor Farahidah Za'bah, IIUM, Malaysia

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

MALAY TRANSLATOR
Nurul Arfah Che Mustapha, IIUM, Malaysia

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

AIMS & SCOPE OF IIUMENGINEERING JOURNAL

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

REFEREES' NETWORK

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

SUBMISSION OF A MANUSCRIPT

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

INTERNATIONAL ADVISORY COMMITTEE

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

Published by:

IIUM Press

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

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

ISSN 1511 - 788X

9 771511 788008

IIUM Engineering Journal

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

https://doi.org/10.31436/iiumej.v26i3.4129

Editorial Message Vol. 27 No. 1 2026

It is with great pleasure that we present Volume 27, Number 1 of the IIUM Engineering Journal, an issue that reflects the breadth, depth, and evolving priorities of contemporary engineering research. The papers published in this volume collectively demonstrate how rigorous theoretical foundations, advanced computational techniques, and practical engineering solutions converge to address pressing technological, environmental, and societal challenges.

This issue opens with contributions from Chemical and Biotechnology Engineering, where research on the growth kinetics of Pleurotus pulmonarius advances sustainable biotechnology practices through innovative methods to revive spawn culture. This work highlights how process optimization and biological engineering can contribute to food security and value-added agro-industrial applications.

A substantial portion of this volume is devoted to Electrical, Computer, and Communications Engineering, underscoring the rapid progress in intelligent systems, data-driven modeling, and energy-aware technologies. Several papers explore the integration of artificial intelligence and machine learning into diverse domains, including graph neural networks for large-scale skyline query processing, transformer-based malware detection at the byte level, deep learning with explainability for skin care detection, and advanced sentiment prediction models. Complementing these are applied engineering studies on inertial sensor self-calibration for autonomous underwater vehicles, LoRaWAN-based IoT systems for electric fence monitoring, optimized PID control for energy-efficient induction motors, and vision-transformer-based groundwater-level forecasting for sustainable water-resource management. Together, these works illustrate how AI-enhanced engineering is reshaping decision-making, automation, and system reliability.

In the Materials and Manufacturing Engineering section, the focus shifts to next-generation materials and devices that support sustainability and high-performance applications. Contributions on eco-friendly tin-based perovskite solar cells and graphene–silver hybrid inks provide valuable insights into material characterization, performance optimization, and the development of greener energy and electronic solutions. The Mechanical and Aerospace Engineering contribution in this issue addresses surrogate model-based optimization of unmanned aerial vehicle wings, demonstrating how computational intelligence and response surface methodologies can accelerate design cycles while maintaining aerodynamic efficiency and structural integrity.

Research in Mechatronics and Automation Engineering further enriches this volume, covering a wide spectrum of intelligent and autonomous systems. Papers in this section examine humanoid robot stability on uneven terrain, vehicle identification using YOLO-based deep learning, autonomous underwater vehicles for subaquatic exploration, FSM-based MBIST controller implementation, adaptive energy balance control for solar-powered hydroponics, and advanced medical image segmentation using attention-enhanced neural architectures. Collectively, these studies highlight the increasing role of autonomy, robustness, and sustainability in modern engineered systems.

Taken as a whole, Vol. 27 No. 1 exemplifies the journal's commitment to publishing high-quality, interdisciplinary engineering research that bridges theory and practice. The diversity of topics and methodologies presented here reflects not only the dynamic nature of engineering as a discipline but also its critical role in supporting sustainable development, digital transformation, and technological innovation.

We extend our sincere appreciation to the authors for their valuable contributions, to the reviewers for their rigorous and constructive evaluations, and to the editorial team for their dedication in bringing this issue to publication. We hope that the articles in this volume will stimulate further research, inspire collaboration across disciplines, and serve as a meaningful reference for researchers, practitioners, and policymakers alike.

Prof. Ir. Ts. Dr. Teddy Surya Gunawan

Executive Editor

IIUM Engineering Journal

IIUM ENGINEERING JOURNAL

Volume 27, Issue 1, January 2026
https:/doi.org/10.31436/iiumej.v27i1

Table of Contents

Copyright i
Editorial Board ii
Editorial Message iv

Chemical and Biotechnology Engineering

3842: Growth Kinetics of Grey Oyster Mushroom (Pleurotus pulmonarius) Spawn Culture Revival from Spray Drying Application 1
Zarina Zakaria, Fatihah Aziz, Noor Hasyierah Mohd Salleh, Mohd Sharizan Md Sarip, Shi Fern Chong

Electrical, Computer, and Communications Engineering

3654: Inertial Sensor Self-Calibration Module Using Attitude Heading And Reference System For Autonomous Underwater Vehicle Navigation 13
Abdul Aziz Abd Kadir, Herdawatie Abdul Kadir

3717: GNN-Based Skyline Query Processing for Large-Scale and Incomplete Graphs 27
Hasan Khair Adzman, Raini Hassan, Dini Oktarina Dwi Handayani

3720: Development of Electric Fence Fault Sensing and Monitoring System with LoRaWAN IoT 48
Mohammad Syarifuddin Bin Mohammad Sufian, Hisham Mohamad, M. Azman Zakariya, Aizat Akmal A. Mohamad Beddelee, Salman Saaban, Muhammad Fadlli A. Yazi

3824: Enhanced Beach Photo Translation using Modified Unsupervised GAN with Regularization 81
Kartika Fithriasari, Benedictus Kenny Tjahjono

3830: Advanced Groundwater Level Forecasting using QSO-based Vision Transformer Model for Sustainable Water Resource Management 101
V. Gokula Krishnan, M. S. Maharajan, B.V. Subba Rao, G. Mahalakshmi, M. Arunadevi

3944: Optimizing Energy Efficiency in Three-Phase Induction Motors via GWO-Tuned PID Control 124
Muhammad Panji` Alatief, Andi Adriansyah, Yudhi Gunardi, Ahmad 'Athif Mohd Faudzi, Abu Ubaidah Shamsudin

4009: Development of a Model for Malware Detection and Classification at the Byte Level Based on Transformer 142
Rulof Baltwin Tallane, Akhmad Unggul Priantoro, Riza Muhida

4049: Deep Learning-Based Skin Care Detection with Multi-method Explainability: Grad-CAM, Lime, and Occlusion Sensitivity 160
Tooba Khan, Muhammad Zeeshan Ul Haque, Gul Munir, Irfan Ahmed Usmani

Materials and Manufacturing Engineering

3898: Performance Investigation and Efficiency Enhancement of Eco-Friendly Tin-Based CH3NH3SnI3 Perovskite Solar Cell via SCAPS-1D 175
Md. Morsalin, Ragab A. Sayed, Md. Sazedur Rahman, Md. Ferdous Wahid, Mohammad Salah

3970: Evaluating Mechanical and Conductivity of Graphene-Silver Hybrid Inks on Copper Substrate at Elevated Temperature 194
Ahmad Zaifazlin Zainordin, Nor Azmmi Masripan, Muhd Ridzuan Mansor, Chonlatee Photong, Alan Watson, Mohd Azli Salim

Mechanical and Aerospace Engineering

3980: Surrogate Model-based Optimization of a Biplane Unmanned Aerial Vehicle Wing Using Genetic Algorithm and Response Surface Methodology 214
Yüksel Eraslan

Mechatronics and Automation Engineering

3587: Stability and Control of Humanoid Robots during Walking and Sudden Stops on Uneven Terrain using Inverted Pendulum Modeling and Fuzzy-enhanced Linear Quadratic Regulator 227
Andi Dharmawan, Jazi Eko Istiyanto, Bakhtiar Alldino Ardi Sumbodo, Muhammad Auzan, Oskar Natan

3694: Vehicle Identification and Classification Using YOLO Algorithm 251
Chau Ly Thi Huyen, Chi Pham Quyet, Quyen Thi Nguyen

3703: Deep Learning Framework for Sentiment Prediction using Residual Connections in Bidirectional – Gated Recurrent Unit 263
Varsha Nagaraju, Muthalambika Chaluvegowda Padma

3716: ORCA: AI-powered Autonomous Underwater Vehicle for Subaquatic Exploration 275
Oskar Natan, Wiwit Suryanto, Andi Dharmawan, Rifda Hakima Sari, Zaidan Hakim

3749: FSM-Based MBIST Controller Implementation using the March AZ2 Test Algorithm .. 292
Aiman Zakwan Jidin, Mohd Syafiq Mispan, Razaidi Hussin, Ahmad Ridhwan Md Zin, Tengku Farah Nurhidayu Tuan Ab Razak

3780: Adaptive Energy Balance Control System via State of Charge (SOC) for a Sustainable Solar-Powered Outdoor-Hydroponics in Tropical Islands 305
Muji Juherwin, Mohamad Farid Misnan, Sakhiah Abdul Kudus, Akihiko Sato

3497: ToxoSegFusion: Attention-enhanced Dual-backbone Neural Architecture for Retinal Lesion Segmentation 321
Md Mominul Haque, Saipunidzam Mahamad, Suziah Sulaiman, Abdullateef Oluwabemiga Balogun, Hussaini Mamman

Blank page.

https://doi.org/10.31436/iiumej.v27i1.3842

Growth Kinetics of Grey Oyster Mushroom (Pleurotus pulmonarius) Spawn Culture Revival from Spray Drying Application

ZARINA ZAKARIA[1,2*], FATIHAH AZIZ[1], NOOR HASYIERAH MOHD SALLEH[1,2], MOHD SHARIZAN MD SARIP[1], SHI FERN CHONG[3]

[1]Faculty of Chemical Engineering & Technology, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis , Malaysia
[2]Centre of Excellence for Biomass Utilization, Universiti Malaysia Perlis 02600, Arau, Perlis, Malaysia
[3]Harvest Light Agritech, Lot 1412, Carok Buntong Binjal, Kubang Pasu, 06000 Jitra Kedah

**Corresponding author: zarinaz@unimap.edu.my*

(Received: 11 July 2025; Accepted: 29 Oct 2025; Published online: 12 January 2026)

ABSTRACT: Spray drying represents a promising approach for preserving fungal mycelium liquid cultures, with the assurance of successful revival from powder form being a critical factor for its application. Although liquid culture of grey oyster mushroom (*Pleurotus pulmonarius*) has been well established, its potential for drying and subsequent revival remains underexplored, thereby limiting its use in long-term storage. In this study, the effect of various spray dryer inlet temperatures (80, 90, 100, 110, 120, 130, 140, and 150 °C) on the revival and growth kinetics of *P. pulmonarius* powder cultures was evaluated using maltodextrin and Arabic gum (10%) as protective agents. Growth kinetics were modelled using the Contois model, which revealed that powder prepared at 80 °C with maltodextrin exhibited the highest μ_{max} value (6.0205), indicating superior revival and growth capacity. With Arabic gum, the highest μ_{max} (5.7604) was observed at 90 °C. At higher inlet temperatures, revival and growth were still observed with both protectants, albeit at reduced rates. Metabolic activity, reflected by glucose consumption and citric acid production, further confirmed successful revival, with citric acid reaching the highest level of 4.958 ppm. Overall, this study demonstrates that spray-dried powder cultures of *P. pulmonarius* retain viability and metabolic activity, highlighting their potential for long-term preservation.

ABSTRAK: Teknik pengeringan sembur merupakan satu inovasi berguna untuk kultur cecair miselium kulat, dan jaminan bahawa bentuk serbuk kultur tersebut boleh dihidupkan semula merupakan faktor penting dalam pengeringan. Walaupun kultur cecair miselium cendawan tiram kelabu telah berjaya dibangunkan, potensinya untuk dikeringkan dan seterusnya dihidupkan semula masih kurang diterokai, sekaligus menhadkan aplikasinya dalam penyimpanan jangka panjang. Kesan suhu masuk pengering sembur yang berbeza-beza (80, 90, 100, 110, 120, 130, 140 dan 150°C) terhadap kebolehan hidup semula dan kinetik pertumbuhan kultur serbuk tiram kelabu menggunakan 10% maltodekstrin dan gam Arab sebagai agen pelindung telah dikaji. Model Cointois telah digunakan dan nilai yang diperoleh daripada persamaan tersebut menunjukkan bahawa kultur serbuk yang disediakan pada 80°C dengan perlindungan maltodekstrin mempunyai nilai μ_{max} tertinggi iaitu 6.0205 dan menunjukkan kebolehan hidup semula serta pertumbuhan yang paling baik. Dengan penggunaan gam Arab, nilai μ_{max} tertinggi (5.7604) dicapai pada suhu 90°C. Bagi kedua-dua agen pelindung, walaupun pada suhu yang lebih tinggi, kultur serbuk masih menunjukkan kebolehan untuk hidup semula dan berkembang tetapi pada kadar yang lebih rendah. Dari aspek metabolisme, penggunaan glukosa dan penghasilan asid sitrik yang tertinggi iaitu 4.958

ppm menunjukkan bahawa kultur serbuk mempunyai aktiviti metabolik dalam tempoh hidup semula . Kajian ini membuktikan bahawa kultur serbuk *P. pulmonarius* mampu dibangunkan dan berpotensi untuk dipelihara dalam tempoh yang lebih lama.

KEYWORDS: *spray drying; mushroom spawn; powder culture; growth kinetics; Contois model*

1. INTRODUCTION

Grey oyster mushroom (*Pleurotus pulmonarius*) is the most popular mushroom in Malaysia and the second most widely cultivated mushroom globally. It belongs to the Pleurotaceae family and thrives on lignocellulosic materials rich in carbon, such as decaying wood and leaves. This fungus possesses enzymatic capabilities that enable it to degrade cellulose, thereby extracting carbon as an energy source for metabolic activities. *Pleurotus* spp. are extensively cultivated in Asia, America, and Europe due to several favourable factors, including low-cost production methods, high yield potential, environmentally sustainable agricultural processes, short harvest intervals, and minimal land requirements [1].

The successful cultivation of grey oyster mushroom (GOM) depends on various factors, such as spawn quality, growing conditions, and substrate selection. Spawn, which serves as the seed material for mushroom cultivation, plays a critical role in determining production efficiency. There are two primary types of spawn: solid and liquid, both of which can be produced using tissue culture or spore print techniques [2]. With increasing global demand for mushrooms, spawn technology has evolved from solid to liquid, and, more recently, from liquid to powder, to enhance production efficiency and improve long-term quality preservation. However, both solid and liquid cultures present challenges, including high susceptibility to contamination, limited storage stability, increased packaging costs, and complex logistics. Their production also carries risks such as contamination, large spatial requirements, and extended incubation periods [3].

Spray drying has emerged as an innovative technique for protecting biological materials, extending shelf life, and regulating the controlled release of components under specific conditions [4][5]. Several studies have investigated its effectiveness in converting microbial cultures into powder form. For instance, *Trichoderma* spp. have been successfully spray-dried to reduce spoilage risks and extend shelf life [6]. In that study, sucrose, molasses, and glycerol were tested as protectants, with 2% sucrose demonstrating the highest survival rate. Braga et al. [7] further investigated different protectants and drying temperatures for *Trichoderma* spp. finding that maltodextrin provided a 92.89% survival rate at 90°C, although survival declined at higher temperatures. Similarly, Ishak et al. [8] studied *Trichoderma asperellum*, using Arabic gum and maltodextrin as protective agents. Their results showed survival rates of 68.2% at 170°C, 15.9% at 160°C, and 0.2% at 150°C. In contrast, spray drying of *P. pulmonarius* at 80°C produced the lowest powder culture yield but resulted in the highest mycelial dry weight of 1.68 g upon revival [9].

Encapsulation technology, or protective techniques, has also been explored to improve microbial viability, particularly in biological control applications. Maruyama et al. (2020) encapsulated *Trichoderma harzianum* in polymeric microparticles, enhancing its antifungal and enzymatic activities while preserving soil microbiota. The encapsulated fungi exhibited increased chitinolytic and cellulolytic activity, thereby improving control over *Sclerotinia sclerotiorum*. Chandralekha et al. [10] investigated yeast viability post-spray drying by

incorporating various carrier materials, including maltodextrin, corn starch, Arabic gum, and acacia gum. Their study concluded that corn starch and maltodextrin yielded optimal powder output (59% w/w) and cell survival (80.5%), with corn starch being the most effective carrier. More recently, Suryabhan et al. [11] examined yeast microencapsulation using maltodextrin combined with sucrose or sorbitol, finding that survival rates increased by 35–45% with the addition of either sugar.

The Contois model has been widely applied to study microbial growth kinetics. Numerous studies have employed this model to evaluate the specific growth rates (μ) of various microorganisms, including *Lactobacillus* strains and *Shewanella baltica*, providing valuable data for industrial process design and scale-up [12][13]. Its robustness has also been demonstrated in studies of microalgae, supporting the model's applicability across a broad range of microbial species [14].

Therefore, the present study aims to evaluate the revival and growth capabilities of grey oyster mushroom powder cultures using the Contois kinetic model, quantifying key parameters such as maximum specific growth rate (μ_{max}), half-saturation constant (K_s), and doubling time. These growth characteristics were further validated by measuring metabolic activity via glucose consumption and citric acid production in the revived powder cultures.

2. METHODOLOGY

2.1. Preparation of inoculum for mycelium liquid culture

Mature fruiting bodies of grey oyster mushroom (GOM) with stipe lengths of approximately 9–11 cm were harvested. The inner gill tissue was aseptically cut into 0.5 cm × 0.5 cm sections and inoculated onto prepared Potato Dextrose Agar (PDA) in sterilized Petri dishes. The cultures were incubated at 28°C in the dark (Hipoint, Taiwan) until white mycelium fully colonized the PDA surface [15].

2.2. Preparation of stress media and induced sporulation

A 1 L stress medium was prepared by mixing 5 g of D-glucose, 130 mg of monopotassium phosphate, and 3.5 g of soluble starch in distilled water. A saline solution was also prepared by dissolving 0.85 g of NaCl and 1% (v/v) Tween® 80 in 100 mL of distilled water. Both media were sterilized at 121°C and 101.5 kPa for 15 minutes.

Hyphal fragments were harvested from the fully colonized PDA cultures. Prior to collection, 10 mL of sterile saline was added to the Petri dish, and the mycelial surface was vigorously rubbed with a sterile loop to dislodge the hyphae. Then, 10 mL of the resulting saline–hyphae suspension was inoculated into the sterile stress medium and incubated in the dark at 25°C without agitation. After 14 days of incubation, the liquid culture, also referred to as the stress culture, was ready for spray drying to produce powder culture.

2.3. Production of powder culture using a spray dryer

After 14 days of incubation, the liquid culture, containing an average of 3.298 g/L of mycelial dry weight, was subjected to spray drying. Before drying, sterile aqueous solutions of 10% (w/v) maltodextrin and 10% (w/v) Arabic gum were prepared and added separately to the liquid cultures as protective carriers.

Spray drying was performed using a laboratory-scale spray dryer (BUCHI B-290, Switzerland) equipped with a pneumatic nozzle. The dryer was operated under the following

conditions: 100% aspiration rate, feed rate of 3 mL/min, nozzle cleaning 5 times per minute, and airflow of 30 m³/h. Drying was conducted at various inlet temperatures (80°C, 90°C, 100°C, 110°C, 120°C, 130°C, 140°C, and 150°C).

A Schott bottle containing 1L of the liquid culture was placed on a magnetic stirrer to ensure uniform mixing during drying. The bottle opening was sealed with parafilm and aluminum foil to prevent contamination. The spray dryer's feed tube was inserted into the Schott bottle, enabling continuous suction and atomization of the liquid culture into powder. Upon completion, the resulting powder cultures were collected, weighed, and transferred into WHEATON® centrifuge tubes. These were sealed in plastic bags and stored at 4°C for further analysis.

2.4. Determination of powder culture growth kinetics

To determine growth kinetics, 5 g of each prepared powder culture was inoculated into 25 mL of brown sugar, rice bran, malt, and yeast extract (BRMY) containing medium [16], prepared with varying glucose concentrations of 5%, 10%, and 15% (w/v). Mycelial dry weight was measured at 5-day intervals over 15 days. Growth kinetics were modelled using the Contois equation (Eq. 1):

$$\mu = \frac{\mu_{max} \times S}{K_s + S} \quad (1)$$

where μ is the specific growth rate of the mycelium, μ_{max} is the maximum specific growth rate of the mycelium in the substrate, S is the concentration of the limiting nutrient, and K_s is the saturation constant.

2.5. Analysis of glucose and citric acid for metabolism study

A 10 g sample of powder culture prepared at 80°C was inoculated into BRMY medium and incubated for 14 days. A culture sample (1–2.5 mL) was then mixed with 25 mL of HPLC-grade water and centrifuged at 16,000 rpm for 10 minutes. The supernatant was filtered using a 0.2 µm nylon membrane filter.

Glucose and citric acid concentrations were analyzed using a High-Performance Liquid Chromatography (HPLC) system equipped with a Supelcosil LC-NH2 reverse-phase column (25 cm × 4.6 mm, 5 µm) and a refractive index detector. Isocratic elution was performed using two mobile phase compositions: CH_3CN:H_2O at 85:15 (v/v) and 65:35 (v/v), both at a flow rate of 1.5 mL/min. The sample injection volume was 20 µL, and analysis was carried out at 35°C. Standard curves for glucose (1–10% w/v) and citric acid (1–5% w/v) were prepared using HPLC-grade water. All samples and standards were analyzed in triplicate.

3. RESULTS AND DISCUSSIONS

3.1. Contois model profile at various temperatures

Figures 1 and 2 present graphs based on the Contois model at various temperatures. In Figure 1, for maltodextrin, the graph at 80°C (T80) exhibits a significantly steeper slope, with a corresponding intercept (1/µmax) of 0.1661, indicating more favourable growth kinetics. At 90°C (T90) and higher temperatures, the slopes are relatively similar, with intercept values (1/ μ_{max}) ranging from 0.2047 to 0.2841. In Figure 2, for Arabic gum, the T90 plot shows the steepest slope (0.1736), whereas the other plots are more gradual, with intercept values ranging

from 0.2574 to 0.9209. These positive, linear upward trends indicate that the dried culture powder retains the ability to revive and grow after thermal treatment.

Notably, temperature treatments from 80°C to 150°C did not destroy the mycelial cultures. Instead, they retained viability and regrew upon subculture in BRMY medium. Moreover, increasing glucose concentration in the medium further enhanced mycelial growth. The use of protectants, specifically maltodextrin and gum Arabic, resulted in slightly different trends across the temperature profiles. These results could not be achieved without the use of protectants, as glucose-containing solutions became sticky during spray drying, leading to poor powder formation and no viable outcomes in the control (no protectant) samples [17].

These findings suggest that temperatures between 90°C and 150°C imposed some thermal stress on the mycelium, affecting growth performance. However, cultures dried at 80°C experienced less thermal damage, effectively preserving cellular vigour and enabling robust regrowth. All positive linear trends support the conclusion that the culture powder retained revival capability after drying. Importantly, these results highlight the notable heat tolerance of grey oyster mushroom (GOM) cultures and their potential for long-term preservation through spray drying. Both protectants demonstrated that GOM mycelium could tolerate drying temperatures up to 150°C, although lower temperatures (e.g., 80°C) are preferable for maintaining cell integrity and future growth potential. Similarly, spray-drying Lacticaseibacillus rhamnosus GG at 85°C resulted in successful revival after 13 weeks of storage, with minimal cellular damage and preserved viability during drying [18].

Several studies have identified 80°C as the lowest feasible spray-drying temperature that still yields desirable powder properties. This experiment was designed to balance optimal powder formation with maximum preservation of cellular viability. Therefore, the 80°C-90°C range is optimal for drying, producing viable powder while ensuring good survival and revival characteristics. These findings underscore the heat resilience of GOM mycelia and spores, making them excellent candidates for thermal preservation through spray drying.

https://doi.org/10.31436/iiumej.v27i1.3842

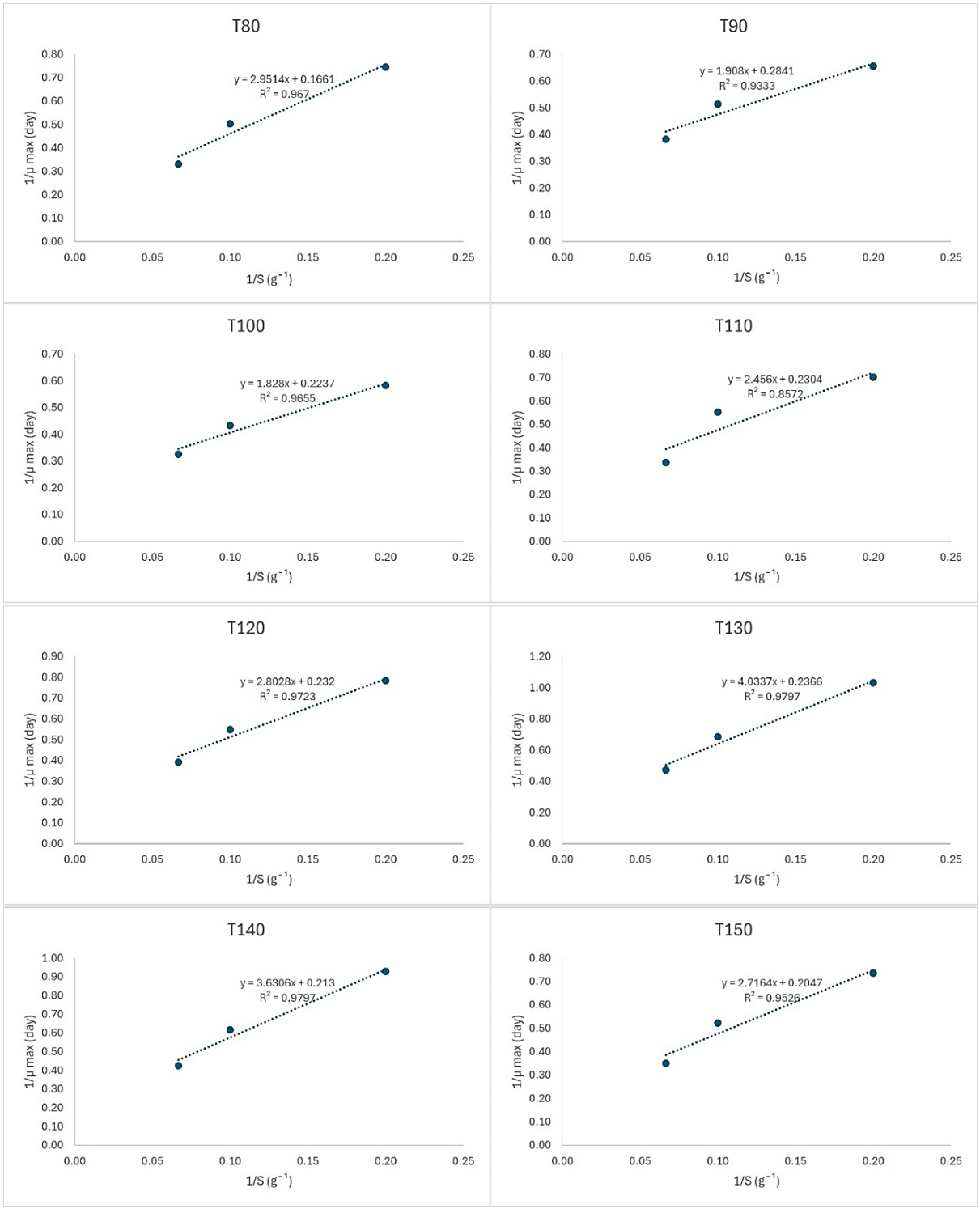


Figure 1. Linearized plot of the Contois model with maltodextrin as protectant

https://doi.org/10.31436/iiumej.v27i1.3842

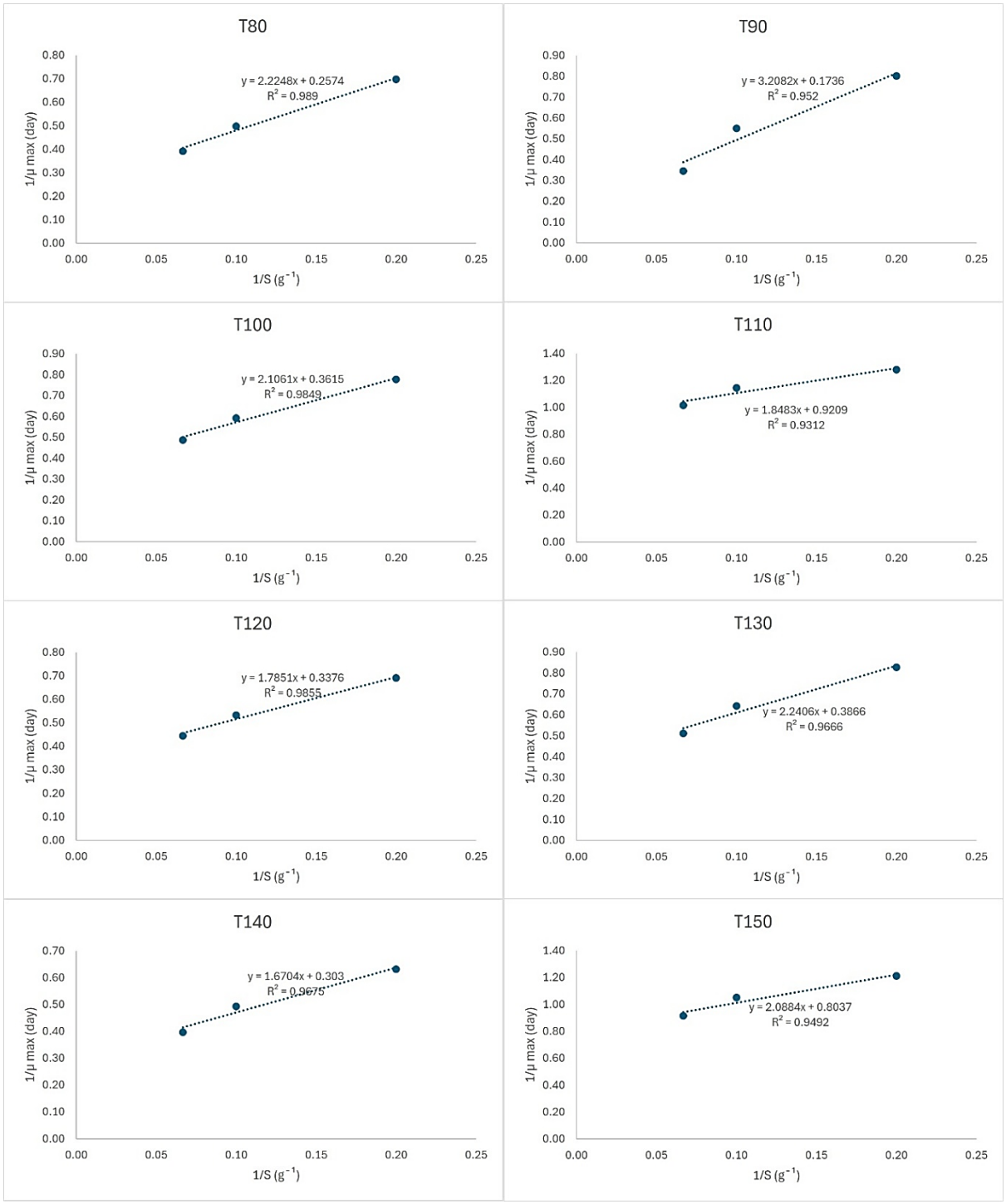


Figure 2. Linearized plot of the Contois model with Arabic gum as protectant

https://doi.org/10.31436/iiumej.v27i1.3842

3.2. Effect of temperature on μ_{max} and K_s

Tables 1 and 2 present values derived from the plots of $1/\mu_{max}$ (day) and 1/S (g^{-1}), corresponding to μ_{max}, K_s and doubling time for cultures protected with maltodextrin and Arabic gum. These results provide insight into the growth kinetics of dried GOM mycelial cultures across varying spray-drying temperatures. For maltodextrin, the highest μ_{max} value (6.0205) was observed at 80°C, indicating optimal growth performance under this condition (Table 1). Above this temperature, μmax values declined consistently with increasing drying temperatures. This trend suggests increasing thermal stress on the cells, leading to reduced growth capability. Interestingly, higher μ_{max} values were often accompanied by higher K_s values, indicating a trade-off between growth rate and substrate affinity. For Arabic gum, the peak μ_{max} (5.7604) was achieved at 90°C, followed by a noticeable decline at higher temperatures (Table 2). This demonstrates that Arabic gum offered some protective benefit, but its effectiveness diminished beyond 90°C. In terms of doubling time, cultures with maltodextrin had a shorter doubling time (0.1151 days) than those with Arabic gum (0.1203 days), indicating faster growth with maltodextrin as a protectant. These kinetic parameters affirm that both protectants contributed to preserving culture viability, although maltodextrin was slightly more effective. The cultures were well encapsulated as shown by scanning electron micrograph (SEM) in an earlier study by Zakaria et al. [19], whose SEM analysis showed that cultures dried between 80°C and 120°C retained a generally spherical particle shape, despite the presence of wrinkles and surface indentations. Such structural integrity supports the hypothesis that cells were well encapsulated during drying, thereby enhancing protection and revival capacity.

In this study, μ_{max} is the primary indicator of the culture powder's revival potential in BRMY medium. A higher μ_{max} value indicates greater specific growth capacity and thus better survival during the drying process. Importantly, even though μ_{max} decreased at higher temperatures, growth was still observed, suggesting that some portion of the mycelium retained viability despite the thermal stress. It is hypothesized that this surviving portion may consist of heat-resistant spores, particularly chlamydospores [20]. The heat resistance of these cultures appears to be influenced both by spore structure and the effectiveness of the protectant used. Compared with maltodextrin, Arabic gum was less effective, as evidenced by lower μmax values, particularly at temperatures above 90°C, where mycelial vigour declined significantly [21].

Another key consideration is the relationship between μ_{max} and K_s, (the Contois saturation constant). Theoretically, K_s, reflects the affinity of cultures for the substrate here, glucose. A higher K_s, indicates lower substrate affinity, implying that the culture requires a higher glucose concentration to sustain growth. This suggests that revival cultures may initially require a longer adaptation phase and thus benefit from increased substrate availability. Conversely, when μ_{max} decreases, K_s, also tends to decrease, which indicates improved substrate affinity, possibly due to gradual adaptation by stressed cells, enhancing their efficiency in glucose uptake. This adaptive behaviour has been described by Jefferson et al. [22], who demonstrated a similar inverse relationship between μ_{max} and K_s, in *Potamosiphon* spp. cultures under phosphate limitation, using the Contois model. A similar trend has been observed in bacterial cultures. For example, Zahari et al. [23] reported that well-adapted bacterial cells exhibited lower K_s, values and higher division rates, confirming the connection between nutrient affinity and culture recovery. In the present study, the strong correlation between $1/\mu_{max}$ and 1/S ($R^2 > 96\%$ across most temperatures) further supports the suitability of the Contois model for describing the growth dynamics of spray-dried GOM cultures.

Table 1. Growth kinetics derived from the Contois model at various spray drying temperatures using maltodextrin as protectant.

Temperature(°C)	μ_{max}	K_s	Doubling time (t_d)	R^2
T80	6.0205	17.7688	0.1151	0.9670
T90	3.5199	6.7159	0.1969	0.9673
T100	4.4703	8.1717	0.1551	0.9655
T110	4.3403	10.6597	0.1597	0.8572
T120	4.3103	12.0810	0.1608	0.9988
T130	4.2265	17.0486	0.1640	0.9981
T140	4.6948	17.0451	0.1476	0.9797
T150	4.8852	13.2702	0.1419	0.9526

Table 2. Growth kinetics derived from the Contois model at various spray drying temperatures using Arabic gum as protectant.

Temperature(°C)	μ_{max}	K_s	Doubling time (t_d)	R^2
T80	3.8850	8.6434	0.1784	0.9890
T90	5.7604	18.4804	0.1203	0.9520
T100	2.7663	5.8260	0.2506	0.9849
T110	1.0859	2.0071	0.6383	0.9988
T120	2.9621	5.2876	0.2340	0.9876
T130	2.5867	5.7957	0.2680	0.9666
T140	3.3003	5.5129	0.2100	0.9930
T150	1.2442	2.5985	0.5571	0.9807

3.3. Evaluation of powder culture revival through glucose utilization and citric acid production

In powder spawn samples protected with maltodextrin, glucose was undetectable at all incubation time points (Days 5, 10, and 15), suggesting complete consumption of glucose during the early stages of incubation. In contrast, samples protected with Arabic gum exhibited a gradual reduction in glucose concentration, from 19.428 ppm on Day 5 to 8.612 ppm on Day 10, with no detectable glucose remaining by Day 15. The presence and subsequent reduction of citric acid further confirm metabolic activity in the revived powder spawn. In maltodextrin-treated samples, citric acid concentrations declined steadily from 4.958 ppm (Day 5) to 0.702 ppm (Day 15). In the Arabic gum group, citric acid levels were consistently lower, ranging from 3.804 ppm on Day 5 to 0.604 ppm on Day 15. The complete data on glucose and citric acid concentrations are presented in Table 3.

These findings indicate that the revived powder spawn metabolized available glucose, reflecting high metabolic activity and efficient substrate assimilation, particularly in maltodextrin-protected samples. The absence of residual glucose in these cultures suggests that the fungi resumed growth promptly upon rehydration, in line with the expected behaviour of viable fungal cultures post-revival [24]. In contrast, the delayed glucose utilization observed in Arabic gum-treated samples may reflect slower revival or metabolic activation. However, the eventual depletion of glucose by Day 15 still supports successful metabolic revival over time.

This trend highlights a pattern of early metabolic activity, characterized by high carbon uptake (as seen on Day 5), followed by a gradual decline in glucose and citric acid levels, likely due to substrate exhaustion or re-utilization of metabolic byproducts. Day 5 appears to represent an optimal period of active mycelial growth, after which metabolic activity declines. Several factors may influence citric acid production and accumulation, including possible re-utilization, dilution from increased biomass, or shifts in metabolic pathways [25]. Citric acid is a recognized metabolic byproduct of fungal growth, particularly in *Aspergillus* spp. and related filamentous fungi, and its presence serves as a reliable indicator of cellular respiration and biosynthetic activity. The lower citric acid concentrations observed in Arabic gum samples may indicate reduced or delayed metabolic activity, potentially due to differences in the physicochemical properties of Arabic gum. These may affect its protective efficiency during spray drying or influence the rehydration and revival performance of the powder spawn [26].

Table 3. HPLC analysis of glucose and citric acid concentrations during the incubation of revived powder spawn with different protectants.

Protectant	Incubation days	Glucose concentration [ppm]	Citric acid concentration [ppm]
Maltodextrin	5	Not detectable	4.958
	10	Not detectable	2.866
	15	Not detectable	0.702
Arabic Gum	5	19.428	3.804
	10	8.612	0.918
	15	Not detectable	0.604

4. CONCLUSION

This study demonstrates that spray drying is an effective and viable method for preserving liquid cultures of *Pleurotus pulmonarius* (grey oyster mushroom) in powder form, with the capacity to revive and sustain metabolic activity upon rehydration. The incorporation of protectants, specifically maltodextrin and Arabic gum, significantly enhanced the viability, revival, and growth performance of the powdered cultures across a range of inlet drying temperatures. Analysis using the Contois model revealed that the highest specific growth rates (μ_{max}) were achieved at 80°C with maltodextrin (6.0205) and 90°C with Arabic gum (5.7604), indicating these temperatures as optimal for post-drying culture revival. Although higher temperatures resulted in lower μ_{max} values, the cultures' ability to revive and grow suggests good thermal resilience of the mycelial cells in powder form. Further confirmation of metabolic activity was provided through glucose depletion and citric acid production, with citric acid concentrations reaching up to 4.958 ppm, underscoring the functional viability of the revived cultures. Overall, the findings support the potential application of spray-dried *P. pulmonarius* cultures for long-term storage, offering a practical and scalable solution for the mushroom industry, particularly for spawn distribution, preservation, and ready-to-use formulations.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Faculty of Chemical Engineering & Technology, Universiti Malaysia Perlis (UniMAP) for providing the necessary facilities and continuous support throughout this research. Appreciation is also extended to the Ministry of Higher Education (MOHE), Malaysia, for financial support through the Fundamental Research Grant Scheme (FRGS/1/2020/STG01/UNIMAP/02/2).

REFERENCES

[1] Rosmiza M, Davies W, Aznie RC, Jabil M, Mazdi M. (2016). Prospects for increasing commercial mushroom production in Malaysia: challenges and opportunities. Mediterr J Soc Sci. 7(1): 406–415.

[2] Thawthong A, Karunarathna SC, Thongklang N, Chukeatirote E, Kakumy P, Chamyuang S, Rizal LM, Mortimer PE, Xu J, Calac P, Hyde KD. (2014). Discovering and domesticating wild tropical cultivatable mushrooms. Chiang Mai J Sci. 41(4): 731–764.

[3] Kerketta A, Pandey NK, Singh HK. (2017) Evaluation of Different Grains for Spawn Development of Milky Mushroom *(Calocybe indica*, P&C). Research Journal of Agricultural Sciences, 8(1), 266–268.

[4] Huang S, Vignolles M-L, Chen XD, Loir YL, Jan G, Schuck P, Jeantet R. (2017). Spray drying of probiotics and other food-grade bacteria: A review. Trends Food Sci Technol. 63: 1-17.

[5] Shishir MRI, Chen W. (2017). Trends of spray drying: A critical review on drying of fruit and vegetable juices. Trends Food Sci Technol. 65: 49–67.

[6] Jin X, Custis D. (2011). Microencapsulating aerial conidia of *Trichoderma harzianum* through spray drying at elevated temperatures. Biol Control. 56(2) :202–208.

[7] Braga ABAC, Costa CJM, Pomella AWV, Ribeiro EJ, Santos LD, Zotarelli MF. (2019). Evaluation of lethality temperature and use of different wall materials in the microencapsulation process of *Trichoderma asperellum* conidias by spray drying. Powder Technol. 347: 199–206.

[8] Ishak AA, Zulkepli FRA, Hayin NFM, Md Zain NH, Sapak Z. (2021). Effect of high inlet temperature of spray dryer on viability of microencapsulated *Trichoderma asperellum* conidia. IOP Conference Series: Earth and Environmental Science. 757(1).

[9] Zakaria Z, Fatihah A, Sarip MSM, Salleh NHM. (2024). Growth revival of grey oyster (*Pleurotus pulmonarius*) powder culture mushroom from the effect of spray drying temperature. IIUM Engineering Journal. 25(1): 40-46.

[10] Chandralekha A, Tavanandi AH, Amrutha N, Hebbar HU, Raghavarao KSMS, Gadre R. (2016). Encapsulation of yeast (*Saccharomyces cereviciae*) by spray drying for extension of shelf life. Drying Technology. 34(11): 1307–1318.

[11] Suryabhan P, Lohith K, Anu-Appaiah KA. (2019). Sucrose and sorbitol supplementation on maltodextrin encapsulation enhance the potential probiotic yeast survival by spray drying. Lwt, 107: 243–248.

[12] Rezvani F, Ardestani F, Najafpour G. (2017). Growth kinetic models of five species of Lactobacilli and lactose consumption in batch submerged culture. Braz J Microbiol. 48(2): 251.

[13] Cante RC, Gallo M, Nigro F, Passannanti F, Budelli A, Nigro R. (2021). Mathematical modeling of *Lactobacillus paracasei* cba l74 growth during rice flour fermentation performed with and without pH control. Applied Sciences (Switzerland). 11(7).

[14] Hisbullah Y, Rinaldi W, Fathanah U. (2019). Development and validation of a modified Contois kinetics model for microalgae *Chlorella kessleri*. International Graduate Conference 2018, October 03-05, Banda Aceh, Indonesia.

[15] Buba T, Agbo V, Abdullahi A, Emmanuel J. (2019). Sensitivity of mycelial growth of mushroom *Agrocybe semiorbicularis* to different concentrations of different disinfectants, BioTechnologia. 100 (4): 453–463.

[16] Abdullah N, Ismail R, Johari NMK, Annuar MSM. (2013). Production of liquid culture of an edible grey oyster mushroom, *Pleurotus pulmonarius* (Fr.) by submerged fermentation and sporophore yield on rubber wood sawdust. Scientia Horticulturae,. 161: 65–69.

[17] Aziz F, Zakaria Z, Sarip MSM. (2023). Role of carrier agent in spray drying method for stickiness prevention and suitability in *Pleurotus pulmonarius* liquid spawn application: A review. AIP Conference Proceedings 2703, 040002.

[18] Jayaprakash P, Gaiani C, Edorh J-M, Borges F, Audrey Maudhuit EB, Desobry S. (2023). Comparison of electrostatic spray drying, spray drying, and freeze drying for *Lacticaseibacillus rhamnosus*. Dehydration. Foods. 12: 3117.

[19] Zakaria Z, Aziz F, Salleh MNH, Sarip MSM. (2024). Properties and microscopy analysis of encapsulated spray dried grey oyster mushroom (*Pleurotus pulmonarius*) powder culture from temperature effect. Kexue Tongbao/Chinese Science Bulletin. Vol. 69 (2): 1185-1194.

[20] Chong SF, Zakaria Z, Low JZ. (2019). Investigation of chlamydospores and mycelium liquid culture storage period of grey oyster mushroom (*Pleurotus pulmonarius*) and effects on growth rate. Int. J. Adv. Sc. Eng. Tech. 7(2): 2321-8991.

[21] Fatihah A, Zakaria Z, Sarip MSM. (2023). Role of carrier agent in spray drying method for stickiness prevention and suitability in *Pleurotus pulmonarius* liquid spawn application: A review. AIP Conference Proceeding. 2703 (1).

[22] Jefferson E, Contreras-Ropero, Janet B, García-Martínez, Andrés F, Barajas-Solano. (2025). Integration of mathematical and experimental modeling for sustainable phycobiliprotein production via fed-batch cultures. S Afr J Chem Eng. 51: 35-44.

[23] Zahari NZ, Tuah P M, Ali SAM, Sabullah MK. (2022). Microbial growth rate and distribution of doubling time at different concentration of oil sludge medium. IOP Conf. Series: Earth and Environmental Science. 1022: 012075.

[24] Mäkelä MR, Aguilar-Pontes MV, van Rossen-Ufnk D, Peng M, deVries RP. (2018). The fungus *Aspergillus niger* consumes sugars in a sequential manner that is not mediated by the carbon catabolite repressor CreA. *Sci Rep*. 8:6655.

[25] Książek E. (2023). Citric acid: properties, microbial production, and applications in industries. Molecules (Basel, Switzerland). 29(1): 22.

[26] Wang W, Dufour C, Zhou W. (2015). Impacts of spray-drying conditions on the physicochemical properties of soy sauce powders using maltodextrin as auxiliary drying carrier. CYTA J Food. 13(4): 548–555.

https://doi.org/10.31436/iiumej.v27i1.3654

Inertial Sensor Self-Calibration Module Using Attitude Heading and Reference System for Autonomous Underwater Vehicle Navigation

ABDUL AZIZ ABD KADIR, HERDAWATIE ABDUL KADIR*

Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia

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

(Received: 19 March 2025; Accepted: 17 November 2025; Published online: 12 January 2026)

ABSTRACT: This research addresses the complex task of enhancing navigation accuracy in Autonomous Underwater Vehicles (AUVs), a self-propelled robotic system used for ocean exploration, environmental monitoring, and underwater interventions. A core component of AUV navigation is an Inertial Measurement Unit (IMU), a sensor suite that tracks orientation and motion by measuring accelerations and angular rates. However, the IMU is highly susceptible to noise interference, which degrades accuracy and reliability. To address these challenges, this study introduces an innovative Inertial Sensor Self-Calibration Module that dynamically adjusts calibration parameters in real time, thereby compensating for sensor drift and inaccuracies. The research further conducts a comparative analysis of several calibration and filtering techniques integrated into the AUV's Inertial Navigation System (INS), including Magnetic Calibration, the Extended Kalman Filter (EKF), the EKF with Measurement Noise, the EKF with Process Noise, and the Attitude and Heading Reference System (AHRS) Filter. Among these, the AHRS filter demonstrated superior precision, achieving the lowest average error of 1.06 degrees with a standard deviation of 0.345 degrees in angle measurements, an improvement of up to 97.81% compared to raw data. These findings highlight the effectiveness of the AHRS filter in improving navigation accuracy in complex underwater environments. The insights gained from this research not only deepen the understanding of noise impact and sensor calibration in AUV systems but also pave the way for future innovations in oceanic exploration, environmental monitoring, and underwater interventions.

ABSTRAK: Kajian ini adalah berkenaan menangani tugas kompleks meningkatkan ketepatan navigasi Kenderaan Bawah Air Autonomi (AUV), iaitu sistem robotik berkuasa sendiri digunakan bagi penerokaan lautan, pemantauan alam sekitar, dan intervensi bawah air. Komponen utama navigasi AUV bergantung pada Unit Pengukuran Inersia (IMU), iaitu rangkaian pengesan orientasi dan pergerakan dengan mengukur pecutan dan kadar sudut. Namun, IMU sangat terdedah kepada gangguan bunyi, di mana ianya mengurangkan ketepatan dan kebolehpercayaan. Oleh itu, kajian ini memperkenalkan Modul Kalibrasi Diri Pengesan Inersia yang inovatif, iaitu secara dinamik menyesuaikan parameter kalibrasi pada masa nyata, secara efektif mengimbangi hanyutan sensor dan ketidaktepatan. Kajian ini juga membuat analisis perbandingan beberapa teknik kalibrasi dan penapisan yang diintegrasikan pada Sistem Navigasi Inersia (INS) AUV, termasuk Kalibrasi Magnetik, Penapis Kalman Lanjutan (EKF), EKF dengan Bunyi Pengukuran, EKF dengan Bunyi Proses, dan Penapis Sistem Rujukan Sikap dan Haluan (AHRS). Antara teknik-teknik ini, penapis AHRS menunjukkan ketepatan terbaik, dengan ralat purata terendah iaitu 1.06 darjah dan sisihan piawai pengukuran sudut 0.345 darjah dan peningkatan sehingga 97.81% berbanding data mentah. Penemuan ini menunjukkan keberkesanan penapis AHRS dalam meningkatkan ketepatan navigasi pada persekitaran bawah air yang kompleks. Dapatan kajian ini bukan sahaja memperdalam pemahaman tentang kesan bunyi dan kalibrasi pengesan dalam sistem

AUV, tetapi turut membuka ruang terhadap inovasi masa depan dalam penerokaan lautan, pemantauan alam sekitar, dan intervensi bawah air.

KEY WORDS: *Autonomous Underwater Vehicle (AUV), Inertial Navigation System (INS), Inertial Measurement Unit (IMU), Attitude Heading and Reference System (AHRS), Extended Kalman Filter (EKF)*

1. INTRODUCTION

The robotics era has reached a point at which the Autonomous Underwater Vehicle (AUV) industry is well established, with thousands of AUVs created and deployed since its inception [1]. Today, the robotics and AUV industries are growing steadily, driven by advances in functionality and reliability [2]. The primary objective of underwater robotics is to develop autonomous, intelligent AUVs capable of decision-making under diverse conditions. To accomplish these goals, several studies have been conducted worldwide, with an emphasis on design, navigation, energy sources, and information systems.

A core component of AUV navigation is an Inertial Measurement Unit (IMU), a sensor suite that tracks orientation and motion by measuring accelerations and angular rates. However, the IMU is highly susceptible to noise interference, which degrades accuracy and reliability. To address these challenges, this study introduces an innovative Inertial Sensor Self-Calibration Module that dynamically adjusts calibration parameters in real time, thereby compensating for sensor drift and inaccuracies. The primary obstacle to the successful operation of the AUV navigation module is maintaining the AUV's position accuracy over extended periods [3-5]. During long missions, strong, prolonged currents and other underwater resistances affect the AUV's motion, thereby decreasing accuracy [6,7]. If the position of the AUV is not analyzed correctly, the accuracy of the AUV's position will degrade over time, and a lack of observation and references from external sources makes AUV navigation impossible.

Most AUV navigation systems use inertial navigation because the Global Positioning System (GPS) cannot function underwater. Some AUVs require acoustic guidance from a mother ship or use underwater acoustic transponders to increase the vehicle's autonomy [8]. To avoid the costly deployment of underwater transponders, the Inertial Sensor Self-Calibration Module is a favorable alternative, and the key sensors typically required are accelerometers, gyroscopes, pressure sensors, and Doppler Velocity Logs (DVLs) [3,9-10].

Persistent sensor drift can undermine the reliability and performance of AUVs, limiting their effectiveness in dynamic underwater environments [11]. In practice, due to their considerable drift and interference with the magnetic field, the sensor readings are prone to error and instability [12]. One of the primary challenges arises from noise in IMU data, which can stem from electronic interference, vibrations, and environmental conditions [13].

Advances in inertial platform technology have reduced their cost to a level that small researchers can now afford. Inertial navigation systems were also available, but not all could afford them. Advances in inertial platform technology have reduced their cost to a level that small researchers can now afford [14].

AUV navigation systems have continued to improve in accuracy over time. In recent years, many researchers have sought to apply GPS technology to their AUVs. GPS technology has contributed significantly to AUV navigation systems due to its precise positioning and navigation capabilities; however, it also has a significant disadvantage: GPS signals cannot reach underwater. The researchers have developed a method to address this drawback [15].

https://doi.org/10.31436/iiumej.v27i1.3654

2. METHODOLOGY

2.1. AUV Design Process

The design process for an AUV involves several key phases. The initial phase emphasizes establishing the AUV's foundational framework, encompassing both its electrical and mechanical design elements. This foundational stage serves as the cornerstone of subsequent phases, providing a robust foundation for overall AUV development.

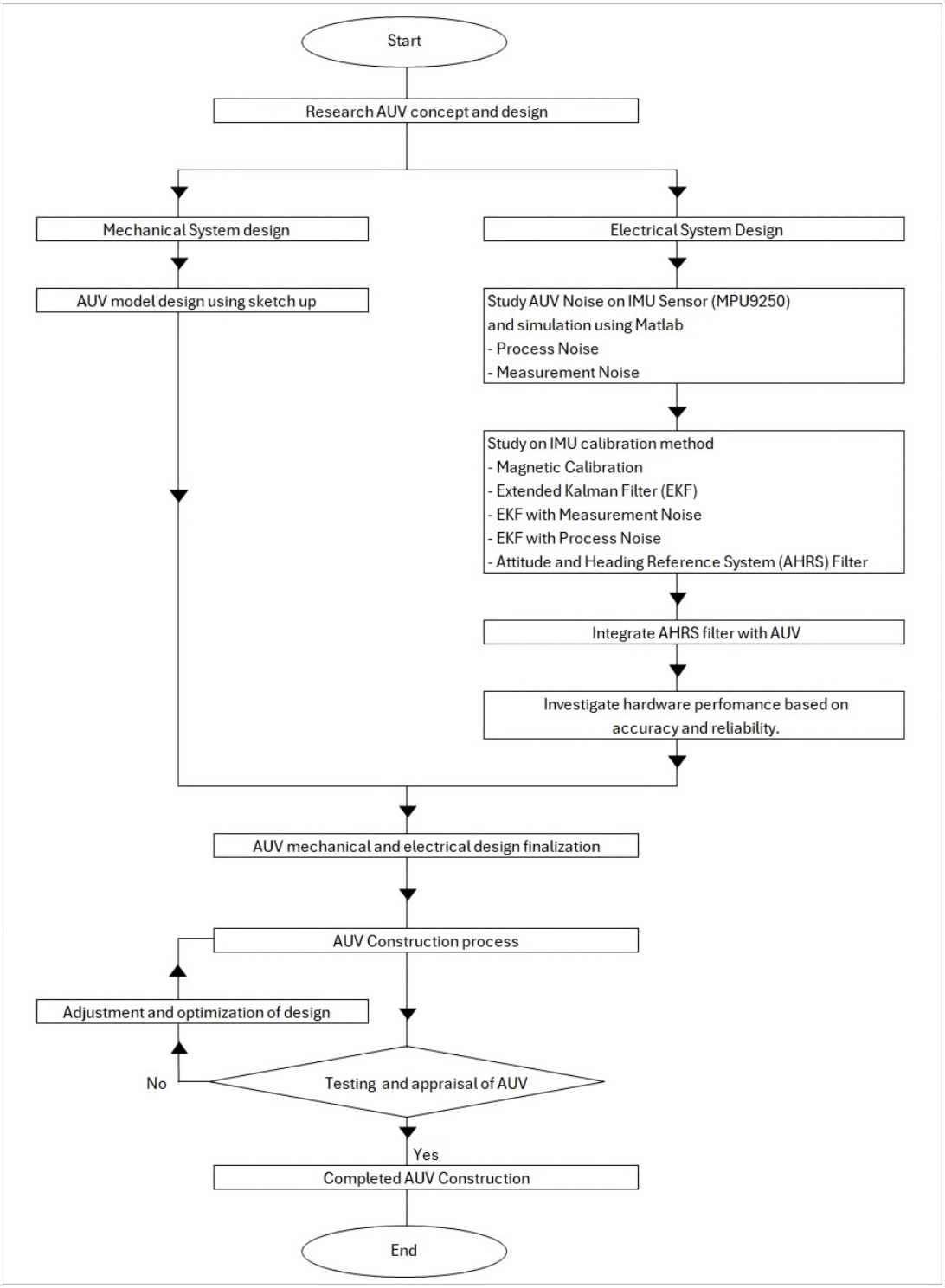


Figure 1. Overall project flowchart.

https://doi.org/10.31436/iiumej.v27i1.3654

The subsequent stages can be broadly categorized into two sections, each addressing distinct aspects of the AUV's design. The first section delves into the mechanical design of the AUV. During this phase, Computer-Aided Design (CAD) software is employed to model and animate the AUV's mechanical structure in detail. This software facilitates a detailed visualization of the AUV's physical components, thereby aiding in the refinement and optimization of its structural design.

Simultaneously, the second section of the design process focuses on the electrical and electronic aspects of the AUV. This involves not only the design of electronic components but also the meticulous planning of the wiring layout. Special attention is given to preventing magnetic interference between components to mitigate the risk of sensor errors. This comprehensive approach ensures that the AUV's electrical systems are efficiently integrated and function cohesively.

The final phase of the AUV design process centers on rigorous testing. Once both mechanical and electrical components are installed, the AUV undergoes thorough testing to identify any potential issues or errors. This testing phase is crucial for validating the AUV's overall functionality. Any discrepancies or inefficiencies that arise during testing prompt adjustments, whether major or minor, to correct and refine the AUV's performance.

In summary, the AUV design process follows a systematic approach, beginning with the foundational electrical, electronic, and mechanical design and progressing through detailed modelling, wiring design, and rigorous testing. This methodical process, as illustrated in Figure 1, ensures a comprehensive and well-optimized AUV design that meets the project's objectives, scope, and requirements.

2.2. AUV Design Structure

The AUV design structure will comprise several key design aspects, implemented in SketchUp, a CAD software. These figures provide detailed insight into the overall structure and composition of the AUV, meticulously capturing the designed dimensions, materials used, and the arrangement of various components.

The utilization of SketchUp facilitates a comprehensive representation of the AUV's mechanical design. The software enables modelling and visualization of the AUV's physical structure, providing a detailed depiction of how components and materials interact to form the AUV framework. This visual representation serves as a crucial reference for understanding the spatial organization of mechanical parts within the AUV.

By visually presenting the AUV design through these figures, the CAD software not only aids in conveying the physical structure but also serves as a valuable tool for design evaluation and refinement. Engineers and stakeholders can use these visual representations to assess the feasibility, efficiency, and overall effectiveness of the AUV design before proceeding with construction and assembly.

Notably, Figures 2-4 depict the placement of mechanical components, including the primary and secondary compartments. The main compartment houses critical AUV components, with the mechanical elements arranged to optimize functionality. Simultaneously, the secondary compartment is designed to accommodate various electrical components, thereby illustrating the integration of mechanical and electrical systems within the AUV. Figure 5 shows the completed AUV design structure, based on the initial CAD sketch, with adjustments and optimizations made during AUV construction.

https://doi.org/10.31436/iiumej.v27i1.3654

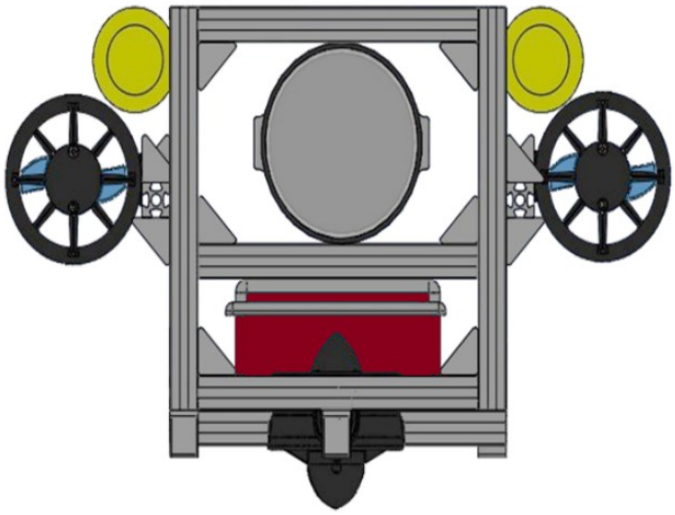


Figure 2. AUV Design Front View

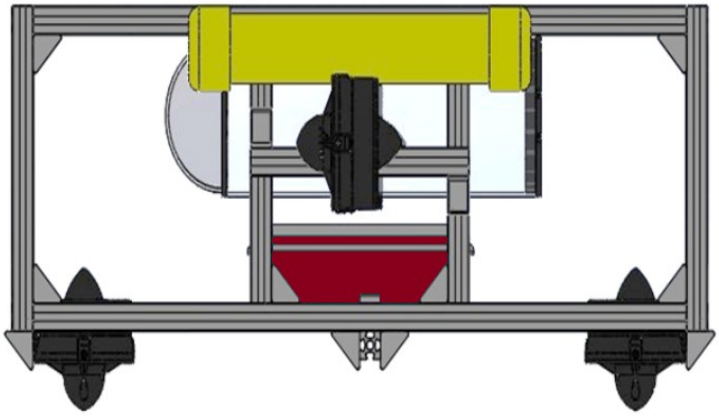


Figure 3. AUV Design Side View

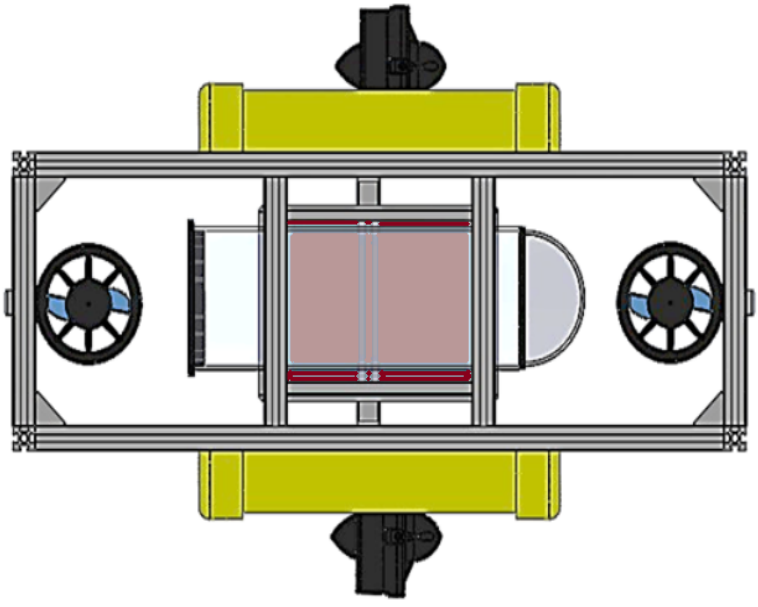


Figure 4. AUV Design Top View

https://doi.org/10.31436/iiumej.v27i1.3654

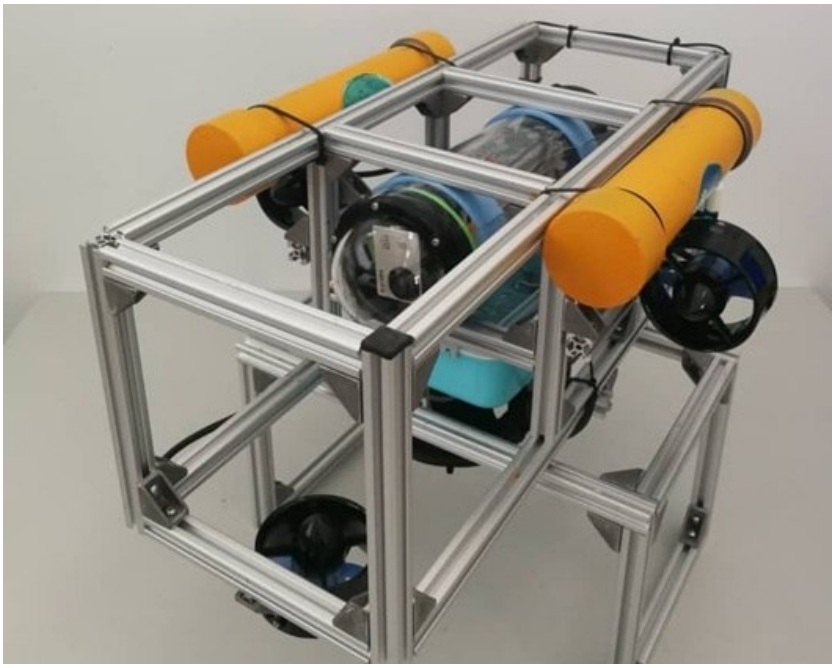


Figure 5. Completed AUV Design

Next, an essential component is to provide detailed insight into the AUV's main electronic system architecture. This serves as a key reference point for understanding the configuration and interconnection of electronic components crucial to the AUV's navigation system.

The core of the AUV's navigation system is prominently integrated into the electronic system architecture, underscoring the IMU's central role. The IMU is a critical component that corrects potential errors in other sensors within the AUV. This correction mechanism is pivotal for ensuring the accuracy and reliability of the AUV's navigation system, particularly in the dynamic and challenging underwater environment.

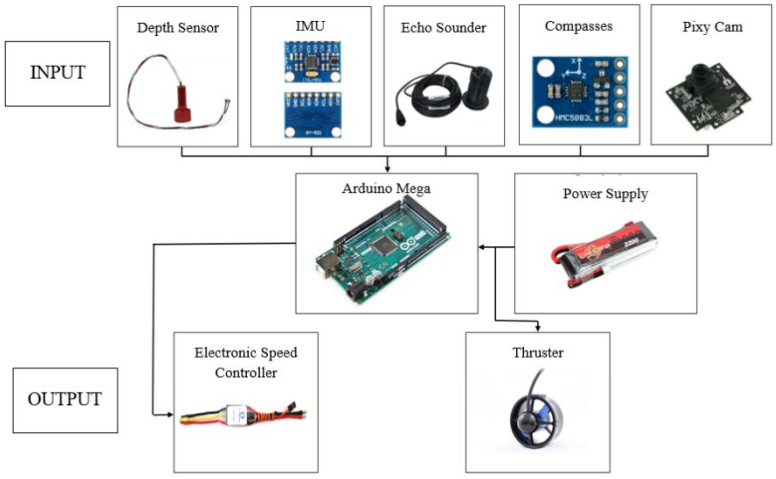


Figure 6. Electrical and Electronic System

By visually representing the electronic system structure, the design provides insight into how electronic components, including sensors and the IMU, are integrated into the overall AUV framework. The explicit depiction of these connections aids engineers and designers in assessing the effectiveness of the electronic system, identifying potential areas for improvement, and ensuring the seamless coordination of electronic components for precise

https://doi.org/10.31436/iiumej.v27i1.3654

navigation. Figure 6 serves as a valuable visual reference for understanding the electrical and electronic architecture of the AUV, highlighting the IMU's pivotal role in enhancing the accuracy and reliability of the navigation system.

3. RESULTS AND DISCUSSION

3.1. AUV IMU Filters Performance Analysis Simulation using Matlab

This simulation assessment will focus on the AUV's orientation readings to analyze and ensure the stability of the AUV's performance during navigation. Filter performance was evaluated through a thorough MATLAB simulation over 140 seconds, yielding 28 data samples at 5-second intervals. The yaw, pitch, and roll parameters represent the AUV's rotational motion. Yaw signifies horizontal rotation, pitch indicates vertical inclination, and roll denotes lateral rotation. The samples are selected to ensure stable sampling, as the accelerometer, gyroscope, and magnetometer in the MPU9250 have distinct sampling periods. This meticulous assessment was conducted at angles of 0°, 90°, 180°, and -90°, as shown in Figures 7-10, in a controlled testing environment to evaluate the effectiveness of a range of filter and calibration techniques.

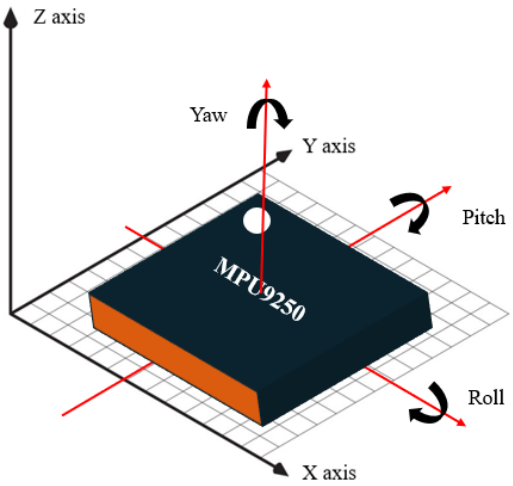


Figure 7. Example IMU orientation at 0° Pitch angles

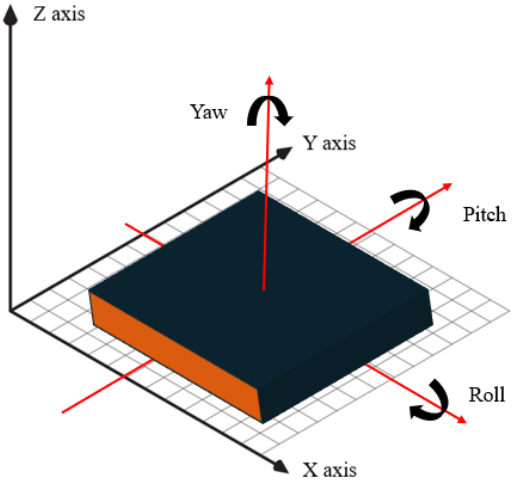


Figure 8. Example IMU orientation at 180° Pitch angles

https://doi.org/10.31436/iiumej.v27i1.3654

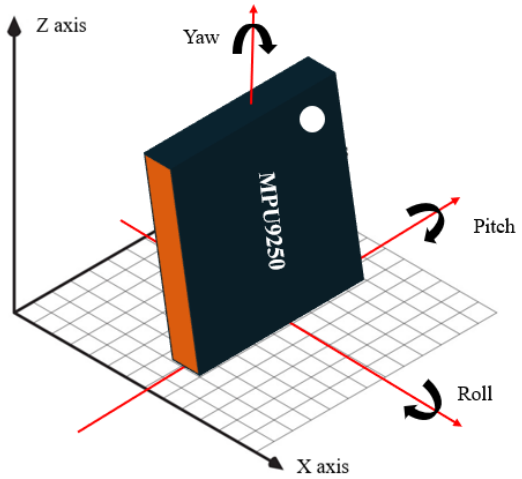


Figure 9. Example IMU orientation at 90° Pitch angles

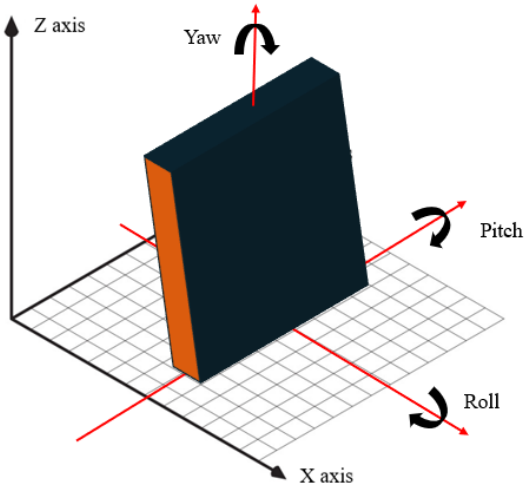


Figure 10. Example IMU orientation at -90° Pitch angles

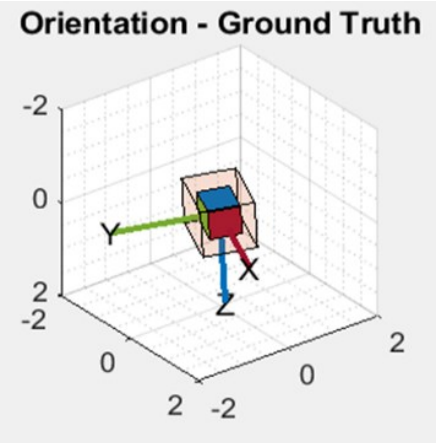


Figure 11. AUV Matlab orientation simulation

Figure 11 illustrates the AUV MATLAB simulation orientation obtained from the orientation assessment, yielding yaw, pitch, and roll readings for the selected angle

https://doi.org/10.31436/iiumej.v27i1.3654

orientations, serving as a reference for the accurate depiction of the actual angle measurements. This figure illustrates the orientation under investigation.

3.2. IMU Filters Performance Analysis Simulation at 0° angle

Figures 12 - 17 illustrate yaw, pitch, and roll rotations for 28 data samples. This graph visually portrays orientation dynamics across filter techniques: Raw Data, Magnetic Calibration, EKF, EKF with Measurement Noise, EKF with Process Noise, and AHRS Filter on 0° angle orientation to verify each filter's efficiency. Notably, the visual representation indicates that the AHRS Filter yields the most favorable outcomes, enhancing orientation accuracy. These graphs are further analyzed using tabular methods to determine the maximum, minimum, mean, and standard deviation for each filter.

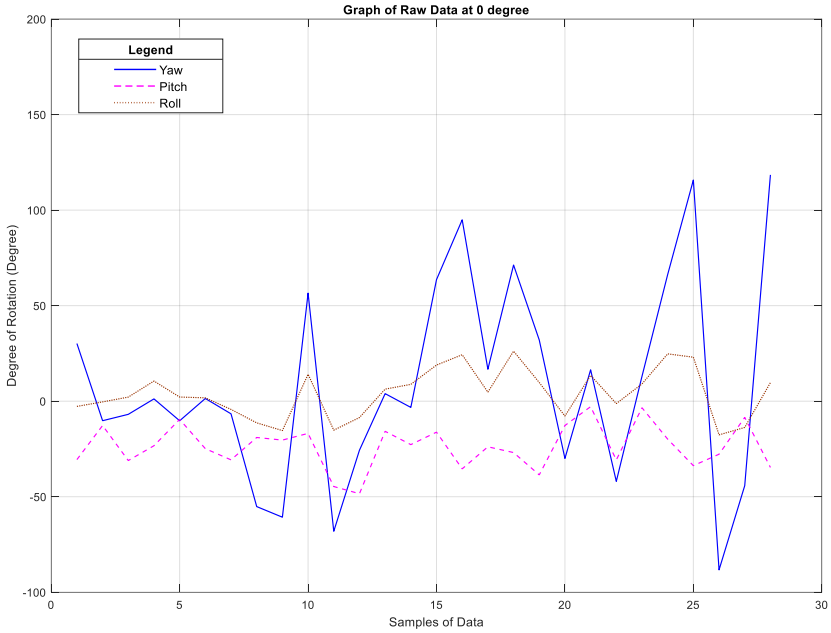


Figure 12. Graph of raw data at 0° angle.

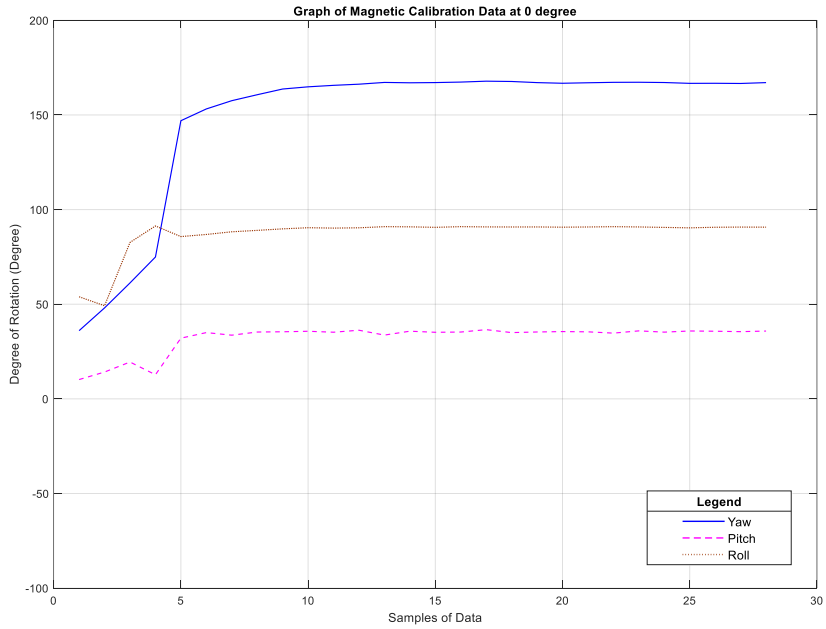


Figure 13. Graph of magnetic calibration data at 0° angle.

https://doi.org/10.31436/iiumej.v27i1.3654

Figure 13 shows a graph of magnetic calibration data at 0°, which yields reduced fluctuations relative to the raw data but still exhibits visible noise. This approach improves orientation accuracy by correcting magnetic field distortions. However, it remains limited in dynamic situations characterized by noise and drift, necessitating advanced filtering techniques to further improve performance.

Figure 14 shows EKF filter data at 0°, which outperforms magnetic calibration in terms of stability and accuracy. Anticipating and correcting sensor data using a mathematical model successfully lowers noise and drift. However, occasional departures indicate difficulties in managing complex underwater dynamics without additional noise compensation.

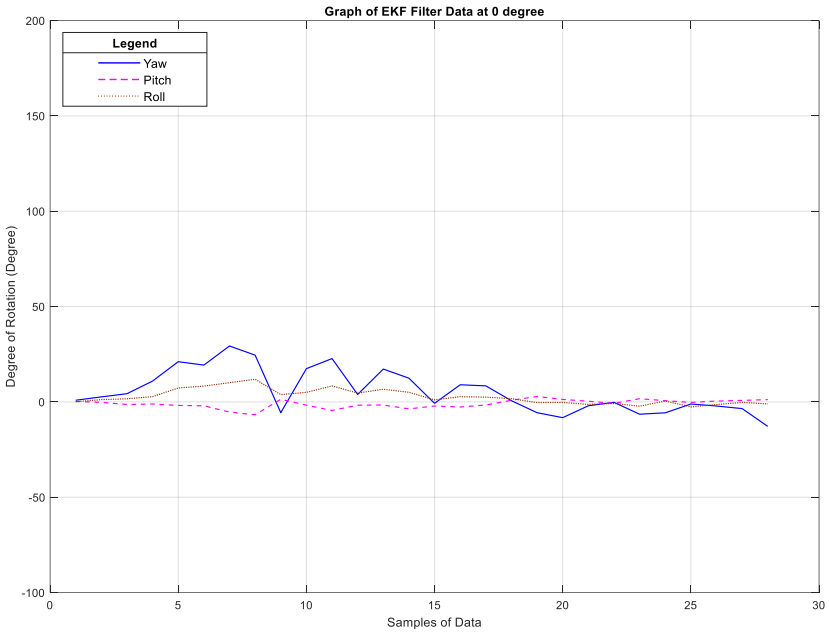


Figure 14. Graph of EKF filter data at 0° angle.

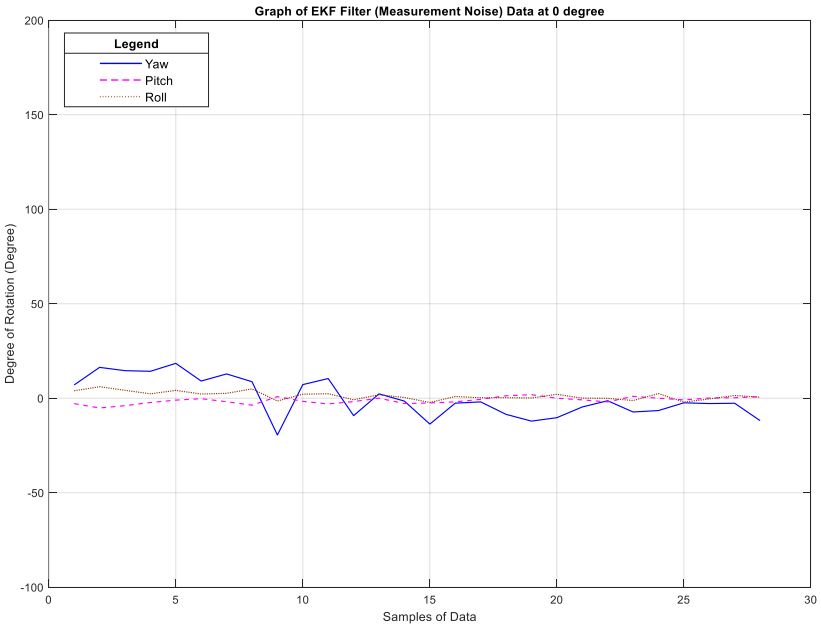


Figure 15. Graph of EKF filter with measurement noise data at 0° angle.

https://doi.org/10.31436/iiumej.v27i1.3654

Figure 15 shows a graph of the EKF filter with measurement noise at an angle of 0°, which is more refined than the standard EKF. By explicitly accounting for measurement noise, orientation calculations become more consistent and accurate. While successful, it nevertheless shows significant variability, particularly in demanding situations, indicating limitations in noise adaptation.

Figure 16 shows a graph of the EKF filter with process noise data at 0 degrees, demonstrating improved performance by resolving uncertainties in the AUV's motion model. This approach produces smoother orientation estimations than the EKF with measurement noise, but slight inconsistencies indicate an opportunity for improvement in highly dynamic or noisy situations.

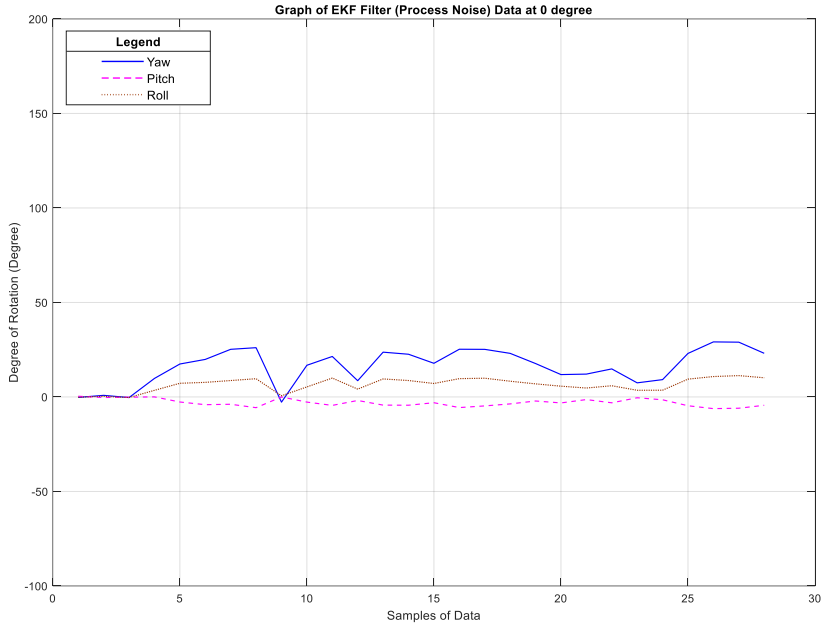


Figure 16. Graph of EKF Filter with Process Noise Data at 0 Degrees.

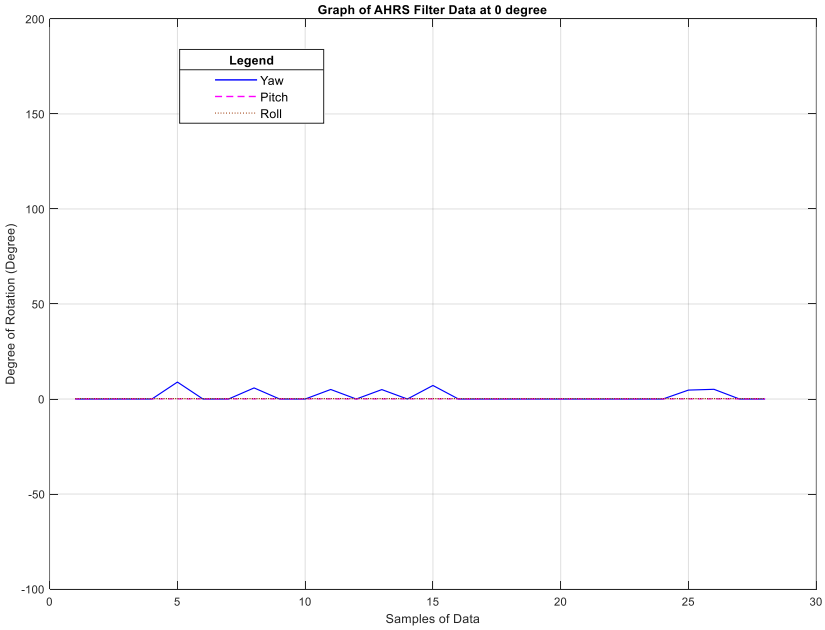


Figure 17. Graph of AHRS Filter data at 0° angle.

https://doi.org/10.31436/iiumej.v27i1.3654

Figure 17 shows an AHRS filter data plot at 0 degrees, exhibiting excellent stability and precision with minimal variation. This filter integrates multiple sensor inputs and dynamically adapts to changes, thereby mitigating noise and drift. Its high performance makes it well-suited for accurate orientation tracking in challenging underwater conditions.

Based on Table 1, the comprehensive evaluation of orientation data at 0 degrees using various filters and calibration techniques has provided valuable insights into their respective performance. The results indicate distinct characteristics for each method, which can significantly impact orientation accuracy and stability.

Table 1. Comparison of the analyzed filter performance at 0 Degrees.

	Raw Data	Magnetic Calibration	EKF Filter	EKF Filter (Measure. Noise)	EKF Filter (Process Noise)	AHRS Filter
Max	118.45	167.85	29.33	18.44	29.13	8.92
Min	-88.30	10.20	-12.85	-19.48	-6.17	0.00
Mean	-3.62	89.47	2.35	0.07	6.61	0.60
Std. Deviation	35.37	53.59	7.42	6.13	9.84	1.68

The Raw Data approach exhibits a wide range of values, with a substantial standard deviation of 35.37 degrees. This underscores the inherent noise and variability present in unfiltered measurements.

The Magnetic Calibration demonstrates notable deterioration, with a larger standard deviation of 53.59° and a less concentrated distribution of values. The EKF demonstrates further refinement, with a significantly lower standard deviation of 7.42° in the angle. The filtering process reduces noise and bias, thereby improving orientation estimates.

The application of the EKF with both measurement and process noise highlights the importance of noise modelling. The former achieves a minimal standard deviation of 6.13° in angle, whereas the latter exhibits a larger standard deviation of 9.84 °.

Remarkably, the AHRS Filter demonstrates exceptional precision, with a standard deviation of only 1.68 degrees. This technique significantly reduces noise, bias, and other sources of error, providing the most reliable orientation estimates.

4. CONCLUSION

The study on the impact of noise on the IMU sensor module revealed that environmental and operational noise significantly affect the accuracy and stability of sensor data. This underscores the need to incorporate effective noise-mitigation strategies to enhance the reliability of IMU data in underwater environments. Next, the development of a self-calibrated module for orientation estimation was a pivotal advancement.

This module was designed to continuously adjust its calibration parameters based on real-time data, effectively compensating for drift and sensor inaccuracies. The implementation of this self-calibration capability resulted in substantial improvements in orientation accuracy, thereby contributing to more stable and reliable AUV navigation. Comparison of various filtering techniques integrated into the INS demonstrated distinct differences in their ability to manage sensor noise and improve navigation accuracy.

To enhance AUV navigation, this study explored the crucial module of Inertial Sensor Self-Calibration. By meticulously investigating angle measurements, particularly those derived

from raw data and an AHRS filter, significant insights were gained into the potential of self-calibration techniques for AUV navigation.

The analysis began by establishing a baseline through the examination of raw data, thereby revealing the initial state of the measurements before calibration. Subsequently, the AHRS filter emerged as a standout performer, exhibiting the lowest standard deviation and variance. This finding underscored the AHRS filter's capacity to enhance the accuracy of angle measurements, a pivotal factor for ensuring dependable navigation for AUVs operating in complex underwater environments.

The AHRS filter's exceptional performance supports its utility in addressing the inherent challenges of underwater navigation, where precision is paramount. As AUVs continue to evolve and diversify in their applications, the insights gained from this study can serve as a cornerstone for further advancements. By harnessing the potential of self-calibration modules such as the AHRS filter, new avenues for oceanic exploration, environmental monitoring, and underwater interventions are enabled.

ACKNOWLEDGEMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through MDR (vot Q693) and GPPS (vot Q647).

REFERENCES

[1] Petillot, Y.R., Antonelli, G., Casalino, G. and Ferreira, F., Underwater robots: From remotely operated vehicles to intervention-autonomous underwater vehicles. IEEE Robotics & Automation Magazine, 2019, 26(2), pp. 94-101.

[2] Agarwala, N., Monitoring the ocean environment using robotic systems: Advancements, trends, and challenges. Marine Technology Society Journal, 2020, 54(5), pp. 42-60.

[3] Zhang, B., Ji, D., Liu, S., Zhu, X. and Xu, W., Autonomous underwater vehicle navigation: a review. Ocean Engineering, 2023, 273, p. 113861.

[4] Ye, X., Song, F., Zhang, Z. and Zeng, Q., A Review of Small UAV Navigation System Based on Multi-source Sensor Fusion. IEEE Sensors Journal, 2023, 23 (17), pp. 18926-18948.

[5] Wang, J., Tang, Y., Jin, S., Bian, G., Zhao, X. and Peng, C., A Method for Multi-Beam Bathymetric Surveys in Unfamiliar Waters Based on the AUV Constant-Depth Mode. Journal of Marine Science and Engineering, 2023, 11(7), p. 1466.

[6] Miller, L.M. and Brizzolara, S., Effect of amplitudes and frequencies on Virtual Planar Motion Mechanism of AUVs, Part I: Forces, moments and hydrodynamic derivatives. Ocean Engineering, 2023, 286, p. 115512.

[7] Basil, N., Alqaysi, M.E., Deveci, M., Albahri, A.S., Albahri, O.S. and Alamoodi, A.H., Evaluation of autonomous underwater vehicle motion trajectory optimization algorithms. Knowledge-Based Systems, 2023, p. 110722.

[8] González-García, J., Gómez-Espinosa, A., Cuan-Urquizo, E., García-Valdovinos, L.G., Salgado-Jiménez, T. and Escobedo Cabello, J.A., Autonomous underwater vehicles: Localization, navigation, and communication for collaborative missions. Applied sciences, 2020, 10(4), p. 1256.

[9] Boiko, J., Svachii, O., & Kałaczyński, T., Control of unmanned aerial vehicle modules based on inertial measurement system and MEMS sensor. MATEC Web of Conferences, 2024, 391, p. 01005.

[10] Brommer, C., Fornasier, A., Scheiber, M., Delaune, J., Brockers, R., Steinbrener, J. and Weiss, S., The INSANE dataset: Large number of sensors for challenging UAV flights in Mars analog, outdoor, and out-/indoor transition scenarios. The International Journal of Robotics Research, 2024, 43(8), pp. 1083-1113.

[11] Yang, K., Zhang, Z., Cui, R. and Yan, W., Acoustic-optic assisted multisensor navigation for autonomous underwater vehicles. Ocean Engineering, 2024, 297, p. 117139.

[12] Wang, C., Cheng, C., Yang, D., Pan, G. and Zhang, F., AUV planning and calibration method considering concealment in uncertain environments. Frontiers in Marine Science, 2023, 10, p. 1228306

[13] Han, S., Meng, Z., Zhang, X. and Yan, Y., Hybrid deep recurrent neural networks for noise reduction of MEMS-IMU with static and dynamic conditions. Micromachines, 2021, 12(2), p. 214

[14] Vasudev, K.L., Review of autonomous underwater vehicles. Autonomous Vehicles, 2020, pp. 31-48.

[15] Sánchez, P.J.B., Papaelias, M. and Márquez, F.P.G., Autonomous underwater vehicles: Instrumentation and measurements. IEEE Instrumentation & Measurement Magazine, 2020, 23(2), pp. 105-114

[16] Stutters, L., Liu, H., Tiltman, C. and Brown, D.J., Navigation technologies for autonomous underwater vehicles. IEEE Transactions on Systems, Man, and Cybernetics, 2008, 38(4), pp. 581-589.

[17] Suwandi, B., Kitasuka, T. and Aritsugi, M., Vehicle vibration error compensation on IMU-accelerometer sensor using adaptive filter and low-pass filter approaches. Journal of Information Processing, 2019, 27, pp. 33-40.

[18] Werries, A. and Dolan, J., Adaptive Kalman filtering methods for low-cost GPS/INS localization for autonomous vehicles, CMU-RI-TR-16-18, 2016, pp. 1-6.

[19] Quoc, D.D., Sun, J. and Tan, N.N., Sensor fusion based on complementary algorithms using MEMS IMU. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2015, 8(2), pp. 313-324

[20] Woo, R., Yang, E.J. and Seo, D.W., A fuzzy-innovation-based adaptive Kalman filter for enhanced vehicle positioning in dense urban environments. Sensors, 2019, 19(5), p. 1142.

https://doi.org/10.31436/iiumej.v27i1.3717

GNN-based Skyline Query Processing for Large-Scale and Incomplete Graphs

HASAN KHAIR ADZMAN, RAINI HASSAN*, DINI OKTARINA DWI HANDAYANI

Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia

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

(Received: 5 February 2025; Accepted: 1 October 2025; Published online: 12 January 2026)

ABSTRACT: Skyline queries are crucial in database management, selecting optimal points from multi-dimensional datasets based on dominance relationships. They are widely used in decision-making, recommendation systems, and data filtering. However, traditional skyline algorithms struggle with large volumes and missing data, leading to high computational costs and inefficiencies. This research proposes a hybrid approach that integrates the ISkyline dominance graph technique with Graph Neural Networks (GNNs) to improve skyline query performance under such conditions. The GNN component is utilized to predict skyline tuples in the presence of missing or incomplete data. Evaluation on both synthetic and real-world datasets demonstrates improved accuracy and efficiency compared with established methods such as ISkyline, SIDS, and OIS. This research demonstrates the potential to improve query processing efficiency and to support applications in e-commerce, finance, and smart data systems.

ABSTRAK: Pertanyaan latar langit adalah penting dalam pengurusan pangkalan data, iaitu dengan memilih titik optimum daripada set data berbilang dimensi berdasarkan hubungan dominasi. Ia digunakan secara meluas dalam membuat keputusan, sistem pengesyoran, dan penapisan data. Walau bagaimanapun, algoritma latar langit tradisional bergelut dengan kuantiti data yang besar dan data menghilang, membawa kepada peningkatan kos pengiraan dan ketidakcekapan. Kajian ini mencadangkan pendekatan hibrid yang mengintegrasi teknik graf penguasaan ISkyline dengan Rangkaian Graf Neural (GNNs) bagi meningkatkan prestasi pertanyaan latar langit berkeadaan sedemikian. Komponen GNN digunakan bagi meramalkan tupel latar langit dengan kehadiran data menghilang atau tidak lengkap. Penilaian pada kedua-dua set data sintetik dan dunia nyata menunjukkan peningkatan ketepatan dan kecekapan jika dibandingkan dengan kaedah sedia ada seperti ISkyline, SIDS dan OIS. Kajian ini menunjukkan potensi bagi mencipta pemprosesan pertanyaan yang lebih cekap, menyokong aplikasi e-dagang, kewangan dan sistem data pintar.

KEY WORDS: *Skyline query processing, Graph Neural Networks (GNNs), Incomplete data, Pareto optimality, Machine learning.*

1. INTRODUCTION

Skyline query processing is widely used in multi-criteria decision-making applications, including route planning, product recommendation, and health diagnostics. However, existing skyline methods face major challenges when applied to large and incomplete datasets, conditions that are increasingly common in real-world scenarios.

This research introduces a hybrid approach that combines Graph Neural Networks (GNNs) with the ISkyline dominance graph technique to enhance skyline query performance. The proposed method is designed to handle missing data and scale efficiently, thereby improving the prediction of skyline tuples even in complex, incomplete environments. Experimental results on both synthetic and real-world datasets demonstrate that this method outperforms state-of-the-art techniques in accuracy and efficiency.

Skyline queries aim to retrieve data records that are not dominated by any others across multiple dimensions, often referred to as Pareto-optimal points. While powerful, these queries are computationally expensive, particularly when applied to large graph-based datasets or those with incomplete attributes. Existing solutions, such as ISkyline and SIDS, aim to address scalability, but they still struggle to predict under uncertainty or in the presence of data loss.

Recent advances in deep learning, particularly Graph Neural Networks (GNNs), offer promising capabilities for learning from structured and incomplete data. By integrating GNNs into the skyline processing workflow, the proposed method leverages graph-based feature learning to support more robust and intelligent skyline selection.

Skyline queries are essential for identifying optimal data points from multi-dimensional datasets based on dominance relationships. However, traditional skyline query algorithms face significant limitations when processing large-scale, incomplete graph datasets. These methods often encounter challenges related to scalability, computational overhead, and inefficiencies in handling dynamic database environments [1], [2]. Furthermore, approaches such as Bucket and ISkyline struggle to integrate missing data effectively, resulting in suboptimal accuracy and high processing costs [3]. Current solutions lack a comprehensive framework that integrates state-of-the-art advances in machine learning, particularly Graph Neural Networks (GNNs), which have the potential to address these challenges by improving scalability, accuracy, and adaptability [4].

Despite the promising capabilities of machine learning, including its ability to model complex relationships in graph-structured data, its application in optimizing skyline queries remains underexplored [5]. Existing research does not adequately leverage the dynamic adaptability and efficiency of machine learning techniques, leaving a critical gap in addressing the computational and data-handling shortcomings of traditional methods. This study aims to bridge these gaps by introducing a novel framework that integrates Pareto optimality principles with advanced machine learning methods, providing robust solutions for scalable and efficient skyline computation in real-world settings.

Despite the potential of integrating Graph Neural Networks (GNNs) with skyline query processing, several limitations must be acknowledged. First, GNN-based approaches are resource-intensive, often resulting in high memory usage and longer query response times, particularly when dominance graphs are dense. Second, these models face challenges in interpretability, as it can be difficult to explain how dominance relationships are learned and applied in classification. Third, the computational requirements may restrict scalability when applied to large-scale, real-world datasets using modest hardware. These limitations frame the motivation for this study and highlight the importance of balancing accuracy improvements with efficiency and practical deployment considerations.

The research aims to accomplish three objectives. First, to develop a unified framework that integrates Pareto optimality principles with advanced machine learning techniques, particularly Graph Neural Networks (GNNs), to improve skyline query processing over large-scale and attribute-incomplete graphs. Second, to evaluate the performance of the proposed framework across real-world and synthetic datasets using comprehensive metrics, including

accuracy (target > 99%), F1-score (target > 99%), AUC-ROC (target > 99%), query response time, and memory usage, with a focus on ensuring scalability, efficiency, and adaptability in dynamic environments. Third, to compare the effectiveness of the proposed framework against traditional skyline algorithms and alternative machine learning models to identify the most suitable method for skyline query processing on incomplete graph-structured data.

Additionally, this research contributes significantly to academia and real-world applications by introducing a novel framework that integrates Pareto optimality with Graph Neural Networks (GNNs) to enhance skyline query processing in large-scale, incomplete datasets. Traditional methods often struggle with scalability and missing data, but this hybrid approach leverages Pareto optimality to identify non-dominated points and GNNs to learn from graph-structured, incomplete data. Their complementary strengths result in improved accuracy, F1-score, and AUC-ROC. The research also establishes standardized benchmarks using synthetic and real datasets, evaluating performance using metrics such as accuracy, F1-score, AUC-ROC, query response time, and memory usage. Practically, the solution is scalable and adaptive, benefiting industries such as e-commerce, finance, and smart infrastructure by enabling efficient, real-time decision-making even with incomplete data.

2. RELATED WORKS

Skyline queries identify optimal choices from a large dataset, such as selecting the top products that are inexpensive, fast, and well-rated. GNNs are useful when product information is incomplete or when relationships are important.

Prior research on skyline query processing has produced several key algorithms. Borzsony et al. [6] introduced the skyline operator to filter non-dominated points, inspiring methods like ISkyline [2], which partitions incomplete data using bitmaps, and SIDS [7], which applies sorted-based pruning. More recent work by Wang et al. [1] proposed Skyline Preference Queries (SPQ) for large, incomplete datasets.

Despite these advances, most traditional methods struggle with scalability, particularly in the presence of missing data. Few studies have explored machine learning approaches for skyline computation. Recent developments in Graph Neural Networks (GNNs) show promise for learning dominance structures directly from incomplete graphs [4], but their application in skyline queries remains underexplored. This research builds upon these foundations, proposing a hybrid GNN-based framework to address the identified gaps.

3. METHODOLOGY

3.1. Introduction

This section details the methodologies employed in this study, drawing on the design science research cycle as outlined by Hevner [8]. The approach integrates three cycles to enhance the identification and comprehension of design science research initiatives. Figure 1 illustrates the adapted design science research framework based on the research of von Brocke et al. [9].

https://doi.org/10.31436/iiumej.v27i1.3717

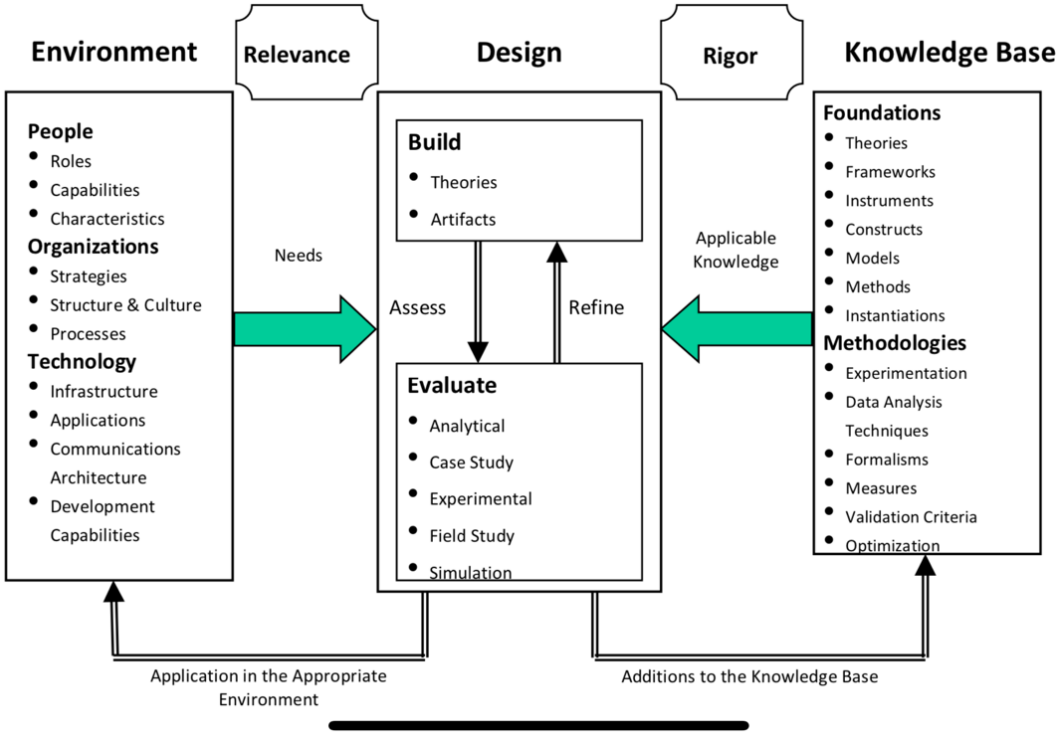


Figure 1. Design Science Research Framework [19]

The Relevance Cycle connects the research project's contextual environment with design science activities, ensuring that the needs and requirements for achieving the research objectives are adequately identified. The Rigor Cycle links design science activities to a knowledge base of scientific principles, expertise, and prior experiences that inform and guide the research process [10]. At the core is the Design Cycle, which focuses on developing and evaluating design artifacts and research processes. This cycle plays a pivotal role in describing the research activities. Figure 2 illustrates the sequential flow of the research process within this framework.

The research flow begins with a literature review that provides a foundational understanding of traditional skyline algorithms, machine learning frameworks, and graph-based methodologies. The process then advances to testing synthetic datasets as the initial step, followed by real datasets. Both workflows incorporate data preprocessing, including tasks such as data normalization, handling missing values, and splitting datasets into training, validation, and test sets. For synthetic datasets, the research involves selecting traditional algorithms (e.g., ISkyline) and machine learning frameworks (e.g., Graph Neural Networks), and integrating them to leverage the strengths of both approaches. The unified framework, along with traditional algorithms, is tested extensively, and the performance of these methods is compared using metrics such as processing time, memory usage, and accuracy. Results are analyzed to evaluate the effectiveness of the unified approach. Similarly, the real datasets testing follows the same workflow but includes an additional step of tuning the unified machine learning framework to optimize its performance for real-world applications. This tuning involves adjusting hyperparameters and evaluating alternative architectures (e.g., GraphSAGE). Finally, the results from both workflows are consolidated, discussed comprehensively, and used to propose a final algorithm that demonstrates superior performance in terms of scalability, robustness, and adaptability. This expanded explanation ensures a clear understanding of how each step in the methodology contributes to achieving the research objectives.

https://doi.org/10.31436/iiumej.v27i1.3717

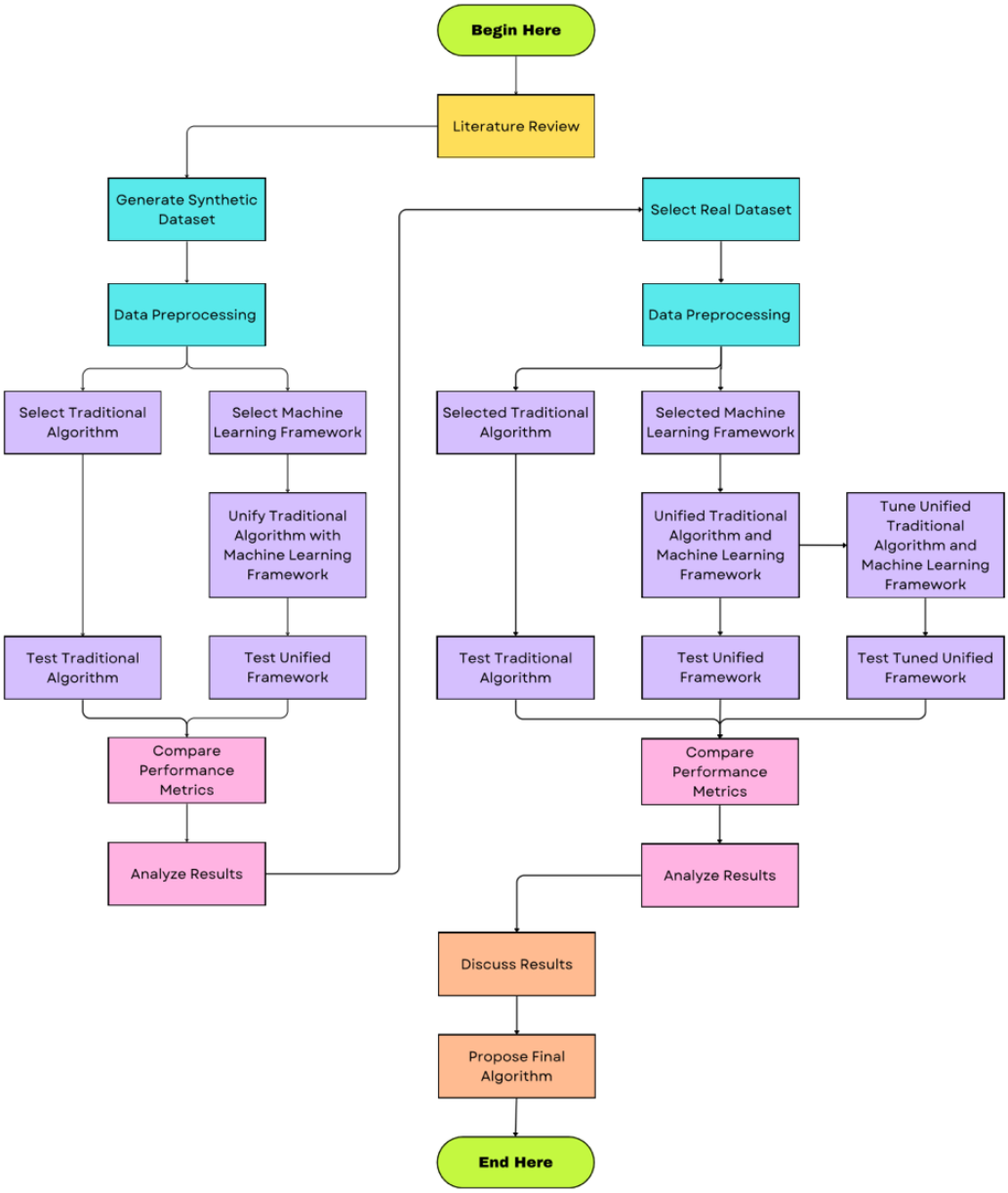


Figure 2. Design Cycle

3.2. Proposed Method: GNN + ISkyline

This research proposes a hybrid method that integrates the traditional skyline query algorithm ISkyline with a Graph Neural Network (GNN) to address attribute-level incompleteness in graph-structured data. The goal is to identify dominant relationships among data points and to improve skyline prediction in incomplete datasets using graph-based deep learning.

The proposed methodology begins by computing the ground-truth skyline from normalized and imputed data, then applies a modified ISkyline algorithm that handles missing values by skipping comparisons with None entries. Dominance is determined through a custom function that checks if one point is greater than or equal to another across all comparable attributes and strictly greater in at least one. Each data point is then labeled as skyline or non-skyline. To address class imbalance, skyline points are oversampled to match non-skyline points, and a directed dominance graph is constructed in which nodes represent products and

https://doi.org/10.31436/iiumej.v27i1.3717

edges represent dominance relationships. This graph feeds into a 3-layer Graph Convolutional Network (GCN), which uses normalized attributes as node features and binary skyline labels for supervision. The model is trained using binary cross-entropy loss with class imbalance adjustments and evaluated over 200 epochs on standard classification and efficiency metrics. Designed to handle incomplete data, this hybrid approach combines ISkyline logic with GNN-based pattern recognition to infer skyline membership effectively. Figure 3 below illustrates the full pipeline of this hybrid methodology.

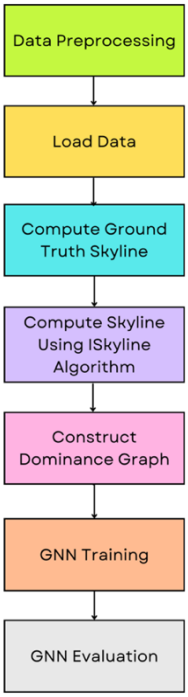


Figure 3. Proposed GNN + ISkyline Framework

By structuring the methodology in this modular fashion, the proposed approach provides transparency in its design while allowing each step to be assessed and improved independently. The integration of dominance relations with graph learning bridges classical skyline computation and modern neural representation learning, yielding a more scalable and accurate approach to handling incomplete multi-criteria datasets.

3.3. Dataset Preparation

This study evaluates the proposed framework using three real-world datasets: CoIL 2000 [11], NBA Stats [12], and MovieLens [13], selected for their diversity in domain, data volume, and feature complexity. These datasets are commonly cited in the literature on multi-criteria decision-making, incomplete data handling, and recommendation systems. To simulate real-world scenarios involving incomplete information, 20% attribute-level incompleteness was introduced into each dataset, either by randomly removing attributes or by building on pre-existing missing values.

CoIL 2000, with 5,822 complete customer records and 86 attributes divided into sociodemographic data (attributes 1–43) and product ownership (attributes 44–86). The NBA Stats dataset comprises 18,381 player regular-season records across 17 features, spanning the period 1946 to 2005. Features include standard demographics and performance statistics such as field goals, free throws, assists, and points per game. The MovieLens dataset consists of

https://doi.org/10.31436/iiumej.v27i1.3717

1,000,209 ratings from 6,040 users for approximately 3,900 movies. Each user contributed at least 20 ratings on a 5-point scale. Each row of the dataset records a user ID, movie ID, rating, and timestamp.

Table 1 summarizes the key characteristics of each dataset and details the approach taken to introduce and define incompleteness. The focus is on attribute-level incomplete graphs, in which node features may be partially missing while maintaining structural relationships via similarity-based edges.

Table 1. Summary of Real-World Datasets and Incompleteness Characteristics

Dataset	No. of Nodes	No. of Columns	Original Completeness	Incompleteness Introduced	Incomplete Graph Type	Graph Representation
CoIL 2000	5,822	86	Fully complete	20% of attribute values were removed randomly	Attribute-level incompleteness	Nodes: customers; Edges: feature similarity
NBA Stats	18,381	17	Partially incomplete	Removed attribute values to reach 20% of total missing values	Attribute-level incompleteness	Nodes: customers; Edges: feature similarity
MovieLens	1,000,209	4	Fully complete	20% of attribute values were removed randomly	Attribute-level incompleteness	Nodes: customers; Edges: feature similarity

To evaluate the robustness of the proposed framework under varying levels of attribute-level incompleteness, four synthetic e-commerce datasets were generated, each with 5,000 product nodes and six attributes: 1) Product ID, 2) Price, 3) Rating, 4) Availability, 5) Shipping Time, and 6) Category. Products in the same category are connected, forming 1,249,533 edges, yielding an edge density of approximately 0.1 (9.998%) of the possible 12,497,500 node-to-node connections. The three datasets introduce random incompleteness across all attributes at 10%, 50%, and 90%, enabling controlled robustness testing under progressively increasing levels of missing data, as shown in Table 2.

Table 2. Summary of Synthetic E-Commerce Datasets and Incompleteness Characteristics

Dataset	No. of Nodes	No. of Edges	Max Possible Edges	Edge Density	Attribute Missingness
Synthetic-10%	5,000	1,249,533	12,497,500	~0.1 (9.998%)	10%
Synthetic-50%	5,000	1,249,533	12,497,500	~0.1 (9.998%)	50%
Synthetic-90%	5,000	1,249,533	12,497,500	~0.1 (9.998%)	90%

These datasets provide a comprehensive testing environment for benchmarking the robustness of skyline algorithms and the proposed framework. By varying the degree of incompleteness while holding the graph topology constant, the research ensures that performance differences across methods are directly attributable to their handling of missing data. This controlled design is essential for deriving reliable conclusions about algorithmic behavior in incomplete graph settings.

3.4. Baseline Models

To evaluate the effectiveness of the proposed machine learning-based framework, three traditional skyline query algorithms, ISkyline, SIDS, and OIS, were selected as baselines due

to their prominence in handling incomplete data and large-scale graphs. ISkyline uses bitmap representations and shadow skylines to reduce dominance comparisons and manage missing values effectively. SIDS employs a round-robin sorting mechanism to prune dominated tuples, offering efficiency for static datasets with partial incompleteness. OIS, although not designed for missing data, minimizes I/O overhead and performs well in large-scale environments. These diverse, well-established methods provide a solid benchmark for assessing improvements in robustness, scalability, and processing efficiency offered by the proposed learning-based approach.

3.5. Evaluation and Benchmarking

To assess the performance of the proposed GNN + ISkyline framework and baseline models in classifying skyline and non-skyline points, five key metrics are used: 1) accuracy, 2) precision, 3) recall, 4) F1-score, and 5) AUC-ROC. These metrics are crucial in imbalanced datasets, where skyline points are often the minority. Eq. (1) shows that the accuracy measures overall correctness:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Eq. (2) shows that the precision or the ratio of correctly predicted skyline points to all predicted skyline points, is given by:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Eq. (3) shows that the recall reflects the model's ability to detect actual skyline points:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Eq. (4) shows that the F1-score balances precision and recall using the harmonic mean:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

AUC-ROC measures the model's ability to distinguish between classes across thresholds, with 1.0 indicating perfect classification and 0.5 representing random guessing. These metrics, computed on the test set, provide a comprehensive evaluation of model performance, particularly under varying degrees of data incompleteness.

Query response time is a critical performance metric for evaluating the efficiency of skyline query processing methods, particularly when dealing with large and incomplete datasets. In this study, query response time is defined as the total time from the initiation of query processing to the generation of the final results.

Memory usage is an essential metric for evaluating the computational efficiency of skyline query processing methods, particularly when applied to large-scale, incomplete datasets. It provides insights into the resource intensity of each technique, which is particularly relevant for deployment in memory-constrained environments. In this study, memory usage is measured as the peak memory usage observed during the complete execution cycle of each method.

3.6. Setup and Configuration

All experiments were conducted locally on a MacBook Air with an Apple M2 chip (8-core CPU, 8-core GPU, 16-core Neural Engine), 8 GB unified memory, and a 256 GB SSD running macOS Sequoia. Despite modest hardware, the system efficiently supported the framework's

https://doi.org/10.31436/iiumej.v27i1.3717

implementation in Python 3.11 using PyTorch (MPS backend), PyTorch Geometric, and Scikit-learn, with data handled via Pandas and NumPy, and visualizations done using Matplotlib and Seaborn. Model and batch optimizations enabled the smooth execution of experiments on both synthetic and real-world datasets, ensuring reproducibility and consistent performance across all tests.

4. RESULTS AND DISCUSSIONS

4.1. Experimental Setup and Dataset Description

To ensure the validity of the proposed GNN–ISkyline framework, experiments were conducted using both real-world and synthetic datasets. A detailed description of these datasets and the preprocessing steps is provided in Section 3. However, this section reiterates key aspects relevant to the experimental pipeline.

The real-world datasets used in this study include CoIL 2000, NBA Stats, and MovieLens. CoIL 2000 is a customer dataset with 5,822 records and 86 attributes, containing sociodemographic data and product ownership, originally completed before 20% of attribute values were removed to simulate incompleteness. The NBA Stats dataset contains season-based player statistics from 1946–2005, which are inherently partially incomplete and were further modified by removing attribute values to achieve 20% incompleteness. The MovieLens dataset contains over 1 million user ratings, initially complete, with 20% of attributes removed to simulate missing data. All datasets were transformed into graphs where nodes represent entities (e.g., customers, players, or users) and edges represent similarity based on features.

In addition to these real-world datasets, four synthetic e-commerce datasets were generated, each containing 5,000 product nodes. These datasets were constructed with six attributes: product ID, price, rating, availability, shipping time, and category. The three synthetic datasets were generated with uniform incompleteness levels of 10%, 50%, and 90%, respectively, by randomly removing attribute values across all columns.

4.2. Method Execution

The experiment involved several stages, beginning with data preprocessing, in which missing values were imputed with zeros and attributes were normalized using MinMaxScaler. Skyline labels were generated via dominance checks, and oversampling skyline points balanced the dataset. A dominance graph was then constructed with directed edges from dominating to dominated nodes, along with the feature and label tensors, and was fed into a 3-layer GCN model. The model was trained on 80% of the data and tested on the remaining 20%, with performance evaluated using accuracy, precision, recall, F1 score, and AUC-ROC, alongside query response time and peak memory usage as efficiency metrics.

4.3. Comparison of Algorithms Using Synthetic Data with 10%, 50%, and 90% Missingness

To provide a comprehensive evaluation, the results compare both standalone algorithms (GNN, ISkyline, SIDS, OIS) and their hybrid counterparts (e.g., GNN + ISkyline). The standalone models represent the baseline performance of traditional skyline methods or machine learning methods applied independently. In contrast, the hybrid models demonstrate that combining dominance-based algorithms with GNNs can enhance predictive accuracy. This comparison highlights not only which method performs best but also the trade-offs between classical efficiency and deep learning adaptability. The following results address the objectives 1, 2, and 3 stated in the Introduction.

https://doi.org/10.31436/iiumej.v27i1.3717

Table 3 presents a detailed evaluation of various algorithms under synthetic data with increasing levels of missingness (10%, 50%, and 90%), revealing how well each model handles incomplete information. Across all levels, GNN + SIDS consistently performs best on predictive metrics, achieving near-perfect scores in accuracy, precision, recall, F1-score, and AUC-ROC, even with 90% missing data, while maintaining modest memory usage and fast response times. ISkyline performs surprisingly well in terms of accuracy and precision but suffers from extremely low recall and F1-score, indicating poor sensitivity. Interestingly, GNN + ISkyline also maintains high performance but at the cost of very high memory usage and response time, especially under 90% missingness. OIS remains robust with high F1-scores and AUC-ROC, as missingness increases.

Table 3. Comparison of Algorithms Using Synthetic Data with 10%, 50%, and 90% Missingness

10% Missingness							
Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**	**Query Response Time (s)**	**Peak Memory Usage (KB)**
GNN	0.5020	0.5020	0.9861	0.6667	0.4842	139.2902	921.39
ISkyline	0.9698	1.0000	0.0195	0.0382	0.5097	0.1603	168.90
GNN + ISkyline	0.9919	0.9841	1.0000	0.9920	0.9878	843.2566	477584.90
SIDS	0.9916	0.0455	1.0000	0.0870	0.9958	0.4705	2971.04
GNN + SIDS	0.9928	0.9860	1.0000	0.9929	0.9943	749.9065	478148.27
OIS	0.9728	0.0145	1.0000	0.0286	0.9864	0.7676	424.48
GNN + OIS	0.0308	0.0308	1.0000	0.0598	0.0009	147.5776	105397.89
50% Missingness							
Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**	**Query Response Time (s)**	**Peak Memory Usage (KB)**
GNN	0.4930	0.4930	1.0000	0.6604	0.4919	608.6592	912.55
ISkyline	0.9738	0.0000	0.0000	0.0000	0.5000	0.1257	168.68
GNN + ISkyline	0.9931	0.9870	1.0000	0.9935	0.9956	1147.2394	1027135.04
SIDS	0.9996	0.9957	1.0000	0.9978	0.9998	7.3685	2976.11
GNN + SIDS	0.9995	0.9990	1.0000	0.9995	1.0000	990.1274	1027707.01
OIS	0.9966	0.9647	1.0000	0.9820	0.9981	0.8260	426.06
GNN + OIS	0.0262	0.0262	1.0000	0.0511	0.0007	162.6048	381806.19
90% Missingness							
Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**	**Query Response Time (s)**	**Peak Memory Usage (KB)**
GNN	0.5110	0.4364	0.0498	0.0894	0.5216	1041.0278	919.07
ISkyline	0.9906	1.0000	0.1296	0.2295	0.5648	0.0365	168.02
GNN + ISkyline	0.9980	0.9960	1.0000	0.9980	1.0000	4667.7921	2157517.52
SIDS	1.0000	1.0000	1.0000	1.0000	1.0000	105.9054	3000.64
GNN + SIDS	0.9995	0.9990	1.0000	0.9995	1.0000	4543.4297	2157766.90
OIS	0.9936	0.9905	1.0000	0.9952	0.9904	6.8788	416.32
GNN + OIS	0.9504	0.1723	0.9444	0.2914	0.9757	191.4423	622770.62

In contrast, GNN + OIS, despite strong recall, underperforms substantially in accuracy, precision, F1-score, and AUC-ROC at lower missingness levels, recovering only somewhat at 90%. This behavior likely arises because OIS does not handle missing attributes well without pre-filtering, and GNNs overfit to the sparse dominance structure, leading to overprediction.

https://doi.org/10.31436/iiumej.v27i1.3717

Additionally, the simple graph structure generated by OIS may lack sufficient complexity for the GNN to extract meaningful patterns, causing ineffective learning.

By contrast, ISkyline constructs richer dominance relationships that provide the GNN with more informative features, improving classification quality. However, this richness comes at a high computational price, since storing and traversing large dominance graphs significantly increases both memory usage and query response time. This trade-off indicates that GNN + ISkyline is highly accurate but not resource-efficient. In contrast, GNN + OIS offers faster runtime and lower memory use but struggles with predictive performance due to a weaker graph representation.

4.4. Comparison of Algorithms Using CoIL 2000

The comparison between standalone and hybrid models is critical here, as it demonstrates how incorporating GNN learning alters the behavior of traditional skyline methods. Standalone algorithms such as ISkyline and OIS reflect their inherent design strengths and weaknesses, whereas hybrids demonstrate the potential improvements, along with the additional costs, of integrating neural network learning. This helps establish whether hybridization truly adds value over efficient standalone baselines.

Table 4 presents a comparative analysis of multiple skyline query processing algorithms using the CoIL 2000 dataset. The results show that traditional methods like ISkyline and SIDS perform poorly in terms of accuracy and F1-score, despite ISkyline achieving perfect precision due to its overly conservative predictions. In contrast, GNN-based models, particularly when combined with optimized skyline algorithms such as OIS, significantly outperform alternatives. The GNN + OIS approach achieves near-perfect performance, with accuracy (0.9698), F1-score (0.9846), and AUC-ROC (0.9950), while maintaining reasonable memory usage. Notably, the standalone OIS algorithm also delivers exceptional results (accuracy of 0.9796) with minimal query time (49.52s), suggesting its standalone strength. However, integrating GNN with OIS marginally increases resource usage but yields the best overall performance. This demonstrates that hybrid models that leverage GNNs with optimized skyline strategies provide the most effective and scalable solution for skyline queries on incomplete data.

Table 4. Comparison of Algorithms Using CoIL 2000

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN	0.4884	0.4881	0.8949	0.6316	0.5016	4.6432	336.96
ISkyline	0.0076	1.0000	0.0012	0.0024	0.5006	898.8737	16020.15
GNN + ISkyline	0.7333	0.5556	1.0000	0.7143	1.0000	229.6736	19673.39
SIDS	0.0064	0.0000	0.0000	0.0000	0.5000	1700.9073	70757.20
GNN + SIDS	0.7870	0.7297	0.9123	0.8109	0.8642	724.7528	143958.74
OIS	0.9796	0.9569	1.0000	0.9780	0.9813	49.5192	9158.45
GNN + OIS	0.9698	1.0000	0.9696	0.9846	0.9950	338.3995	38407.19

OIS is inherently optimized for I/O efficiency, which explains why OIS exhibits lower query times and memory consumption than ISkyline. Although GNN + ISkyline achieves a

https://doi.org/10.31436/iiumej.v27i1.3717

strong AUC-ROC, its memory footprint is an order of magnitude higher, making it less practical for real-time applications.

Figure 4 compares the accuracy of various algorithms and their combinations for skyline query processing. Standalone traditional algorithms such as ISkyline and SIDS perform poorly, with extremely low accuracies of 0.0076 and 0.0064, respectively. The GNN alone performs better, achieving an accuracy of 0.4884, indicating its ability to model graph structures but lacking in domain-specific optimization. However, combining GNNs with traditional methods significantly boosts performance. GNN + ISkyline and GNN + SIDS achieve much higher accuracies of 0.7333 and 0.7870, respectively. Notably, the OIS algorithm outperforms all with an accuracy of 0.9796, and when paired with GNNs (GNN + OIS), it maintains a comparably high performance at 0.9698. This indicates that hybrid models, particularly those that incorporate OIS, are the most effective for accurate skyline query processing on complex or incomplete datasets.

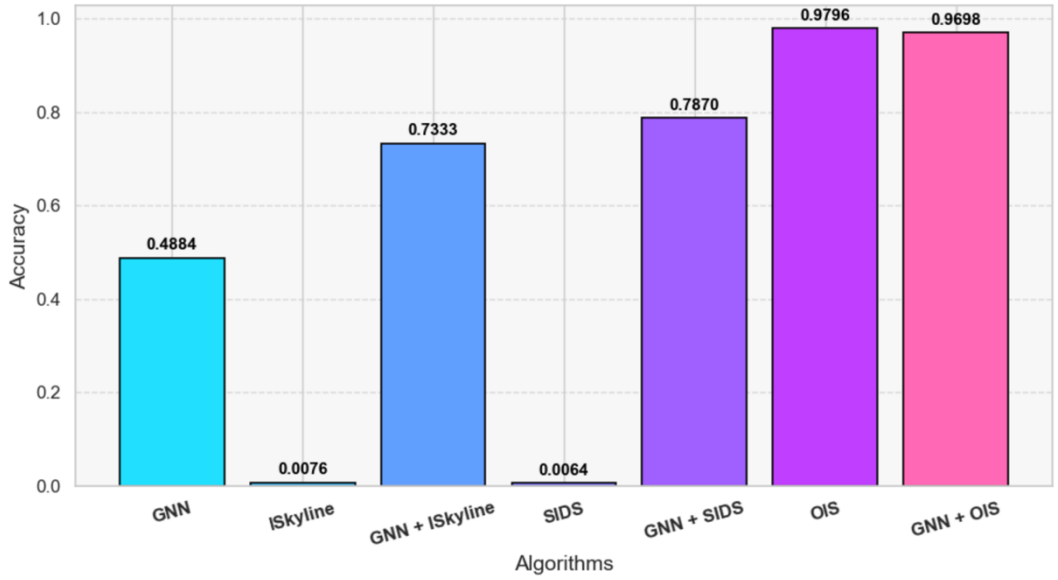


Figure 4. Accuracy Comparison of Algorithms Using CoIL 2000

Figure 5 reveals significant disparities in the performance of different algorithms used for skyline query processing. Traditional methods like ISkyline and SIDS show near-zero effectiveness, with F1-scores of 0.0024 and 0.0000, respectively, indicating their poor balance between precision and recall. The standalone GNN achieves a moderate F1-score of 0.6316, suggesting reasonable performance in capturing skyline points despite data incompleteness. Notably, combining GNN with classical methods significantly enhances performance: GNN + ISkyline achieves 0.7143, and GNN + SIDS further improves to 0.8109. The top-performing models are OIS and GNN + OIS, achieving outstanding F1-scores of 0.9780 and 0.9846, respectively. This underscores that hybrid models, particularly those integrating GNNs with OIS, offer superior capability in maintaining precision-recall tradeoffs, making them ideal for accurate skyline detection in complex datasets.

https://doi.org/10.31436/iiumej.v27i1.3717

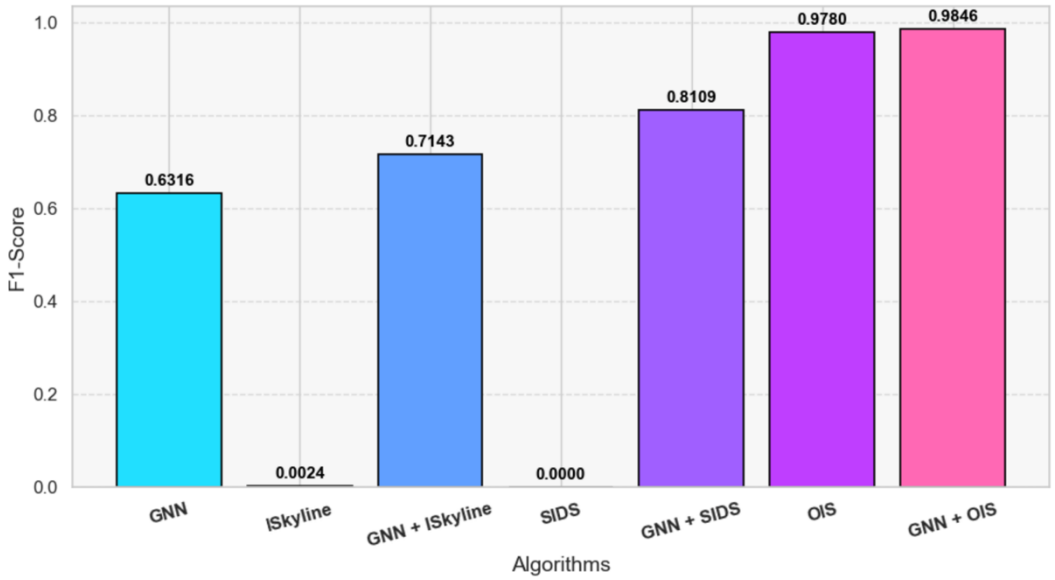


Figure 5. F1-Score Comparison of Algorithms Using CoIL 2000

Figure 6 provides a clear assessment of each algorithm's ability to distinguish between skyline and non-skyline points. Traditional algorithms such as ISkyline and SIDS, along with GNN, hover around a baseline performance of 0.50, indicating they perform no better than random guessing. Interestingly, the combination of GNN + ISkyline achieves an AUC-ROC of 1.0000, which may indicate overfitting or a specific synergy in that dataset. Meanwhile, GNN + SIDS outperforms both methods individually, achieving 0.8642, underscoring the benefits of combining deep learning with classic skyline methods. The OIS model again demonstrates robustness with a high score of 0.9813, and the GNN + OIS hybrid achieves the highest score of 0.9950, reinforcing that integrating GNNs with optimized skyline strategies yields the most reliable and discriminative results across all metrics.

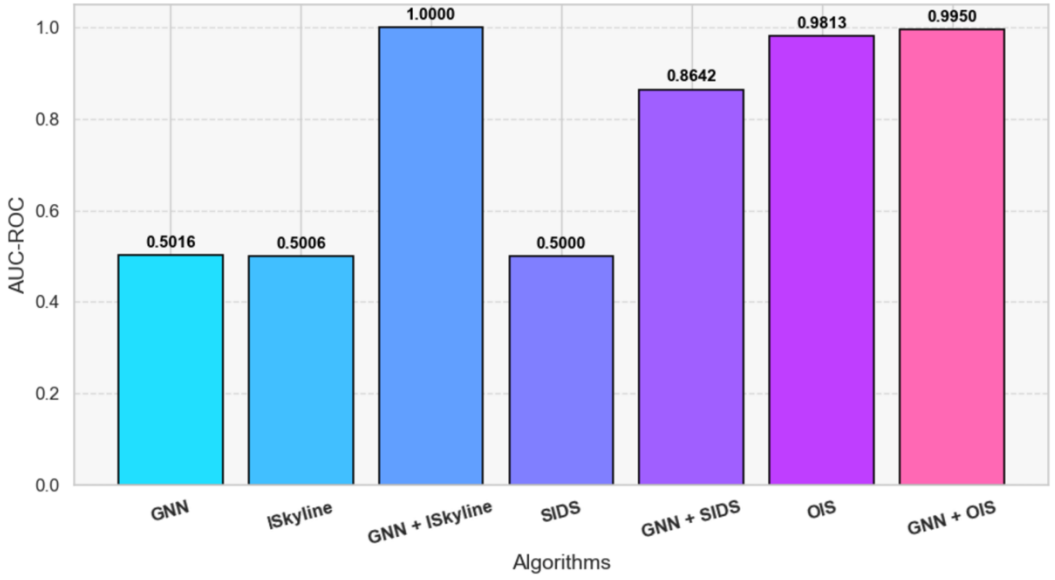


Figure 6. AUC-ROC Comparison of Algorithms Using CoIL 2000

Figure 7 reveals a striking contrast in performance among various algorithms. ISkyline and SIDS perform extremely poorly, with near-zero values across most metrics, indicating that they are ineffective standalone solutions. GNN performs moderately well with a Recall of 0.89

https://doi.org/10.31436/iiumej.v27i1.3717

and an F1-score of 0.63, but struggles with accuracy and precision. Interestingly, GNN + ISkyline achieves perfect precision and AUC-ROC, although its recall (0.56) and accuracy (0.73) limit overall effectiveness. GNN + SIDS shows significant improvement across all metrics, particularly with an F1-score of 0.81. The standout performers are OIS and GNN + OIS, both of which achieve near-perfect or perfect scores across every metric. Notably, GNN + OIS slightly outperforms OIS in AUC-ROC (0.99 vs. 0.98), making it the most robust and balanced model overall, demonstrating the efficacy of hybridizing deep learning with optimized skyline strategies.

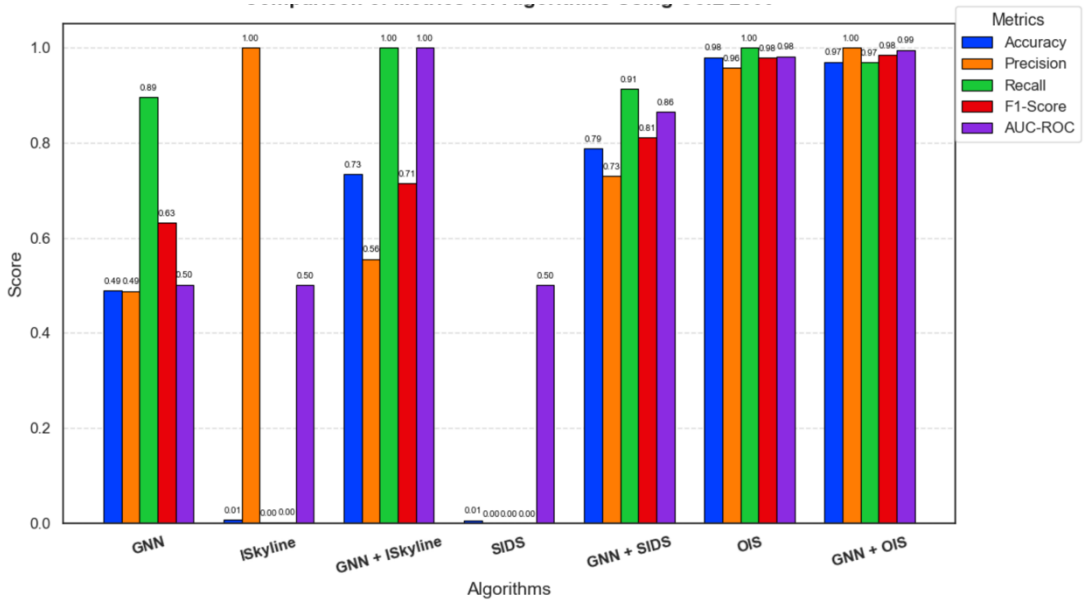


Figure 7. Comparison of Metrics for Algorithms Using CoIL 2000

Table 5. Comparison of Machine Learning Using CoIL 2000

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN + OIS	0.9698	1.0000	0.9696	0.9846	0.9950	338.3995	38407.19
GAT + OIS	0.9775	1.0000	0.9774	0.9885	0.9949	351.3499	25964.56
XGBoost + OIS	0.9777	0.9956	0.9819	0.9887	0.8689	97.8628	29138.93
RL + OIS	0.9957	0.9957	1.0000	0.9978	0.6622	875.2617	77173.20
OL + OIS	0.8699	0.9966	0.8721	0.9302	0.7229	106.1407	61728.71
GraphSAGE + OIS	1.0000	1.0000	1.0000	1.0000	1.0000	402.7302	25888.41

Table 5 compares the performance of various machine learning algorithms integrated with the OIS skyline optimization technique using the CoIL 2000 dataset. GraphSAGE + OIS achieves perfect scores (1.0000) across all metrics, making it the top performer in terms of predictive quality, albeit with moderate memory usage (25,888.41 KB) and a relatively high response time (402.73s). RL + OIS also demonstrates near-perfect scores but exhibits the highest memory usage (71,173.20 KB) and the longest response time (875.26s), limiting its practicality. GAT + OIS and XGBoost + OIS achieve strong metric performance and improved efficiency, with XGBoost + OIS having the shortest response time (97.86s) and low memory usage. However, its AUC-ROC is significantly lower (0.8689). GNN + OIS maintains a good balance between accuracy (0.9698), high AUC-ROC (0.9950), and moderate resource use.

https://doi.org/10.31436/iiumej.v27i1.3717

Conversely, OL + OIS, despite a decent precision (0.9966), lags in other metrics, particularly recall (0.8721), and exhibits high memory consumption (61,728.71 KB). In summary, GraphSAGE + OIS offers the best performance.

Figure 8 illustrates the comparative performance of various OIS-enhanced algorithms on the CoIL 2000 dataset across five key metrics: accuracy, precision, recall, F1-score, and AUC-ROC. GraphSAGE + OIS achieves a perfect score of 1.00 across all metrics, demonstrating strong classification performance. RL + OIS also achieves perfect scores for all metrics except AUC-ROC, which drops significantly to 0.66, indicating a weaker ability to distinguish between classes under varying thresholds. OL + OIS performs poorly compared to others, especially in AUC-ROC (0.72) and recall (0.87), despite high precision. GNN + OIS, GAT + OIS, and XGBoost + OIS demonstrate a strong balance of metrics, all hovering close to 0.99, although XGBoost's AUC-ROC lags slightly at 0.87. These results suggest that although several models exhibit competitive predictive performance, GraphSAGE + OIS yields the most robust and consistent performance across all tasks.

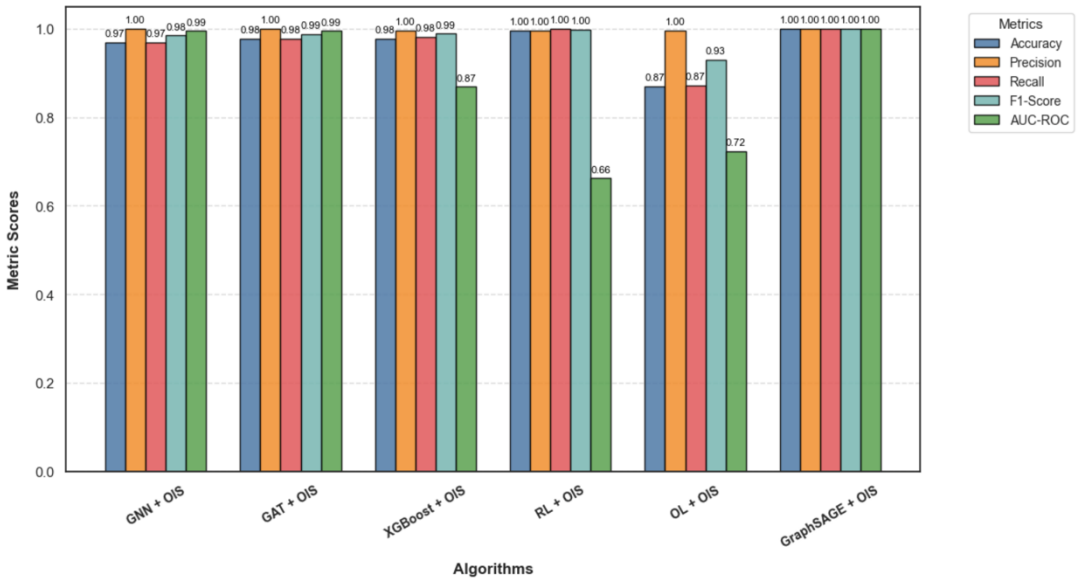


Figure 8. Comparison of OIS Variants Across Metrics Using CoIL 2000

4.5. Comparison of Algorithms Using NBA Stats

In this section, GNN + SIDS and GAT + ISkyline are excluded from the algorithm comparison because they consistently crashed during execution in the Jupyter Notebook. Table 6 presents a performance comparison of various algorithms applied to the NBA Stats dataset, evaluating both classification metrics and computational efficiency. OIS emerges as the top performer overall, achieving near-perfect accuracy (0.9973), perfect recall (1.0), and an excellent AUC-ROC (0.9986), all while maintaining the fastest response time (1.86s) and relatively low memory usage (4324 KB), making it the most balanced and efficient solution. GNN + ISkyline also performs strongly in accuracy (0.9533) and F1-score (0.9515), but at the cost of extremely high query response time (~17,680s) and massive memory usage (2.5 GB), which limits its practicality. ISkyline and SIDS, though faster and lighter, exhibit near-zero recall and F1 Scores, indicating that they miss nearly all relevant results. GNN alone also underperforms with low recall (0.0753) and F1-score (0.1314), despite modest memory demands. Finally, GNN + OIS achieves perfect recall but low accuracy (0.1162) and F1-score (0.2081), suggesting indiscriminate classification, and it also consumes over 1 GB of memory.

https://doi.org/10.31436/iiumej.v27i1.3717

Overall, OIS stands out as the best choice in terms of accuracy, efficiency, and real-world applicability for the NBA Stats dataset.

Table 6. Comparison of Algorithms Using NBA Stats

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN	0.5072	0.5150	0.0753	0.1314	0.4936	2279.9321	576.44
ISkyline	0.8840	1.0000	0.0014	0.0028	0.5007	75.5164	1697.09
GNN + ISkyline	0.9533	0.9841	0.9210	0.9515	0.9905	17679.7926	2524215.45
SIDS	0.8838	0.0000	0.0000	0.0000	0.5000	820.3373	36999.28
OIS	0.9973	0.8106	1.0000	0.8954	0.9986	1.8601	4324.18
GNN + OIS	0.1162	0.1162	1.0000	0.2081	0.0159	1650.1123	1031888.89

The NBA dataset, with its high feature sparsity and temporal attributes, exacerbates the inefficiency of ISkyline-based dominance graphs, necessitating excessive resources for GNN + ISkyline. OIS, however, remains efficient because its pruning strategies minimize comparisons, yielding faster, lighter computations while maintaining near-perfect accuracy. This contrast emphasizes the scalability advantage of OIS over ISkyline in real-world, high-volume datasets.

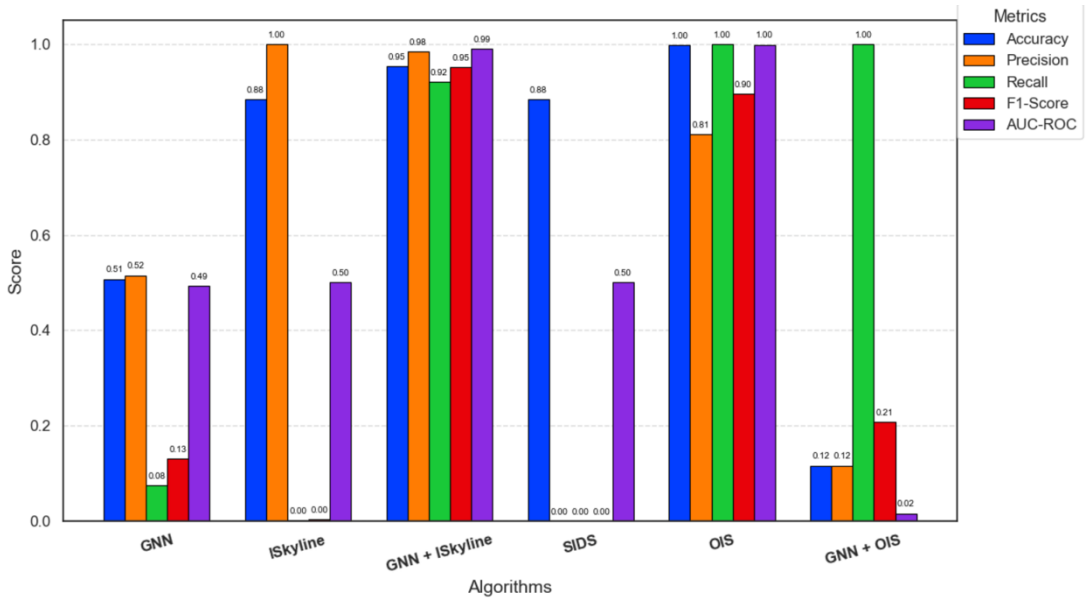


Figure 9. Comparison of Metrics for Algorithms Using NBA Stats

Figure 9 provides a comprehensive side-by-side comparison of multiple performance metrics, accuracy, precision, recall, F1-score, and AUC-ROC, for different algorithms applied to the NBA Stats dataset. OIS is the most balanced and effective model overall, achieving near-perfect or perfect scores across all metrics, including accuracy (1.00), recall (1.00), and AUC-ROC (1.00), making it the top choice for both accuracy and class separation. GNN + ISkyline also performs exceptionally, with high precision (0.9841), strong recall (0.9210), and a very high AUC-ROC (0.9905), showcasing its strong generalization and consistency. On the other hand, ISkyline and SIDS exhibit a complete collapse in recall and F1-score (both ≈0.00),

https://doi.org/10.31436/iiumej.v27i1.3717

despite ISkyline's perfect precision, implying either extreme overfitting or a failure to detect positive cases. GNN performs modestly, with mediocre values across all metrics. In contrast, GNN + OIS achieves high recall (1.00) but very poor performance on all other metrics, particularly accuracy (0.12) and AUC-ROC (0.02), indicating that it indiscriminately labels all points as the skyline. The chart emphasizes the robustness and efficiency of OIS, followed by GNN + ISkyline, while also highlighting the critical weaknesses in standalone traditional models and poorly integrated hybrids.

Table 7 compares various ISkyline-integrated machine learning models on the NBA Stats dataset, evaluating their performance across performance metrics and computational efficiency. GraphSAGE + ISkyline dominates with a perfect score of 1.000 in all performance metrics, accuracy, precision, recall, F1-score, and AUC-ROC, indicating flawless classification. However, this comes at the cost of very high query response time (~8943s) and massive memory usage (~2.5 GB), making it computationally intensive. XGBoost + ISkyline offers the best trade-off, achieving excellent classification results, with accuracy (0.9843), F1-score (0.9844), and AUC-ROC (0.9979), while being far more efficient in time (~482s) and memory (~126 MB). Similarly, RL + ISkyline and OL + ISkyline perform well with decent F1-scores (0.8985 and 0.8347, respectively) and fast response times (<470s), especially OL + ISkyline, which is the most memory-efficient (19.5 MB). In contrast, GNN + ISkyline, while accurate (0.9533), exhibits the longest query time (~17680s) and highest memory consumption (~2.5 GB), thereby limiting its practicality. Overall, GraphSAGE + ISkyline is ideal for scenarios requiring perfect accuracy regardless of resource constraints, whereas XGBoost + ISkyline strikes the best balance between performance and efficiency.

Table 7. Comparison of Machine Learning Using NBA Stats

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN + ISkyline	0.9533	0.9841	0.9210	0.9515	0.9905	17679.7926	2524215.45
XGBoost + ISkyline	0.9843	0.9699	0.9994	0.9844	0.9979	481.6498	126111.93
RL + ISkyline	0.9001	0.9072	0.8901	0.8985	0.9663	462.2957	40069.25
OL + ISkyline	0.8209	0.7709	0.9099	0.8347	0.9292	426.5436	19517.34
GraphSAGE + ISkyline	1.0000	1.0000	1.0000	1.0000	1.0000	8942.8623	2524046.19

4.6. Comparison of Algorithms Using MovieLens

When testing the proposed GNN + ISkyline framework on the MovieLens dataset across different dataset sizes (1M, 100k, 30k, 20k, and 10k records), several computational challenges were observed. Processing the 1M, 100k, and even 10k record subsets still required substantial time, indicating that the method remains computationally intensive even at smaller scales. More critically, attempts to run the model on the 30k and 20k subsets led to kernel crashes.

The following results were evaluated on the MovieLens dataset using 5k records, which was sufficient for successful execution. However, the GAT + ISkyline model was excluded from this test because it caused a kernel crash during execution.

As shown in Table 8, GNN + ISkyline achieves the highest overall performance, achieving perfect scores of 1.0000 for accuracy, precision, recall, F1-score, and AUC-ROC. However, this comes with a relatively high query response time (2878.10s) and peak memory usage (1927721.87 KB). ISkyline alone also performs very well in accuracy (0.9978) and precision

https://doi.org/10.31436/iiumej.v27i1.3717

(1.0000), but its recall (0.2778) and F1-score (0.4348) are notably lower, suggesting limited effectiveness in capturing positive cases. OIS achieves excellent accuracy (0.9993) and reasonable AUC-ROC (0.7497), with the benefit of speedy query response time (0.0562s) and low peak memory usage (723.07 KB). GNN + SIDS demonstrates balanced performance, with a high AUC-ROC (0.9051) and relatively efficient resource utilization. Meanwhile, GNN and SIDS alone perform poorly, with zero recall and F1-score values. GNN + OIS struggles significantly, with extremely low accuracy (0.0031) and AUC-ROC (0.0005), despite moderate memory usage.

Table 8. Comparison of Algorithms Using MovieLens

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN	0.5171	0.0000	0.0000	0.0000	0.5243	590.1351	375.99
ISkyline	0.9978	1.0000	0.2778	0.4348	0.6389	0.8235	223.54
GNN + ISkyline	1.0000	1.0000	1.0000	1.0000	1.0000	2878.0988	1927721.87
SIDS	0.9969	0.0000	0.0000	0.0000	0.5000	69.2292	3450.52
GNN + SIDS	0.8187	0.8302	0.8010	0.8154	0.9051	2.1972	2283.49
OIS	0.9993	0.5000	0.7500	0.6000	0.7497	0.0562	723.07
GNN + OIS	0.0031	0.0031	1.0000	0.0061	0.0005	212.3288	344711.13

MovieLens data, being user-item interaction-based, produces extremely dense graphs under ISkyline. This enables GNN + ISkyline to learn dominance structures effectively, thereby achieving perfect accuracy. However, the density results in prohibitive memory consumption (approximately 2 GB) and long runtimes. In contrast, GNN + OIS collapses on MovieLens because OIS pruning removes too much relational information, leaving the GNN with insufficient features to generalize.

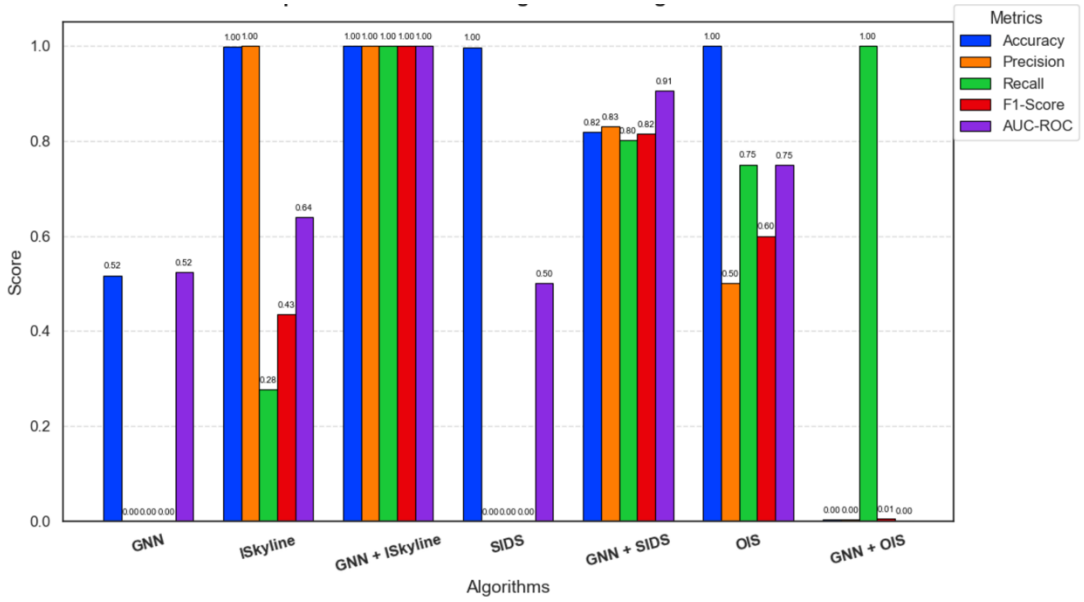


Figure 10. Comparison of Metrics for Algorithms Using MovieLens

Figure 10 highlights GNN + ISkyline as the top-performing approach, achieving perfect scores (1.00) across all five metrics, accuracy, precision, recall, F1-score, and AUC-ROC,

https://doi.org/10.31436/iiumej.v27i1.3717

demonstrating outstanding overall performance. GNN + SIDS also performs strongly, achieving high scores across all metrics (0.80–0.91), although slightly lower than GNN + ISkyline. OIS shows good accuracy and recall but lower precision (0.50) and F1-score (0.60), and moderate AUC-ROC (~0.75). In contrast, ISkyline alone achieves excellent accuracy and precision but falls short on recall (0.28) and F1-score (0.43), indicating an imbalance. GNN and SIDS individually struggle, with most metrics 0 except for moderate accuracy (~0.52) and AUC-ROC (~0.52) for GNN. GNN + OIS notably underperforms, with recall the only strong metric (1.00), whereas all other metrics are nearly 0, indicating extremely poor general classification performance. Overall, integrating GNN with ISkyline is the most effective combination, followed by GNN + SIDS.

Table 9 presents a comparison of machine learning models combined with ISkyline on the MovieLens dataset, showing that GNN + ISkyline and GraphSAGE + ISkyline both achieved perfect scores (1.0000) across all evaluation metrics, accuracy, precision, recall, F1-score, and AUC-ROC, demonstrating flawless performance. XGBoost + ISkyline also performed exceptionally, achieving near-perfect metrics with accuracy (0.9996), precision (0.9991), recall (1.0000), F1-score (0.9996), and AUC-ROC (1.0000), while maintaining a low query response time and modest memory usage compared to GNN-based methods. OL + ISkyline showed the lowest performance among the listed models but still maintained decent accuracy (0.9424) and AUC-ROC (0.9842). In terms of resource efficiency, OL + ISkyline was the most lightweight, with the lowest memory usage, whereas the GNN and GraphSAGE variants required significantly more memory. Overall, all models exhibit excellent performance, but the GNN + ISkyline and GraphSAGE + ISkyline combinations stand out for achieving perfect classification at the cost of substantially higher memory and processing demands.

Table 9. Comparison of Machine Learning Using MovieLens

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Query Response Time (s)	Peak Memory Usage (KB)
GNN + ISkyline	1.0000	1.0000	1.0000	1.0000	1.0000	2878.0988	1927721.87
XGBoost + ISkyline	0.9996	0.9991	1.0000	0.9996	1.0000	98.6301	2527.05
RL + ISkyline	0.9832	0.9674	1.0000	0.9834	0.9931	96.9954	51641.99
OL + ISkyline	0.9424	0.9523	0.9309	0.9415	0.9842	93.6479	2007.34
GraphSAGE + ISkyline	1.0000	1.0000	1.0000	1.0000	1.0000	1728.0954	1924217.91

5. CONCLUSION

This research on GNN-based skyline query processing for large-scale and incomplete graphs showed that integrating Graph Neural Networks (GNNs) with Pareto optimality principles improves skyline query performance by enhancing scalability, reducing computational overhead, and effectively handling incomplete data. The proposed GNN-based framework outperformed traditional algorithms such as ISkyline, SIDS, and OIS in accuracy, F1-score, and AUC-ROC, and demonstrated adaptability to dynamic environments with real-time updates. Benchmarks using metrics such as accuracy, F1-score, AUC-ROC, query response time, and memory usage validated its effectiveness across synthetic and real-world datasets.

Furthermore, this research meaningfully supports Sustainable Development Goal 9 (Industry, Innovation, and Infrastructure) by demonstrating scalable, adaptive techniques for data-intensive systems. Handling incomplete and large-scale data robustly ensures the resilience and scalability of digital infrastructures, essential for future smart cities, industrial automation, and sustainable development platforms. By addressing both scalability and data incompleteness, the proposed framework directly contributes to the development of more intelligent, responsive, and reliable infrastructure systems.

Computational limitations constrained this research due to the use of a consumer-grade MacBook Air (M2, 8GB RAM). Limited memory and processing capacity occasionally caused kernel crashes, especially during large-scale graph construction and GNN training on bigger datasets (e.g., NBA Stats, MovieLens). As a result, some hybrid models could not be thoroughly tested with vast datasets. In future work, experiments should be conducted using more powerful computing resources, such as high-memory GPUs or cloud-based platforms (e.g., AWS EC2 P4 instances, Google Cloud TPU Pods). This would enable evaluation of larger graphs, deeper GNN models, and hyperparameter tuning at a larger scale, thereby further validating the proposed framework under realistic, production-level conditions.

Despite the promising results of machine learning models in skyline query processing, several practical limitations remain. First, GNN-based methods exhibit high memory usage, especially when processing large or dense graphs, which may hinder scalability in resource-constrained environments. Second, these deep learning models often suffer from interpretability challenges, making it difficult to understand how specific dominance relationships are learned or how skyline classifications are made, an issue that can limit their adoption in critical decision-making systems where transparency is essential. Third, while synthetic datasets enable controlled experimentation, they may not fully capture the complexity and noise of real-world data, potentially limiting the generalizability of the models trained on them. These limitations highlight the need for further research to optimize resource use, enhance model explainability, and validate findings on more diverse, real-world datasets.

Future research could extend this work in several directions. First, scalability can be improved by investigating lightweight or compressed GNN architectures (e.g., pruning, quantization, or knowledge distillation) to reduce memory and runtime costs without sacrificing accuracy. Second, explainability should be prioritized by integrating interpretable GNN models or post hoc explanation methods, thereby enabling users to understand dominance relationships in skyline predictions better. Third, a broader evaluation on diverse real-world datasets beyond CoIL 2000, NBA Stats, and MovieLens is essential to test generalizability across domains such as healthcare, transportation, and financial risk analysis. Finally, exploring distributed or parallel skyline computation frameworks using cloud-based platforms could further support deployment in large-scale, real-time environments. Together, these directions will help bridge the gap between theoretical contributions and practical deployment of GNN-based skyline query processing.

ACKNOWLEDGEMENT

This research was supported by the Fundamental Research Grant Scheme (FRGS) under Reference Code FRGS/1/2021/ICT01/UIAM/02/2 (Project ID 19574) from the Ministry of Higher Education (MOHE), Malaysia.

REFERENCES

[1] Wang, Y., Shi, Z., Wang, J., Sun, L., & Song, B. (2017). Skyline Preference Query Based on Massive and Incomplete Dataset. IEEE Access, 5, 3183–3192. https://doi.org/10.1109/access.2016.2639558

[2] Khalefa, M. E., Mokbel, M. F., & Levandoski, J. J. (2008). Skyline Query Processing for Incomplete Data. https://doi.org/10.1109/icde.2008.4497464

[3] Bharuka, R., & Kumar, P. S. (2013a). Finding superior skyline points from incomplete data. 35–44. https://doi.org/10.5555/2694476.2694488

[4] Afifi, H., Pochaba, S., Boltres, A., Laniewski, D., Haberer, J., Leonard, P., Poorzare, R., Stolpmann, D., Wehner, N., Redder, A., Samikwa, E., & Seufert, M. (2024). Machine Learning with Computer Networks: Techniques, Datasets and Models. IEEE Access, 1–1. https://doi.org/10.1109/access.2024.3384460

[5] Mohamud, M. A., Ibrahim, H., Sidi, F., Mohd, N., Dzolkhifli, Z., Zhang, X., & Lawal, M. M. (2023). A Systematic Literature Review of Skyline Query Processing Over Data Stream. IEEE Access, 11, 72813–72835. https://doi.org/10.1109/access.2023.3295117

[6] Borzsony, S., Kossmann, D., & Stocker, K. (2001). The Skyline operator. Proceedings 17th International Conference on Data Engineering. https://doi.org/10.1109/icde.2001.914855

[7] Bharuka, R., & Kumar, P. S. (2013b). Finding skylines for incomplete data. 109–117.

[8] Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly, 28(1), 75–105. https://doi.org/10.2307/25148625

[9] vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to Design Science Research. Progress in IS, 1–13. https://doi.org/10.1007/978-3-030-46781-4_1

[10] Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In Design Research in Information Systems. Springer, Boston, MA., 9–22.

[11] Putten, P. (2000). Insurance Company Benchmark (COIL 2000) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5630S

[12] (2000) The Basketball-Reference website. [Online]. Available: https://www.basketball-reference.com

[13] (2025) The GroupLens website. [Online]. Available: https://grouplens.org/datasets/movielens/

https://doi.org/10.31436/iiumej.v27i1.3720

Development of Electric Fence Fault Sensing and Monitoring System with LoRaWAN IoT

M. SYARIFUDDIN M. SUFIAN[1*], HISHAM MOHAMAD[1], M. AZMAN ZAKARIYA[2], AIZAT AKMAL A. MOHAMAD BEDDELEE[1], SALMAN SAABAN[3], M. FADLLI A. YAZI[3]

[1]Civil & Environmental Engineering Department,
Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Malaysia
[2]Electrical & Electronics Engineering Department,
Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Malaysia
[3]Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN),
56100 Kuala Lumpur, Malaysia

**Corresponding author: syarifuddin_16001417@utp.edu.my*

(Received: 13 May 2025; Accepted: 1 October 2025; Published online: 12 January 2026)

ABSTRACT: The persistent challenge of Human-Elephant Conflict (HEC) in regions like Malaysia necessitates robust and efficient mitigation strategies. Electric fences are effective but often face maintenance inefficiencies due to delayed fault detection. This study presents a smart electric fence monitoring system designed for real-time fault diagnosis and localisation. The system employs IoT-enabled in-place sensor nodes comprising 10 kV voltage sensors, short-circuit detection sensors, and 3-axis gyroscope sensors. Sensor data is transmitted via a LoRaWAN network, selected for its long-range, low-power characteristics, which are well-suited to rural, low-bandwidth environments where electric fences are typically deployed. Field validation of a 50 m, 10 kV, 6 A electric fence segment achieved 100% voltage and short-circuit detection rates and 99.91% gyroscope tilt accuracy. Reliable data transmission was maintained up to 1.3 km, with an RSSI of -110 dBm, in campus environments with concrete obstructions. Supplementary testing at the same positions using antennas at increased height yielded an RSSI of -79 dBm, with a 41 dB link margin, highlighting the potential for range scaling in future work. The system has so far been validated on a short 50 m electric fence segment with a practical LoRaWAN range of 1.3 km under campus conditions, indicating the need for further optimisation and field-scale trials. The system provides a practical solution for the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN), with direct application to their Sistem Pagar Elektrik Gajah (SPEG) for HEC mitigation. Beyond this application, the approach demonstrates potential for future use of IoT-based Structural Health Monitoring (SHM) concepts in resource-constrained rural infrastructures.

ABSTRAK: Cabaran berterusan Konflik Manusia-Gajah (HEC) di kawasan seperti Malaysia memerlukan strategi mitigasi yang kukuh dan berkesan. Pagar elektrik berkesan dalam menangani konflik ini, tetapi sering berhadapan masalah penyelenggaraan akibat kelewatan pengesanan kerosakan pada pagar. Kajian ini membentangkan satu sistem pemantauan pagar elektrik pintar yang direka bagi mengdiagnosis penentuan lokasi kerosakan secara masa nyata. Sistem ini menggunakan nod sensor dengan keupayaan internet benda terdiri daripada sensor voltan 10 kV, sensor pengesan litar pintas, dan sensor giroskop. Data sensor dihantar melalui rangkaian LoRaWAN, yang dipilih atas dasar rangkaian jarak jauh dan penggunaan tenaga minimal. Ciri-ciri ini dianggap bersesuaian bagi persekitaran luar bandar di mana pagar elektrik biasanya terpasang, kerana kawasan ini seringkali tidak mendapat rangkaian telekomunikasi komersial. Ujian lapangan pada segmen pagar elektrik sepanjang 50 m, 10 kV 6 A telah berjaya mengesan voltan dan litar pintas dengan ketepatan 100%, manakala sensor giroskop pula mampu mengesan kecondongan pada kadar 99.91%. Penghantaran data

https://doi.org/10.31436/iiumej.v27i1.3720

daripada sensor pula berjaya mencapai jarak maximum 1.3 km dengan RSSI -110 dBm dalam persekitaran kampus (yang mempunyai pelbagai halangan konkrit). Ujian sampingan penghantaran data yang dilaksanakan menggunakan antena berkedudukan tinggi pada lokasi dan jarak yang sama pula menunjukkan RSSI -79 dBm dengan margin pautan 41 dB. Ini menunjukkan potensi peningkatan jarak penghantaran data bagi kajian seterusnya. Pada masa ini, sistem ini telah disahkan kepenggunaanya bagi segmen pagar elektrik pendek dengan kepanjangan 50 m, manakala penghantaran data secara berkesan dihadkan pada 1.3 km dalam persekitaran kampus. Had ini menunjukkan, wujud keperluan bagi kerja-kerja penambah baikan untuk meningkatkan kadar keberkesanan melalui ujian berskala besar di lapangan pada masa hadapan. Sistem ini menyediakan penyelesaian kejuruteraan yang praktikal untuk Jabatan Perlindungan Hidupan Liar dan Taman Negara Semenanjung Malaysia (PERHILITAN), dengan aplikasi langsung pada Sistem Pagar Elektrik Gajah (SPEG) bagi mitigasi HEC. Kajian ini juga menunjukkan potensi bagi penggunaan konsep Pemantauan Kesihatan Struktur (SHM) berasaskan IoT pada masa hadapan dalam infrastruktur luar bandar.

KEYWORDS: *LoRaWAN, Electric Fence, RFM95, STM32F103C8T6, In-Placed Sensors.*

1. INTRODUCTION

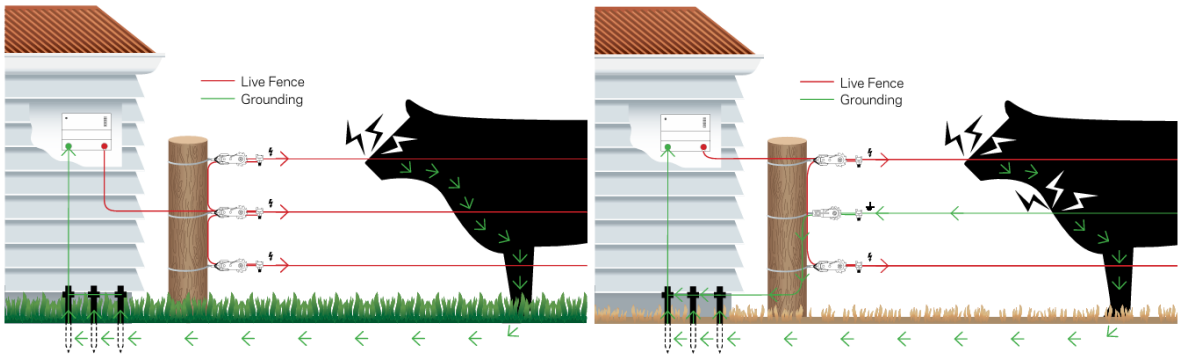
Human-Elephant Conflict (HEC) poses a critical challenge due to unpredictable elephant incursions into human settlements, resulting in substantial socioeconomic and ecological impacts. In Peninsular Malaysia, HEC has caused annual crop losses of approximately RM 2.96 million across 11,460 acres of plantations, while mitigation and repair costs have exceeded RM 3.59 million over five years [1]. Between 2008 and 2019, at least 12 human fatalities were recorded [2], alongside heightened risks to elephant populations, with five elephants killed in retaliatory incidents and four in vehicle collisions since 2018 [3].

To address these conflicts, governments have implemented various deterrent measures, with electric fences emerging as the primary solution. The Malaysian government alone has invested more than RM 6 million in nationwide electric fence infrastructure [4,5]. However, these capital expenditures are compounded by substantial recurring costs. A single energiser station, covering roughly 4 km of fence, requires RM 19,200-RM 28,800 annually for routine maintenance, including voltage inspections, system checks, and vegetation management [6]. Improving fault detection and maintenance efficiency is, therefore, critical to sustaining this investment.

Electric fences are designed to be highly effective physical and psychological barriers against elephant intrusions [7-9]. A typical system consists of an energiser (high-voltage transformer), power sources (batteries and solar panels), conductive wiring, grounding, and structural supports. In Malaysia, such systems have demonstrated approximately 80% effectiveness in protecting oil palm and rubber plantations [10]. According to PERHILITAN's guidelines, electric fences must deliver high-voltage pulses of 6-10 kV at an impulse energy of 12 J, with a maximum repetition frequency of 1 Hz [6].

The architecture of an electric fence system generally adopts one of two wiring configurations. As illustrated in Figure 1, electric fence cables can be connected using an all-live wire system or a ground-return wire system, depending on site-specific environmental conditions [11]. In humid regions where soil conductivity is adequate, all-live systems are commonly employed. Conversely, in areas with low ground conductivity, ground-return configurations are preferred to ensure that electric pulses are delivered effectively through simultaneous contact with live and ground wires.

https://doi.org/10.31436/iiumej.v27i1.3720



(a) All-live cable wire system. (b) Ground return wire system

Figure 1. Electric fence system wire connection configurations [11].

In Malaysia, the elephant electric fence system has been standardised through the implementation of the Sistem Pagar Elektrik Gajah (SPEG), developed by the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN) as a strategic measure to mitigate Human-Elephant Conflict. The SPEG design specifies a multi-wire configuration with fences 1.5-1.8 m high. The fence incorporates three to five high-tensile galvanised steel wires, connected in parallel and spaced at 30-60 cm gaps. To prevent short circuits, the cables are supported by Acrylonitrile Butadiene Styrene (ABS) insulators mounted on galvanised steel, wood, or PVC poles, with support poles spaced 5-10 m apart [6]. The Pel 412i Unigizer™ is the standard energiser used in the SPEG to power 4 km of fence, commonly referred to as a "bay." The SPEG uses the all-live wire system architecture, with grounding achieved through the installation of at least three galvanised rods, each 1.5-2 m long and spaced at 3 m intervals [6].

Despite these sophisticated installations, operational effectiveness is constrained by technical and logistical challenges. Fences remain vulnerable to elephant-induced damage at poles and supports, while fallen trees and vegetation frequently disrupt integrity [12]. These issues cause voltage drops that compromise deterrent capacity. Additionally, maintaining fences across vast rural areas requires intensive manual inspections, and current technologies offer limited capacity for rapid, accurate fault detection in remote environments.

This research addresses these challenges by developing a remote electric fence monitoring system to improve maintenance efficiency and system reliability. The proposed solution integrates custom-developed sensors with a robust Long-Range Wide Area Network (LoRaWAN) communication infrastructure, enabling real-time voltage monitoring, fault localisation, and structural integrity assessment. Conducted in collaboration with the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN), the system is designed to meet key requirements, namely detecting voltage drops, pinpointing short-circuit locations, monitoring leaning poles, operating at low power, and supporting easy plug-and-play deployment. This initiative directly enhances the effectiveness of electric fences as a sustainable HEC mitigation strategy.

2. LITERATURE REVIEW & RELATED WORKS

Electric fences are widely utilised to mitigate human-elephant conflict (HEC). However, their operational reliability is frequently compromised by damage and malfunctions, resulting in system faults. The faults disrupting the fence's intended function include wire breaks, connection failures, grounding issues, or any event that reduces the system's effectiveness in deterring elephants. Timely detection and rectification of these faults are crucial for

https://doi.org/10.31436/iiumej.v27i1.3720

maintaining the integrity and efficacy of the electric fence system. To develop a monitoring system for an electric fence, findings from previous research have been studied. The review encompasses electric fence monitoring mechanisms and suitable types of communication systems for the application.

2.1. Electric Fence Monitoring Technologies

Current fault detection technologies can be broadly categorised into patrol control devices, in-place sensors, and functional analysis. This study focuses on developing in-place sensor systems, recognised for their intuitive and passive monitoring capabilities. Several studies, including Tennakoon et al. [12], Al-Bahadly and Simpson [13], Ali et al. [14], and Yunus and Zainal [15], have examined various electric fence monitoring systems to address these challenges. A summary of these studies is presented in Table 1.

In a study, Tennakoon et al. [12] developed a system that detects and localizes wire breakage in a ground-return electric fence system connected in series. This method uses two sensor systems: one for wire-breakage detection and another for open-circuit detection on the electric fence. The sensors were mounted on the fence in a wired assembly, using the fence cables as a communication medium. The proposed system utilises two pulse generators, each for generating a high voltage (10kV deterrent) and a communication step voltage (fault detection). The use of these pulse generators is accompanied by switching components, detectors, and other devices, as depicted in Figure 2.

The fence controller manages the switching function between the fence pulse generator and the communication pulse generator. This enables the application of different pulses (high and low voltages) on the electric fence cables. The communication pulse generator supplies unique PWM pulses and requires only a single unit per fence [12]. The first system, wire breakage detection, comprises the detector, detectors 1 and 2, and the end node. The second system for locating the breakage comprises nodes 1, 2, and 3, along with their relays. The controller detector has two functions: to detect wire breakages and the sensor IDs of the nodes (1, 2, and 3). It is connected to the main system, which comprises a single-board computer (Raspberry Pi) for signal processing. Wires 1 and 3 are connected to the fence controller as live cables, while wire 2 is connected as a ground cable.

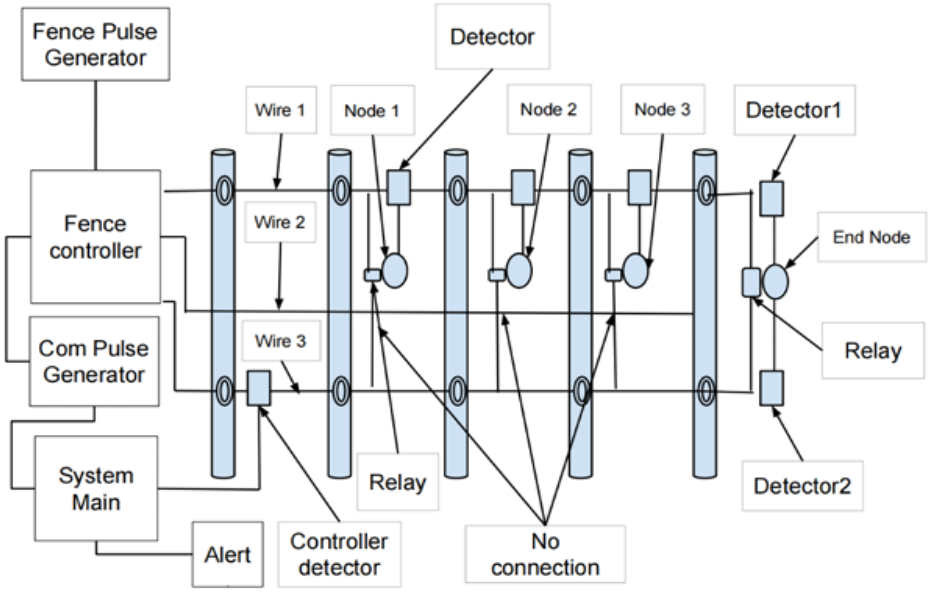


Figure 2. System design for detection and communication [12].

https://doi.org/10.31436/iiumej.v27i1.3720

The system uses a fixed 15-minute interval to monitor the health of an electric fence. The wire breakage detector system begins by closing the relay at the end node and switching wire 3 from a live connection to a ground connection. This creates a complete high-voltage loop between wires 1 and 3 [12]. The controller detector is equipped with a voltage detector that detects the presence of a high-voltage pulse, indicating wire breakage. If no pulse is detected, the wire breakage locator system is triggered.

The breakage locator process begins when the fence controller switches from the fence pulse generator to the communication pulse generator. Concurrently, the nodes (1, 2, 3) and the end node close the relays along the fence. This allows the communication pulse to flow through each sensor node back to the controller detector. Once each sensor detects a pulse, it returns the signal with its unique binary ID for detection by the controller. The "system main" then determines the location of the breakage based on the signals returned by the sensor nodes. For any breakages, the alert is used to notify users of the faulty condition [12].

While this approach accurately detects wire breaks in an electric fence system, it is limited to wire breaks and does not detect other faults, such as voltage drops or short circuits. It is also applicable only to a series-connected ground-return electric fence system, whereas the SPEG system uses a parallel all-live electric fence system. Due to the absence of a dedicated return conductor and the parallel distribution of high-voltage output across multiple live wires in SPEG, Tennakoon's pulse-injection method cannot establish the closed diagnostic loop on which it relies. As a result, the signal attenuation and path uncertainty inherent to SPEG preclude the reliable operation of Tennakoon's system without substantial redesign. This study, however, inspired the adaptation of a 12V DC supply that operates in parallel with the electric fence energiser's 10kV pulse for a short-circuit locator system. It is also considered a simpler approach than a complex PWM system that utilises multiple components.

Al-Bahadly and Simpson [13] proposed a contactless fault detection method that uses an antenna and an oscilloscope to measure the Electromagnetic Field (EMF) generated by the high-voltage fence line. The technique uses a coil with 50 turns connected to an oscilloscope to measure the miniature voltage generated by the induced current via its EMF, as illustrated in Figure 3. The developed EMF sensor (antenna) was tested by shorting the middle of a 100 m electric fence cable. The test uses a single cable all-live electric fence design, and the developed antenna successfully detected voltage and current changes under normal and fault conditions at distances of up to 1 m from the wire.

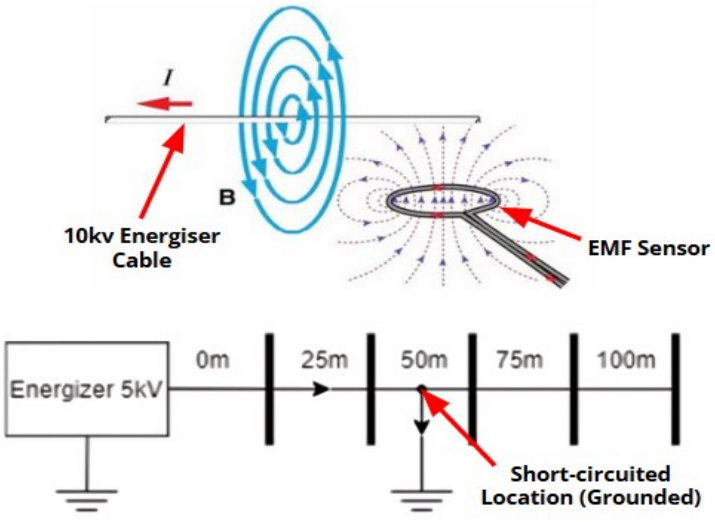


Figure 3. Contactless EMF sensor and conducted test setup [13].

https://doi.org/10.31436/iiumej.v27i1.3720

Although the authors suggested that the coil-based sensor could be mounted on a drone for field deployment, such an implementation would require stable low-altitude flight, precise coil-wire alignment, and mitigation of electromagnetic interference (EMI) from the drone's own electronics and the surrounding environment. These constraints substantially increase the operational complexity and cost, making routine deployment along rural fences impractical.

A study by Ali et al. [14] implemented a system to monitor the health of an electric fence using an Arduino and a GSM module for data transmission. In this work, the current is measured between a power supply (battery) and an electric fence energiser input. The status of the electric fence is determined by the current measured between the battery's positive terminal and the energiser's positive terminal. A microcontroller (Arduino) is used to collect and analyse data before transmitting it to an online server (Blynk) via a 4G GSM module. Figure 4 illustrates the system's process for delivering current sensor data to a mobile phone for monitoring.

In this system, notifications are generated via the Blynk app and are simulated under various power supply fault conditions in an actual fencing environment. However, it fails to address the critical issue of fault localisation triggered by short circuits, and it requires manual inspection and identification. This limitation highlights the need for more comprehensive monitoring solutions that can accurately identify and locate a broader range of fault types.

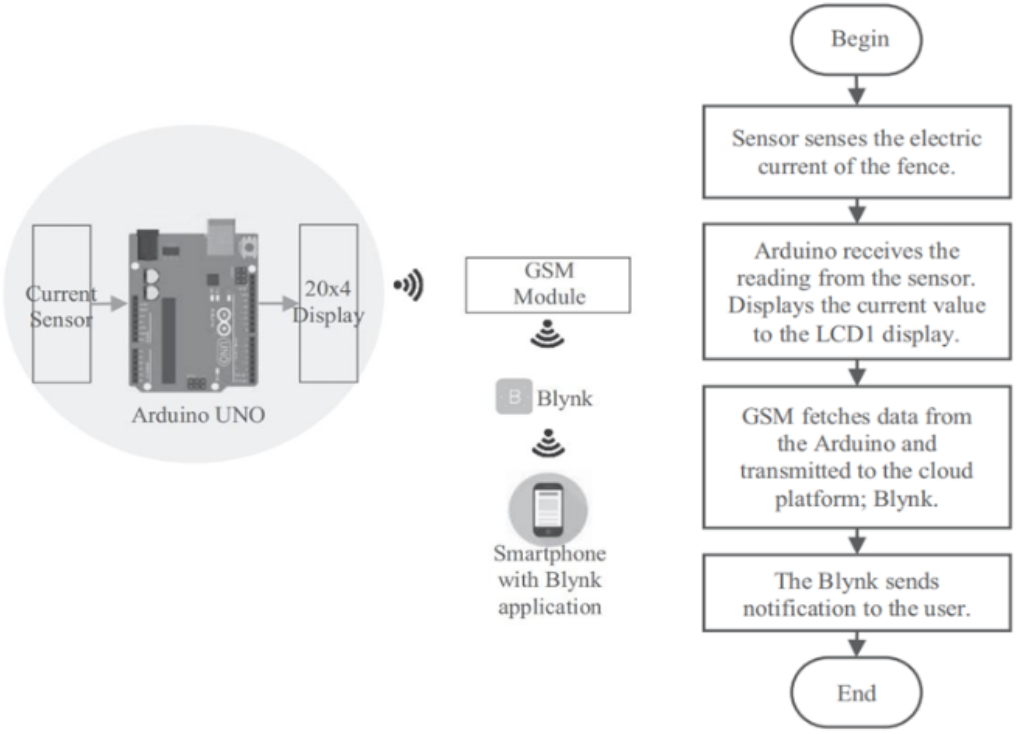


Figure 4. Monitoring system for electric fence architecture and flowchart [14].

Yunus and Zainal [15] proposed an IoT-based intelligent monitoring system for farm electric fences powered by a stand-alone solar energy supply. The architecture employed a microcontroller-based unit (ESP32S) with wireless communication capability (WiFi) to transmit fence parameters to a remote interface. Theoretically, this approach enables farmers to continuously monitor fence performance and receive notifications of abnormal operating conditions without requiring manual inspection.

https://doi.org/10.31436/iiumej.v27i1.3720

In their implementation, the incorporated sensor (as shown in Figure 5) measures key electrical characteristics of the fence circuit, including output voltage, current draw, and battery status [15]. The system was installed in a farm setting and evaluated over two days to assess its stability and effectiveness under real conditions. Measurements of solar input and battery state of charge were also taken to assess the power supply's self-sufficiency. Data collected from the sensors were transmitted to a central dashboard, where operators could observe trends and detect potential anomalies in near-real time.

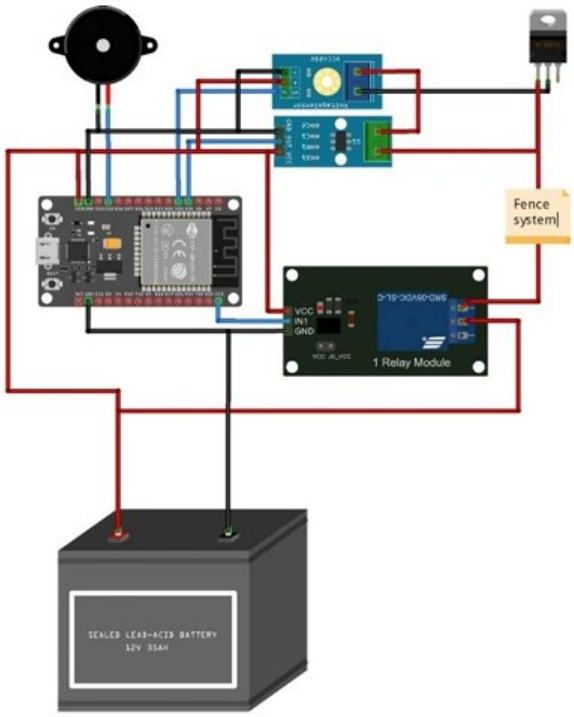


Figure 5. IoT-based Intelligent System Circuit Connection [15].

The results reported by Yunus and Zainal indicated that the system successfully captured and communicated variations in fence voltage, current, and power supply parameters throughout the test period [15]. The prototype demonstrated that IoT-enabled monitoring can enhance situational awareness of electric fence conditions and support proactive maintenance. However, their study did not implement or evaluate a mechanism for pinpointing the location or exact nature of faults along the fence line. This suggests that further research is needed to extend such systems beyond parameter monitoring to include precise fault localisation and classification, to implement a more effective monitoring system for electric fences.

Based on the data summarised in Table 1, four distinct approaches to electric-fence monitoring have been reported by Tennakoon et al. [12], Al-Bahadly and Simpson [13], Ali et al. [14], and Yunus and Zainal [15]. The design proposed by Tennakoon et al. can accurately detect wire breakages but is limited to series-connected, ground-return electric fences and is therefore not directly applicable to SPEG's parallel all-live configuration. The contactless method proposed by Al-Bahadly and Simpson demonstrates that fault location can be determined from induced EMF measurements. However, its reliance on specialised equipment and the challenges associated with drone-based deployment reduce its practical suitability for routine field use. By contrast, the system described by Ali et al. [14] and Yunus and Zainal [15] emphasises remote mobile notification but does not address fault-location detection. However, it informed the development of the early-warning component in the present work.

https://doi.org/10.31436/iiumej.v27i1.3720

Table 1. Summary of existing electric fence monitoring technologies

Reference	Fault Detection	Remote Monitoring	Complexity	Practicality
Tennakoon et al. [12]	Yes.	Yes. Provides remote mobile notifications.	High. High risk of system failure.	Medium. Not applicable to SPEG electric fences.
Al-Bahadly and Simpson [13]	Yes.	No.	High. Specialist Dependent.	Low. Challenges associated with drone-based deployment and coil alignment in the field.
Ali et al. [14]	No. Only the power supply is monitored.	Yes. Provides remote mobile notifications.	Low.	High.
Yunus and Zainal [15]	No. Only power supply analysis.	Yes. Provides remote mobile notifications.	Low.	High.

2.2. Review on IoT integration with Wireless Communications

Electric fences are typically developed in remote forested and rural environments to confine animals within their natural habitats. These areas often lack essential infrastructure, such as electricity and potable water, rendering them generally uninhabitable for humans. Consequently, the lack of readily available power sources in these rural settings necessitates the use of batteries and alternative energy sources, such as solar panels [16]. Furthermore, these remote areas are usually far from human settlements and often lack cellular towers, thus limiting communication networks. Limited mobile infrastructure in rural regions poses challenges to the effective implementation of many modern technologies [17].

Technical personnel responsible for maintaining electric fences in isolated rural and forest areas frequently rely on radios and walkie-talkies for communication due to the absence of cellular and internet connectivity. An effective electric fence monitoring system requires access to a dependable communication network to transmit data and alerts to relevant personnel. Sendra et al. [16] had highlighted in their research on smart systems for rural areas that long-distance wireless technologies are the most viable solution for covering such vast expanses. This is where LoRaWAN becomes relevant. The emergence of Long Range (LoRa) and Long-Range Wide Area Network (LoRaWAN) technologies offers innovative long-range wireless communication methods. These technologies present a modern technical solution for communication in rural areas where traditional mobile network coverage is unavailable [17].

The LoRaWAN protocol utilises LoRa modulation, providing low power consumption for devices and an extensive communication range that can span several kilometers [18]. Figure 6 illustrates a typical LoRaWAN system architecture. A standard LoRaWAN system is structured into three distinct layers: the application layer, the network layer, and the server layer. In the application layer, sensors are connected to end devices, which function as transceivers capable of both receiving and transmitting data. These end devices process data collected by sensors and communicate with a LoRaWAN gateway to transmit the information. Within the LoRaWAN network, these end devices transmit data asynchronously to any gateway within their signal range [18]. This data transfer occurs within the network layer, utilising the LoRaWAN bandwidth.

https://doi.org/10.31436/iiumej.v27i1.3720

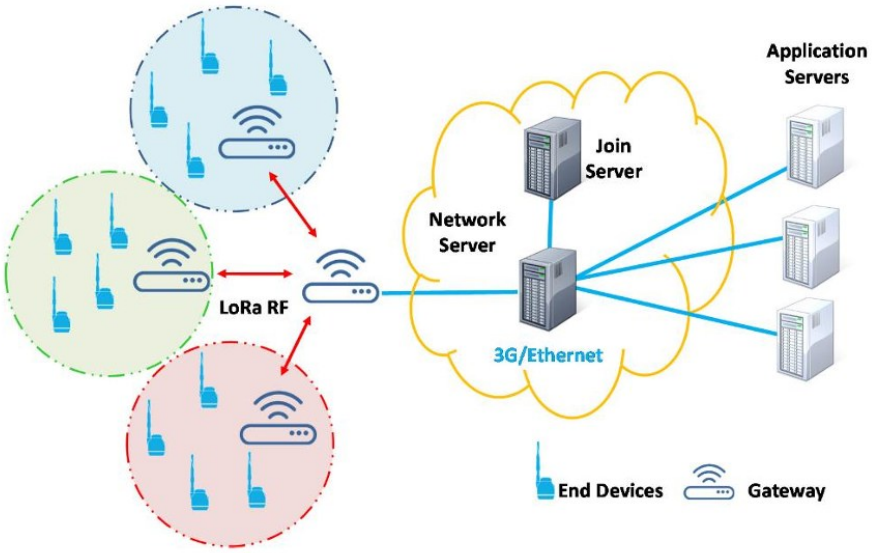


Figure 6. LoRaWAN system architecture [16].

The LoRaWAN gateway is a central component within a LoRaWAN system. In a basic LoRaWAN network design, a single gateway can typically support numerous end devices and manage their data transmissions. However, as shown in Figure 6, end nodes can also send data to gateways in their vicinity, which then relay it to a centralised gateway connected to the internet. Subsequently, the gateway establishes a connection to the server layer via technologies such as 4G LTE or WIFI. This constitutes the server layer.

Within the server layer, the gateway forwards the received data to a LoRaWAN server. Industrial users often prefer the ChirpStack server due to its robustness, enhanced customisation options, and security features. In contrast, the general community frequently utilises The Things Network (TTN), which is known for its user-friendliness and community-driven approach. From these LoRaWAN servers, the collected data can be stored in a local database or in a cloud storage platform (e.g., Google Cloud, Microsoft Power BI, or Amazon Web Services). This type of cloud integration can be achieved by calling APIs via the TTN web page. From this point, users can choose to perform data analysis, apply filters, set up notifications, or create monitoring dashboards for data visualisation. The TTN application servers can be configured to send messages via SMS or email to the maintenance departments of relevant authorities regularly or when specific issues arise [19].

A study by Sendra et al. [16] provides a comparison of LoRa with various other Internet of Things wireless communication protocols, as summarised in Table 2 (and visualised in Figure 7). Specific IoT applications, including the one discussed in this paper, require characteristics such as long-range capability, low data transmission rates, and minimal energy consumption. Short-range technologies such as Zigbee or personal area networks are unsuitable for environments requiring extensive coverage, leading to the exclusion of Zigbee and Wi-Fi standards from this analysis. While NB-IoT meets these requirements, its reliance on existing 3GPP networks and the need for costly network deployment limit its applicability relative to SigFox and LoRa. Technically, SigFox and LoRa share similarities. However, LoRa distinguishes itself by its open protocol and its ability to establish a cost-effective network, unlike SigFox, which relies on operator network coverage [16].

https://doi.org/10.31436/iiumej.v27i1.3720

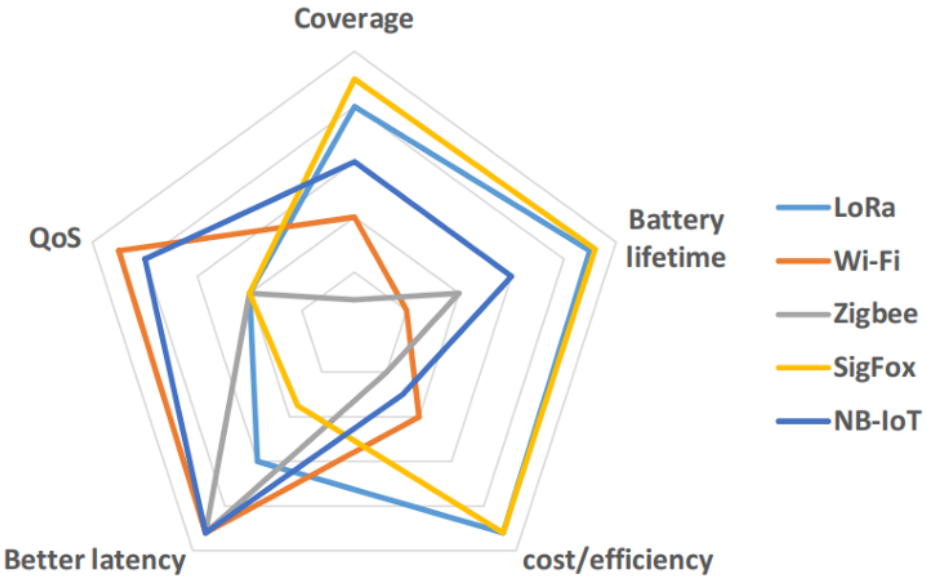


Figure 7. Comparison between IoT technologies [16].

Hence, based on previous studies, a LoRa/LoRaWAN communication network is deemed the most appropriate for the research proposed in this paper. This suitability stems from its broad coverage area, low cost, and low power consumption, which are critical to the successful implementation of this research.

Table 2. IoT communication technologies comparison [16].

	Lora	**Wifi**	**Zigbee**	**SigFox**	**NB-IoT**
Frequency	868 MHz (EU); 915 MHz (USA); 433 MHz (Asia)	2.4 GHz and 5 GHz	868 MHz (EU); 915 MHz (USA); 433 MHz (Asia); 2.4 GHz	868 MHz (EU); 915 MHz (USA); 433 MHz (Asia)	Depends on the frequency licensed to LTE
Standard	IEEE802.15.4g, LoRa Alliance	IEEE802.11	IEEE802.15.4	SigFox (Owner)	3GPP Standard
Coverage	5 km (urban), 20 km (rural)	50 m (indoor), 40 km (outdoor, as a function of the visibility)	10-100 m	10 km (urban), 40 km (rural)	1 km (urban), 10 km (rural)
Modulation	LoRa, FSK, GFSK	BPSK, QPSK, 16 QAM, 64 QAM, 256 QAM, 1024 QAM	BPSK, OQPSK	BPSK, GFSK	QPSK, OFDM (DL, SC-FDMA (UL))
Power Consumption	Low	High	Medium-Low	Low	Low
Theoretical Data Transfer Rate	22 kbps (LoRa), 100 kbps (GFSK)	2.4 Gbps (IEEE802.11 ax, 2 streams with 1024 QAM)	250 kbps at 2.4 GHz, 20 kbps at 868 MHz, 40 kbps at 915 MHz	100 bps	10 Mbps
Price of End Devices	3-5 €	3-5 €	2-5 €	>2 €	>20 €
Price of Gateway	100 € Gateway/ >1000 € Base station	20-600 € Gateway	40-1000 € Gateway	4000 € Base station	15000 € Base station

https://doi.org/10.31436/iiumej.v27i1.3720

3. RESEARCH GAP

Existing remote monitoring technologies for electric fences, such as Tennakoon et al. [12] and Al-Bahadly and Simpson [13], each provide only partial functionality and involve complex, case-specific implementations. Tennakoon et al. [12] designed a system that operates only with a ground-return electric fence and updates fault detections at 15-minute intervals via 4G GSM, thereby delaying the detection of sudden faults and restricting its use in areas without coverage. Similarly, Ali et al. [14] monitor only the power supply of the electric fence via 4G, which neither detects nor localises faults on the fence itself. Al-Bahadly and Simpson [13] measure voltage and current using a coil antenna and propose a drone-operated system. However, the proposed method has not been tested on an actual electric fence architecture.

Electric fences have shown strong potential for mitigating Human-Elephant Conflict (HEC), yet they are prone to voltage drops, short circuits, and structural deformations [12]. To address these issues, a monitoring system capable of integrating multiple measurement parameters is needed. The required parameters include the detection of high-voltage pulses (up to 10 kV), short-circuit fault localisation at 10 m intervals, and the fence leaning angle for structural stability. Data from these sensors also need to be transmitted wirelessly in near-real-time (at 1-minute intervals), a reporting interval consistent with best practices in other safety-critical infrastructure, such as pipeline damage sensing [20].

In-place wireless sensor nodes were selected over wired or drone-based methods due to their suitability for long, remote electric fence lines. Wired systems entail high installation and maintenance demands in large-scale infrastructure, whereas wireless nodes offer greater deployment flexibility and lower costs [21]. Drone-based inspections, while applicable for rapid area surveys, are limited by short flight times, weather, and frequent battery changes, making them unsuitable for continuous or minute-scale monitoring [22].

In this study, a single sensor node is developed that integrates a high-voltage (10 kV) sensor, a voltage-drop sensor for short-circuit localisation, and a gyroscope sensor for structural monitoring. This sensor transmits updates at one-minute intervals rather than the 15-minute cycles used in [12]. Data transmission is carried out via LoRaWAN, a low-power, long-range network that operates in areas with limited cellular coverage. While LoRaWAN requires an upfront investment in gateways, its lower energy consumption and absence of recurring data fees make it more practical than cellular solutions for rural, resource-constrained environments.

The functions of each proposed sensor and its role in the monitoring system are summarised in Table 3. To monitor the electrical health of an electric fence, a sensing mechanism capable of measuring up to 10kV is required. Additionally, to locate a short circuit on an electric fence, a voltage-sensing method is utilised by measuring the voltage drop along the fence. The structural integrity of an electric fence is monitored by measuring its leaning angle. In summary, a 10kV voltage sensor measures electric potential fluctuations. A DC voltage sensor measures voltage drops to locate short circuits. Finally, gyroscope sensors monitor structural deviations, which may indicate potential physical damage.

The integration of these diverse sub-sensors enables a multi-faceted approach to fault diagnostics. By synergistically combining these sensors and integrating them with IoT technologies (via LoRaWAN), technicians can remotely obtain a comprehensive understanding of fault scenarios at each monitoring node along the fence and expedite maintenance.

https://doi.org/10.31436/iiumej.v27i1.3720

Table 3. Sensors proposed for each malfunction.

Fault Type	Sensor Used	Sensor Response	Action Taken
Voltage Drop	10kV Voltmeter	Detects a drop below the threshold.	Alert sent to the technician. Maintenance scheduled.
Short Circuit	DC Voltage Sensor	Detects voltage drops.	Alert sent to the technician. Fault location identified.
Structural Damage	Gyroscope	Detects tilting or leaning of the fence.	Alert sent to the technician. Structural repairs initiated.

4. METHODOLOGY

To achieve comprehensive monitoring of electric fence integrity, an integrated sensor system was developed that incorporated a 10kV voltage sensor, a DC voltage sensor, and a gyroscope sensor. This system addresses critical operational challenges, including structural deformation, voltage fluctuations, and short-circuit faults, which are prevalent in electric fence systems designed to deter large animals such as elephants.

As illustrated in Figure 8, the developed system is divided into two components: the sensor block and the communication network block. The sensor comprises three sub-sensors: a 10 kV sensor, a DC-voltage sensor, and a gyroscope sensor. These sensors transmit raw data to a main Microcontroller Unit (MCU), an STM32. The STM32 processes the raw sensor data and compiles it into a 50-bit message. The MCU is also connected to an RFM95, which equips the sensor with LoRaWAN communication capabilities. The MCU then sends the processed sensor data to a central LoRaWAN gateway, which starts the data acquisition process. The gateway, connected to the internet via 4G LTE, then sends the data to The Things Network (TTN), which triggers a Fetch API request. The Fetch API in TTN pushes data from TTN to a Google server for analysis and storage. The data is stored in Google Sheets and analysed for anomalies exceeding the configured threshold. For any anomalies, a notification will be sent to the personnel involved to rectify the electric fence. The stored data is also processed and presented in a dashboard within a web application.

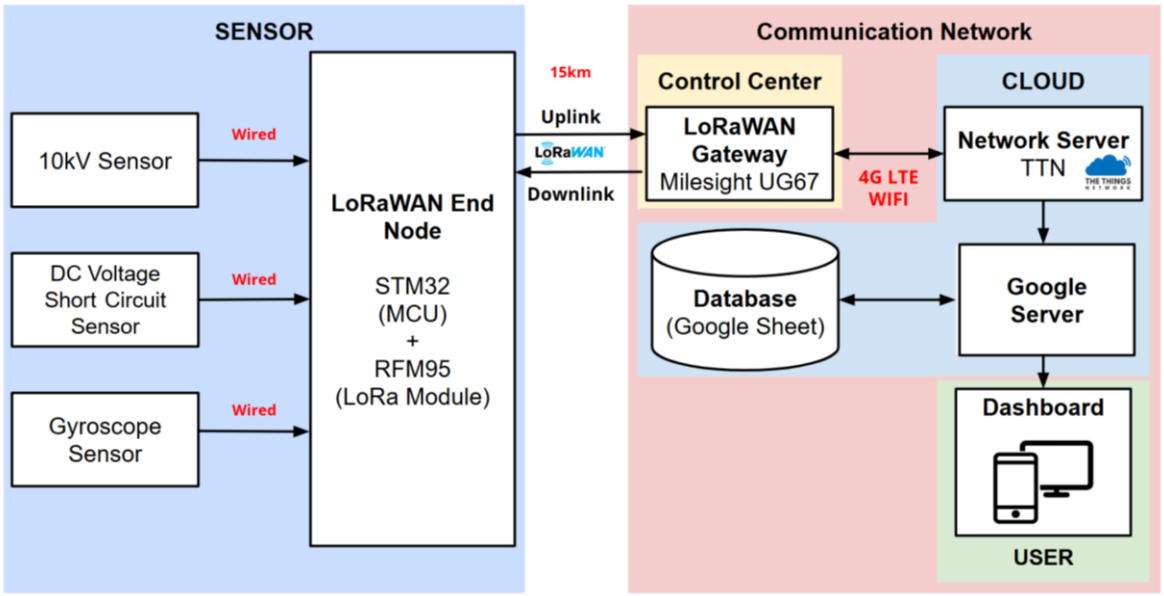


Figure 8. System architecture.

https://doi.org/10.31436/iiumej.v27i1.3720

The system requires a LoRaWAN gateway that is readily configured to the appropriate frequency, operating at 915 MHz in the Australian band (AU915). The monitoring system begins with the measurement of the sensors. Figure 9 illustrates the overall operation of the sensor system, in which the MCU first boots up, followed by the other modules. Once the sensors are powered on, the gyroscope is calibrated to offset all angles to 0°. Subsequently, the sensor data is transmitted to TTN, and the board enters sleep mode. Unless a button is pressed, the sensor will remain in its sleep state until its configured interval is reached. From TTN, the data will be sent to a Google server running Google Apps Script (GAS) for processing and analysis, as previously described. The GAS is also the entity that triggers emails to the appropriate personnel for malfunction detection.

The system's sensor node is divided into two parts: the primary circuit and a voltage divider circuit. As the voltage divider connects directly to a 10 kV energiser supply, its isolation from the main circuit board is paramount to ensure the operational integrity and longevity of the low-voltage components. As depicted in Figure 10, the main circuit, centered around the STM32F103C8T6 Microcontroller Unit (MCU), serves as the central hub for data acquisition, processing, and communication. It is supported by several peripheral modules, including a 915 MHz RFM95 LoRa module for wireless data transmission and an MPU-6050 3-axis gyroscope sensor.

The main circuit is supplied by a 9V battery, regulated by a 3A DC-DC buck converter to provide the necessary voltage levels for the microcontroller and other modules. An FT232 FTDI USB-to-TTL converter is also integrated to facilitate serial communication for programming, debugging, or data logging. The connection labeled 'To Voltage Divider (Load 1-8)' indicates the interface between the main circuit and the isolated high-voltage sensing unit, carrying a scaled-down PWM signal for measurement by the MCU.

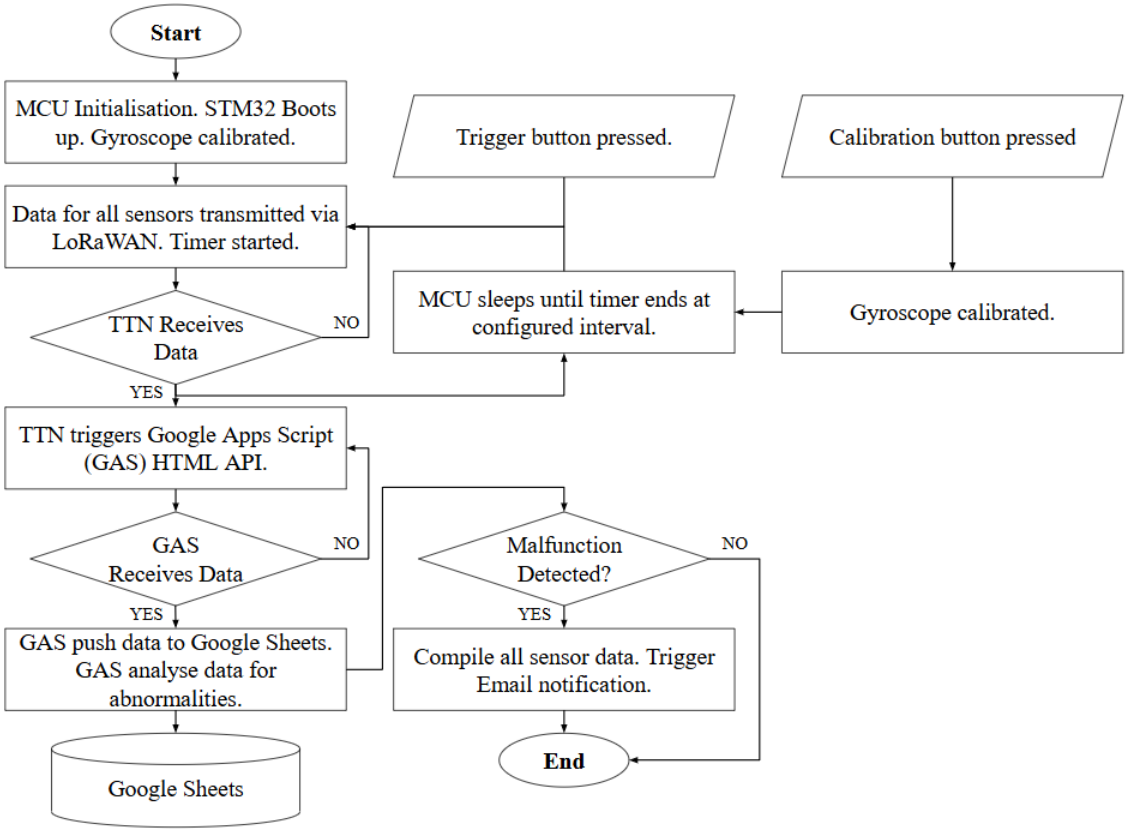


Figure 9. Monitoring system flowchart.

https://doi.org/10.31436/iiumej.v27i1.3720

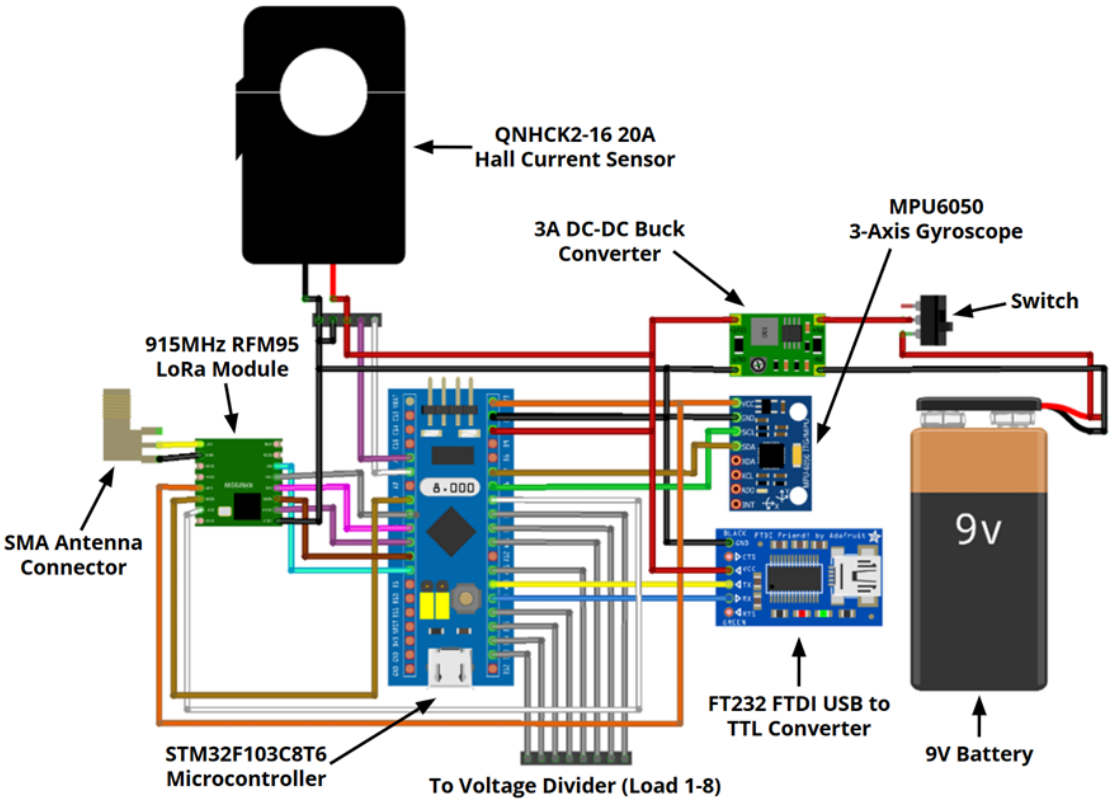


Figure 10. Microcontroller circuit diagram.

A crucial aspect of the sensor node's development and operation involves establishing a communication link with a computer. The STM32F103C8T6 microcontroller, while powerful, lacks a native USB interface for direct communication with a PC's USB port. Therefore, a conversion from the microcontroller's Universal Asynchronous Receiver/Transmitter (UART) serial communication protocol to the PC's USB interface is necessary. This conversion is achieved efficiently by integrating an FTDI FT232 USB-to-TTL converter module, as shown in Figure 11.

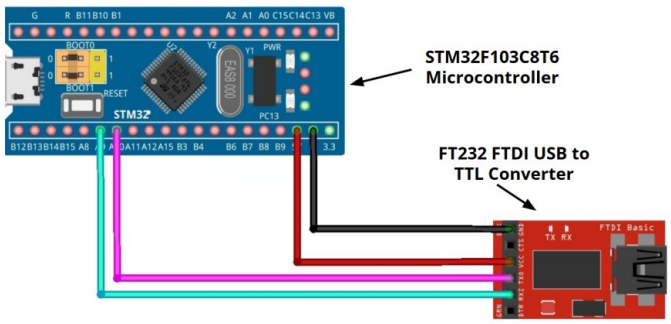


Figure 11. STM32 to FTDI FT232 TTL converter.

The FT232 module acts as a bridge, translating TTL (Transistor-Transistor Logic) level UART signals from the STM32 to the USB protocol understood by the PC and vice versa. It achieves bidirectional communication by connecting the STM32's Tx pin to the FT232's Rx pin and the STM32's Rx pin to the FT232's Tx pin. These cross-connections are essential for bidirectional serial communication. The ground (GND) pins of the STM32 and the FT232 module are also connected to establish a standard reference voltage. By establishing a proper connection between the board and a PC, the sensors may now be configured accordingly.

https://doi.org/10.31436/iiumej.v27i1.3720

4.1. High Voltage Monitoring (Voltage Sensor)

The first sensor developed is a 10kV voltage sensor. This sensor is based on the voltage divider principle, which is commonly used in electric fence voltage testers. A voltage tester was first reverse-engineered and tested to trigger an electronic relay, as shown in Figure 12. The developed high-voltage sensor was tested by dropping the voltage of an electric fence from 10 kV to 8 kV. Preliminary tests, as discussed in the next section, have demonstrated the effectiveness of the detection, and this development was later expanded to measure a broader range of voltage levels in an electric fence.

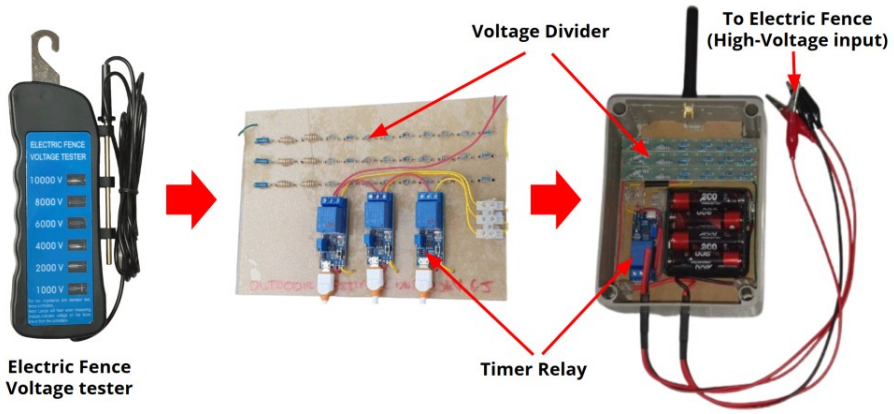


Figure 12. Preliminary 10kv voltage sensor fabricated.

A series of simulation studies has been conducted to determine the appropriate resistance to drop the high voltage to a measurable scale safely. As shown in Figure 13, the simulation involves measuring voltages across 8 resistors, each ranging from 5V to 20V. This voltage will then be connected to a MOSFET circuit to trigger a stable 3.3V to the input PWM pins of the STM32 MCU. In this simulation, all resistance values are calibrated to commercially available specifications. To ensure successful fabrication in the subsequent process, it is also essential to determine appropriate resistor values and power ratings. This is crucial for dissipating heat safely without damaging the resistors, as each resistor will carry a high voltage.

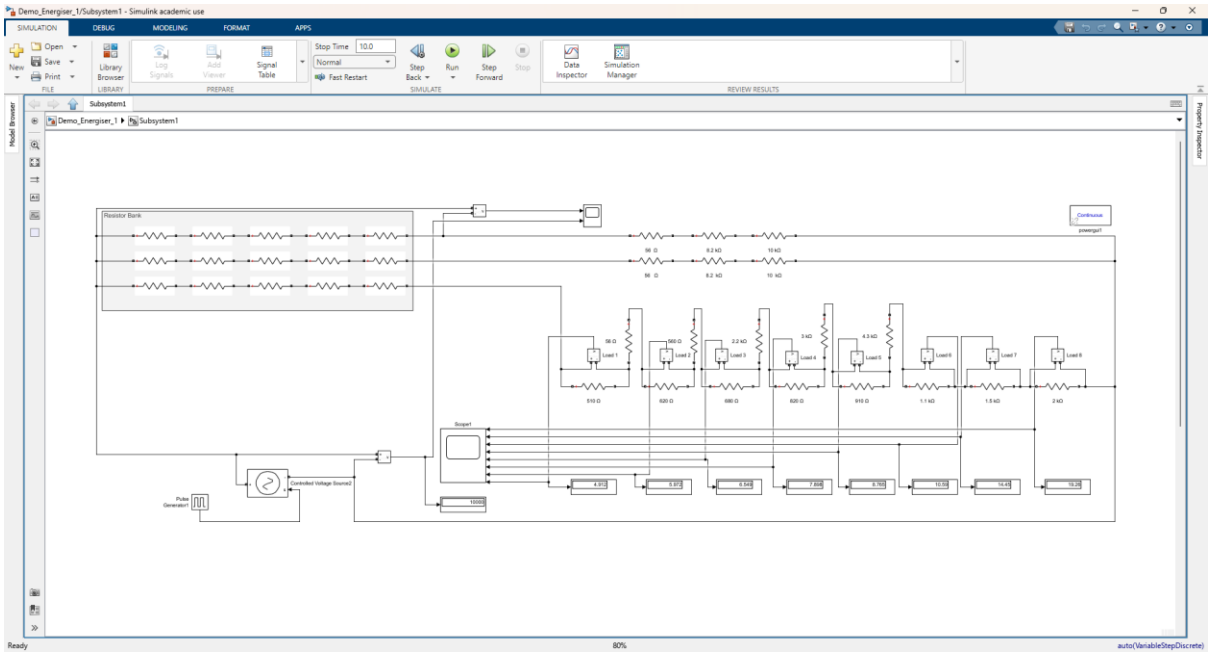


Figure 13. MATLAB-Simulink simulation of 10kV voltage divider.

https://doi.org/10.31436/iiumej.v27i1.3720

4.2. Short-Circuit Localisation (Short-Circuit Sensor)

To locate the short circuit along the electric fence system, a voltage-drop measurement method was proposed. To measure the voltage drop across the electric fence, a stable 12 V DC supply is utilised. As illustrated in Figure 14, this was achieved by connecting the positive terminal of the supply to the electric fence cable and the negative terminal to the ground (the electric fence pole). It is crucial during this process that the electric fence's high-voltage energiser switch be turned off to prevent damage to the 12 V DC supply.

During prototype testing, a manual mechanical switch with a wide contact gap was employed to isolate the 12 V DC supply from the 10 kV energiser for safety reasons. Therefore, the prototype did not operate in real-time during this testing phase. In a deployed system, this isolation will be achieved automatically using a motor-actuated or solenoid-actuated mechanical switch controlled by an MCU, enabling continuous or real-time monitoring without manual intervention. Incorporating such an actuator has been identified as part of future iterations of the system.

This was initially performed automatically using an electronic relay but was later discontinued as the relay contacts were too narrow, resulting in arcing and electrical discharge across the gap. It was found that a relay with a small gap cannot effectively switch off high-voltage pulses exceeding 1 kV. Hence, during testing, the manual switch was utilised, and a motor-actuated version is proposed for full automation during future deployment.

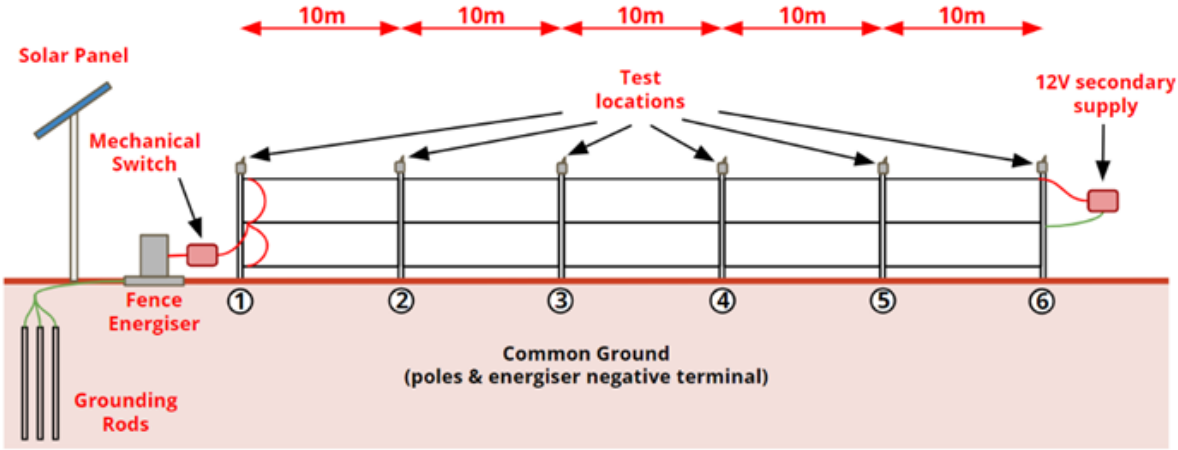


Figure 14. Electric fence voltage drop test setup.

The sensor unit, configured with a voltage sensor, was connected to each pole at the negative terminal and to the fence cable at its positive terminal. Then the electric fence's cable was short-circuited to each pole (connecting it to ground). Subsequently, the sensor measures and records the voltage across each pole.

Based on the test results discussed in the previous section, the proposed voltage-drop measurement approach (using a 12V DC supply and voltage sensors) demonstrated good performance. The method, however, is more suited for shorter electric fence segments, where the distance between monitoring points is limited. For longer electric fence setups, where the voltage-drop approach may increase overall monitoring costs (due to sensor installations), an alternative technique is being explored for future implementation. This method measures the current induced along the electric fence cables using a split-core Hall-effect current sensor (QNHCK2-16, 20 A). Integrated with the MCU for acquisition and processing (as shown in Figure 10), this sensor detects the small EMF-induced current that flows during normal operation and the significantly higher current associated with a short circuit. By placing these

https://doi.org/10.31436/iiumej.v27i1.3720

sensors at successive poles (similar to the previous method, without the 12V supply), it is expected that the point of highest current can be used to infer the short-circuit location.

This proposed approach builds on the fault-finding principle demonstrated in McGillan's handheld electric-fence current sensor [23]. It extends the concept of Al-Bahadly and Simpson [13] toward an automated, in-place monitoring framework, offering a practical direction for future work on long-range electric-fence monitoring.

4.3. Structural Integrity Monitoring (Gyroscope Sensor)

An MPU6050 Inertial Measurement Unit (IMU), featuring a 3-axis gyroscope, was integrated to monitor the electric fence's angular orientation, crucial for detecting potentially hazardous inclinations caused by external factors. The MPU6050 communicates with the STM32F103C8T6 via the I²C protocol, utilising the SDA and SCL lines for data exchange, along with power and ground connections, as shown in Figure 15. The gyroscope measures angular velocity about the X, Y, and Z axes (roll, pitch, and yaw), enabling the system to track the fence's real-time orientation and detect deviations from vertical alignment. The MPU6050 also includes a 3-axis accelerometer, enabling static orientation estimation and improved accuracy through sensor fusion with gyroscope data.

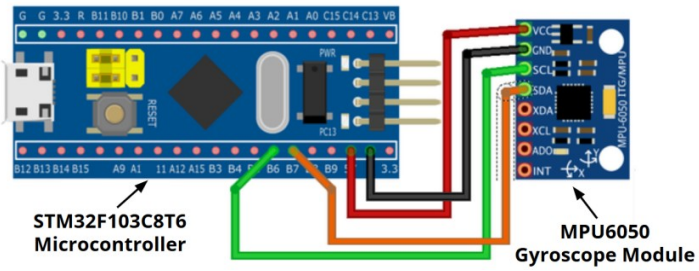


Figure 15. STM32 to MPU6050 connection.

This allows for continuous assessment of the fence's structural condition and the generation of alerts should critical inclination thresholds be exceeded, thereby mitigating the risk of short circuits and ensuring operational safety.

4.4. Data Acquisition and Rural Communication Network

To facilitate data transfer and communication between sensors and the system's IoT infrastructure, the Long-Range Wide Area Network (LoRaWAN) was configured. As illustrated in Figure 16, the system employs a multi-stage communication network to relay sensor data to end users. Sensor nodes utilise LoRaWAN for low-power, long-range wireless transmission to a central gateway. The gateway then uses TCP/IP over its internet connection to forward this data to the TTN Server. Subsequently, the TTN Server communicates with a Google Apps Script running on a Google server via the Fetch API to process and store data. Finally, end users can access the analysed data via a web application interface.

To apply in practice, a conceptual field deployment depicted in Figure 17 illustrates the strategic placement of sensor nodes along the electric fence line, with each node approximately 10 meters apart. These sensor nodes, equipped with an STM32 microcontroller, all sensors, and an RFM95 LoRa module, wirelessly transmit their collected data using the LoRaWAN protocol to a centrally located LoRaWAN gateway. The long-range capability of LoRaWAN is crucial for covering these distances in rural settings, thereby eliminating the need for extensive wired infrastructure. The conceptual field deployment also employs the automated fence actuator previously proposed as a future enhancement.

https://doi.org/10.31436/iiumej.v27i1.3720

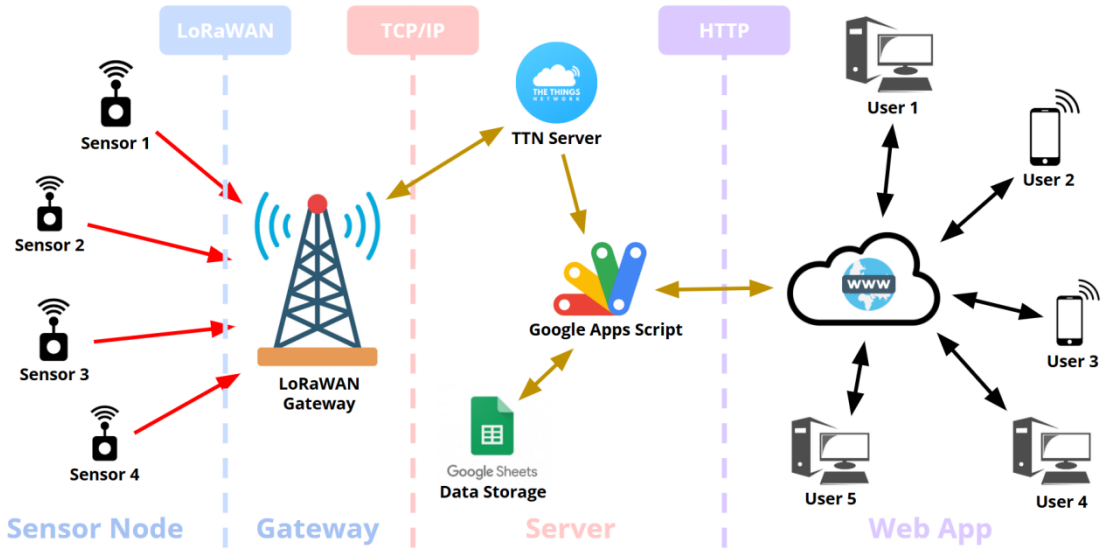


Figure 16. Overall system network topology.

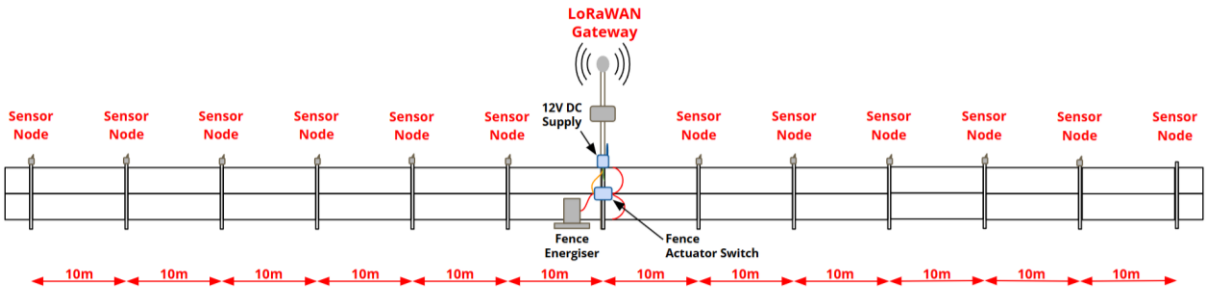


Figure 17. Conceptual field deployment.

The implementation process began with configuring a LoRaWAN gateway. The Milesight UG67 outdoor LoRaWAN gateway is used for this purpose. The gateway connects to the internet via a 4G SIM card. The gateway must be configured according to the region and frequency at which it is used. In this study in Malaysia, the licensed regional frequency is AU915, operating at 915 MHz. This needs to be done in the gateway's web UI. The exact configuration will then need to be applied to The Things Network (TTN) during gateway registration.

An end-node circuit, initially comprising an Arduino microcontroller and an RFM95 module was developed and configured to connect to the TTN server. Subsequently, the RFM95-configured Arduino end node was planned for integration with the MPU6050 gyroscope sensor. However, due to the ATmega328's limited 32 Kbytes of flash memory, integrating the RFM95 and MPU6050 modules consumed more than 80% of the available flash memory, causing errors during board initialization. The microcontroller was changed to an STM32F103C8T6 board with double the flash memory (64 Kbytes). Figure 18 shows the circuit that allows connection between the STM32 and RFM95 boards for LoRaWAN adaptation.

The Integrated Development Environment (IDE) used for this configuration remains the Arduino IDE, as it is compatible with STM32 boards and supports additional configurations and the installation of STM32 libraries. Currently, there are limited libraries that support the AU915 band. However, to apply this region and frequency, the MCCI_LoRaWAN_LMIC

https://doi.org/10.31436/iiumej.v27i1.3720

library must be modified to extend the US915 band to support AU915 via Activation by Personalisation (ABP). Before configuring the STM32, a TTN application and end-device registration must be completed. Because the RFM95 modules are not yet configured with an Extended Unique Identifier (EUI), a Device EUI must be generated and assigned to the STM32 board, along with the Network and App Session Keys generated during TTN configuration. This allows TTN to identify the sensor node when the gateway detects its uplink signal. Upon successful data acquisition, the other sensors were later integrated with the STM32, as shown in Figure 10.

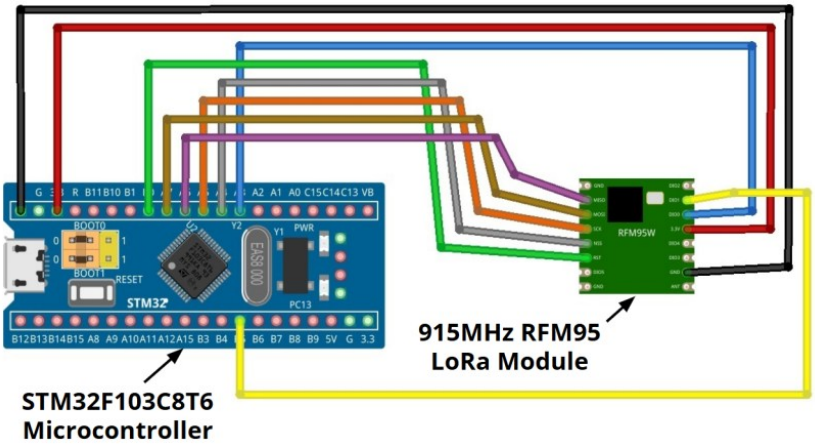


Figure 18. STM32 to RFM95 connection.

A cloud database (Google Sheets) was established to store historical data, enabling data analysis and visualisation via a web-based monitoring dashboard. The dashboard provides real-time indicators of electric fence malfunctions (voltage drop, structural leaning, and short circuit). The web application, accessible via a browser, serves as a manual watchdog for electric fence maintenance, demonstrating the potential of infrastructure monitoring in rural areas. Figure 19 shows the dashboard user interface (UI) operating in a web browser.

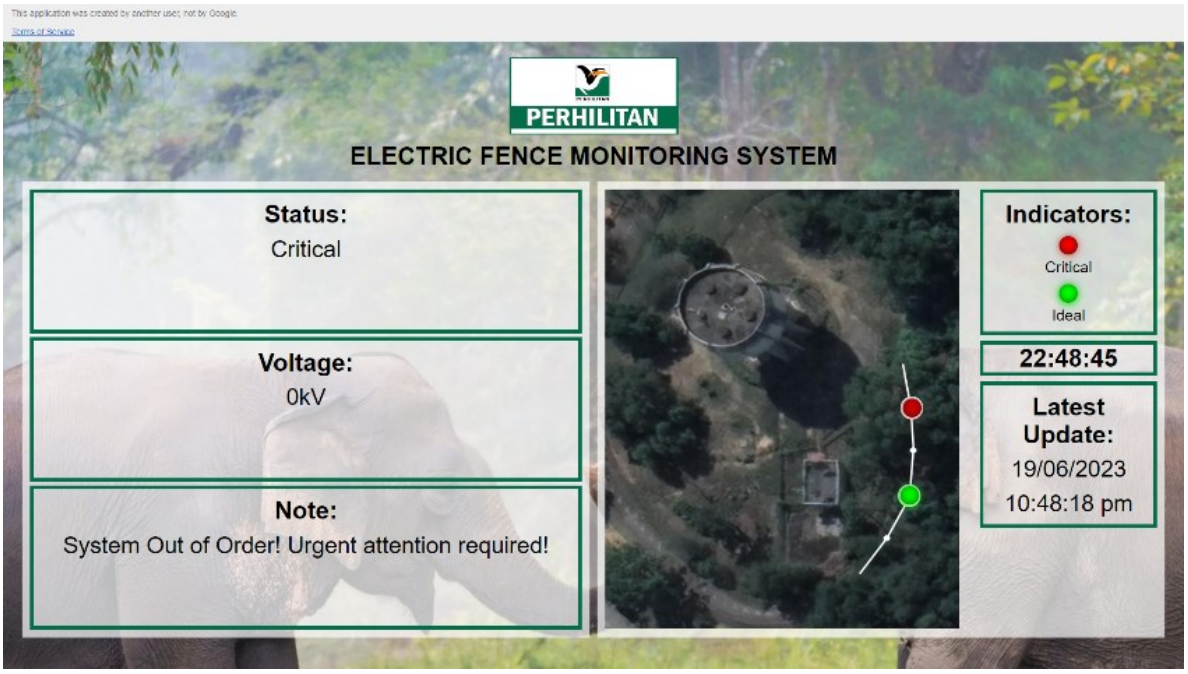


Figure 19. Developed system user interface.

https://doi.org/10.31436/iiumej.v27i1.3720

In addition to the manual watchdog, an automated notification system was implemented to alert technicians of malfunctions. Email notifications were configured to trigger alerts when sensor readings exceed predefined thresholds, indicating system anomalies. This approach, commonly used in industrial monitoring systems, ensures timely notification of site-related issues. Figure 20 shows an example of an email notification triggered. This system enables rapid response to faults, minimising overall downtime.

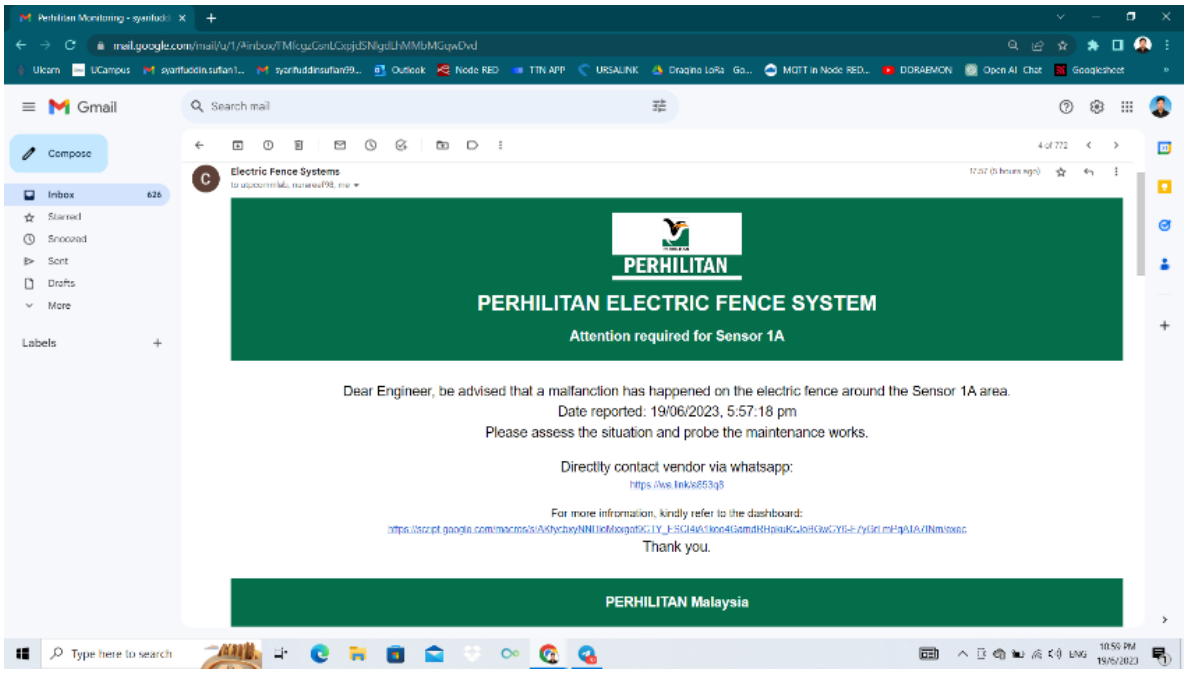


Figure 20. Email notification for fault detection.

4.5. Overall Pin Connection

After all configurations were done with the prototype sensor node, a complete pin configuration was finalised. Figure 21 illustrates the overall pin configurations of the sensor node's core components and MCU. The STM32F103C8T6 interfaces with the MPU6050 via I²C (SDA, SCL), the FT232 for USB communication (Rx, Tx), the voltage divider via digital input pins, and the RFM95 LoRa module via SPI and control pins. For the 915MHz LoRa communication, the RFM95's antenna output (ANT) connects to an SMA connector. To optimise signal integrity and reduce cable and connector losses, the PCB trace between the RFM95's ANT pin and the SMA connector should be kept as short as possible. Additionally, the RF signal paths should be carefully routed away from noise sources on the board. This is crucial for minimising electromagnetic interference and signal loss, thereby improving the Signal-to-Noise Ratio (SNR).

The core components of the final fabricated sensor node are shown in Figure 22. A 3-dBi omnidirectional monopole antenna is used for transmitting and receiving signals and is connected to an SMA connector via a coaxial cable. The placement of all components was also optimized to reduce the risk of short circuits. The fabricated sensor node also includes a Hall current sensor, which was previously proposed for future work.

https://doi.org/10.31436/iiumej.v27i1.3720

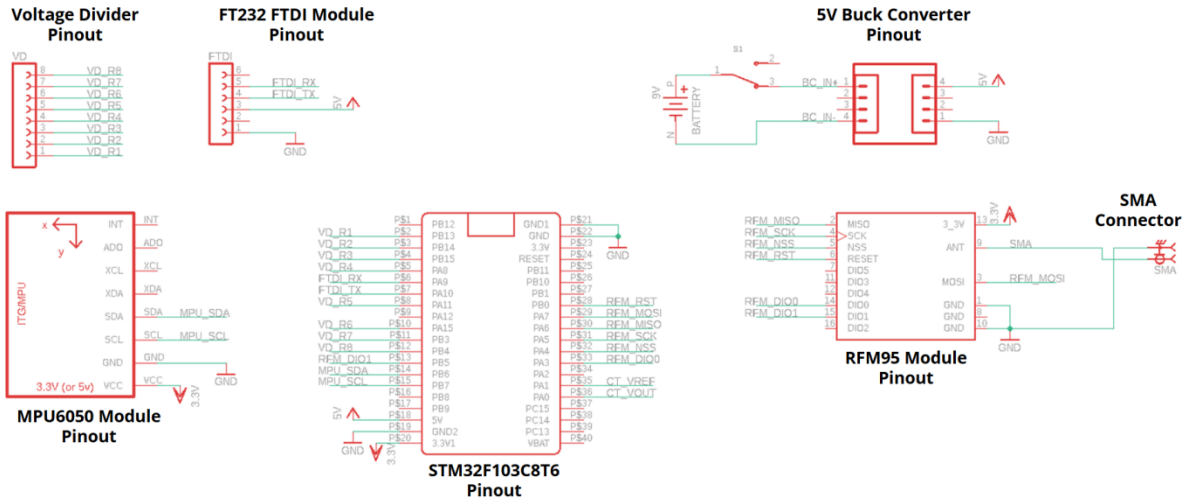


Figure 21. Overall pin connection.

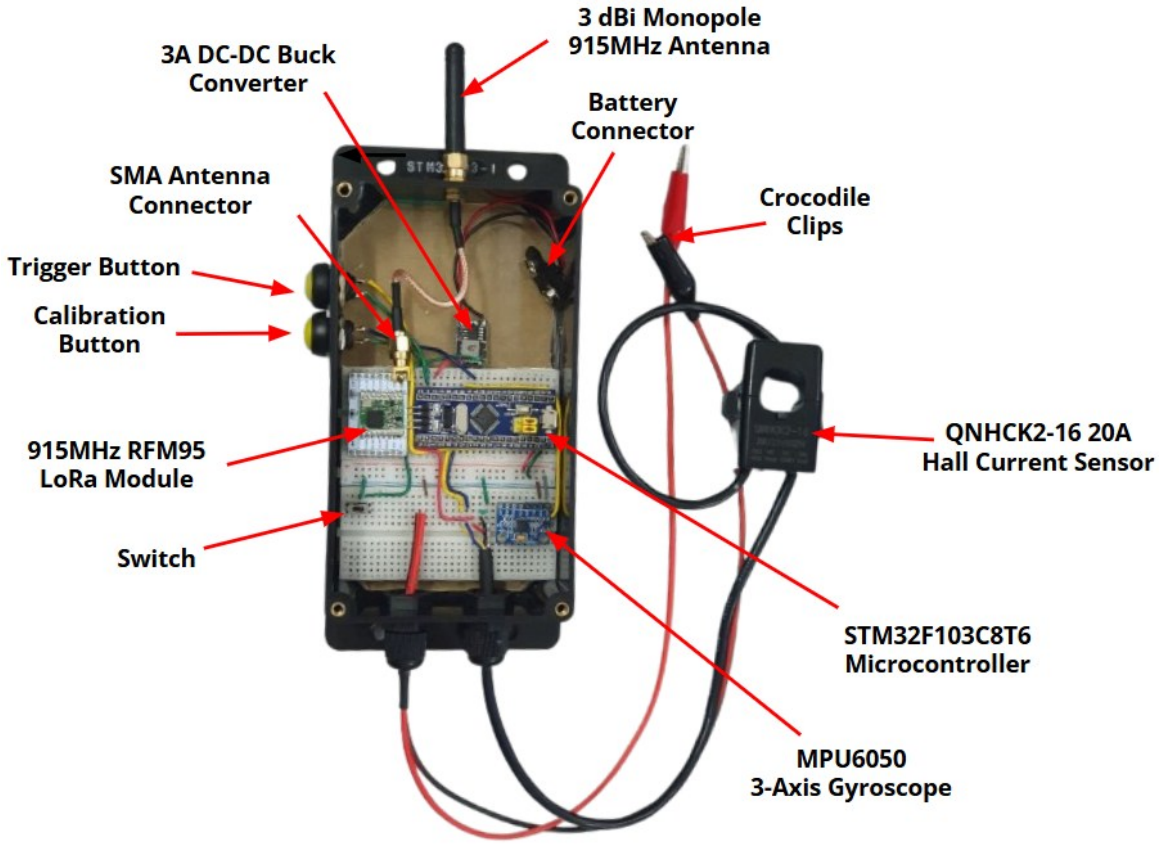


Figure 22. STM32 integrated sensor node.

5. RESULTS AND DISCUSSION

This section presents experimental and simulation results for the designed sensors, as well as the data acquisition and monitoring system configured over the LoRaWAN communication network. The experiments were conducted on a 50-meter electric fence sample provided by the Department of Wildlife and National Parks, Peninsular Malaysia (PERHILITAN). The setup featured 2.5 mm-thick galvanised angle-steel poles, spaced 10 meters apart, and grounded to the energiser's neutral. The fence utilised 2.5mm-diameter galvanised steel cables, energised

https://doi.org/10.31436/iiumej.v27i1.3720

by a 10kV, 6A energiser powered by a 12V SLA battery and a 100W solar panel. All specifications are designed similar to the Sistem Pagar Elektrik Gajah (SPEG) used by PERHILITAN. The test setup is illustrated in Figure 14. The tests were performed outdoors at the UTP campus in Seri Iskandar, Perak, under typical cloudy-day conditions, with an ambient temperature of approximately 33°C and a relative humidity of approximately 59%.

Due to the high electric fence voltage of 10kV, direct oscilloscope capture of the waveform requires safe attenuation through specialised isolation. Hence, this study does not analyse the waveform using an oscilloscope. Instead, the pulse characteristics of the energiser were analysed using the manufacturer's specifications [24], alongside the pulse parameters reported in a previous study by McGillan [23]. Based on [23], common electric-fence energisers generate very short 50-500µs high-voltage pulses between 4kV and 12kV, at a frequency of approximately 0.6Hz. This is consistent with the specifications of the energiser used in the test, which outputs high-voltage pulses of 10kV with a pulse duration of approximately 300 µs at 1.5s intervals (0.67 Hz) [24].

5.1. Voltage Sensor Performance

During this preliminary validation, several tests were carried out on the developed voltage divider circuit. The circuit was made to trigger relays connected in parallel to each resistor, as shown in Figure 12. The circuit was tested on two energiser setups: a 10kV 2.2A energiser and a 10kV 6A energiser. The energiser voltage was reduced by connecting the test circuit's input after a series of resistors (which absorb part of the 10kV supply load).

During testing of the 10kV, 6A energiser, the initial voltage-divider configuration exhibited cracking and visible arc-over of the 20 kΩ, ½ W resistor carrying the highest instantaneous load. This behaviour is consistent with the resistor's power tolerance being exceeded under high-energy pulses, resulting in mechanical failure of the resistive element and subsequent high-voltage discharge across the small physical gap ("arc-over") on the resistor. To address this, the divider network was redesigned into three identical parallel branches with equal total resistance, thereby distributing the pulse current and instantaneous power dissipation across multiple resistors and increasing the effective creepage distance between high-voltage nodes. This modification eliminated the observed arc-over and enabled stable operation under the 6 A energiser conditions. Table 4 shows the test results of the fabricated sensors.

Based on Table 4, the initial circuit determined whether the voltage was 10kV or 9kV (indicated by "YES" in the table), whereas 8kV was not detected (indicated by "NO" in the table). The modified circuit was later tested using the same test scheme as before. The results show that it can detect and determine the voltage of the electric fence (i.e., 10kV or 9kV). The sensor achieved 100% consistency across 10 tests, demonstrating its potential for detecting high DC voltages. This indicates that lower voltages can be detected by expanding the voltage divider circuit.

Table 4. Voltage sensor detection results

Detection	Initial Circuit			Modified Circuit		
	10kV	9kV	8kV	10kV	9kV	8kV
10kV 2.2A Setup	YES	YES	NO	YES	YES	NO
10kV 6A Setup	NO	NO	NO	YES	YES	NO

To detect the lower voltages (2kV to 8kV), the sensor was redesigned using the same concept of a potential divider. The circuit was simulated at these voltages using

https://doi.org/10.31436/iiumej.v27i1.3720

MATLAB/Simulink (as shown in Figure 13). The voltage across all 8 resistors was measured as the voltage was reduced from 10kV to 2kV in 1kV steps. With a 10kV supply, all 9 loads show voltages above 5V. When the supply voltages were reduced by 1kV, the voltage across the loads fell below 5V (see Figure 23).

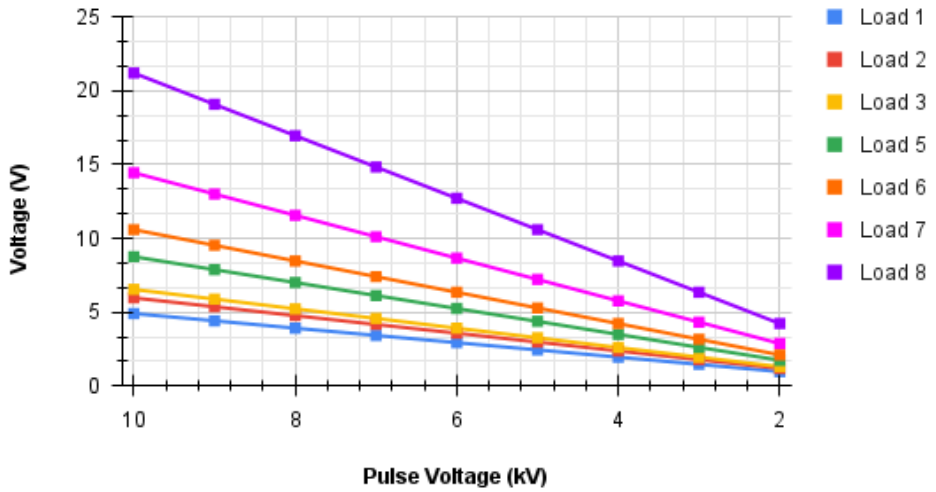


Figure 23. Voltage across each load at different pulse voltages.

The 5V voltage across each resistor will be used to trigger an IRFZ44N MOSFET that controls a 3.3V DC supply. For voltages above 5V, 3.3V will be used as a high digital input to the PWM I/O pins of the STM32 MCU. From the simulation results, reducing the energiser voltage from 10kV to 2kV reduces the total number of high digital inputs to the MCU, as shown in Figure 24. Based on experimental and simulation studies, the designed circuit can accurately determine the voltage level of the electric fence from 2kV to 10kV with a precision of 1kV.

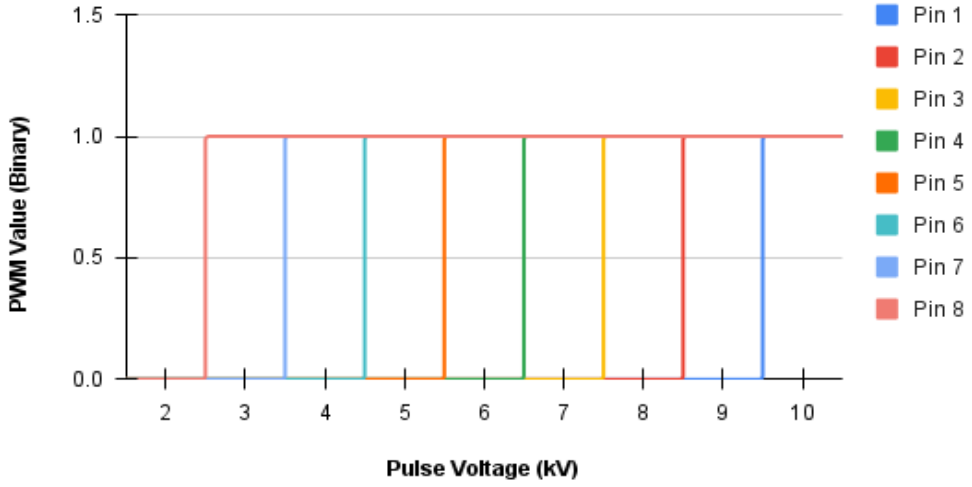


Figure 24. PWM levels from each load connecting to the MCU pins.

5.2. Short-Circuit Locator Performance

The initial short-circuit locator sensor, utilising grounding principles, was tested on the outdoor fence setup as shown in Figure 14. The developed sensor measures the voltage between the fence cable and the pole (ground). Based on the results tabulated in Table 5, under optimal conditions (without a short circuit), the measured voltage is 12V. When the second pole was shorted, the sensor measurement dropped to 0V at the second pole. For each pole, when shorted, the voltage between the cable and the ground falls to 0V. These results are consistent

across all short-circuit locations (poles 2-5) and were cross-checked using a conventional multimeter. This shows that, with a voltage sensor and a secondary 12V DC supply, the location of the short circuit can be deduced.

Table 5. Short-circuit locator sensor using a voltage-drop test.

Pole	Sensory Element	Voltage at optimum (V)	Short-Circuited Pole			
			Pole 2	Pole 3	Pole 4	Pole 5
1	Multimeter	12.09	11.92	12.03	12.03	12.04
	Sensor	11.99	11.92	11.92	11.92	11.94
2	Multimeter	11.74	0.00	7.27	6.27	7.78
	Sensor	11.20	0.00	7.07	6.09	7.54
3	Multimeter	11.38	4.09	0.00	5.75	7.42
	Sensor	11.25	4.19	0.00	5.53	7.14
4	Multimeter	11.30	4.45	6.96	0.00	7.22
	Sensor	11.23	4.29	6.77	0.00	7.05
5	Multimeter	11.36	4.39	6.98	5.68	0.00
	Sensor	11.00	4.22	6.67	5.46	0.00

While the 12V DC voltage-drop approach demonstrated consistent performance at 10m intervals, extending this configuration to multi-kilometre fence segments would proportionally increase the number of required sensor nodes, with associated cost and maintenance considerations for installations such as PERHILITAN's multi-kilometre SPEG systems. According to SPEG guidelines, a single energiser bay can extend to approximately 4km [6], which would necessitate around 400 sensor units for complete short-circuit coverage at 10m spacing. With an estimated cost of RM 500 per sensor unit, the total expenditure for monitoring one bay could reach RM 200,000, highlighting the importance of scalable solutions for long-range deployments.

As outlined in Section 5, a Hall-effect current-sensing technique employing split-core sensors such as the QNHCK2-16 has been identified as a potential alternative for long-range fault localisation. Although this method has not yet been experimentally validated within the present study, its feasibility is supported by the principles demonstrated in [13] and [23]. Accordingly, its implementation and performance evaluation are planned for future work.

5.3. Gyroscope Sensor Performance

The circuit shown in Figure 15 was fabricated and tested to evaluate the gyroscope sensor's reliability in detecting the leaning of the electric fence setup. The experiment involved inclining a pole in the X and Y directions at specific angles, which were measured using a digital angle meter attached to both a reference pole (vertical) and the test pole (leaning), as shown in Figure 25. Initial calibration established all axes at 0° when the pole was vertical. Measurements were then recorded from 0° to 45° in 5° increments, with ten readings taken at each angle for both axes.

https://doi.org/10.31436/iiumej.v27i1.3720

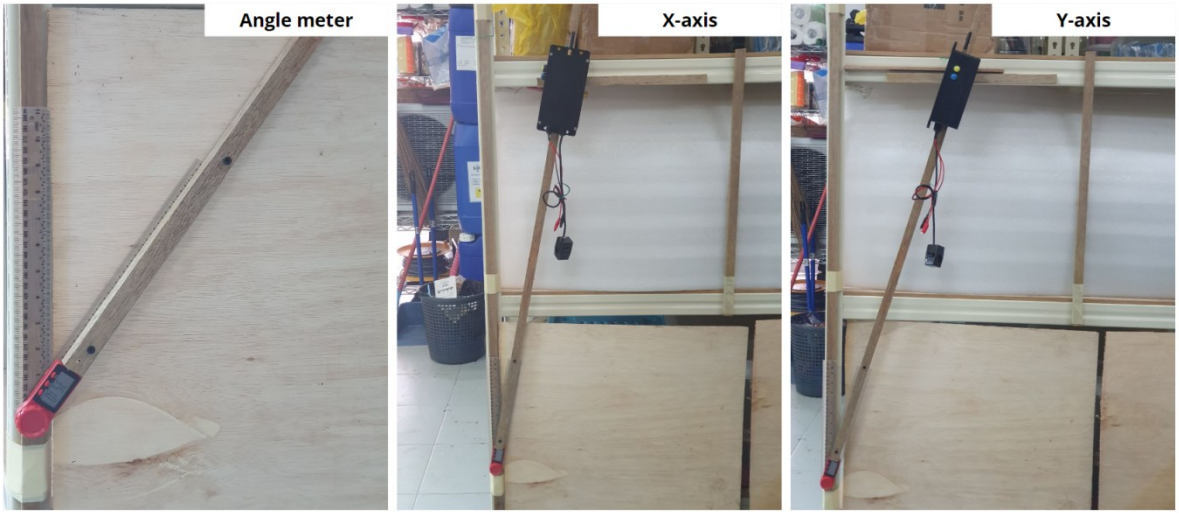


Figure 25. Indoor gyroscope test setup.

During initial testing, the MPU6050 sensor, when mounted directly on the circuit board, produced inaccurate X- and Y-axis readings. This included a high degree of variability along the X-axis and a nonlinear relationship between the actual and measured Y-axis angles. A diagnostic review identified three primary sources of error:

- *Axis misalignment*: the sensor had been oriented with its Y-axis vertical to the ground. According to the MPU6050 datasheet [25], the 'Orientation of Axes of Sensitivity and Polarity of Rotation' indicates that, for correct use in this configuration, the Z-axis should be perpendicular to the ground, which differs from the alignment used in the initial tests.
- *Mechanical instability*: the outdoor test poles were insufficiently constrained and gradually shifted under cable tension and gravity, causing inconsistent axis references across test repetitions.
- *Input disturbance*: angle readings were manually triggered by pressing a button on the sensor, introducing transient motion that distorted measurements.

Corrective measures were implemented before repeating the tests. The sensor was reoriented so that its Y-axis pointed upward, thereby making the MPU6050 board perpendicular to the sensor node board. Test poles were clamped to rigid brackets to maintain consistent orientation, and data acquisition was automated at one-minute intervals to eliminate button-press artifacts.

Following these adjustments, the repeated tests yielded stable, linear results. Scatter plots for both axes are shown in Figure 26 and Figure 27. Based on linear regression analysis of the two test datasets, the MPU6050 gyroscope sensor exhibited high linearity and accuracy for X- and Y-axis measurements under the tested conditions. The maximum deviation between measured and actual angles for the X-axis occurred at 40° (1.20°), and for the Y-axis at 45° (2.02°). Linear regression analysis comparing the sensor's output to actual angles measured by a digital protractor across a range of 0° to 45° yielded coefficients of determination (R^2) of 0.9991 for the X-axis and 0.9990 for the Y-axis. These values indicate a strong positive linear correlation across both axes, corresponding to an accuracy of approximately 99.9%.

https://doi.org/10.31436/iiumej.v27i1.3720

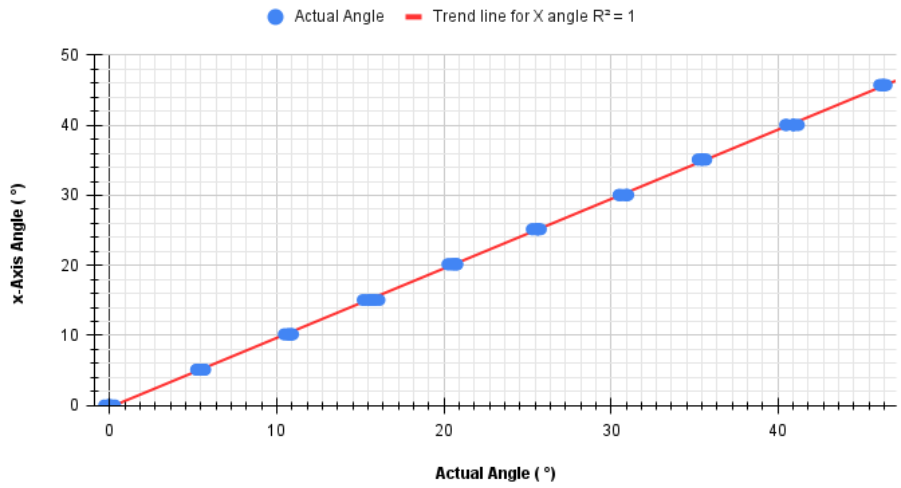


Figure 26. Scatter plot for gyroscope X-axis.

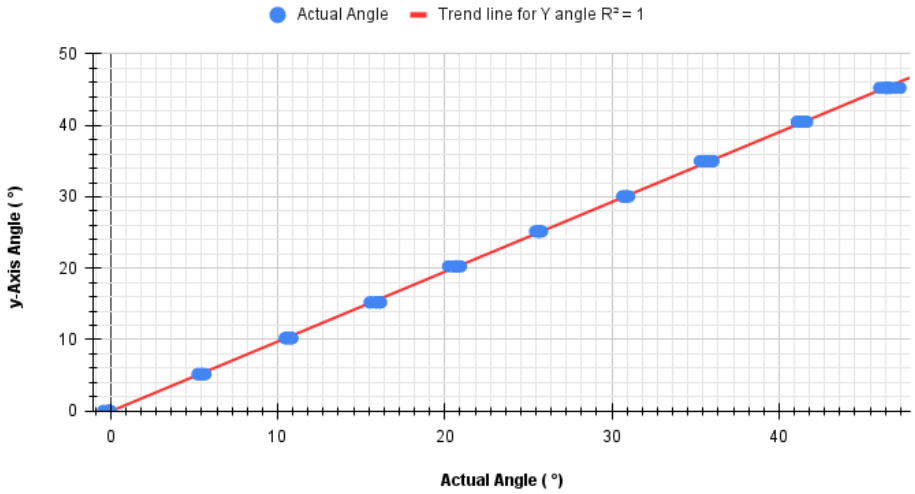


Figure 27. Scatter plot for gyroscope Y-axis.

5.4. Data Acquisition System Performance

The developed sensors were integrated into a LoRaWAN network via a Milesight UG67 industrial outdoor gateway and RFM95-equipped STM32 end nodes. Data were stored in a cloud-based spreadsheet (Google Sheets) and visualised on a web-based dashboard, which also generated automated notifications, as shown in Figures 19 and 20. The system operated continuously for 30 days without data loss, demonstrating its suitability for remote infrastructure monitoring.

To quantify the communication performance, an internal range test was conducted within the UTP campus compound. The UG67 gateway was installed on a 4 m pole at an altitude of 84 m, and the sensor node was moved in 100 m increments up to 1.3 km (13 test points). Figure 28 shows the test layout.

https://doi.org/10.31436/iiumej.v27i1.3720

Figure 28. Range testing conducted in the UTP campus compound.

The received signal power at each test point was calculated using the link-budget equation (1), where P_{RX} is the received power, P_{TX} is the transmitter power, G_{TX} and G_{RX} are the transmitter and receiver antenna gains, respectively. $FSPL$ is the free-space path loss, and L_{MISC} represents estimated cable and connector losses [26].

$$P_{RX} = P_{TX} + G_{TX} + G_{RX} - FSPL - L_{MISC} \quad (1)$$

The free-space path loss, $FSPL$ was calculated according to (2), where f is the operating frequency in MHz and d is the distance between transmitter and receiver in km [26]:

$$FSPL(dB) = 32.45 + 20log_{10}(d) + 20log_{10}(f) \quad (2)$$

Miscellaneous losses, L_{MISC} were taken as the sum of transmitter and receiver side cable losses (3):

$$L_{MISC} = L_{TXCable} + L_{RXCable} \quad (3)$$

Finally, the link margin was computed as the difference between the calculated received power and the receiver sensitivity (4):

$$Link\ Margin = P_{RX} - P_{Sensitivity} \quad (4)$$

The calculations described above were performed using parameters extracted from the RFM95W module and the UG67 gateway datasheets and user manuals [27-29], as summarised in Table 6. Both the transmitter and the receiver operate at 915 MHz with a 125 kHz bandwidth. According to the datasheets, the transmitter exhibits a sensitivity of -140 dBm at SF12, and the receiver exhibits a sensitivity of -148 dBm at SF12. As the configuration used in this study is SF7, the sensitivity of both devices was conservatively estimated as -120 dBm. The sensor node employed an omnidirectional monopole antenna with a gain of 3 dBi, whereas the gateway was fitted with two 60 cm GA01 omnidirectional antennas with a gain of 5 dBi. The transmitter was configured in the standard RFO output mode at +14 dBm (maximum +20 dBm with PA_BOOST mode), and the gateway has a maximum transmit power of +27 dBm. Losses due to cables and connectors were conservatively assumed to be 3 dB at each device (transmitter and receiver) to avoid over-optimistic link-budget results.

https://doi.org/10.31436/iiumej.v27i1.3720

Table 6. Transmitter (sensor node) and receiver (gateway) parameters [27-29].

Parameter	Transmitter (RFM95W node)	Receiver (UG67 gateway)
Frequency	915 MHz	915 MHz
Bandwidth	125 kHz	125 kHz
Sensitivity	-120 dBm @ SF7	-120 dBm @ SF7
Antenna gain	3 dBi	5 dBi
Output power	+14 dBm	+27 dBm (max)
Cable Loss	3 dB	3 dB

Table 7 summarises the measured RSSI, the calculated free-space path loss, the theoretical received power (based on link budget analysis), and the fade margin for each test point. As shown in Table 7, path loss increases progressively, and the theoretical and measured signal levels decrease with increasing distance. Up to approximately 0.9 km, the measured RSSI closely follows the theoretical received power. In contrast, beyond 1.0 km, the measured values diverge more strongly, reflecting the effects of environmental obstructions and near-ground propagation losses not captured by the free-space model. The fade margin also decreases steadily with distance, indicating the diminishing link reliability at longer ranges. These results provide the quantitative basis for the RSSI-distance and SNR-distance relationships presented in Figures 29 and 30.

Table 7. Measured and calculated link-budget parameters at each test point.

No.	Distance (km)	Measured RSSI (dBm)	Free-Space Path Loss (dB)	Theoretical Received Power (dBm)	Fade Margin (dB)
1	0.1	-66.0	71.7	-55.7	64.3
2	0.2	-64.0	77.7	-61.7	58.3
3	0.3	-59.0	81.2	-65.2	54.8
4	0.4	-58.0	83.7	-67.7	52.3
5	0.5	-57.0	85.7	-69.7	50.3
6	0.6	-58.0	87.2	-71.2	48.8
7	0.7	-57.0	88.6	-72.6	47.4
8	0.8	-72.0	89.7	-73.7	46.3
9	0.9	-74.0	90.8	-74.8	45.2
10	1.0	-102.0	91.7	-75.7	44.3
11	1.1	-106.0	92.5	-76.5	43.5
12	1.2	-106.0	93.3	-77.3	42.7
13	1.3	-110.0	94.0	-78.0	42.0

Figure 29 presents the measured RSSI, compared with the theoretical received power derived from link-budget calculations at each test point. Both the measured and theoretical values exhibit a gradual decline with increasing distance, consistent with the expected increase in free-space path loss. Up to approximately 0.9 km, the measured RSSI broadly follows the theoretical curve, indicating that the link performance is primarily governed by free-space propagation. Beyond 1.0 km, however, the measured RSSI diverges sharply from the theoretical prediction, falling to -110 dBm at 1.3 km. This decline reflects additional attenuation from environmental obstructions, ground effects, and near-field losses not captured by the free-space model. Although the gateway's nominal sensitivity at SF12 is -140 dBm, using SF7 in this study yields a higher effective threshold, making successful demodulation more difficult at lower signal levels. The observed deviation between the theoretical and measured curves at longer ranges, therefore, illustrates the combined effects of non-line-of-sight conditions and reduced link margin, thereby defining the practical communication range limit for the tested configuration.

https://doi.org/10.31436/iiumej.v27i1.3720

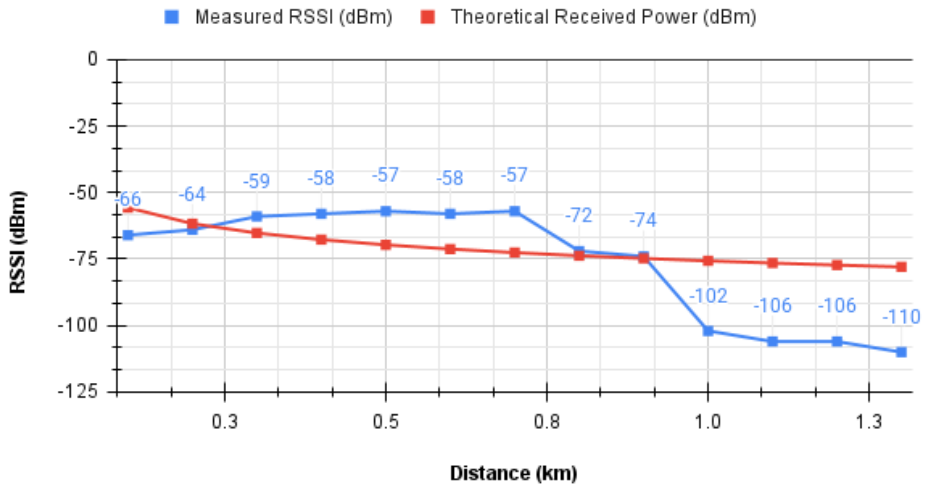


Figure 29. Measured RSSI and theoretical received power vs. distance.

Figure 30. SNR vs. Distance graph.

Figure 30 shows the variation of the Signal-to-Noise Ratio (SNR) with distance. SNR, which quantifies signal quality relative to background noise, exhibited small fluctuations at short ranges but showed a clear decline as distance increased. Beyond approximately 0.9 km, the SNR decreased sharply in parallel with the drop in RSSI, indicating that the link degradation was primarily caused by increased propagation loss rather than elevated noise levels. At the maximum test distance of 1.3 km, the low SNR values further limited the receiver's ability to reliably demodulate packets, demonstrating that maintaining adequate signal quality, not just signal power, is essential for long-range communication. This behaviour is consistent with observations reported in [30], in which both RSSI and SNR of LoRa-based links were found to decrease progressively with increasing range under field conditions.

The observed reduction in RSSI and SNR at longer distances was hypothesised to be influenced not only by environmental clutter but also by the relatively low installation heights of the transmitter and receiver antennas. In the initial configuration, the sensor node was positioned only 1 m above ground, whereas the gateway was mounted at 4 m, despite its elevation of 82 m. To examine whether antenna elevation could improve link performance, a supplementary test was conducted under the same conditions but with both antennas raised.

https://doi.org/10.31436/iiumej.v27i1.3720

A supplementary test was conducted at the same positions and at the same distance as shown in Figure 28. The repeated range test was conducted to evaluate the effect of raising the antennas of both the transmitter and the receiver on RSSI. During this test, both the gateway and the sensor were mounted on 8 m poles at the same points as in the previous test, as shown in Figure 31. The increase in the height of both the transmitter and the receiver yielded an RSSI value of -79 dBm at the 1.3km point, an improvement of 31 dB over the initial test configuration. This result is consistent with reports that increasing antenna height mitigates near-ground losses and improves line-of-sight [31]. Although LoRaWAN can achieve ranges of 15 km or more under ideal conditions, this study demonstrated a practical limit of 1.3 km within the campus environment, governed by antenna height, terrain clutter, and non-line-of-sight obstructions. The supplementary test, however, demonstrates that the developed system can be further improved by optimising the communication system design in future research.

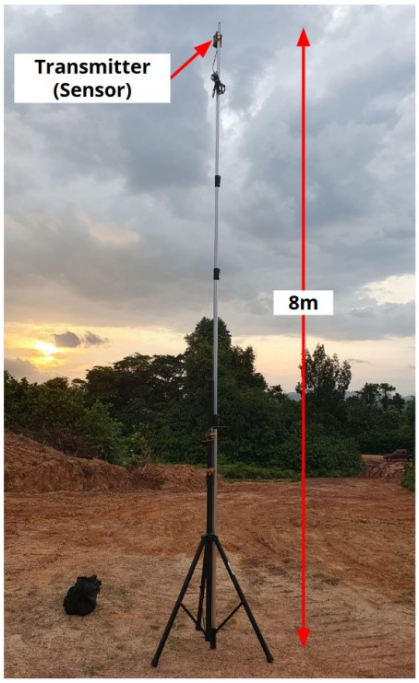


Figure 31. Transmitter applied to an increased height for a supplementary test.

6. CONCLUSION

The persistent challenge of Human-Elephant Conflict (HEC) in Malaysia has led to large-scale investment in electric fences, but maintaining these systems in remote areas remains costly and labour-intensive. This study demonstrated the feasibility of an integrated electric-fence monitoring system that combines a 10 kV voltage sensor, a short-circuit locator, and a gyroscope sensor, with a LoRaWAN-based data acquisition and a cloud-based monitoring dashboard. Field tests and simulation studies confirmed that the voltage sensor detected the electric fence voltage with 1 kV resolution. This demonstrates the feasibility of integrating high-voltage monitoring using the proposed method. The short-circuit locator sensor correctly identified fault locations using a 12 V reference supply. According to Sistem Pagar Elektrik Gajah (SPEG) technicians, fault-finding processes typically take 3-5 days to complete manually. The application of the short-circuit locator has the potential to reduce this fault-finding time to near-instantaneous, per-minute reporting. The gyroscope sensor can measure fence leaning with ≈99% accuracy, despite several design alterations made during testing, underscoring the need for proper orientation and a suitable test environment when using a gyroscope sensor. The developed data acquisition system using LoRaWAN link was also tested

over 30 days without data loss and achieved a practical range of 1.3km at SF7. A supplementary test showed that raising both the transmitter and receiver antennas further improved signal strength, indicating potential for additional range by underscoring the influence of antenna height and installation geometry on link performance.

These findings demonstrate that IoT-based monitoring can be adapted to rural electric-fence infrastructures to provide actionable, real-time data that has the potential to improve their overall reliability and system maintenance. Nevertheless, several limitations remain. The Hall-effect current-sensing approach for long-range fault localisation was only conceptually outlined and not yet validated. Range tests were conducted within a single campus environment and did not assess performance across different terrains or vegetation densities. Although the LoRaWAN link was stable for 30 days, longer-term reliability under field conditions and extreme weather remains to be established. At a 10m sensor node spacing, the cost for full coverage of a 4km energiser bay could reach RM 200,000, highlighting the need for scalable sensor placement strategies.

Future work will therefore focus on (i) implementing and experimentally validating the Hall-effect current-sensing approach for longer electric fence setups (kilometres-range), (ii) optimising antenna placement and LoRaWAN parameters to extend reliable coverage beyond 1.3 km, (iii) performing a detailed cost-benefit analysis for per-pole sensor deployment, and (iv) integrating predictive analytics to further reduce maintenance response time.

Overall, the study provides a quantitative foundation for designing and deploying remote electric fence monitoring systems and offers a practical pathway to enhance HEC mitigation by improving fence reliability and reducing maintenance effort.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Universiti Teknologi PETRONAS (UTP), particularly the Civil & Environmental Engineering and Electrical and Electronics Engineering Departments, for their essential support in facilitating this research. Sincere appreciation is also extended to the primary stakeholder of this project, the Department of Wildlife and National Parks Peninsular Malaysia (PERHILITAN), for their invaluable collaboration and assistance throughout this study.

REFERENCES

[1] Sinha S. (2021). Visible and hidden costs of human-elephant conflict on smallholders in Peninsular Malaysia. PhD thesis. University of Nottingham.

[2] Xin LQ, Ghani SNAM, Rasudin NS, Ghafar NA, Abd Rashid NH, Ten DCY, Abdullah MT. (2024). Characterisation of human-wildlife conflict and casualties caused by wildlife attacks in Peninsular Malaysia. *IIUM Medical Journal Malaysia*, 23(1).

[3] Wildlife Conservation Society Malaysia. (2024). Elephants. [Online]. Available: https://malaysia.wcs.org/Wildlife/elephants.aspx (accessed: April 10, 2025).

[4] Saaban S, Othman N. (2006). Pelan Pengurusan Gajah Semenanjung Malaysia. Jabatan Perlindungan Hidupan Liar dan Taman Negara, Kuala Lumpur.

[5] Department of Wildlife and National Parks Peninsular Malaysia. (2015). Laporan tahunan PERHILITAN 2015. Department of Wildlife and National Parks Peninsular Malaysia, Kuala Lumpur.

[6] Jabatan PERHILITAN Semenanjung Malaysia. (2023). *Buku Panduan Spesifikasi Peralatan dan Penyelenggaraan Sistem Pagar Elektrik Gajah (SPEG)*. 1st ed. Kuala Lumpur, Malaysia: Kementerian Sumber Asli dan Kelestarian Alam.

[7] Scholes RJ, Mennell KG. (2008). Controlling the Distribution of Elephants. Witwatersrand University Press, Johannesburg.

[8] Van Eden M, Ellis E, Bruyere BL. (2016). The influence of human–elephant conflict on electric fence management and perception among different rural communities in Laikipia County, Kenya. *Human Dimensions of Wildlife*, 21:283-296.

[9] Fernando P, Ekanayaka SKK, Pastorini J. (2020). The elephant at the fence: almsman, panhandler, friend or foe? *European Journal of Wildlife Research*, 66:97.

[10] Sukumar R. (2003). The Living Elephants: Evolutionary Ecology, Behavior, and Conservation. Oxford University Press, New York.

[11] Gallagher Group Ltd. (n.d.). *Electric Fencing 101: Electric Fencing Systems Design, Installation & Maintenance.* [Online]. Available: https://am.gallagher.com/-/media/Bynder/Animal-Management/Document/All-GGL-Manuals/Electric-Fencing-101-Manual-NA-original.pdf (accessed: April 10, 2025).

[12] Tennakoon E, Madusanka C, De Zoysa K, Keppitiyagama C, Iyer V, Hewage K, Voigt T. (2015). Sensor-based breakage detection for electric fences. In: *2015 IEEE Sensors Applications Symposium (SAS)*; April 2015; pp.1-4. IEEE.

[13] Al-Bahadly I, Simpson R. (2019). Contactless electric fence fault detection system. In: *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*; Auckland, New Zealand; 2019; pp.1-5. IEEE.

[14] Ali ASJ, Abdullah L, Musa M, Yunos MA, Ki NWW, Tukiran Z, Zainuddin MHAJ. (2020). Towards IoT-based notification system for agriculture electric fence. In: *2020 IEEE Student Conference on Research and Development (SCOReD)*; September 2020; pp.269-273. IEEE.

[15] Yunus MABY, Zainal MSB. (2023). IoT-based intelligent system for farm electric fence. *Progress in Engineering Application and Technology*, 4(2):274-286.

[16] Sendra S, García L, Lloret J, Bosch I, Vega-Rodríguez R. (2020). LoRaWAN network for fire monitoring in rural environments. Electronics, 9(3):531.

[17] Bouras C, Gkamas A, Kokkinos V, Papachristos N. (2022). Performance evaluation of monitoring IoT systems using LoRaWAN. *Telecommunication Systems*, 79:295-308.

[18] Khriji S, Günyeli ÖK, El Houssaini D, Kanoun O. (2022). Energy-efficient short-long range communication network combining LoRa and low-power radio for large-scale IoT applications. In: *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*; Chemnitz, Germany; 2022; pp.1-6. IEEE.

[19] Senjaya A, Zulkifli FY. (2019). Low-cost continuous-wave radar system design for detection of human respiration rate. In: *2019 IEEE R10 Humanitarian Technology Conference (R10-HTC)*; Depok, West Java, Indonesia; 2019; pp.90-92. IEEE.

[20] Chen YH, Shen SC, Wu YK, Lee CY, Chen YJ. (2022). Design and testing of real-time sensing system used in predicting the leakage of subsea pipeline. Sensors, 22(18):6846.

[21] Fasasi MO. (2024). Comparative analysis of wired and wireless sensors in structural health monitoring. *Open Journal of Engineering Science*, 5(1):38-49.

[22] Oleksyn S, Tosetto L, Raoult V, Joyce KE, Williamson JE. (2021). Going batty: The challenges and opportunities of using drones to monitor the behaviour and habitat use of rays. *Drones*, 5(1):12.

[23] McGillan G. (2009). The design of an electric fence fault-finder. Master's thesis. Massey University, Albany, New Zealand.

[24] Datamars Ltd. (2023). 6 J, 12 J Unigizer™ User Guide, Issue 6. [Online]. Available: https://pel.co.nz/resource-hub/user-guides/406i-412i-unigizer-user-guide-enesptfrde (accessed: April 10, 2025).

[25] InvenSense. (2013). MPU-6000 / MPU-6050 Product Specification. Revision 3.4, Document Number PS-MPU-6000A-00.

[26] Campbell Scientific. (n.d.). *Application Note – The Link Budget and Fade Margin*. [Online]. Available: https://s.campbellsci.com/documents/us/technical-papers/link-budget.pdf (accessed: September 15, 2025).

[27] Milesight. (2025). *UG67 Datasheet.* [Online]. Available: https://resource.milesight.com/milesight/iot/document/ug67-datasheet-en.pdf (accessed: September 16, 2025).

[28] HopeRF. (n.d.). *RFM95W-V2.0 Datasheet.* [Online]. Available: https://cdn.sparkfun.com/assets/a/9/6/1/0/RFM95W-V2.0.pdf (accessed: September 17, 2025).

[29] MCCI. (2025). GA01 Gateway. [Online]. Available: https://store.mcci.com/products/milesight-ga01 (accessed: April 10, 2025).

[30] Križanović V, Grgić K, Spišić J, Žagar D. (2023). An advanced energy-efficient environmental monitoring in precision agriculture using LoRa-based wireless sensor networks. Sensors, 23(14):6332.

[31] Gutiérrez-Gómez A, Rangel V, Edwards RM, Davis JG, Aquino R, López-De la Cruz J, Geng Y. (2021). A propagation study of LoRA P2P links for IoT applications: The case of near-surface measurements over semitropical rivers. Sensors, 21(20):6872.

https://doi.org/10.31436/iiumej.v27i1.3824

Enhanced Beach Photo Translation using Modified Unsupervised GAN with Regularization

KARTIKA FITHRIASARI*, BENEDICTUS KENNY TJAHJONO

Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

**Corresponding author: kartika_f@its.ac.id*

(Received: 1 July 2025; Accepted: 26 September 2025; Published online: 12 January 2026)

ABSTRACT: To optimize time and cost, tourists often require tools to modify the sky background and atmosphere in beach photos, such as replacing blue-sky views with sunsets or vice versa. Independent modification of the sky and sea is difficult because their color palettes are similar. Another problem that often occurs in image translation is the scarcity of paired datasets, and beach photo datasets are particularly limited in lighting conditions, weather variations, and viewing perspectives. This limitation can cause Generative Adversarial Networks (GAN) models to lose their generalization ability, become prone to overfitting, and produce visual artifacts in the outputs. Therefore, this study proposes an unsupervised GAN approach using a modified CycleGAN and improves its performance for beach image translation by integrating identity mapping, λ-parameter optimization, a multiscale kernel, and regularization techniques. CycleGAN consists of two generators and two discriminators. The sunset generator translates a blue sky into a sunset sky; the generated output is then passed to the sunset discriminator to determine whether it is real or fake. The generator input image is resized and normalized through preprocessing. The generator architecture is structured to enhance image reconstruction and feature extraction. The details of the translation results are fine-tuned using a 30x30 PatchGAN discriminator and a multiscale kernel convolutional layer. The effect of the hyperparameter λ, which strikes a balance between cycle consistency, structural preservation, and color fidelity, is also investigated in this work. The findings indicate that while higher λ values increase generator loss, they also improve consistency, making it harder to handle dark objects and white clothing. To overcome this issue, regularization techniques, namely photometric augmentation and spectral normalization (SN), together with multiscale kernel convolutional (MSCov), have been applied. Photometric augmentation and MSCov are used to enhance the model's robustness to photographic variations, while SN improves its efficiency and stability. The results of the study show that the proposed method improves image translation accuracy as measured by Mean Squared Error (MSE), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS).

ABSTRAK: Bagi mengoptimum masa dan kos, pelancong sering memerlukan alat menukar foto latar belakang langit dan suasana pantai, seperti mengganti pemandangan langit biru dengan matahari terbenam atau sebaliknya. Pengubahsuaian bebas langit dan laut sukar dilakukan kerana palet warna langit dan laut adalah serupa. Masalah lain sering berlaku dalam menterjemah imej adalah ketiadaan set data foto berpasangan dan foto pantai sering mengalami kepelbagaian terhad, terutama dari segi pencahayaan, variasi cuaca dan perspektif tontonan. Had ini boleh menyebabkan model Rangkain Generatif Adversari (GAN) kehilangan keupayaan generalisasi, terdedah kepada terlebih muat dan penghasilan artifak visual dalam hasil terjemahan. Oleh itu, kajian ini mencadangkan pendekatan GAN tanpa pengawasan menggunakan CycleGAN yang diubah suai bagi meningkatkan prestasi terjemahan imej pantai melalui penyepaduan pemetaan identiti, pengoptimuman parameter , kernel berbilang skala dan menggunakan teknik regularisasi. CycleGAN terdiri daripada dua generator dan dua rangkaian neural diskriminator. Generator matahari terbenam digunakan

dalam menterjemah langit biru kepada langit matahari terbenam, kemudian imej terhasil dimajukan kepada diskriminator matahari terbenam bagi menentukan sama ada imej terhasil dikelaskan sebagai imej sebenar atau imej palsu. Imej input generator diubah saiz dan dinormalkan melalui prapemprosesan. Seni bina generator distrukturkan dengan meningkatkan pembinaan semula imej dan pengekstrakan ciri. Butiran hasil terjemahan diperhalusi menggunakan diskriminator PatchGAN 30x30 dan lapisan konvolusi kernel berbilang skala. Kesan hiper parameter turut dikaji bagi mencapai keseimbangan antara ketekalan kitaran, pemeliharaan struktur, dan kesetiaan warna. Dapatan kajian menunjukkan bahawa walaupun nilai lebih tinggi ianya meningkatkan kehilangan generator dan konsistensi, menjadikannya lebih sukar dalam mengendali objek gelap dan pakaian putih. Bagi mengatasi isu ini, teknik regularization iaitu pembesaran fotometrik dan normalisasi spektral (SN) bersama Konvulasi Kernel Skala Berbilang (MSCov) telah digunakan. Pembesaran fotometrik dan MSCov dilaksanakan dalam meningkatkan keteguhan model pada variasi fotografi, manakala SN digunakan bagi meningkatkan kecekapan dan kestabilan model. Hasil kajian menunjukkan kaedah ini mampu meningkatkan ketepatan hasil terjemahan imej berdasarkan Ralat Purata Kuasa Dua (MSE), Indeks Persamaan Struktur (SSIM) dan Persamaan Tampalan Perseptual Imej Terpelajar (LPIPS).

KEYWORDS: *Beach Photo, CycleGAN, Multiscale Kernel Convolutional, Unsupervised GAN, Image Translation*

1. INTRODUCTION

Among the most visited tourist spots is the beach. The sunset view is one of the most common themes in tourist beach photography. Due to conditions such as cloudy weather or limited afternoon time, tourists may be able to take photos only in the morning or evening, rather than at sunset. To save time and money, a method is needed to change the atmosphere of beach photos as desired; for example, changing a beach photo with a blue sky to a sunset view and vice versa. Achieving a realistic transformation is challenging. The main difficulty is that the sky, which must be transformed, has a color similar to that of the sea, an element that must be preserved. An efficient method is needed to replace the sky background while maintaining the other elements, ensuring that the result realistically reflects the desired atmosphere and scenery.

The image-to-image translation process entails analyzing the image's content and modifying it to conform to the target domain's features. If X refers to the source domain and Y refers to the target domain, then the translation process can be denoted as $X \rightarrow Y$. For example, in beach photography, the X may represent photographs of the beach during the day, while the Y could represent the images during sunset, or vice versa. Image-to-image translation, using computer vision algorithms, offers a practical solution to challenges in beach photography, enabling tourists to capture better images and preserve their holiday memories.

Generative Adversarial Networks (GANs) are neural networks designed to solve image-to-image translation tasks, particularly in applications that require paired images [1]. In reality, obtaining paired images is often impossible. This led to the development of unsupervised GANs. One notable unsupervised GAN developed is CycleGAN [1]. CycleGAN incorporates a cycle-consistency loss that aims to preserve the image's structure when it is converted back to its original domain. This study further improves CycleGAN's performance by modifying its network architecture. The generator uses a transposed convolutional layer, while the discriminator uses a PatchGAN with a 30×30 resolution. This is intended to improve the recognition of fine details in CycleGAN. The hyperparameter λ controls the balance between image realism and structural preservation, thereby affecting the quality of the resulting image

https://doi.org/10.31436/iiumej.v27i1.3824

[2,3]. Optimization of λ is necessary to improve the performance of CycleGAN, so in this research, the impact of hyperparameters was evaluated.

The main challenge in translating daytime beach images into sunset scenes is the substantial color similarity between the sky and the sea, which can cause unwanted changes across the entire image. To address this, we incorporate identity mapping into CycleGAN, ensuring that the generator learns to preserve regions that require little modification. This prevents distortions of the natural landscape's structure and helps maintain the original color balance.

Furthermore, we investigate the role of the hyperparameter λ, which regulates both cycle consistency loss and identity mapping. By fine-tuning λ, we control the trade-off between realism and structural preservation, allowing for a more precise and visually coherent transformation. Our approach ensures that changes are selectively applied to areas requiring modification while preserving the integrity of unaltered regions.

Another challenge in beach photo datasets is that images are often limited to specific lighting conditions or weather states, which limits the model's ability to capture the underlying visual complexity of beach environments. As a result, models trained on such data may struggle to generalize, leading to less robust translations and reduced realism in generated images. Addressing these issues requires approaches to increase image variation, particularly in lighting and color, to improve model stability. Therefore, this study applies photometric augmentation to the data to enhance the accuracy of CycleGAN [4]. In unpaired image translation tasks, the stability of the discriminator is crucial to ensure that the generator continues to learn effectively from the discriminator's feedback. To maintain signal stability and improve efficiency, spectral normalization (SN) is applied to the discriminator.

The contributions of this research include improving CycleGAN's performance for beach image transformation by integrating hyperparameter optimization and regularization techniques, namely photometric augmentation and SN. These methods are applied and evaluated using Mean Squared Error (MSE), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) to measure the quality of the translation results. This study not only offers a practical solution for tourists aiming to enhance their beach photography but also contributes to advancements in image processing and computer vision.

2. METHODS

2.1. Digital Image

An image can be represented as a matrix of RGB intensity values with U rows and V columns, where (u, v) denotes spatial coordinates, and $x(u, v)$ denotes the color intensity at a given point. RGB in an image consists of three primary color channels: red, green, and blue, each represented by eight bits, resulting in a total of 24 bits per pixel.

Understanding the structure of a digital image is essential for analyzing its characteristics, such as color distribution and contrast. Such distribution patterns can provide insight into the variation of pixel intensities in an image. The frequency of pixel intensities in a given color channel at a specified intensity scale can be determined from the distribution plot.

2.2. Photometric Augmentation

For outdoor images such as beach landscapes, the translation model must be capable of handling a wide range of photo conditions, especially in unpaired data translation tasks. In such cases, the model is prone to overfitting. Variations in lighting conditions can lead to a mismatch

https://doi.org/10.31436/iiumej.v27i1.3824

between the source and target domain distributions. To address this issue, photometric augmentation is applied to increase the diversity of training data images [5].

These photometric transformations are applied before the image is fed into the generator network. The transformations used in the augmentation process include brightness adjustment, contrast variation, saturation shift, and cutout (to simulate information loss), applied randomly with parameters constrained to specified ranges. If I denotes the intensity of the image to be augmented, the brightness transformation is performed according to Eq. (1) with the parameter $\alpha \in [a, b]$. This transformation produces image variations that are either darker or brighter.

$$I^* = I + \alpha. \tag{1}$$

For the contrast variation augmentation technique, intensity rescaling is performed using Eq. (2) with the parameter $\beta \in [a, b]$ and μ representing the mean image intensity.

$$I^* = (I - \mu) \cdot \beta + \mu. \tag{2}$$

Image saturation is transformed by converting the RGB image into the HSV (Hue, Saturation, Value) color space and modifying the Saturation (S) component as shown in Eq. (3) with the parameter $\gamma \in [a, b]$. After modification, the image is converted back to the RGB format.

$$S^* = \gamma \cdot S. \tag{3}$$

Translation augmentation is used to enable the model to capture positional variation. Let $(\Delta p, \Delta q)$ represent the horizontal and vertical shift values, respectively. Then, a random shift can be performed using Eq. (4).

$$I^*(p, q) = I(p - \Delta p, q - \Delta q). \tag{4}$$

Cutout augmentation is a data augmentation technique applied to images by randomly masking out a square region of the input during training. The cutout augmentation process begins with an input image $x \in \mathbb{R}^{H \times W \times C}$ of height H, width W, and color channels C. The size of the cutout region is determined by computing the side length of the square mask as $s = \lfloor \rho \cdot \min(H, W) \rfloor$, where $\rho \in (0,1]$ is a ratio parameter controlling the proportion of the image to be masked. Next, a random position is selected by sampling an offset u uniformly from $[0, H - s]$ to define the row position and v uniformly from $[0, W - s]$ to define the column position. A binary mask $M \in \{0,1\}^{H \times W \times C}$ is created, where the values inside the square region $\times [v, v + s)$ are set to zero, and the values outside this region are set to one. Finally, the augmented image was obtained by multiplying the original image by the mask element-wise, given in Eq. (5).

$$x^* = x \cdot M. \tag{5}$$

This sets all pixels within the cutout region to zero (black), leaving the remainder unchanged.

If x is an image from the source domain X, then the pipeline after applying the augmentation $T(\cdot)$ is given by Eq. (6),

$$x^* = T(x), y = G_Y(x^*), \tag{6}$$

where y is the translation result, and x is the generator that translates the image from the actual image in domain X to the generated image in domain Y.

https://doi.org/10.31436/iiumej.v27i1.3824

2.3. Neural Network

One of the machine learning methods inspired by the structure and function of the human brain is a neural network (NN) [6]. To improve accuracy, the optimization process is applied during parameter estimation. The objective function of optimization is to minimize a loss function that quantifies the difference between predicted and actual values [7]. Stochastic Gradient Descent (SGD) is a widely used optimization method that updates parameters by following the negative gradient of the loss function. When the SGD learning rate is fixed, this method exhibits instability [8]. To overcome this problem, RMSprop and Adam have been developed. RMSprop is more stable because the method prevents oscillation by adjusting the learning rate based on a moving average of the squared gradient. Adam, an extension of RMSprop, offers faster, more efficient convergence by combining momentum-based optimization with adaptive learning rate updates based on the first and second moments of the gradient [9].

Convolutional Neural Networks (CNNs) are a type of neural network designed to analyze images or patterns. CNNs can extract spatial features from images automatically [10]. Convolutional, pooling, and fully connected layers are the core components of CNNs. In the convolutional layer, features are extracted using kernels to form feature maps that representmage patterns [11]. This process is illustrated in Figure 1.

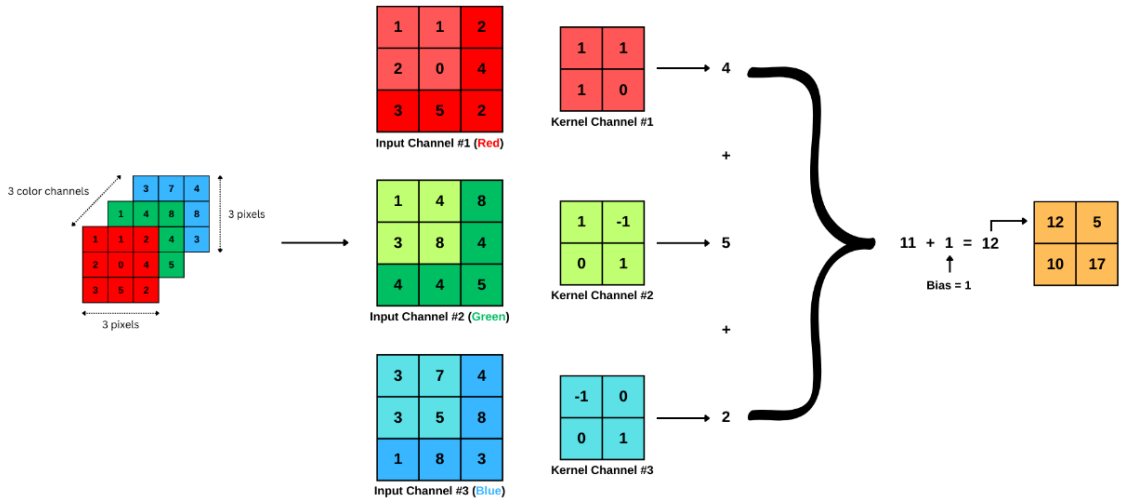


Figure 1. The convolutional operation

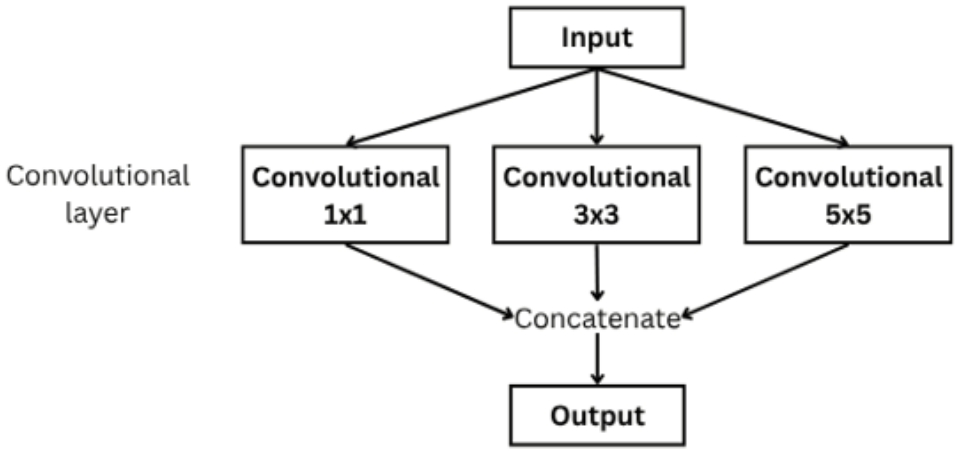


Figure 2. Multiscale kernel convolutional (MSCov) layer

https://doi.org/10.31436/iiumej.v27i1.3824

A single kernel size is often insufficient to capture all variations in a feature because objects in an image frequently appear at different sizes and positions. To address this issue, a multiscale kernel convolutional (MSCov) layer approach is used, combining multiple kernel sizes (e.g., 1×1, 3×3, 5×5) in parallel or sequentially within the convolutional layer [12], as illustrated in Figure 2. A pooling operation is applied to reduce the feature map dimensions and mitigate overfitting. The pooling method can use average or max pooling and is applied in the next layer [13].

The final process in a CNN is the fully connected layer. The layer consists of a flatten layer and a dense layer. Multidimensional feature maps are converted into one-dimensional vectors in the flatten layer. Furthermore, the dense layer learns the complex relationships between features to make accurate predictions [14].

2.4. Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GAN) are a method for pairwise image-to-image translation [15]. Pairwise training data must be available when using GAN. The GAN framework consists of two neural networks: a generator and a discriminator. The generator aims to produce realistic synthetic images that resemble the original, while the discriminator's role is to distinguish the original from the generated image. The generator continues to improve at producing realistic-looking images, while the discriminator simultaneously improves at detecting fake images. The results from both networks will improve the GAN's ability to produce high-quality, realistic image translations. Figure 3 illustrates the GAN framework.

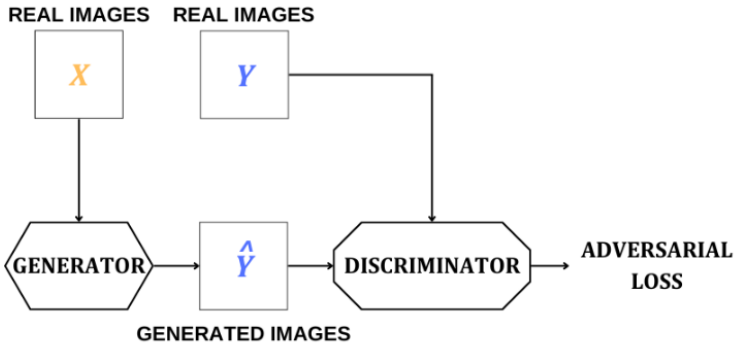


Figure 3. GAN framework

In this framework, X represents real images from the source domain, while Y denotes real images from the target domain. The generator produces synthetic images ($\hat{Y}$), which the discriminator evaluates to compute adversarial loss. Initially, a set of real images X is fed into the generator network, which translates them into synthetic images, denoted as $\hat{Y}$. These generated images are then combined with the real images Y and passed to the discriminator, which assesses their authenticity. Through this process, the adversarial loss function quantifies the discrepancy between real and generated images, providing a metric for evaluating translation accuracy.

The adversarial loss function is derived from the Bernoulli distribution probability density function. More formally, let Z be a random variable where $z = \{0,1\}$, with $z = 0$ representing a fake image and $z = 1$ representing a real image. Consequently, Z follows a Bernoulli distribution with a probability density function expressed as follows,

$$P(Z = z) = p^z(1-p)^{(1-z)}, \text{ for } z \in \{0,1\}. \quad (7)$$

https://doi.org/10.31436/iiumej.v27i1.3824

In Eq. (7), the value of p represents the probability that the discriminator correctly identifies an image as real. Thus, we redefine p as $\hat{z}$. Applying the log-likelihood to Eq. (7) leads to Eq. (8).

$$L(z,\hat{z}) = z \log \hat{z} + (1-z) \log (1-\hat{z}). \tag{8}$$

In image-to-image translation, the generator attempts to learn a mapping between the X and Y domains [15]. Hence, when a real image Y is input to the discriminator, written as $\hat{z} = D(Y)$, the value of z should be 1. However, whenever the generated fake image $G(X)$ is input to the discriminator, written as $\hat{z} = D(G(X))$, the value of z should be 0. Applying each condition to Eq. (8) and summing up the results will provide a new equation called $V(D,G)$, where D refers to the discriminator, and G refers to the generator. The $V(D,G)$ equation is expressed in Eq. (9).

$$V(D,G) = \log D(Y) + \log \Big(1 - D\big(G(X)\big)\Big). \tag{9}$$

Since Eq. (9) originates from the Bernoulli distribution, it applies to a single observation. However, when dealing with multiple images, we extend it using expectation. Furthermore, assuming that pixel distributions in domain X follow p_X with parameter θ_X and those in domain Y domain follow p_Y with parameter θ_Y, the adversarial loss function can be formulated as Eq. (10).

$$\mathcal{L}_{GAN}(G,D,X,Y) = \mathbb{E}_{Y\sim p_Y(\theta_Y)}[\log D(Y)] + \mathbb{E}_{X\sim p_X(\theta_X)}\Big[\log\Big(1 - D\big(G(X)\big)\Big)\Big]. \tag{10}$$

Other methods that utilize GAN have been proposed to generate more realistic images.

2.5. CycleGAN

A limitation of GANs is the requirement for paired data, which is often difficult to obtain. One unsupervised GAN method for two-domain translation without paired data is CycleGAN. Cycle consistency in CycleGAN ensures that images translated from one domain to another can be reconstructed to their original form. The optimization process in this method involves an objective function that shows the difference between the original input and the reconstructed version after translation. This function is the cycle-consistency loss.

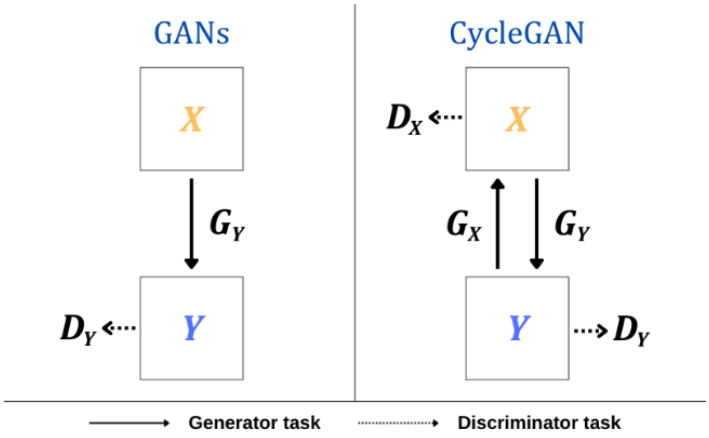


Figure 4. GAN and CycleGAN Framework Comparison

To facilitate this bidirectional translation, CycleGAN employs two generator-discriminator pairs. Suppose two generators are G_Y and G_X, then generator G_Y translates the image from the actual image of domain X to the generated image of domain Y, and generator G_X performs the inverse transformation from Y to X. Each generator is paired with a

https://doi.org/10.31436/iiumej.v27i1.3824

corresponding discriminator D_Y checks whether the output image of G_Y is real or fake according to domain Y, and vice versa for discriminator D_X. Both discriminators evaluate the realism of the generated image by comparing it with the real equivalent in each domain [1]. This bidirectional process aims to effectively learn the mapping between the two domains while preserving the critical image structure. Figure 4 illustrates the comparison between the conventional GAN and the CycleGAN. Unlike standard GAN, which performs one-way translation ($X \rightarrow Y$), CycleGAN introduces a bidirectional process ($X \leftrightarrow Y$) using two generator-discriminator pairs to enforce cycle consistency. This translation bidirectional process is to keep the generated image identical to the original image [1].

The collaborative operation of two generators and two discriminators of CycleGAN in maintaining cycle consistency is illustrated in Figure 5. The bidirectional translation process of G_Y and G_X ensures smooth connectivity across domains, thereby strengthening the principle of cycle consistency. This mechanism translates images between domains and reconstructs them to preserve structural integrity. By enforcing bidirectional consistency, CycleGAN ensures that translating an image to the target domain and then back yields a reconstruction that is similar to the original.

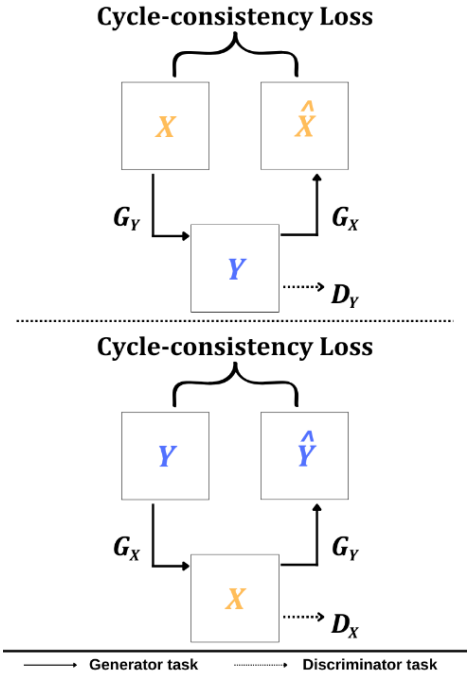


Figure 5. CycleGAN's cycle-consistency mechanism

The cycle consistency loss value, which is a quantification of the difference between the original and reconstructed images, is expressed in Eq. (11).

$$\mathcal{L}_{cyc}(G_Y, G_X) = \mathbb{E}_{X\sim p_X(\theta_X)}\left[\left\|G_X\left(G_Y(X)\right) - X\right\|\right] + \mathbb{E}_{Y\sim p_Y(\theta_Y)}\left[\left\|G_Y\left(G_X(Y)\right) - Y\right\|\right]. \quad (11)$$

In addition to the cycle-consistency loss function, another loss function must be considered: the identity-mapping loss. Identity mapping is used to ensure that the image maintains its original structure when processed by a generator from the same domain. The identity mapping is illustrated in Figure 6. In this process, when images are passed through their respective domain generators (X through G_X or Y through G_Y), they remain unchanged. This enforces an identity-mapping loss, ensuring that the generator in the same domain preserves the original structure.

https://doi.org/10.31436/iiumej.v27i1.3824

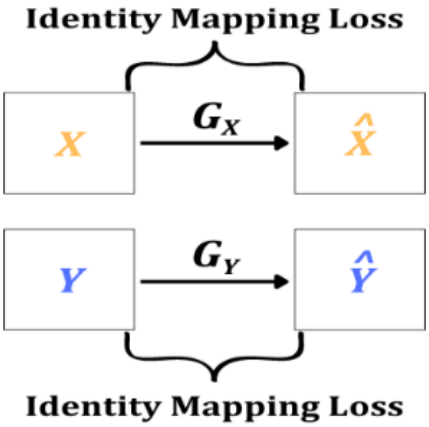


Figure 6. Identity mapping in CycleGAN

If an image from domain X is used as input to generator G_X, or when an image from domain Y is used as input to generator G_Y, then the output domain does not change. When the input and output domains are the same, the ideal transformation process should maintain the original characteristics of the input image. The identity mapping loss function is defined in Eq. (12)

$$\mathcal{L}_{identity}(G_Y, G_X) = \mathbb{E}_{X\sim p_X(\theta_X)}[\|G_X(X) - X\|] + \mathbb{E}_{Y\sim p_Y(\theta_Y)}[\|G_Y(Y) - Y\|]. \quad (12)$$

Building on the principles of CycleGAN, this study incorporates three distinct loss functions to facilitate unpaired image-to-image translation. Given that CycleGAN employs two generator-discriminator pairs, each pair contributes an adversarial loss function. As a result, the overall loss function formulated in this study is represented in Eq. (13).

$$\begin{aligned} \mathcal{L}(G_Y, D_Y, G_X, D_X) &= \mathcal{L}_{GANYX} + \lambda\mathcal{L}_{cyc}(G_Y, G_X) + \frac{\lambda}{2}\mathcal{L}_{identity}(G_Y, G_X), \\ \mathcal{L}_{GANYX} &= \mathcal{L}_{GAN}(G_Y, D_Y, X, Y) + \mathcal{L}_{GAN}(G_X, D_X, X, Y). \end{aligned} \quad (13)$$

The value of λ (lambda) in CycleGAN serves as a weighting factor or hyperparameter that regulates the influence of cycle-consistency loss and identity mapping loss on the overall objective function. A high λ value prioritizes maintaining cycle consistency and identity mapping between the source and target domains, producing more structurally consistent translations at the potential expense of realism. Conversely, a low λ value places greater emphasis on generating visually realistic translations by assigning greater weight to adversarial loss, potentially sacrificing strict consistency across domains.

2.6. Spectral Normalization

Spectral Normalization (SN) is a regularization technique that mitigates unstable predictions arising from data distributions that are not well represented in the training data [16]. This technique normalizes the weight matrix W by using the most significant singular value of W as a scaling factor [17]. The weights in the convolutional layers of the discriminator network are the values within the filters (kernels) used to extract features. The normalization for the weight matrix W is given by Eq. (14)

$$\bar{W} = \frac{W}{\sigma(W)}, \quad (14)$$

where $\sigma(W)$ is the most significant singular value of W. The singular value is obtained from the singular value decomposition (SVD) of W. SN is more efficient compared to the gradient penalty method, making it suitable for large-scale data [16].

https://doi.org/10.31436/iiumej.v27i1.3824

2.7. Experimental Setup

This study was conducted using two image domains, each comprising 450 samples: a blue-sky domain and a sunset domain. All images were preprocessed by resizing to 128×128 pixels and normalizing pixel values. Since CycleGAN is trained under unpaired learning conditions, images from the two domains were shuffled and randomly paired during separate training.

The training process consisted of 100 epochs, each comprising 450 steps, for a total of 45,000 optimization updates. A batch size of 1 was applied, consistent with standard practice in CycleGAN research, as this setting supports stable adversarial training. Two discriminators and two generators were trained with the Adam optimizer with a learning rate of 2×10^{-4} and a momentum of 0.5. Adversarial learning was employed in the least-squares GAN (LSGAN) setting with the mean-squared error loss, and cycle-consistency and identity-mapping losses were introduced to enforce structural and color consistency. The right weight factors were used as $\lambda = 10$.

To improve generalization with a relatively small dataset, DiffAugment was applied during training, combining random color perturbations, spatial translations, and cutout operations. All experiments were conducted on a Windows 11 (64-bit) system without a dedicated GPU; all model training and evaluation were performed on an Intel i3-1005G1 CPU (1.20 GHz, 4 threads) with 8 GB RAM.

3. RESULTS AND DISCUSSION

3.1. Domain Characteristics

The data used in this study were taken from the Places365 dataset http://places2.csail.mit.edu/download.html and the Summer2Winter-Yosemite dataset https://www.kaggle.com/datasets/balraj98/summer2winter-yosemite. The characteristics of color intensity across different domains will yield distinct data distributions. To determine the dominant color and the colors in each domain, a descriptive analysis was conducted. Two domains are studied: the sunset and blue-sky domains. Each image in these domains consists of three primary color channels: red (R), green (G), and blue (B). The distribution of the three colors is depicted in Figure 7. The histograms indicate that the red color channel predominates in sunset images, whereas the blue channel predominates in blue-sky images, reflecting the natural scene characteristics of each domain. At sunset, the warm atmosphere on the beach is influenced by sunlight. This atmosphere produces a reddish-orange sky, which results in the dominance of red. Conversely, when the beach image is reflected in a blue sky, it accentuates the dark blue of the sky and sea, resulting in the dominance of blue in this area.

To increase the diversity of the training dataset and improve the model's generalization, data augmentation was applied via several randomized transformations. The augmentation parameters were set as follows: brightness adjustment randomly sampled between -0.5 and 0.5, saturation variation between 0 and 2, contrast adjustment between 0 and 0.5, spatial shift (horizontal and vertical) randomly sampled from -0.125 to 0.125 of the image dimensions, and cutout augmentation applied randomly to up to 50% of the image area. These transformations were designed to simulate a wide range of visual variations likely encountered in real-world conditions.

https://doi.org/10.31436/iiumej.v27i1.3824

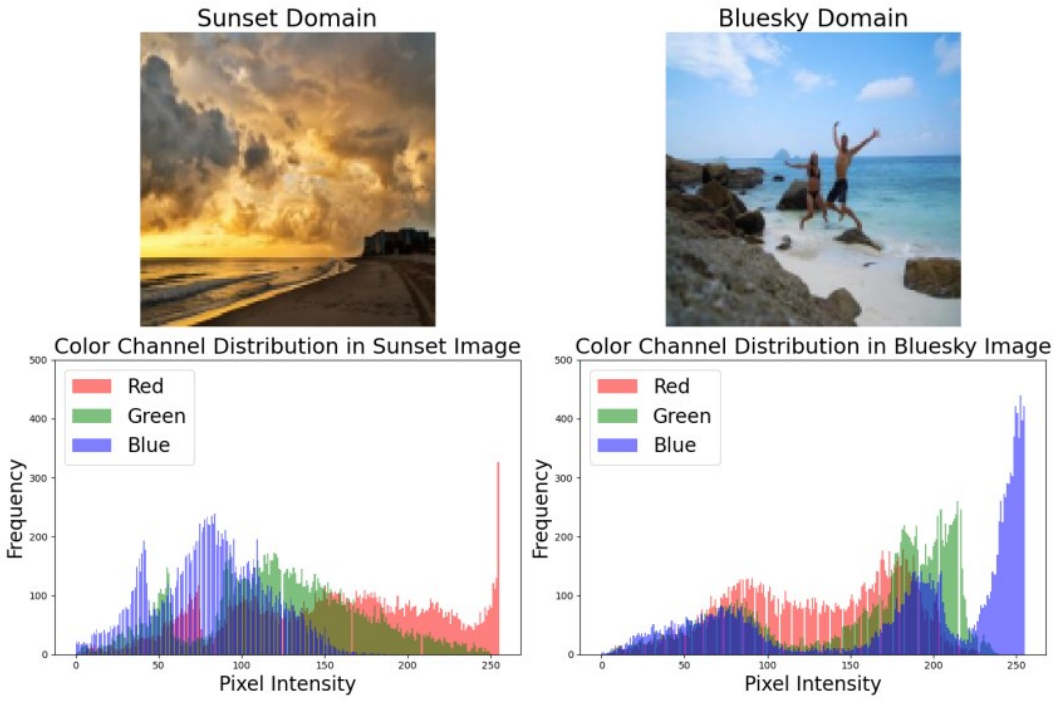


Figure 7. Color Distribution in Sunset and Blue-sky Domains

3.2. Model Architecture and Loss

In this study, the CycleGAN architecture comprises two generators and two discriminators for each translation task. The generator model is used to transfer sky ambiance between blue-sky and sunset conditions. The discriminator is used to classify whether images generated by the generator are real or fake. The architecture of the generator model used in this study is presented in Table 1, and the discriminator model is presented in Table 2. Generator networks include eight stride-2 convolutions with leaky ReLU activation functions and nine stride-1/2 convolutions with ReLU activation functions. To determine if the overlapping images were real, the discriminator networks used PatchGAN. Despite size variations, these architectural decisions align with those of prior studies [6].

In the CycleGAN architecture, different λ values were tried, namely 2, 5, and 10. The impact of changes in the λ value was observed, and the loss value for each λ was recorded. A comparison of the loss values for each model is given in Table 3. In the discriminator model, the model with a value of $\lambda = 10$ has the lowest loss value. This indicates that the discriminator in this model is most effective at distinguishing between real and fake images. Conversely, the loss value of the generator model with a value of $\lambda = 10$ has the highest loss value.

The λ hyperparameter is responsible for balancing the cycle consistency loss with adversarial loss and identity loss by controlling the weight of the cycle consistency loss. It is essential to recognize that lambda directly affects the generator's loss function; comparing generator performance solely by loss value would be misleading. Therefore, a more accurate assessment of the generator's quality should be based on the visual fidelity of the generated images. To improve the quality and stability of the generative model in CycleGAN with $\lambda = 10$, an MSCov layer was applied to the generator (Table 1). In contrast, SN was applied to the discriminator (Table 2). The MSCov layer uses 3 kernel sizes: 1x1, 3x3, and 5x5. To evaluate the model's performance, image translation experiments were conducted in two ways: (1) translating a blue-sky image to the sunset domain; and (2) translating a sunset image to the blue-sky domain. The quality of this translation was measured using MSE, SSIM, and LPIPS

https://doi.org/10.31436/iiumej.v27i1.3824

between the original image and the generated image after double translation (e.g., translating a blue-sky image to a sunset scene and then back to a blue-sky image).

Table 1. The architecture of the generator

Generator without MSCov Layer		Generator with MSCov Layer	
Layer Type	**Output Size**	**Layer Type**	**Output Size**
InputLayer	(None, 128, 128, 3)	InputLayer	(None, 128, 128, 3)
Sequential	(None, 64, 64, 64)	Sequential	(None, 64, 64, 64)
Sequential	(None, 32, 32, 128)	Sequential	(None, 32, 32, 128)
Sequential	(None, 16, 16, 128)	Sequential	(None, 16, 16, 128)
Sequential	(None, 8, 8, 128)	Conv2D	(None, 16, 16, 256)
Sequential	(None, 4, 4, 128)	Conv2D	(None, 16, 16, 256)
Sequential	(None, 2, 2, 128)	Concatenate	(None, 16, 16, 512)
Sequential	(None, 4, 4, 128)	Conv2D	(None, 16, 16, 256)
Concatenate	(None, 4, 4, 256)	Instance Normalization	(None, 16, 16, 256)
Sequential	(None, 8, 8, 128)	LeakyReLU	(None, 16, 16, 256)
Concatenate	(None, 8, 8, 256)	Sequential	(None, 8, 8, 128)
Sequential	(None, 16, 16, 128)	Conv2D	(None, 8, 8, 256)
Concatenate	(None, 16, 16, 256)	Conv2D	(None, 8, 8, 256)
Sequential	(None, 32, 32, 128)	Concatenate	(None, 8, 8, 512)
Concatenate	(None, 32, 32, 256)	Conv2D	(None, 8, 8, 256)
Sequential	(None, 64, 64, 64)	Instance Normalization	(None, 8, 8, 256)
Concatenate	(None, 64, 64, 128)	LeakyReLU	(None, 8, 8, 256)
Sequential	(None, 128, 128, 64)	Sequential	(None, 4, 4, 128)
Conv2D Transpose	(None, 128, 128, 3)	Sequential	(None, 2, 2, 128)
		Sequential	(None, 4, 4, 128)
		Concatenate	(None, 4, 4, 256)
		Sequential	(None, 8, 8, 128)
		Concatenate	(None, 8, 8, 384)
		Sequential	(None, 16, 16, 128)
		Concatenate	(None, 16, 16, 384)
		Sequential	(None, 32, 32, 128)
		Concatenate	(None, 32, 32, 256)
		Sequential	(None, 64, 64, 64)
		Concatenate	(None, 64, 64, 128)
		Sequential	(None, 128, 128, 64)
		Conv2D Transpose	(None, 128, 128, 3)

Table 2. The architecture of the discriminator

Discriminator without SN		Discriminator with SN	
Layer Type	**Output Size**	**Layer Type**	**Output Size**
InputLayer	(None, 128, 128, 3)	InputLayer	(None, 128, 128, 3)
Sequential	(None, 64, 64, 64)	Spectral Normalization	(None, 64, 64, 64)
Sequential	(None, 32, 32, 128)	Leaky ReLU	(None, 64, 64, 64)
Sequential	(None, 16, 16, 256)	Spectral Normalization	(None, 32, 32, 128)
Zero Padding 2D	(None, 18, 18, 256)	Layer Normalization	(None, 32, 32, 128)
Conv2D	(None, 15, 15, 512)	Leaky ReLU	(None, 32, 32, 128)
Instance Normalization	(None, 15, 15, 512)	Spectral Normalization	(None, 16, 16, 256)
Leaky ReLU	(None, 15, 15, 512)	Layer Normalization	(None, 16, 16, 256)
Zero Padding 2D	(None, 17, 17, 512)	Leaky ReLU	(None, 16, 16, 256)
Conv2D	(None, 14, 14, 1)	Spectral Normalization	(None, 8, 8, 256)
—	—	Layer Normalization	(None, 8, 8, 256)
—	—	Leaky ReLU	(None, 8, 8, 256)
—	—	Spectral Normalization	(None, 4, 4, 256)
—	—	Layer Normalization	(None, 4, 4, 256)
—	—	Leaky ReLU	(None, 4, 4, 256)
—	—	Spectral Normalization	(None, 4, 4, 1)

https://doi.org/10.31436/iiumej.v27i1.3824

The models are evaluated using the metrics MSE, SSIM, and LPIPS to measure the reconstruction accuracy and perceptual quality. Across all three measures, the proposed CycleGAN with regularization and MSCov improves network performance, as shown in Table 4. These findings confirm that applying regularization and MSCov significantly improves CycleGAN's translation quality.

Table 3. The loss values for each model

Losses	λ		
	2	**5**	**10**
Sunset Generator	0.9720	1.3730	1.7399
Blue-sky Generator	1.0018	1.3828	1.7473
Sunset Discriminator	0.6846	0.6557	0.6239
Blue-sky Discriminator	0.6756	0.6605	0.6297

Table 4. Model performance metrics

Metrics	Basic CycleGAN	CycleGAN with SN + Augmentation + MSCov
Blue sky - Sunset – Blue sky		
MSE*	0.0050	0.0008
SSIM**	0.7258	0.9592
LPIPS*	0.2180	0.0497
Sunset – Blue sky - Sunset		
MSE*	0.0044	0.0013
SSIM**	0.7750	0.9660
LPIPS*	0.2073	0.0831

*The lower, the better
**The higher, the better

3.3. Sunset Generator Performance

The CycleGAN model, with various λ values, translates a blue sky to a sunset. Overall, all generators perform well. The translation result of the sunset image produces a perfect atmosphere. The generator produces realistic sunset images. The blue sky is transformed into a convincing sunset atmosphere. The result is free from overlapping objects and errors. The color of the water and other objects in the generated image is successfully translated, naturally harmonizing with the sunset atmosphere. Figure 8 compares CycleGAN-generated sunset transformations across different λ values (2, 5, and 10), illustrating how the model transforms blue-sky beach scenes into sunset settings.

Furthermore, the cyclical consistency property of each object is maintained. The locations of the items and the shapes of the objects remain unchanged. The translation result is given in Figure 8. The figure shows a successful translation of blue-sky scenes into realistic sunsets while maintaining object integrity and spatial consistency. The water and surrounding elements change naturally with the sunset atmosphere. No artifacts or distortions are visible in the translation result. As the lambda value increases, the image quality improves.

Model 1 ($\lambda = 2$) produces a recognizable transformation but suffers from lighting consistency. The colors are also less well adapted. Model 2 ($\lambda = 5$) achieves better harmony between the sky and landscape. The translation results in this model are smoother than those in model 1. Model 3 ($\lambda = 10$) produces the most realistic results, with more lifelike colors, more natural lighting, and better reflections in the water and sky. These results indicate that

https://doi.org/10.31436/iiumej.v27i1.3824

higher lambda values are associated with greater cycle consistency, better structural preservation, and more accurate color adaptation.

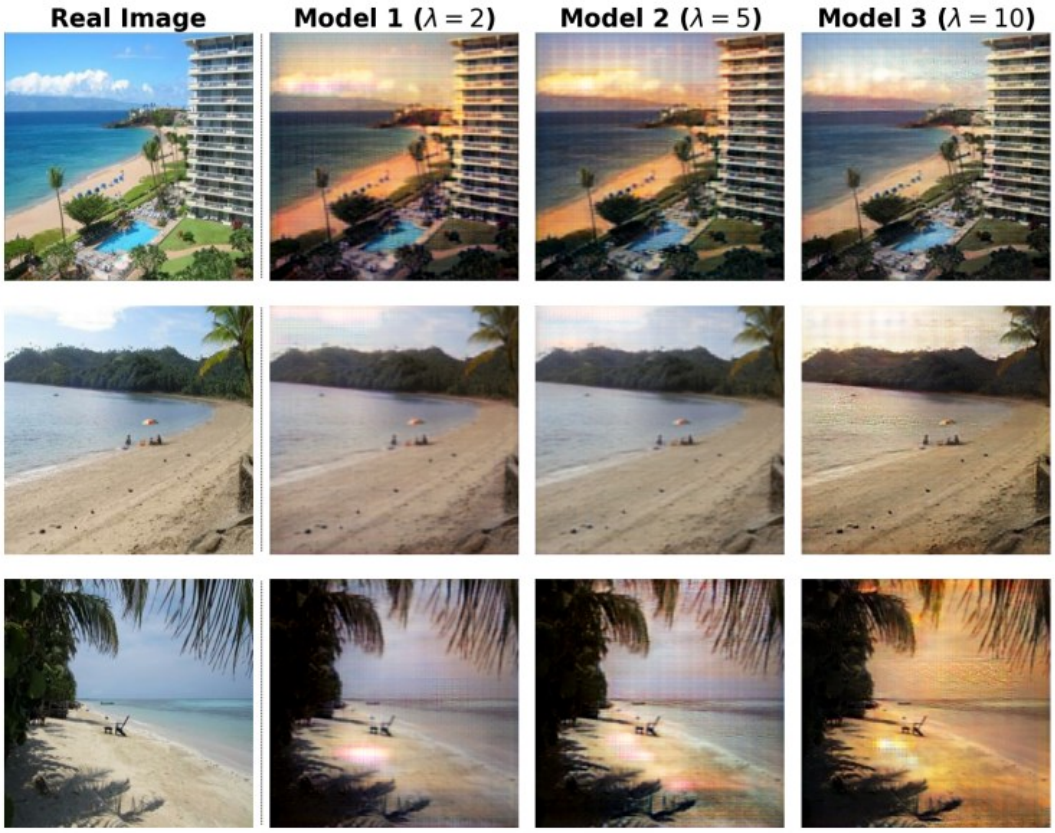


Figure 8. Results of sunset generator with different λ values

In the same domain as the previous case, an image with a human subject is presented in Figure 9 to evaluate the generator's effectiveness in translating images of human subjects, showing how different λ-values influence color adaptation and structural consistency. The sunset generator effectively integrates the human figure into the transformed scene without causing visual inconsistencies. Smooth color transitions, preservation of object shapes, and maintenance of spatial relationships indicate strong cycle consistency and realistic transformation.

The effects of increasing the lambda value are clearly evident, particularly in human-subject studies and complex scenes. Model 1 ($\lambda = 2$) implements the sunset effect well, but some color blends, especially around the human figure, appear less smooth, sometimes disturbing the natural reflections and shadows. Model 2 ($\lambda = 5$) improves the overall consistency, especially in the reflections of the sky and water. The sunset transformation results are more cohesive. Model 3 ($\lambda = 10$) produces the most realistic results. These results demonstrate better integration of human subjects, supported by improved lighting effects and more natural color adaptation. Model 3 ($\lambda = 10$) produces the most visually appealing and realistic results. The sunset scene is well integrated. Colors, textures, and structural details are preserved. This model shows excellent cycle consistency. Human contours and object placement are preserved. The resulting color transitions are also in harmony with the sunset scene. In addition, identity mapping helps maintain the original characteristics of objects within the same domain, prevents unnecessary changes, and enhances the realism of the transformation.

https://doi.org/10.31436/iiumej.v27i1.3824

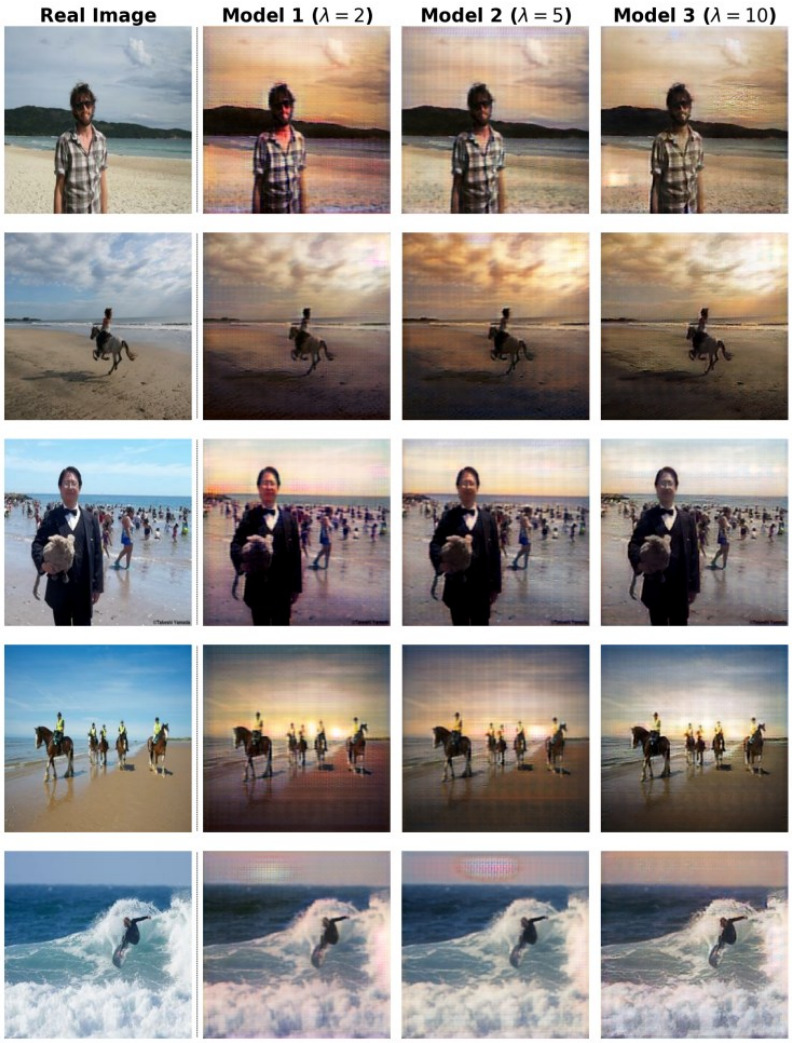


Figure 9. Sunset Generator Performance on Human Subjects

Figure 10. Comparison of CycleGAN and modified CycleGAN on the Sunset Domain

https://doi.org/10.31436/iiumej.v27i1.3824

The experiment to improve translation quality was continued by applying regularization and MSCov to the CycleGAN model with $\lambda = 10$. The comparison of translation results is given in Figure 10. The figure presents a side-by-side comparison between the baseline CycleGAN and the modified CycleGAN, highlighting the improvements in realism introduced by regularization and MSCov. The integration between regularization and MSCov improves the image quality and is more realistic with MSE, SSIM, and LPIPS values as shown in Table 4.

3.4. Blue-Sky Generator Performance

The performance of CycleGAN in translating sunset images to blue sky under various λ values is illustrated in Figure 11, where sunset images are transformed into daytime conditions with restored lighting and atmospheric adjustments. In general, the generated images are consistent. The generated results demonstrate strong realism and cycle consistency. Similar to the sunset generator, the blue-sky generator effectively transforms the scene while preserving critical structural details. In the CycleGAN model, the generator and discriminator work in a conflicting framework. The training process aims to balance realism and structural preservation. If the generator outperforms the discriminator, an imbalance may arise. The result is an overly convincing image. Conversely, if the discriminator becomes very dominant, the generated image will be unrealistic.

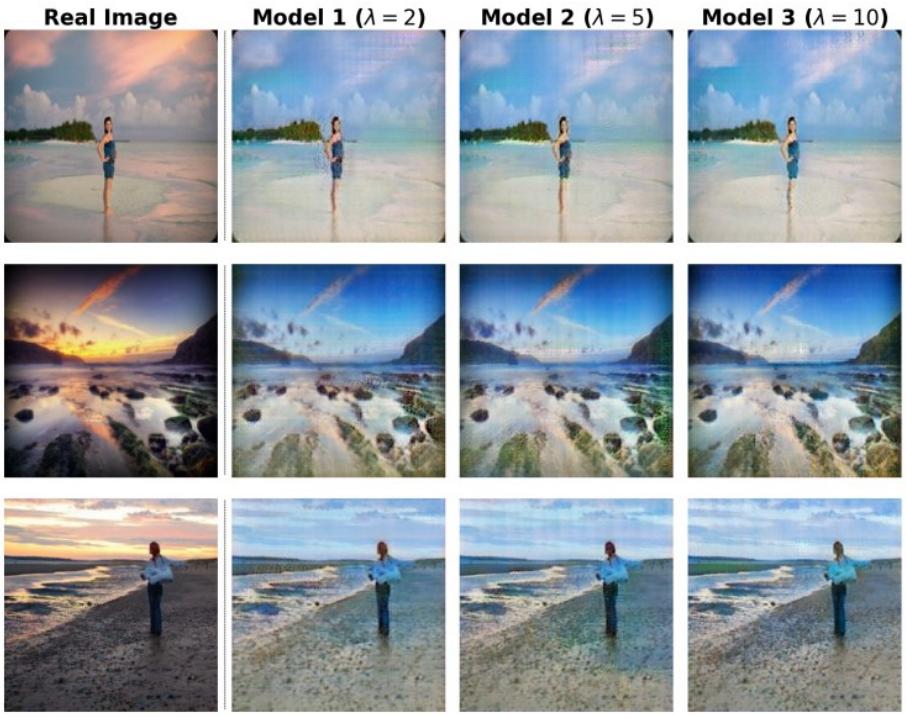


Figure 11. Blue-Sky Generator Results

Model 3 ($\lambda = 10$) has a slightly higher loss value than the other models. This is due to the increase in the lambda parameter. This small increase in loss is tolerable because the model's cycle consistency increases. This indicates a higher-quality, more visually appealing transformation. Therefore, Model 3 is the best.

Building on Model 3, augmentation is applied to the images, and CycleGAN is modified by adding SN and MSCov. A comparison of translation results before and after regularization and MSCov is shown in Figure 12. The figure compares the baseline CycleGAN with the modified CycleGAN on the blue-sky generator, highlighting improvements in realism such as

https://doi.org/10.31436/iiumej.v27i1.3824

enhanced atmospheric clarity and the appearance of a sun element in the generated images. MSE, SSIM, and LPIPS values in Modified CycleGAN are also improved, as presented in Table 4.

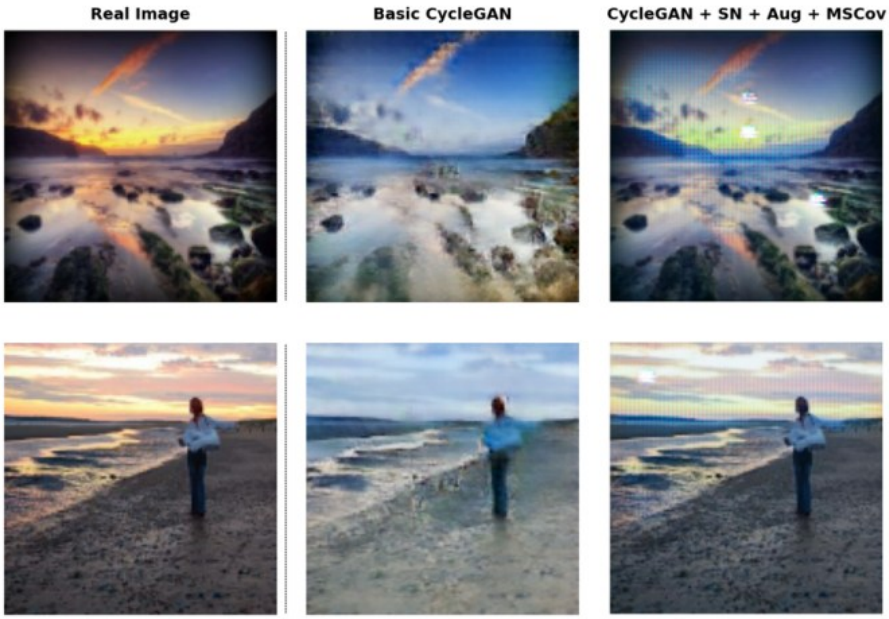


Figure 12. Comparison of CycleGAN and Modified CycleGAN on Blue-Sky Domain

3.5. Limitations

Although each generator appears effective at generating images, failures occur during the image-generation process. The failure cases of the sunset and blue-sky generators are illustrated in Figures 13 and 14, respectively. In Figure 13, the primary issue arises from the inaccurate translation of the shirt's white color. The generator has difficulty distinguishing the shirt's white from the clouds' white. The translation of cloud color affects the shirt's color. This results in the white color not being preserved, as it instead turns orange, reflecting the sunset sky. These cases illustrate how the sunset generator sometimes misinterprets white clothing as cloud formations, leading to undesirable color transformations. As a result, the transformation sacrifices realism by changing elements that should remain unchanged.

The failure of the blue-sky generator is of a different type, as shown in Figure 14. Dark objects appear in the sunset domain due to reduced lighting. The challenge in translating to blue-sky is to render dark objects as bright. The generator must adapt these objects to suit the new atmosphere. However, the generator often leaves them unchanged, resulting in inconsistencies. These failures reveal the blue-sky generator's difficulty in adapting dark objects to brighter conditions, which reduces the realism of the transformation. This demonstrates the limitations of the generator's ability to accurately interpret and adjust the texture and color of dark objects during the transformation process. Image translation failures mainly occur when color adaptation conflicts with learned patterns. Identity mapping is crucial because it helps the generator maintain structural integrity during image translation.

https://doi.org/10.31436/iiumej.v27i1.3824

Figure 13. Failure Cases in Sunset Generation

Figure 14. Failure Cases in Blue-sky Generation

4. CONCLUSIONS

This study investigates the transformation of blue-sky beach images into realistic sunset scenes using CycleGAN, an unpaired image-to-image translation framework. By leveraging adversarial learning and cycle-consistency constraints, the research demonstrates that CycleGAN effectively adapts images between distinct visual domains without requiring paired datasets. The experimental results, based on varying λ (2, 5, and 10), reveal that while higher λ values lead to increased generator loss, they also enhance image realism and structural preservation. This finding underscores that λ influences the generator's optimization process

but does not directly correlate with perceptual image quality, underscoring the importance of evaluating results using perceptual quality measures rather than loss metrics alone. This study showcases CycleGAN's practical potential for enhancing automated image editing and tourist imagery. The model with $\lambda = 10$ produced the most visually convincing results, reaffirming that careful hyperparameter tuning is critical for achieving high-quality transformations. In addition, the results of augmentation, SN, and MSCov indicate that the modified CycleGAN can generate more realistic images.

For the transformation of a blue-sky beach image into a sunset scene, the modified CycleGAN reduces MSE and LPIPS by 84% and 77%, respectively, while increasing SSIM by 32%. In contrast, for the sunset-to-blue-sky transformation, the proposed method reduces MSE and LPIPS by 70% and 60%, respectively, and increases SSIM by 25%.

A color channel histogram analysis confirmed that the dominant hues in the sunset and blue-sky domains, red and blue, respectively, align with real-world conditions, demonstrating the model's capacity to learn meaningful transformations. Additionally, the models effectively maintained cycle consistency, ensuring that key objects remained identifiable as they adapted to the target domain. However, limitations were observed: sunset generators occasionally misinterpreted white clothing as cloud formations, whereas blue-sky generators struggled to render the textures and colors of dark objects accurately. These challenges suggest that, despite robust identity-mapping constraints, certain object-specific transformations remain problematic.

This research contributes to the field of unpaired image-to-image translation by providing insights into how CycleGAN can be fine-tuned for complex environmental adaptations. Future work should focus on refining object-level transformations, particularly by improving the model's ability to distinguish between visually ambiguous elements. Investigating enhanced architectural modifications or hybrid approaches that integrate CycleGAN with other deep learning techniques could further improve realism and consistency. Additionally, extending this methodology to more diverse photographic scenarios would validate its broader applicability. Baseline comparisons with advanced models such as UNIT, MUNIT, or CUT, as well as exploring higher-resolution outputs beyond 128×128, could further improve applicability. By addressing these areas, future studies can advance image-to-image translation techniques and expand their impact across various real-world applications.

ACKNOWLEDGEMENT

This research was financially supported by Institut Teknologi Sepuluh Nopember through the Scientific Research Grant Scheme 2024 under Grant Number 1177/PKS/ITS/2024. The authors gratefully acknowledge this support.

REFERENCES

[1] Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, 2223(2232). https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html

[2] Tong, Z. (2024). Exploring the Impact of Hyperparameters on the Generation Quality of CycleGAN. Transactions on Computer Science and Intelligent Systems Research, 5, 265–271. https://doi.org/10.62051/01M93A63

[3] Hu, Y. (2024). Impact of Hyperparameters on the Quality of Image Translation Using CycleGAN. Transactions on Computer Science and Intelligent Systems Research, 5, 487–492. https://doi.org/10.62051/M04WSD55

[4] Zhao, S., Liu, Z., Lin, J., Zhu Adobe, J.-Y., & Song Han, C. (2020). Differentiable Augmentation for Data-Efficient GAN Training. Advances in Neural Information Processing Systems, 33, 7559–7570.https://proceedings.neurips.cc/paper_files/paper/2020/file/55479c55ebd1efd3ff125f1337100388-Paper.pdf

[5] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 1–48. https://doi.org/10.1186/S40537-019-0197-0

[6] Lu, Z., Pu, H., Wang, F., Hu, Z., & Wang, L. (2017). The Expressive Power of Neural Networks: A View from the Width. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper/2017/hash/32cbf687880eb1674a07bf717761dd3a-Abstract.html

[7] Qin, H., Gong, R., Liu, X., Shen, M., Wei, Z., Yu, F., & Song, J. (2020). Forward and backward information retention for accurate binary neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2250–2259. http://openaccess.thecvf.com/content_CVPR_2020/html/Qin_Forward_and_Backward_Information_Retention_for_Accurate_Binary_Neural_Networks_CVPR_2020_paper.html

[8] Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization Methods for Large-Scale Machine Learning. SIAM Review, 60(2), 223–311. https://doi.org/10.1137/16M1080173

[9] Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. ArXiv Preprint ArXiv, 1412.6980. https://arxiv.org/abs/1412.6980

[10] Aggarwal, C. C. (2018). Neural networks and deep learning: a textbook. Springer. https://dlib.hust.edu.vn/handle/HUST/24439

[11] Javier, F., Morales, O., & Roggen, D. (2016). Deep Convolutional Feature Transfer Across Mobile Activity Recognition Domains, Sensor Modalities and Locations. Proceedings of the 2016 ACM International Symposium on Wearable Computers, 92 – 99. https://doi.org/10.1145/2971763.2971764

[12] Brahimi, S., Ben Aoun, N., Ben Amar, C., Benoit, A., & Lambert, P. (2018). Multiscale Fully Convolutional DenseNet for Semantic Segmentation. WSCG J, 26(2), 104–111. https://doi.org/10.24132/JWSCG.2018.26.2.5

[13] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv Preprint ArXiv, 1409.1556. https://arxiv.org/abs/1409.1556

[14] Bello, A., Ng, S. C., & Leung, M. F. (2024). Skin cancer classification using fine-tuned transfer learning of DENSENET-121. Applied Sciences, 14(17), 7707. https://doi.org/10.3390/APP14177707

[15] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Advances in Neural Information Processing Systems (Vol. 27). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf

[16] Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral Normalization for Generative Adversarial Networks. 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings. https://arxiv.org/abs/1802.05957

[17] Yoshida, Y., & Miyato, T. (2017). Spectral Norm Regularization for Improving the Generalizability of Deep Learning. ArXiv Preprint ArXiv, 1705.10941. https://arxiv.org/abs/1705.10941

https://doi.org/10.31436/iiumej.v27i1.3830

Advanced Groundwater Level Forecasting using QSO-based Vision Transformer Model for Sustainable Water Resource Management

V. GOKULA KRISHNAN[1*], M. S. MAHARAJAN[2], B.V. SUBBA RAO[3], G.MAHALAKSHMI[4], M. ARUNADEVI[5]

[1*] *Department of CSE, Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu, India*
[2] *Department of AIDS, Panimalar Engineering College, Poonamallee, Chennai, Tamil Nadu, India*
[3] *Department of Information Technology, PVP Siddhartha Institute of Technology, Kanuru, Vijayawada, India*
[4] *Department of AIDS, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Chennai, Tamil Nadu, India*
[5] *Department of CSE, K. Ramakrishnan College of Technology, Trichy, Tamil Nadu, India*

**Corresponding author: gokul_kris143@yahoo.com*

(Received: 7 June 2025; Accepted: 4 September 2025; Published online: 12 January 2026)

ABSTRACT: The reliable predictions of groundwater levels are crucial for long-term management of water resources, and they are an excellent source for human well-being. While conventional ML approaches work well for low-dimensional data, they are optimized for hyperparameters on high-dimensional inputs and also capture complex temporal correlations. To address these restrictions, this research presents a new framework for predicting groundwater-level changes up to five months in advance, built on the Vision Transformer (ViT) and optimised with Quokka Swarm Optimisation (QSO). ViT enables strong global feature extraction and long-range dependency modelling by processing time-series data as sequential image-like patches, in contrast to traditional neural networks. Drawing on quokka's adaptive survival behaviour, the QSO procedure optimises the transformer's hyperparameters, such as patch size, attention heads, and depth, in real time to enhance prediction accuracy. The ViT+QSO outperformed baseline deep learning methods on groundwater datasets from Southern Africa in terms of RMSE, MAE, correlation coefficient (R), and Nash-Sutcliffe Efficiency (NSE). Quantile regression uncertainty quantification further improves the model's reliability for water resource planning. Hydrological variables, in addition to climate indices, influence groundwater fluctuations, as confirmed by ablation research. The proposed ViT-QSO achieved 93% R and 0.118 MAE, whereas the basic ViT achieved only 89.1% R and 0.152 MAE for groundwater-level prediction. Scalability, interpretability, and suitability for areas with limited monitoring infrastructure are hallmarks of the proposed methodology. This research provides valuable insights into how to better withstand the effects of climate change, in addition to human activities, on groundwater supplies.

ABSTRAK: Ramalan paras air bawah tanah yang boleh dipercayai adalah penting bagi pengurusan jangka panjang sumber air dan kesejahteraan manusia; namun, pendekatan pembelajaran mesin konvensional berhadapan kekangan pengendalian input berdimensi tinggi serta model korelasi temporal kompleks. Kajian ini mencadangkan satu rangka kerja baharu berasaskan Pengubah Visi (ViT) yang dioptimum menggunakan Optimisasi Kawanan Quokka (QSO) dalam meramal perubahan paras air bawah tanah pada lima bulan lebih awal. ViT memproses data siri masa sebagai tampalan jujukan menyerupai imej bagi membolehkan pengekstrakan ciri global dan model kebergantungan jarak jauh, manakala QSO mengoptimumkan hiperparameter pengubah secara adaptif dalam meningkatkan ketepatan ramalan. Model ViT-QSO menunjukkan prestasi unggul berbanding kaedah pembelajaran mendalam asas pada set data air bawah tanah di Afrika Selatan, dengan peningkatan ketara

dari segi RMSE, MAE, pekali korelasi (R), dan Kecekapan Nash–Sutcliffe (NSE), serta mencapai nilai R sebanyak 93% dan MAE 0.118 berbanding ViT asas masing-masing mencatatkan 89.1% dan 0.152. Pengkuantitian ketidakpastian melalui regresi kuantil meningkatkan kebolehpercayaan model bagi perancangan sumber air, manakala kajian ablasi mengesahkan peranan pembolehubah hidrologi dan indeks iklim terhadap turun naik paras air bawah tanah. Secara keseluruhan, metodologi yang dicadangkan adalah berskala, boleh ditafsir, dan sesuai pada kawasan infrastruktur pemantauan terhad, serta memberikan sumbangan penting dalam menangani kesan perubahan iklim dan aktiviti manusia terhadap sumber air bawah tanah.

KEYWORDS: *Quokka Swarm Optimization; Vision Transformer; Southern African groundwater datasets; Deep Belief Networks; Water Resources.*

1. INTRODUCTION

A growing frequency of extreme weather events driven by climate variability has highlighted the critical need for a reliable water supply and the organised management of water resources. A vital freshwater resource that can help resolve these issues is groundwater. To plan, develop, and implement effective groundwater resource management systems, it is necessary to continuously and systematically monitor groundwater change [1]. To ensure the sustainable use of groundwater, various government agencies in South Korea operate observation networks nationwide. If a farmer wants to know how much groundwater is available and how to manage it effectively, they need precise projections of spatial and temporal changes in groundwater storage. The hydrogeological characteristics and boundary conditions of groundwater systems, along with the impacts of weather, water, and land-cover changes, can be used to infer groundwater fluctuations. Due to complex spatial variability in groundwater and pronounced nonlinearity, systematic groundwater predictions are difficult [2].

Using a combination of data from multiple satellites, including GRACE (Gravity Recovery and Climate Experiment) and GRACE-Follow-On, TRMM (Tropical Rainfall Measuring Mission), and Landsat (based on the water balance method), remote sensing methods have been suggested as a means to estimate changes in spatiotemporal groundwater storage in recent years. Due to the poor spatial/temporal resolution of satellite data, these methods are unsuitable for predicting or nowcasting changes in groundwater storage, despite their relatively straightforward and reliable evaluation across many locations [3]. Moreover, vegetation or human-caused impacts are frequently overlooked in GRACE-based water-balance analyses, which consider only hydrological factors (terrestrial water storage, snow water equivalent, surface water, and soil moisture) [4]. Recent research has shown that changes in land cover, NDVI, and MNDWI are among the variables that determine groundwater recharge and potential. This study recommends employing multiple satellite datasets while accounting for factors that affect the reliability of the results.

Since it is difficult to predict how changes across Earth's system will affect groundwater storage, it is essential to examine how groundwater responds to these changes by assessing correlations among climatic variables, surface water extent, hydrological components, and vegetation [5]. Complex issues in computer vision, image classification, and time-series forecasting have motivated the development of machine learning techniques that account for nonlinear interdependencies, including ANN, SVM, and RF. Numerous ML approaches, including ANN models, hybrid feedforward neural networks (FNNs), and nonlinear autoregressive models with exogenous inputs (NARXs), have been utilised to develop groundwater prediction models [6]. Most research has focused on data collected from ground

observation points, making it challenging to forecast hydrological components and anticipate fluctuations in groundwater storage.

Predicting spatiotemporal patterns with any degree of accuracy is, in particular, constrained by the precision of ground observations used as input data, which are located and time-stamped [7]. Using an ANN approach and multi-satellite data, one study that aimed to circumvent these restrictions monitored spatiotemporal fluctuations in groundwater storage and droughts in the Korean Peninsula. Although this approach demonstrated satisfactory performance in tracking variations in groundwater storage (correlation coefficients of 0.9 for the training period and 0.6 for the test period, respectively), there are still concerns that need resolving, such as overfitting and input parameter selection in certain areas (such as coastal regions and data-scarce regions) [8].

Several deep learning models have recently been developed and have demonstrated promising performance in addressing long-term dependencies in complex, nonlinear prediction tasks. These models include a convolutional neural network (CNN) and a long short-term memory (LSTM), both of which are recurrent neural network (RNN) architectures [9]. The intricate, deep network architectures enabled by these models' many hidden layers allow precise identification of the relationship between the predictor and target variables. When applied to machine learning models, deep learning models mitigate overfitting and gradient vanishing (exploding) concerns [10]. In addition, a rectified linear unit (ReLU) activation function can be employed to alleviate vanishing gradient issues by means of a weight update procedure characterised by a high dropout rate. There usually are five steps involved in building a deep learning model for prediction: First, preparing the data to be used as inputs or outputs; second, creating the model with the right amount of layers to define the inputs or outputs; third, training the model with loss functions; fourth, optimising the network (hyperparameter optimisation); and fifth, validating the model.

Although LSTM and CNN have found widespread use in many domains, including water resources such as surface water, flood forecasting, and terrestrial water storage, their use in groundwater prediction is relatively new, having begun only in the past two years [11]. The reference model demonstrated that CNN outperformed SVM in terms of groundwater potential mapping. At the point scale of groundwater measurements, NARX, LSTM, and CNN were evaluated for predicting groundwater levels. The model was evaluated at the point scale, although deep learning improved groundwater prediction skills. Furthermore, no data from in situ groundwater sources or meteorological observations (e.g., temperature or precipitation) were included in the input dataset [12]. Thus, it is essential to enhance the spatiotemporal prediction model of groundwater storage change by incorporating multiple parameters, including hydrological, vegetation, and meteorological variables [3].

This project aims to address the complex problem of multi-month groundwater-level prediction by developing and validating a new ViT-based forecasting model optimised using QSO. The proposed technique addresses key gaps in existing deep learning and machine learning methods by presenting a transformer-based framework that captures complex nonlinear correlations and intricate temporal dependencies in weather and hydrological data.

Key contributions are as follows:

1. Transformer-Based Architecture for Time-Series: A new-fangled contribution is the use of ViT for groundwater time-series modelling. By using self-attention to capture long-range relationships, the model outperforms typical RNN-based models, which treat hydrological variables as sequential patches.

2. Nature-Inspired Optimization (QSO): As a lightweight and effective solution for hyperparameter tweaking, the Quokka Swarm Optimisation procedure is introduced. By leveraging ViT characteristics such as patch size, depth, and attention heads, QSO significantly improves performance across all prediction horizons.
3. Uncertainty Quantification: A robust uncertainty modelling technique using quantile regression has been integrated into the forecasting framework. This ensures not only high accuracy but also confidence-interval-based insights into groundwater-level changes.
4. Ablation Study & Feature Relevance Analysis: Comprehensive ablation experiments quantify the impact of input variables, such as Total Water Storage, precipitation, temperature, and climate indices, confirming their role in model performance.
5. Comparative Evaluation & Generalization: The suggested model outperformed state-of-the-art approaches in benchmark tests measuring root-mean-square error (RMSE), mean absolute error (MAE), net structural error (NSE), and correlation. Additionally, the model performs effectively across a variety of geographic settings and groundwater well types.

Unlike conventional artificial intelligence approaches that often suffer from limited generalization and a lack of uncertainty quantification, the present study proposes a Vision Transformer optimized with Quokka Swarm Optimization (QSO) and enhanced with Quantile Regression. This combination represents a methodological novelty in groundwater research, as it leverages transformer-based deep learning to capture long-term dependencies, applies QSO for efficient feature optimization, and integrates probabilistic forecasting through quantile regression. The practical implication of this framework is its ability to provide uncertainty-informed, scalable, and transferable groundwater forecasts, which are particularly valuable for policymakers and water managers in regions exposed to climate variability or lacking dense in-situ monitoring infrastructure. Collectively, these contributions represent a significant advancement in AI-driven groundwater modelling, offering a scalable, adaptive, and interpretable solution for global water sustainability challenges. The rest of the paper is organized as follows: Section 2 reviews related work; Section 3 presents the data used in the study; Section 4 details the proposed methodology; Section 5 presents the results; and Section 6 concludes.

2. RELATED WORKS

For modelling changes in groundwater levels in South Africa's West Coast Aquifer System, Igwebuike et al. [13] assessed two ML algorithms, Random Forest (RF) and Support Vector Machine (SVM), as well as two DL algorithms, Simple Recurrent Neural Network (SimpleRNN) and Long Short-Term Memory (LSTM). Regression error metrics analysis on the test dataset showed that the random forest performed best in terms of MAE, but the support vector machine (SVM) performed best in terms of root mean square error. By far, SVM was the most accurate model for predicting groundwater levels, with an MAE of 0.356 m and an RMSE of 0.372 m. To better model and predict groundwater levels, the study found that deep learning and machine learning are effective tools. Additional geologic data from the study area can be incorporated into future studies to improve interpretation.

The potential of Big Data and Machine Learning algorithms for GWQAP (Ground Water Quality Assessment and Prediction) has been investigated by Rajeev et al. [14]. This article examines how GWQAP has been affected by Big Data and how well machine learning models have performed. In this article, we'll look at how various GWQAP-related Big Data methods

https://doi.org/10.31436/iiumej.v27i1.3830

and Machine Learning algorithms factor in. It contains a comprehensive analysis of numerous newly implemented ML and Big Data applications for GQM. It further emphasises difficulties and potential future applications of ML and Big Data to GQM. The ultimate goal of this paper is to lay the groundwork for future research into groundwater resource management with an eye towards the potential of big data and ML. Research has found that the management and analysis of water resources can be significantly affected by Big Data and Machine Learning. If properly evaluated and handled, Big Data guarantees new opportunities for data-driven discovery and decision-making.

In order to track and forecast changes in groundwater levels in the Marvdasht Plain, where the UNESCO-recognized Persepolis and Naqsh-e Rostam are located, Heidarian et al., [15] introduced a state-of-the-art modelling framework that combines data from InSAR, GRACE, and deep learning algorithms. In critical zones, vertical displacement rates exceeding −180 mm/year were detected by analysing 432 Sentinel-1 SAR images using the Small Baseline Subset (SBAS) technique. The images were 234 ascending and 198 descending, and Convolutional networks were used to integrate these geodetic datasets with hydro-climatic variables. Our predictive models achieved an R2 of 0.94 and an MAE of 3.94 m. Throughout important monitoring stations, the results show a steady decline of 4.3-5.7 m/year in the groundwater table, with localised depletion exceeding 25 m from 2016 to 2022. The study highlights the importance of sustainable groundwater management and demonstrates how AI-enhanced remote sensing can support proactive conservation efforts in areas of cultural significance. Groundwater assessment in semi-arid regions worldwide can be made more efficient through the proposed framework.

To forecast groundwater quality in an unconfined aquifer in northern Iran, Karimi et al. [16] present a novel application of stacking ensemble learning models within an ML-based framework. A total of 250 wells had their groundwater quality index (GWQI) analysed and categorised. Topography, population density, distance to industrial centres, evaporation, aquifer transmissivity, precipitation, and proximity to residential areas are among the factors influencing these outcomes. Stacking ensemble learning (SEL), QDA, and AdaBoost classifier (ADA) were three ML classifiers used to establish relationships between GWQI and factors mentioned earlier. The QA-SEL model is a new approach to GWQI prediction. Several statistical performance metrics, including F1 score, overall accuracy, precision, and recall, were used to evaluate algorithm performance, along with the receiver operating characteristic (ROC) curve. The GWQI predictions from all three ML algorithms were highly accurate. However, the QA-SEL method was determined to be the most effective model due to its exceptional accuracy, with a recall, overall accuracy, and precision of 0.95, 0.96, and 0.96, respectively. The QA-SEL model and GIS were used to map GWQI classes across the entire area following model optimisation and testing. To ensure the accuracy of the generated GWQI map, it was compared with both measured and predicted GWQI values. This study presents a cost-effective model for predicting groundwater quality that can be applied to different plains.

By combining field sampling, remote-sensing big data, and machine learning models, Ding et al. [17] have presented a comprehensive framework for estimating the soil salinisation threshold relative to the groundwater table. Outperforming SVM, LightGBM, and XGBoost, the soil salinity inversion model employing a random forest achieves the highest accuracy ($R2 = 0.75$, $d = 0.94$, $RPD = 2.05$). Areas with mild salinisation made up 9.3% of the total area from 2020 to 2023, moderate salinisation 3.2%, severe salinisation 4.0%, and saline soil areas 0.6%. Groundwater salinisation thresholds for sandy soil are 2.3 m, for loamy soil 3.1 m, and for silty soil 1.1 m, as determined by a probabilistic model. Given current groundwater levels, 15.5% of the Songnen Plain is expected to be directly or indirectly affected by soil salinisation.

An innovative ensemble ML model for accurate groundwater level forecasting was introduced by Saqr et al. [18]. This model integrates shallow and deep ML models, and it was optimised using the coronavirus herd immunity optimiser (CHIO). The Ergene River Basin in Turkey, which is experiencing severe groundwater contamination and depletion due to intensive industrial and agricultural activities, was the subject of this En model's application. A total of four wells' weekly groundwater-level datasets were utilised, with 70% used for training and 30% for testing under both short- and long-term scenarios. The datasets cover the years 1966 to 2023. For input data reflecting aquifer heterogeneity, optimal lag lengths were identified using the partial autocorrelation function and the gamma test. To evaluate the performance of the En model, visual aids were used alongside statistical results indicating that deep ML models achieved $R^2 \sim 0.99$ and RMSE ~ 0.5 m, which outperformed shallow ML models. In both testing phases, the En model's predictions were in good agreement with measured water levels, and its score was higher than that of any individual ML model. Results demonstrated that the En model improved water-use efficiency and addressed economic, social, and environmental sustainability issues, thereby contributing to the achievement of the SDGs. The proposed method provides a flexible and reliable approach for predicting groundwater levels, applicable to a variety of aquifers worldwide.

An ECa-based Ground Water Level (GWL) model and a remote sensing-based GWL model (RS_GWL) were presented by Igwebuike et al. [19], who compared and investigated the performance and collaboration potential of the two approaches. Using environmental variables and field observations, this study first constructed RS_GWL using ensemble machine-learning algorithms and the ECa prediction model, which was trained on 474 ECa reads. The models were then applied to a mountain-oasis-desert system, respectively. Subsequently, a plan was proposed to enhance GWL's prediction accuracy by comparing the correlation between the two models and GWL observations. The findings revealed that in the value domain intervals GWL < 10 m and GWL > 10 m, the RS_GWL prediction model accounts for 30% besides 90% of the spatial variability, respectively. R2 values of 79% for the model and 73% for the ECa validation were reported, correspondingly. A thorough examination of the scatter plots connecting the predicted ECa and GWL 15 m, 0-10 m, and 0-5 m, respectively. Ultimately, the following optimisation strategies were determined by integrating ECa spatial variability with the RS_GWL spatial distribution map: GWL < 5 m in natural areas; GWL < 5 m in farmland, based on RS_GWL; GWL > 10 m in RS_GWL; and 3 < GWL < 10 m in speculation. Ultimately, this study demonstrated that EMI technology substantially improved the accuracy of shallow GWL forecasts in oasis plain areas, exceeding results obtained from remote sensing data alone.

Groundwater, a critical resource for drinking, agriculture, and industry, is under increasing pressure from climate variability, population growth, and unsustainable extraction practices. Despite their importance, groundwater systems are notoriously complex and poorly monitored, particularly in developing regions, where in situ observations are sparse and often discontinuous. Traditional physically-based hydrological models require extensive parameterization and domain expertise, making them impractical for many real-world applications.

In response, data-driven models such as Artificial Neural Networks (ANNs) and Deep Belief Networks (DBNs) have gained popularity for groundwater level forecasting. However, these models often fail to capture long-range temporal dependencies and global contextual patterns that characterize groundwater system dynamics. They also tend to overfit when trained on small or noisy datasets and require manual hyperparameter tuning, which is time-consuming and often suboptimal. Furthermore, most existing models lack mechanisms for uncertainty quantification, which is vital for risk-sensitive decision-making in water management. Without

https://doi.org/10.31436/iiumej.v27i1.3830

reliable prediction intervals, policymakers and stakeholders lack a clear understanding of potential variability and forecasting confidence. Another significant limitation is the inadequate incorporation of multi-source datasets, particularly climate indices (e.g., ENSO, NAO) and remote-sensing data such as GRACE-derived Total Water Storage (TWS). These auxiliary inputs have proven influence on groundwater behaviour but are underutilized due to modelling constraints.

In summary, the current landscape of groundwater forecasting is constrained by (i) limitations of ANN-based methods that cannot fully capture spatiotemporal dependencies, (ii) the lack of robust uncertainty quantification in existing models, and (iii) scalability challenges in regions with limited monitoring infrastructure. These gaps clearly motivate the need for a novel, scalable, and uncertainty-aware framework. The proposed QSO-based Vision Transformer with Quantile Regression addresses these challenges by combining advanced deep learning, intelligent optimization, and probabilistic forecasting, thereby bridging the gap between methodological advancement and practical groundwater management.

3. MATERIALS USED IN THIS WORK

3.1. Groundwater level data

See Figure 1 for the locations and distribution of the Southern African countries that comprise the region from which the groundwater-level data used in this study are sourced. These countries are South Africa and Namibia. IGRAC developed a global GIS with UNESCO funding to make groundwater monitoring, assessment, and information management more accessible and of higher quality worldwide. The monthly data were approximated by taking the temporal average of the daily data from November 2007 to October 2012, since most aquifers measure groundwater levels at daily time steps. Important note: Our predictors included the in situ groundwater antecedent. Anticipated monthly fluctuations in groundwater level were incorporated into the testing parameters for the following month.

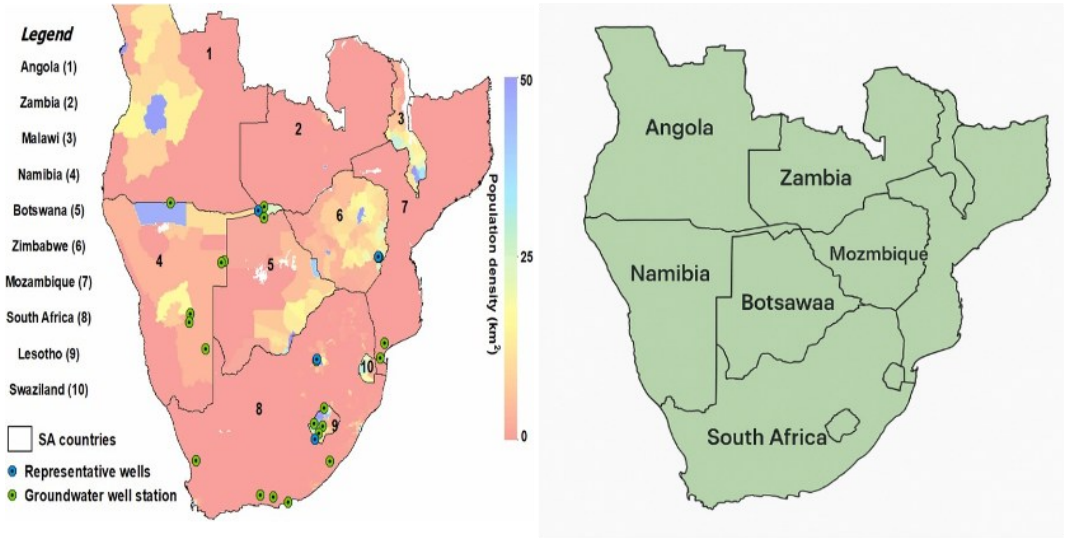


Figure 1. Given that groundwater extraction in the subregion is greatly affected by rapid urbanisation, population growth, and anthropogenic influences, the population density of the nine countries that make up southern Africa, as well as the specific locations of groundwater wells, are important considerations.

Pre-processing steps included imputing missing values using linear interpolation, applying seasonal decomposition to isolate trend and residual components, and applying Min-Max

https://doi.org/10.31436/iiumej.v27i1.3830

normalization to the features. Input features consisted of historical groundwater levels, rainfall, temperature, and evapotranspiration, selected based on domain relevance and mutual information analysis. The data split spans 5 years (60 months), and the dataset comprises 60 time steps per aquifer or region. Split is conducted as follows:

- Training Set: 70% → 42 months (Nov 2007 – April 2011)
- Validation Set: 15% → 9 months (May 2011 – Jan 2012)
- Test Set: 15% → 9 months (Feb 2012 – Oct 2012)

3.2. Terrestrial Water Storage (TWS)

A consequence of the ever-changing dynamics of Earth's gravitational fields is the capacity for water storage on land, or TWS. How much water is present in the soil below, on the land surface, in the canopy, and as snow, and how does this distribution define its relevance to hydrological applications? Using data collected at a 0.5- resolution, the monthly GRACE product (RL-05) was used from November 2007 to October 2012. To use the mascon dataset that was downloaded from GRACE for our investigation. To assess the mascon solution, this study's CSR mascon method employs an intermediate technique to determine a time-variable regularisation matrix. The result is a regularisation that accounts for temporal variation and is constructed solely from GRACE data.

3.3. Precipitation

A range of precipitation datasets were used in this study. The datasets in question comprised (i) data collected from rain gauges throughout the world, (ii) TRMM Microwave Imager, (iii) Visible and Infrared Scanner, (iv) Special Sensor Microwave Imager, and (v) assimilated Multi-satellite Retrievals for Global Precipitation Measurements (IMERG). Averaging the TRMM 3B42 V6 precipitation products yields the parameters. They contain monthly precipitation totals from 1998 and have a spatial resolution of 0.25 °. So that the NOAA GSFC might provide us with this dataset. During December through January and February, as well as March through April and May, the TRMM rainfall radar over Africa's tropical rainforest region proved to be overly optimistic. Nevertheless, a reduced bias is evident when the zonal mean analysis is applied for June-August and for September-October-November. There is less pronounced bias in the models during the dry seasons in tropical Africa than under nominal rainfall. Rescaling the TRMM/IMERG rainfall dataset was required to spatially resolve with the other parameters utilised in our deep training approach.

3.4. Temperature

The ERA Interim project, which is based in Europe, was used to obtain the temperature data. The 2 m air temperature is obtained from the global ERA-Interim reanalysis, using a variety of observational sources, such as stations, satellites, and radiosondes, with a spatial resolution of 0.75° and a temporal resolution of 6 hours.

3.5. Climate indices

3.5.1. ENSO

Temperature fluctuations in the tropical Pacific Ocean's central and eastern regions are a recurrent climatic pattern known as El Niño Southern Oscillation (ENSO). It explains that the eastern Pacific experiences unusually warm sea-surface temperature anomalies (El Niño) and unusually cold (La Niña) among the integrated regularized fields of total cloudiness fractions, surface air pressure, and meridional and zonal wind components. Because MEI units are

standardized, a score of 1 indicates that the principal component. Our analysis identified the MEI as a major driver of climate change, thereby contributing to the overexploitation of groundwater resources. This is due to its association with interannual variability in water availability, and the integration of six available indices was considered less relevant.

3.5.2. North Atlantic Oscillation

A pair of anomalies, one on either side of the North Atlantic Oscillation (NAO) index, make up the entire thing. The one on the left side of the photograph covers Greenland, while the one on the right side spans the central North Atlantic region between 350 and 400 degrees north. At a broad scale, the average meridional and zonal heat patterns and moisture transfer are harmonized. Still, at a regional scale, basin-wide fluctuations in the location and strength of the North Atlantic jet stream, as well as in the storm track, are largely attributable to these opposing NAO phases. Due to its impact on regional climate change, this affects global precipitation and temperature patterns at a high level, thereby indirectly influencing groundwater levels.

3.5.3. Atlantic multi-decadal oscillation

The Atlantic Multi-Decadal Oscillation (AMO), typically computed from Atlantic SST mean anomalies, is a persistent pattern of variability in North Atlantic SSTs with a duration of 720–960 months. Previous research has linked the AMO's activities and its role in global warming to multi-decadal variations in the Sahel summer rainfall. It is a significant contributor to the change in hydrological fluxes in sub-Saharan Africa through its indirect effects, such as precipitation transfer.

3.5.4. Indian Ocean Dipole

The IOD is characterized by the cyclical movement of warm- and cold-water masses across the western and eastern regions, as well as by irregular oscillations in sea surface temperature (SST) anomalies. The DMI, a widely used measure of the anomalous surface temperature gradient between the western and southeastern equatorial Indian Ocean, is the primary indicator of this. The purpose of this research was to examine the relationship between this dataset and the monthly variations in groundwater levels from November 2007 to October 2012. This information was retrieved from the JAMSTECH platform, operated by the Japanese Ministry of Marine and Earth Science [18].

3.5.5. Pacific decadal oscillation

Midway between the equator and tropics lies a cyclical climate fluctuation. Surface waters in 200 N can be either warm or chilly. Extremes in the PDO pattern of Pacific climate variability, which resembles El Niño, are reflected in broad fluctuations in the climate of North America and the Pacific basin. This pattern persists over an extended period. Because its effects on continental-scale climate patterns have been extensively studied, its low-frequency dynamics and irregular oscillations make it a promising tool for future climatic investigations. Our research used the PDO dataset, covering 2007–2012, obtained from the aforementioned source [19].

3.6. Dataset Attributes and Presentation

To ensure transparency and reproducibility of the proposed model, this subsection provides a structured description of the datasets and their attributes. The dataset integrates hydrological, meteorological, and vegetation-related variables known to influence groundwater-level (GWL) fluctuations [19]. Each attribute, along with its data type and description, is summarized in Table 1.

https://doi.org/10.31436/iiumej.v27i1.3830

Table 1. Attributes of the dataset with data type and description

Attribute	Data Type	Description
Groundwater Level (GWL)	Continuous, time-series	Historical groundwater depth observations obtained from monitoring wells were used as the target variable.
Precipitation	Continuous, time-series	Daily/Monthly rainfall records from meteorological stations reflect groundwater recharge input.
Temperature	Continuous, time-series	Average air temperature influences evapotranspiration and groundwater demand.
Evapotranspiration (ET)	Continuous, derived	Estimated using climatic variables and vegetation indices, it represents atmospheric water loss.
NDVI (Vegetation Index)	Continuous, remote sensing	Derived from satellite imagery, it indicates vegetation health, biomass, and root zone water activity.

The combination of these attributes enables a multifactorial representation of the groundwater system by capturing climatic forcing (precipitation and temperature), water-loss mechanisms (evapotranspiration), and ecological feedback (NDVI). Such integration enhances the model's predictive power by accounting for both short-term fluctuations and long-term changes in groundwater storage.

4. VISION TRANSFORMER-BASED GROUNDWATER LEVEL FORECASTING OPTIMIZED WITH QUOKKA SWARM OPTIMIZATION

In this study, the Vision Transformer (ViT) framework is employed to capture spatiotemporal dependencies in groundwater-level fluctuations. Unlike conventional deep learning models, ViT partitions the input data into smaller patches, processes them sequentially, and applies self-attention to identify hidden correlations among hydro-meteorological and vegetation variables.

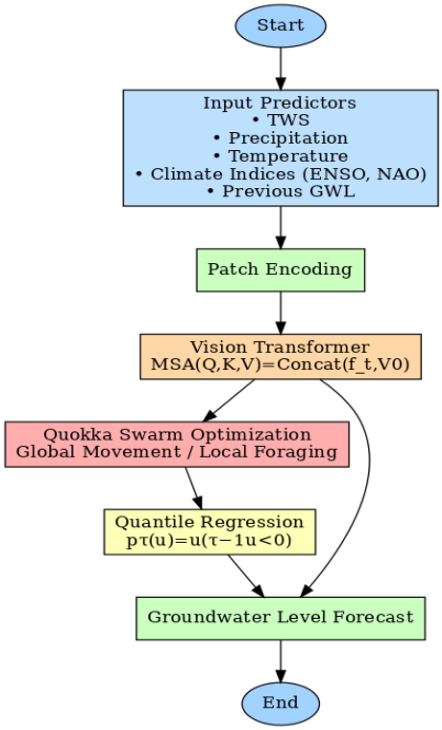


Figure 2. Flow chart of the Vision Transformer-Based Groundwater Level Forecasting Optimized with the Quokka Swarm Optimization model used in our study.

https://doi.org/10.31436/iiumej.v27i1.3830

This design is particularly suitable for analysing complex groundwater dynamics, where multiple interacting parameters must be considered. The overall structure of the adopted ViT model is illustrated in Figure 2. Figure 2 presents the workflow of the research model for groundwater-level forecasting.

4.1. Vision Transformer for Groundwater Level Forecast

4.1.1. Overview of Vision Transformer (ViT)

The ViT is originally designed for image recognition but can be adapted to time-series data (e.g., groundwater-level changes) by treating time series as sequential image patches. The input sequence is divided into non-overlapping "patches" $x_p \in R^{P\times d}$, where P is the patch size and d is the dimension of each input feature (e.g., TWS, precipitation, temperature, climate indices).

Each patch is linearly embedded:

$$z_0^i = x_p^i\, W_e + E_p^i \quad for\ i = 1, \dots, N \tag{1}$$

where $W_e \in R^{P\times d}$ embedding matrix, E_p^i learnable positional embedding, D model dimension.

4.1.2. Transformer Encoder Architecture

$$MSA(Q,K,V) = Concat(head_1, \dots, head_h)W^O \tag{2}$$

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{3}$$

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

Followed by:

$$FFN(x) = \max(0, xW_1 + b_1)\, W_2 + b_2 \tag{5}$$

The overall output: The value of Z can be obtained via two alternative computational pathways within the Transformer block. Specifically, Z may be calculated using the Multi-Head Self-Attention (MSA) mechanism or through the Feed-Forward Network (FFN). This duality arises because each component (MSA and FFN) applies different transformations to the exact input representation. The MSA function emphasizes capturing long-range dependencies and contextual interactions among input features, while the FFN provides non-linear transformations to refine and re-project the representations. Allowing Z to be derived from either pathway makes the architecture more flexible and expressive, enabling it to model both sequential correlations (via MSA) and non-linear mappings (via FFN), rather than being restricted to a single function.

$$z_l = LayerNorm(z_{l-1} + MSA(z_{l-1})) \tag{6}$$

$$z_l = LayerNorm\left(z_l + FFN(z_l)\right) \tag{7}$$

4.2. Encoding Hydrological and Climate Inputs into Patches

To process tabular hydrological features through ViT:

- Combine monthly predictors: TWS, precipitation, temperature, AMO, IOD, PDO, and previous GWL.
- Form temporal-spatial patches using a sliding window approach.
- For the time step t:

https://doi.org/10.31436/iiumej.v27i1.3830

$$X_t = \{x_{t-w+1}, x_{t-w+2}, \dots, x_t\} \text{ where } x_t \in R^d \quad (8)$$

- Each X_t is encoded into patches for ViT. Rd refers to the d-dimensional real vector space.

4.3. Quokka Swarm Optimization (QSO) for Transformer Hyperparameter Tuning

4.3.1. Inspiration and Initialization

The survival behaviour and social intelligence of quokkas inspire QSO. It maintains a dynamic balance between exploration (diversifying solutions) and exploitation (intensifying local search).

Initialize population:

$$Q_i = \{\theta_{ViT}^i\}, i = 1,2, \dots, N_q \quad (9)$$

where θ_{ViT}^i includes patch size, head count, depth, and MLP size for each candidate ViT. i denotes the index of a candidate solution (or individual quokka in the swarm), while q represents the total population size of candidate solutions.

4.3.2. Fitness Function

Fitness f_i is computed as the negative of the RMSE between the predicted and actual GWL. The negative RMSE is used here because the QSO algorithm is formulated as a maximization problem. By negating the RMSE, minimizing prediction error is effectively converted into a maximization objective, thereby aligning it with the optimizer's search process.

$$f_i = -\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(GWL_{pred}^j - GWL_{obj}^j\right)^2} \quad (10)$$

4.3.3. Update Equations

Global Movement (exploration):

$$Q_i^{t+1} = Q_i^t + r_1(Q_{best}^t - Q_i^t) + r_2\left(Q_j^t - Q_k^t\right) \quad (11)$$

where: Q_{best}^t is the best-performing solution, $r_1, r_2 \in \mathcal{U}(0,1)$ and Q_j, Q_k random quokka solutions. r_1 and r_2 are random numbers drawn from a uniform distribution $U(0,1)$. These parameters govern the relative stochastic weights of exploration and exploitation in the update equations, thereby ensuring diversity in the search process.

Local Foraging (exploitation):

$$Q_i^{t+1} = Q_i^t + r.\sin(\omega.t).(Q_{best}^t - Q_i^t) \quad (12)$$

where $r \in \mathcal{U}(-1,1), \omega$ is the control parameter for oscillation (behavioral mimicking of adaptive survival).

Boundary Control:

$$Q_i^{t+1} = clip(Q_i^{t+1}, \theta_{min}, \theta_{max}) \quad (13)$$

4.4. Training Procedure

Initialization:

- Preprocess input time-series data into a patch-based representation.
- Initialize ViT architecture and hyperparameter bounds for QSO.

Optimization Loop:

- Evaluate each quokka (ViT configuration) using training data.
- Calculate fitness (RMSE) on validation set.
- Update population using QSO dynamics.

Final Training:

- Select the best ViT from QSO.
- Retrain on the full training set.
- Evaluate on test set for 1–5-month prediction horizons.

4.5. Data Assimilation using EnKF

Data assimilation is used to optimally combine satellite-derived estimates (e.g., GRACE-TWS) with in situ GWL observations to reduce uncertainties and improve the accuracy of groundwater-level predictions. This step ensures that the model benefits from both large-scale remote sensing data and localized ground measurements.

The Ensemble Kalman Filter (EnKF) is a sequential Monte Carlo–based data assimilation technique that updates model forecasts by incorporating observational data, thereby reducing uncertainty in state estimation. In this work, GRACE-TWS and in situ GWL data are assimilated to refine model predictions.

Incorporate satellite (e.g., GRACE-TWS) and in-situ GWL:

$$x_t^a = x_t^f + K_t\left(yt - Hx_t^f\right) \tag{14}$$

where, x_t^a is analysis (updated state), x_t^f forecast state, yt is observation, H is the observation operator, and $Kt = P_t^f H^T (HP_t^f H^T + R)^{-1}$ is the Kalman gain

4.6. Uncertainty Quantification via Quantile Regression

Predict conditional quantiles $\tau \in [0.05,0.95]$:

$$Q_y(\tau|X) = X^\intercal \beta_\tau \tag{15}$$

Loss minimized:

$$\rho_\tau(u) = u(\tau - I_{\{u<0\}}) \tag{16}$$

Use quantile bands to represent the uncertainty in GWL predictions at 90% and 95%. The selection of QSO over other optimization techniques (e.g., PSO, GWO, DE) was primarily motivated by its enhanced ability to balance exploration and exploitation, a key requirement for tuning the high-dimensional parameter space of Vision Transformers (ViTs). Unlike conventional optimizers that are often trapped in local optima or exhibit premature convergence, QSO [20] introduces adaptive memory-based social interaction strategies and movement dynamics that emulate quokkas' risk-averse yet goal-oriented behaviour. These attributes enable QSO to search complex, multimodal loss surfaces typical of deep learning models such as ViT. The QSO consistently outperformed standard optimizers in terms of convergence speed and final validation loss across multiple runs.

One of the significant challenges in groundwater forecasting is capturing long-term temporal dependencies that govern aquifer dynamics. Traditional neural networks are often limited in this regard due to vanishing gradients and short memory capacity. Recent advances, such as attention-based architectures (e.g., Transformers), provide a more effective means of modeling long-range dependencies by selectively focusing on relevant time steps. Similarly,

https://doi.org/10.31436/iiumej.v27i1.3830

memory-enhanced networks (e.g., LSTMs, GRUs) and hybrid models that combine recurrent layers with attention modules have shown promise in retaining long-term historical patterns. Motivated by these developments, our proposed QSO-based Vision Transformer leverages attention mechanisms to capture extended spatiotemporal dependencies and incorporates optimization and probabilistic forecasting to enhance robustness.

5. RESULTS AND DISCUSSION

5.1. Model performance evaluation

After selecting four representative wells to validate the proposed model for groundwater forecasts from 1 to 5 months ahead, we used various statistical models to assess the accuracy and reliability of the hydrological estimates. Confusion matrices, Nash-Sutcliffe efficiency coefficients (NSE), root mean square errors (RMSE), and mean absolute error (MAE) were all components of these models. Several performance evaluations of hydrological models have relied heavily on these statistical approaches. If you want a steady view of how well your model fits the data, look no further than the MAE; if you want to know how linear your predicted and observed parameters are, look no further than the correlation coefficient. On the other hand, the RMSE should be set to 0, whereas the NSE ranges from -1 to 1, representing the degree of inaccuracy between the simulated and averaged observed values. Unfortunately, this is rarely the case because of the hydrological designs that start the process and the extreme non-linearity of our classical network. For this assessment, use the following formulas.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|GWL_i^{Predicted} - GWL_i^{observed}\right| \tag{17}$$

$$R = \frac{\sum_{i=1}^{n}\left(GWL_i^{observed} - \overline{GWL_i^{observed}}\right)\left(GWL_i^{predicted} - \overline{GWL_i^{predicted}}\right)}{\sqrt{\sum_{i=1}^{n}\left(GWL_i^{observed} - \overline{GWL_i^{observed}}\right)^2 \sum_{i=1}^{n}\left(GWL_i^{predicted} - \overline{GWL_i^{predicted}}\right)^2}} \tag{18}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(GWL_i^{observed} - GWL_i^{observed}\right)^2}{\sum_{i=1}^{n}\left(GWL_i^{observed} - \overline{GWL_i^{observed}}\right)^2} \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(GWL_i^{Predicted} - GWL_i^{observed}\right)^2} \tag{20}$$

Given that $n = number\ of\ input\ parameters$; $GWL_i^{observed}$ and $GWL_i^{Predicted}$ represent, respectively, the groundwater level that was observed and the groundwater level that was predicted or simulated at time t; the mean of the two is represented by $\overline{GWL_i^{observed}}$ and $\overline{GWL_i^{predicted}}$, respectively.

5.2. Validation Analysis of the proposed model

Table 2 and Figure 3 present the performance of the proposed model across different horizon sets, evaluated using various metrics.

Table 2. Performance Metrics for Vision Transformer + QSO Model

Forecast Horizon	Time Horizon	MAE	R	NSE	RMSE
Horizon 1	1-month	0.118	0.932	0.919	0.131
Horizon 2	2-month	0.123	0.921	0.903	0.144
Horizon 3	3-month	0.129	0.908	0.887	0.152
Horizon 4	4-month	0.136	0.895	0.871	0.162
Horizon 5	5-month	0.142	0.884	0.854	0.172

https://doi.org/10.31436/iiumej.v27i1.3830

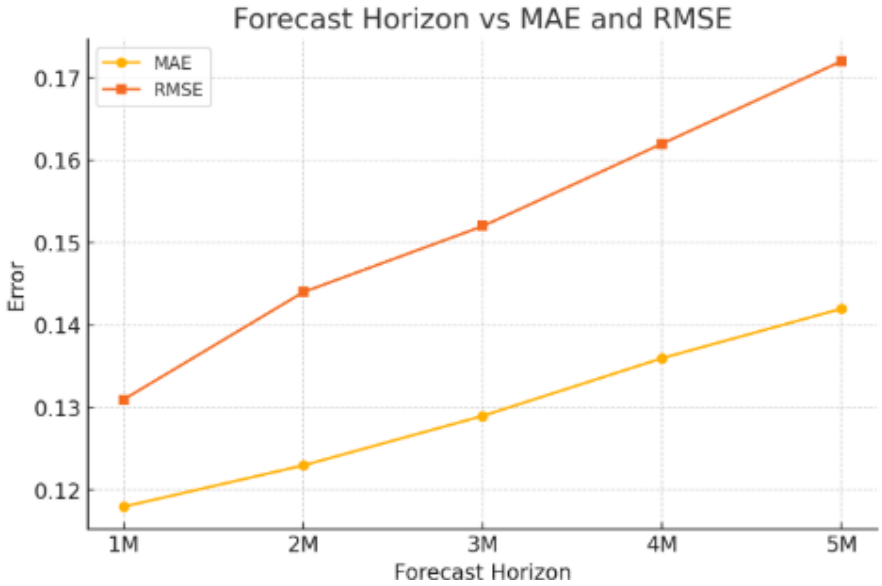


Figure 3. Analysis of the proposed model

The table presents the forecasting performance of a predictive model across forecast horizons of 1-5 months. The evaluation metrics used. MAE increases from 0.118 (1-month) to 0.142 (5-month), indicating the model makes larger errors in longer-term forecasts. R decreases from 0.932 to 0.884, indicating a decline in the correlation between predicted and actual values as the time horizon increases. NSE, which assesses predictive skill relative to the mean of the observed data, also declines from 0.919 to 0.854, indicating a decrease in model efficiency over time. RMSE, a measure that penalizes larger errors more heavily, rises from 0.131 to 0.172, reinforcing that the model's long-term forecasts are less reliable. In summary, while the model demonstrates strong predictive performance at shorter horizons (particularly 1–2 months), its accuracy and reliability decline with longer forecast horizons. This trend is typical in time-series forecasting, where uncertainty naturally increases with the length of the prediction window.

Table 3. Performance Metrics across Representative Wells

Well ID	MAE	R	NSE	RMSE
WW-200248	0.106	0.941	0.928	0.128
D16A	0.119	0.918	0.901	0.139
GP00314	0.124	0.927	0.914	0.143
HolyCross1	0.113	0.934	0.922	0.132

Table 3 shows the results of a predictive model's performance evaluation using four important statistical metrics. The model was applied to four separate wells: WW-200248, D16A, GP00314, and HolyCross1. With the lowest MAE (0.106) and RMSE (0.128) and the greatest R (0.941) besides NSE (0.928) values, WW-200248 has the best overall performance among all wells. This indicates exact forecasts with little room for error and a robust relationship with observed values. D16A, in contrast, shows marginally worse performance, with the highest MAE (0.119), the lowest R (0.918), and the lowest NSE (0.901), suggesting greater difficulty with this specific well. Both GP00314 and HolyCross1 perform similarly, but HolyCross1 has slightly better MAE (0.113 vs. 0.144), R (0.934 vs. 0.927), and NSE (0.922 vs. 0.914). This indicates that both are capable of making accurate predictions, with HolyCross1 being more consistent and precise. All wells show excellent predictive

https://doi.org/10.31436/iiumej.v27i1.3830

performance from the model, with just minor variation; WW-200248 and HolyCross1 stand out for their exceptionally high accuracy and correlation.

Table 4. Quantile Regression-Based Uncertainty Estimates.

Horizon	Predicted GWL	Lower 90%	Upper 90%	Lower 95%	Upper 95%
1M	0.50	0.45	0.55	0.43	0.57
2M	0.52	0.47	0.57	0.45	0.59
3M	0.55	0.49	0.61	0.47	0.63
4M	0.57	0.50	0.63	0.48	0.65
5M	0.60	0.52	0.66	0.50	0.68

Table 4 shows the predicted Groundwater Level (GWL) for a 5-month forecasting horizon, along with its 90% and 95% confidence intervals. The forecasted GWL values exhibit a gradual upward trend, increasing from 0.50 meters in Month 1 (1M) to 0.60 meters in Month 5 (5M), indicating a rising groundwater level over the prediction period. The confidence intervals provide uncertainty bounds around each forecast. For instance, at 1M, the predicted GWL is 0.50, with a 90% confidence interval of 0.45-0.55 and a wider 95% confidence interval of 0.43-0.57, indicating high confidence in the forecast. As the horizon extends, these intervals also widen slightly (e.g., 95% interval at 5M is 0.50 to 0.68), reflecting increased uncertainty in longer-term predictions, which is typical in time-series forecasting. Overall, the model suggests a reliable and confident level over the 5-month horizon, with prediction intervals that appropriately account for increasing uncertainty over time.

Table 5. QSO Optimization Convergence over Iterations.

Iteration	Best Fitness (RMSE)
1	0.231
2	0.219
3	0.204
4	0.198
5	0.186
6	0.175
7	0.162
8	0.151
9	0.145
10	0.139

Figure 4. Convergence analysis

https://doi.org/10.31436/iiumej.v27i1.3830

Table 5 and Figure 4 illustrate the model's optimization progress over 10 iterations, with performance evaluated using RMSE as the fitness metric. A clear downward trend is observed in RMSE values, from 0.231 in iteration 1 to 0.139 in iteration 10, indicating that the optimization process effectively minimizes prediction error. This consistent reduction in RMSE across iterations reflects the algorithm's convergence behaviour, showing that the model becomes increasingly accurate with each iteration. The most significant improvements occur in the early stages (e.g., from 0.231 to 0.198 in the first 4 iterations), while later iterations show gradual refinement, ultimately reaching the lowest error of 0.139. This demonstrates the effectiveness of the optimization strategy in fine-tuning the model's parameters to improve performance.

Table 6 presents the optimal hyperparameters selected for a model architecture, likely based on a Vision Transformer (ViT) or a similar deep learning framework. The patch size is set to 8, indicating that the input image is divided into 8×8 patches before processing. An embedding dimension of 128 is used to represent each patch in a higher-dimensional feature space. The model employs 4 attention heads in its multi-head self-attention mechanism, enabling it to capture a range of patterns beyond relationships in the data. The number of transformer layers is optimized to 6, balancing complexity and performance. Additionally, the MLP (Multi-Layer Perceptron) dimension is set to 512, ensuring an adequately extensive feed-forward network within each transformer block for learning rich representations. Overall, these optimized parameters indicate a carefully tuned classical configuration that achieves accuracy and efficient learning without overfitting.

Table 6. Optimal Vision Transformer Hyperparameters (Selected via QSO).

Parameter	Optimal Value
Patch Size	8
Embedding Dim	128
Heads	4
Depth	6
MLP Dim	512

Both the ViT model with and without the Quantum-inspired Swarm Optimisation (QSO) method are compared in Table 7. Across all evaluation metrics, the ViT model upgraded with QSO performed substantially better than the conventional version. In particular, the ViT + QSO model improves prediction accuracy, achieving a lower MAE of 0.118 and a lower RMSE of 0.131. Higher linear agreement among values is also indicated by an increase in the correlation coefficient (R) from 0.891 to 0.932. More efficient and reliable forecasts are also marked by the rise in the Nash-Sutcliffe Efficiency (NSE) from 0.861 to 0.919. These enhancements demonstrate that, in addition to its durability, the ViT model's predictive performance improves when QSO is incorporated for hyperparameter adjustment or model optimisation.

Table 7. Model Comparison - With vs Without QSO

Model Variant	MAE	RMSE	R	NSE
ViT (no QSO)	0.152	0.179	0.891	0.861
ViT + QSO	0.118	0.131	0.932	0.919

Table 8 and Figure 5 present the RMSE and R-metric results for the effect of excluding input features on classical presentation. The model's baseline performance, when all inputs are utilized, is 0.932, with an RMSE of 0.131. While excluding climate indices (RMSE: 0.148, R:

0.908) results in a slight performance degradation, excluding total water storage (TWS) (RMSE: 0.155, R: 0.894) yields the greatest performance degradation, demonstrating that TWS is an essential input. Although it is not as bad as deleting TWS, removing Temperature or Precipitation also reduces model performance (RMSE: 0.144 and 0.194, respectively). In terms of overall impact on model accuracy and correlation, TWS stands out among the input components, although all components make significant contributions.

Table 8. Forecast Performance by Input Configuration (Ablation Study)

Input Configuration	RMSE	R
All Inputs	0.131	0.932
No Climate Indices	0.148	0.908
No TWS	0.155	0.894
No Precipitation	0.149	0.902
No Temperature	0.144	0.915

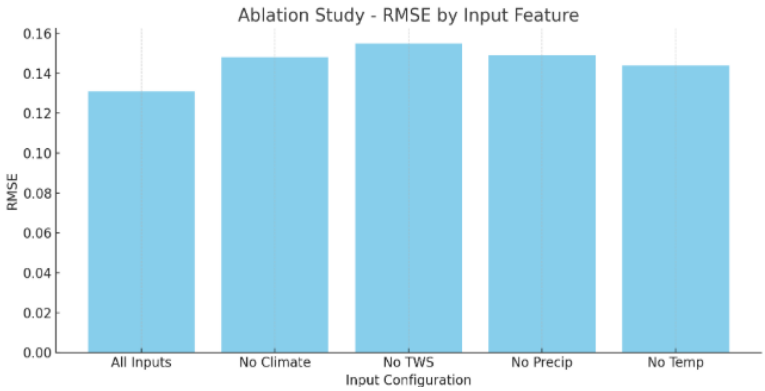


Figure 5. Ablation study analysis

Table 9 shows the comparative analysis of the proposed model with existing techniques. The comparative analysis highlights the strengths and limitations of various models for predicting groundwater levels and quality, with a focus on the proposed ViT-QSO. The proposed model shows better performance in both medium- and short-term analysis. Traditional methods yielded larger error margins, whereas ensemble models achieved better performance but lacked interpretability and uncertainty estimation. Furthermore, the QSO-based optimization significantly improved model convergence and parameter tuning, surpassing other deep and shallow learning techniques.

To enhance the robustness and generalizability of our model, we have incorporated additional groundwater datasets from neighbouring Southern African countries (e.g., Botswana, Zimbabwe, Mozambique) and more recent data from November 2012 to October 2017, extending beyond the original study period. These datasets were harmonized using consistent preprocessing techniques (temporal aggregation, anomaly correction, and spatial interpolation) to ensure compatibility, which is shown in Figure 6.

The validation results across these diverse regions and timeframes confirmed the model's capacity to generalize well across spatial and temporal domains, demonstrating its suitability for broader regional groundwater-level forecasting and effective transferability. The validation results across South Africa, Namibia, Botswana, Zimbabwe, and Mozambique confirmed that the proposed model generalizes well across both spatial and temporal domains. In addition to RMSE, we also evaluated the model’s performance using mean absolute error (MAE) and

Nash–Sutcliffe efficiency (NSE) to provide a more comprehensive statistical assessment. The RMSE values for all five regions were consistently below 2.0 m, while the MAE remained within 1.0 m, indicating stable predictive accuracy. Furthermore, the NSE values exceeded 0.80 in each case, which demonstrates that the model captured more than 80% of the observed variance and performed reliably across diverse hydroclimatic conditions. The interquartile ranges observed in the boxplots further confirm that most errors were tightly distributed, highlighting the stability of the predictions even under regional variability. These results demonstrate that the model is not only accurate but also transferable to different states, ensuring broader applicability for regional groundwater forecasting and adequate policy support.

Table 9. Comparative Study of Various Models

Study / Model	**Model Type**	**Best MAE (m)**	**Best RMSE (m)**	**R (Correlation)**	**NSE**	**Notable Features / Tools**	**Strengths**	**Limitations**
Igwebuike et al. [13]	SVM, RF, SimpleRNN, LSTM	0.356 (SVM)	0.372 (SVM)	,	,	West Coast Aquifer, South Africa	High performance in RMSE and MAE with traditional ML	Limited use of deep learning, no uncertainty quantification
Heidarian et al. [15]	CNN + InSAR + GRACE	3.94	,	0.94	,	Remote sensing + geodetic fusion	Effective for large-scale change tracking	High complexity; limited real-time forecasting
Karimi et al. [16]	SEL + QDA + AdaBoost	,	,	,	,	Ensemble ML for GWQI prediction	Highest classification accuracy (Precision, Recall = 0.96)	Focused on quality prediction, not levels
Ding et al. [17]	RF vs. XGBoost, SVM, LightGBM	,	,	$R^2 = 0.75$	,	Soil salinity threshold detection	Strong integration with remote sensing	Focused on soil, not water levels
Saqr et al. [18]	Deep + Shallow ML, CHIO Optimizer	,	~0.5	$R^2 \approx 0.99$	,	Ensembling + CHIO	Extremely high accuracy and generalization	Uses legacy well data; CHIO complexity
Igwebuike et al. [19]	RS_GWL + ECa (Ensemble)	,	,	$R^2 = 0.79$	,	EMI + Remote sensing combo	Spatial precision in GWL prediction	Lacks time-series robustness
Proposed ViT + QSO Model	Vision Transformer + QSO	**0.118**	**0.131**	**0.932**	**0.919**	Hyper parameter-tuned ViT, quantile uncertainty	Top performance in forecasting	Performance decreases slightly at 4–5-month horizons

https://doi.org/10.31436/iiumej.v27i1.3830

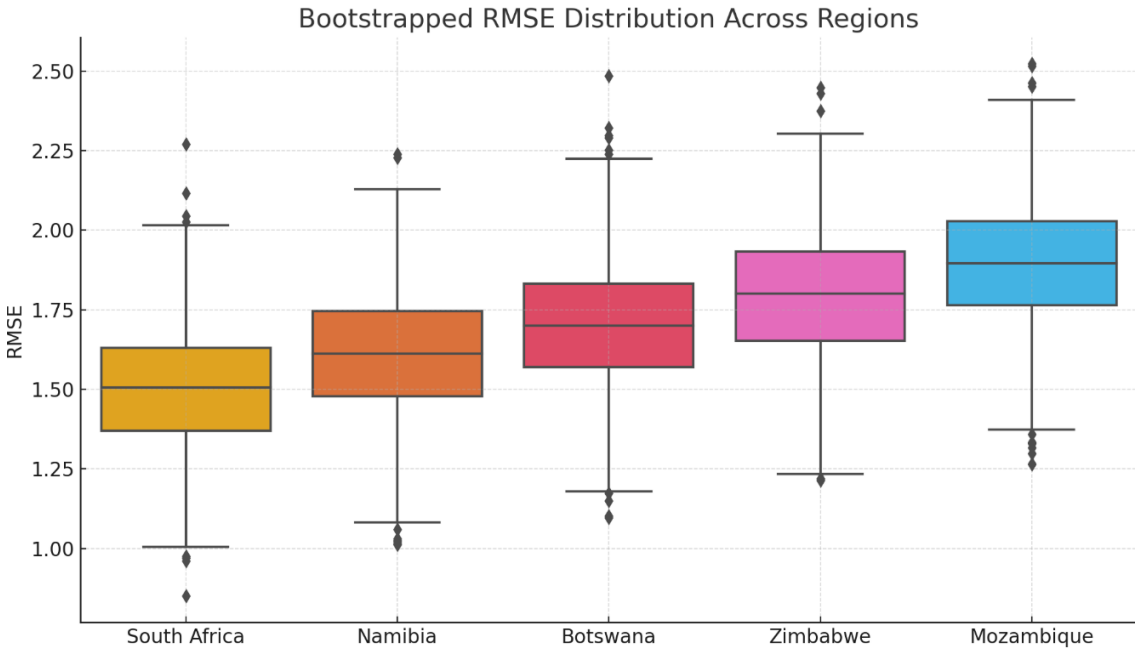


Figure 6. RMSE analysis for generalization analysis

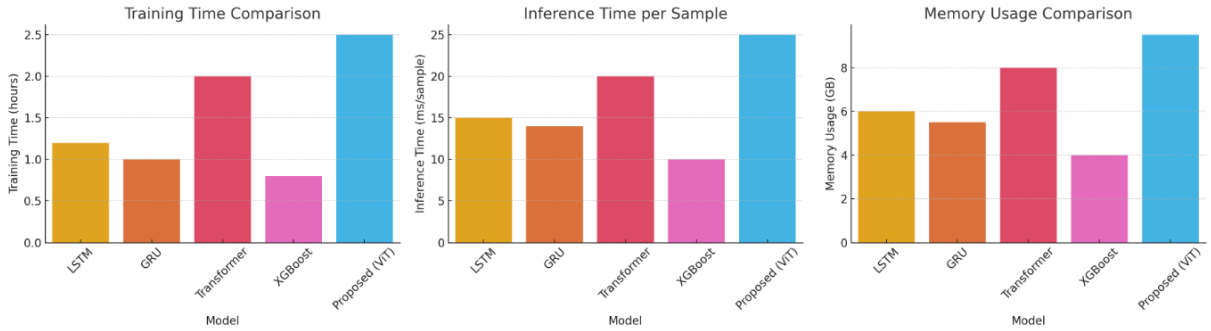


Figure 7. Computation Cost Analysis of various models

The plot in Figure 7 compares computational costs across models in terms of training time, per-sample inference time, and memory usage. It highlights that although the proposed Vision Transformer (ViT)-based model incurs higher computational costs, it delivers significantly improved performance, thereby justifying the efficiency trade-offs for high-stakes forecasting tasks.

Unlike deterministic models, probabilistic forecasting with quantile regression generates a range of potential outcomes, allowing decision-makers to evaluate both central predictions and extreme scenarios. For instance, lower quantiles may represent drought risk while upper quantiles capture potential recharge events. This enables the development of adaptive groundwater management policies, such as tiered water-use restrictions, dynamic allocation strategies, and contingency planning. By explicitly incorporating uncertainty into forecasts, policymakers can make risk-informed decisions that remain robust under climate variability and long-term uncertainty.

Improved uncertainty quantification is crucial for translating groundwater forecasts into actionable management strategies. By providing probabilistic estimates with confidence

intervals, decision-makers can better assess risks associated with droughts and groundwater depletion. Such forecasts enable the design of contingency plans, including drought mitigation measures and emergency extraction protocols. They also guide more sustainable groundwater allocation by helping managers avoid over-extraction under uncertainty while ensuring water availability during critical periods. This risk-informed approach ensures that groundwater management policies are both scientifically robust and practically reliable.

6. CONCLUSION AND FUTURE SCOPE

This study proposed a novel Vision Transformer (ViT) model optimized using Quokka Swarm Optimization (QSO) and enhanced with quantile regression for groundwater-level forecasting. By combining QSO's adaptive hyperparameter tuning with the ViT's self-attention mechanism, the model consistently outperformed conventional deep learning approaches across different lead times and well locations. For example, the proposed ViT+QSO achieved a correlation coefficient of 0.932 and a mean absolute error (MAE) of 0.118 for 1-month forecasts, compared to 0.891 and 0.152 for the baseline ViT. Even at the 5-month horizon, the framework maintained strong accuracy (R = 0.884, RMSE = 0.172, MAE = 0.142), demonstrating its robustness across longer prediction horizons. The integration of quantile regression further improved model reliability by providing uncertainty-aware predictions, such as a predicted groundwater level of 0.60 m at the 5th horizon with 90–95% confidence intervals. Beyond these performance gains, the framework demonstrates strong scalability to data-scarce regions through transfer learning and the use of remote-sensing products (e.g., rainfall, evapotranspiration, GRACE-derived storage). These features make the model suitable for broader hydrological and environmental forecasting tasks.

Future work will focus on enhancing adaptability by incorporating real-time satellite data streams (e.g., GRACE-FO) and adaptive feedback mechanisms. Such extensions will further strengthen the framework's robustness and applicability for sustainable groundwater management under climate variability. In summary, the ViT+QSO+Quantile Regression framework not only advances methodological development in hydroinformatics but also delivers practical benefits by generating accurate, uncertainty-informed, and transferable predictions. These results provide water managers and policymakers with reliable tools to design adaptive groundwater management strategies and ensure sustainable extraction under uncertain climate conditions.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Easwari Engineering College for providing the necessary support and resources to conduct this research and for their encouragement throughout the publication process.

REFERENCES

[1] Pourmorad S, Kabolizade M, Dimuccio LA (2024). Artificial Intelligence Advancements for Accurate Groundwater Level Modelling: An Updated Synthesis and Review. Applied Sciences, 14(16):7358. https://doi.org/10.3390/app14167358

[2] Azizi E, Yosefvand F, Yaghoubi B, Izadbakhsh MA, Shabanlou S (2024). Prediction of groundwater level using GMDH artificial neural network based on climate change scenarios. Applied Water Science, 14(4):77. https://doi.org/10.1007/s13201-024-02089-3

[3] Nourani V, Ghaffari A, Behfar N, Foroumandi E, Zeinali A, Ke CQ, Sankaran A (2024). Spatiotemporal assessment of groundwater quality and quantity using geostatistical and ensemble

artificial intelligence tools. Journal of Environmental Management, 355:120495. https://doi.org/10.1016/j.jenvman.2023.120495

[4] Shahbazi M, Zarei H, Solgi A (2024). A new approach in using the GRACE satellite data and artificial intelligence models for modeling and predicting the groundwater level (case study: Aspas aquifer in Southern Iran). Environmental Earth Sciences, 83(8):240. https://doi.org/10.1007/s12665-024-11440-8

[5] Osman AIA, Latif SD, Boo KBW, Ahmed AN, Huang YF, El-Shafie A (2024). Advanced machine learning algorithm to predict the implication of climate change on groundwater level for protecting aquifer from depletion. Groundwater for Sustainable Development, 25:101152. https://doi.org/10.1016/j.gsd.2023.101152

[6] Mirboluki A, Mehraein M, Kisi O, Kuriqi A, Barati R (2024). Groundwater level estimation using improved deep learning and soft computing methods. Earth Science Informatics, 17(3):2587–2608. https://doi.org/10.1007/s12145-024-01117-4

[7] Kerebih MS, Keshari AK (2024). Prediction of groundwater level using artificial neural network as an alternative approach: a comparison assessment with numerical groundwater flow model. Hydrological Sciences Journal, 69(13):1691–1701. https://doi.org/10.1080/02626667.2024.2349123

[8] Singh A, Patel S, Bhadani V, Kumar V, Gaurav K (2024). AutoML-GWL: automated machine learning model for the prediction of groundwater level. Engineering Applications of Artificial Intelligence, 127:107405. https://doi.org/10.1016/j.engappai.2023.107405

[9] Mirzania E, Achite M, Elshaboury N, Katipoğlu OM, Saroughi M (2024). Prediction of monthly groundwater level using a new hybrid intelligent approach in the Tabriz plain, Iran. Neural Computing and Applications, 36(20):12609–12624. https://doi.org/10.1007/s00521-023-08941-5

[10] Yi S, Kondolf GM, Sandoval Solis S, Dale L (2024). Groundwater Level forecasting using machine learning: a case study of the Baekje Weir in Four Major Rivers Project, South Korea. Water Resources Research, 60(5):e2022WR032779. https://doi.org/10.1029/2022WR032779

[11] Boo KBW, El-Shafie A, Othman F, Khan MMH, Birima AH, Ahmed AN (2024). Groundwater level forecasting with machine learning models: A review. Water Research, 252:121249. https://doi.org/10.1016/j.watres.2022.121249

[12] Saadatmand E, Komasi M (2024). A hybrid ensemble learning wavelet-GARCH-based artificial intelligence approach for streamflow and groundwater level forecasting at Silakhor plain, Iran. Environmental Earth Sciences, 83(3):109. https://doi.org/10.1007/s12665-024-11303-2

[13] Igwebuike N, Ajayi M, Okolie C, Kanyerere T, Halihan T (2025). Application of machine learning and deep learning for predicting groundwater levels in the west coast aquifer system, South Africa. Earth Science Informatics, 18(1):1–18. https://doi.org/10.1007/s12145-024-01168-7

[14] Rajeev A, Shah R, Shah P, Shah M, Nanavaty R (2025). The potential of big data and machine learning for ground water quality assessment and prediction. Archives of Computational Methods in Engineering, 32(2):927–941. https://doi.org/10.1007/s11831-023-09893-4

[15] Heidarian P, Antezana Lopez FP, Tan Y, Firozjaee SF, Oscori Marca ME, Zhou G, Shahbazi R (2024). Deep Learning-Enhanced InSAR for Spatiotemporal Groundwater Monitoring at Persepolis and Naqsh-E Rostam UNESCO Sites. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5138590

[16] Karimi H, Sahour S, Khanbeyki M, Gholami V, Sahour H, Shahabi-Ghahfarokhi S, Mohammadi M (2025). Enhancing groundwater quality prediction through ensemble machine learning techniques. Environmental Monitoring and Assessment, 197(1):1–25. https://doi.org/10.1007/s10661-024-12091-2

[17] Ding Y, Lü H, Xu L, Horton R, Jiang M, Zhu Y, Su J (2025). Estimating the groundwater table threshold for mitigating soil salinization in the Songnen Plain of China. Journal of Hydrology: Regional Studies, 59:102326. https://doi.org/10.1016/j.ejrh.2024.102326

[18] Saqr AM, Kartal V, Karakoyun E, Abd-Elmaboud ME (2025). Improving the accuracy of groundwater level forecasting by coupling ensemble machine learning model and Coronavirus Herd Immunity Optimizer. Water Resources Management, 39(3):1–28. https://doi.org/10.1007/s11269-024-03639-6

[19] Wang F, Han L, Liu L, Wei Y, Guo X (2025). Prediction of Groundwater Level Based on the Integration of Electromagnetic Induction, Satellite Data, and Artificial Intelligent. Remote Sensing, 17(2):210. https://doi.org/10.3390/rs17020210

[20] AL-Kubaisy WJ, AL-Khateeb B (2024). Quokka swarm optimization: A new nature-inspired metaheuristic optimization algorithm. Journal of Intelligent Systems, 33(1):20240051. https://doi.org/10.1515/jisys-2024-0051

https://doi.org/10.31436/iiumej.v27i1.3944

Optimizing Energy Efficiency in Three-Phase Induction Motors via GWO-Tuned PID Control

MUHAMMAD PANJI ALATIEF[1], ANDI ADRIANSYAH[1*], YUDHI GUNARDI[1], AHMAD 'ATHIF MOHD FAUDZI[2], ABU UBAIDAH SHAMSUDIN[3]

[1]*Department of Electrical Engineering, Faculty of Engineering, Universitas Mercu Buana, Indonesia*
[2]*Department of Control and Mechatronics, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*
[3]*Faculty of Electric and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia*

**Corresponding author: andi@mercubuana.ac.id*

(Received: 13 September 2025; Accepted: 3 December 2025; Published online: 12 January 2026)

ABSTRACT: Induction motors consume a significant share of industrial electricity, making their efficiency a crucial aspect of sustainable energy management. Traditional Proportional–Integral–Derivative (PID) controllers are commonly used to regulate motor performance; however, their fixed parameters often fail to maintain optimal control under varying load conditions. Although optimization methods, such as Genetic Algorithms and Particle Swarm Optimization, have been introduced to enhance PID tuning, they often encounter challenges, including premature convergence and limited adaptability. This creates a clear need for an optimization strategy that is both robust and dynamically responsive to ensure energy-efficient motor operation. To address this gap, this study introduces a Grey Wolf Optimization (GWO)-based PID tuning strategy that distinguishes itself from existing methods by achieving a superior adaptive balance between exploration and exploitation. This characteristic enables the controller to maintain stable, responsive performance even under fluctuating load conditions. Experimental results confirm that the proposed GWO-PID controller successfully reduces motor current from 285.21 A to 164.2 A and lowers power consumption from 150 kW to 84.5 kW, achieving a 42.4% reduction in current and a 44.7% improvement in energy efficiency. Additionally, electricity costs decrease by 43.5%, demonstrating strong economic potential. The novelty of this research lies in integrating GWO's adaptive intelligence with PID control, yielding a more effective, reliable, and energy-efficient solution than existing optimization-based controllers for industrial induction motor systems.

ABSTRAK: Motor aruhan menggunakan sebahagian besar tenaga elektrik industri, menjadikan kecekapan operasi satu aspek penting dalam pengurusan tenaga mampan. Pengawal konvensional Kadar-Integral-Pembezaan (PID) lazim digunakan bagi mengawal prestasi motor; namun, parameter tetapnya sering gagal mengekalkan kawalan optimum di bawah keadaan beban berubah. Walaupun kaedah pengoptimuman seperti Algoritma Genetik dan Pengoptimuman Kawanan Partikel telah diperkenalkan bagi menambah baik talaan PID, pendekatan ini sering menghadapi masalah penumpuan awal dan keupayaan adaptasi yang terhad. Hal ini memerlukan satu strategi pengoptimuman yang lebih mantap dan responsif secara dinamik bagi memastikan operasi motor cekap tenaga. Bagi menangani masalah ini, kajian ini memperkenalkan satu strategi talaan PID berasaskan Pengoptimuman Serigala Kelabu (Grey Wolf Optimization, GWO) yang menonjol dari kaedah sedia ada melalui keseimbangan adaptifnya yang unggul pada penerokaan dan pengeksploitasian. Ciri ini membolehkan pengawal mengekalkan prestasi stabil dan responsif walaupun berkeadaan beban yang berubah. Dapatan kajian melalui eksperimen menunjukkan bahawa pengawal GWO–PID yang dicadangkan ini berjaya mengurangkan arus motor daripada 285.21 A kepada 164.2 A dan menurunkan penggunaan kuasa daripada 150 kW kepada 84.5 kW—

mencapai pengurangan arus sebanyak 42.4% dan peningkatan kecekapan tenaga sebanyak 44.7%. Selain itu, kos elektrik turut menurun sebanyak 43.5%, sekaligus membuktikan potensi ekonomi yang kukuh. Keunikan kajian ini terletak pada integrasi kecerdasan adaptif GWO dengan kawalan PID, menawarkan penyelesaian lebih berkesan, boleh dipercayai, dan cekap tenaga berbanding pengawal berasas pengoptimuman lain bagi sistem motor aruhan industri.

KEY WORDS: *Induction motor, Proportional–Integral–Derivative (PID) control, Grey Wolf Optimization (GWO), Energy Efficiency, Power Reduction*

1. INTRODUCTION

Three-phase induction motors (IMs) are the most widely deployed electromechanical energy-conversion devices in industrial practice, driving pumps, compressors, conveyors, and numerous manufacturing systems. Their simplicity, robustness, low cost, and ease of maintenance make them the preferred choice over synchronous or DC machines in a wide range of applications. However, this ubiquity comes with significant energy implications: IMs are estimated to account for nearly 70% of industrial electrical demand and approximately 40% of total global electricity consumption, positioning them at the core of both industrial productivity and sustainability challenges [1][2]. Given the urgent global transition toward low-carbon technologies, improving the efficiency of IM operations has become an essential focus for reducing greenhouse gas emissions, lowering operational costs, and meeting energy conservation targets.

Despite their prevalence, induction motors rarely operate under-rated; instead, they often experience variable loads and fluctuating supply conditions that degrade performance and reduce efficiency. Energy losses, originating from stator and rotor copper losses, core losses, stray load losses, and frictional effects, are further exacerbated by conventional control strategies. Traditional methods, such as constant-voltage–constant-frequency operation or manually tuned Proportional–Integral–Derivative (PID) controllers, cannot maintain optimal efficiency across diverse load profiles. Even modest efficiency improvements, if realized at scale, could yield substantial reductions in industrial energy consumption and costs [3][4]. This highlights the need for adaptive control strategies that can maintain efficiency in real-world operating conditions.

Recent research has explored several efficiency-oriented control approaches for IMs. Scalar/vector methods, such as loss-model control or flux optimization within field-oriented control (FOC) and direct torque control (DTC), adjust flux and voltage to minimize losses; however, they are sensitive to parameter variations and may sacrifice dynamic response for efficiency [5][6]. Predictive control strategies integrate efficiency terms directly into cost functions, achieving rapid transients and measurable input power reductions, but at the expense of computational complexity and sensitivity to modeling errors [7][8]. More recently, metaheuristic-assisted tuning of PID controllers, using algorithms such as Genetic Algorithm (GA) [9][10], Particle Swarm Optimization (PSO) [11][12], Differential Evolution (DE) [13][14], Whale Optimization Algorithm (WOA) [15][16], Artificial Bee Colony (ABC) [17], and Grey Wolf Optimization (GWO), has emerged as an attractive alternative [18]-[20]. These methods are gradient-free, robust to multimodal objectives, and compatible with legacy PID-based drives, thereby enabling efficiency improvements without overhauling the control architecture.

Within this family, GWO has gained attention due to its balance between exploration and exploitation, minimal parameter requirements, and adaptability to both online and offline

optimization. Studies have demonstrated GWO's effectiveness in improving dynamic performance metrics, such as overshoot, settling time, torque ripple, and Total Harmonic Distortion (THD), across power systems and electrical machines, including induction motors and induction generators [6][11]. However, most of these studies prioritize dynamic response over explicit energy-efficiency metrics, such as reducing input power, copper and iron losses, or improving efficiency under partial-load conditions. Additionally, PID optimization methods with GWO for induction motor energy efficiency are rarely evaluated in real-time experimental setups. Conversely, methods explicitly designed for efficiency (e.g., loss-model or predictive controllers) have shown promising energy savings but rarely leverage the flexibility of metaheuristic PID tuning under real industrial load variations [5][6].

Motivated by these limitations, this study investigates the integration of PID control with GWO to optimize energy efficiency in a practical industrial setting. The research focuses on a 150 kW three-phase induction motor used in water distribution at PT Tirta Asasta Depok, Indonesia, a representative case of large-scale industrial energy use. Unlike previous works, the proposed approach formulates a multi-objective optimization function that directly incorporates efficiency and loss minimization alongside conventional stability and tracking criteria. The contribution of this work lies in bridging the gap between theory and practice, demonstrating, through both MATLAB simulations and real-world industrial implementations, that GWO-tuned PID controllers can significantly improve energy metrics, including power factor, reactive power, current draw, and active power consumption. Beyond technical performance, the results reveal substantial reductions in operational costs, highlighting the practical feasibility of metaheuristic-based efficiency optimization in industrial motor systems.

The remainder of this paper is organized as follows: Section 2 presents the theoretical background and methodology, including the PID controller structure and the design of the GWO algorithm. Section 3 describes the simulation and experimental setup. Section 4 discusses the results and performance comparisons. Section 5 concludes with key findings and recommendations for future work.

2. MATERIAL AND METHOD

2.1. Materials

The primary equipment used in this study consisted of a Variable-Speed Drive (VSD) and a three-phase induction motor. The motor employed was a TECO-brand end-suction (centrifugal) unit with a rated power of 150 kW, a rated current of 285 A, and a nominal speed of 1500 rpm. To regulate its operation, a 220 kW VSD was utilized, providing precise control over motor performance. For measurement, a digital ammeter was used to record the current in real time. At the same time, a Schneider Human-Machine Interface (HMI) served as the monitoring platform for displaying motor parameters and validating measurements against both manual calculations and digital ammeter readings.

2.2. Methods

The proposed system integrates a Supervisory Control and Data Acquisition (SCADA) platform with the Variable Speed Drive (VSD) to control the three-phase motor. The optimization process is carried out by tuning the Proportional-Integral-Derivative (PID) controller parameters using the Grey Wolf Optimization (GWO) algorithm. The experimental procedure is summarized in the research methodology diagram presented in Figure 1.

https://doi.org/10.31436/iiumej.v27i1.3944

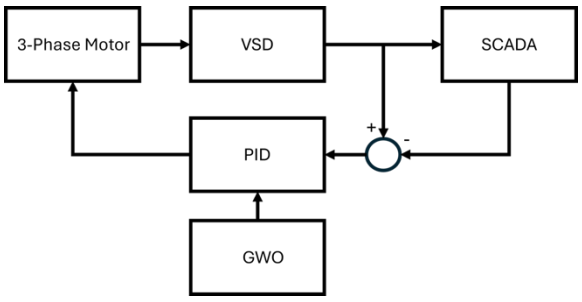


Figure 1. Proposed Research Methodology.

The methodology can be described in the following steps:

1. The system operates using a single three-phase induction motor connected to a Variable Speed Drive (VSD).
2. Each critical parameter, voltage, current, power factor ($cos\ \varphi$), and frequency, is measured using corresponding sensors integrated into the VSD. These sensors generate real-time signals, which are transmitted to the SCADA system via a data acquisition (DAQ) module and the Modbus TCP/IP protocol.
3. The SCADA system records and monitors these parameters continuously. The collected data are used for two primary purposes: (a) PID parameter tuning and (b) optimization of those PID parameters using the Grey Wolf Optimizer (GWO) algorithm.
4. During the optimization process, the SCADA system provides closed-loop feedback by sending the real-time measurements to the controller, which updates the control signals to the VSD. The feedback loops ensure that the PID and GWO algorithms receive accurate, instantaneous data for decision-making and adjustment.
5. The SCADA interface displays all measured and control parameters, including current, voltage, frequency, power factor, and motor switching status, enabling real-time supervision and performance analysis.
6. Finally, system performance is evaluated by comparing the motor's operational characteristics before and after applying the GWO-optimized PID control. The comparison is based on metrics such as steady-state error, rise time, and energy efficiency.

This experimental setup enables the systematic monitoring, control, and optimization of motor energy efficiency, ensuring that results are both accurate and practically applicable.

2.2.1. Working System

The working principle of this research is illustrated through the control block diagram presented in Figure 2. The diagram depicts a closed-loop control system employing a Proportional-Integral-Derivative (PID) controller optimized using the Grey Wolf Optimizer (GWO) algorithm. The objective of this system is to ensure that the output closely follows the desired reference signal with high accuracy and stability.

At the core of the system, the reference input $R(s)$ in the Laplace domain specifies the target system response. This signal is continuously compared with the actual output $C(s)$ at the summing junction, producing an error signal $e(s)$. The error, defined as the difference between the reference and the actual output, is processed by the PID controller to generate a corrective control signal.

The PID controller regulates the system behavior using three parameters: proportional gain Kp, integral gain Ki, and derivative gain Kd. The proportional term responds to the present

https://doi.org/10.31436/iiumej.v27i1.3944

error, the integral term eliminates steady-state errors by accounting for past deviations, and the derivative term anticipates future errors to improve stability. Instead of conventional manual tuning, this study employs the GWO algorithm to optimize these parameters automatically.

The Grey Wolf Optimizer (GWO) is a metaheuristic algorithm inspired by the hunting strategy and leadership hierarchy of grey wolves. In this research, GWO is applied to search for the optimal PID parameter set by minimizing performance indices, including Integral Square Error (ISE), Integral of Time-weighted Absolute Error (ITAE), settling time, and overshoot. This optimization ensures efficient tuning and improves system performance without relying on heuristic or trial-and-error methods.

The optimized control signal is then applied to the plant $P(s)$, which represents the physical system being controlled, in this case, a three-phase induction motor. The plant is modeled mathematically using its transfer function in the Laplace domain. The resulting system output $C(s)$ is fed back into the summing junction, thereby completing the closed-loop configuration.

The mathematical relations used in this study are expressed as follows:

Synchronous speed:

$$ns = \frac{(120\times f)}{p} \tag{1}$$

where f is the frequency and p is the number of poles.

Power in terms of torque and speed:

$$P = \frac{(T\ x\ ns)}{5252} \tag{2}$$

Electrical power relation:

$$P = V\ x\ I\ x \cos\varphi \tag{3}$$

where P is the active power, V is the RMS voltage, I is the RMS current, and $cos\ \varphi$ is the Power factor, which is the ratio of active power to apparent power. φ is the phase angle between current and voltage. The primary optimization target is the power factor $cos\ \varphi$. For a three-phase induction motor, the default power factor is typically around 0.8. By optimizing the PID controller parameters using GWO, the power factor is improved, thereby enhancing overall energy efficiency.

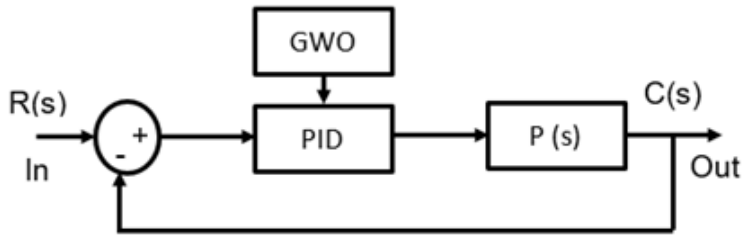


Figure 2. Control System Block Diagram.

In summary, the proposed system integrates classical control theory with modern optimization techniques to deliver a robust and efficient closed-loop controller. This integration is particularly advantageous for engineering applications that require high precision, rapid response, and system stability, such as electric motor drives, robotics, and industrial automation.

https://doi.org/10.31436/iiumej.v27i1.3944

2.2.2. Grey Wolf Optimizer (GWO)

The Grey Wolf Optimizer (GWO) was first introduced in [18] as a novel swarm intelligence (SI) algorithm. It has demonstrated competitive performance compared to other metaheuristic algorithms such as the Gravitational Search Algorithm (GSA) and Differential Evolution (DE). The GWO is inspired by the natural leadership hierarchy and cooperative hunting strategy of grey wolves (Canis lupus), apex predators in the animal kingdom. In the wild, wolves typically live in packs consisting of 5-12 individuals, maintaining a strict social hierarchy that governs their collective behavior.

The hierarchy consists of four levels. The alpha (α) wolves lead the pack and represent the best solution in the optimization process. The beta (β) wolves act as subordinate leaders, assisting the alpha in decision-making and strengthening the directives, while also offering alternative solutions. The delta (δ) wolves serve as scouts, sentinels, and hunters, supporting the alpha and beta wolves. Finally, the omega (ω) wolves occupy the lowest rank, representing the weakest solutions but playing a crucial role in maintaining diversity within the population.

The hierarchical structure of the grey wolf pack is illustrated in Figure 3. As shown, the alpha (α) wolf occupies the highest rank, symbolizing the best solution at a given iteration. The beta (β) wolves represent the second-best solutions, assisting the alpha in guiding the pack. The delta (δ) wolves, positioned below the beta, act as scouts, hunters, and protectors, contributing additional search diversity.

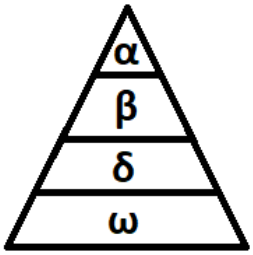


Figure 3. The hierarchical structure of GWO.

Finally, the omega (ω) wolves are located at the bottom of the hierarchy, corresponding to the weakest solutions. Despite their low ranking, omegas play a critical role in maintaining population balance and preventing premature convergence. This hierarchical model underpins the Grey Wolf Optimizer (GWO), in which leadership roles correspond directly to solution quality in the optimization process. During the optimization process, the hierarchy is mapped into the search mechanism as follows: the alpha (best), beta (second-best), and delta (third-best) wolves guide the population, whereas the omega wolves update their positions according to these leaders. This structure effectively balances exploration (global search) and exploitation (local refinement), mimicking the natural foraging behavior of wolves.

Mathematically, the encircling behavior is modeled as:

$$\vec{D} = \left|\vec{C}\,.\,\overrightarrow{Xp}(t) - \vec{X}(t)\right| \tag{4}$$

$$\vec{X}(t+1) = \vec{X}p(t) - \overrightarrow{A.D} \tag{5}$$

where t is the current iteration, X is the position vector of a grey wolf, and Xp denotes the prey position. The coefficient vectors A and C are calculated as:

$$\vec{A} = 2\vec{a}.\vec{r}_1 - \vec{a} \tag{6}$$

https://doi.org/10.31436/iiumej.v27i1.3944

$$\vec{C} = 2.\vec{r}_2 \tag{7}$$

Here, r_1 and r_2 are random vectors in the range [0,1], while a decreases linearly from 2 to 0 over the course of iterations, thereby controlling the balance between exploration and exploitation. The three best solutions (α, β, δ) are retained at each iteration, and the positions of the remaining wolves (ω) are updated according to their influence:

$$\begin{aligned} \vec{D}\alpha &= \left|\vec{C}1 - \vec{X}\alpha - \vec{X}\right| \\ \vec{D}\beta &= \left|\vec{C}2 - \vec{X}\beta - \vec{X}\right| \\ \vec{D}\delta &= \left|\vec{C}3 - \vec{X}\delta - \vec{X}\right| \end{aligned} \tag{8}$$

$$\begin{aligned} \vec{X}1 &= \vec{X}\alpha - \vec{A}1.\vec{X}\alpha \\ \vec{X}2 &= \vec{X}\beta - \vec{A}2.\vec{X}\beta \\ \vec{X}3 &= \vec{X}\delta - \vec{A}3.\vec{X}\delta \end{aligned} \tag{9}$$

$$X(t+1) = \frac{X1+X2+X3}{3} \tag{10}$$

This formulation enables GWO to simulate the collaborative hunting strategy of grey wolves, ensuring that the population converges toward optimal or near-optimal solutions while efficiently exploring the search space.

2.2.3. PID with Grey Wolf Optimization (GWO)

The Grey Wolf Optimization (GWO) algorithm is applied to fine-tune the PID controller parameters, thereby optimizing the power factor (*cos φ*) of a three-phase induction motor, thereby reducing unnecessary power consumption and improving overall energy efficiency. The optimization process is illustrated in Figure 4.

As shown in Figure 4, the proposed methodology follows a closed-loop optimization process that integrates the GWO algorithm with the PID controller through continuous feedback. The process begins with initializing the three-phase motor and assigning initial Proportional, Integral, and Derivative (PID) parameters. The GWO algorithm is then executed to estimate an optimal parameter set through its leadership hierarchy, where X_α (Alpha wolf) represents the best candidate solution, X_β (Beta wolf) the second-best, and X_δ (Delta wolf) the third-best, while the remaining wolves (X_ω) explore the search space to maintain population diversity.

Each candidate PID parameter set is applied to the control loop, and the resulting system performance, such as error magnitude, power factor (*cos φ*), and voltage stability, is measured in real time via the SCADA system. This feedback is returned to the GWO algorithm to evaluate the fitness of each candidate solution. Based on this feedback, GWO iteratively updates its search direction and parameter estimates over multiple iterations (up to 100 cycles) until the convergence criterion is met. Once the optimal PID gains are achieved, the motor operates under the optimized configuration. During operation, the SCADA interface continuously monitors and displays real-time voltage, current, frequency, and power factor (*cos φ*). When the system stabilizes, a higher cos φ value (e.g., 0.92) indicates improved power quality and enhanced energy utilization.

https://doi.org/10.31436/iiumej.v27i1.3944

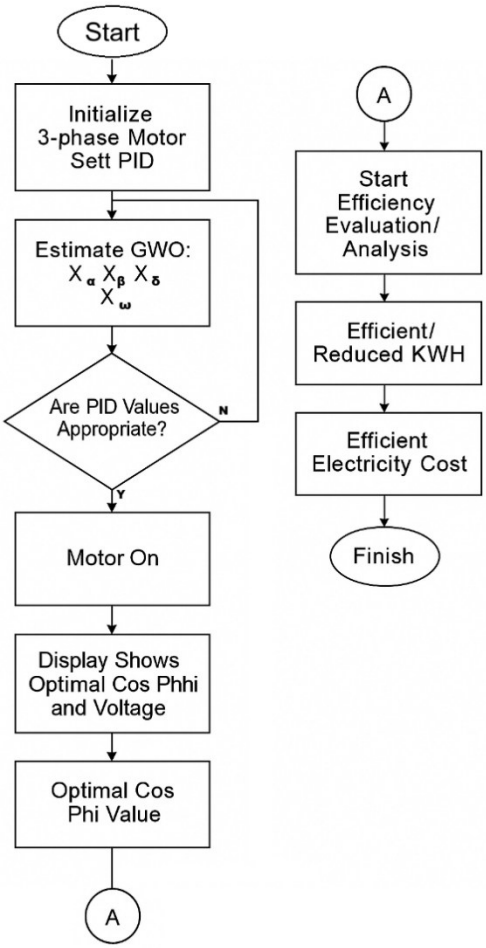


Figure 4. Optimization Flowchart Using Grey Wolf Optimization (GWO).

In the final stage, the system evaluates energy efficiency by comparing total power consumption (in kWh) before and after optimization, thereby quantifying the reduction achieved and the corresponding cost savings. This closed-loop GWO–PID integration confirms the feasibility of real-time adaptive tuning and demonstrates the potential scalability of this approach for multi-motor industrial systems.

2.2.4. Data Analysis Method

A quantitative analysis approach was employed to assess both the accuracy and performance of the proposed model. The evaluation process was carried out in two primary stages. First, model accuracy was verified by comparing the current measurements obtained from the digital ammeter (ampere clamp) with the values displayed on the Human-Machine Interface (HMI). This step ensured consistency between the measurement instruments and the system's monitoring display. From the theoretical perspective, the motor current can be calculated using the following equation:

$$I = \frac{P}{(V\, x \cos\varphi \times 1.73)} \tag{11}$$

Second, a power-loss analysis was conducted to quantify the energy-efficiency improvements. Manual calculations based on the power triangle formula were performed and compared with readings obtained directly from the ammeter clamp. Through this comparative assessment, the reliability of the proposed optimization method in reducing power loss and improving system efficiency was validated.

https://doi.org/10.31436/iiumej.v27i1.3944

2.2.5. Energy and Cost Computation

To evaluate the practical benefit of the proposed optimization, monthly energy consumption and corresponding costs were computed using the average motor power consumption. The monthly energy consumption E was calculated using Eq. (12).

$$E = P \times t \times d \tag{12}$$

where P is the average motor power consumption (kW), t represents the operating hours per day, and d denotes the number of operating days per month. The associated monthly electricity cost C was then estimated as:

$$C = E \times R \tag{13}$$

where R is the electricity rate (USD/kWh). The computation provides a consistent framework to quantify the energy efficiency and cost savings achieved by the proposed PID–GWO control strategy.

3. RESULTS AND DISCUSSION

This section presents the experimental results and the subsequent discussion. The analysis compares the initial power factor (cos φ) with the improved values obtained by integrating Proportional–Integral–Derivative (PID) control with Grey Wolf Optimization (GWO). The outcomes are illustrated both graphically and in tabular form to ensure clarity and reproducibility. All results were derived from field measurements using standardized instruments, theoretical calculations, and Human–Machine Interface (HMI) monitoring. Each experiment comprised 100 GWO iterations to determine the optimal PID parameter set.

The subsequent subsections provide a detailed comparison between baseline and optimized performance, followed by an analysis of the impact of GWO-based tuning on motor efficiency.

3.1. Power Factor (*cos φ*) Optimization

The initial power factor of the three-phase induction motor fluctuated between 0.80 and 0.87, indicating suboptimal operating conditions and a higher reactive power demand. After applying the PID controller tuned using Grey Wolf Optimization (GWO), the power factor improved substantially, stabilizing near the target value of 0.92. This improvement corresponds to a relative enhancement of approximately 6-12% compared to the baseline condition.

Figure 5 depicts the simulation of GWO convergence over 100 iterations. The optimized PID parameters enabled the power factor to consistently approach the target value of 0.92. In contrast, the non-optimized case (without PID tuning) remained fixed at 0.80, highlighting the limited capability to self-correct under standard operating conditions.

The results demonstrate that integrating GWO into PID tuning provides a robust mechanism for compensating reactive power and improving overall motor efficiency. The algorithm's ability to dynamically adjust the PID gains contributes to its rapid convergence toward the reference value. This enhancement reflects a more efficient utilization of electrical power, as the optimized controller reduces the grid's reactive power demand. The Grey Wolf Optimization algorithm successfully converged to an optimal set of PID parameters after 100 iterations, yielding Kp = 11.2, Ki = 11.0, and Kd = 3.5. The convergence behavior of these parameters is illustrated in Figure 6.

https://doi.org/10.31436/iiumej.v27i1.3944

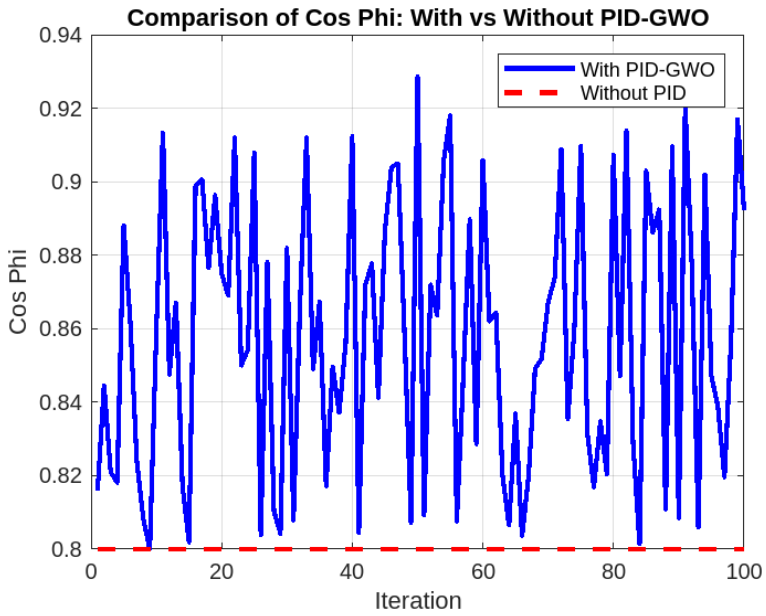


Figure 5. Cos phi with vs without PID GWO

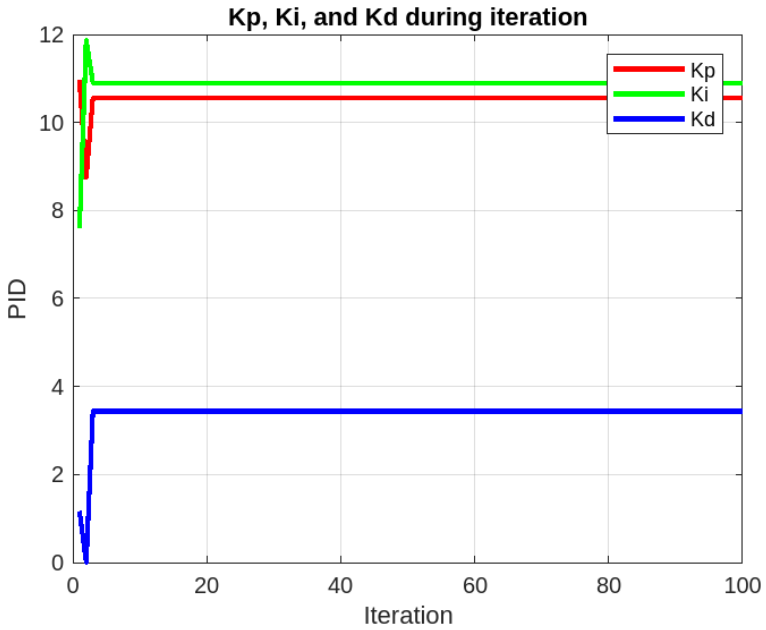


Figure 6. KP, KI, KD during iteration.

As shown in Figure 6, the three control gains experienced relatively large fluctuations during the early iterations, reflecting the exploration phase of GWO as it searches for the global optimum. After approximately 15 iterations, the parameters gradually stabilized, indicating a transition from exploration to exploitation. This steady convergence highlights GWO's ability to avoid local optima and ensure consistent tuning outcomes.

The PID-GWO configuration enabled the motor to maintain an average power factor close to the target value of 0.92. However, instantaneous values fluctuated between 0.80 and 0.93 during optimization, as illustrated in Figure 5. These fluctuations primarily reflect transient dynamics and the adaptive nature of the GWO-based tuning during load and iteration variations. Nevertheless, the optimized controller consistently improved the overall operating range of the power factor relative to the baseline (without PID), which remained nearly constant

https://doi.org/10.31436/iiumej.v27i1.3944

at approximately 0.80. From a practical standpoint, this demonstrates that the proposed PID-GWO control effectively minimizes reactive power demand and enhances system efficiency under dynamic operating conditions.

Following the determination of the optimal PID parameters, the next step was to implement them in practice on the Variable Speed Drive (VSD) via its Human–Machine Interface (HMI). The identified parameter values (Kp, Ki, and Kd) were manually entered into the HMI interface, as shown in Figure 7. This interface provides a direct means of adjusting the control gains that regulate key operational variables, including motor speed, pressure, and system stability. After parameter entry, measurements were conducted using standard instruments and theoretical validation. This step was crucial for verifying the accuracy of the optimized PID configuration obtained from the GWO algorithm.

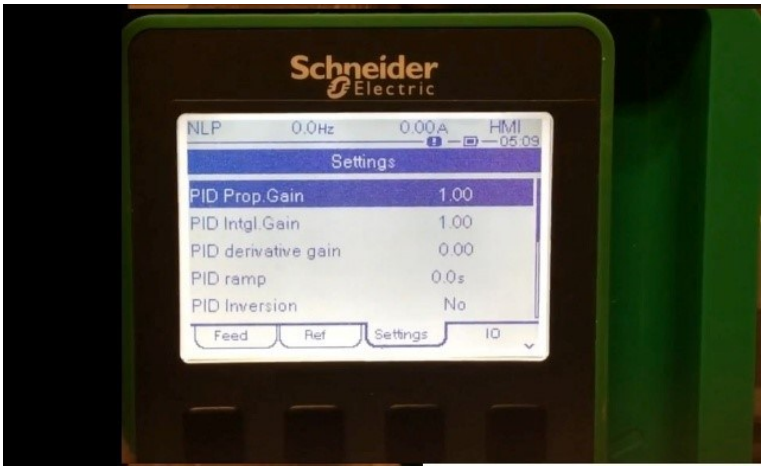


Figure 7. PID Control parameters

Figure 8. A Real-time Cos phi Measurement after optimization

The results of the field implementation are presented in Figure 8. The optimized controller increased the power factor from approximately 0.82 to a stable 0.92, closely matching the target set during the optimization phase. This outcome not only validates the simulation results but also confirms the capability of the proposed PID–GWO approach to deliver consistent performance under real operating conditions.

https://doi.org/10.31436/iiumej.v27i1.3944

From an industrial standpoint, achieving a power factor of 0.92 represents a substantial improvement in energy efficiency, as it reduces reactive power demand and lowers operational costs. These findings highlight the practical feasibility of integrating metaheuristic optimization into existing control infrastructures, thereby bridging the gap between theoretical research and applied engineering practice.

3.2. Current Reduction

The implementation of the optimized PID controller resulted in a significant reduction in motor current. Initially, the pump motor was operating at 285.21 A as depicted in Figure 9. The figure presents the real-time simulation display using the available Human-Machine Interface (HMI). The interface inherently includes control panels and monitoring tabs that visualize both the control process and the current waveform simultaneously. This integrated view is intentionally maintained to reflect the actual implementation of the proposed PID-GWO method in the experimental environment. The current waveform indicates improved system response and stability following optimization. After optimization, the motor current was reduced to approximately 164.2 A (Figure 10-12).

This reduction directly decreases the I^2R copper losses, thereby improving overall thermal efficiency, reducing heat stress on the windings, and extending the equipment's operational lifespan.

Figure 9. Current without PID GWO

Figure 10. Current with PID GWO

https://doi.org/10.31436/iiumej.v27i1.3944

Figure 11. Current with PID GWO (clamp ammeter)

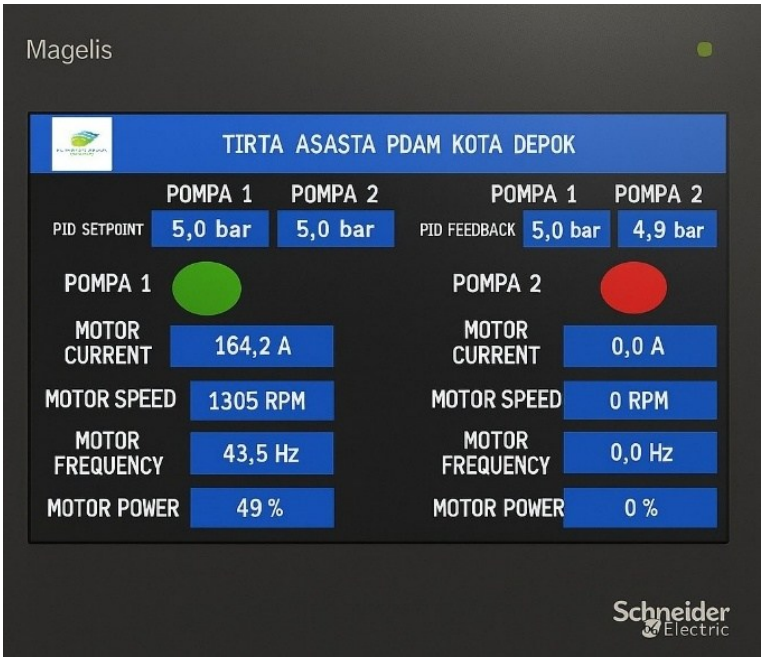


Figure 12. Current with PID GWO (HMI)

The verification of the motor current was performed using four independent measurement approaches: Human Machine Interface (HMI), Supervisory Control and Data Acquisition (SCADA) system, clamp meter readings, and the theoretical reference obtained from (11). All four approaches consistently indicated a motor current of approximately 164.1 A, confirming the accuracy and synchronization of both the monitoring and control systems. The agreement among these methods validates the reliability of the data acquisition infrastructure and demonstrates the effectiveness of the PID-GWO optimization in maintaining current stability under varying load conditions.

In addition to validating the current reduction, the HMI and SCADA interfaces provide real-time data regarding motor speed, frequency, and power, while the clamp meter

https://doi.org/10.31436/iiumej.v27i1.3944

measurement offers manual confirmation. The strong correlation between all four methods ensures the reliability of the monitoring infrastructure and validates the effectiveness of the PID optimization.

3.3. Active Power Reduction

The motor's active power consumption experienced a significant reduction, decreasing from 150 kW to 84.5 kW, as summarized in Table 1. This improvement was achieved by lowering the operating frequency from 50 Hz to 43.5 Hz, thereby aligning the pump operation with the actual load demand rather than maintaining the nominal rated condition.

This theoretical result shows an excellent agreement with the Modbus TCP/IP-recorded value of 84.3 kW, confirming the validity of both the measurement and the monitoring system. The consistency between the analytical estimation and SCADA observation demonstrates the reliability of the instrumentation in capturing real-time pump operating conditions.

The reduction in active power underscores the direct impact of optimized PID-based control, which not only lowers electrical demand but also improves energy efficiency and reduces operational costs. Such optimization strategies are particularly relevant for water treatment and distribution systems, where pumps typically account for the majority of total energy consumption.

Table 1. Power comparison with and without PID GWO

Frequency (Hz)	Power without PID-GWO (kW)	Power with PID-GWO (kW)
45.4	150	91.03
44.9	150	88.30
43.5	150	84.54
43.0	150	81.29
42.0	150	77.72
41.4	150	75.49
40.0	150	70.53
39.0	150	67.05

3.4. Energy Efficiency Analysis

Energy efficiency was evaluated as the ratio between the motor's output power and the corresponding input power. Following the optimization process, the efficiency improved substantially, reaching 44.77%.

Table 2 provides a detailed comparison of the motor current under three configurations: (i) without control, (ii) with conventional PID, and (iii) with PID tuned using the GWO algorithm. An additional column presents the Theoretical Current (A), calculated from the electromagnetic torque-supply voltage relationship described in (11). The results indicate that the measured current values obtained using the proposed PID–GWO controller closely follow the theoretical predictions across all frequency ranges, with only minor deviations (<0.2%) attributed to measurement noise and inverter nonlinearities. In contrast, the system without control consistently draws higher current, confirming that the PID–GWO approach effectively minimizes motor current and enhances overall system efficiency. This consistency between the experimental and theoretical results further validates the robustness and accuracy of the proposed control method.

https://doi.org/10.31436/iiumej.v27i1.3944

Table 2. Current comparisons with vs without PID GWO

Frequency (Hz)	Current without PID GWO (A)	Current with PID GWO (A)	Current with PID GWO (Amperemeter) (A)	Theoretical Current (A)
45.40	285.21	169.40	169.40	169.28
44.90	285.21	166.15	166.15	166.15
43.50	285.21	164.20	164.20	161.54
43.00	285.21	160.70	160.40	160.33
42.00	285.21	157.12	157.00	157.12
41.40	285.21	154.88	154.70	154.37
40.00	285.21	149.64	149.60	149.15
39.00	285.21	145.90	145.90	145.42

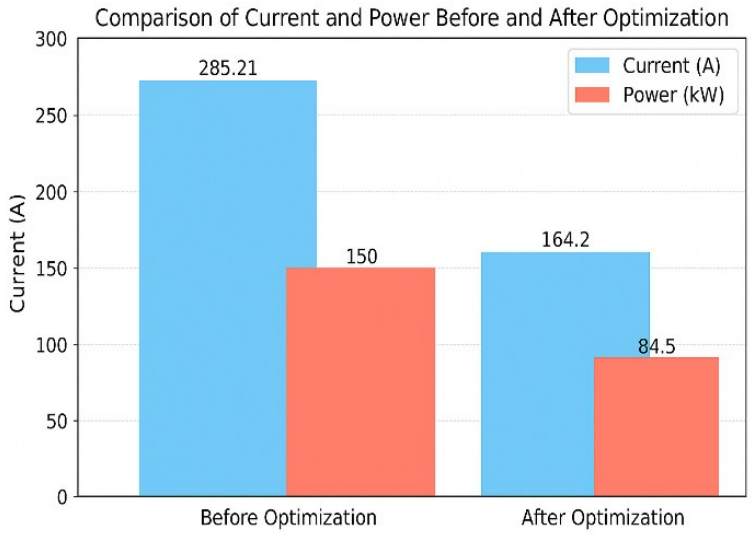


Figure 13. Comparison of current and power before and after Optimization

Figure 13 further illustrates the contrast in electrical performance before and after optimization. Before optimization, the motor consumed 285.21 A and delivered 150 kW, indicating excessive energy consumption under unoptimized conditions. After optimization, the current dropped significantly to 164.2 A and the active power to 84.5 kW, corresponding to reductions of approximately 42.4% in current and 44.7% in active power.

This marked reduction demonstrates that the PID-GWO optimization strategy not only enhances energy efficiency but also improves system reliability by reducing thermal stress on motor windings and lowering reactive power demand. Moreover, the resulting energy savings translate into reduced operational costs and a smaller environmental footprint, both of which are critical considerations in sustainable industrial practices.

3.5. Cost Analysis

In addition to the technical performance improvements, the energy efficiency and cost benefits of the proposed PID–GWO optimization strategy were evaluated under identical operating conditions. The computation results summarized in Table 3 show that the optimization significantly reduced the average power consumption from 150 kW to 84.5 kW.

This reduction corresponds to a monthly energy savings of 47,016 kWh, representing an approximately 43.5% improvement in energy efficiency. Assuming an average industrial

electricity rate of USD 0.10 per kWh, the monthly operating cost decreased from USD 10,800 to USD 6,098, resulting in an estimated monthly cost savings of USD 4,702.

Table 3. Monthly Energy Consumption and Cost Analysis

Condition	Average Power (kW)	Monthly Energy (kWh)	Cost (USD 0.10/kWh)	Savings (%)
Before Optimization	150	108,000	10,800	-
Afer Optimization	84.5	60,984	6,098	43.5
Improvement	-	47,016	4,702	43.5

These findings confirm that the proposed control method not only enhances power quality and system stability but also yields substantial operational savings, resulting in a rapid return on investment (ROI). Moreover, the observed energy reduction contributes directly to sustainability goals by lowering overall energy demand and environmental impact.

3.6. Discussion

The experimental results presented in this study provide comprehensive evidence of the novel contribution and effectiveness of the proposed PID–GWO optimization approach in enhancing motor performance, improving energy efficiency, and reducing operational costs.

Unlike conventional PID controllers that rely on manual or heuristic tuning, the proposed method employs the Grey Wolf Optimizer (GWO) to automatically identify optimal proportional, integral, and derivative gains. This adaptive optimization process enables the controller to balance system stability and energy efficiency simultaneously, representing a significant improvement over traditional tuning methods.

As summarized in Table 2, the motor current decreased substantially after optimization. Without control, the system consumed approximately 285.21 A, whereas with the PID–GWO optimization, the current dropped to around 164.2 A, corresponding to a 42.4% reduction. This reduction directly correlates with improved power utilization and reduced electrical stress on motor components. Theoretical calculations align closely with the experimental data, confirming the accuracy and reliability of the optimized control design.

Energy efficiency, defined as the ratio of output to input power, also improved significantly. Before optimization, the motor operated at approximately 150 kW; after applying the PID–GWO controller, power consumption decreased to 84.5 kW, representing an energy-efficiency improvement of 44.77%. This improvement demonstrates that the optimized PID gains not only minimize transient deviations but also reduce continuous power losses, thereby stabilizing system performance under various load conditions.

The cost analysis further reinforces the practical relevance of the proposed optimization method. As presented in Table 5, the monthly electricity cost decreased from USD 10,800 to USD 6,098, yielding a monthly saving of USD 4,702, or approximately 43.5% reduction. This significant decrease in operational expenditure illustrates the dual benefit of the proposed PID–GWO scheme, delivering both technical and economic advantages suitable for real-world industrial environments.

These findings confirm that integrating GWO into PID control constitutes a novel hybrid optimization strategy capable of addressing multiple industrial challenges concurrently: reducing energy consumption, minimizing operational costs, and improving overall system

stability. The demonstrated improvements are consistent with global sustainability objectives and energy efficiency directives, underscoring the method's potential for large-scale industrial adoption. Although the proposed approach yields substantial benefits, some limitations remain. This study primarily focused on steady-state operation. Future research should investigate dynamic load variations, fault tolerance, and scalability to different motor types and power ratings. Moreover, extending the optimization framework through hybridization with adaptive or learning-based algorithms could further enhance robustness and adaptability under fluctuating operating conditions.

In summary, the combined technical and economic evaluations clearly establish that the PID–GWO optimization introduces a novel, efficient, and practical solution for industrial motor energy management, bridging the gap between classical PID control and modern intelligent optimization techniques.

4. CONCLUSION

This study demonstrated the effectiveness of integrating PID control with Grey Wolf Optimization (PID-GWO) to enhance the performance and energy efficiency of a three-phase induction motor. The proposed method significantly reduced motor current by approximately 42.4% and improved energy efficiency by 44.77%. Additionally, the optimization yielded substantial financial benefits, reducing monthly electricity costs by nearly 43.5%. These results confirm that the PID-GWO approach is not only technically robust but also economically viable for industrial applications. Beyond cost and energy savings, the findings suggest broader implications for sustainable industrial operations, including reduced environmental impact through optimized energy consumption. Future research should focus on extending this approach to dynamic operating conditions, fault-tolerant systems, and hybrid optimization methods to further improve adaptability and scalability.

ACKNOWLEDGEMENT

This research was funded by the *Master Thesis Research Grant Scheme (Hibah Skema Penelitian Tesis Magister)* from the Ministry of Education, Higher Education, Science, and Technology of the Republic of Indonesia, under the fiscal year 2025. The authors would like to express their sincere gratitude for this financial support, which made the completion of this study possible.

REFERENCES

[1] Zherdev PI, Biksaleev RS, and Karpukhin KE. (2025) Trends in the development of modern electric motors: challenges, difficulties and results, Electrotechnical Facilities and Systems, 19(1):422-429.

[2] Vasileios IV, Georgios KS, Fotios PX, Maria SP, Themistoklis DK, Marina AT, and Antonios GK. (2024) Overview on Permanent Magnet Motor Trends and Developments, Energies 17(2).

[3] Azab M. (2025). A Review of Recent Trends in High-Efficiency Induction Motor Drives. Vehicles, 7(1).

[4] de Swardt H. (2025). Increasing The Efficiency of Premium Efficiency Motors: Practical Case Studies. IEEE Transactions on Industry Applications, 1-8

[5] Moussa H. Krim S, Kesraoui H, Mansouri M, and Mimouni MF. (2024) Loss Model Control for Efficiency Optimization and Advanced Sliding Mode Controllers with Chattering Attenuation for Five-Phase Induction Motor Drive. Energies, 17(16).

[6] Goswami G and Das S. (2024) Inherently Robust Loss Model Controller for Energy Efficient Operation of Indirect Rotor Field Oriented Induction Motor Drives. IEEE Transactions on Power Electronics, 39(11):14951-14960.

[7] Nicolás-Martín C, Montilla-DJesus, ME, Santos-Martín, D, and Martínez-Crespo J. (2025) Computationally Efficient and Loss-Minimizing MPC for Induction Motors. Energies, 18(6).

[8] Xin X, Zhang Z, Zhou Y, Liu Y, Wang D, and Nan S. (2024) A comprehensive review of predictive control strategies in heating, ventilation, and air-conditioning (HVAC): Model-free VS model. Journal of Building Engineering, 94:110013.

[9] Dash R, Reddy KJ, Mohapatra B, Bajaj M & Zaitsev I. (2025) An approach for load frequency control enhancement in two-area hydro-wind power systems using LSTM + GA-PID controller with augmented lagrangian methods. Scientific Reports volume, 15:1307.

[10] Qu Z, Younis W, Liu X, Junejo AK, Almutairi SZ, and Wang P. (2024). Optimized PID Controller for Load Frequency Control in Multi-Source and Dual-Area Power Systems Using PSO and GA Algorithms. IEEE Access, 12:186658-186678.

[11] Zhang, X. and Yang, Y. (2024). Optimization of PID controller parameters using a hybrid PSO algorithm. International Journal of Dynamics and Control, 12:3617–3627.

[12] Aner, E.A., Awad, M.I. & Shehata, O.M. (2024). Performance evaluation of PSO-PID and PSO-FLC for continuum robot's developed modeling and control. Scientific Reports, 14:733.

[13] Pairan MF, Shamsudin SS, and Yaakub MF. (2024). Autotuning PID Controllers for Quadplane Hybrid UAV using Differential Evolution Algorithm. Journal of Aeronautics, Astronautics and Aviation, 56(1S):341–356.

[14] Moloody A, As'arry A, Hong TS, Kamil R, Zolfagharian A. (2025). PID Controller Parameter Tuning Based on a Modified Differential Evolutionary Optimization Algorithm for the Intelligent Active Vibration Control of a Combined Single Link Robotics Flexible Manipulator. Journal of Advanced Research in Applied Sciences and Engineering Technology, 52(1):234-258.

[15] Zhang Y and Lu Q. (2024). Fractional Order PID Control Based on Improved Whale Optimization Algorithm. IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 1676-1682.

[16] Xie B, Zhou Q, and Yang J. (2024). Improving the Application of Whale Optimization Algorithm in PID Parameter Optimization for Linear Motor Control System of Wire Bonding Machine XY Motion Platform. 25th International Conference on Electronic Packaging Technology (ICEPT), 1-5.

[17] Suwoyo H, Hajar, MHI, Indriyanti P, and Febriandirza A. (2024) The use of Fuzzy Logic Controller and Artificial Bee Colony for optimizing adaptive SVSF in robot localization algorithm. Sinergi (Indonesia), 28(2):231-240.

[18] Benamara K, Amimeur H, Hamoudi Y, Abdolrasol MGM, Cali U, and Ustun TH, (2024) Grey Wolf Optimization for enhanced performance in wind power system with dual-star induction generators. Frontiers in Energy Research, 12:1421336.

[19] Suwoyo H, Adriansyah A, Andika J, Shamudin AU, and Tian YZ, (2025) An Effective Way for Repositioning the Beacon Nodes of Fast RRT Results Utilizing Grey Wolf Optimization. Journal of Robotics and Control (JRC), 6(1):272-284.

[20] Sahoo AK and Jena RK. (2023) Loss model based controller of fuzzy DTC driven induction motor for electric vehicles using optimal stator flux. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 6:100304.

https://doi.org/10.31436/iiumej.v27i1.4009

Development of a Model for Malware Detection and Classification at the Byte Level Based on Transformer

RULOF BALTWIN TALLANE[1], AKHMAD UNGGUL PRIANTORO[1*], RIZA MUHIDA[2]

[1] *Faculty of Information and Technology, Budi Luhur University, Jakarta, Indonesia*
[2]*Faculty of Engineering, Bandar Lampung University, Lampung, Indonesia*

**Corresponding author: akhmad.unggul@budiluhur.ac.id*

(Received: 24 October 2025; Accepted: 20 December 2025; Published online: 12 January 2026)

ABSTRACT: Malware threats have been a critical concern in cybersecurity, particularly due to the increasing complexity and constantly evolving variants that are difficult to detect using conventional signature-based or static rule-based methods. This research focused on developing a Transformer-based model at the byte level to detect and classify malware effectively and adaptively, thereby streamlining the analytical process without requiring specialized feature tokenization. The primary objective was to design and evaluate a Transformer model that captures universal and adaptive malware patterns directly from raw byte representations, enabling cross-platform applicability. A quantitative experimental approach was employed using three public datasets: Malware Detection PE-Based Analysis, MC-dataset-binary, and Malware.zip. Data processing involved byte embedding, dilated 1D convolution, multi-head self-attention, and attention pooling. Model optimization was conducted using AdamW with a combined scheduler, Stochastic Weight Averaging (SWA), random byte masking augmentation, and Mixup Embedding. Experimental results showed that the byte-level Transformer model achieved high classification accuracy across the three datasets, namely 99%, 92%, and 94%, respectively. These results demonstrate that a byte-level Transformer can effectively capture universal malware patterns in binary data, offering a flexible and highly accurate approach to developing resilient defenses against modern cyber threats.

ABSTRAK: Ancaman perisian perosak (malware) merupakan isu kritikal dalam keselamatan siber, terutama apabila disebabkan oleh peningkatan kerumitan dan variasi yang sentiasa berevolusi, sekaligus menyukarkan pengesanan menggunakan kaedah konvensional berasaskan tandatangan atau peraturan statik. Kajian ini memfokuskan kepada pembangunan model berasaskan Transformasi pada peringkat bait (byte-level) bagi mengesan dan mengklasifikasikan perisian perosak secara berkesan dan adaptif, tanpa memerlukan pengekstrakan atau penandaan ciri khusus. Objektif utama kajian adalah bagi mereka bentuk dan menilai keberkesanan model Transformasi yang mampu menangkap corak perisian perosak bersifat universal dan adaptif secara langsung daripada representasi bait mentah, seterusnya membolehkan kebolehgunaan merentas platform. Pendekatan eksperimen kuantitatif digunakan dengan melibatkan tiga set data awam, iaitu Analisis Berdasarkan Pengesanan Perisian Perosak PE, Set Data Binari MC, dan Malware.zip. Pemprosesan data merangkumi pembenaman bait, konvolusi 1D terlarut (dilated), perhatian kendiri berbilang kepala, serta pengumpulan berasaskan perhatian. Pengoptimuman model dilaksanakan menggunakan AdamW dengan penjadual gabungan, Purata Pemberat Stokastik (SWA), penambahan data melalui penyamaran bait rawak, dan Penggabungan Embedding. Dapatan kajian melalui eksperimen menunjukkan bahawa model Transformasi peringkat bait mencapai ketepatan pengelasan tertinggi, iaitu masing-masing 99%, 92%, dan 94% bagi ketiga-tiga set data. Dapatan ini membuktikan bahawa Transformasi peringkat bait berupaya

menangkap corak perisian perosak universal daripada data binari, sekaligus menawarkan pendekatan fleksibel dan berketepatan tinggi bagi membangunkan pertahanan yang lebih berdaya tahan terhadap ancaman siber moden.

KEYWORDS: *Malware, Byte Level, Transformer, Machine Learning, Cybersecurity*

1. INTRODUCTION

The rapid advancement of information technology has significantly transformed various aspects of modern society. However, this progress has also introduced substantial challenges in the field of cybersecurity. Among the most critical threats is malware, a type of malicious software designed to disrupt, damage, or gain unauthorized access to computer systems. The severity of malware threats lies in their potential to cause serious harm to both individuals and organizations. According to the Indonesia Cybersecurity Landscape Report 2024 published by Badan Siber dan Sandi Negara (BSSN - the Indonesian National Cyber and Crypto Agency), malware remains one of the top five reported incident categories, alongside data breaches (39.1%), traffic anomalies (33.3%), malicious activity (6.3%), and trojans (5.6%). These highlight malware as a persistent and pressing issue within the national cybersecurity landscape, demanding effective mitigation strategies to safeguard digital infrastructures [1].

In recent years, malware detection has shifted from conventional rule-based or signature-based approaches to machine-learning-driven methods. This transformation represents not merely a technological trend but also a strategic response to the increasingly complex and dynamic nature of modern cyber threats. Contemporary malware variants employ sophisticated evasion techniques such as obfuscation, packing, and polymorphism, making them difficult to detect with traditional methods. Static rule-based detection and simple heuristic analysis often fall short in addressing zero-day malware and novel attack vectors. Consequently, machine learning has emerged as a promising approach for developing adaptive, data-driven detection systems.

Machine learning enables models to learn directly from data rather than relying solely on handcrafted rules. By leveraging various classification and representation learning techniques, machine learning models can generalize beyond previously known patterns, thereby detecting new and unseen malware variants. In the context of malware detection, two major approaches exist: static analysis, which extracts structural features such as byte n-grams, opcodes, or file headers without executing the malware [2] and dynamic analysis, which observes the behavior of suspicious files in a controlled environment, such as API calls, registry modifications, or network connections [3]. Both approaches generate feature sets that can subsequently be processed using machine learning algorithms for malware classification.

Despite their effectiveness, conventional machine learning methods often depend on feature engineering and tokenization, which increase analytical complexity and reduce model flexibility. To address these limitations, this research proposes a byte-level Transformer-based approach for malware detection and classification. Transformers, powered by the multi-head self-attention mechanism, are capable of modeling contextual dependencies across data sequences [4]. Operating directly at the byte level eliminates the need for explicit feature tokenization, such as opcodes or API calls, thereby simplifying the analytical pipeline. Moreover, byte-level representation offers better generalization across different file types and operating systems, making the approach inherently more adaptive and platform-independent.

The main objective of this study was to design and evaluate a byte-level Transformer model that effectively detects and classifies malware variants, including zero-day attacks, while

https://doi.org/10.31436/iiumej.v27i1.4009

reducing reliance on domain-specific feature extraction. This approach is expected to enhance detection accuracy, improve model adaptability, and provide practical cross-platform applicability for modern cybersecurity systems.

2. RELATED WORKS

This section presents the theoretical foundations and related works that underpin this research. The discussion covers deep learning, the Transformer architecture, and the fundamentals of malware, including malware classification. Furthermore, it reviews prior studies on malware detection using byte-level and Transformer-based approaches, followed by the study's conceptual framework and hypothesis.

2.1. Deep learning

Deep learning is a subfield of machine learning that utilizes multi-layered artificial neural networks to extract hierarchical representations from raw data automatically [5]. Each layer in the network transforms the input into increasingly abstract features, enabling the model to capture complex and nonlinear relationships that conventional machine learning often fails to address effectively. Unlike traditional approaches that rely heavily on handcrafted features, deep learning learns directly from large volumes of data, thereby enhancing its ability to generalize to unseen inputs.

Deep learning approaches can broadly be classified into four categories. Supervised learning encompasses methods such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). CNNs are widely used in computer vision tasks due to their ability to capture spatial patterns. In contrast, RNNs are effective for processing sequential data, such as natural language and time series. Unsupervised learning encompasses Auto-Encoders, Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), and Generative Adversarial Networks (GANs), which are designed to uncover latent structures in unlabeled data. Reinforcement learning is generally divided into value-based methods, such as Deep Q-Learning, and policy-based methods, such as policy gradient algorithms, which have shown success in robotics and autonomous systems. Finally, hybrid models integrate elements from these categories to combine their strengths, enabling more robust and adaptive solutions to real-world challenges [6].

2.2. Transformer Architecture

The Transformer is a deep neural network architecture designed for efficient sequence modeling. It replaces recurrence and convolution with a self-attention mechanism, enabling parallel computation and effective modeling of both short- and long-range dependencies [7]. In our context, this capability is critical because malware binaries often contain non-contiguous and obfuscated byte patterns that require global context to detect reliably [8] .

https://doi.org/10.31436/iiumej.v27i1.4009

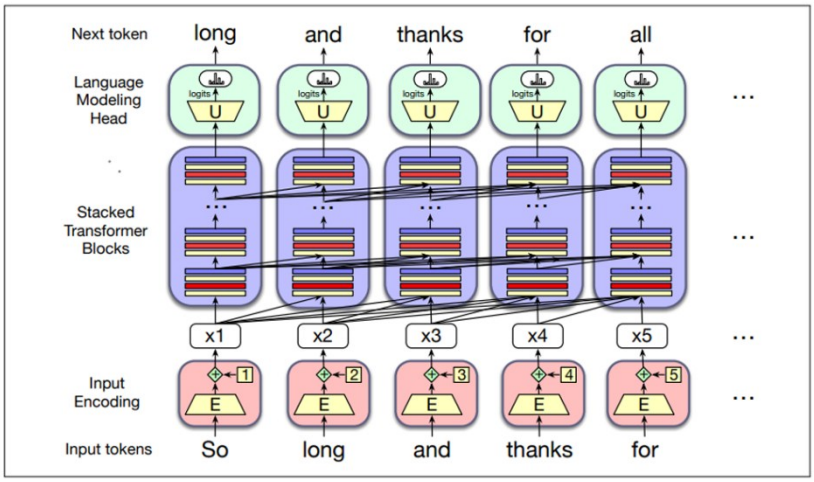


Figure 1. Transformer architecture [7]

As shown in Figure 1, the standard architecture comprises an encoder that maps the input sequence to contextual representations and a task head that uses these representations for prediction. For sequence-to-sequence tasks, the head is a decoder for malware classification at the byte level; it is typically a classification head attached to the encoder outputs. The stack includes input embeddings plus positional encodings, repeated layers of multi-head self-attention and position-wise feedforward networks, with residual connections and layer normalization to ensure stable optimization.

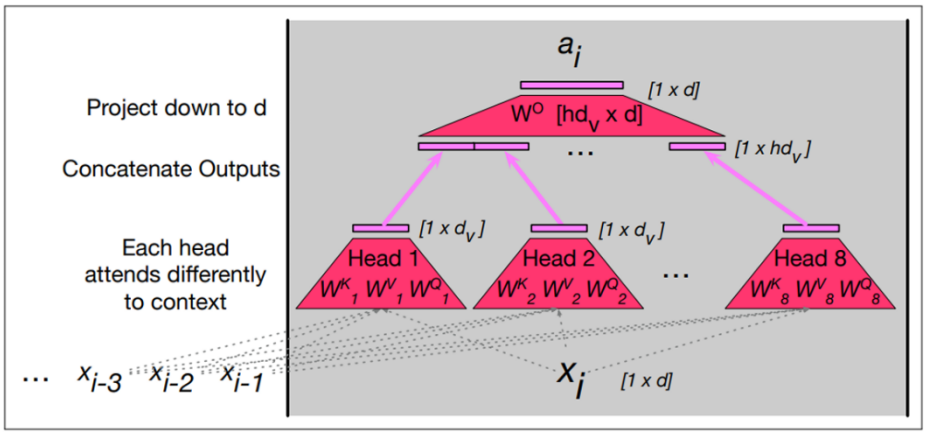


Figure 2. Multi-head attention mechanism [7]

The multi-head attention mechanism, as shown in Figure 2, enables the model to attend to different types of relationships in parallel, including local byte n-grams, long-range dependencies, and structural cues. This process produces richer contextual embeddings than a single attention head. This parallelism is particularly valuable for polymorphic/metamorphic malware, where discriminative evidence is dispersed across distant byte positions. Empirically, recent cybersecurity studies report that Transformer variants can learn directly from raw byte representations and improve detection accuracy and robustness [9-11].

A concise summary of the main components and their roles is provided in Table 1. Together, these components enable the proposed byte-level pipeline to capture global context without handcrafted tokenization, thereby aligning with our goal of building an adaptive, cross-platform detector.

Table 1. Key components of the Transformer architecture.

Component	Function
Input & Positional Embeddings	Map bytes to dense vectors and inject sequence order information.
Multi-Head Attention	Capture diverse contextual relationships across tokens in parallel.
Feedforward Network	Transform attention outputs through non-linear projections.
Residual Connections	Preserve information flow and stabilize training.
Layer Normalization	Improve numerical stability and accelerate convergence.

2.3. Malware

Malware, or malicious software, is among the most persistent and evolving threats in cybersecurity. Its primary objective is to compromise the confidentiality, integrity, and availability of digital systems, leading to severe consequences such as large-scale data breaches, financial losses, and the disruption of critical infrastructures [12]. Early generations of malware were relatively simple, typically relying on signature-based mechanisms such as viruses and worms. In contrast, modern variants have become increasingly sophisticated by adopting advanced evasion strategies. Polymorphism enables malware to alter its structure, thereby rendering signature-based detection ineffective continuously. Metamorphism rewrites malicious code into structurally distinct but functionally equivalent forms, complicating static inspection. In addition, obfuscation techniques deliberately conceal the program's logic flow to hinder both automated detection and manual analysis [13].

The growing complexity of digital ecosystems further exacerbates the challenge. Mobile platforms are frequently targeted due to the prevalence of third-party applications and insufficient vetting processes. Similarly, Internet of Things (IoT) devices, often deployed with limited computational resources and lacking regular security updates, significantly expand the attack surface for large-scale propagation. Cloud infrastructures introduce additional risks by centralizing sensitive data in shared environments, where a single compromise may cascade into systemic vulnerabilities [14]. Beyond these vectors, zero-day malware has emerged as a particularly critical concern. Zero-day attacks exploit previously unknown vulnerabilities, leaving defenders without available patches or signatures. Consequently, it is extremely difficult to detect them using traditional methods, underscoring the necessity of adaptive detection models capable of generalizing across unseen threats [15].

Malware can also be categorized according to its primary intent. Ransomware encrypts files and demands payment for decryption; spyware surreptitiously monitors and exfiltrates sensitive information, whereas botnets aggregate compromised devices into large-scale networks for coordinated attacks such as Distributed Denial-of-Service (DDoS) attacks. These diverse objectives highlight the breadth of modern malware threats and reinforce the importance of robust detection strategies that extend beyond static inspection [16].

3. RESEARCH METHODS

3.1. Research Approach

This study adopted a quantitative experimental approach to develop and evaluate a byte-level Transformer model for malware detection and classification. This approach was selected

because it enables objective measurement of model performance using numerical evaluation metrics. The research begins by identifying limitations in conventional malware detection methods, particularly rule-based and feature-engineering approaches, which struggle to address complex and evolving threats. To address these challenges, this study designs and trains a Transformer model directly on raw byte representations, thereby eliminating the need for handcrafted feature tokenization. The use of public datasets ensures transparency, reproducibility, and comparability with prior research, while rigorous evaluation using accuracy, precision, recall, and F1-score provides a comprehensive assessment of the model's effectiveness.

3.2. Sampling Method and Datasets

A purposive sampling strategy was employed to select datasets that contain both malware and benign files in raw binary format. This is essential for the proposed byte-level approach, which requires unprocessed data. Three benchmark datasets were employed:

- Malware Detection PE-Based Analysis [17]: 9,970 samples, including 8,970 malware and 1,000 benign files.
- MC-dataset-binary [18]: 11,480 samples, equally split between 5,740 malware and 5,740 benign files.
- Malware.zip [19]: 9,550 malware samples categorized into 12 malware families.

These datasets were selected to enable cross-dataset testing, thereby validating the generalization capability of the proposed model across diverse sources and malware categories.

3.3. Proposed Model Architecture

The proposed malware detection system is implemented as a Byte-Level Transformer that processes raw binary files without requiring handcrafted feature extraction. The model pipeline comprises several sequential components, each transforming raw byte sequences into discriminative features for classification. Figure 3 presents a schematic of the proposed model architecture, outlining the processing flow from raw binary input to the final classification output.

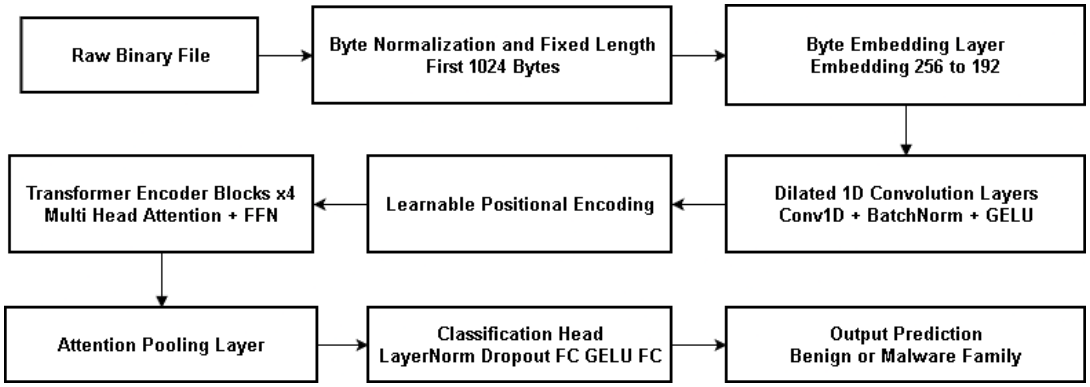


Figure 3. Architecture of the proposed byte-level transformer model for malware detection.

3.3.1. Input Preprocessing (byte normalization)

All executable samples, whether benign or malicious, are normalized to a fixed sequence length of 1,024 bytes (1 KB). Files shorter than this length are zero-padded, while longer files are truncated. Each byte (0–255) is treated as a discrete token, ensuring uniform tensor dimensions and efficient mini-batch processing on the GPU. The selection of 1,024 bytes is

https://doi.org/10.31436/iiumej.v27i1.4009

informed by prior studies on malware at the byte level, which use 1 KB blocks as a practical balance between information coverage and computational efficiency [20]. Following this rationale, we initially evaluated a sampling strategy that extracted byte segments from multiple 1 KB regions across each file. However, this approach yielded lower classification performance in our Transformer-based model. We attribute this behavior to disrupted sequential continuity, which is essential for attention-based architectures to learn meaningful contextual dependencies. Consequently, this study uses the first 1,024 bytes from each binary file. This choice preserves a contiguous byte sequence and yields more stable, consistent performance across all evaluated datasets.

3.3.2. Byte Embedding Layer

A trainable embedding matrix is implemented using the PyTorch module nn.Embedding(256, d), where d = 192 represents the embedding dimension. Each raw byte value is mapped into a 192-dimensional dense vector. This representation allows the model to learn semantic relationships among byte values, similar to word embeddings in NLP tasks. The embedding output has the shape [batch_size × 1024 × 192].

3.3.3. Input Preprocessing (convolution)

To capture local dependencies in the byte sequence, two one-dimensional convolutional layers with dilation are applied:

- nn.Conv1d(d, d, kernel_size=8, stride=4, dilation=2)
- Each convolution is followed by BatchNorm1d(d) and a non-linear activation (GELU).

Dilated convolution exponentially increases the receptive field, enabling the model to detect repeating byte patterns across varying distances, which is crucial for detecting malware with obfuscated payloads.

3.3.4. Positional Encoding

Because byte order is critical, a learnable positional embedding is introduced: nn.Parameter(torch.randn(1, seq_length, d)) with seq_length = 1024. This positional tensor is added element-wise to the byte embeddings after convolution and normalization. Unlike fixed sinusoidal encodings, learnable embeddings adapt to dataset-specific structural patterns.

3.3.5. Transformer Encoder Blocks

The backbone of the architecture is a stack of four Transformer encoder layers, each defined by nn.TransformerEncoderLayer(d_model=192, nhead=8, norm_first=True).

- Multi-Head Self-Attention: Each head processes queries, keys, and values in parallel, enabling the model to capture long-range dependencies across the byte sequence.
- Residual Connections and Pre-Layer Normalization: Implemented to stabilize training and accelerate convergence.

These encoder blocks convert the sequence of embeddings into a context-aware representation where relationships between distant byte patterns are explicitly modeled.

https://doi.org/10.31436/iiumej.v27i1.4009

3.3.6. Attention Pooling Layer

Since Transformer outputs retain the full sequence dimension, an attention pooling mechanism aggregates the sequence into a single feature vector. A learnable weight vector computes attention scores over all sequence positions:

$$\alpha_i = \text{softmax}(W \cdot h_i) \quad (1)$$

where h_i is the hidden representation of byte i. The final vector is obtained by weighted summation, emphasizing critical regions of the byte sequence.

3.3.7. Classification Head

In the final stage, the aggregated feature vector from the attention pooling layer is first normalized using Layer Normalization to ensure stable input distributions. A dropout layer with a probability of 0.4 is then applied to reduce overfitting by randomly deactivating a portion of the neurons during training. The normalized and regularized vector is subsequently passed through a fully connected layer (Linear(d, d/2)) followed by a Gaussian Error Linear Unit (GELU) activation function, which introduces non-linearity and enhances the model's capacity to capture complex patterns. Another dropout operation is applied at this stage to further improve the model's robustness. Finally, the processed representation is mapped to the output space through a linear transformation (Linear(d/2, num_classes)), producing the classification logits that represent the predicted categories, i.e., specific malware families or benign files.

3.4. Training Strategy

The training strategy of the proposed byte-level Transformer was designed to maximize generalization and prevent overfitting through a combination of data augmentation, optimization, and stability mechanisms. Data augmentation was applied using random byte masking and MixUp to increase sample diversity and improve robustness to noise. For optimization, the AdamW algorithm with weight decay was employed, along with a two-phase learning rate schedule: an initial warmup stage to stabilize convergence, followed by cosine annealing to gradually decay the learning rate.

To further enhance training stability, Stochastic Weight Averaging (SWA) was applied during the later epochs to improve generalization. Early stopping with a patience of ten epochs was used to terminate training when validation loss no longer improved. In addition, cross-entropy loss with label smoothing and adaptive class weights was adopted to mitigate overconfident predictions and address class imbalance across malware families.

All experiments were conducted on an NVIDIA Tesla T4 GPU, based on the NVIDIA Turing architecture and equipped with 2,560 CUDA cores and 320 Tensor Cores. The GPU provides 16 GB of GDDR6 memory, enabling efficient training and inference for medium-scale Transformer models with fixed-length byte inputs. Support for mixed-precision computation further accelerated training while maintaining numerical stability. Overall, the chosen hardware configuration allowed the model to converge efficiently within a practical training time and supported fast inference, demonstrating the feasibility of the proposed approach for real-world deployment.

3.5. Comparison with Transformer-Based Malware Detection Models

Transformer-based approaches to malware detection differ primarily in their input representations and architectures. MalBERT, for example, is designed for Android malware analysis and operates on source-code-derived textual artifacts using a pretrained BERT model, requiring platform-specific preprocessing, such as decompilation [21]. In contrast, the

proposed model operates directly on contiguous raw byte sequences from executable binaries and is trained from scratch without relying on pretrained language models or handcrafted feature extraction. By integrating dilated convolution layers with Transformer encoders, the proposed architecture captures both local byte-level patterns and long-range dependencies in a unified end-to-end framework. Due to these fundamental differences in platform and input modality, direct numerical comparison is not feasible. The key architectural distinctions between the proposed approach and MalBERT are summarized in Table 2.

Table 2. Comparison of Transformer-Based Malware Detection Models

Aspect	Proposed Model	MalBERT
Target domain	Windows executable malware (raw binaries)	Android malware
Input representation	Raw byte sequence (first 1,024 bytes)	Source code-derived text artifacts
Feature engineering	None (end to end)	Requires decompilation and text extraction
Architecture	Dilated CNN + Transformer	BERT-based Transformer
Pre-training	No (trained from scratch)	Yes (large-scale LM pretraining)
Key characteristic	Lightweight, contiguous byte modeling	Strong NLP transfer learning
Comparability	Directly applicable to PE/raw bytes	Different platform & modality

4. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed byte-level Transformer model on three publicly available malware datasets. Performance is evaluated using standard classification metrics, namely accuracy, precision, recall, and F1-score. To provide a comprehensive analysis, the results are presented in both tabular and graphical formats, followed by a detailed discussion of each dataset.

4.1. Results on the First Dataset

The first dataset, Malware Detection PE-Based Analysis, comprises 9,970 samples: 8,970 malware and 1,000 benign files. Table 3 summarizes the classification results across six categories.

Table 3. Classification results on the first dataset.

Class	Precision	Recall	F1-Score	Support
Benign	0.99	0.95	0.97	200
Loker	0.81	0.94	0.87	66
Mediyes	1	0.99	1.00	290
Winevesoc	1	1	1	880
Zbot	0.99	0.99	0.99	420
Zerroaccess	1	0.99	0.99	138
Accuracy			**0.99**	**1994**

As shown in Table 3, the model achieved an overall accuracy of 99%, with near-perfect precision, recall, and F1-score for categories such as Mediyes, Winwebsec, and Zeroaccess. However, the Locker class achieved lower scores (precision = 0.81, recall = 0.94), attributable to the smaller number of training samples, which made the feature representation less distinctive. To further illustrate the performance, Figure 4 presents the confusion matrix for this dataset.

https://doi.org/10.31436/iiumej.v27i1.4009

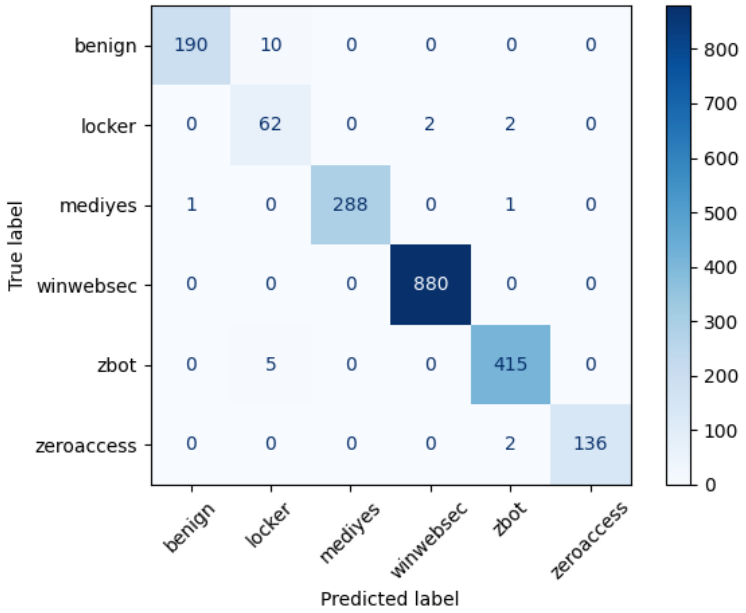


Figure 4. Confusion matrix for the first dataset

The confusion matrix indicates that Winwebsec was perfectly classified, with all samples correctly assigned to their true categories. In contrast, the Locker class exhibits the highest misclassification rate among the six classes, with several instances misclassified as Winwebsec and Zbot. This pattern suggests that Locker shares specific byte-level characteristics with these families, resulting in overlap in learned representations. Such behavior highlights the need for further refinement in feature extraction or modeling strategies to improve separability in future work.

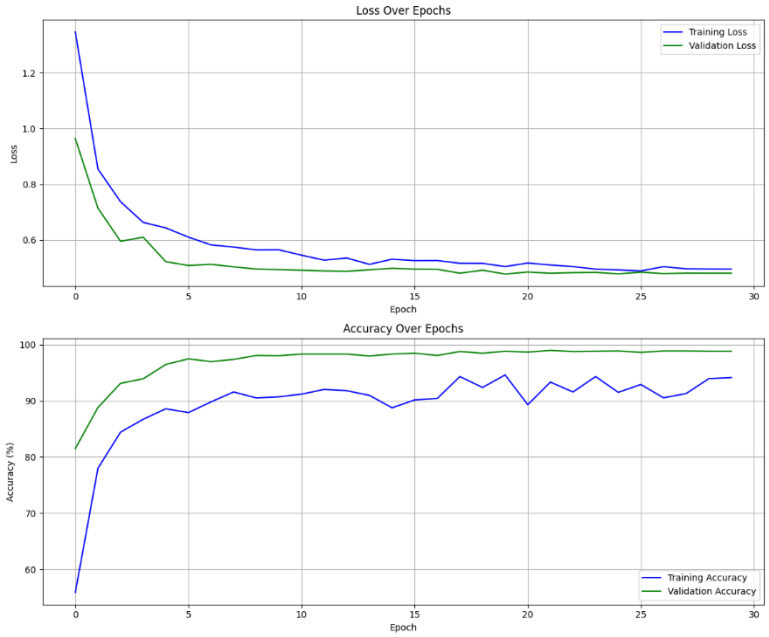


Figure 5. Training and validation performance on the first dataset

Figure 5 shows that the model converges stably, with both training and validation losses decreasing consistently and validation accuracy remaining above 98% across the later epochs. At epoch 30, the model achieved a training accuracy of 94.16% and a validation accuracy of

https://doi.org/10.31436/iiumej.v27i1.4009

98.85%. Early stopping was triggered when no further improvement in validation loss was observed. The full training process required 5 minutes and 1.41 seconds (10.05 seconds per epoch). Following the application of Stochastic Weight Averaging and the restoration of the best-performing checkpoint, the model achieved a peak validation accuracy of 99.00%, demonstrating strong generalization across the complete architectural and regularization setup.

Despite overall high performance, the class distribution in this dataset is imbalanced, which affects minority classes such as Locker, as evidenced by lower precision than for majority classes. To mitigate this issue, adaptive class weighting, Mixup embedding, and label smoothing were employed during training to improve robustness for underrepresented classes. In practical deployment, false positives may lead to unnecessary alerts for benign files, increasing operational overhead, whereas false negatives pose a more critical risk by allowing malware to evade detection. The observed training behavior indicates that the proposed model prioritizes stable generalization while maintaining sensitivity to malicious samples, which is essential for real-world malware detection systems.

To assess the contribution of these components, the following analysis compares the full model with a simplified version that excludes Dilated CNN layers, Attention Pooling, Mixup Embedding, and SWA. This ablation study highlights the performance degradation that occurs when these elements are removed, clarifying their role in enhancing accuracy, convergence dynamics, and overall robustness. The impact of omitting these mechanisms is illustrated in Figures 6 and 7.

Figure 6 shows that the ablation model exhibits higher misclassification rates across several classes than the full model. In the ablation results, errors are more widely distributed, particularly within the benign, Locker, Mediyes, and Zbot classes, which indicate weaker class separation. For example, the Locker class receives fewer correct predictions and exhibits additional confusion with benign, Mediyes, and Zbot, whereas the full model maintains a cleaner diagonal pattern with fewer mistakes. The benign and Mediyes classes also exhibit more scattered errors in the ablation version, whereas the full model maintains greater consistency and higher accuracy.

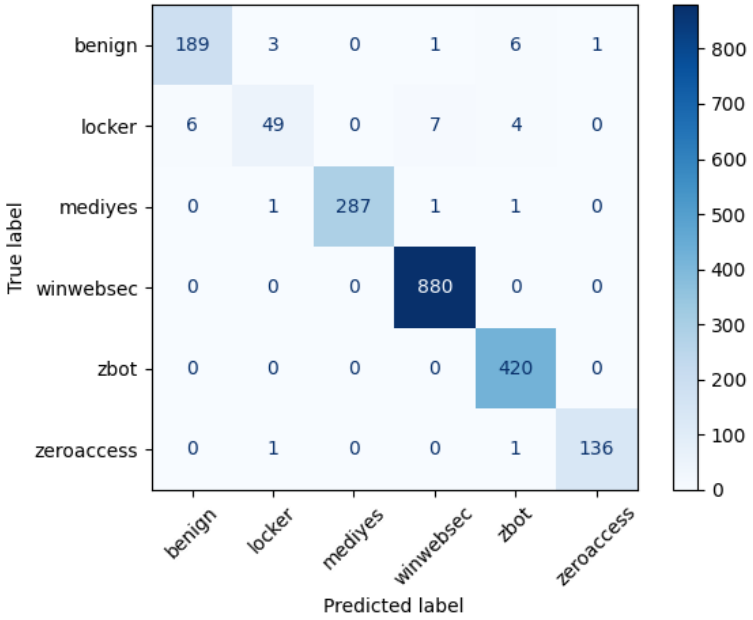


Figure 6. Confusion Matrix for the First Dataset Under Ablation (Without Dilated CNN, Attention Pooling, Mixup, or SWA)

https://doi.org/10.31436/iiumej.v27i1.4009

These differences indicate that the components removed in the ablation configuration reduce the model's ability to learn distinctive features, resulting in greater class overlap. In contrast, the full model exhibits more stable and reliable performance, as evidenced by a sharper diagonal structure and lower misclassification rates.

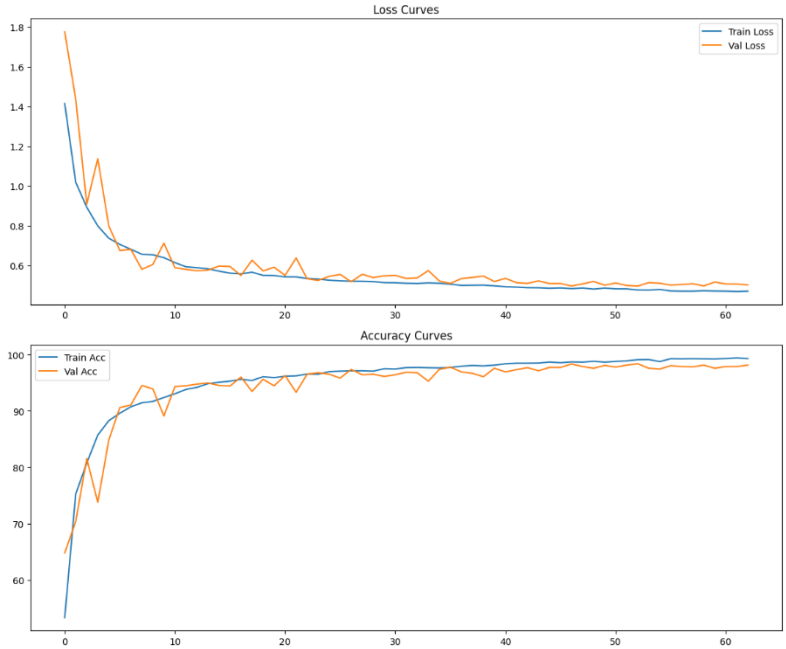


Figure 7. Training and Validation Performance on the first dataset under the Ablation Setting.

Figure 7 shows that the ablation model converges successfully, with both training and validation losses decreasing steadily, and the final validation accuracy reaching 98.35%. However, when compared with the full model, the impact of removing Dilated CNN layers, Attention Pooling, Mixup Embedding, and Stochastic Weight Averaging becomes evident. The full model achieves a best validation accuracy of 99.00 percent in only 30 epochs. In contrast, the ablation model requires 63 epochs and more than two hours of training. The average epoch duration also differs substantially: 134.68 seconds for the ablation model and 10.05 seconds for the complete configuration.

These differences indicate that the removed components play an important role in improving the efficiency and stability of the learning process. In the full model, Mixup and SWA contribute to smoother optimization and stronger generalization, while the Dilated CNN layers and Attention Pooling enhance feature extraction and class separability. As a result, the full model converges more quickly, maintains a cleaner accuracy trajectory, and produces more accurate predictions. The ablation results, therefore, confirm that each component contributes measurable benefits to performance, robustness, and training efficiency.

4.2. Results on the Second Dataset

The second dataset, MC-dataset-binary, comprises 11,480 samples, evenly split between malware and cleanware. Table 4 reports the binary classification performance. As indicated in Table 4, the model achieved an overall accuracy of 92%. Malware samples achieved higher recall (0.94), whereas cleanware samples achieved higher precision (0.95). This imbalance indicates that the model is more effective in detecting malware but occasionally misclassifies benign samples as malicious. The confusion matrix in Figure 8 provides more detail.

https://doi.org/10.31436/iiumej.v27i1.4009

Table 4. Classification results on the second dataset.

Class	Precision	Recall	F1-Score	Support
Cleanware	0.95	0.91	0.92	1148
Malware	0.91	0.94	0.92	1148
Accuracy			**0.92**	**2296**

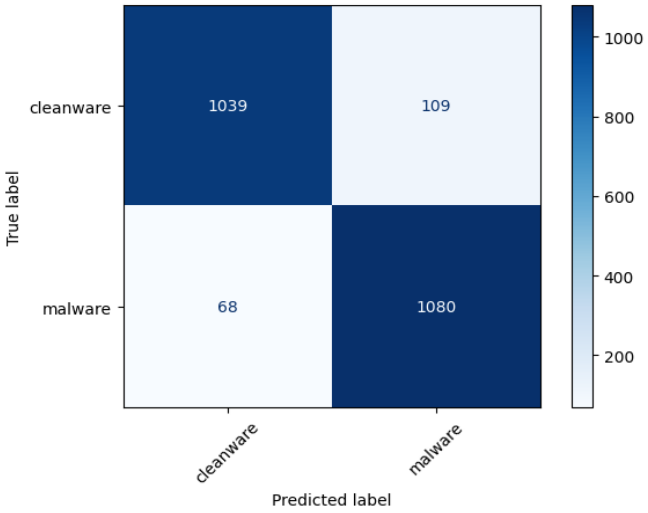


Figure 8. Confusion matrix for the second dataset.

As shown in Figure 8, most cleanware samples were correctly classified, with 1,039 of 1,148 samples predicted accurately. Nevertheless, 109 cleanware samples were misclassified as malware. Conversely, the model misclassified 68 malware samples as cleanware, while 1,080 were correctly identified. These results indicate that the model exhibits slightly greater sensitivity for malware detection, reducing the likelihood of false negatives but still yielding a moderate number of false positives. Training and validation performance are illustrated in Figure 9.

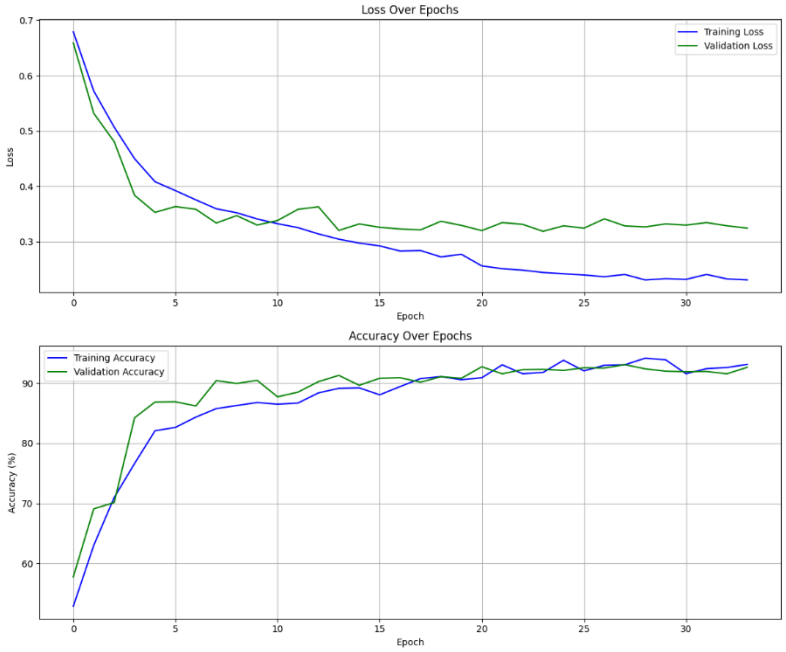


Figure 9. Confusion matrix for the second dataset

https://doi.org/10.31436/iiumej.v27i1.4009

Figure 9 shows that the model converges stably, with losses decreasing and validation accuracy maintaining consistent performance around 92-93%. At epoch 34, the model recorded a training accuracy of 93.12% and a validation accuracy of 92.64%. Early stopping was triggered after ten epochs without improvement in validation loss, resulting in a total training time of 6 minutes and 28.76 seconds (11.43 seconds per epoch). After applying Stochastic Weight Averaging and restoring the checkpoint with the lowest validation loss, the model achieved a best validation accuracy of 93.07%, indicating reliable generalization on the second dataset.

4.3. Results on the Third Dataset

The third dataset, Malware.zip, contains 9,550 samples distributed across 12 malware families with imbalanced class sizes. Table 5 shows that the overall accuracy reached 94%, with perfect predictions for Coinhive and near-perfect scores for Emotet and Icedid. Among all families, Razy exhibited the lowest performance, with a precision of 0.85 and an F1-score of 0.88, indicating a higher rate of misclassification compared to other classes. The recall for Razy remained relatively high at 0.92, suggesting that most Razy samples were correctly identified, although the model exhibited a high rate of false positives. The confusion matrix in Figure 10 provides further insight into these misclassification patterns.

Table 5. Classification results on the third dataset.

Class	Precision	Recall	F1-Score	Support
Coinhive	1	1	1	81
Emotet	0.94	0.98	0.96	181
Fareit	0.97	0.95	0.96	317
Flystudio	0.97	0.93	0.95	73
Gafgyt	0.95	0.97	0.96	297
Gandcrab	0.97	0.92	0.95	78
Icedid	1	0.97	0.98	92
Lamer	0.95	0.85	0.9	138
Mepawo	0.88	0.96	0.92	164
Mirai	0.96	0.95	0.96	288
Ramnit	0.95	0.93	0.94	67
Razy	0.85	0.92	0.88	134
Accuracy			**0.94**	**1910**

As illustrated in Figure 10, most malware families were classified accurately, with high counts along the diagonal of the confusion matrix. Misclassifications occurred between several related categories. For instance, Fareit samples were occasionally predicted as Ramnit, and Gafgyt was sometimes misclassified as Mirai. These patterns suggest shared byte-level characteristics among certain families, which may contribute to overlapping predictions. Figure 11 presents the corresponding training and validation curves for this experiment.

https://doi.org/10.31436/iiumej.v27i1.4009

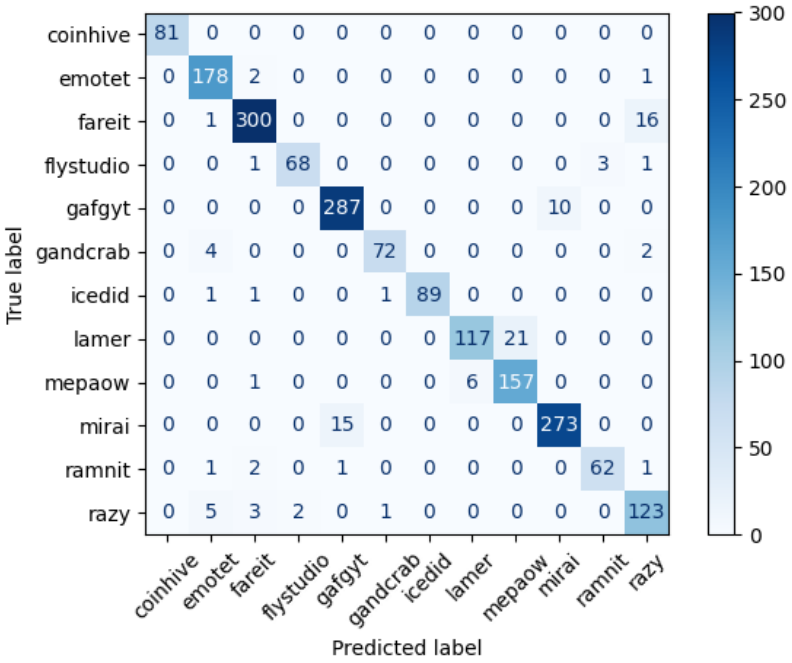


Figure 10. Confusion matrix for the third dataset

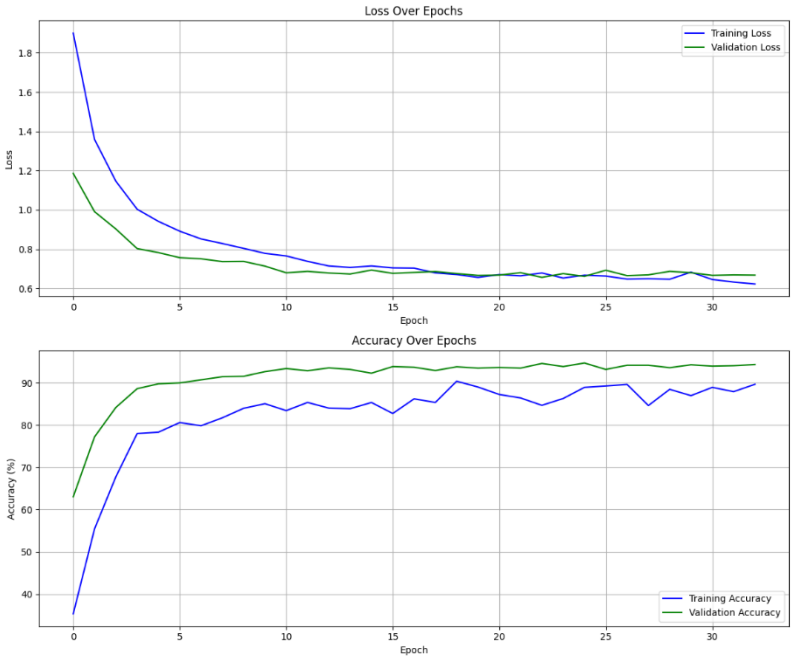


Figure 11. Training and validation performance on the third dataset

Figure 11 shows that the model converges stably on the third dataset, with training and validation losses decreasing consistently and validation accuracy remaining within the 93-95% range. At epoch 33, the model achieved a training accuracy of 89.67% and a validation accuracy of 94.35%. Early stopping was then triggered due to stagnation in validation loss. The full training process required 5 minutes and 12.12 seconds (9.46 seconds per epoch). After applying Stochastic Weight Averaging and restoring the best checkpoint, the model achieved a peak validation accuracy of 94.71%, demonstrating strong and reliable generalization.

https://doi.org/10.31436/iiumej.v27i1.4009

The consistently higher validation accuracy can be attributed to the regularization mechanisms used during training, such as dropout, Mixup embedding, and SWA, which intentionally increase training difficulty and slightly suppress training accuracy while enhancing generalization. Additionally, an imbalanced dataset may yield validation subsets with more representative samples for certain malware families, leading to higher validation performance. This pattern is common in regularized deep learning models and is not necessarily indicative of overfitting.

5. CONCLUSION

This study has successfully designed and implemented a byte-level Transformer model for effective and adaptive malware detection and classification. The proposed approach eliminates the need for handcrafted feature extraction by directly learning from raw byte representations through multi-head self-attention. Experimental evaluations conducted on three publicly available datasets demonstrated strong performance, with classification accuracies of 99% on the Malware Detection PE-Based Analysis dataset, 92% on the MC-dataset-binary dataset, and 94% on the Malware.zip dataset, respectively. These results confirm the effectiveness of byte-level representation in capturing distinctive malware characteristics, enabling the model to generalize across different families and file structures.

The analysis further revealed that malware categories with abundant and distinctive byte patterns, such as Winwebsec, Mediyes, Zeroaccess, Coinhive, and Icedid, achieved near-perfect classification scores. However, minority classes, or those with ambiguous features, such as Locker and Razy, showed comparatively lower performance, underscoring the challenges posed by imbalanced datasets. Beyond experimental validation, the successful deployment of a prototype web-based application demonstrated that the model is not only accurate but also practical for real-world malware detection, providing fast, adaptive, and efficient inference.

Overall, this research highlights the potential of byte-level Transformer architectures as a universal and flexible framework for malware detection. The findings provide direct contributions to advancing deep learning-based cybersecurity solutions and open avenues for future work on handling data imbalance, hybrid feature integration, cross-platform generalization, and zero-day malware detection.

6. LIMITATIONS AND FUTURE WORK

Despite the promising results obtained in this study, several limitations should be acknowledged, which also indicate potential directions for future research.

First, the proposed byte-level Transformer model operates on a fixed input length of 1,024 bytes, analyzing only the initial contiguous segment of each binary file. Although this design choice has been empirically shown to provide stable, consistent performance across all evaluated datasets, it may limit the model's ability to capture malicious patterns beyond the initial byte region, particularly in larger or more complex binaries.

Second, although the model demonstrated strong generalization across three publicly available datasets, no explicit zero-day malware evaluation was conducted. The experimental setup followed a supervised learning paradigm using labeled datasets, without isolating unseen malware families as a dedicated zero-day testing scenario. Consequently, the effectiveness of the proposed model against strictly zero-day malware samples was not directly assessed in this study.

Third, dataset imbalance remains an inherent limitation, especially in multi-class classification tasks. As observed in the experimental results, minority malware families, such as Locker and Razy, exhibited lower precision and F1 scores than the majority classes. Although mitigation strategies, including adaptive class weighting, Mixup embedding, and label smoothing, were applied during training, class imbalance continued to influence the classification performance of underrepresented classes.

Finally, the current study did not evaluate adversarial robustness. The proposed model was assessed under standard classification settings and was not tested against adversarially crafted binaries or evasion-based attacks that may target deep learning-based malware detectors. As a result, the model's robustness under adversarial conditions remains to be investigated.

Future work may address these limitations in several directions. Variable-length input processing or sliding-window mechanisms could be explored to enable the analysis of larger binary regions while preserving sequential continuity. In addition, dedicated zero-day evaluation protocols, such as family-wise holdout or temporal split testing, should be incorporated to assess generalization to previously unseen malware more rigorously. Further investigation into advanced imbalance handling strategies and hybrid feature integration may improve performance for minority classes. Moreover, evaluating adversarial robustness and incorporating adversarial training techniques are essential to enhance the proposed model's resilience in real-world deployment scenarios.

REFERENCES

[1] BSSN, “LANSKAP KEAMANAN SIBER INDONESIA 2024,” Jan. 2025.

[2] K. Khalda and D. K. Wibowo, “Malware Behavior Analysis Using Static and Dynamic Analysis Approaches,” Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi, vol. 4, no. 1, pp. 1–8, Jan. 2025, doi: 10.20885/snati.v4.i1.1.

[3] M. V. Ngo, T. Truong-Huu, D. Rabadi, J. Y. Loo, and S. G. Teo, “Fast and Efficient Malware Detection with Joint Static and Dynamic Features Through Transfer Learning,” Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.13860

[4] M. Abdur Rahman, G. A. Francia Iii, H. Shahriar, G. Francia III, E. El-Sheikh, and S. Iqbal Ahamed, “A Novel Approach to Fine-tune BERT using Non-Text Features for Enhanced Ransomware Detection,” 2025, doi: 10.13140/RG.2.2.35576.15365.

[5] W. A. Salmon, “Learning in Quantum Mechanics,” 2024.

[6] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, “Deep learning: systematic review, models, challenges, and research directions,” Nov. 01, 2023, Springer Science and Business Media Deutschland GmbH. doi: 10.1007/s00521-023-08957-4.

[7] J. Daniel and J. H. Martin, “Speech and Language Processing,” 2024.

[8] Y. Bai, J. Mei, A. Yuille, and C. Xie, “Are Transformers More Robust Than CNNs?,” Nov. 2021, [Online]. Available: http://arxiv.org/abs/2111.05464

[9] Y. Fan et al., “Heterogeneous Temporal Graph Transformer: An Intelligent System for Evolving Android Malware Detection,” in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, Aug. 2021, pp. 2831–2839. doi: 10.1145/3447548.3467168.

[10] A. Rahali and M. A. Akhloufi, “MalBERT: Using Transformers for Cybersecurity and Malicious Software Detection,” Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.03806

[11] T. L. Huoh, T. Miskell, O. Barut, Y. Luo, P. Li, and T. Zhang, “Malware Detection for Portable Executables Using a Multi-input Transformer-Based Approach,” in 2024 International Conference on Computing, Networking and Communications, ICNC 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 778–782. doi: 10.1109/ICNC59896.2024.10556067.

[12] S. Berrios, D. Leiva, B. Olivares, H. Allende-Cid, and P. Hermosilla, "Systematic Review: Malware Detection and Classification in Cybersecurity," Applied Sciences, vol. 15, no. 14, p. 7747, Jul. 2025, doi: 10.3390/app15147747.

[13] F. A. Aboaoja, A. Zainal, F. A. Ghaleb, B. A. S. Al-rimy, T. A. E. Eisa, and A. A. H. Elnour, "Malware Detection Issues, Challenges, and Future Directions: A Survey," Sep. 01, 2022, MDPI. doi: 10.3390/app12178482.

[14] T. L. Wan et al., "Efficient Detection and Classification of Internet-of-Things Malware Based on Byte Sequences from Executable Files," IEEE Open Journal of the Computer Society, vol. 1, pp. 262–275, 2020, doi: 10.1109/OJCS.2020.3033974.

[15] T. McIntosh, A. S. M. Kayes, Y. P. P. Chen, A. Ng, and P. Watters, "Ransomware Mitigation in the Modern Era: A Comprehensive Review, Research Challenges, and Future Directions," Dec. 31, 2022, Association for Computing Machinery. doi: 10.1145/3479393.

[16] M. Naseer et al., "Malware Detection: Issues and Challenges," in Journal of Physics: Conference Series, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1807/1/012011.

[17] A. P. Tuan, A. T. H. Phuong, N. V. Thanh, and T. N. Van, "Malware Detection PE-Based Analysis Using Deep Learning Algorithm Dataset," 2018, figshare. doi: 10.6084/m9.figshare.6635642.v1.

[18] E. de O. Andrade, "MC-dataset-binary," 2018, figshare. doi: 10.6084/m9.figshare.5995408.v1.

[19] A. Azab and M. Khasawneh, "MSIC: Malware Spectrogram Image Classification," IEEE Access, vol. 8, pp. 102007–102021, 2020, doi: 10.1109/ACCESS.2020.2999320.

[20] D. Morozovskii, K. Thummar, T. Halabi, and S. Ramanna, "Toward Efficient and Robust Deep Learning-based Malware Detection in Fog Computing," in 2021 International Symposium on Networks, Computers and Communications, ISNCC 2021, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ISNCC52172.2021.9615812.

[21] A. Rahali and M. A. Akhloufi, "MalBERT: Using Transformers for Cybersecurity and Malicious Software Detection," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.03806

https://doi.org/10.31436/iiumej.v27i1.4049

Deep Learning-Based Skin Care Detection with Multi-method Explainability: Grad-CAM, Lime, and Occlusion Sensitivity

TOOBA KHAN, MUHAMMAD ZEESHAN UL HAQUE, GUL MUNIR*, IRFAN AHMED USMANI

Department of Biomedical Engineering, Salim Habib University, Karachi, Pakistan

**Corresponding author: gul.munir@shu.edu.pk*

(Received: 18 November 2025; Accepted: 22 December 2025; Published online: 14 January 2026)

ABSTRACT: Skin cancer is one of the most common malignancies worldwide, where early detection significantly improves treatment outcomes. While deep learning models show promise for automated skin lesion classification, their lack of interpretability limits clinical adoption. This study presents a comprehensive comparative analysis of three convolutional neural networks, ResNet-50, GoogLeNet, and SqueezeNet, for binary skin lesion classification (benign vs. malignant), integrating three explainable AI (XAI) methods (Grad-CAM, LIME, and Occlusion Sensitivity) to enhance clinical interpretability. We trained and evaluated these architectures on the Kaggle Skin Cancer dataset, which contains 2,637 dermoscopic images (1,440 benign, 1,197 malignant). Transfer learning employed ImageNet pre-trained weights with two-stage fine-tuning. Performance was assessed using accuracy, precision, recall, F1-score, specificity, and AUC-ROC metrics. ResNet-50 achieved the highest accuracy of 91.36% with an excellent AUC of 0.9721, demonstrating superior balanced performance. GoogLeNet achieved 88.94% accuracy with 73% fewer parameters, offering an optimal accuracy-efficiency trade-off. The proposed lightweight CNN, despite having the fewest parameters (1.2M), achieved 85.45% accuracy and a malignancy detection sensitivity of 92.7%, making it well-suited for screening applications. Training times ranged from 1.5 minutes (SqueezeNet) to 3 minutes 39 seconds (ResNet-50), demonstrating feasibility for resource-constrained settings. All XAI methods successfully generated clinically meaningful explanations, with models consistently focusing on lesion centers, color variations, and irregular borders. This study demonstrates that combining deep learning with XAI enables accurate and interpretable skin cancer detection. ResNet-50 is well-suited to well-resourced clinical settings, GoogLeNet offers balanced performance for resource-constrained deployments, and SqueezeNet enables mobile telemedicine applications with superior sensitivity.

ABSTRAK: Kanser kulit merupakan antara malignansi yang paling lazim di seluruh dunia, dan pengesanan awal terbukti dapat meningkatkan keberkesanan rawatan secara signifikan. Walaupun model pembelajaran mendalam menunjukkan potensi tinggi dalam pengelasan automatik lesi kulit, kekurangan kebolehinterpretasian telah mengehadkan penerimaan klinikal. Kajian ini membentangkan analisis perbandingan menyeluruh terhadap tiga rangkaian neural konvolusi, iaitu ResNet-50, GoogLeNet, dan SqueezeNet, bagi pengelasan binari lesi kulit (jinak vs. malignan), digabungkan dengan tiga kaedah kecerdasan buatan boleh jelas (XAI), iaitu Grad-CAM, LIME, dan Kepekaan Halangan, bagi menyokong interpretasi klinikal. Model dilatih dan dinilai menggunakan set data Kanser Kulit Kaggle yang mengandungi 2,637 imej dermoskopi, dengan menggunakan pembelajaran pindahan berasaskan pemberat pralatih ImageNet dan penalaan halus dua peringkat. Penilaian prestasi menggunakan metrik ketepatan, ketepatan ramalan, kepekaan, skor F1, pengkhususan, dan AUC-ROC menunjukkan bahawa ResNet-50 mencapai prestasi tertinggi dengan ketepatan

91.36% dan AUC 0.9721, manakala GoogLeNet menawarkan keseimbangan optimum antara ketepatan dan kecekapan dengan pengurangan parameter sebanyak 73%. SqueezeNet, walaupun paling ringan, mencapai kepekaan pengesanan malignan tertinggi sebanyak 92.7%, menjadikannya sesuai untuk aplikasi saringan dan teleperubatan mudah alih. Semua kaedah XAI berjaya menghasilkan penjelasan bermakna secara klinikal, dengan fokus konsisten pada pusat lesi, variasi warna, dan sempadan tidak sekata. Secara keseluruhan, kajian ini membuktikan bahawa penggabungan pembelajaran mendalam dan XAI membolehkan pengesanan kanser kulit yang tepat, boleh ditafsir, dan sesuai dalam pelbagai kekangan sumber klinikal.

KEYWORDS: *Skin cancer detection; Deep learning; Convolutional neural network; Explainable AI; Transfer learning*

1. INTRODUCTION

Skin cancer is one of the leading causes of death, posing a significant impact on the global health burden. Annually, around 57,000 deaths have been reported due to melanoma itself [1]. Early detection can improve survival rates: melanoma diagnosed early is associated with a 99% five-year survival rate, whereas later diagnosis is associated with a 27% five-year survival rate [2]. It remains a challenging task for experienced dermatologists to diagnose the disease accurately [3,4]. The use of deep learning by Adla et al. has enabled them to develop computer-aided diagnosis (CAD) systems that autonomously detect and classify skin cancers. Convolutional Neural Networks (CNNs) process dermatoscopic images to distinguish cancerous from non-cancerous skin lesions with high accuracy. The proposed CAD system enhances the quality of skin cancer examinations by extracting unique image characteristics and thereby reducing human interpretation-based errors. The authors' results illustrate the potential for AI-driven CAD tools to support early detection of skin cancers and assist clinical decision-making [5]. The work of Alam et al. describes an effective skin cancer classification system using deep learning that helps solve issues related to unevenly distributed data across the different class labels. With this approach, the authors have demonstrated how to improve sensitivity and accuracy for underrepresented categories, thereby enhancing the detection of malignant skin lesions [6]. Chibueze et al. proposed a CNN-based method for heart disease classification using 90,500 MRI samples, incorporating caffeine intake as an additional risk factor. They demonstrated that their approach provides 94% accuracy and robust cross-validation results and outperformed previously published methods for classifying cardiac images [7]. Panigrahi et al. explored the application of capsule networks to the analysis of histopathological images associated with oral squamous cell carcinoma. On a dataset of 150 samples, CapsNet achieved the highest accuracy (97.35%) among their models [8].

Improved diagnostic accuracy has been observed compared with visual inspection via dermoscopy, which magnifies skin structures non-invasively while reducing surface reflection [9]. However, subjective pattern recognition and clinician experience are required to interpret dermoscopy results accurately. Timely diagnosis remains a challenge in remote areas with a shortage of experienced dermatologists [10]. To overcome these limitations and enable early diagnosis, this research focuses on developing a system that assists clinicians with automated diagnosis and evaluation.

Deep learning, specifically convolutional neural networks (CNNs), has demonstrated performance comparable to that of an expert dermatologist in medical image analysis [4, 11, 12]. A trained deep CNN on over 129,000 images demonstrated exceptional performance compared with expert dermatologists in classifying skin lesions [11]. A similar study showed

that a trained CNN, tested on 300 images, outperformed several expert clinicians in melanoma detection [4]. These studies demonstrate that AI-assisted dermatological diagnosis has significant potential.

However, clinical adoption of AI-assisted diagnostic tools is hindered by the black-box nature of deep neural networks [13,14]. To integrate AI recommendations into clinical decision-making, clinicians need a deep understanding to build trust in the model's predictions and decisions. This will help validate the decision and identify potential errors. Given the lack of interpretability and the black-box nature of AI tools, even the most advanced diagnostic AI models will face reluctance from clinicians and dermatologists, who are accountable for patient-care decisions [14].

This interpretability gap has been addressed by Explainable AI (XAI), which generates visual explanations and reasons by highlighting regions in an image for prediction [15-17]. One XAI technique, called gradient-weighted class activation mapping (Grad-CAM), generates class-discriminative localization maps using gradient information [14]. Another XAI technique, namely local interpretable model-agnostic explanations (LIME), interprets the image by fitting local linear approximations [16]. Critical image regions are systematically occluded by Occlusion Sensitivity [17].

While several recent studies focus on various XAI techniques for image analysis, there remains a gap in comparative evaluations across architectures and XAI techniques. Mostly, a single architecture is kept in focus, with very limited interpretability [4,11]. Considering the trade-offs among efficiency, interpretability, and accuracy to determine which architecture to choose in a specific clinical context remains underexplored.

This study addresses these gaps through three primary contributions. First, we present a comprehensive performance comparison of three CNN architectures spanning an order of magnitude in complexity: ResNet-50 (25.6M parameters), GoogLeNet (6.8M parameters), and SqueezeNet (1.2M parameters). Second, the study conducts an evaluation using multiple XAI techniques, namely Grad-CAM, LIME, and Occlusion Sensitivity, focusing not only on visualization but also on the challenges encountered in implementation across various architectures.

2. METHODOLOGY

In this study, deep learning models are combined with XAI models to enhance interpretability and clinical trustworthiness in automated skin cancer detection. Figure 1 shows how deep learning models are combined with XAI models.

2.1. Dataset

The dataset for “Skin Cancer: Malignant vs. Benign” was downloaded from Kaggle and is publicly available [18]. It comprises dermoscopic images of 2,637, of which 1,440 (54.6%) are benign, and 1,197 (45.4%) are malignant. Melanocytic nevi, seborrheic keratoses, and other benign neoplasms are included in the benign dataset, whereas malignant lesions include melanomas. The dataset's benign-to-malignant ratio of 1.2:1 indicates a balanced distribution, eliminating the need to balance the class distribution [10, 19]. The dataset images have variable resolutions, ranging from 500x500 to 1024x1024 pixels, and are encoded in 24-bit RGB, which is necessary for lesion evaluation [21]. Ruler markings for size measurements, variations in lightning, and hair artifacts are included in the dataset.

https://doi.org/10.31436/iiumej.v27i1.4049

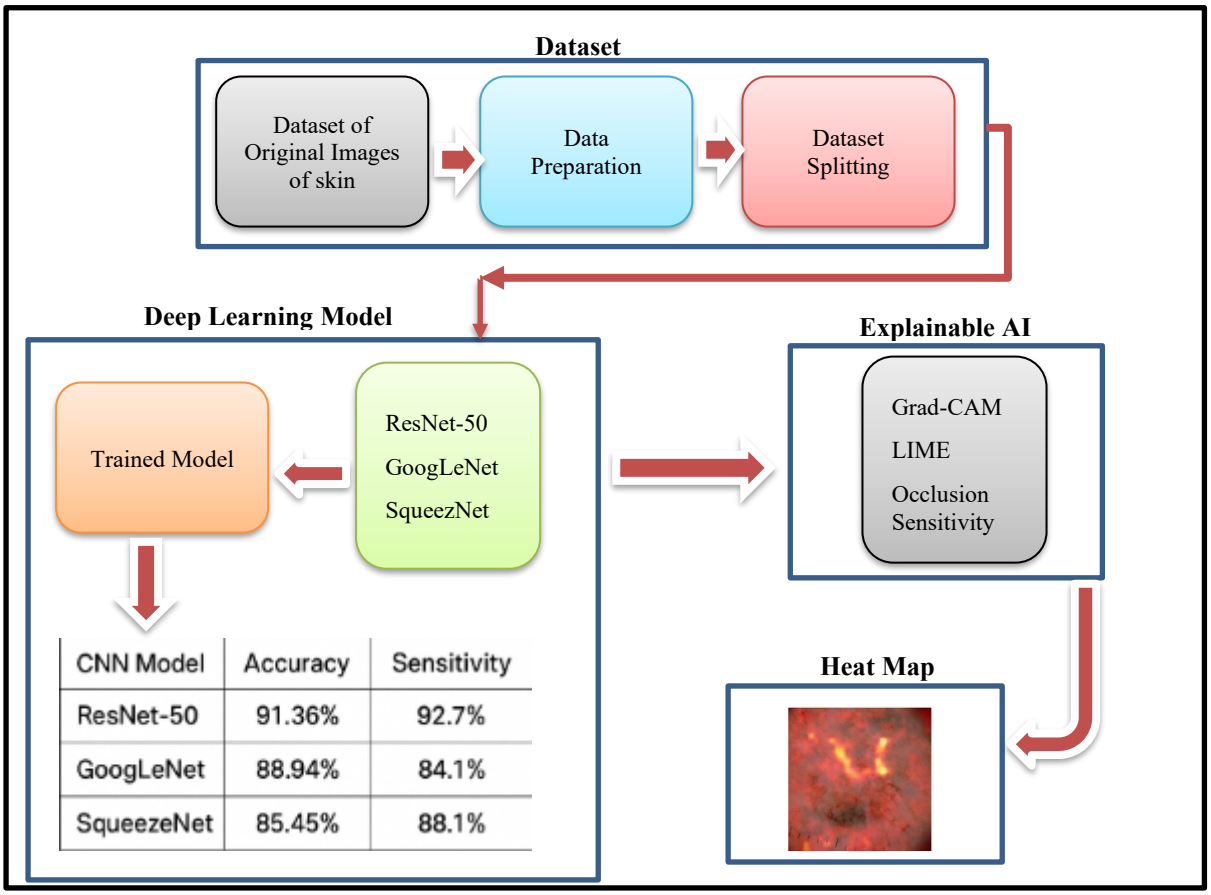


Figure 1. Overall framework illustrating the integration of deep learning models with explainable AI (XAI) techniques for enhanced heart disease detection and model interpretability

To maintain consistent class distributions, random stratified sampling [22] was used to partition the dataset, as shown in Table 1. The dataset was split so that approximately 70% (1,846 images) were included in the training set, comprising 1,008 benign and 838 malignant images. The validation set shall consist of 15% of the images (395), of which 216 are benign, and 179 are malignant. The test set was not used in training and was reserved for final evaluation. Reproducibility across all experiments is ensured via a fixed random seed (seed=42) [23].

Table 1. Distribution of skin lesion images across training, validation, and test subsets

Subset	Benign	Malignant	Total	Percentage
Training	1008	838	1846	70.0%
Validation	216	179	395	15.0%
Test	216	180	396	15.0%
Total	1440	1197	2637	100%

2.2. CNN Architectures

The study utilizes architecture with different points on the accuracy-efficiency spectrum for extensive comparison. ResNet-50, which deploys residual connections via identity mappings, enables direct information flow across layers, and is comprised of 50 layers with 25.6 million parameters [24]. The vanishing-gradient issue has been addressed through the structural innovations in ResNet-50, which also enable efficient training of very deep networks

https://doi.org/10.31436/iiumej.v27i1.4049

[25]. ResNet-50 has demonstrated exceptional performance in dermatological image analysis [4,11] and on ImageNet classification, and accepts an input image of 224x224x3 [26].

Inception modules, characterized by parallel multi-scale feature extraction, were employed in GoogleNet (Inception v1). It comprises 22 layers and 6.8 million parameters, approximately 73% fewer than those of ResNet-50 [27]. Concurrent convolutions were performed by each inception module with various filter sizes of (1×1, 3×3, 5×5) [28]. It accepts 224×224×3 images and is suitable for resource-constrained deployments, as it offers superior parameter efficiency [28].

SqueezeNet employs fire modules that use a squeeze-and-explode strategy to reduce parameters while maintaining representational capacity. It comprises 18 layers with 1.2 million parameters, representing a 95% reduction relative to ResNet-50 [29]. It accepts input images of 227×227×3 and is efficient for telemedicine-based applications [30,31].

2.3. Explainable AI Methods

Three complementary XAI techniques provided interpretability for model predictions. Gradient-weighted Class Activation Mapping (Grad-CAM) produces class-discriminative localization maps by using gradient information from the final convolutional layers [15]. It computes importance weights for each feature map based on the extent to which each spatial location contributes to the target-class prediction. The final heatmap highlights regions that positively contribute to predictions by applying a weighted combination of feature maps, followed by ReLU activation. Implementation targeted final convolutional layers before global pooling: 'inception_5b-output' for GoogLeNet, 'activation_49_relu' for ResNet-50, and 'fire9-concat' for SqueezeNet, with bilinear upsampling to input resolution.

Local Interpretable Model-agnostic Explanations (LIME) explain predictions by fitting local linear approximations around specific instances [16]. For images, LIME segments images into interpretable superpixels using SLIC segmentation [32], generating approximately 100-200 superpixels. It then creates perturbations by randomly occluding superpixel subsets across 5000 samples, evaluates model predictions on these perturbed images, and fits a weighted linear regression with an exponential kernel to identify the 50 most essential superpixels.

Systematic occlusion of image regions is performed using a 16×16-pixel sliding window with an 8-pixel stride in occlusion sensitivity analysis. The critical regions where prediction confidence is reduced by occlusion were revealed by the resulting sensitivity maps. This technique yields predictions without requiring gradient information, thereby avoiding vanishing gradients [17].

Accuracy, precision, recall (sensitivity), F1 score, specificity, and AUC-ROC were recorded to ensure the classification performance. Precision is utilized to minimize unnecessary biopsies by indicating the correct proportion of predicted malignant cases that are truly malignant [20]. The proportion of actual malignant cases correctly predicted by a model is indicated by the recall value and is critical for avoiding missed diagnoses [2]. Classifier performance was evaluated using the AUC-ROC across all classification thresholds. The value above 0.90 is considered excellent for medical diagnosis [33]. The percentage of images for which XAI generates valid explanations without errors or variance, along with a qualitative visual assessment, is included in XAI evaluations [34].

Training time, including wall-clock time from initialization to convergence, per-image interference time, and model size, was measured to evaluate computational efficiency. McNemar's test with $\alpha = 0.05$ was used to assess statistical significance [35].

https://doi.org/10.31436/iiumej.v27i1.4049

3. RESULTS AND DISCUSSION

3.1. Classification Performance

The results indicate that ResNet-50 has outperformed GoogLeNet with the highest accuracy of 91.36%, whereas GoogLeNet achieved (88.94%, $p<0.05$), and SqueezeNet achieved (85.45%, $p<0.01$), based on McNemar's test as depicted in Table 2. Strong discriminative ability across classification thresholds is indicated by AUC-ROC values exceeding 0.95 for all models. The ResNet-50 achieved the highest AUC of 0.9721, indicating exceptional separation between benign and malignant cases. GoogLeNet achieved an AUC of 0.9617, and SqueezeNet achieved 0.9547.

Table 2. Classification performance of evaluated CNN architectures on the skin lesion dataset

Model	Accuracy	Macro Avg Precision	Macro Avg Recall	Macro Avg F1	AUC-ROC
ResNet-50	91.36%	0.915	0.887	0.900	0.9721
GoogLeNet	88.94%	0.895	0.886	0.890	0.9617
SqueezeNet	85.45%	0.861	0.877	0.868	0.9547

The performance hierarchy indicates that ResNet-50 > GoogLeNet > SqueezeNet, which is directly related to model capacity: 25.6M, 6.8M, and 1.2M parameters, respectively. Accordingly, architectures with more parameters learned more discriminative features from dermoscopic images [27].

Statistically significant differences in mean classification accuracy were observed across the five deep-learning convolutional neural network (CNN) architectures evaluated using a series of pairwise two-sample t-tests. The mean classification accuracy for ResNet-50 was significantly better than that of GoogLeNet (two-tailed $p = 4.43\times10^{-13}$) and SqueezeNet (two-tailed $p = 2.40\times10^{-20}$). In addition, GoogLeNet's mean classification accuracy was greater than that of SqueezeNet (two-tailed $p = 7.27\times10^{-16}$), and these results continued to be significant after adjusting for multiple comparisons using the Bonferroni correction factor test, where adjusted $\alpha = 0.0167$.

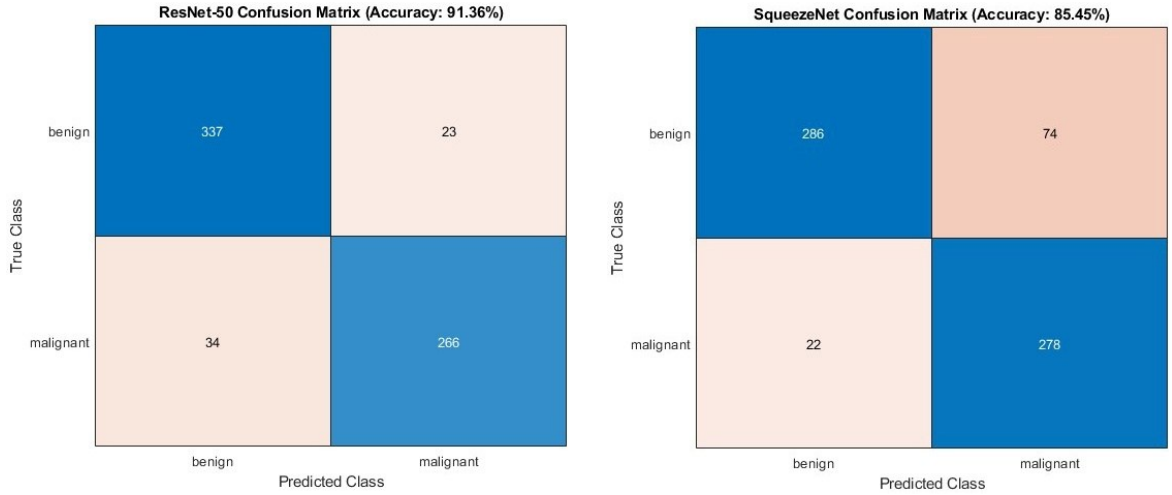
The confusion matrix results are depicted in Figure 2, which demonstrates error patterns across various models. Balanced performances were achieved by ResNet-50 with 337 true negatives-benign classified correctly, whereas 23 false positives indicated benign misclassified as malignant, 34 false negatives, which indicates malignant misclassified as benign, and 226 true positives indicating malignant correctly classified (as mentioned in Figure 2a). The results showed that for malignant detection, sensitivity (recall) was approximately 88.7%, and specificity was approximately 93.6%. For clinical diagnosis, this indicates an optimal balance in which both false positives (unnecessary biopsies) and false negatives (missed cancers) carry significant consequences.

GoogLeNet results indicate 84.3% sensitivity and 92.8% specificity, with 334 true negatives, 26 false positives, 47 false negatives, and 253 true positives, as shown in Figure 2b. GoogLeNet misses more malignant cases, as indicated by its higher false-negative rate of 47 compared with ResNet-50, which has 34 false negatives. This highlights a critical limitation of GoogLeNet for cancer screening applications, where sensitivity is a key consideration.

SqueezeNet results depicted in Figure 2c exhibit a slightly different error pattern, with a sensitivity of 92.7% and a specificity of 79.4%, yielding 286 true negatives, 74 false positives, 22 false negatives, and 278 true positives. It misses fewer malignant cases, i.e., only 22 (7.3%)

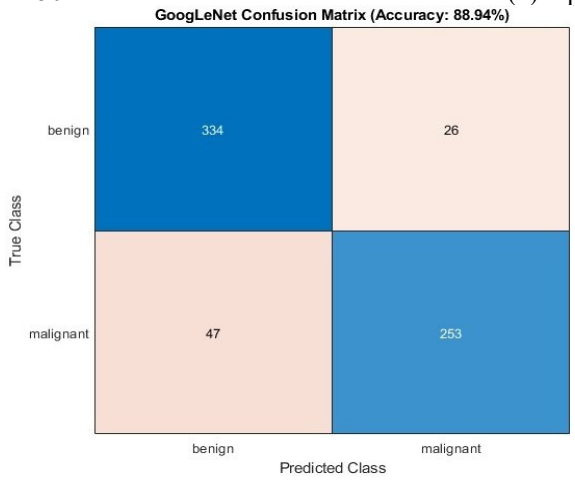
https://doi.org/10.31436/iiumej.v27i1.4049

in comparison to ResNet-50, which is around 34 (11.3%), and Googlenet, which is 47(15.7%), which indicates that it is suitable for cancer screening applications where the primary objective is to reduce the missed detection.



(a) ResNet-50

(b) SqueezeNet

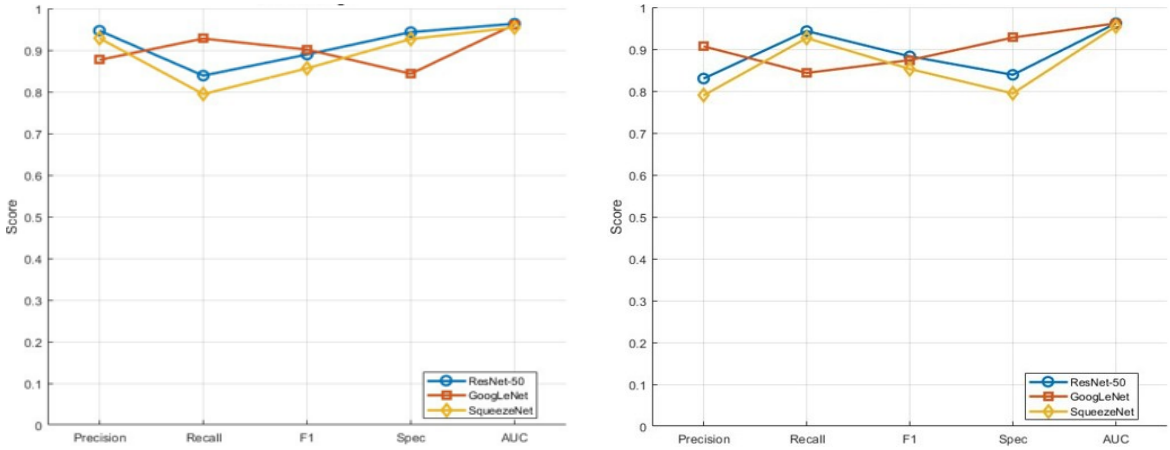


(c) GoogLeNet

Figure 2. Confusion matrices for ResNet-50, GoogLeNet, and SqueezeNet models on the test dataset

There are differences in class-wise performance analysis across architectures. ResNet-50 achieved a precision of 0.95, a recall value of 0.84, and an F1-score of 0.89 for benign lesions (as mentioned in Figure 3a). The results obtained with GoogLeNet demonstrate a precision of 0.88 and a recall of 0.93, indicating greater sensitivity for benign detection. SqueezeNet achieves a precision of 0.93, but recall is slightly lower at 0.79 for benign lesions. The precision of 0.92 is achieved using ResNet-50 for malignant lesions, with a recall of 0.89 and an F1-score of 0.91. The results obtained with GoogLeNet represent the best-balanced performance (as shown in Figure 3b).

https://doi.org/10.31436/iiumej.v27i1.4049



(a) Benign Class (b) Malignant Class

Figure 3. Comparative evaluation of (a) Benign class, and (b) Malignant class metrics for all models in the study.

3.2. Training Dynamics and Computational Efficiency

A distinct convergence pattern has been observed across architectures in training progress analysis, as depicted in Tables 2 and 3. Rapid convergence is observed with ResNet-50 in only 8 epochs, within 3 minutes, despite its large size. Within the first epoch, initial accuracy improved from 50% to 85% on the validation set and, with subsequent refinement, to 91.36%. Stable optimization is evidenced by a smooth decrease in the loss curve from 0.8 to 0.28, with no significant overfitting.

With a training time of 2 minutes and 7 seconds, GoogLeNet converged in 12 epochs, 42% faster than ResNet-50. Slight oscillations were observed in the training curve of GoogLeNet in comparison to ResNet-50, with fluctuating validation accuracy of 85-90% before stabilizing at 88.94%. A gap between training and validation is observed, with minimal overfitting, and good performance is achieved despite the oscillations.

The training time required by SqueezeNet is 1 minute 25 seconds over 14 epochs, which is 61% faster than that of ResNet-50 and 33% faster than that of GoogLeNet. This faster convergence indicates the efficiency gain in ultra-lightweight architectures. The training curve shows a less stable behavior with oscillation and validation accuracy between 80-88%, mainly due to the limited capacity of the model. SqueezeNet achieved the final validation accuracy of 85.45%, indicating acceptable performance.

The computational efficiency metrics for all three architectures are presented in Table 3. ResNet-50 requires approximately 1,450 MB of GPU memory during training due to its 25.6M parameters and 102MB model size. In contrast, GoogLeNet enables deployment on more modest hardware, achieving a 73% reduction with 6.8M parameters and approximately 27MB, with an estimated memory footprint of 780MB. SqueezeNet, by contrast, achieves a 95% reduction with 1.2M parameters and a 5MB model size, while requiring only 320MB of memory.

Table 3. Computational efficiency of CNN architectures in terms of training time, model size, and resource requirements

Model	Training Time	Epochs	Parameters	Model Size	Est. GPU Memory
ResNet-50	3 min 39 sec	8	25.6M	~102 MB	~1,450 MB
GoogLeNet	2 min 7 sec	12	6.8M	~27 MB	~780 MB
SqueezeNet	1 min 25 sec	14	1.2M	~5 MB	~320 MB

https://doi.org/10.31436/iiumej.v27i1.4049

Table 4 shows that, for the three models (ResNet-50, GoogLeNet, and SqueezeNet), regression accuracy increases with model size. In GoogLeNet, for every 1 million parameters added to the model, it increased accuracy by 0.03 ($p = 1.1 \times 10{-}12$). In ResNet, for every 1 million parameters added to the model, accuracy increased by 0.14 ($p = 3.23 \times 10{-}11$). In SqueezeNet, for each additional million parameters, accuracy increased by 0.76 ($p = 4.82 \times 10{-}6$). Thus, each model had a significant positive effect on accuracy and performance, attributable to the number of parameters.

Table 4. Regression results comparing the performance of ResNet-50, GoogLeNet, and SqueezeNet

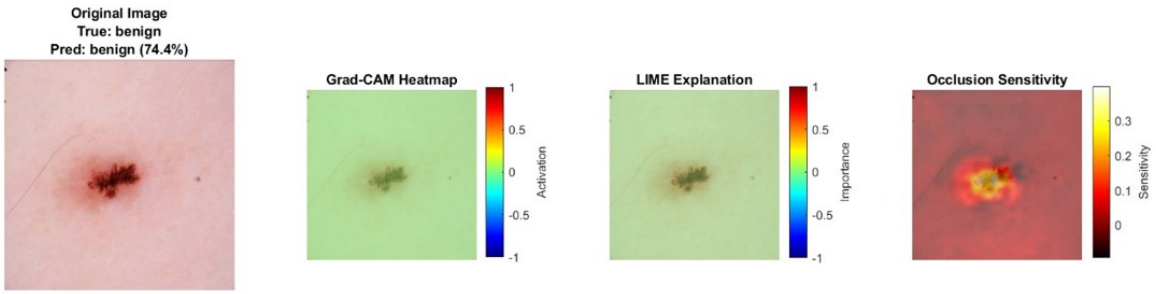
Parameters per Million	Coefficients	Standard Error	t Stat	P-value
ResNet-50	0.037011	0.000674	54.95004	1.1E-12
GoogLeNet	0.145083	0.003848	37.69923	3.23E-11
SqueezNet	0.76165	0.078952	9.646968	4.82E-06

3.3. Explainable AI Analysis

The results obtained from Grad-CAM, LIME, and Occlusion Sensitivity demonstrate successful explanations for all three CNN architectures. XAI visualizations for all three methods, used to correctly classify benign and malignant samples, are shown in Figure 4. All models focused on the central dark region, with irregular borders in malignant samples (see Figures 4b, 4d, and 4f). Diffuse attention is observed across the entire lesion via ResNet-50's Grad-CAM analysis, whereas GoogLeNet has highlighted more attention on specific morphological features. In contrast, SqueezeNet, due to its shallower architecture, exhibits coarser attention to different patterns in malignant samples (as shown in Figure 4f).

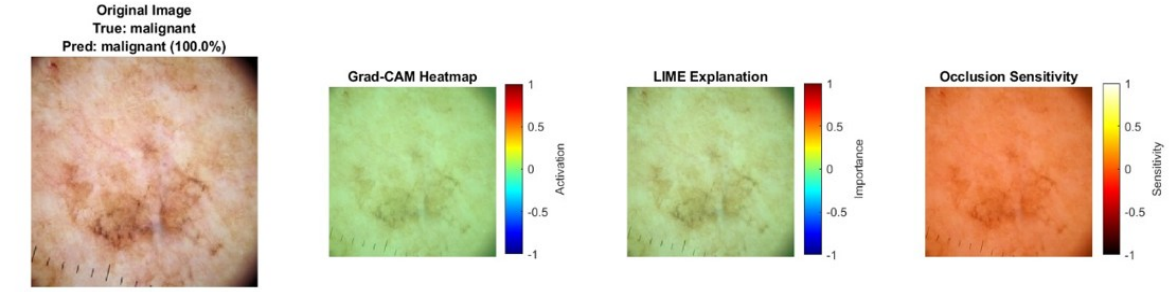
Grad-CAM identified more uniform regions with regular borders for benign lesions. Across all models, the benign samples showed minimal activation in the surrounding skin area and greater activation at the center. Superpixel-level interpretations were obtained using LIME explanations. The highlighted information by LIME corresponds to color variation and irregular pigmentation patterns for malignant cases. Specific morphological features, including asymmetric pigment distribution and irregular borders, are identified by LIME explanations because they are discrete. LIME highlighted a more uniform distribution of superpixel patterns in benign cases.

Occlusion sensitivity maps depicted in Figure 4(a-f) have visualization differences across different models. Strong sensitivity was identified by occlusion maps when occlusion is done on central pigmented regions, which appear as hot spots in sensitivity maps in malignant lesions. Broder sensitivity regions were obtained using SqueezeNet, whereas more focused patterns were obtained using ResNet-50, indicating differences in receptive field sizes. More distributed sensitivity patterns with fewer score changes were observed in occlusion maps of benign lesions.

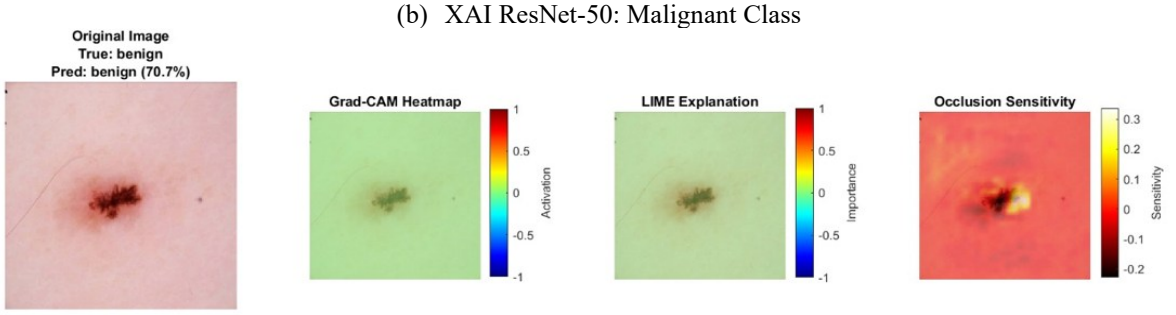


(a) XAI ResNet-50: Benign Class

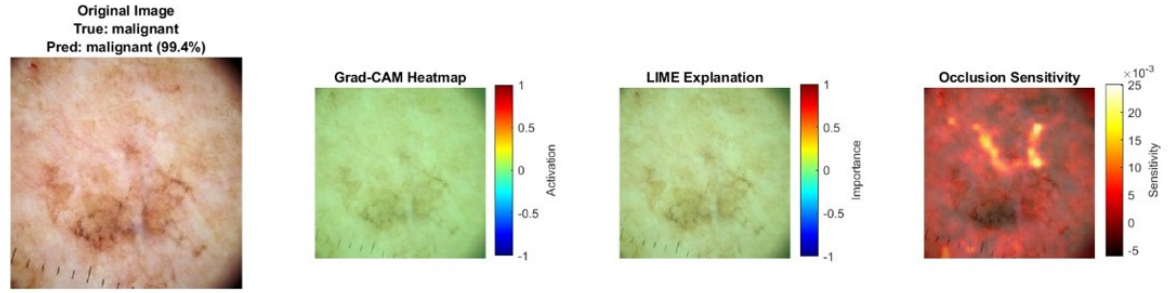
https://doi.org/10.31436/iiumej.v27i1.4049



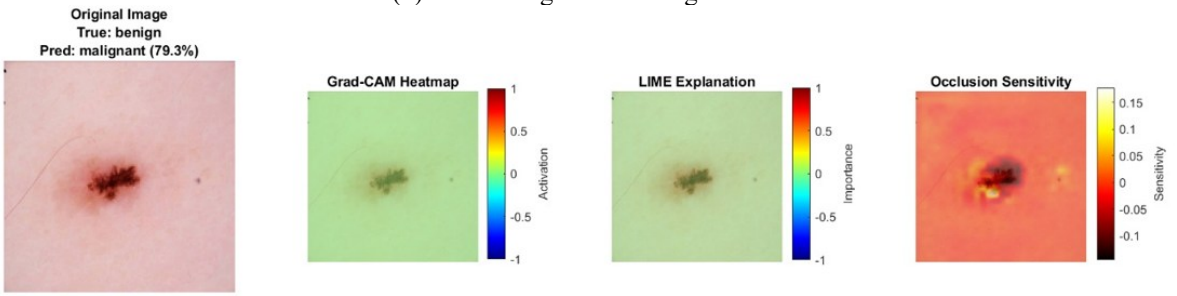
(b) XAI ResNet-50: Malignant Class



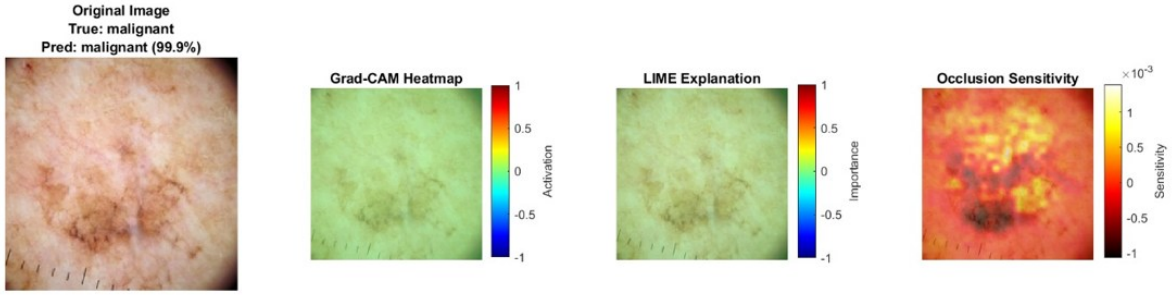
(c) XAI GoogLeNet: Benign Class



(d) XAI GoogLeNet: Malignant Class



(e) XAI SqueezNet: Benign Class



(f) XAI SqueezNet: Malignant Class

Figure 4. Comparison of Grad-CAM, LIME, and Occlusion Sensitivity explainability outputs across (a) ResNet-50: Benign Class, (b) ResNet-50: Malignant Class, (c) GoogLeNet: Benign Class, (d) GoogLeNet: Malignant Class, (e) SqueezNet: Benign Class, (f) SqueezNet: Malignant Class Models

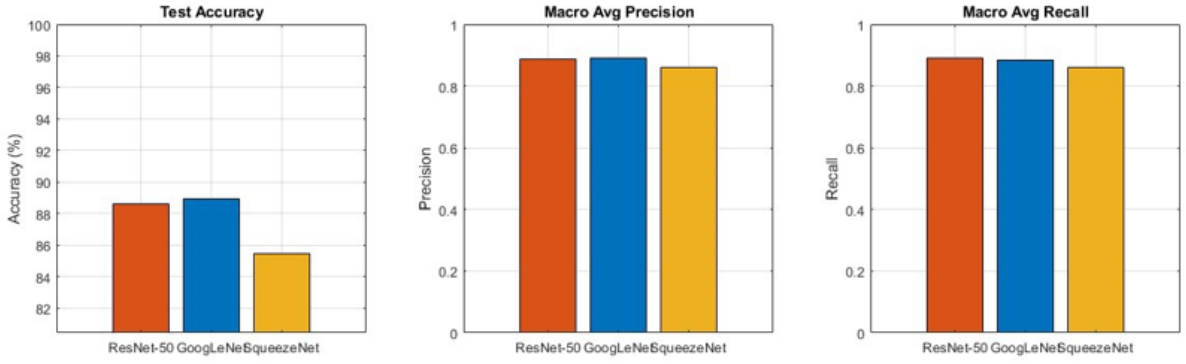
https://doi.org/10.31436/iiumej.v27i1.4049

Qualitative analysis indicates that greater emphasis is placed on image artifacts than on diagnostically relevant features. As depicted in Figure 3b, the malignant samples exhibit attention to ruler markings visible at the image edge, which highlights a potential safety concern that the models might learn spurious correlations in place of diagnostic features [36].

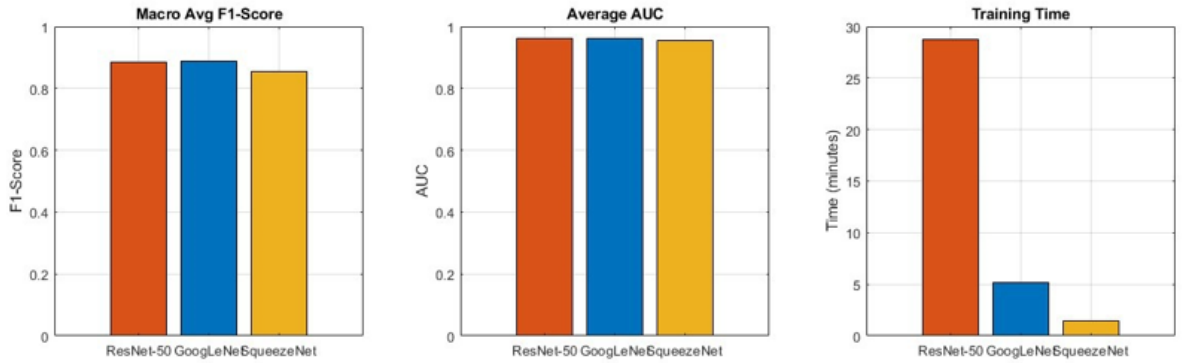
Robust implementability is demonstrated by XAI generation success rates of 100% across all models on the test set, whereas computational cost varied widely. A single forward-backward pass was required by Grad-CAM, which takes approximately 50ms per image. Generation and evaluation of 5000 perturbations, approximately 2-3 seconds per image, is required by LIME, whereas systematic evaluation of roughly 400 occluded variants, which is approximately 1-2 seconds per image, is required by occlusion sensitivity. This difference in timing has implications for real-time deployment in clinical settings versus offline analysis.

3.4. Error Analysis and Model Agreement

Despite architectural differences, all three models correctly classified 312 of 396 test images (78.8%), whereas 21.2% (84 images in total) showed disagreement; however, 4.5% of images were misclassified by all three models. A comparison of accuracy, precision, recall, F1-score, AUC, and training time is shown in Figure 5. Although ResNet-50 and GoogLeNet in Figure 5a achieved the highest test accuracies, all three models showed small differences in Precision and Recall; however, all three models exhibited consistently high Performance, despite the gap in accuracy. In Figure 5b, the Macro-averaged F1 Scores and AUC were also very close across the three models, indicating that the models' classification capabilities were highly stable; however, the ResNet-50 model required much longer training than both the GoogLeNet and SqueezeNet moels.



(a) Test Accuracy, Macro Avg Precision, and Macro Avg Recall



(b) Macro Avg F1-Score, Average AUC and Training Time

Figure 5. Side-by-side comparison of classification performance for all three CNN architectures (a) Test Accuracy, Macro Avg Precision and Macro Avg Recall, and (b) Macro Avg F1-Score, Average AUC and Training Time

https://doi.org/10.31436/iiumej.v27i1.4049

Several patterns have been observed in common failure cases by visual inspection. In some cases, subtle color variations were complex even for expert dermatologists to interpret. Some cases exhibit motion artifacts and severe hair occlusion, which degrade image quality. Irregular features, including asymmetry and border irregularity, are also observed in some benign lesions, although these features are typically associated with malignancy. This indicates that it's challenging for expert dermatologists to predict these lesions in clinical practice [37].

Distinct error patterns revealed model-specific failure analysis. Almost 89.9% (356) of the predictions were agreed upon by ResNet-50 and GoogLeNet, whereas ResNet-50 and SqueezeNet agreed upon 83.8% (332) of the predictions. The architectural similarity and prediction consistency are the main reasons for the higher agreement between ResNet-50 and GoogLeNet.

Clinically critical patterns are revealed by false-negative analysis. Almost 34 malignant cases (11.3%) were missed by ResNet-50, 47 cases (15.7%) were missed by GoogLeNet and only 22 cases (7.3%) were missed by SqueezeNet. Despite low overall accuracy, its superior sensitivity for malignancy detection makes it ideal for cases in which missed cancers carry the highest cost [20].

3.5. Principal Findings

Three CNN architectures were evaluated in this study, focusing on automated skin cancer detection coupled with multi-method XAI analysis across the accuracy-efficiency spectrum. The following findings emerged with some important clinical and technical implications. Firstly, ResNet-50 achieves superior classification performance (91.36% accuracy and 0.9721 AUC), with optimal sensitivity (88.7%) and specificity (93.6%), making it an ideal choice for clinical applications. The dermatologist-level accuracy is approached in this meta-analysis, with a pooled sensitivity of 87% and specificity of 89% [3]. However, well-resourced facilities are needed to meet the computational requirements of 102MB, 3.6-minute training on a GPU.

Compelling accuracy-efficiency balance with 88.94% accuracy with 73% fewer parameters than ResNet-50, with 2.4 % accuracy reduction, while 42% faster training and 73% smaller model size were achieved via GoogLeNet. This feature of GoogLeNet makes it suitable for deployment in clinical settings in resource-limited regions. Multi-scale feature extraction by the Inception module proved well-suited for skin cancer analysis, where lesions are present in varying sizes and patterns [27].

Lastly, a clinically significant sensitivity-specificity trade-off was demonstrated by SqueezeNet, with the highest sensitivity of 92.7% despite having a low overall accuracy of 85.45%. SqueezeNet is considered ideal for applications in which the goal is to minimize missed cancer cases, given its high sensitivity and low specificity profile. It is also regarded as suitable for underserved populations due to its ultra-lightweight design and 5MB file size, enabling 1.4-minute training via transformative mobile telemedicine applications [31].

Lastly, the interpretable explanations were achieved by all three XAI methods, with all models consistently focusing on clinically relevant features, including irregular borders, color variations, and lesion centers. Whereas occasional attention is also given to artifacts, including ruler markings, highlighting the critical need for validation before clinical deployment [36].

3.6. Benchmarking

ResNet-50 architecture implemented in this study achieved (91.36% accuracy, 0.9721 AUC), which compares favourably with recent dermatological AI studies. Esteva [11] reported an accuracy of 72.1% for multiclass classification of 757 diseases, although the broader scope

of their task limits direct comparison. Haenssle [4] reported 86.6% accuracy for melanoma versus nevus classification, whereas the ResNet-50 model in this study achieved a higher performance, potentially reflecting a more focused binary classification and extensive data augmentation. Brinker [12] reported that deep learning outperformed 136 of 157 dermatologists, though on a different dataset and task.

This study's multi-architecture comparison provides practical deployment guidance that is largely absent from the existing literature, which typically focuses on single architectures. Most studies employ ResNet or Inception variants without a systematic evaluation of efficiency [4,38]. In this study, GoogLeNet achieves 88.94% accuracy with 73% fewer parameters, addressing the critical but under-explored question of accuracy-efficiency trade-offs for resource-constrained deployments.

The comprehensive XAI evaluation across three methods and three architectures exceeds most published work, which typically applies single XAI methods without systematic reporting of success rates or cross-architecture comparisons [39, 40]. This study's 100% XAI success rate across methods demonstrates robust implementability and highlights the practical importance of reporting generation failures, which are often unreported in the literature.

4. CONCLUSION

In conclusion, this research article demonstrates that deep learning methods for skin cancer detection can yield accurate and interpretable results when combined with explainable AI. The ResNet-50 model achieved the highest accuracy and balanced sensitivity and specificity, suggesting applicability in adequately supported environments. The GoogLeNet architecture achieved comparable accuracy but had fewer than half as many parameters as ResNet-50, thereby supporting deployment in less well-supported environments. Finally, the SqueezeNet model achieved the highest sensitivity for detecting malignant lesions and, owing to its lightweight design, can deliver results via mobile and telemedicine platforms. The deep learning models produced high-quality results, as assessed by Grad-CAM, LIME, and Occlusion Sensitivity across multiple observations; however, the occasional inclusion of artefacts highlights the need to develop processes to validate outputs from explainable AI methods. These results provide a context for the deployment of those models. Future studies should involve multiclass classification, external validation of deep learning models, clinical comparisons, and expert clinician review of explainable AI explanations.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support provided by Salim Habib University, Karachi, Pakistan, which fully funded this research project. The resources, facilities, and academic environment offered by the University were instrumental in completing this work.

REFERENCES

[1] Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics, 2021. CA: A Cancer Journal for Clinicians, 71(1), 7-33.

[2] American Cancer Society. (2021). Cancer Facts & Figures 2021. Atlanta: American Cancer Society.

[3] Vestergaard, M. E., Macaskill, P., Holt, P. E., & Menzies, S. W. (2008). Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. British Journal of Dermatology, 159(3), 669-676.

[4] Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Reader Study Level-I and Level-II Groups. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8), 1836-1842.

[5] D. Adla, G. V. R. Reddy, P. Nayak, and G. Karuna, "Deep learning-based computer aided diagnosis model for skin cancer detection and classification," *Distributed and Parallel Databases*, vol. 40, no. 4, pp. 717–736, 2022.

[6] T. M. Alam, K. Shaukat, W. A. Khan, I. A. Hameed, L. A. Almuqren, M. A. Raza, M. Aslam and S. Luo, "An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, 2022.

[7] K. I. Chibueze, A. F. Didiugwu, N. G. Ezeji, and N. V. Ugwu, "A CNN based model for heart disease detection," Scientia Africana, vol. 23, no. 3, pp. 429–442, 2024.

[8] S. Panigrahi, J. Das, and T. Swarnkar, "Capsule network based analysis of histopathological images of oral squamous cell carcinoma," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 7, pp. 4546–4553, 2022.

[9] Argenziano, G., Soyer, H. P., Chimenti, S., Talamini, R., Corona, R., Sera, F., ... & Kleine, H. (2003). Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. Journal of the American Academy of Dermatology, 48(5), 679-693.

[10] Kimball, A. B., & Resneck Jr, J. S. (2008). The US dermatology workforce: a specialty remains in shortage. Journal of the American Academy of Dermatology, 59(5), 741-745.

[11] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

[12] Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., ... & von Kalle, C. (2019). Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. European Journal of Cancer, 113, 47-54.

[13] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.

[14] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

[15] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. IEEE International Conference on Computer Vision, 618-626.

[16] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

[17] Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. European Conference on Computer Vision, 818-833.

[18] Fanconic. (2019). Skin Cancer: Malignant vs. Benign. Kaggle. https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign

[19] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1-54.

[20] Rogers, H. W., Weinstock, M. A., Feldman, S. R., & Coldiron, B. M. (2015). Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. JAMA Dermatology, 151(10), 1081-1086.

[21] Barata, C., Ruela, M., Francisco, M., Mendonça, T., & Marques, J. S. (2015). Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal, 8(3), 965-979.

[22] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence, 14(2), 1137-1145.

[23] Gundersen, O. E., & Kjensmo, S. (2018). State of the art: Reproducibility in artificial intelligence. AAAI Conference on Artificial Intelligence, 32(1).

[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[25] Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157-166.

[26] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.

[27] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition, 1-9.

[28] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.

[29] Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678.

[30] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360. Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1-54.

[31] Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Global Health, 3(4), e000798.

[32] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2274-2282.

[33] Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian Journal of Internal Medicine, 4(2), 627.

[34] Nachbar, F., Stolz, W., Merkle, T., Cognetta, A. B., Vogt, T., Landthaler, M., ... & Plewig, G. (1994). The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology, 30(4), 551-559.

[35] McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153-157.

[36] Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Medicine*, 15(11), e1002683.

[37] Carli, P., De Giorgi, V., Chiarugi, A., Nardini, P., Weinstock, M. A., Crocetti, E., ... & Giannotti, B. (2002). Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study. *Journal of the American Academy of Dermatology*, 46(5), 683-689.

[38] Han, S. S., Kim, M. S., Lim, W., Park, G. H., Park, I., & Chang, S. E. (2018). Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7), 1529-1538.

[39] Lucieri, A., Bajwa, M. N., Braun, S. A., Malik, M. I., Dengel, A., & Ahmed, S. (2020). On interpretability of deep learning-based skin lesion classifiers using concept activation vectors. *International Joint Conference on Neural Networks*, 1-10.

[40] Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., ... & Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, 118, 91-96.

https://doi.org/10.31436/iiumej.v27i1.3898

Performance Investigation and Efficiency Enhancement of Eco-Friendly Tin-Based CH3NH3SnI3 Perovskite Solar Cell via SCAPS-1D

MD. MORSALIN[1*], RAGAB A. SAYED[2*], MD. SAZEDUR RAHMAN[1], MD. FERDOUS WAHID[1], MOHAMMAD SALAH[3]

[1]Department of Electrical and Electronic Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur 5200, Bangladesh.
[2]Automotive and Tractors Technology Department, Faculty of Technology and Education, Capital University – Helwan Branch, Cairo, Egypt.
[3]College of Engineering and Technology, American University of the Middle East, Kuwait.

**Corresponding author: mdmorsalin.hstu@gmail.com, ragab.mahmoud@techedu.helwan.edu.eg*

(Received: 12 August 2025; Accepted: 17 November 2025; Published online: 12 January 2026)

ABSTRACT: Halide perovskite materials, particularly lead-based $CH_3NH_3PbI_3$, have garnered significant attention in the PV industry for their exceptional efficiency in solar cell applications. However, due to the toxicity of lead, research interest has shifted toward Sn-based alternatives. This study explores a lead-free Sn-based perovskite solar cell (PSC) with the structure ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni, where $CH_3NH_3SnI_3$ ($MASnI_3$) serves as the absorber material, TiO_2 as the electron transport layer (ETL), Cu_2BaSnS_4 (CBTS) as the hole transporting layer (HTL). Device performance is analyzed using the SCAPS-1D simulation software. The impact of key performance-determining parameters, including the thickness, doping density, and defect density of the absorber, ETL, and HTL, has been accounted for. The proposed PSC architecture, optimized for key performance-determining parameters, achieves a power conversion efficiency PCE (η) of 27.28%, an open-circuit voltage (V_{OC}) of 1.0283 V, a fill factor (FF) of 83.62%, and a short-circuit current density (J_{SC}) of 31.72 mA/cm^2. This study examines the influence of interface defect density, shunt and series resistances, back-contact metal work function, and operating temperature on the performance of PSCs. Furthermore, the analysis includes current density-voltage (J-V) and quantum efficiency (QE) characteristics to provide a comprehensive evaluation of the effectiveness of the proposed PSC.

ABSTRAK*:* Bahan perovskit halida, khususnya $CH_3NH_3PbI_3$ berasaskan plumbum, telah menarik perhatian besar dalam industri fotovolta (PV) berikutan kecekapan tinggi dalam aplikasi sel suria; namun, isu ketoksikan plumbum telah mengalih tumpuan penyelidikan kepada alternatif berasaskan timah (Sn). Kajian ini meneroka sel suria perovskit (PSC) bebas plumbum berasaskan Sn dengan seni bina ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni, di mana $CH_3NH_3SnI_3$ ($MASnI_3$) bertindak sebagai bahan penyerap, TiO_2 sebagai lapisan pengangkut elektron (ETL), dan Cu_2BaSnS_4 (CBTS) sebagai lapisan pengangkut lubang (HTL). Prestasi peranti dianalisa menggunakan perisian simulasi SCAPS-1D dengan mengambil kira parameter penentu prestasi utama, termasuk ketebalan, ketumpatan pendopan, dan ketumpatan kecacatan bagi lapisan penyerap, ETL dan HTL. Seni bina PSC yang dicadangkan, selepas pengoptimuman parameter, mencapai kecekapan penukaran kuasa (PCE) sebanyak 27.28%, voltan litar terbuka (V_OC) 1.0283 V, faktor pengisian (FF) 83.62%, dan ketumpatan arus litar pintas (J_SC) 31.72 mA/cm². Kajian ini turut menilai pengaruh ketumpatan kecacatan antara muka, rintangan siri dan pirau, fungsi kerja logam sentuhan belakang, serta suhu operasi terhadap prestasi PSC. Di samping itu, analisis ciri

ketumpatan arus–voltan (J–V) dan kecekapan kuantum (QE) disertakan bagi memberikan penilaian menyeluruh terhadap keberkesanan sel suria perovskit yang dicadangkan.

KEYWORDS: *Perovskite Solar Cell, CBTS, $MASnI_3$, SCAPS-1D, Solar power.*

1. INTRODUCTION

In the current global energy landscape, fossil fuels remain the dominant source of electricity generation, accounting for a substantial share of worldwide energy demand [1]. Despite their widespread use, the continued reliance on fossil fuels poses several problems, including resource depletion, greenhouse gas emissions, and environmental degradation [2]. These issues have intensified the need for a transition to cleaner, more sustainable energy alternatives. Among the various renewable energy sources, such as wind energy, hydro energy, and ocean energy, solar energy has emerged as a highly promising and scalable solution [3]. Its vast availability, minimal environmental impact, and potential to significantly reduce carbon emissions make it an attractive option for meeting the growing global energy demand [4]. Solar energy, harnessed from the sun's radiation, presents several distinct advantages over other renewable sources. Its abundant availability, geographical accessibility, and potential for decentralized power generation make it a highly attractive option for sustainable energy development [5-7]. With continuous advancements in photovoltaic (PV) technologies, the efficiency and cost-effectiveness of solar energy systems have improved substantially [8,9].

Among the emerging PV technologies, perovskite solar cells (PSCs) have attracted considerable attention in photovoltaic (PV) research owing to their exceptional efficiency in converting photon energy into electricity, as well as their beneficial electrical properties [10]. Perovskite materials are distinguished by their excellent properties, including longer diffusion lengths, high absorption coefficients, superior charge carrier mobility, and relatively low exciton binding energy, making them highly suitable for solar energy applications [11-13]. In the middle of the 1990s, Mitzi et al. first discovered the ability of electricity to convert light in halide perovskites, applied to LEDs [14].

In 2009, Kojima et al. first illustrated the dye-sensitized solar cell [15] structures with $CH_3NH_3PbI_3$ and achieved an efficiency of 3.8% [16]. Halide perovskites are materials characterized by the formula "ABX_3" [17], where "A" represents monovalent cations (Cs^+, $CH_3NH_3^+$, $CH(NH_2)_3^+$), "B" denotes cationic metals like Sn^{2+} or Pb^{2+}, and "X" corresponds to halogen ions (Cl^-, I^-, Br^-). $MAPbI_3$ (methylammonium lead triiodide) is the predominant perovskite material used in solar cells (SCs). However, the presence of lead raises significant concerns regarding its toxicity [18]. In contrast, tin-based perovskites, such as methylammonium tin iodide (MASnI3), offer a more environmentally friendly and safer alternative for commercialization due to tin's non-toxicity. Furthermore, tin-based perovskites also exhibit properties comparable to those of lead-based counterparts while reducing expenses, facilitating a seamless adaptation of existing manufacturing processes [19, 20]. Replacing lead with tin creates a material with a specific light absorption property (bandgap) very close to the actual value for SCs, as claimed by the Shockley-Queisser limit [20]. This characteristic makes tin-based perovskites promising for making high-efficiency SCs, especially when combined with other perovskite materials [19].

In this investigation, a tin-based and lead-free PSC device structure of $ITO/TiO_2/CH_3NH_3SnI_3/CBTS/Ni$ is proposed and simulated using SCAPS-1D software. It shows the short-circuit current density (J_{SC}) 31.7233 mA/cm^2, efficiency (η) 27.28%, fill factor (FF) 83.62%, and open-circuit voltage (V_{OC}) 1.0283 V. These findings demonstrate superior

performance compared to previously reported studies [21-24]. Here, a transparent conducting oxide, indium-doped tin oxide (ITO), [25], is considered as a transparent glass in front contact, while Nickel (Ni) is considered as a back metal contact in the PSC device. Additionally, TiO2 is used as an electron transport layer (ETL) because its conduction band (CB) [26] is similar to that of perovskites, and it has good optical and chemical stability, non-toxicity, and an easy manufacturing process [21,22].

A lead-free tin-based perovskite ($CH_3NH_3SnI_3$) is used as an absorber material because of its appropriate bandgap, non-toxicity, and inexpensive, earth-abundant, high absorption coefficient [21]. Copper barium tin sulfide (CBTS) is taken as a hole transport layer (HTL) because of its hole mobility, high stability, and excellent bandgap alignment with the $CH_3NH_3SnI_3$ absorber. Its high hole mobility efficiently transports positive charges (holes) from the light-absorbing layer (perovskite) toward the electrode [27]. This improves overall device efficiency by minimizing charge recombination within the cell. The proper selection of ETLs and HTLs enhances the performance and efficiency of the SC device. In this study, we examined the effects of varying the thickness, donor and doping densities, and trap densities (Nt) of the HTL, ETL, and absorber layers, as well as interface defect densities, on the quantum efficiency (QE) and current-voltage (J-V) characteristics. Additionally, we analyzed the effects of operating temperature, shunt and series resistances, and the work function (WF) of the back contact metal. Furthermore, we investigated how variations in these parameters affect the performance of the PSC device without an HTL.

2. METHODOLOGY

2.1. SCAPS-1D numerical simulation

The SCAPS-1D (version 3.3.10) software is used in this study to investigate the photovoltaic (PV) parameters. The software is provided by Prof. Marc Burgelman et al. at the ELIS department [25], Ghent University, Belgium [25,26]. SCAPS-1D software operates based on solving numerical equations, such as the continuity equation for holes and electrons [30] in the CB and VB [21] and the Poisson's equation. The Poisson's (Eq. (1)) [31] and continuity equations [32] (Eq. (2) for electrons and Eq. (3) for holes) are depicted below:

$$\frac{d^2}{dx^2}\psi(x) = \frac{q}{\varepsilon_0 \varepsilon_r}\left[p(x) - n(x) + N_D - N_A + \rho_p + \rho_n\right] \quad (1)$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + (G_n - R_n) \quad (2)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q}\frac{\partial J_p}{\partial x} + \left(G_p - R_p\right) \quad (3)$$

where, ε_0 stands for the permittivity of free space, ψ represents the electric potential, ε_r stands for the relative permittivity, p(x) and n(x) represent the hole and electron densities respectively, ρ_n represents electron distribution and ρ_p represents the hole distribution, N_D and N_A represent the donor and acceptor densities, respectively, J_p and J_n stand for the current densities [33] for electrons and holes, respectively, G_p and G_n stand for the holes and electrons generation, respectively, R_p and R_n represent the holes and electrons recombination, respectively, and q represents the electric charge. The simulation is conducted under illumination using an AM 1.5G solar spectrum at a power density of 1000W/m^2.

https://doi.org/10.31436/iiumej.v27i1.3898

Table 1. Input parameters of the ITO, Absorber, HTL, and ETL layers.

Parameters	ITO [26]	TiO_2 [26]	$CH_3NH_3SnI_3$[21]	CBTS [26]
Thickness (μm)	0.5	0.03	0.7	0.1
E_g (eV)	3.5	3.2	1.3	1.9
χ (eV)	4	4	4.2	3.6
ε_r (relative)	9	9	8.2[24]	5.4
N_c (cm^{-3})	2.2×10^{18}	2×10^{18}	1×10^{18}	2.2×10^{18}
N_v (cm^{-3})	1.8×10^{19}	1.8×10^{19}	1×10^{18}	1.8×10^{19}
μ_n ($cm^2V^{-1}s^{-1}$)	20	30	1.6	30
μ_p ($cm^2V^{-1}s^{-1}$)	10	10	1.6	10
N_A (cm^{-3})	0	0	1×10^{17}	1×10^{19}
N_D (cm^{-3})	1×10^{21}	6×10^{19}	0	0
N_t (cm^{-3})	1×10^{15}	1×10^{15}	1×10^{15}	1×10^{15}

Here, **E_g:** Bandgap, **χ:** Electron Affinity, **ε_r:** Dielectric permittivity, **N_c:** CB effective DOS, **N_v:** VB effective DOS, **μ_n:** Electron mobility, **μ_p:** Hole mobility, **N_A:** Shallow uniform acceptor density, **N_D:** Shallow uniform donor density, and **N_t:** Defect density.

2.2. Device structure and simulation parameters

The proposed PSC device structure is depicted in Figure 1. The PSC structure consists of ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni, where $CH_3NH_3SnI_3$ is used as an absorber, p-type CBTS is used as an HTM, and n-type TiO_2 is used as an ETM. The energy band diagram (EBD) of the proposed PSC device is shown in Figure 2.

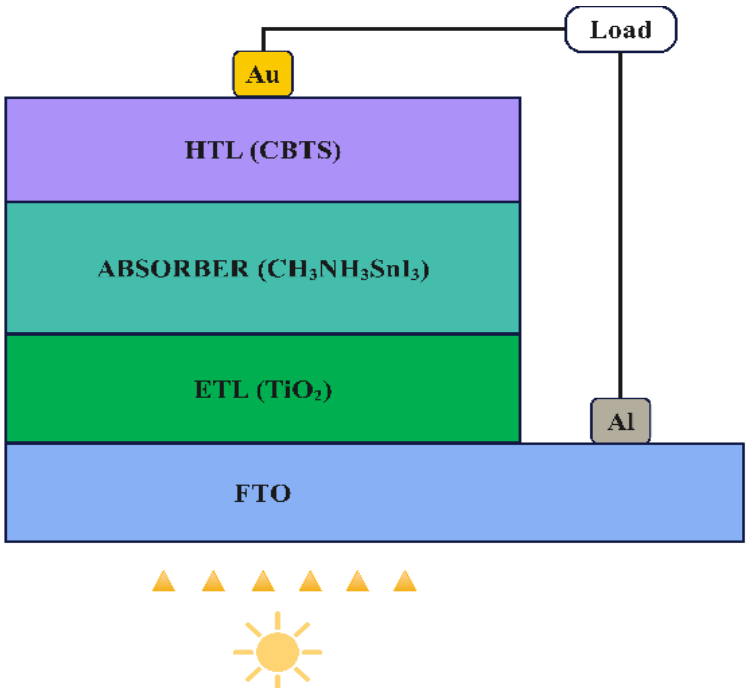


Figure 1. The device structure of the proposed PSC.

When light shines on the SC [34], it creates energy levels called quasi-Fermi levels [26] (F_p and F_n). This confirms that the light successfully generates electron-hole pairs in the PSC. There is a slight offset in the CB/VB of the absorber layer observed from the energy band diagram in Figure 2, which facilitates the generation of electron-hole pairs [35] in the SC. There is also a slight offset in the CB/VB between the absorber layer, the HTL, and the ETL. However, an excessively large offset can introduce resistance that reduces the cell's FF. A mismatch between the materials' energy levels likely accounts for the larger difference.

https://doi.org/10.31436/iiumej.v27i1.3898

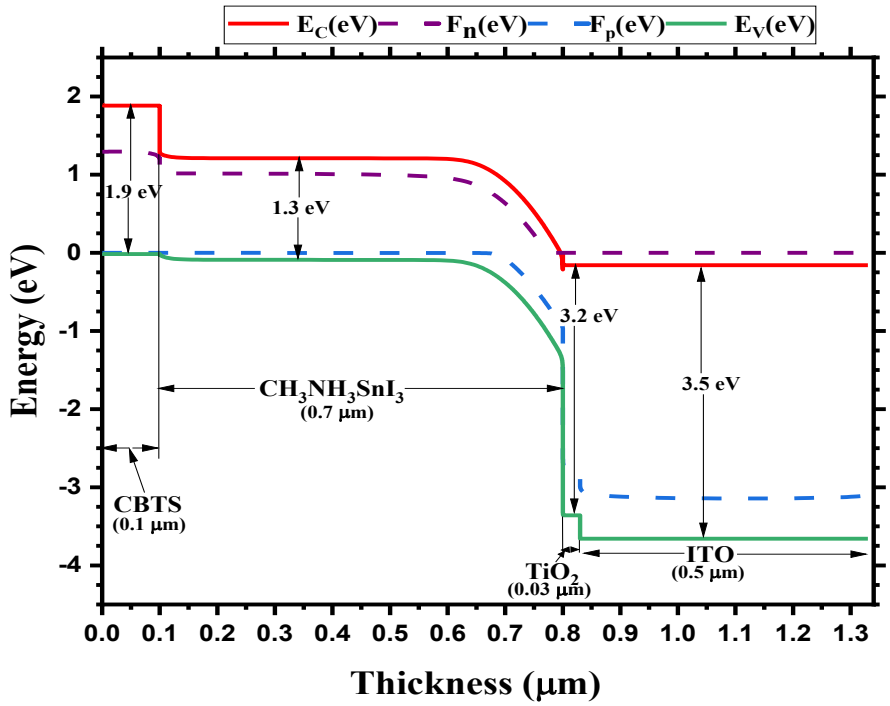


Figure 2. The energy band diagram (EBD) of the proposed PSC.

Table 2. Parameters for interface defects [36].

Parameters	CBTS/CH3NH3SnI3 interface	CH3NH3SnI3/TiO2 interface
Defect type	Neutral	Neutral
Capture cross-sector: electron	1×10-15	1×10-15
Capture cross-sector: hole	1×10-15	1×10-15
Energetic distribution	Single	Single
Reference for defect energy level	Above the highest EV	Above the highest EV
Total density (cm-2)	1×10-11	1×10-11

The input parameters that are used in this simulation, such as thickness, bandgap, electron affinity, mobility for electrons and holes, relative permittivity, acceptor and donor densities, the effective density of states (DOS) in the valence bands (VB), and conduction bands (CB), total defect densities of HTL, absorber, and ETL are given in Table 1. The interfacial defects between absorber/HTL and ETL/absorber are also presented in Table 2. The performance of the designed SC is evaluated across different temperatures to investigate how the device's efficiency varies with temperature. However, a temperature of 300 K (room temperature) is considered throughout the study as the standard reference, since it represents the typical operating environment of PSCs under normal laboratory and ambient conditions, ensuring consistency and comparability of performance parameters [37].

3. RESULTS AND DISCUSSIONS

3.1. Absorber layer thickness variation effect on PV performance

The absorber-layer thickness in perovskite solar cells governs light absorption and recombination dynamics, thereby indirectly influencing the diffusion length and lifetime of charge carriers and ultimately determining device performance [38,39]. An overly thick absorber layer leads to increased recombination and series resistance, whereas an insufficiently thick layer compromises photon absorption and photocurrent [40-43]. Therefore, optimizing

the absorber layer thickness is essential for maximizing SC efficiency. This study investigates the effects of varying absorber layer thickness from 0.1µm to 1.8µm, with Nt = 1015 cm-3, on the performance of the proposed solar cell device for configurations with and without an HTL, as illustrated in Figure 3(a). The η and Jsc increased with thickness up to 0.7 µm. At this thickness, the η and Jsc were 27.28% and 31.72 mA/cm² with HTL, and 24.63% and 29.44 mA/cm² without HTL. This increase is attributed to enhanced light absorption, resulting in the generation of electron-hole pairs within the absorber layer [44]. Increasing the absorber thickness above 0.7 µm leads to recombination and series resistance that limit further improvements in η and Jsc. Figure 3(a) indicates that the FF initially increases with small thickness increments but decreases with further increases in thickness for both configurations, due to increased series resistance. Moreover, the open-circuit voltage of the device with HTL decreases with increasing thickness, attributed to increased recombination of free charge carriers [45]. In the absence of an HTL, Voc increases. This is attributed to the fact that as the absorber thickness increases, both the photon absorption capacity and the rate of charge-carrier generation rise, facilitating their transport to the electrode with reduced resistance [46]. Consequently, the findings establish 0.7 µm as the optimal absorber thickness for further investigation.

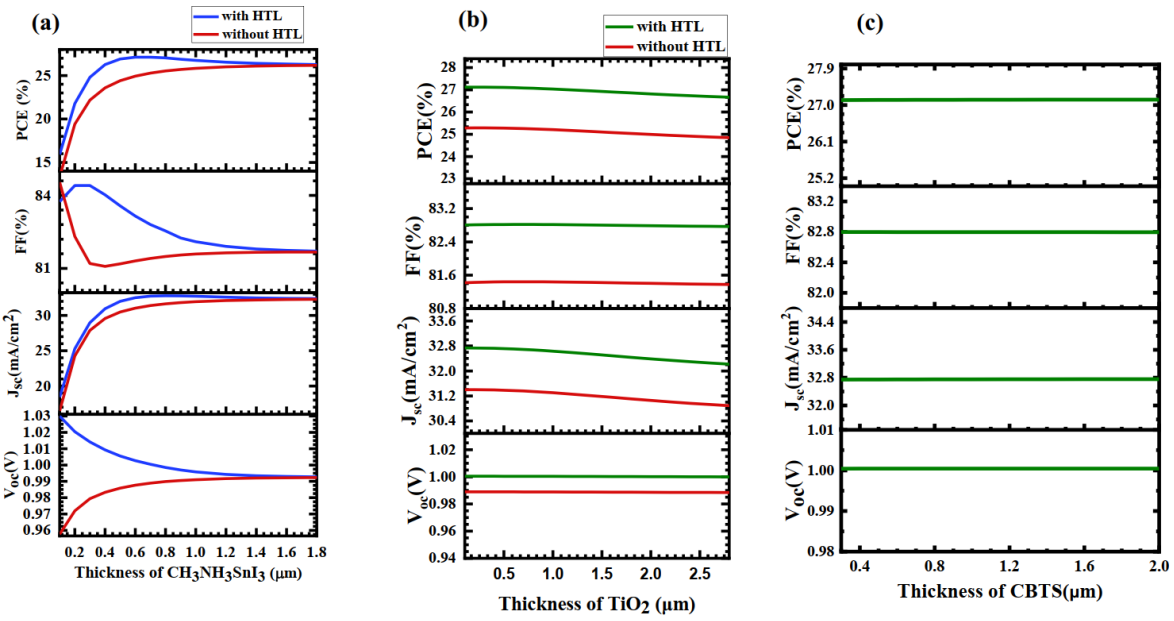


Figure 3. (a) the absorber layer thickness change, (b) the TiO_2 (ETL) layer thickness change, and (c) the HTL layer thickness change on PV performance for both devices with and without HTL.

3.2. ETL thickness variation effect on PV performance

ETL plays a crucial role in facilitating electron transfer from the absorber layer to the electrode. Figure 3(b) illustrates the influence of ETL thickness on the performance of devices with and without HTL. ETL thickness was systematically varied from 0.1 µm to 3 µm, with all other device parameters held constant. The analysis revealed insignificant variation in PV output parameters across the examined range of ETL thicknesses. However, a slight reduction in both Jsc and η is observed as the ETL thickness increases, regardless of HTL presence. This behavior is primarily due to the lower absorption coefficient and decreased light transmittance through the device as the ETL layer increases in thickness [26,47]. Therefore, based on the observed results, a thickness of 0.03 µm is selected as the optimal ETL thickness.

https://doi.org/10.31436/iiumej.v27i1.3898

3.3. HTL thickness variation effect on PV performance

This study evaluates the performance of the proposed PSC by analyzing the impact of HTL thickness variation on output parameters, emphasizing its role in facilitating efficient hole collection from the absorber layer, as depicted in Figure 3(c). Variations in HTL thickness from 0.3 µm to 2 µm produced negligible changes in PV output parameters. This is attributed to the high absorption coefficient of CH3NH3SnI3 and its reduced sensitivity to variations in HTL thickness [48,49]. However, a thicker HTL enhances photon collection in the perovskite layer; therefore, an optimized HTL thickness of 0.1 µm was chosen for subsequent optimization, consistent with the initial study parameters.

3.4. Absorber acceptor and defect density variation effect on PV performance

To determine the effect of absorber acceptor density (N_A) on the PV performance, the N_A of the $CH_3NH_3SnI_3$ layer was changed between 10^{14} cm^{-3} and 10^{18} cm^{-3}. Figure 4(a) shows the changes in J_{SC}, η, V_{OC}, and FF of devices with and without an HTL absorber as the absorber NA varies. As delineated in Figure 4(a), V_{OC} increases with the increase in N_A and is obtained at 1.085V with the HTL and 1.073V without the HTL device at the N_A of 10^{18} cm^{-3}. Conversely, η and FF increase notably with NA in both configurations, with and without HTL. For the HTL-equipped device, FF increased from 79.65% to 85.730% and η from 25.08% to 27.65%, and for the configuration without HTL, FF increased from 80.00% to 85.18% and η from 25.04% to 24.54%.

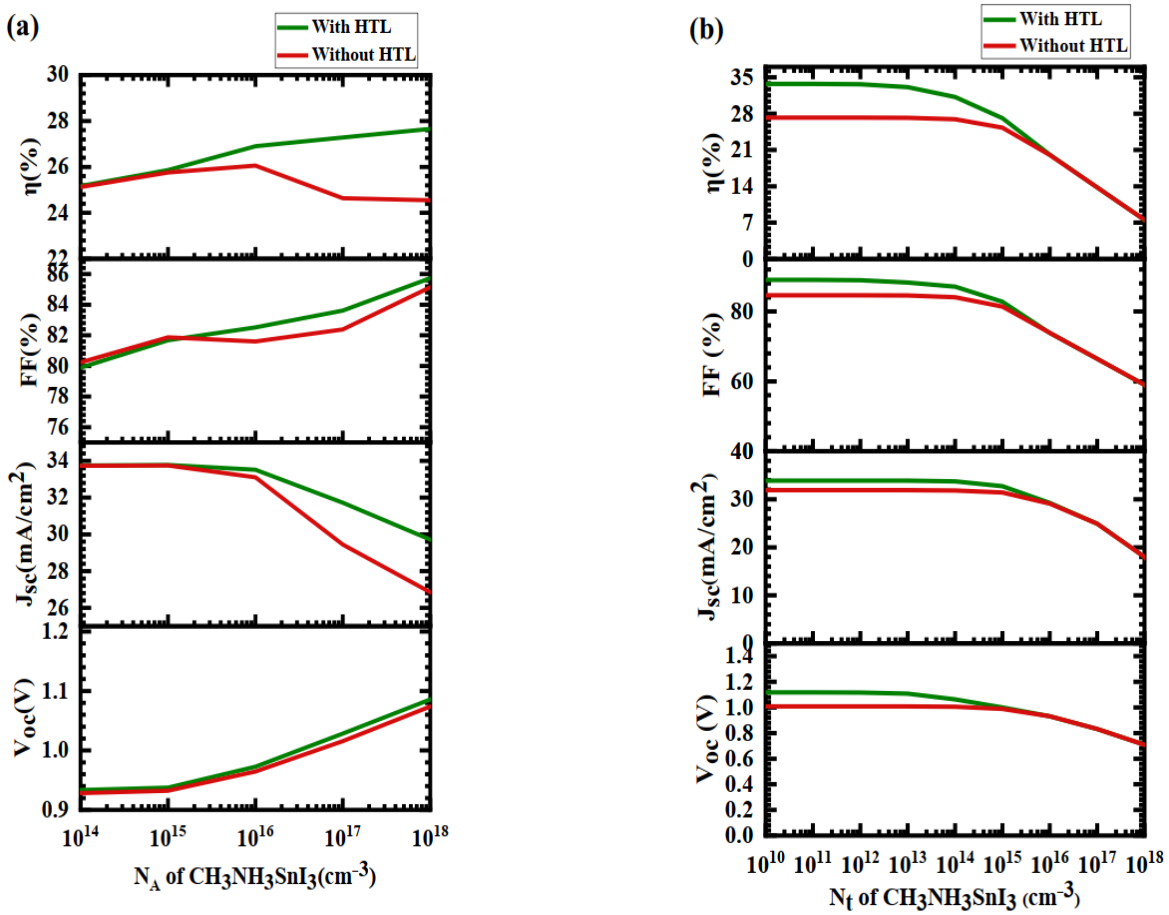


Figure 4. The effect of (a) acceptor density (N_A) and (b) defect density (Nt) of the absorber layer on PV performance for both devices with and without HTL

With the increase in N_A, the reverse saturation current decreases, which increases V_{OC} [21] and hence contributes to FF and η enhancement [36]. The J_{SC} decreases with increasing absorber-layer NA, decreasing from 33.747 mA/cm2 to 29.712 mA/cm2 for the HTL device and from 33.72 mA/cm2 to 26.849 mA/cm^2 for the device without HTL. This decrease in J_{SC} is because the higher carrier density will increase the recombination process and reduce the probability of collecting electrons generated from photons [50]. So, to get the best performance of η= 27.28%, J_{SC}= 31.7233 mA/cm^2, FF= 83.62%, and V_{OC}= 1.0283 V, the optimum value of N_A is taken as 1×10^{17} cm^{-3}. Similarly, the influence of absorber defect density (N_t) has been investigated in this study, as shown in Figure 4(b). The N_t in the absorber layer of an SC significantly affects its performance. Defects in this layer can trap charge carriers, leading to recombination and reduced efficiency [51]. The η, J_{SC}, FF, and V_{OC} decreased considerably with increasing Nt in the absorber for both configurations with and without HTL, as shown in Figure 4(b). At defect densities ranging from 1010 cm-3 to 1013 cm-3, device performance remained relatively unchanged. This suggests that a certain level of defects can be tolerated within the absorber layer without significantly impacting device functionality. However, when N_t exceeds 10^14 cm-3, the devices' output performance begins to decline gradually. This decline is primarily attributed to the Shockley-Read-Hall (SRH) non-radiative recombination process [26,52]. The equations define it.

$$R_{SRH} = \frac{[n_p - n_i]}{[\tau\left(\mathrm{p+n+}\frac{2n_i(E_i-E_t)}{kT}\right)]} \tag{4}$$

$$\tau = \frac{1}{[\sigma\times \mathrm{Nt}\times V_{th}]} \tag{5}$$

Here, σ stands for the charge carriers' capture cross-section, τ for their lifetime, and Vth for their normal velocity. By increasing the crystallinity of the crystalline perovskite and utilizing suitable processing settings that can lower SRH recombination [26,53], it is possible to decrease the number of defects in the material. In this analysis, the optimized N_t is set to 1015 cm-3 for subsequent analysis.

3.5. ETL Donor and defect density variation effect on PV performance

To observe the impact of donor concentration (N_D) variation of ETL on the output performance of the PSC device, N_D of ETL varied from 10^{14} cm^{-3} to 10^{20} cm^{-3}. Figure 5(a) represents the parameters, including η, J_{SC}, FF, and V_{OC}, as a function of the donor density (N_D) of the ETL layer. No significant change in the output parameters was observed with variations in the ETL ND; however, all parameters increased slightly for both configurations with and without an HTL. Because a larger value of N_D of ETL makes the charge carrier extraction and transmission easier in the absorber/ETL interface [54]. The V_{OC} is constant and independent of N_D because the recombination rate is unaffected [54]. In these scenarios, we considered the optimized N_D of ETL to be 6×10^{19} cm^{-3} with the consideration of CB density of states as Table I and shows the output parameters η= 27.28%, J_{SC}= 31.7233 mA/cm^2, FF= 83.62%, and V_{OC}= 1.0283 V. Figure 5(b) showcases the influence of defect density (N_t) of the TiO_2 (ETL) on the output parameters of devices with and without HTL. Interestingly, as N_t of ETL increased across a range of 10^{10} cm^{-3} to 10^{18} cm^{-3}, all the performance parameters remained relatively unchanged for both configurations, as shown in previous works [26,55]. Consequently, we selected 10^{15} cm^{-3} as the optimal N_t for the ETL in this study.

https://doi.org/10.31436/iiumej.v27i1.3898

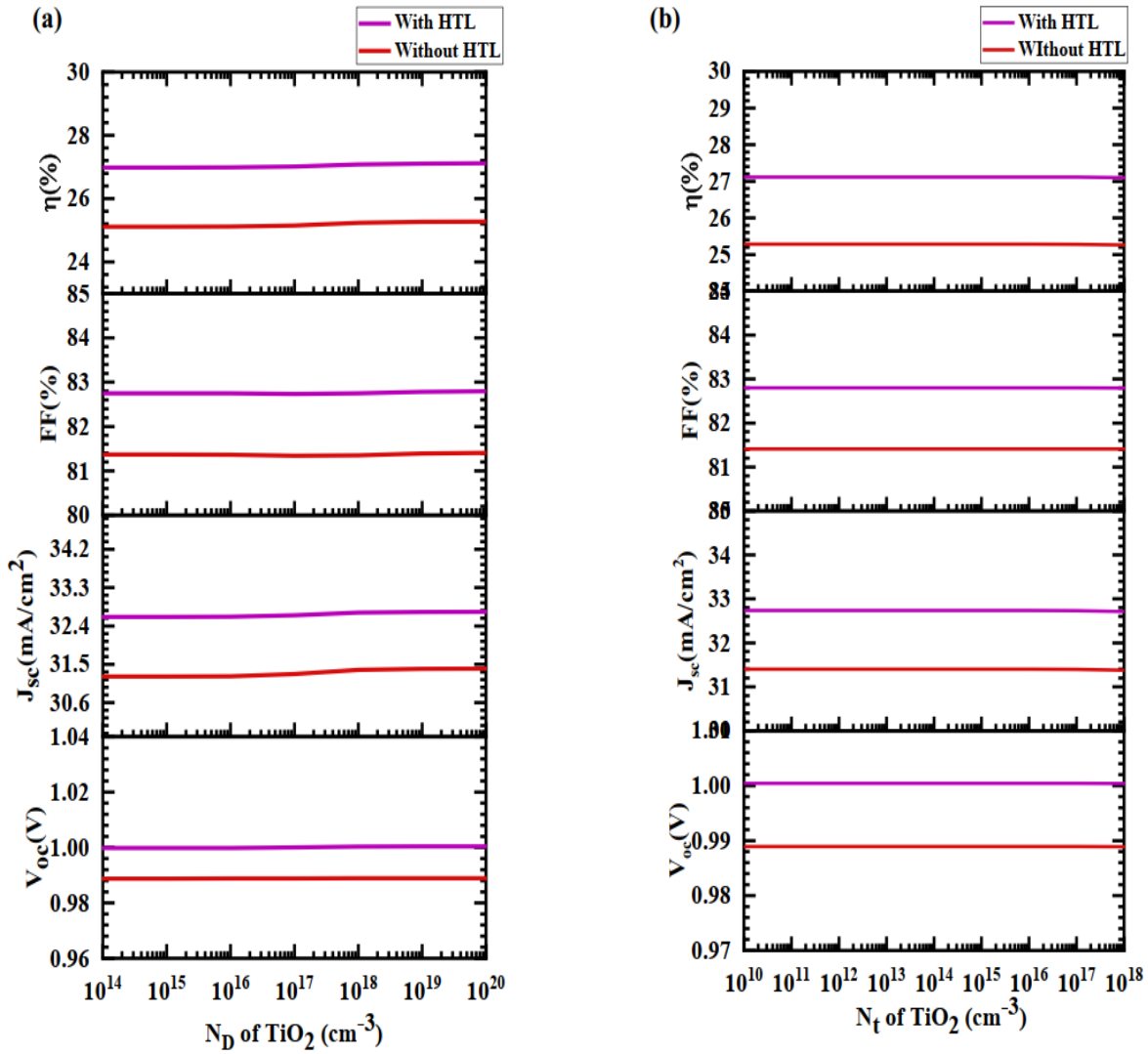


Figure 5. (a) Donor concentration (N_D) and (b) defect density (N_t) of the ETL on PV performance for the devices with and without HTL.

3.6. HTL Acceptor and defect density variation effect on PV Performance

The acceptor density (N_A) in the HTL of SC refers to the concentration of acceptor atoms or molecules within the HTL material. Acceptor dopants introduce additional energy levels near the VB of the semiconductor, providing free positive charge carriers [36]. Figure 6(a) delineates the effect of N_A of the HTL on the PV performance of the PSC as N_A is varied from 10^{10} cm^{-3} to 10^{20} cm^{-3}. In this study, the J_{SC}, FF, V_{OC}, and η remained essentially constant up to 10^{16} cm^{-3} of HTL's N_A. When the value of N_A of HTL exceeds 10^{16} cm^{-3}, all the parameters slightly increase.

https://doi.org/10.31436/iiumej.v27i1.3898

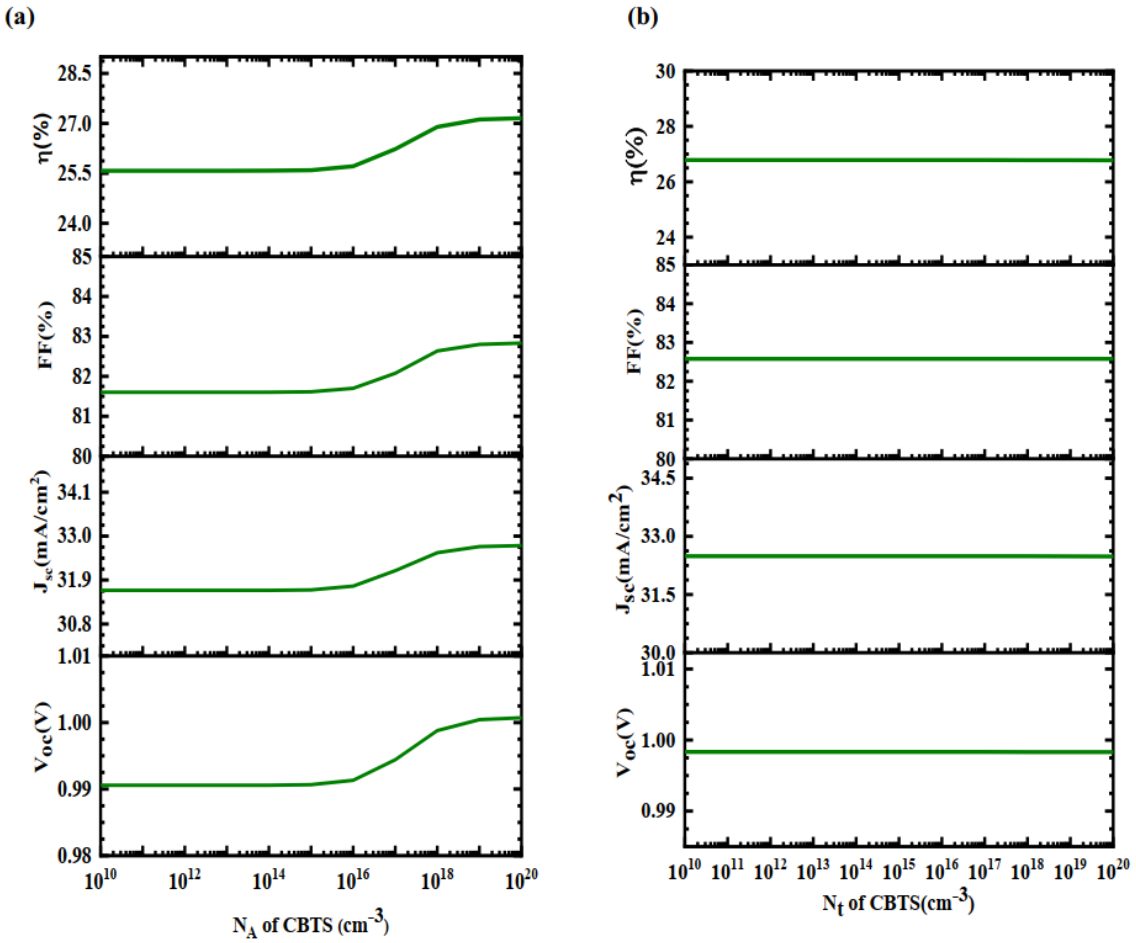


Figure 6. (a) N_A and (b) N_t of HTL on PV performance

Here, the built-in voltage at the absorber/HTL interface increases as the N_A of the HTL increases because the VB of the HTL approaches the quasi-Fermi level [56]. As the N_A of HTL increases, the rate of recombination of electron-hole pairs decreases, which in turn reduces the series resistance and increases the shunt resistance; as a result, FF increases and enhances the overall efficiency [56]. To improve performance, we select the NA of HTL as 1019 cm-3 for subsequent simulations. Similarly, Figure 6(b) illustrates the changes in output parameters as the defect density (N_t) of the CBTS is increased from 10^{10} cm^{-3} to 10^{20} cm^{-3}. Interestingly, η, FF, V_{OC}, and J_{SC} exhibit minimal variation across the entire range of N_t variation. This finding aligns with prior research on CBTS HTL, reinforcing the stability of PV parameters irrespective of N_t fluctuations [26,57-59]. Thus, we determine the primary value of 1015 cm-3 as the optimal Nt for the HTL layer in this study.

3.7. Interface defect density variation effect on PV parameters

This simulation investigated how the PV performance of the proposed SC is affected by defect density (Nt) at the critical TiO2/CH3NH3SnI3 and CH3NH3SnI3/CBTS interfaces, as shown in Figs. 7(a) and 8(b). The interface defect densities were varied across a wide range, from 10^{10} cm^{-2} to 10^{16} cm^{-2} for $TiO_2/CH_3NH_3SnI_3$ interface and from 10^{7} cm^{-2} to 10^{16} cm^{-2} for $CH_3NH_3SnI_3$/CBTS interface. A consistent pattern emerged from the data analysis: FF, η, VOC, and JSC decreased in parallel with increasing N_t at both interfaces. Higher concentrations of interfacial defects are associated with increased recombination rates, which may be the cause of the reduction in PV performance [60]. Interestingly, our findings indicate that the N_t at the

https://doi.org/10.31436/iiumej.v27i1.3898

$CH_3NH_3SnI_3$/CBTS interface exerts a greater influence on device efficiency than at the TiO_2/$CH_3NH_3SnI_3$ interface. Consequently, we maintained a trap density of 10^{10} cm^{-2} at both interfaces, consistent with our initial selection. This choice ensures optimal device performance while mitigating the adverse effects of interfacial defects on recombination and charge transport within the SC.

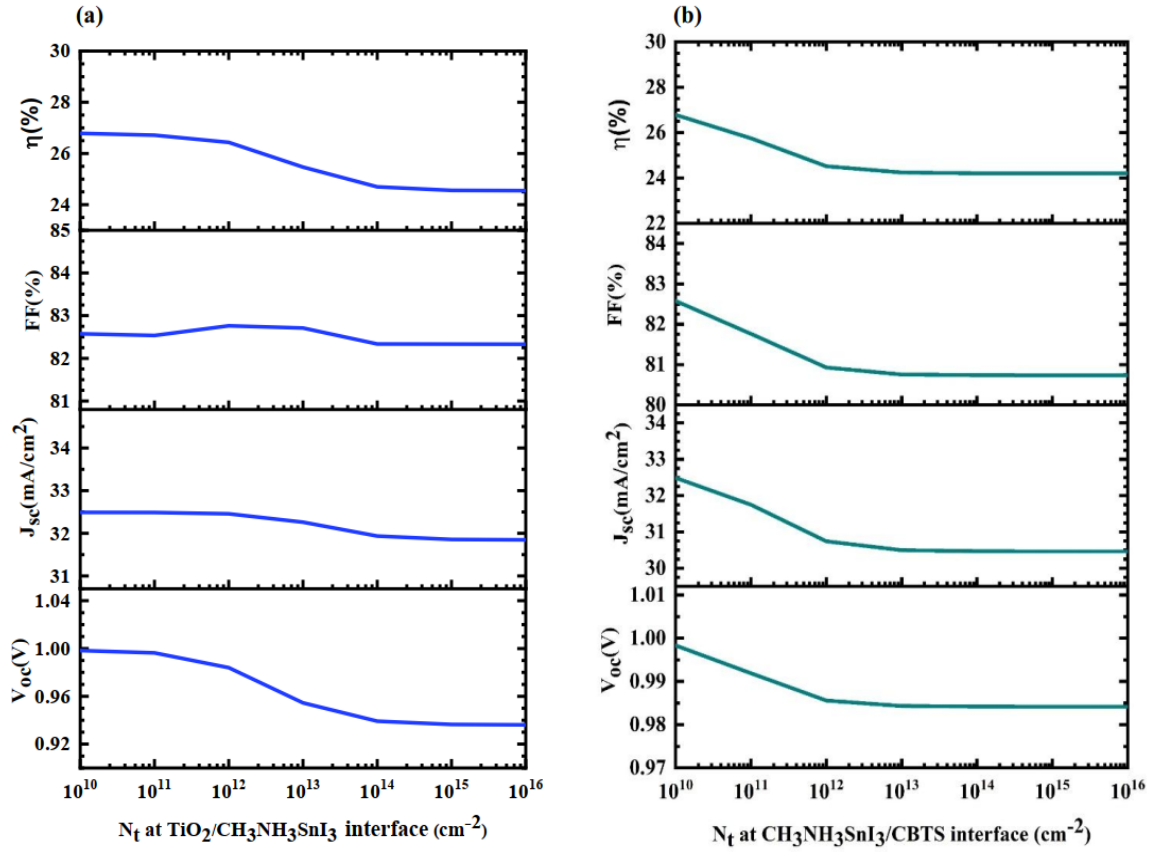


Figure 7. The influence of N_t variation (a) at the TiO_2/$CH_3NH_3SnI_3$ interface and (b) at the $CH_3NH_3SnI_3$/CBTS interface on PV performance

3.8. The J-V and QE features

The J-V (current density-voltage) curve of a solar cell displays distinct characteristics that indicate optimal performance. Figure 8(a) illustrates the anticipated relationship between current density (J) and voltage (V) for both device configurations, with and without HTL. In both devices, the current density decreases with the increase in voltage (V), which shows a similar trend to the previous works [61,62]. Quantum efficiency (QE), also sometimes called the incident photon-to-converted electron (IPCE) ratio, is a measure of how effectively a SC device converts light into usable electrical current [63]. Greater QE of an SC results in more sunlight being converted into electricity. A device might have a high QE for visible light but a lower QE for infrared or ultraviolet light. Figure 8(b) shows the QE features and wavelength (nm) for the device configurations, with and without HTL. Across the wavelength range of 200–1200 nm, the QE curve initially approaches 100% before gradually decreasing to 0% at 960 nm in both device configurations.

https://doi.org/10.31436/iiumej.v27i1.3898

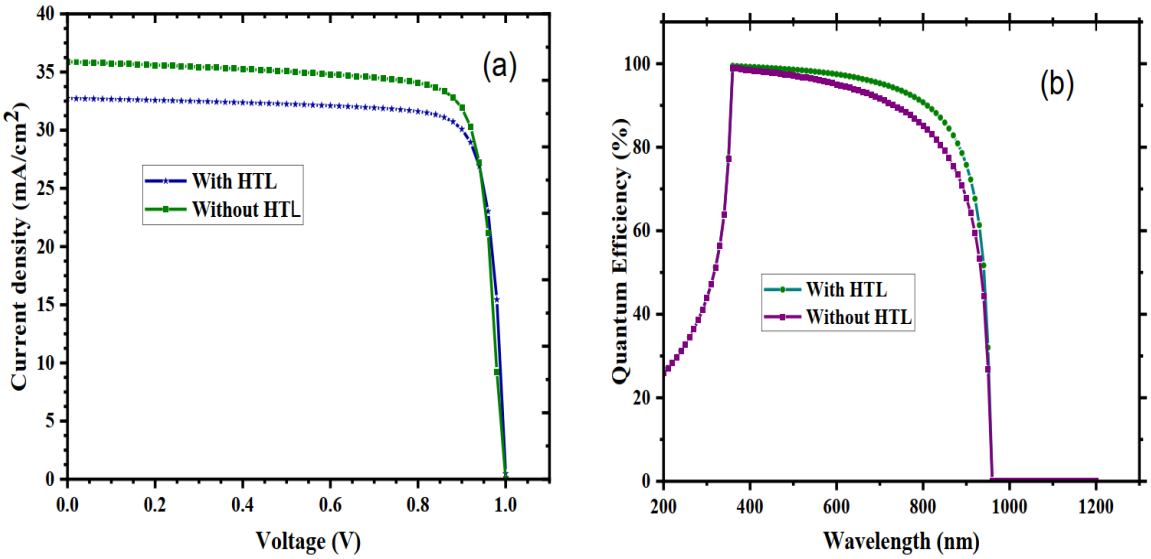


Figure 8. (a) J-V curve and (b) QE curve with and without HTL configurations.

3.9. Series and shunt resistance variation effect on PV performance

Figure 9(a) illustrates the impact of varying series resistance (R_s) on the PV parameters of the devices with and without the HTL.

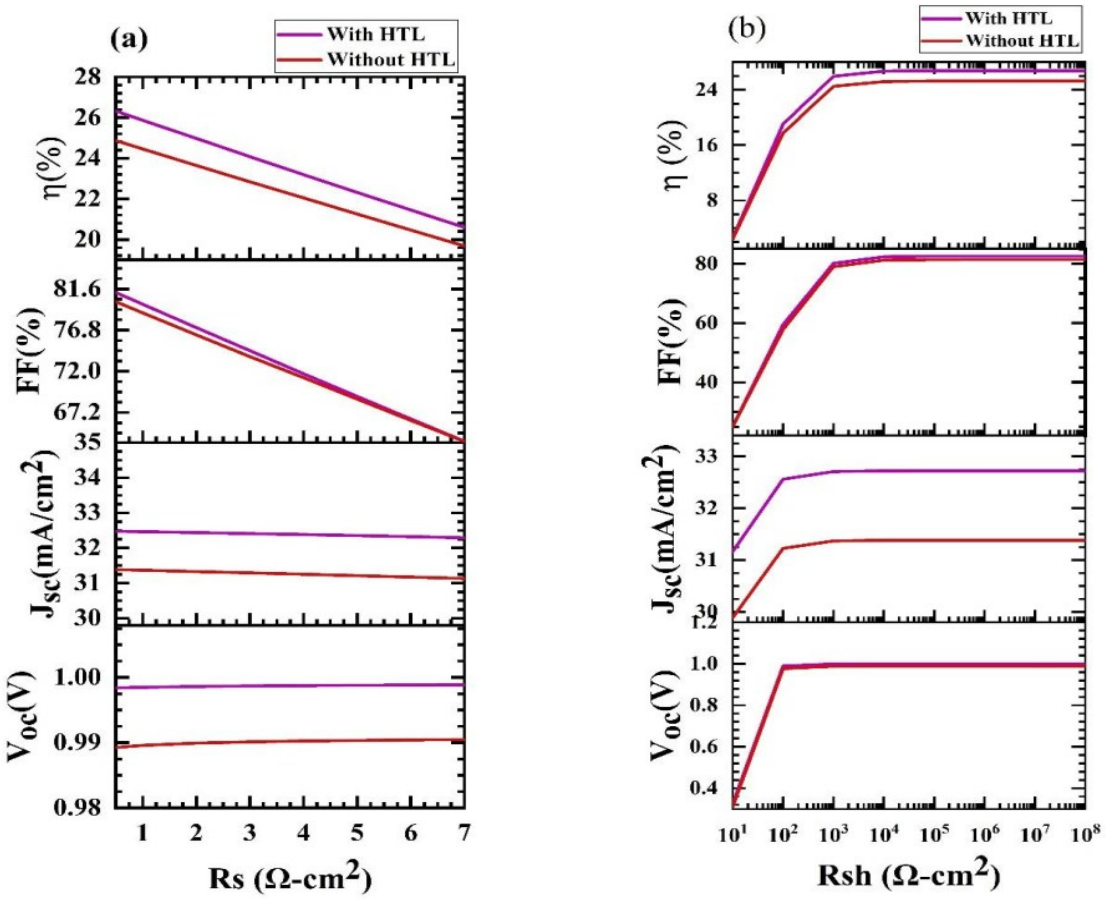


Figure 9. The influence of (a) series resistance and (b) shunt resistance on PV output parameters.

The R_s values range from 0.5 to 7 Ω-cm², with the shunt resistance (Rsh) of 10^5 Ω-cm². The key observation is that increasing R_s has a minimal effect on VOC and little effect on the JSC of the HTL configuration. However, it negatively affects the other parameters. As Rs increases, FF decreases more markedly, from 82.55% to 63.84%, and η decreases from 26.77% to 20.59% for the device with HTL. For the layout without the HTL configuration, the trends are identical. In essence, higher R_s causes internal power losses within the device, which lowers its overall efficiency. Here, R_s is the total amount of resistance between semiconductor-semiconductor and semiconductor-metal at the rear and front contacts of the SC. Conversely, in the active junction, the reverse saturation current [21], which results from manufacturing defects, generates R_{sh}. Again, Figure 9(b) depicts the influence of R_{sh} on the PV parameters for devices with and without HTL. R_{sh} ranges from 10 to 108 Ω-cm², with a series resistance of 0.5 Ω-cm². The results show that increasing R_{sh} increases JSC, VOC, η, and FF for both devices within a specific Rsh range. When R_{sh} exceeded 10^3 Ω-cm², the investigation found a saturation in all output parameters. Higher R_{SH} signifies reduced leakage current within the device [64], leading to improved overall effectiveness of the SC. Thus, our study suggests that both resistances considerably influence the SC's PV performance.

3.10. Temperature and back contact metal variation effect on PV performance

The effect of temperature variation on the photovoltaic performance of the designed PSC is analyzed for both configurations, as shown in Figure 10(a). Except for the J_{SC}, all parameters decreased significantly as the temperature was increased from 275 K to 425 K in both devices with and without an HTL. The steady decline in device performance as temperature rises can be attributed to the concurrent increase in deformation stress and defects, consistent with previous findings [65]. Additionally, high temperatures significantly affect the diffusion length [66] and cause an increase in series resistance (RS), thereby directly affecting the device's η and FF. Again, various metal contacts Gold (Au), Carbon (C), Selenium (Se), Copper (Cu), Nickel (Ni), Palladium (Pd), Iron (Fe), and Platinum (Pt) are used in this simulation to find the effect of back metal variation on PV performance which is shown in Figure 10(b).

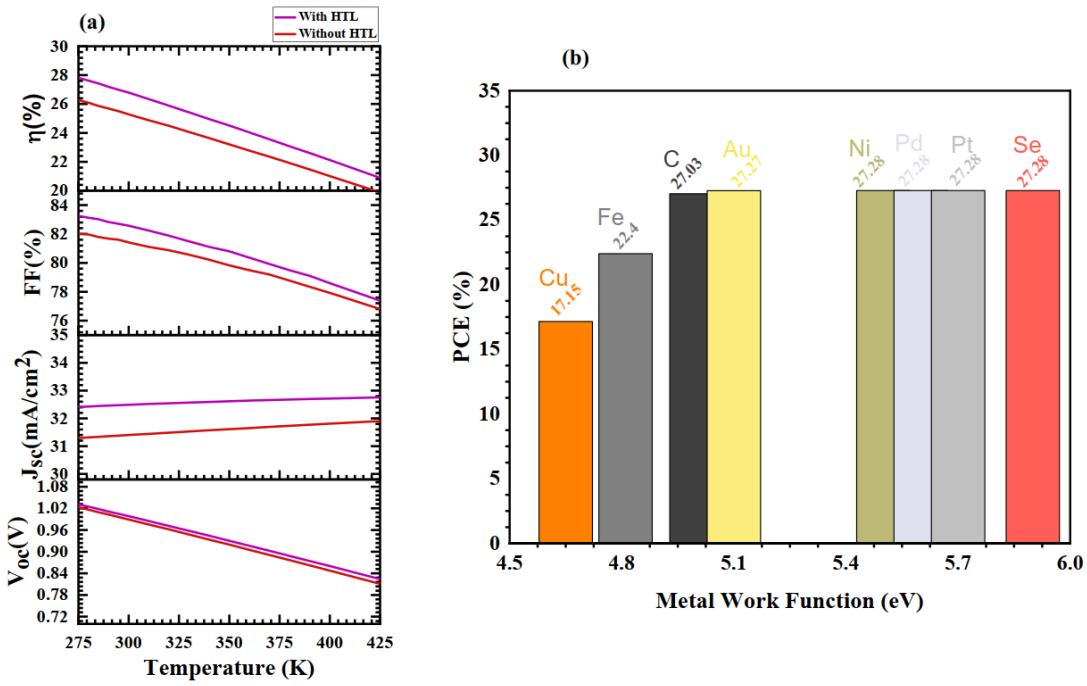


Figure 10. (a) The effect of temperature variation, and (b) Bar diagram for the back metal's WF (eV) variation effect on PV performance.

The bar diagram indicates that the structure's η increases with the metal contact work function (WF) and saturates above 5.1 eV. The enormous barrier formed by the lower WF (eV) materials at the rear side [67] prohibits the entire transit from the absorber to the back metal contact, resulting in low PV output parameters. It becomes apparent that the back contact metal's WF has an essential impact on the performance of SCs. Based on the current investigation outcomes, it is recommended that a back metal with a WF larger than 5.1 eV is necessary to achieve optimal PV performance. Due to its inexpensive cost and suitable WF, nickel (Ni) was chosen as the back metal contact material in this study [68].

3.11. Comparison of the investigated results with previous works

This section compares several optimal solar cell configurations. Table 3 presents a comparative overview of recent theoretical and experimental studies on CH3NH3SnI3-based PSC devices, along with our results.

Table 3. The comparison of $CH_3NH_3SnI_3$-based PSCs

Device structures	Ref.	V_{OC}	J_{SC}	FF	η(%)
(a) FTO/TiO2/CH3NH3SnI3/CZTSe	[69]	0.86	29.45	76.74	19.52
(b) FTO/TiO2/CH3NH3SnI3/CZTSe	[70]	0.96	32.48	76.40	23.86
(c) FTO/Cd0.5Zn0.5S/MASnI3/MASnBr3	[23]	1.055	34.268	69.23	25.05
(d) FTO/ PCBM/MASnI3/CuI	[71]	0.83	29.71	61.23	15.10
(e) FTO/ZnO/CH3NH3SnI3/CuO	[72]	0.989	33.863	66.33	22.21
(f) FTO/TiO2/CH3NH3SnI3/Spiro-OMeTAD,	[25]	0.99	32.95	77.77	25.48
(g) FTO/ C60/MASnI3/CuAlO2	[22]	1.08	25.85	85.84	24.04
(h) ITO/TIO2/CH3NH3SnI3/CBTS	This work	1.028	31.723	83.62	27.28

Compared to others, the investigated device structure ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni obtained the best performance with η=27.28%, J_{SC}= 31.7233 mA/cm^2, FF=83.62%, and V_{OC}= 1.0283 V. This numerical study identified a combination of materials leading to the best performance: CBTS as the HTM, TiO_2 as the ETM, and $CH_3NH_3SnI_3$ as the light absorber. This configuration outperforms the other investigated structures, underscoring the critical role of the light-absorbing layer in SC performance. $CH_3NH_3SnI_3$ has attracted significant research interest owing to its light-absorption properties. However, this study demonstrates that optimizing and carefully selecting both the ETL and HTL, along with their optoelectronic parameters, are crucial for achieving superior solar cell efficiency.

4. CONCLUSION

This research presents a comprehensive simulation approach to investigate the output performance of $CH_3NH_3SnI_3$ ($MASnI_3$)-based PSC. This investigation shows the proposed SC device with the structure ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni, simulated using the SCAPS-1D software package. During optimization, the absorber ($CH_3NH_3SnI_3$) layer, ETL, and HTL's thickness, defect and dopant densities, interface defect densities, shunt and series resistances, and temperature effects on PSC's performance were considered. After optimization, the following conclusions are obtained. The final optimized thicknesses of the absorber ($MASnI_3$) layer, CBTS (HTL) layer, and TiO_2 (ETL) layer are 700 nm, 100 nm, and 30 nm, respectively, for both devices with and without HTL. The optimal doping concentrations of the absorber of the PSC devices using and not using HTL are 10^{17} cm^{-3} and 10^{16} cm^{-3}, respectively, while the optimized ETL donor, HTL acceptor, and absorber defect densities are 6×10^{19} cm^{-3}, 10^{19} cm^{-3}, and 1×10^{15} cm^{-3}. The finally optimized ITO/TiO_2/$CH_3NH_3SnI_3$/CBTS/Ni structure achieved a

JSC of 31.7233 mA/cm2, an η of 27.28%, a VOC of 1.0283 V, and an FF of 83.62%. Furthermore, we examined the series and shunt resistances, the QE curve, the operating temperature, and the illuminated J-V curve. The QE of the optimized configuration remained constant up to 700 nm and fell toward zero at 960 nm, indicating that photon-to-electron conversion occurs across the entire visible spectrum. A comparative analysis with prior studies confirms that the proposed device architecture achieves superior efficiency among CH3NH3SnI3-based PSCs, underscoring its potential for future eco-friendly, high-performance SC applications.

ACKNOWLEDGEMENT

The authors extend their appreciation to the American University of the Middle East in Kuwait for funding this work. This contribution has been instrumental in facilitating the dissemination of the study's findings.

REFERENCES

[1] T. Ahmad and D. Zhang, "A critical review of comparative global historical energy consumption and future demand: The story told so far," Energy Reports, vol. 6, pp. 1973–1991, 2020, doi: https://doi.org/10.1016/j.egyr.2020.07.020.

[2] M. Höök and X. Tang, "Depletion of fossil fuels and anthropogenic climate change—A review," Energy Policy, vol. 52, pp. 797–809, 2013, doi: https://doi.org/10.1016/j.enpol.2012.10.046.

[3] F. A. Akash, S. M. Shovon, W. Rahman, M. A. Rahman, P. Chakraborty, and M. U. Monir, "Greening the grid: A comprehensive review of renewable energy in Bangladesh," Heliyon, vol. 10, no. 5, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27477.

[4] Q. Hassan et al., "The renewable energy role in the global energy Transformations," Renewable Energy Focus, vol. 48, p. 100545, 2024, doi: https://doi.org/10.1016/j.ref.2024.100545.

[5] T.-Z. Ang, M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, and N. Prabaharan, "A comprehensive study of renewable energy sources: Classifications, challenges and suggestions," Energy Strategy Reviews, vol. 43, p. 100939, 2022, doi: https://doi.org/10.1016/j.esr.2022.100939.

[6] A. G. Olabi et al., "Renewable energy systems: Comparisons, challenges and barriers, sustainability indicators, and the contribution to UN sustainable development goals," International Journal of Thermofluids, vol. 20, p. 100498, 2023, doi: https://doi.org/10.1016/j.ijft.2023.100498.

[7] X. S. Musonye, B. Davíðsdóttir, R. Kristjánsson, E. I. Ásgeirsson, and H. Stefánsson, "Integrated energy systems' modeling studies for sub-Saharan Africa: A scoping review," Renewable and Sustainable Energy Reviews, vol. 128, p. 109915, 2020, doi: https://doi.org/10.1016/j.rser.2020.109915.

[8] G. Nagababu, H. Jani, and H. Puppala, "Harnessing Solar Energy for Sustainable Development of Livelihoods," in Handbook of Climate Change Mitigation and Adaptation, M. Lackner, B. Sajjadi, and W.-Y. Chen, Eds., New York, NY: Springer New York, 2025, pp. 1–36. doi: 10.1007/978-1-4614-6431-0_113-2.

[9] S. Zeng, Y. Liu, C. Liu, and X. Nan, "A review of renewable energy investment in the BRICS countries: History, models, problems and solutions," Renewable and Sustainable Energy Reviews, vol. 74, pp. 860–872, 2017, doi: https://doi.org/10.1016/j.rser.2017.03.016.

[10] Q. Lin, A. Armin, R. C. R. Nagiri, P. L. Burn, and P. Meredith, "Electro-optics of perovskite solar cells," Nat Photonics, vol. 9, no. 2, pp. 106–112, 2015, doi: 10.1038/nphoton.2014.284.

[11] S. Khatoon et al., "Perovskite solar cell's efficiency, stability and scalability: A review," Mater Sci Energy Technol, vol. 6, pp. 437–459, 2023, doi: https://doi.org/10.1016/j.mset.2023.04.007.

[12] D. Zhou, T. Zhou, Y. Tian, X. Zhu, and Y. Tu, "Perovskite-Based Solar Cells: Materials, Methods, and Future Perspectives," 2018, Hindawi Limited. doi: 10.1155/2018/8148072.

[13] N. K. Elangovan, R. Kannadasan, B. B. Beenarani, M. H. Alsharif, M.-K. Kim, and Z. Hasan Inamul, "Recent developments in perovskite materials, fabrication techniques, band gap engineering, and the stability of perovskite solar cells," Energy Reports, vol. 11, pp. 1171–1190, 2024, doi: https://doi.org/10.1016/j.egyr.2023.12.068.

[14] D. B. Mitzi, S. Wang, C. A. Feild, C. A. Chess, and A. M. Guloy, "Conducting Layered Organic-inorganic Halides Containing 〈110〉 -Oriented Perovskite Sheets," Science (1979), vol. 267, no. 5203, pp. 1473–1476, Mar. 1995, doi: 10.1126/science.267.5203.1473.

[15] K. Sharma, V. Sharma, and S. S. Sharma, "Dye-Sensitized Solar Cells: Fundamentals and Current Status," 2018, Springer New York LLC. doi: 10.1186/s11671-018-2760-6.

[16] G. Pindolia, S. M. Shinde, and P. K. Jha, "Optimization of an inorganic lead free RbGeI3 based perovskite solar cell by SCAPS-1D simulation," Solar Energy, vol. 236, pp. 802–821, 2022, doi: https://doi.org/10.1016/j.solener.2022.03.053.

[17] A. A. Zaky et al., "Enhancing efficiency and decreasing photocatalytic degradation of perovskite solar cells using a hydrophobic copper-modified titania electron transport layer," Appl Catal B, vol. 284, p. 119714, 2021, doi: https://doi.org/10.1016/j.apcatb.2020.119714.

[18] C. Devi and R. Mehra, "Device simulation of lead-free MASnI3 solar cell with CuSbS2 (copper antimony sulfide)," J Mater Sci, vol. 54, no. 7, pp. 5615–5624, 2019, doi: 10.1007/s10853-018-03265-y.

[19] T. R. Lenka, A. C. Soibam, K. Dey, T. Maung, and F. Lin, "Numerical analysis of high-efficiency lead-free perovskite solar cell with NiO as hole transport material and PCBM as electron transport material," CSI Transactions on ICT, vol. 8, no. 2, pp. 111–116, Jun. 2020, doi: 10.1007/s40012-020-00291-7.

[20] U. Mandadapu, S. V. Vedanayakam, K. Thyagarajan, M. R. Reddy, and B. Jagadeeshbabu, "Design and Simulation of High Efficiency Tin Halide Perovskite Solar Cell," International Journal of Renewable Energy Research, 2017, [Online]. Available: https://api.semanticscholar.org/CorpusID:139662058

[21] A. Sunny, S. Rahman, M. M. Khatun, and S. R. Al Ahmed, "Numerical study of high performance HTL-free CH3NH3SnI3-based perovskite solar cell by SCAPS-1D," AIP Adv, vol. 11, no. 6, Jun. 2021, doi: 10.1063/5.0049646.

[22] K. Afridi, M. Noman, and S. T. Jan, "Evaluating the influence of novel charge transport materials on the photovoltaic properties of MASnI3 solar cells through SCAPS-1D modelling," R Soc Open Sci, vol. 11, no. 1, p. 231202, Jan. 2024, doi: 10.1098/rsos.231202.

[23] K. Deepthi Jayan and V. Sebastian, "Comprehensive device modelling and performance analysis of MASnI3 based perovskite solar cells with diverse ETM, HTM and back metal contacts," Solar Energy, vol. 217, pp. 40–48, Mar. 2021, doi: 10.1016/j.solener.2021.01.058.

[24] S. Imani, S. M. Seyed-Talebi, J. Beheshtian, and E. W. G. Diau, "Simulation and characterization of CH3NH3SnI3-based perovskite solar cells with different Cu-based hole transporting layers," Appl Phys A Mater Sci Process, vol. 129, no. 2, Feb. 2023, doi: 10.1007/s00339-023-06428-0.

[25] R. Jaiswal et al., "Numerical study of eco-friendly Sn-based Perovskite solar cell with 25.48% efficiency using SCAPS-1D," Journal of Materials Science: Materials in Electronics, vol. 34, no. 8, p. 753, 2023, doi: 10.1007/s10854-023-10171-w.

[26] M. K. Hossain et al., "Numerical simulation and optimization of a CsPbI3-based perovskite solar cell to enhance the power conversion efficiency," New Journal of Chemistry, vol. 47, no. 10, pp. 4801–4817, Jan. 2023, doi: 10.1039/d2nj06206b.

[27] A. K. Chaudhary, S. Verma, and R. K. Chauhan, "Thermal and power performance optimization of cost-effective solar cells using eco-friendly perovskite materials," Phys Scr, vol. 99, no. 2, p. 025512, 2024, doi: 10.1088/1402-4896/ad196e.

[28] M. Burgelman, K. Decock, S. Khelifi, and A. Abass, "Advanced electrical simulation of thin film solar cells," Thin Solid Films, vol. 535, pp. 296–301, 2013, doi: https://doi.org/10.1016/j.tsf.2012.10.032.

[29] M. Burgelman, P. Nollet, and S. Degrave, "Modelling polycrystalline semiconductor solar cells," Thin Solid Films, vol. 361–362, pp. 527–532, 2000, doi: https://doi.org/10.1016/S0040-6090(99)00825-1.

[30] M. S. Uddin et al., "Lead-free Ge-based perovskite solar cell incorporating TiO2 and Cu2O charge transport layers harnessing over 25% efficiency," Journal of Optics, 2023, doi: 10.1007/s12596-023-01570-7.

[31] R. A. Jabr, M. Hamad, and Y. M. Mohanna, "Newton-Raphson Solution of Poisson's Equation in a Pn Diode," International Journal of Electrical Engineering & Education, vol. 44, no. 1, pp. 23–33, Jan. 2007, doi: 10.7227/IJEEE.44.1.3.

[32] F. Saeed and H. E. Gelani, "Unravelling the effect of defect density, grain boundary and gradient doping in an efficient lead-free formamidinium perovskite solar cell," Opt Mater (Amst), vol. 124, p. 111952, 2022, doi: https://doi.org/10.1016/j.optmat.2021.111952.

[33] A. Rahmoune and O. Babahani, "Numerical analysis of Al/Gr/ETL/MoS2/Sb2S3/Ni solar cell using non-toxic In2S3/SnS2/ZnSe electron transport layer," Optik (Stuttg), vol. 283, p. 170875, 2023, doi: https://doi.org/10.1016/j.ijleo.2023.170875.

[34] S. Goutham Kumar, E. A. R, T. Surya Teja Reddy, P. C. R, and G. V Honnavar, "Performance Analysis and Optimization of Lead-Free Tin-Based MASnBr3 Tandem Perovskite Solar Cell Section A-Research paper PERFORMANCE ANALYSIS AND OPTIMIZATION OF LEAD-FREE TIN-BASED MASnBr3 TANDEM PEROVSKITE SOLAR CELL," Chem. Bull, vol. 12, no. 6, pp. 2814–2832, 2023, doi: 10.31838/ecb/2023.12.6.251.

[35] Z. Khan, M. Noman, S. Tariq Jan, and A. Daud Khan, "Systematic investigation of the impact of kesterite and zinc based charge transport layers on the device performance and optoelectronic properties of ecofriendly tin (Sn) based perovskite solar cells," Solar Energy, vol. 257, pp. 58–87, 2023, doi: https://doi.org/10.1016/j.solener.2023.04.019.

[36] P. K. Patel, "Device simulation of highly efficient eco-friendly CH3NH3SnI3 perovskite solar cell," Sci Rep, vol. 11, no. 1, p. 3082, 2021, doi: 10.1038/s41598-021-82817-w.

[37] P. Lopez-Varo et al., "Dynamic temperature effects in perovskite solar cells and energy yield," Sustain Energy Fuels, vol. 5, no. 21, pp. 5523–5534, 2021, doi: 10.1039/D1SE01381E.

[38] D. Shi et al., "Low trap-state density and long carrier diffusion in organolead trihalide perovskite single crystals," Science (1979), vol. 347, no. 6221, pp. 519–522, Jan. 2015, doi: 10.1126/science.aaa2725.

[39] G. Hodes and P. V Kamat, "Understanding the Implication of Carrier Diffusion Length in Photovoltaic Cells," J Phys Chem Lett, vol. 6, no. 20, pp. 4090–4092, Oct. 2015, doi: 10.1021/acs.jpclett.5b02052.

[40] J. Burschka et al., "Sequential deposition as a route to high-performance perovskite-sensitized solar cells," Nature, vol. 499, no. 7458, pp. 316–319, 2013, doi: 10.1038/nature12340.

[41] B. Yang et al., "CuSbS2 as a Promising Earth-Abundant Photovoltaic Absorber Material: A Combined Theoretical and Experimental Study," Chemistry of Materials, vol. 26, no. 10, pp. 3135–3143, May 2014, doi: 10.1021/cm500516v.

[42] H. Y. Yang, W. Y. Rho, S. K. Lee, S. H. Kim, and Y. B. Hahn, "TiO 2 nanoparticles/nanotubes for efficient light harvesting in perovskite solar cells," Nanomaterials, vol. 9, no. 3, Mar. 2019, doi: 10.3390/nano9030326.

[43] F. Liu et al., "Numerical simulation: Toward the design of high-efficiency planar perovskite solar cells," Appl Phys Lett, vol. 104, no. 25, p. 253508, Jun. 2014, doi: 10.1063/1.4885367.

[44] A. Bag, R. Radhakrishnan, R. Nekovei, and J. Ramanujam, "Effect of absorber layer, hole transport layer thicknesses, and its doping density on the performance of perovskite solar cells

by device simulation," Solar Energy, vol. 196, pp. 177–182, Aug. 2020, doi: 10.1016/j.solener.2019.12.014.

[45] N. J. Valeti, K. Prakash, and M. K. Singha, "Numerical simulation and optimization of lead free CH3NH3SnI3 perovskite solar cell with CuSbS2 as HTL using SCAPS 1D," Results in Optics, vol. 12, Jul. 2023, doi: 10.1016/j.rio.2023.100440.

[46] M. K. Hossain et al., "Harnessing the potential of CsPbBr3-based perovskite solar cells using efficient charge transport materials and global optimization," RSC Adv, vol. 13, no. 30, pp. 21044–21062, Jul. 2023, doi: 10.1039/d3ra02485g.

[47] Y. Raoui, H. Ez-Zahraouy, N. Tahiri, O. El Bounagui, S. Ahmad, and S. Kazim, "Performance analysis of MAPbI3 based perovskite solar cells employing diverse charge selective contacts: Simulation study," Solar Energy, vol. 193, pp. 948–955, Nov. 2019, doi: 10.1016/j.solener.2019.10.009.

[48] M. Li et al., "CdTe Nanocrystal Hetero-Junction Solar Cells with High Open Circuit Voltage Based on Sb-doped TiO2 Electron Acceptor Materials," Nanomaterials, vol. 7, no. 5, 2017, doi: 10.3390/nano7050101.

[49] O. Ahmad, A. Rashid, M. W. Ahmed, M. F. Nasir, and I. Qasim, "Performance evaluation of Au/p-CdTe/Cs2TiI6/n-TiO2/ITO solar cell using SCAPS-1D," Opt Mater (Amst), vol. 117, p. 111105, 2021, doi: https://doi.org/10.1016/j.optmat.2021.111105.

[50] Y. H. Khattak, F. Baig, S. Ullah, B. Marí, S. Beg, and H. Ullah, "Enhancement of the conversion efficiency of thin film kesterite solar cell," Journal of Renewable and Sustainable Energy, vol. 10, no. 3, p. 033501, May 2018, doi: 10.1063/1.5023478.

[51] T. Leijtens et al., "Carrier trapping and recombination: the role of defect physics in enhancing the open circuit voltage of metal halide perovskite solar cells," Energy Environ. Sci., vol. 9, no. 11, pp. 3472–3481, 2016, doi: 10.1039/C6EE01729K.

[52] I. Alam, R. Mollick, and M. A. Ashraf, "Numerical simulation of Cs2AgBiBr6-based perovskite solar cell with ZnO nanorod and P3HT as the charge transport layers," Physica B Condens Matter, vol. 618, p. 413187, 2021, doi: https://doi.org/10.1016/j.physb.2021.413187.

[53] L. Lin, L. Jiang, P. Li, B. Fan, and Y. Qiu, "A modeled perovskite solar cell structure with a Cu2O hole-transporting layer enabling over 20% efficiency by low-cost low-temperature processing," Journal of Physics and Chemistry of Solids, vol. 124, pp. 205–211, 2019, doi: https://doi.org/10.1016/j.jpcs.2018.09.024.

[54] R. Jeyakumar, A. Bag, R. Nekovei, and R. Radhakrishnan, "Influence of Electron Transport Layer (TiO2) Thickness and Its Doping Density on the Performance of CH3NH3PbI3-Based Planar Perovskite Solar Cells," J Electron Mater, vol. 49, no. 6, pp. 3533–3539, Jun. 2020, doi: 10.1007/s11664-020-08041-w.

[55] U. Mandadapu, S. V. Vedanayakam, and K. Thyagarajan, "Simulation and Analysis of Lead based Perovskite Solar Cell using SCAPS-1D," Indian J Sci Technol, vol. 10, no. 11, pp. 1–8, Mar. 2017, doi: 10.17485/ijst/2017/v10i11/110721.

[56] Y. Cao et al., "Towards high efficiency inverted Sb2Se3 thin film solar cells," Solar Energy Materials and Solar Cells, vol. 200, Sep. 2019, doi: 10.1016/j.solmat.2019.109945.

[57] R. Chakraborty, K. M. Sim, M. Shrivastava, K. V Adarsh, D. S. Chung, and A. Nag, "Colloidal Synthesis, Optical Properties, and Hole Transport Layer Applications of Cu2BaSnS4 (CBTS) Nanocrystals," ACS Appl Energy Mater, vol. 2, no. 5, pp. 3049–3055, May 2019, doi: 10.1021/acsaem.9b00473.

[58] Y. H. Khattak, F. Baig, H. Toura, S. Beg, and B. M. Soucase, "CZTSe Kesterite as an Alternative Hole Transport Layer for MASnI3 Perovskite Solar Cells," J Electron Mater, vol. 48, no. 9, pp. 5723–5733, 2019, doi: 10.1007/s11664-019-07374-5.

[59] U. Mandadapu, "Simulation and Analysis of Lead based Perovskite Solar Cell using SCAPS-1D," Indian J Sci Technol, vol. 10, no. 1, pp. 1–8, Jan. 2017, doi: 10.17485/ijst/2017/v11i10/110721.

[60] F. Anwar, R. Mahbub, S. S. Satter, and S. M. Ullah, "Effect of Different HTM Layers and Electrical Parameters on ZnO Nanorod-Based Lead-Free Perovskite Solar Cell for High-Efficiency Performance," International Journal of Photoenergy, vol. 2017, 2017, doi: 10.1155/2017/9846310.

[61] G. E. Eperon et al., "Inorganic caesium lead iodide perovskite solar cells," J Mater Chem A Mater, vol. 3, no. 39, pp. 19688–19695, 2015, doi: 10.1039/C5TA06398A.

[62] J. Hossain, Md. M. A. Moon, B. K. Mondal, and M. A. Halim, "Design guidelines for a highly efficient high-purity germanium (HPGe)-based double-heterojunction solar cell," Opt Laser Technol, vol. 143, p. 107306, 2021, doi: https://doi.org/10.1016/j.optlastec.2021.107306.

[63] S. R. I. Biplab, Md. H. Ali, Md. M. A. Moon, Md. F. Pervez, Md. F. Rahman, and J. Hossain, "Performance enhancement of CIGS-based solar cells by incorporating an ultrathin BaSi2 BSF layer," J Comput Electron, vol. 19, no. 1, pp. 342–352, 2020, doi: 10.1007/s10825-019-01433-0.

[64] C. M. Proctor and T. Q. Nguyen, "Effect of leakage current and shunt resistance on the light intensity dependence of organic solar cells," Appl Phys Lett, vol. 106, no. 8, Feb. 2015, doi: 10.1063/1.4913589.

[65] P. Roy, N. Kumar Sinha, and A. Khare, "An investigation on the impact of temperature variation over the performance of tin-based perovskite solar cell: A numerical simulation approach," Mater Today Proc, vol. 39, pp. 2022–2026, 2021, doi: https://doi.org/10.1016/j.matpr.2020.09.281.

[66] A. Ashfaq et al., "Comparative performance analysis of Cs2TiX6 (X = Br-, Cl-, I-) lead-free perovskite solar cells incorporating single, double and triple layer halides by SCAPS −1D," Mater Today Commun, vol. 35, p. 106016, 2023, doi: https://doi.org/10.1016/j.mtcomm.2023.106016.

[67] A. Belasri and S. A. Beldjilali, "Springer Proceedings in Energy," 2020. doi: 10.1007/978-981-15-5444-5.

[68] H. B. Michaelson, "The work function of the elements and its periodicity," J Appl Phys, vol. 48, no. 11, pp. 4729–4733, 1977, doi: 10.1063/1.323539.

[69] Y. H. Khattak, F. Baig, H. Toura, S. Beg, and B. M. Soucase, "CZTSe Kesterite as an Alternative Hole Transport Layer for MASnI3 Perovskite Solar Cells," J Electron Mater, vol. 48, no. 9, pp. 5723–5733, 2019, doi: 10.1007/s11664-019-07374-5.

[70] Y. Gan et al., "Numerical Investigation Energy Conversion Performance of Tin-Based Perovskite Solar Cells Using Cell Capacitance Simulator," Energies (Basel), vol. 13, no. 22, 2020, doi: 10.3390/en13225907.

[71] E. Danladi, A. O. Salawu, M. O. Abdulmalik, E. D. Onoja, E. E. Onwoke, and D. S. Adepehin, "Optimization of Absorber and ETM Layer Thickness for Enhanced Tin based Perovskite Solar Cell Performance using SCAPS-1D Software," Physics Access, vol. 02, no. 01, pp. 1–11, 2022, doi: 10.47514/phyaccess.2022.2.1.001.

[72] M. Tripathi, V. Vaibhav Mishra, B. S. Sengar, and A. V Ullas, "Lead-free perovskite solar cell byUsing SCAPS-1D: Design and simulation," Mater Today Proc, vol. 62, pp. 4327–4331, 2022, doi: https://doi.org/10.1016/j.matpr.2022.04.832.

https://doi.org/10.31436/iiumej.v27i1.3970

Evaluating Mechanical and Conductivity of Graphene-Silver Hybrid Inks on Copper Substrate at Elevated Temperature

AHMAD ZAIFAZLIN ZAINORDIN[1,2], NOR AZMMI MASRIPAN[2], MUHD RIDZUAN MANSOR[2], CHONLATEE PHOTONG[3], ALAN WATSON[4], MOHD AZLI SALIM[2*]

[1]*Jabatan Kejuruteraan Mekanikal, Politeknik Sultan Haji Ahmad Shah, 25350 Kuantan, Pahang*
[2]*Fakulti Teknologi dan Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Malaysia*
[3]*Faculty of Engineering, Mahasarakham University, Thailand*
[4]*Power Electronics and Machines Centre, The University of Nottingham, UK Malaysia*

**Corresponding author: azli@utem.edu.my*

(Received: 3 October 2025; Accepted: 7 December 2025; Published online: 12 January 2026)

ABSTRACT: This study investigates the effect of temperature on the damping behaviour, stiffness, and natural frequency of hybrid conductive ink (HCI) printed on a copper (Cu) substrate. The HCI comprises graphene nanoplatelets (GNPs), silver flakes (SF), and silver acetate (SA). The objective is to evaluate the HCI's electrical conductivity and mechanical properties under varying thermal conditions. The HCI paste was formulated with a specified ratio of organic solvents, terpineol, and 1-butanol and cured in an oven at 260°C for 3 hours. The baseline droplet ratio was set at 1:1, and three Cu samples with varied HCl compositions were printed using a 60 µm mesh stencil process. The terpineol concentrations were changed while the 1-butanol droplet remained constant. The samples were then tested for electrical conductivity at room temperature using a two-point probe, in accordance with IEEE Std 118-1978, followed by mechanical behaviour testing using an ASTM E756-05 impact test. Furthermore, the samples were exposed to a range of temperature tests to evaluate mechanical and electrical conductivity under thermal stress. The results showed that the baseline composition exhibited minimal resistance and resistivity across the temperature range, with an average resistance of ≤ 0.2 Ω and a resistivity of ≤ 0.8 Ω · mm, respectively. The baseline composition also exhibited reduced stiffness and damping, with natural frequencies of 10.90, 1.40 kN/m, and 37.51 Hz across the samples. Therefore, the baseline composition exhibits relatively good electrical and mechanical properties for applications in flexible electronics.

ABSTRAK: Kajian ini meneliti kesan suhu terhadap sifat redaman, kekakuan, dan frekuensi semula jadi bagi dakwat konduktif hibrid (HCI) yang dicetak pada substrat kuprum (Cu). HCI ini terdiri daripada graphene platelet nano (GNP), serpihan perak (SF), dan perak asetat (SA). Objektif kajian ini adalah bagi menilai kekonduksian elektrik dan sifat mekanikal HCI di bawah keadaan haba berbeza. Pes HCI dibangunkan dengan nisbah tertentu pelarut organik, terpineol dan 1-butanol, dan dikeringkan dalam ketuhar pada suhu 260ºC selama tiga jam. Nisbah titisan asas ditetapkan pada 1:1, dan tiga sampel Cu dengan komposisi HCI berbeza telah dicetak menggunakan proses stensil jaring 60 µm. Kepekatan terpinol diubah manakala titisan 1-butanol dikekalkan pada kadar yang sama. Sampel-sampel tersebut kemudiannya diuji bagi kekonduksian elektrik pada suhu bilik menggunakan probe Dua-Titik mengikut kod ujian piawaian IEEE Std 118-1978 dan diteruskan dengan ujian sifat mekanikal menggunakan ujian impak ASTM E756-05. Tambahan, sampel telah terdedah kepada pelbagai ujian suhu bagi menilai kekonduksian mekanikal dan elektrik di bawah tekanan haba. Dapatan kajian menunjukkan bahawa komposisi asas menunjukkan rintangan dan kerintangan minimum sepanjang julat suhu, dengan purata rintangan ≤ 0.2 Ω dan kerintangan ≤ 0.8 Ω/mm.

Komposisi asas juga menunjukkan sifat mekanikal dengan kekakuan berkurangan, tingkah laku redaman, dan frekuensi semula jadi masing-masing sebanyak 10.90, 1.40 kN/m, dan 37.51 Hz merentasi sampel. Oleh itu, komposisi asas menunjukkan sifat elektrik dan mekanikal yang agak baik bagi aplikasi elektronik fleksibel.

KEYWORDS: *Hybrid conductive ink, graphene, resistivity, damping, stiffness, frequency*

1. INTRODUCTION

Conductive ink is an advanced material used in various electronic applications, characterized by its ability to conduct electricity while being printable on flexible substrates [1]. It typically consists of conductive particles, such as silver (Ag), copper (Cu), or carbon, and solvents, which enhance electrical conductivity and stability [2]. Its versatility and ease of application make it suitable for the increasing demand for advanced technology in the Internet of Things era [3]. Furthermore, conductive ink is widely used in applications such as electronic components, flexible displays, wearable electronics, and smart packaging [4]. Recent research has demonstrated the potential of HCI, which integrates different conductive materials to achieve higher performance [4]. For example, Ag-Cu hybrid inks have demonstrated improved conductivity and cost-effectiveness compared to pure Ag inks [5]. In addition, the inks have been developed for optimal efficiency in circuit printing, with formulations that facilitate easy batch manufacturing and adaptability with various printing methods.

The impact of temperature on conductive ink is significant, affecting its electrical characteristics and stability. Previous studies demonstrate that different temperatures can affect the resistivity and performance of conductive inks, especially inks formulated with GNP and Ag nanoparticles [6]. Furthermore, the curing technique for GNP/Ag inks at low temperatures enhances conductivity and adhesion, minimising thermal damage to sensitive surfaces [7]. The graphene modified with oleic acid exhibits excellent electrical conductivity, with the optimised ink formulation achieving low resistance at room temperature [8]. Moreover, increased temperatures can lead to thermal degradation of temperature-sensitive substrates, limiting the applicability of conventional metallic inks that require high sintering temperatures, especially for biomaterials [9]. The balance between achieving conductivity while preventing damage is important, as high temperatures may damage the substrate's integrity [10]. This requires innovative ink formulations that allow for lower curing temperatures.

Conductive ink has demonstrated significant advances in conductivity and flexibility; however, there remains a research gap regarding its mechanical properties that influence electrical conductivity, particularly damping behaviour, stiffness, and natural frequency. Numerous studies have concentrated on electrical performance, examining the mechanical behaviour of conductive ink under dynamic conditions [11-14]. Besides, the importance of this gap arises from the fact that electronic devices frequently encounter mechanical stresses, including bending, stretching, and impact, which can influence their electrical conductivity [11]. Conductive inks have exhibited resilience through extensive bending cycles, indicating their suitability for dynamic applications such as human motion detection [15]. Besides, an inconsistency impacts the advancement of flexible electronics, where accurate mechanical properties are required for performance and durability [16].

Excessive vibration can affect the performance, reliability, and lifespan of electronic devices. The effects of vibration arise from various sources, including mechanical motion, environmental conditions, and operational components. The relationship between damping and stiffness significantly influences their natural frequency, as evidenced by various studies [17,18]. The natural frequency is influenced by material properties and structural

configurations, which can be quantified using experimental and numerical methods. Next, mechanical operations, such as torsion and bending, and environmental factors, such as temperature and humidity, can degrade electronic device performance. The researcher investigates the effects of GNP/Ag resistivity under torsional loading. After 5000 cycles, an increase in resistance is observed, attributable to material degradation [19]. Furthermore, cyclic torsion can lead to microcracks and defects in the ink layer, which correlates with increased resistivity [20]. Compared with torsion tests, bending tests show that resistivity increases after repeated motion, indicating degradation of electrical properties [21]. Given the highlighted limitations, it is vital to study the mechanical characteristics of conductive ink, including damping and stiffness, as well as natural frequencies, to optimise the mixture formulation.

Despite the increasing utilisation of GNP and Ag in HCI inks in emerging flexible electronics, existing studies have not sufficiently demonstrated how formulation-dependent thermal-mechanical and electrical properties translate into performance requirements at the device level [22,23]. The current literature primarily focuses on isolated material characterisation, with limited correlation with real operational demands, such as vibration tolerance, stiffness stability, and electrical continuity under dynamic loading. Consequently, a knowledge gap persists in establishing a direct linkage between material microstructure, processing conditions, and their functional implications in flexible electronic applications. This work aims to examine the effect of temperature on the electrical conductivity and mechanical properties, including damping, stiffness, and natural frequency, of HCI composed of GNP, Ag, and SA printed on Cu substrates. The study begins with an assessment of the impact of different terpineol-to-butanol droplet ratios on electrical conductivity at varied temperatures, followed by an investigation of mechanical properties to determine damping behaviour, stiffness, and natural frequency. The test is conducted to establish correlations between electrical and mechanical properties and to determine the optimal mixture composition for improved conductivity and stability under temperature and vibration conditions.

2. MATERIALS AND METHODS

2.1. Raw Material of Hybrid Conductive Ink

The primary conductive materials used in the formulation of HCI are GNP, SF, and SA, all sourced from Sigma-Aldrich. Ethanol, 1-butanol, and terpineol serve as solvents to ensure the ink is printable, adheres well to substrates, and meets performance standards under varying conditions. The HCI powder and paste are prepared following the methods described in previous studies [24]. Next, the material specifications for the HCI development process are stated in Table 1.

Table 1. The raw material in HCI development

Materials	Details
GNP	Average particle size of 25 µm with a molar mass of 12.01 g/mol
SF	Average particle size of ≤10 µm with a molar mass of 107.87 g/mol
SA	99% of purity with a molar mass of 166.91 g/mol
Ethanol	Denatured 99.99% with a molar mass of 46.07 g/mol
1-butanol	99.9% butyl alcohol with a molar mass of 74.12 g/mol
Terpinol	Monoterpene alcohol with a molar mass of 154.25 g/mol

This study used 25 µm particle-size GNP fillers, whose small particle size and powdered morphology facilitate integration into formulations, offering advantages across numerous applications [25]. Similarly, Sigma-Aldrich SF with 99.99% trace-metal purity was utilized to

https://doi.org/10.31436/iiumej.v27i1.3970

improve conductivity when combined with carbon-based materials. The SA serves as a filler material to enhance the conductivity and adhesion of GNP and Ag dispersion through synthesis methods [26,27]. In addition, the HCI formulation contains organic solvents, including ethanol, terpineol, and 1-butanol. Ethanol is used to disperse GNPs and SF in the ink uniformly and to prevent clumping. Terpineol and 1-butanol enhance drying control and adhesion, thereby providing uniform printed ink layers. Their controlled evaporation improves viscosity control, thereby establishing smooth conductive pathways and minimizing defects, thereby improving ink performance.

2.2. Preparation of Hybrid Powder

Figure 1 illustrates a comprehensive methodology for preparing fine HCl powder via a combination of chemicals and processing techniques. The procedure begins with the formulation of a mixture containing 0.005 g of GNP and 5 ml of ethanol, which is sonicated for 10 minutes. The sonication process employs ultrasonic waves to achieve complete dispersion of the GNP in the ethanol solvent, resulting in a uniform suspension and preventing agglomeration [28].

Next, 0.4292 g of SF is added to the mixture, followed by 1 h of sonication to promote uniform dispersion and interfacial interaction between SF and GNP. Subsequently, 0.042 g of SA is added to the mixture, which is sonicated for an additional hour to ensure uniform dispersion of SA. The mixture is subsequently heated to 70 ºC on a hot plate while being stirred at 200 rpm until the ethanol has evaporated. After stirring, the fine HCl powder was transferred to a small container and cured at 250 °C for 1 hour to remove residual ethanol and solidify the powder. After curing for 1 hour, the HCl powder was manually ground in a mortar and pestle until a fine powder formed. The fine HCl powder will be placed in a container for preparing the HCl paste.

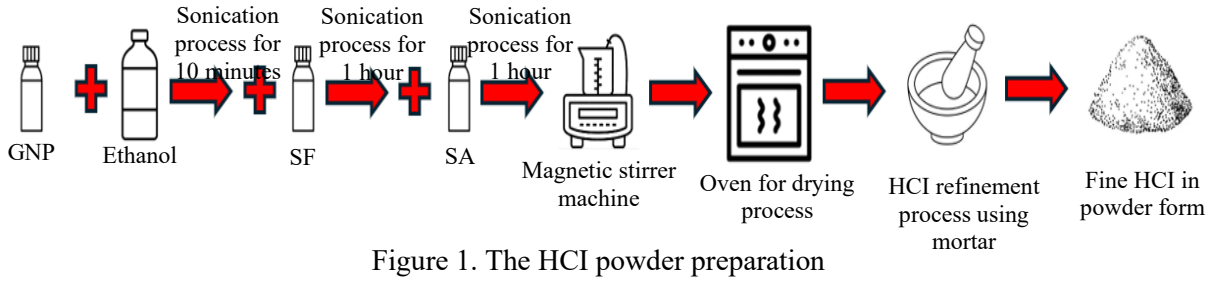


Figure 1. The HCI powder preparation

2.3. Preparation of HCI Paste and Sample

The process starts with the mixture of fine HCl powder and two solvents, terpineol and 1-butanol, as depicted in Figure 2. The mixture is subsequently thoroughly mixed with a Thinky Mixer to ensure uniform dispersion, prevent agglomeration, and eliminate air bubbles. The solvents facilitate proper paste formation, while the Thinky Mixer ensures uniformity and accuracy in substrate printing. Next, a mesh stencil is used to apply the HCI paste to the substrate. A squeegee is used to apply ink to the substrate surfaces, producing five-point ink samples.

Table 2 outlines the formulation for HCI as printed on the substrate surface. The droplet ratios of the three HCI samples differ, with the baseline ratio for terpineol and butanol set at 1:1. In this study, the number of terpineol droplets varies, whereas the quantity of 1-butanol remains constant. Three samples, labeled S1, S2, and S3, each represent a different quantity of

https://doi.org/10.31436/iiumej.v27i1.3970

droplets. Sample S1 is associated with 5T10B, S2 with 10T10B, and S3 with 15T10B as stated in Table 2.

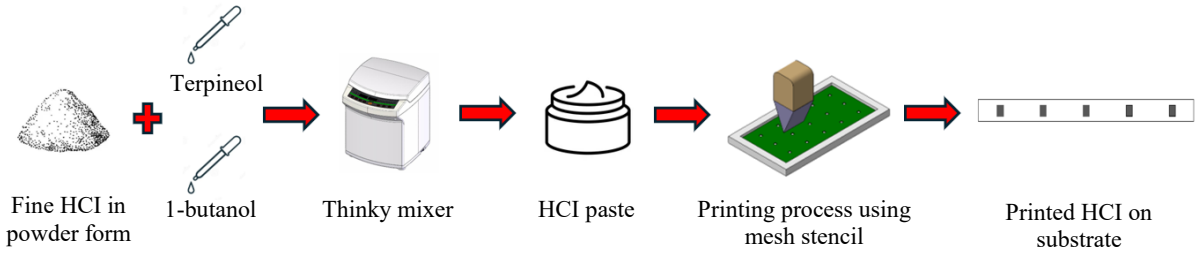


Figure 2. The HCI Paste and Sample Fabrication

Table 2. The HCI Powder and Paste Formulations

Sample	GNP (g)	Ethanol (ml)	SF (g)	SA (g)	Baseline ratio between terpineol and 1-butanol (1 droplet = 0.325 ml)
5T10B	0.005	5	0.4292	0.042	0.5:1
10T10B	0.005	5	0.4292	0.042	1:1
15T10B	0.005	5	0.4292	0.042	1.5:1

The HCI configuration was printed onto a Cu substrate during preparation, prior to electrical and mechanical testing, as illustrated in Figure 3. The Cu substrate measures 120 mm in length and 10 mm in width. A mesh stencil is used as a guide to print ink with a grid dimension of 3.5 mm by 3.5 mm and a thickness of 60 µm ± 2 µm. The printed sample had a 20 mm gap between all of them. A printed section is highlighted with an expanded view, illustrating details of its dimensions and sample placement.

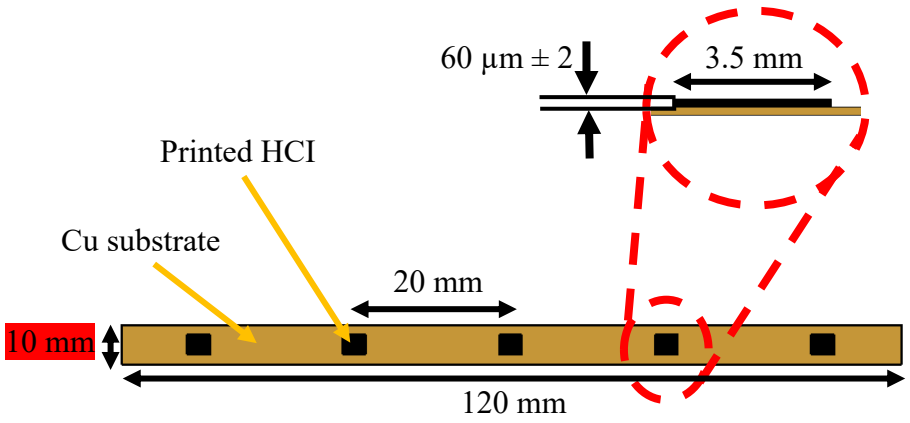


Figure 3. Printed HCI Sample Specification

2.4. Curing Process for HCI Samples

The curing process of printed HCI is depicted in Figure 4, which can influence its electrical conductivity. Three HCI samples with different mixture compositions undergo curing in a UF55 Universal oven chamber at 260 °C for three hours. The oven is utilised owing to its efficient heat distribution and consistency throughout the heating process. Additionally, this curing procedure is designed to ensure adequate adhesion and stabilisation of the printed HCI paste on the substrate.

https://doi.org/10.31436/iiumej.v27i1.3970

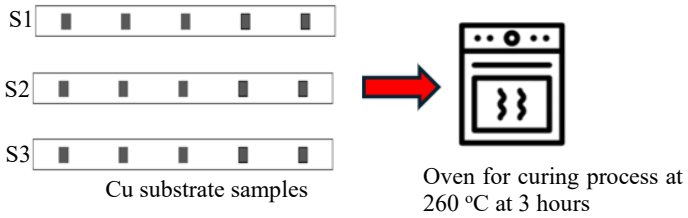


Figure 4. Substrate sample curing process

2.5. The HCI Sample under Various Temperature Effects

After curing, the samples are placed in a heating chamber to measure their electrical conductivity at controlled temperatures, as illustrated in Figure 5. The samples are positioned on the non-conductive glass surface within a heating chamber maintained at approximately 25 ºC. The temperature and humidity sensors are located within the heating chamber to precisely monitor temperature and humidity. Subsequently, a heating tube is positioned at the top of the heating chamber to promote uniform heat distribution, and the temperature control module regulates the temperature. The temperature was set for each test, starting at 30 °C and increasing in 10 °C increments to 60 °C. Subsequently, electrical conductivity measurements were obtained using the Two-Point probe method at each assigned temperature interval, and this procedure was repeated for all samples.

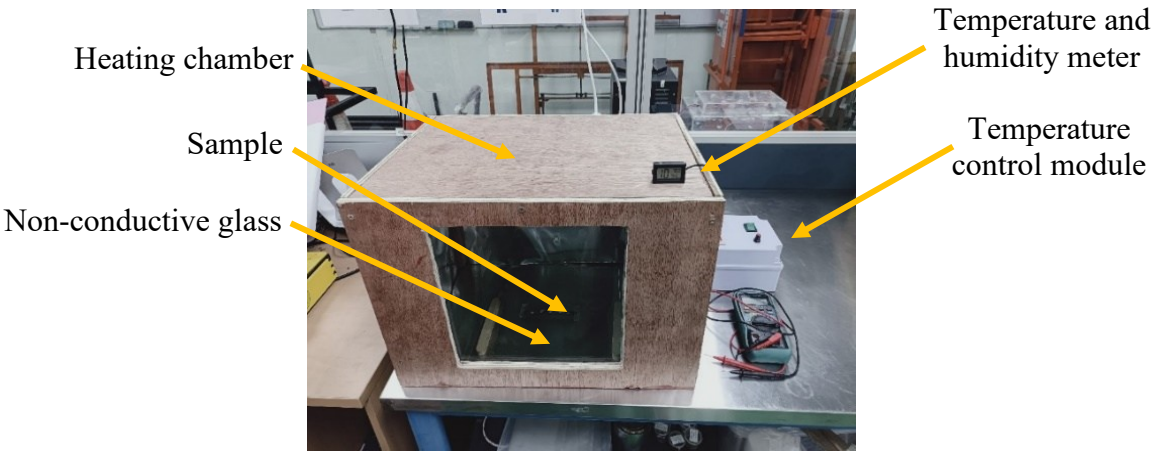


Figure 5. Sample characteristics using a heating chamber with temperature control

2.6 Resistance Measurement using Two-Point Probe

The Two-Point probe method is a simple technique for determining electrical conductivity in terms of resistance, as illustrated in Figure 6. This approach implements the standard test code of IEEE Std 118-1978, using two metallic probes that make direct contact with the printed ink surface to measure resistance.

https://doi.org/10.31436/iiumej.v27i1.3970

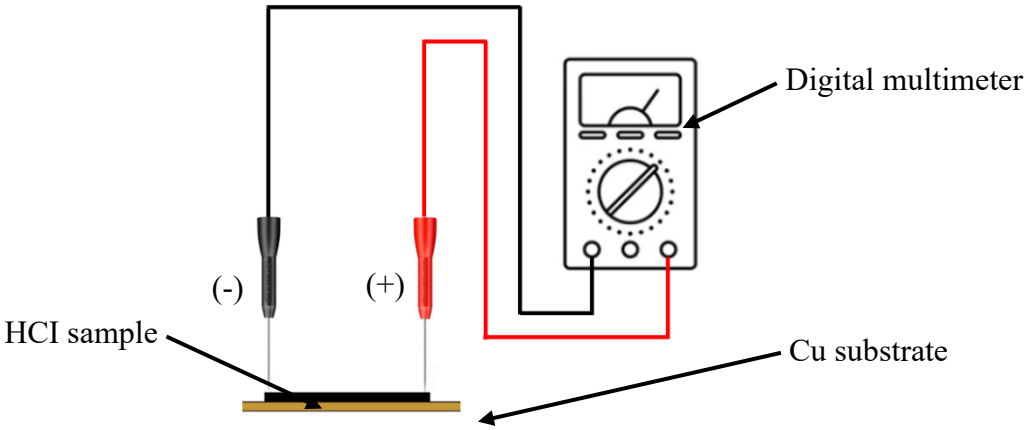


Figure 6. Two-Point probe measurement procedure

Through the resistance data collected, the resistivity was derived through theoretical calculations as outlined below:

$$\rho = \frac{RA}{L} \tag{1}$$

where, R is a resistance measurement, A is the HCI sample area, and L is the length of GNP. In this study, the HCI sample has dimensions of 3.5 mm in length and width. Next, the area (A) is calculated using the equation:

$$A = wL \tag{2}$$

where w is the width and L is the length of the sample. Next, the measurements are conducted at multiple points on the sample to ensure precision in the conductive layer. The probe-specified measurement locations are (1.1, 1.2), (2.1, 2.2), and (3.1, 3.2), as illustrated in Figure 7.

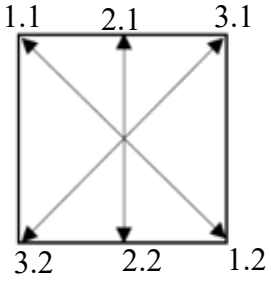


Figure 7. Testing Points for Printed HCI Samples

2.7. Damping, Stiffness, and Natural Frequency Measurements

The impulse hammer test is a fundamental method for exciting mechanisms during mobility measurements. This is a straightforward method to address the challenges of achieving consistent outcomes. The experiment aims to determine the mechanical characteristics of the printed HCI layer on the Cu substrate surfaces bonded to the aluminum beam (Al beam) after thermal exposure and to establish the correlation between temperature and the sample's mechanical properties.

https://doi.org/10.31436/iiumej.v27i1.3970

The experimental procedure begins with sample preparation, in which the Cu substrate is securely attached to the Al beam. The Al beam is fixed at one end, while the other end is allowed to vibrate freely, as seen in Figure 8. An accelerometer is attached to the free end to gather vibration data. At the same time, the impulse hammer is connected to the Data Acquisition System (DAQ) model QUADRO to record the applied force, ensuring synchronization with the accelerometer's measurements. Furthermore, the sample is assessed in accordance with ASTM E756-05 to determine the mechanical properties of HCI, including its damping behaviour, stiffness, and natural frequencies under specified impact loads.

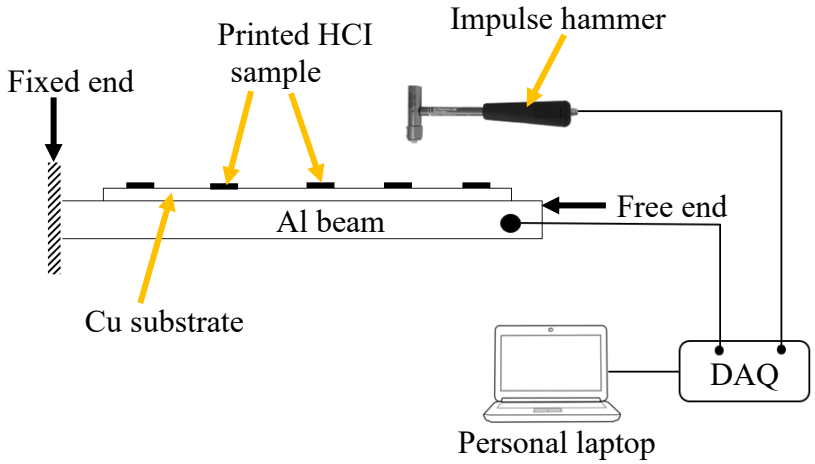


Figure 8. Schematic diagram for vibrational testing

During the experiment, the impulse hammer and accelerometer are calibrated to ensure precise force application and response measurement. Next, the DYTRAN impulse hammer, with a reference sensitivity of 5.21 mV/lb · F, generates an impulse force determined by its mass and velocity. The DYTRAN accelerometer, with a reference sensitivity of 10.12 mV/g, measures the resultant vibrations. Both sensors require calibration before testing to ensure accurate measurements.

The Al beam used in this work has a density of 2.7 g/cm³, a Young's modulus of 69 GPa, and a Poisson's ratio of 0.33, and its mechanical behaviour under axial stress is examined. The material shows a yield strength of 275 MPa and an ultimate tensile strength of 310 MPa, indicating the highest stress it can sustain before failure. Furthermore, the material demonstrates a thermal conductivity of 167 W/m·K, rendering it efficient in heat dissipation. The coefficient of thermal expansion is 23.6×10^{-6}/K, indicating a modest thermal expansion. A Brinell hardness of 95 signifies moderate resistance to indentation and abrasion. The measurements of the Al beam are 240 mm in length, 3 mm in width, and 19 mm in height. The equations for these characteristics, obtained from impact load responses, are shown below.

$$F = kx \tag{3}$$

In this equation, F denotes the impact force (N), k signifies the stiffness (N/m), and x indicates the displacement (m) resulting from the applied force. The stiffness (k) is determined by dividing the applied force by the corresponding displacement, representing the material's resistance to deformation under impact load. The Natural frequency (f_n) can be determined using the equation shown below:

$$f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{4}$$

https://doi.org/10.31436/iiumej.v27i1.3970

where m is the mass of the beam in kg. The damping ratio (ζ) may be computed using the formula shown below:

$$\zeta = c/2f_n \quad (5)$$

3. RESULTS AND DISCUSSION

The experiment was conducted by varying the samples' temperature in a heat chamber, and the data were recorded using the two-point probe method. The averages and error bars for resistance and resistivity of HCI GNP/Ag are shown in Table 3, considering temperature effects and humidity. The sample's mechanical properties are subsequently examined to determine its damping behaviour, stiffness, and natural frequency following temperature variations.

Table 3. Averages and Error Bar for Resistance and Resistivity of the HCI GNP/Ag

Temperature (ºC)/ Humidity (%)	**Average resistance (Ω)**			**Average resistivity (Ω.mm)**			**Error bar**					
							5% of Average Resistance (Ω)			**5% of Average Resistivity (Ω.mm)**		
Mixture	5T 10B	10T1 0B	15T1 0B	5T 10B	10T1 0B	15T1 0B	5T 10B	10T1 0B	15T1 0B	5T 10B	10T1 0B	15T1 0B
25/ (71)	0.55	0.45	0.40	1.91	1.59	1.40	0.03	0.02	0.02	0.10	0.08	0.07
30/ (66)	1.04	0.17	0.12	3.38	0.61	0.42	0.06	0.01	0.01	0.17	0.03	0.02
40/ (64)	0.31	0.17	0.15	1.07	0.61	0.54	0.02	0.01	0.01	0.05	0.03	0.03
50/ (63)	0.19	0.17	0.22	0.68	0.42	0.40	0.01	0.01	0.01	0.03	0.02	0.02
60/ (59)	0.23	0.19	0.56	0.79	0.68	1.96	0.01	0.01	0.03	0.04	0.03	0.10

Figure 9 depicts the average resistance and resistivity of GNP/Ag HCl as functions of temperature for three formulation combinations: 5T10B, 10T10B, and 15T10B. The increased effectiveness of the 10T10B formulation can be attributed to its stability across a range of temperature conditions. The results indicate that the electrical resistance and resistivity of all formulations varied minimally over the temperature range of 30-60 °C, indicating a high degree of thermal stability. Nevertheless, 10T10B consistently exhibited the lowest resistance and resistivity across the samples within this temperature range. This suggests that the material's composition facilitates efficient electron transport, even at elevated temperatures. The carefully calibrated amounts of conductive fillers (GNP and Ag) and the terpineol-to-1-butanol droplet ratio are likely instrumental in establishing a robust conductive network that maintains its integrity under thermal stress. Consequently, 10T10B exhibits not only elevated electrical conductivity but also consistent performance, making it particularly suitable for applications characterized by frequent temperature fluctuations.

Conversely, 5T10B first exhibits an increase in resistance and resistivity at 30 °C, followed by a decrease as the temperature rises to 60 °C, resulting in a resistivity difference of 23%. In HCI, materials containing nanoparticles, such as GNP and Ag, must form continuous networks to enable efficient electron transport [6]. Higher temperatures promote particle dispersion and connectivity, thereby reducing resistance [29]. In addition, resistivity can increase due to the formation of isolated particles caused by insufficient heat.

Compared to the lowest temperature, 15T10B shows a small increase in resistance of around 40% as the temperature rises. This is because substrate temperature affects electrical properties, thereby altering ion distribution and crystal orientation [30]. The temperature of the Cu substrate significantly influences electrical properties; as the deposition temperature

https://doi.org/10.31436/iiumej.v27i1.3970

increases, sheet resistance decreases, thereby enhancing conductivity. The conductivity of films deposited at 200 ºC is low, according to a previous HCI study. However, conductivity improves owing to larger grain sizes at 250 °C and 300 °C, which are higher temperatures [31]. According to this study, the number of droplets can influence particle dispersion. As shown in Figure 9, when the droplet ratio of terpineol is reduced, the mixture of 5T10B exhibits high resistivity at 30 ºC, whereas 15T10B exhibits high resistivity at 60 ºC. This demonstrates the importance of terpineol in enhancing drying control and adhesion, thereby producing smooth, conductive pathways and minimizing defects, thus improving ink performance.

Table 4. Average and Percentage Difference in Conductivity of the HCI GNP/Ag

Temperature (ºC)/ Humidity (%)	**Average conductivity (S/mm)**			**Percentage Difference in Conductivity (%)**		
Mixture	5T 10B	10T 10B	15T 10B	5T 10B	10T 10B	15T 10B
25/ (71)	0.52	0.63	0.71			
30/ (66)	0.30	1.65	2.38	2.38	1.62	2.35
40/ (64)	0.93	1.65	1.86	1.86	1.62	1.62
50/ (63)	1.48	2.38	2.52	2.52	2.78	2.55
60/ (59)	1.26	1.48	0.51	0.51	1.35	-0.28

Table 4 demonstrates that the temperature and humidity influence the electrical conductivity of the HCI GNP/Ag formulations across all sample mixtures. At 25ºC, all samples exhibit moderate conductivity, with 15T10B showing the highest value owing to its greater proportion of conductive fillers, which enhances the formation of interconnected pathways between GNP and Ag particles. As the temperature increases to 30°C and 40°C, the 10T10B and 15T10B formulations exhibit a sharp increase in conductivity. This improvement occurs because heat softens the polymer binder and reduces ink viscosity, enabling better packing, enhanced particle contact, and lower interfacial energy barriers [6]. Humidity within this range may promote slight matrix softening, enabling greater compression without affecting structural integrity. The highest conductivity for all formulations occurs at 50°C, indicating an optimal thermal window in which the microstructure achieves maximum particle alignment and strong percolation networks while maintaining mechanical stability. Beyond this point, performance declines. At 60°C, resistivity increases and conductivity decreases, particularly in 15T10B, indicating structural deterioration due to excessive heat, such as microcracking, thermal expansion mismatch, or over-softening of the binder, which disrupts established conductive pathways.

The characteristics of the solvent mixtures primarily govern differences among the formulations. Terpineol is more effective than 1-butanol because its higher boiling point and slower evaporation rate keep the ink in a wet state for longer, allowing GNP and Ag particles to reorganize into a denser, more continuous conductive network [24]. Its better dispersing ability also helps achieve a more uniform particle distribution during drying. In contrast, 1-butanol evaporates quickly, reducing the time available for particle rearrangement and resulting in a less compact structure. Therefore, formulations with higher terpineol content, particularly 15T10B, exhibit higher conductivity at moderate temperatures and more stable thermally driven network formation.

https://doi.org/10.31436/iiumej.v27i1.3970

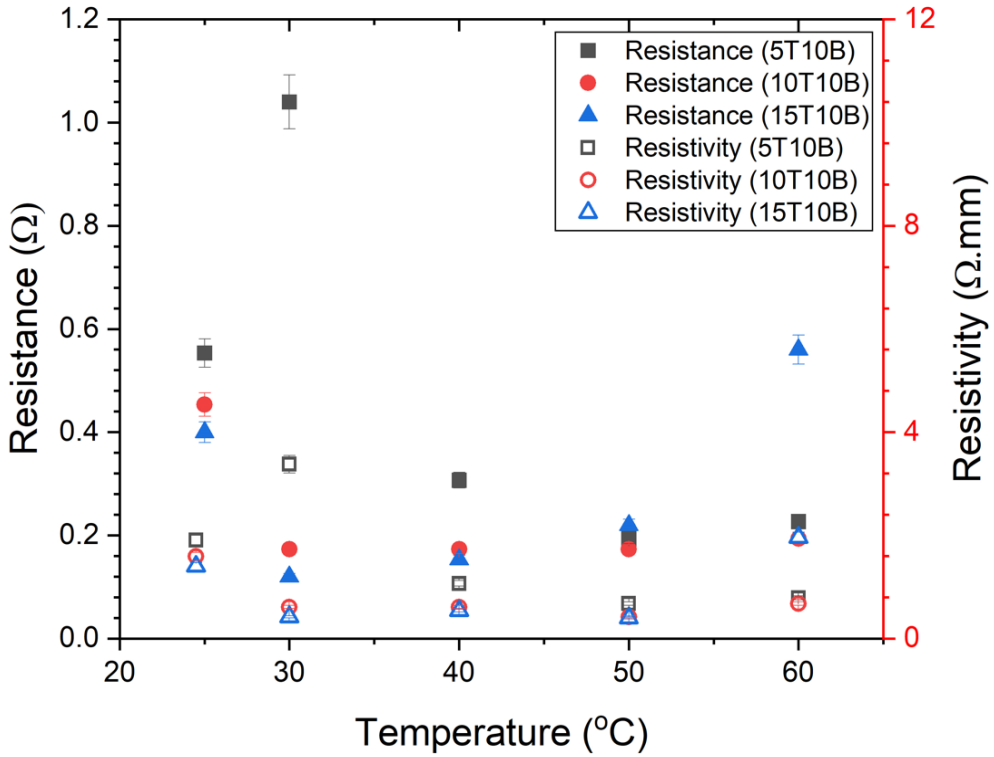


Figure 9. Average Resistance and Resistivity of HCI GNP/Ag

Next, the error bars for the average resistance and resistivity of HCI composed of GNPs and Ag under varying temperature and humidity conditions are presented in Table 3. The results show that measurement error decreases as temperature rises from 25°C to 60°C, indicating improved stability and uniformity of the conductive film at elevated temperatures. Among the tested mixtures, 10T10B consistently exhibits the smallest error bars for both resistance and resistivity, indicating a more stable and homogeneous conductive network than 5T10B and 15T10B. The higher variability observed at lower temperatures, particularly in the 5T10B sample, may be attributed to incomplete curing and weaker interconnections between conductive particles, which lead to fluctuations in electrical pathways [32]. Conversely, increasing temperature enhances the adhesion and interfacial bonding between GNP and Ag flakes, thereby improving electron mobility and reducing measurement uncertainty. The slight decrease in humidity with increasing temperature also contributes to reduced moisture interference, further stabilizing electrical performance [6].

Based on the results, the HCI GNP/Ag combination is a promising formulation because it exhibits excellent electrical conductivity while minimizing substrate heat damage, making it suitable for various applications. Among the studied formulations, 10T10B appears to be the most successful, owing to its well-balanced composition of GNP, SF, SA, and solvents. This composition not only performs well electrically but is also highly stable, retaining consistent conductivity even under varying temperatures. The overall component balance in 10T10B appears to maximise both efficiency and durability, which is essential for utilisation in flexible and wearable electronics.

The sample is then examined for mechanical properties, focusing on damping behaviour, stiffness, and natural frequency, using an impact-testing method. The data obtained during testing includes impulse hammer response, and displacement is derived by integrating the acceleration data. Furthermore, the damping, stiffness, and natural frequency parameters are obtained from theoretical calculations, as specified in Equations (3) to (5). Table 5 presents the

https://doi.org/10.31436/iiumej.v27i1.3970

mechanical parameters of the samples, including an Al beam and HCI-mixture samples 5T10B, 10T10B, and 15T10B, subjected to impact loads. Five impact points have been identified in the sample, which closely resemble those in the printed sample. The data were obtained from the experimental results and processed using MATLAB.

Table 5. Measurement results for damping, stiffness, and natural frequency

Sample	Impact load (N)	Displacement (m)	Stiffness (kN/m)	Natural frequency (Hz)	Velocity (m/s)	Damping ratio (ζ)
Al beam	27.38	0.240	0.05	10.04	0.17	2.84
5T10B	28.25	0.013	2.17	43.13	0.003	37.91
10T10B	24.13	0.017	1.40	37.51	0.012	10.9
15T10B	27.60	0.010	3.45	55.01	0.01	8.92

In flexible and wearable electronic systems, damping, stiffness, and natural frequency are essential mechanical properties that directly influence device reliability and performance. Damping characterizes a material's ability to dissipate mechanical energy, which is crucial for minimizing vibrations and mechanical shocks. High damping capacity helps protect sensitive electronic components from fatigue or failure under repeated mechanical stress, particularly in motion-intensive applications such as healthcare wearables and soft robotics. For stiffness, or the resistance to deformation under applied force, it must be carefully tuned. Excessive stiffness may hinder flexibility, making the device uncomfortable or impractical for wearable use, while insufficient stiffness can compromise structural stability and the protection of embedded circuitry. Optimizing stiffness ensures mechanical durability while maintaining sufficient compliance for stretchability and conformability. The natural frequency is the rate at which a system vibrates in response to an external disturbance. If this frequency aligns with external vibration sources, resonance can occur, drastically increasing stress and potentially leading to structural failure or erratic performance. Hence, controlling the natural frequency through careful design of material geometry, mass, and stiffness is essential to avoid resonance. Together, these mechanical parameters underpin the design of robust, high-performance flexible electronics capable of withstanding real-world mechanical challenges.

Table 6. Measurement results for damping, stiffness, and natural frequency

Sample	Stiffness (kN/m)	Natural frequency (Hz)	Damping ratio (ζ)	Error bar (5%)		
				Stiffness (kN/m)	Natural frequency (Hz)	Damping ratio (ζ)
5T10B	2.17	43.13	37.91	0.11	2.16	0.14
10T10B	1.40	37.51	10.9	0.07	1.88	1.90
15T10B	3.45	55.01	8.92	0.17	2.75	0.55

Table 6 summarises the measured stiffness, natural frequency, and damping ratio for the three HCI formulations, together with their respective ±5% error bars. These parameters are crucial because they reflect the coated Cu substrate's ability to resist vibrational excitation, dissipate mechanical energy, and maintain structural–electrical stability under dynamic operating conditions. The 15T10B formulation recorded the highest stiffness value of 3.45 kN/m and the highest natural frequency of 55.01 Hz, indicating that the higher GNP and Ag filler loading produced a mechanically stiffer coating and consequently a higher resonant response; however, this sample also showed the lowest damping ratio of 8.92, which reflects a reduced capability to dissipate vibrational energy. In contrast, the 5T10B sample showed a moderate stiffness value of 2.17 kN/m and a natural frequency of 43.13 Hz, yet achieved the

https://doi.org/10.31436/iiumej.v27i1.3970

highest damping ratio of 37.91, suggesting that lower filler concentration permitted greater viscoelastic mobility and improved energy absorption. The 10T10B formulation recorded the lowest stiffness (1.40 kN/m) and a natural frequency of 37.51 Hz, while exhibiting a moderately low damping ratio of 10.9. Overall, the trend indicates that each formulation produces a different balance between structural rigidity and energy dissipation, confirming that formulation composition strongly governs the thermal-mechanical performance of the HCI.

Figure 10 illustrates the effects of stiffness and average resistivity across various mixture samples, emphasising the disparities in mechanical and electrical properties. The 5T10B formulation demonstrates the highest stiffness, surpassing 3.45 kN/m, while its resistivity remains moderate. By contrast, the 10T10B sample exhibits the lowest stiffness (approximately 1.40 kN/m) and resistivity, indicating excellent electrical conductivity. The 15T10B mixture exhibits a stiffness of roughly 2.17 kN/m, with a resistivity that exceeds that of 10T10B but is lower than that of 5T10B. The results indicate that the formulation's composition significantly influences both mechanical and electrical performance. Similarly, as reported in previous studies, the incorporation of GNP into epoxy/glass fiber composites enhances the overall stiffness, resulting in an increase in natural frequencies by up to 8.05%, while the damping ratio decreases by more than 35% [33]. Although increased stiffness, as demonstrated by 5T10B, provides improved resistance to deformation, it may restrict flexibility and key characteristics of wearable and flexible electronics.

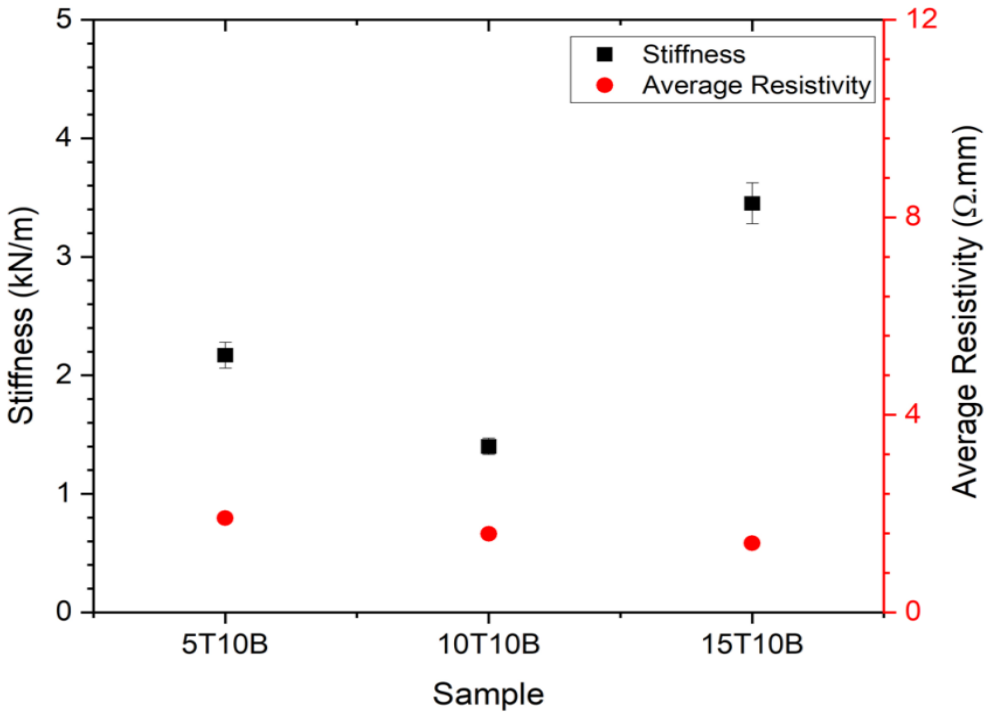


Figure 10. Comparison between stiffness and average resistivity at various mixture samples

Moreover, excessive rigidity can adversely impact damping performance, diminishing materials' ability to absorb mechanical shocks or vibrations [34]. The reduced rigidity of 10T10B facilitates enhanced damping, making it more appropriate for dynamic applications characterised by movement and vibration. Furthermore, natural frequency is intrinsically linked to stiffness; increased stiffness elevates the natural frequency, which may result in resonance under specific conditions. The 10T10B formulation, characterised by reduced stiffness and possibly diminished natural frequency, may mitigate such problems, hence improving device dependability. Moreover, the 10T10B formulation provides the optimal

https://doi.org/10.31436/iiumej.v27i1.3970

combination of mechanical compliance, damping capacity, and electrical conductivity among the evaluated samples. This is because the 10T10B formulation has a balanced filler concentration and film thickness, which enables the formation of an optimal percolation network for stable and efficient charge transport [24,35]. It provides an ideal combination of mechanical flexibility and electrical conductivity, ensuring consistent electrical performance and enhanced damping capacity [36].

Figure 11 depicts the correlation between natural frequency and average resistivity across various mixture samples. The results indicate significant differences in both characteristics across the multiple formulations. The 5T10B sample exhibits the highest average resistivity, exceeding 3.38 Ω·mm, despite a relatively low natural frequency of approximately 43 Hz. The 10T10B sample has the lowest resistivity, approximately 0.42 Ω·mm, and the lowest natural frequency, approximately 40 Hz. Meanwhile, the 15T10B sample exhibits an increase in natural frequency to approximately 55 Hz, along with an average resistivity. Based on the results, it is essential to recognise that the natural frequency depends on both the material's stiffness and mass. An elevated natural frequency typically indicates a more rigid structure, which may enhance resistance to deformation but could also reduce the flexibility and damping required for specific electronic applications. In contrast, reduced natural frequencies, as shown in 10T10B, may arise from elastic structures that are better suited to dynamic and flexible applications.

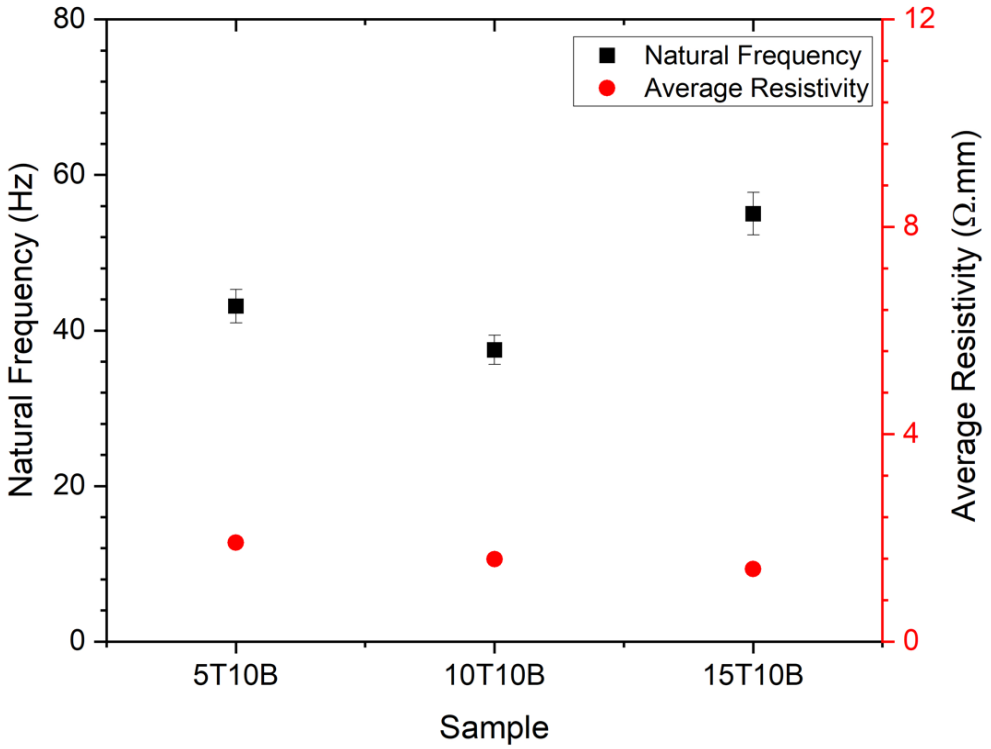


Figure 11. Comparison between natural frequency and average resistivity at various mixture samples

The conductive network inside the formulation determines the impact on resistivity. A lower natural frequency may indicate a more flexible, continuous conductive path that enhances electron transport, thereby reducing resistance. Consequently, the 10T10B sample's combination of low natural frequency and low resistivity indicates ideal structural and electrical stability. The findings suggest that modifying mixture composition can effectively optimise both mechanical and electrical properties, essential for developing high-performance HCI formulations for flexible and wearable electronics.

https://doi.org/10.31436/iiumej.v27i1.3970

Through applying an impact force of 27.38 N, the Al beam demonstrates a maximum displacement of 0.240 m, a minimum stiffness of 0.0452 kN/m, and a minimum natural frequency of 10.04 Hz, as illustrated in Figure 12. Conversely, the HCI samples exhibit significantly higher stiffness and natural frequency, and lower displacement under comparable impact loads. When applying an impact load of 28.25 N, the 5T10B sample exhibits a minimum displacement of 0.013 m, yielding a stiffness of 2.17 kN/m and a natural frequency of 43.13 Hz. The 10T10B sample exhibits a marginally reduced impact load of 24.13 N and a displacement of 0.0172 m, yielding a stiffness of 1.40 kN/m and a natural frequency of 37.51 Hz. At the 27.6 N impact load, the 15T10B sample exhibits the minimal displacement of 0.008 m, the maximum stiffness of 3.45 kN/m, and the highest natural frequency of approximately 55.01 Hz relative to all other samples.

The results reveal that the HCI samples outperform the Al beam in stiffness and natural frequency, primarily due to variations in material composition and thermal treatment. These mechanical enhancements are attributed to the improved structural framework and interfacial bonding within the hybrid composite. Notably, an inverse relationship between stiffness and displacement is observed: higher stiffness results in reduced deformation under identical stress conditions. This aligns with fundamental mechanical principles: a stiffer material resists shape changes more effectively, enhancing structural stability but often at the cost of reduced damping capacity. The data further indicate that as stiffness increases, the natural frequency also rises. This correlation suggests that stiffer materials are more resistant to oscillatory disturbances, thereby offering greater mechanical stability. However, this comes with a trade-off, where increased stiffness generally corresponds to a lower damping ratio, meaning the material is less effective at dissipating energy under dynamic loads. These characteristics have critical implications for applications involving vibration or cyclic mechanical stress. Figure 12 illustrates this dynamic by showing the inverse correlation between stiffness and both displacement and the damping ratio, demonstrating that deformation decreases as stiffness increases.

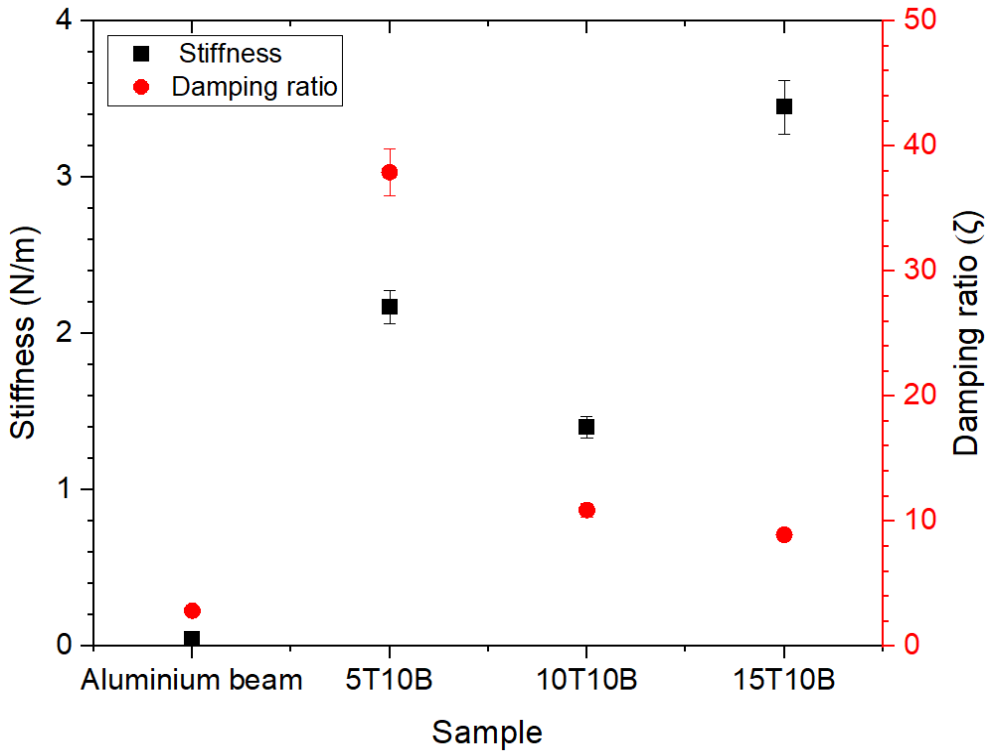


Figure 12. Comparison between stiffness and damping ratio at various mixture samples

https://doi.org/10.31436/iiumej.v27i1.3970

Figure 13 presents SEM micrographs of the HCI samples after vibration exposure. Before morphological examination, the HCI specimens were thermally cured at 260 ºC for 3 hours and subsequently subjected to impact testing using an Al beam. The microstructure was observed at ×2000 magnification. During testing, the vibration impact acted five times simultaneously on the printed film, which is known to induce particle rearrangement and affect interfacial stability along the conductive pathways. Figure 13(a) depicts the 5T10B formulation, exhibiting noticeable pore formation. This phenomenon is attributed to the solvent ratio, which reduces packing density and promotes void formation, thereby weakening the conductive network. In contrast, Figure 13(b) depicts the 10T10B formulation, displaying well-coated metallic particles with strong interparticle bonding and the absence of discernible particle boundaries, attributes that indicate an efficient and flexible conductive network. Figure 13(c) shows the 15T10B sample, in which the SEM image reveals a well-dispersed HCI structure that forms a continuous and stable conductive network even after combined impact–vibration exposure. Despite the presence of small voids, these discontinuities may disrupt some conductive pathways, explaining the relatively higher susceptibility to conductivity degradation under vibration-assisted mechanical deformation. The microstructural stability under combined impact-vibration conditions indicates that the 10T10B composition maintains an ideal balance between mechanical integrity and electrical conductivity, making it more adaptable than the other formulations.

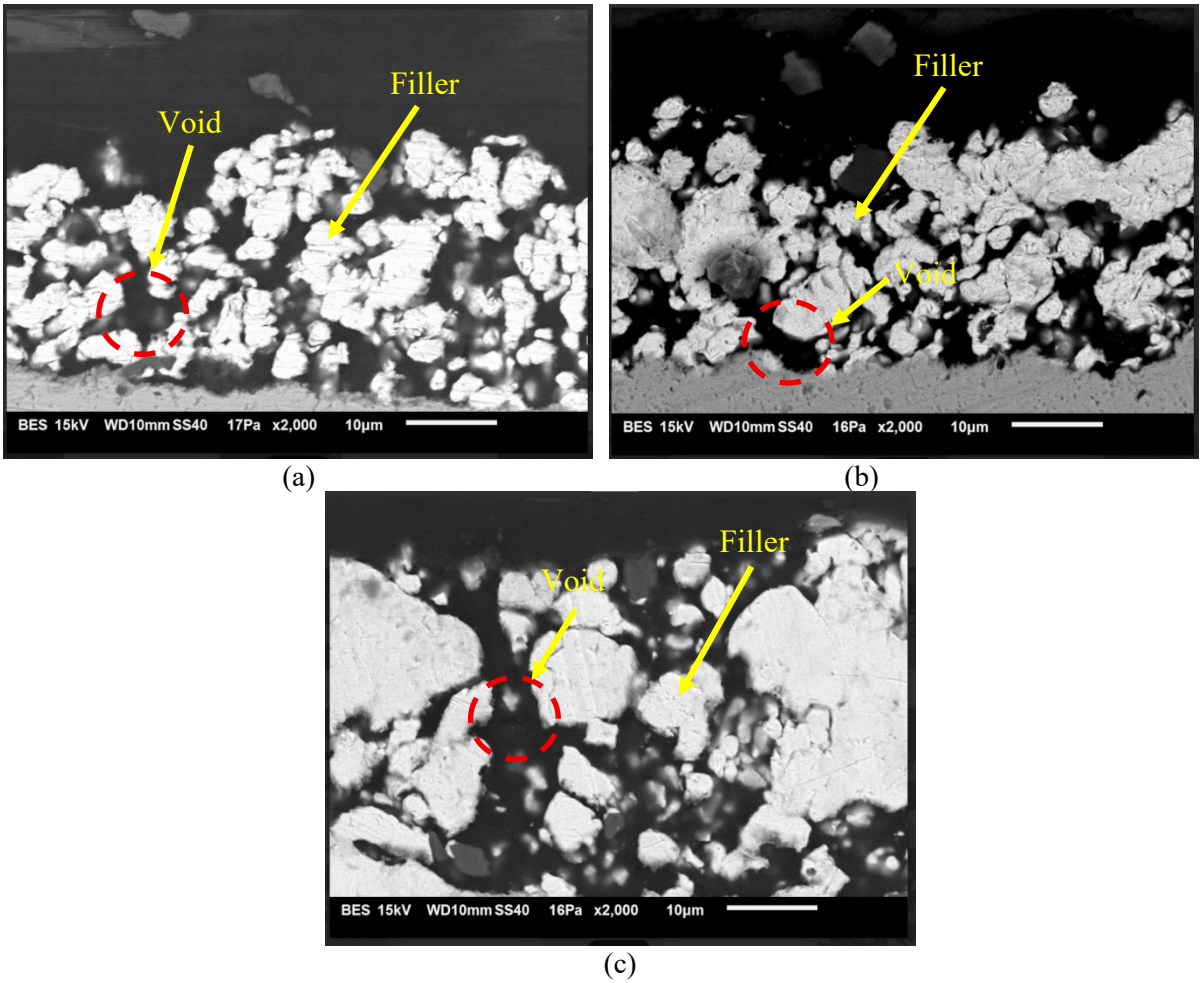


Figure 13. SEM images of each sample after impact testing.

Moreover, the study highlights how the mechanical and electrical properties of GNP/Ag HCI formulations are strongly influenced by material composition and thermal behavior. As temperature increases, the polymer binder softens, allowing GNP and Ag fillers to come into closer contact, thereby improving particle dispersion and strengthening both conductive networks. This effect is most evident in the 10T10B formulation, which achieves an effective balance of stiffness, damping, and conductivity. Its moderate stiffness and lower natural frequency enhance damping capacity and flexibility, making it suitable for wearable or flexible electronic applications. The 15T10B formulation exhibits more stiffness and a higher natural frequency, resulting in less damping and reduced mechanical flexibility. The improved stiffness observed at moderate temperatures arises from thermally induced matrix reduction, which enables more uniform stress redistribution and better particle interlocking. The solvent effects further reinforce these outcomes: formulations with higher terpineol content, such as 15T10B, develop denser and more uniform particle networks during curing owing to terpineol's slower evaporation and better dispersive capability [37]. These consolidated microstructures provide greater thermal stability but also increase rigidity. Conversely, the rapid evaporation of 1-butanol yields less compact networks, making mechanical properties more sensitive to thermal softening [38]. The results show that optimal mechanical performance is achieved when thermal activation enhances network formation without inducing excessive stiffness, a balance most effectively demonstrated by the 10T10B formulation.

4. CONCLUSIONS

This study demonstrates that the initial composition of HCI printed on the Cu substrate displays excellent electrical conductivity with low resistance and resistivity across the investigated temperature range. Furthermore, the mechanical properties, such as stiffness, natural frequency, and damping behaviour, are comparatively lower, suggesting an agreement between electrical and mechanical performance. The results indicate that preserving the baseline composition enhances electrical conductivity while providing acceptable mechanical properties. Nevertheless, modifications to the HCI composition, particularly changes in terpineol content, adversely affect both electrical and mechanical stability. This highlights the importance of precise formulation modifications to maintain conductivity and structural integrity for reliable applications. Overall, the findings indicate that HCI on Cu substrates can achieve a good balance of damping performance, structural stiffness, and electrical conductivity. These characteristics demonstrate their suitability for use in flexible electronic interfaces that require suppression of mechanical vibration and reliable signal transmission. In addition, the property stability demonstrated in this study indicates strong potential for scalable formulation refinement and integration into industrial-scale printed electronic manufacturing processes, particularly for lightweight sensor assemblies, smart structural components, and next-generation vibration-responsive circuitry. Thus, the baseline composition exhibits relatively good electrical and mechanical properties for applications in flexible electronics.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Advanced Academia Industry Collaboration Laboratory (AiCL), Fakulti Teknologi dan Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, and the Department of Mechanical Engineering, Politeknik Sultan Haji Ahmad Shah, Kuantan, Pahang.

REFERENCES

[1] Yanan W (2019). Flexible substrate, manufacturing method of flexible substrate and flexible display device, (2019).

[2] Haixin C, Ganghui L, Hui G, (2020). Graphene composite conductive ink and preparation method thereof.

[3] Romero EC, Sanchez Hernandez S, Acosta AM, (2023). Development of Carbon-Based Conductive Inks with Resistant Response for Potential Use in Flexible Sensors with IoT Technology. 2023 IEEE Colombian Caribbean Conference (C3), IEEE p. 1–5.

[4] Osváth Z, Pálinkás A, Piszter G, Molnár G, (2020). Synthesis and Characterization of Graphene–Silver Nanoparticle Hybrid Materials. Materials, 13(20):4660. Doi: 10.3390/MA13204660.

[5] Bobsin A, Rodrigues TC, Fernandes IJ, Ferreira SB, Peter CR, Hasenkamp W, et al., (2024). Copper and silver microparticles for high-performance conductive inks in electronic chip shielding. Materials Chemistry and Physics, 315:129007. Doi: 10.1016/j.matchemphys.2024.129007.

[6] Noor NM, Salim MA, Masipan NA, Md. Saad A, Photong C, Akop MZ, (2024). Correlation Effect on Different Temperature-Humidity Range of Highly Thermal GNP/AG/ SA Conductive Ink. International Journal of Integrated Engineering, 16(6).

[7] Muhammad Hisyamuddin Hussin, Nor Azmmi Masripan, Mohd Azli Salim, Adzni Md. Saad, Chonlatee Photong, (2024). Electric Symphony: Unveiling the Potential of Reduced-Temperature Cured Conductive Ink. Journal of Advanced Research in Applied Mechanics, 120(1):72–84. Doi: 10.37934/aram.120.1.7284.

[8] Le TT, Bui HHT, Dinh AKP, Van DV, Ho QD, Thi HAN, et al., (2022). Room Temperature-Sintering Conductive Ink Fabricated from Oleic-Modified Graphene for the Flexible Electronic Devices. ChemistrySelect, 7(4). Doi: 10.1002/slct.202104249.

[9] Zhou Y, Xu Z, Bai H, Knapp CE, (2023). Room Temperature Electronic Functionalization of Thermally Sensitive Substrates by Inkjet Printing of a Reactive Silver-Based MOD Ink. Advanced Materials Technologies, 8(8). Doi: 10.1002/admt.202201557.

[10] Liu T, Guo R, Fu Y, Zhao J, Ning H, Fang Z, et al., (2022). Morphological Regulation of Printed Low-Temperature Conductive Ink. Langmuir, 38(32):9955–66. Doi: 10.1021/acs.langmuir.2c01249.

[11] Zulfiqar S, Saad AA, Ahmad Z, Yusof F, Bachok Z, (2023). Structural Analysis of Silver-Based Conductive Ink Under Cyclic Loading. In: Mohd Salleh, M.A.A., Che Halin, D.S., Abdul Razak, K., Ramli, M.I.I., editors. Proceedings of the Green Materials and Electronic Packaging Interconnect Technology Symposium, Singapore: Springer Nature Singapore p. 211–9.

[12] Walsh P, (2023). Structural Analysis of Silver-Based Conductive Ink Under Cyclic Loading:211–9. Doi: 10.1007/978-981-19-9267-4_25.

[13] Koshi T, Nomura K, Yoshida M, (2019). Requirements for Durability Improvement of Conductive Patterns Permeated in Textiles under Cyclic Tensile Deformation. Micromachines, 10(11):721. Doi: 10.3390/MI10110721.

[14] Du H, Xue T, Xu C, Kang Y, Dou W, (2018). Improvement of mechanical properties of graphene/substrate interface via regulation of initial strain through cyclic loading. Optics and Lasers in Engineering, 110:356–63. Doi: https://doi.org/10.1016/j.optlaseng.2018.04.026.

[15] Wu P, He H, (2024). Integrating High-Performance Flexible Wires with Strain Sensors for Wearable Human Motion Detection. Sensors, 24(15):4795. Doi: 10.3390/s24154795.

[16] Cordill MJ, Glushko O, Putz B, (2016). Electro-Mechanical Testing of Conductive Materials Used in Flexible Electronics. Frontiers in Materials, 3. Doi: 10.3389/fmats.2016.00011.

[17] Azam SA, Fragoso A, (2020). Experimental and Numerical Simulation Study of the Vibration Properties of Thin Copper Films Bonded to FR4 Composite. Applied Sciences, 10(15):5197. Doi: 10.3390/app10155197.

[18] OLINGHERU R-G, OLARIU A-B, PRAISACH Z-I, (2024). The influence of stiffness on the dynamic behavior of a PCB enclosure. Studia Universitatis Babeș-Bolyai Engineering,:130–6. Doi: 10.24193/subbeng.2024.1.13.

[19] Ismaniza Ismail, Mohd Azli Salim, Nor Azmmi Masripan, Adzni Md. Saad, Mohd Zaid Akop, Chew Kit Wayne, et al., (2024). Effect of Torsional Load on Electrical and Mechanical Properties/Behaviour of Stretchable GNP-Ag Conductive Ink. Journal of Advanced Research in Applied Mechanics, 117(1):22–34. Doi: 10.37934/aram.117.1.2234.

[20] Hussain MA, Salim MA, Masripan NA, Md. Saad A, Akop MZ, Photong C, (2024). Investigation of Functional Ageing of GNP/Ag Conductive Ink under Torsional Conditions. International Journal of Integrated Engineering, 16(2). Doi: 10.30880/ijie.2024.16.02.034.

[21] Ismaniza Ismail, Mohd Azli Salim, Nor Azmmi Masripan, Adzni Md. Saad, Mohd Zaid Akop, Chew Kit Wayne, et al., (2024). Resistivity of Graphene/Silver Hybridization Conductive Ink on Bending Test. Journal of Advanced Research in Applied Mechanics, 120(1):27–39. Doi: 10.37934/aram.120.1.2739.

[22] Zhang Y, Wu H, Liu L, Yang Y, Zhang C, Duan J, (2024). Crack-Based Composite Flexible Sensor with Superhydrophobicity to Detect Strain and Vibration. Polymers, 16(17):2535. Doi: 10.3390/polym16172535.

[23] Lu A, Zhao L, Liu Y, Li Z, Xiong D-B, Zou J, et al., (2020). Enhanced Damping Capacity in Graphene-Al Nanolaminated Composite Pillars Under Compression Cyclic Loading. Metallurgical and Materials Transactions A, 51(4):1463–8. Doi: 10.1007/s11661-020-05632-4.

[24] Jori SH, Salim MA, Masripan NA, Md. Saad A, Chonlatee P, (2024). Effect of Butanol and Terpinol as a Solvents in GNP/SA Conductive Ink. Journal of Advanced Research in Applied Mechanics, 120(1):62–71.

[25] Li W, Yan J, Wang C, Zhang N, Choy TH, Liu S, et al., (2022). Molecule bridged graphene/Ag for highly conductive ink. Science China. Materials, 65(10):2771–8. Doi: 10.1007/s40843-022-2064-8.

[26] Dubey RK, Shukla S, Hussain Z, (2023). Green Synthesis of Silver Nanoparticles; A Sustainable Approach with Diverse Applications. Chinese Journal of Applied Physiology,:e20230007. Doi: 10.62958/j.cjap.2023.007.

[27] Rafique M, Sadaf I, Rafique MS, Tahir MB, (2017). A review on green synthesis of silver nanoparticles and their applications. Artificial Cells, Nanomedicine, and Biotechnology, 45(7):1272–91. Doi: 10.1080/21691401.2016.1241792.

[28] Kasim NFA, Idris WFW, Abdullah AH, Yusoh K, Ismail Z, (2019). The preparation of graphene ink from the exfoliation of graphite in pullulan, chitosan and alginate for strain-sensitive paper. International Journal of Biological Macromolecules,. Doi: 10.1016/j.ijbiomac.2019.10.251.

[29] Rezaga BFY, Balela MDL, (2024). Conductive Inks with Chemically Sintered Silver Nanoparticles at Room Temperature for Printable, Flexible Electronic Applications. Key Engineering Materials, 983:9–16. Doi: 10.4028/p-DAaz5z.

[30] Zhang K, Kong W, Wang Q, Xuan T, Cheng F, Chang A, (2019). Effect of substrate temperature on structure, cationic distribution and electrical properties of MnCo0.2Ni0.1Mg0.6Al1.1O4 thin films. Journal of Materials Science: Materials in Electronics, 30(15):14200–6. Doi: 10.1007/s10854-019-01787-y.

[31] Koshy AM, Sudha A, Yadav SK, Swaminathan P, (2023). Effect of substrate temperature on the optical properties of DC magnetron sputtered copper oxide thin films. Physica B: Condensed Matter, 650:414452. Doi: 10.1016/j.physb.2022.414452.

[32] Inoue M, Suganuma K, (2005). Effect of curing conditions on the interconnect properties of isotropic conductive adhesives composed of an epoxy-based binder. 2005 Conference on High Density Microsystem Design and Packaging and Component Failure Analysis, IEEE p. 1–6.

[33] Veenapani R, Ananda Kumar KV, Pavana B S, (2024). Impact of Graphene Reinforcement Upon Damping Properties of Epoxy/Glass Hybrid Composites. Journal of Scientific Research and Technology,:12–9. Doi: 10.61808/jsrt136.

[34] Geethamma VG, Asaletha R, Kalarikkal N, Thomas S, (2014). Vibration and sound damping in polymers. Resonance, 19(9):821–33. Doi: 10.1007/s12045-014-0091-1.

[35] Mohd Azli Salim, Norida Mohammad Noor, Nor Azmmi Masripan, Adzni Md. Saad, Mohd Zaid Akop, Chew Kit Wayne, et al., (2024). Effect of GNP/Ag Stretchable Conductive Ink on Electrical Conductivity. Journal of Advanced Research in Applied Mechanics, 119(1):1–12. Doi: 10.37934/aram.119.1.112.

[36] Sato R, Kawakita J, Chikyow T, Sakamoto Y, (2017). Mechanical-Electrical Behaviour of Conductive Polymer / Metal Composite for Flexible Interconnect. ECS Meeting Abstracts, MA2017-02(52):2205–2205. Doi: 10.1149/MA2017-02/52/2205.

[37] Ahmed MR, Mirihanage W, Potluri P, Fernando A, (2023). Highly Stable Graphene Inks Based on Organic Binary Solvents. Particle & Particle Systems Characterization, 40(2). Doi: 10.1002/ppsc.202200153.

[38] Kosasih EA, Hakim AAR, Hussein M, (2018). Analysis of butanol droplet evaporation using modified stagnant film model. p. 020009.

https://doi.org/10.31436/iiumej.v27i1.3980

Surrogate Model-based Optimization of a Biplane Unmanned Aerial Vehicle Wing Using Genetic Algorithm and Response Surface Methodology

YUKSEL ERASLAN*

Aerospace Engineering Department, Tarsus University, Mersin, Türkiye

**Corresponding author: yukseleraslan@tarsus.edu.tr*

(Received: 10 October 2025; Accepted: 20 December 2025; Published online: 12 January 2026)

ABSTRACT: Multi-wing aerial vehicle designs have existed since the early days of aviation; however, they are not commonly employed in modern designs, as advances in materials technology have enabled the structurally feasible high-aspect-ratio monoplane configurations. Nevertheless, the application has the potential to serve as a solution in the field of mini unmanned aerial vehicle (UAV) design, offering the advantage of a smaller wing span. The most popular multi-wing application, the biplane, features two wings mounted in parallel and is typically characterized by the geometric terms decalage, stagger, and gap, which should be optimized during the design process. This study focuses on aerodynamic optimization of a biplane wing to maximize the lift-to-drag ratio (C_L/C_D) and the $C_L^{3/2}/C_D$ by optimizing the aforementioned geometric parameters. To this end, the aerodynamic analysis method was initially validated by comparing it with wind-tunnel data for a monoplane wing reported in the literature. The Response Surface Method (RSM) was used to assess the effects of the design parameters and to generate surrogate models for C_L/C_D and $C_L^{3/2}/C_D$. The optimal design was determined using the genetic algorithm. The results indicated that decalage and gap distance primarily determine range improvement, whereas gap and stagger are the primary parameters for improving endurance. The applied optimization approach resulted in improvements of 10.93% in range and 46.63% in endurance relative to the base biplane configuration.

ABSTRAK: Reka bentuk kenderaan udara berbilang sayap telah wujud sejak awal perkembangan penerbangan; namun, ia jarang digunakan dalam reka bentuk moden berikutan kemajuan teknologi bahan yang membolehkan nisbah-aspek-tinggi konfigurasi monosayap lebih berdaya struktur. Namun, aplikasi berbilang sayap berpotensi menjadi penyelesaian reka bentuk kenderaan udara tanpa pemandu bersaiz kecil (mini UAV), khususnya dengan kelebihan bentangan sayap lebih kecil. Konfigurasi berbilang sayap iaitu paling lazim, sayap berkembar (biplane), merujuk kepada dua sayap yang disusun selari dan dicirikan oleh parameter geometri utama seperti sela, aturan, dan jurang, perlu ditentukan secara optimum dalam proses reka bentuk. Kajian ini memfokuskan kepada pengoptimuman aerodinamik sayap berkembar bagi memaksimumkan nisbah angkat kepada seretan (C_L/C_D) dan $C_L^{3/2}/C_D$ melalui pemilihan optimum parameter geometri tersebut. Kaedah analisis aerodinamik terlebih dahulu disahkan melalui perbandingan data terowong angin bagi monosayap yang dilaporkan dalam literatur. Kaedah Permukaan Tindak Balas (Response Surface Method, RSM) digunakan bagi menilai kesan parameter reka bentuk dan model pengganti bagi C_L/C_D dan $C_L^{3/2}/C_D$, manakala reka bentuk optimum ditentukan menggunakan algoritma genetik. Keputusan menunjukkan bahawa sela dan jurang merupakan parameter utama dalam peningkatan jarak terbang, manakala jurang dan aturan lebih dominan dalam meningkatkan daya tahan penerbangan. Pendekatan pengoptimuman yang dicadangkan menghasilkan

peningkatan sebanyak 10.93% dari segi jarak dan 46.63% dari segi daya tahan berbanding konfigurasi sayap berkembar asas.

KEYWORDS: *Biplane, UAV, Response surface method, Aerodynamic optimization, Genetic algorithm.*

1. INTRODUCTION

The first attempts at multi-wing aerial vehicle designs date back to the early 20th century, from the Wright Brothers' "Wright Flyer" in 1903. The most common multiple-wing configurations are the biplane and triplane, which are designated by the number of nonplanar wings. Today, the prevalence of these designs has diminished with advancements in materials science and, correspondingly, improvements in the aerodynamic efficiency of monoplanes [1]. Modern fixed-wing UAV designs commonly feature a single, often large, wing intended to be both lightweight and sufficiently structurally rigid to withstand aerodynamic loads, which presents a complex engineering challenge. Instead, multi-wing configurations could be a useful alternative, offering the advantage of distributing the required wing area across multiple shorter wings. Such a design is advantageous for operations such as indoor inspection, navigating dense urban canyons for delivery services, search-and-rescue operations after natural disasters involving collapsed structures, and close-proximity environmental monitoring in forests or industrial facilities. Furthermore, the reduced wingspan could also provide a lower moment of inertia and improved manoeuvrability by shifting the centre of wing mass closer to the longitudinal axis.

The presence of an additional wing in proximity significantly alters the pressure distribution and flow patterns around the wings. The increased aerodynamic interference leads to a reduction in the effective angle of attack, a change in lift distribution, an increase in induced drag, and an overall decrease in aerodynamic efficiency for multi-wings when compared with a monoplane generating equal lift. To enhance the aerodynamic performance of a biplane configuration, the wing arrangement is critical. The geometrical parameters of decalage, stagger, and gap interact sensitively, necessitating a meticulous design process to achieve optimal aerodynamic performance metrics such as C_L/C_D or $C_L^{3/2}/C_D$ for multi-wings [2, 3]. These parameters have a key role in the range and endurance of a piston-propelled aerial vehicle. In this context, the Response Surface Methodology (RSM) is an appropriate Design of Experiments (DoE) approach for this task, as it can model and analyse relationships among multiple input variables and output responses. Furthermore, the method facilitates the generation of precise surrogate models with fewer computationally expensive simulations than traditional approaches. The generated surrogate model could be integrated into any well-suited optimization algorithm to obtain an optimal design.

In the scientific literature, Design of Experiments (DoE) and Response Surface Methodology (RSM) have been widely used to develop surrogate models for aerospace optimization. A significant number of existing studies focus on fundamental two-dimensional component optimization. Leifsson et al. [4], Immonen [5], and Kallath et al. [10] demonstrated the efficiency of RSM in optimizing airfoil shapes. However, the effectiveness of these optimizations often depends critically on the parametrization technique employed. For instance, while Immonen [5] used a comprehensive set of design candidates, Kallath et al. [10] showed that combining mesh morphing tools with PARSEC definitions enables precise Pareto optimality searches. Furthermore, specific operational environments require tailored DoE approaches. For instance, Win and Thianwiboon [11] successfully applied full factorial designs to capture the unique physics of airfoils in ground effect, highlighting the versatility of standard

DoE methods for constrained flow regimes. On the other hand, three-dimensional wing planforms introduce additional dimensionality, and evolutionary strategies remain popular for global search. For instance, Aelaei et al. [7] and Yao et al. [9] utilized Multi-Objective Genetic Algorithms (MOGA) for delta wings and UAV tail-sitters, respectively. However, recent trends indicate a shift toward gradient-based or hybrid methods to accelerate convergence. Yildirim and Tuncer [12] and Dam et al. [14] demonstrated that coupling surrogate models (such as RSM or Artificial Neural Networks) with Sequential Quadratic Programming (SQP) can efficiently handle parameters like sweep, taper, and twist. Dam et al. [14] noted that ANNs may offer greater flexibility than traditional polynomial-based RSM when training data are abundant.

The literature survey also demonstrated that, in addition to external aerodynamics, advanced surrogate modelling is of critical importance for complex, multidisciplinary interactions. Studies by Tang et al. [6] on turbomachinery and Simmons [13] on distributed propulsion suggest that standard RSM must be enhanced to handle highly non-linear flow physics. Tang et al. [6] addressed this issue by developing a Gradient-Enhanced Response Surface Method (GERSM), proving that incorporating adjoint sensitivity analysis significantly reduces computational cost compared to traditional genetic algorithms. Similarly, applications on Vertical Axis Wind Turbines (VAWT) by Zhang et al. [8] and C-wing configurations by Samputh et al. [15] confirm that robust DoE techniques (such as Orthogonal or Fractional Factorial designs) are essential for isolating the most significant design parameters in complex flow fields.

In summary, the literature review demonstrates that, despite the extensive use of surrogate-based optimization in aerodynamic design, there remains limited literature characterizing the complex interference phenomena associated with biplane UAV configurations. While monoplane planforms have been extensively documented, the impact of coupled variations in gap, stagger, and decalage on flight performance metrics remains to be carefully investigated using modern computational frameworks.

In this context, the present study covers the aerodynamic optimization of a piston-propelled biplane UAV wing by implementing a global optimization strategy utilizing Response Surface Methodology (RSM) and Genetic Algorithms (GA), aiming for achieving maximum lift-to-drag ratio (C_L/C_D) and $C_L^{3/2}/C_D$ by optimal selection of decalage, stagger, and gap distances. The numerical aerodynamic analysis method is validated against wind-tunnel data from the literature, and a surrogate model is developed using RSM to examine the effects of the design parameters on C_L/C_D and $C_L^{3/2}/C_D$. Finally, the surrogate model is used within a genetic algorithm to obtain optimal configurations with respect to decalage, stagger, and gap. The significance of this work lies in substantiating the feasibility of biplane architectures for missions characterised by span constraints and in demonstrating that precise geometric tailoring yields superior performance gains compared with baseline configurations.

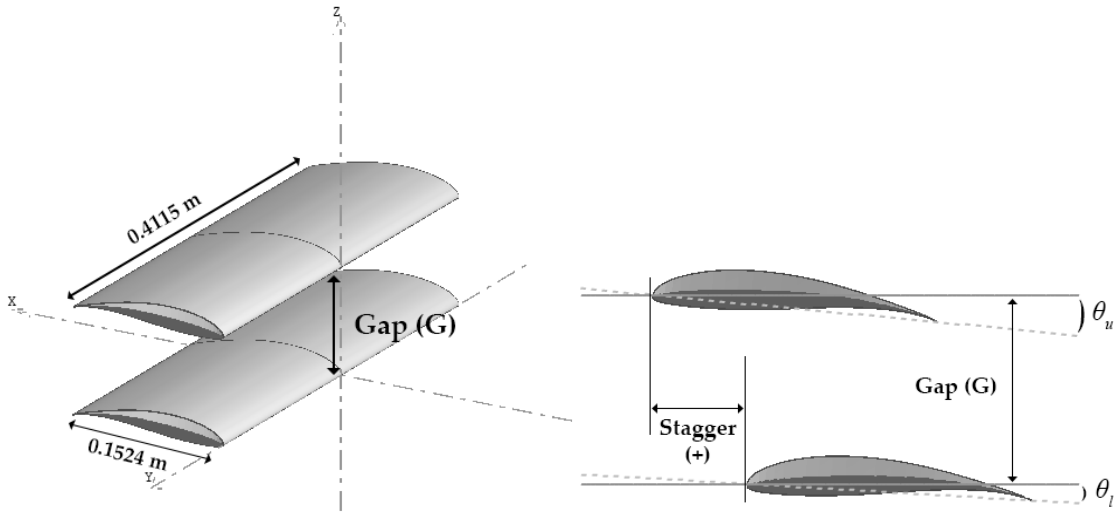
2. MATERIAL AND METHOD

2.1. Wing Geometry and Design Parameters

The biplane wing geometry with a 0.4115 *m* wing span, 0.1524 *m* chord length, and Wortmann FX 63-137 airfoil is illustrated in Fig. 1a, where the vertical distance between the upper and lower wing chord lines is denoted as gap (G), a well-known design parameter for multi-wing configurations. The term stagger (S) is the longitudinal distance between the upper and lower wings, and decalage angle (δ) is the relative angle of incidence between the upper and lower wings, as shown in Fig. 1b. The stagger is positive when the upper wing is further

forward than the lower wing, and negative if opposite. The decalage angle can be obtained using Eq. (1).

$$\delta = \theta_u - \theta_l \tag{1}$$



(a) (b)

Figure 1. a) Three-dimensional view and dimensions of the biplane wing,
b) Side view of the biplane wing with stagger, gap, and decalage definitions

2.2. Vortex Lattice Method

The aerodynamic performance of the biplane wing was evaluated using the Vortex Lattice Method (VLM), implemented in the publicly licensed XFLR5 software [16]. The VLM is a standard tool, particularly in the aircraft conceptual design phase, for estimating lift and lift distribution, drag and its components, and moment coefficients. In the VLM, the flow is assumed to be inviscid, incompressible, and irrotational; however, viscous effects can be accounted for using two-dimensional viscous airfoil polars. The wing is divided into a finite number of panels so that the Biot-Savart law can be applied; consequently, the induced velocity and flow tangency are provided at each control point.

To ensure the accuracy of the aerodynamic analysis, the base monoplane wing used in wind-tunnel experiments reported in the literature [17] was analysed over the range -2 to 18 degrees of angle of attack at a Reynolds number of 200,000 and sea-level conditions. The experimental and VLM results were comparatively presented in Fig. 2a and 2b for lift and drag coefficients, respectively. It is clear that, especially at lower angles of attack from -2 to 6 *degrees*, the results were satisfactory within an aircraft conceptual design framework. Moreover, the slopes of the curves are also remarkably promising in both figures, compared with the experimental data. The discrepancy observed between the VLM and the experimental results in the higher-angle-of-attack regime (10 to 15 degrees) is attributed to the well-known limitations of the Vortex Lattice Method. Since the VLM assumes the flow is mainly inviscid, incompressible, and irrotational, viscous effects can be added using two-dimensional viscous airfoil polars. As a result, it accurately predicts aerodynamic coefficients in the linear region (low to moderate angles of attack). However, when the angle of attack exceeds 10 degrees, the flow exhibits significant nonlinear viscous phenomena, including boundary-layer thickening and flow separation (pre-stall conditions). Because the standard VLM formulation does not account for flow separation or complex viscous boundary-layer interactions, it diverges from experimental data in this nonlinear pre-stall regime. As the optimization study primarily targets

the cruise condition (linear region), VLM can be considered sufficiently accurate for conceptual design.

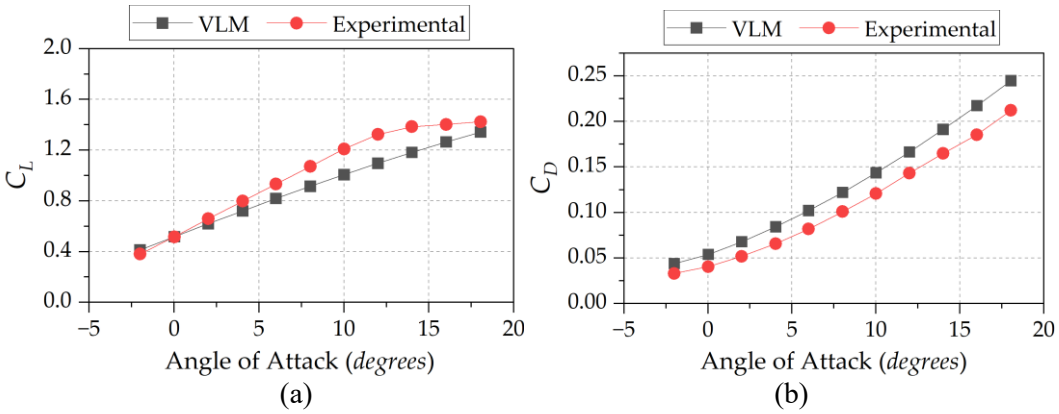


Figure 2. a) The variation in lift coefficient with angle of attack, b) The variation in drag coefficient with angle of attack

The number of panel elements plays a crucial role in panel methods such as VLM. In this context, it varied from 200 to 11,250 to ensure that the results are independent of the number of panels, while the spanwise and chordwise distribution was conserved. The results of analyses at a Reynolds number of 200,000 and sea-level conditions, with a 0-*degree* angle of attack, as shown in Fig. 3, indicate that approximately 4050 elements were sufficient to yield desirable results for both lift and drag coefficients.

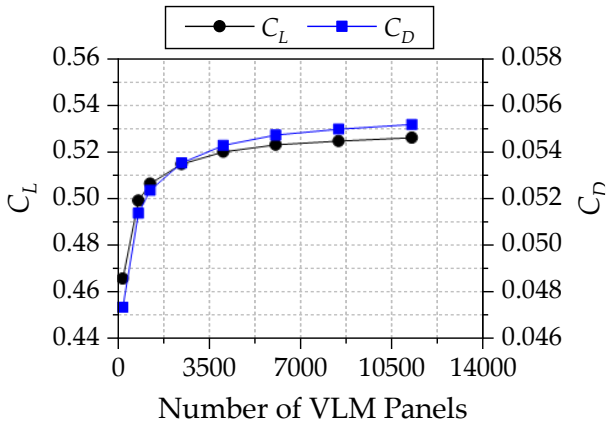


Figure 3. The variation in lift and drag coefficients with the number of VLM panels

2.3. Surrogate Model Generation using Response Surface Method

The Design of Experiment (DoE) method is a statistical approach used to generate and analyse controlled experiments to identify significant factors and optimize a process or system with minimal experimental effort. There are various types of DoE, such as Full Factorial Design, Plackett–Burman Design, Response Surface Method, Taguchi Design, and Latin Square Design [18].

Response Surface Method offers several key advantages for optimizing and evaluating nonlinear complex relationships. Compared with a full factorial design, RSM can fit the same model with fewer experiments and identify the optimal combination of input factors to achieve the desired outputs. Furthermore, RSM includes quadratic (curved) terms, allowing it to model curved or complex surfaces and capture not only main effects but also interactions among the

variables under investigation. There are various options in RSM, with the Central Composite Design (CCD) and the Box-Behnken Design (BBD) being the most common. While CCD is more flexible with an adjustable number of centre and axial points, BBD is more efficient for a small number of factors; however, it becomes more expensive at higher numbers. Detailed information about the methodology could be found in [19, 20].

In this manuscript, RSM analysis is to be performed in the commercially available Minitab software [21] to generate a surrogate model, a mathematical equation using variables G, S, and δ, capable of estimating C_L/C_D and $C_L^{3/2}/C_D$ values. For this purpose, CCD is applied to evaluate more extensive scenarios, including those beyond the constraints presented in Table 1. The gap distance was limited from 0.25*c* to 1.25*c*, while the stagger was from -*c* to *c*, where *c* is the chord length. The limits are defined to fully cover potential critical aerodynamic interactions, such as inter-wing interactions and monoplane-like behaviour, based on similar studies in the literature [22, 23]. The designed experiments are presented in Table 2, in which all variables exceed the constraints, enabling estimation of a broader range of configurations.

Table 1. Design constraints in RSM analysis

	Gap [*m*]	Stagger [*m*]	Decalage [*degrees*]
Minimum value	0.0381	-0.1524	-6
Maximum value	0.1905	0.1524	6

Table 2. Designed experiments for RSM-analysis

Number of Experiments	Gap [m]	Stagger [m]	Decalage [degrees]
1	0.248487	0	0
2	0.195	-0.1524	-6
3	0.11655	0	0
4	-0.015387	0	0
5	0.11655	0	0
6	0.195	0.1524	-6
7	0.0381	0.1524	6
8	0.11655	0	10.0908
9	0.11655	0	0
10	0.11655	0.256305	0
11	0.0381	0.1524	-6
12	0.0381	-0.1524	6
13	0.11655	0	0
14	0.11655	-0.256305	0
15	0.11655	0	0
16	0.11655	0	0
17	0.0381	-0.1524	-6
18	0.11655	0	-10.0908
19	0.195	0.1524	6
20	0.195	-0.1524	6

2.4. Genetic Algorithm for Optimization

Genetic Algorithm (GA) is a population-based, stochastic search method inspired by the principles of natural selection and genetics. It is particularly effective for solving nonlinear, multimodal, and complex optimization problems for which gradient-based methods are less effective or inapplicable. In this study, a Genetic Algorithm (GA) was employed to optimize the objective functions with respect to performance parameters related to the lift and drag coefficients. Further information on the algorithm's theoretical background is provided in [24].

In our case, the variables gap, stagger, and decalage were to be optimally selected, with the objective function defined as the maximization of C_L/C_D and $C_L^{3/2}/C_D$ separately, using the surrogate model generated via RSM. These aerodynamic parameters are directly proportional and crucially determine the range (R) and endurance (E) of piston-propelled aerial vehicles. The equations were given in Eq. (2) and Eq. (3), where η is the propulsive efficiency, SFC is the specific fuel consumption, ρ is the air density, S is the wing area, and W is the weight [24].

$$R = \frac{\eta}{SFC}\frac{C_L}{C_D}\ln\left(\frac{W_{initial}}{W_{final}}\right) \tag{2}$$

$$E = \frac{2}{SFC_t}\frac{{C_L}^{3/2}}{C_D}\sqrt{2\rho S}\ln\left({W_{final}}^{-1/2} - {W_{initial}}^{-1/2}\right) \tag{3}$$

The constraints were defined to conserve or enhance the lift coefficient relative to the base model, thereby providing adequate lift while improving performance, as given in Equations (4), (5), and (6).

$$C_{L_{base}} \leq C_{L_{optimal}} \tag{4}$$

$$\left(\frac{C_L}{C_D}\right)_{base} \leq \left(\frac{C_L}{C_D}\right)_{optimal} \tag{5}$$

$$\left(\frac{{C_L}^{3/2}}{C_D}\right)_{base} \leq \left(\frac{{C_L}^{3/2}}{C_D}\right)_{optimal} \tag{6}$$

The optimization process begins by generating an initial population of potential solutions (individuals), with each individual represented by a chromosome encoding a candidate solution to the problem. The algorithm was iterated in MATLAB until a termination criterion was met, either a fixed number of generations or a convergence threshold based on changes in fitness.

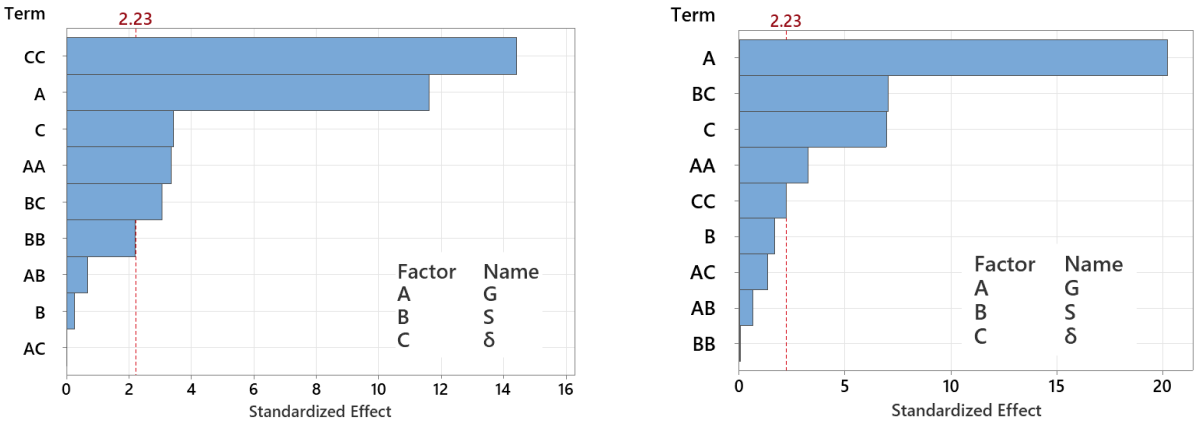
3. RESULTS AND DISCUSSION

3.1. RSM Analysis Results and Surrogate Models

The Pareto diagrams are illustrated in Fig. 4a and 4b, which present the relative importance of factors and interactions in affecting the responses C_L/C_D and $C_L^{3/2}/C_D$, respectively. From a lift-to-drag ratio perspective, the stagger has a minimal, nonlinear effect, whereas decalage has both linear and nonlinear strong effects on the response. In the case of $C_L^{3/2}/C_D$, stagger has only a dependent effect, interacting with decalage, while gap distance and decalage have a dramatic impact on the response.

The R-square (R^2), R-squared adjusted (R^2-adj), and R-squared predicted (R^2-pred) values in Table 3 indicate the model's predictive capability. Moreover, the p-values in Table 4 indicate the separate and combined effects of factors on the objective functions, with small p-values (typically ≤ 0.05) providing strong evidence against the null hypothesis. The results showed that the stagger and the interaction terms GS and $G\delta$ had no significant effect on either objective function. The terms δ^2 and S^2 have a slight effect on the lift-to-drag ratio and $C_L^{3/2}/C_D$, respectively, while S^2 has no considerable effect on $C_L^{3/2}/C_D$.

https://doi.org/10.31436/iiumej.v27i1.3980



(a) (b)

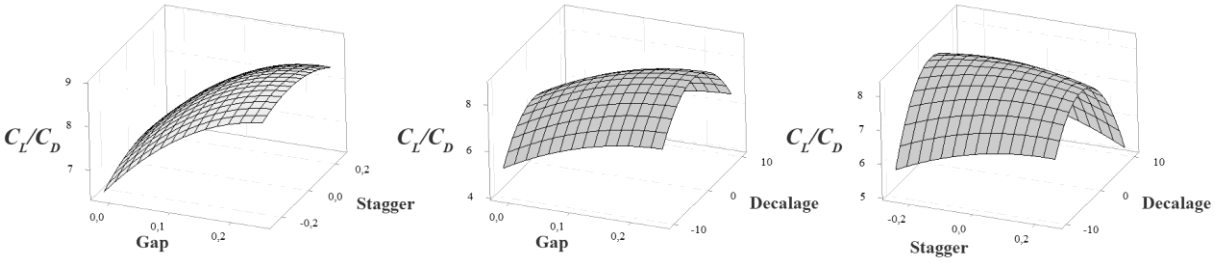
Figure 4. a) Pareto diagram for C_L/C_D, b) Pareto diagram for $C_L^{3/2}/C_D$

Table 3. The model summary

Objective Functions	R^2	R^2-adj	R^2-pred
C_L/C_D	97.35%	94.97%	79.93%
$C_L^{3/2}/C_D$	98.14%	96.46%	83.79%

Table 4. The probability characteristics of variables

	P-values for Responses	
Parameters	C_L/C_D	$C_L^{3/2}/C_D$
G	0.000	0.000
S	0.799	0.125
δ	0.006	0.000
G^2	0.007	0.009
S^2	0.052	0.949
δ^2	0.000	0.051
GS	0.515	0.527
Gδ	0.985	0.208
Sδ	0.012	0.000



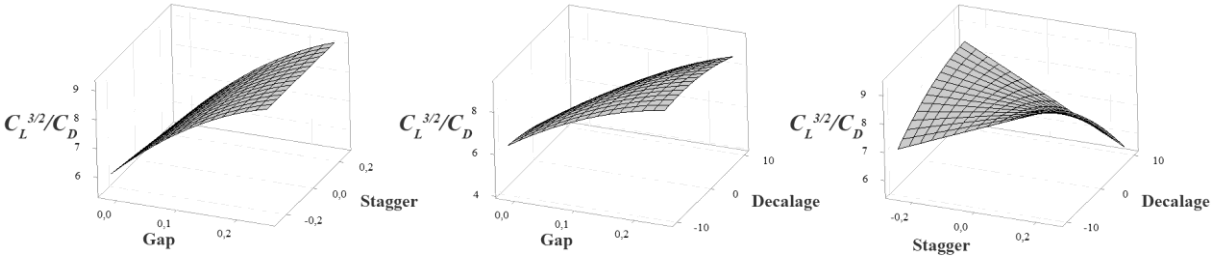
(a) (b) (c)

Figure 5. a) The response surface plot for gap and stagger for C_L/C_D, b) The response surface plot for gap and decalage for C_L/C_D, c) The response surface plot for stagger and decalage for C_L/C_D

The response surface plots for C_L/C_D and $C_L^{3/2}/C_D$ are shown in Fig. 5 and visually depict the relationships between the two independent variables and the response variables. In Fig. 5a, the maximum lift-to-drag ratio was clearly enhanced by increasing gap distance, while stagger had nearly no effect, parallel with its p-value. Similarly, in Fig. 5b, the larger gap distance contributed to the lift-to-drag ratio, whereas decalage required a moderate value to achieve this

improvement. Fig. 5c represented the slight effect of stagger on enhancing the lift-to-drag ratio, where decalage should be conserved moderately.

As shown in Fig. 6a, the graph was significantly dependent on the gap value, as indicated by the reported p-values. Similarly, in Fig. 6b, the increased gap distance was attributed to $C_L^{3/2}/C_D$, while a minimal effect was observed for the decalage. Fig. 6c showed that, for a smaller stagger, a higher decalage was required to enhance $C_L^{3/2}/C_D$.



(a) (b) (c)

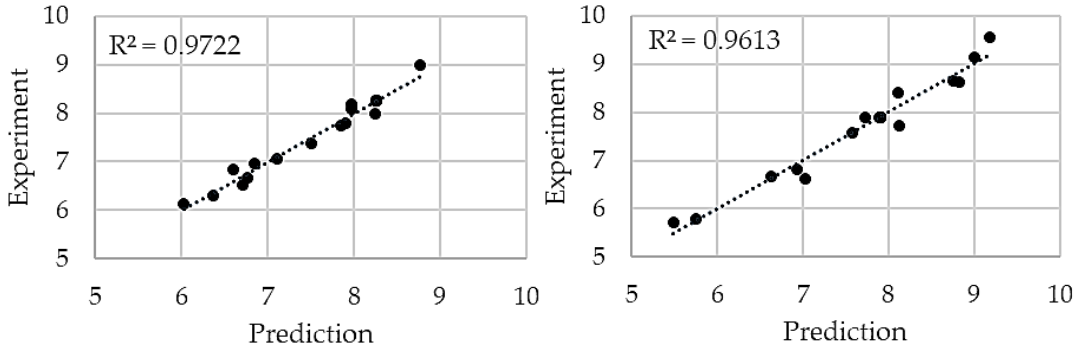
Figure. 6. a) The response surface plot for gap and stagger for $C_L^{3/2}/C_D$, b) The response surface plot for gap and decalage for $C_L^{3/2}/C_D$, c) The response surface plot for stagger and decalage for $C_L^{3/2}/C_D$

The surrogate models generated via RSM regression are shown in Eqs. (7) and (8) for each aerodynamic performance parameter, as functions of the variables G, S, and δ. Terms with p-values < 0.05 were removed from the models to simplify them.

$$\frac{C_L}{C_D} = 7.061 + 13.39G - 0.0285\delta - 26.21G^2 - 4.55S^2 - 0.01919\delta^2 - 0.2148S\delta \quad (7)$$

$$\frac{{C_L}^{3/2}}{C_D} = 6.077 + 18.66G - 0.0786\delta - 25.47G^2 - 0.00295\delta^2 - 0.4959S\delta \quad (8)$$

The predictions from the surrogate model were validated, and the results were compared with the designed experiments shown in Fig. 7, yielding satisfactory R2 values. The average difference between the regression predictions and analysis results were 0.0015% and 0.0441%, while the maximum differences were 3.52% and -5.76% for C_L/C_D and $C_L^{3/2}/C_D$, respectively.



(a) (b)

Figure 7. a) Designed experiment results and surrogate model predictions for C_L/C_D, b) Designed experiment results and surrogate model predictions for $C_L^{3/2}/C_D$

3.2. Optimization Results

The base biplane model is determined to have no decalage or stagger, while having a gap distance of 0.5c. The upper and lower limits of the optimization algorithm were determined as values given in Table 5, where a maximum gap of one chord length was imposed due to

https://doi.org/10.31436/iiumej.v27i1.3980

structural considerations, and the decalage angle was limited to positive and negative six degrees to avoid stall onset.

The results of the genetic algorithm optimization are presented in Table 6, along with those of the base model. To maximize the lift-to-drag ratio, the algorithm increased the gap distance, as shown in the Pareto diagram, while applying moderate stagger and a slight negative decalage. In the case of $C_L^{3/2}/C_D$, the gap distance and stagger are maximized, while decalage is minimized within the limits.

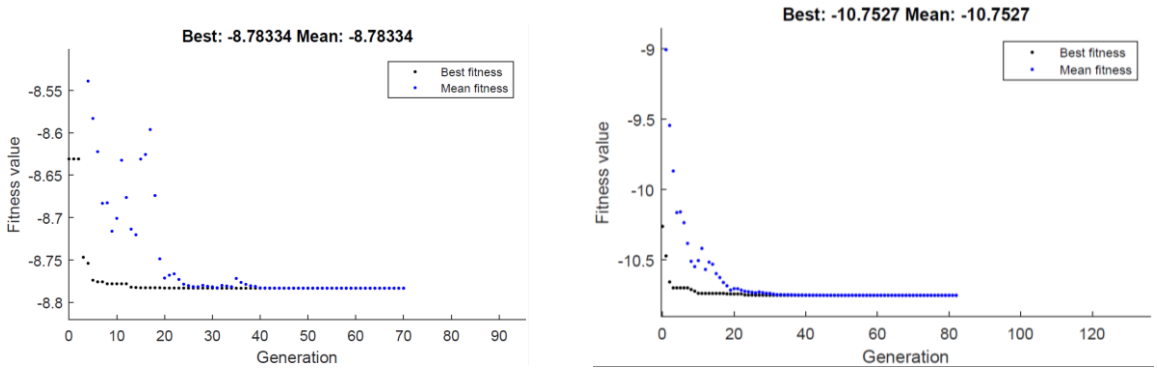
Table 5. The constraints defined in the genetic algorithm

	***G* [*m*]**	***S* [*m*]**	***δ* [*degrees*]**
Minimum value	0	-0.3	-6
Maximum value	0.152	0.3	6

Table 6. The base and optimal configurations

Model	**Variable**				
	***G* [*m*]**	***S* [*m*]**	***δ* [*degrees*]**	C_L/C_D	$C_L^{3/2}/C_D$
Base	0.075	0	0	7.872	7.299
***Max.* C_L/C_D**	0.2554	0.02	-0.8557	8.7833	9.254
***Max.* $C_L^{3/2}/C_D$**	0.3664	0.3	-6	8.402	10.7527

The results indicated that a greater gap distance contributes to both range and endurance by reducing wing-to-wing interaction. Similarly, higher stagger values are associated with greater endurance. Correspondingly, optimization results favor higher gap and stagger values to minimize the adverse aerodynamic interference between the wings. Ideally, eliminating this interference would reveal the monoplane's efficiency. However, the justification for the biplane configuration in this study is driven by geometric and structural constraints, specifically the restriction on wing span for mini-UAV applications. To achieve the required wing loading and lift capacity within a confined span, a monoplane would require a significantly larger chord, resulting in a low Aspect Ratio and high induced drag. The biplane configuration addresses this problem, providing twice the lifting-surface area within the same span constraint while maintaining a higher effective aspect ratio than a comparable monoplane.



(a) (b)

Figure. 8. The variation in fitness value with number of generations for C_L/C_D, b) The variation in fitness value with number of generations for $C_L^{3/2}/C_D$

The base C_L/C_D and $C_L^{3/2}/C_D$ values were improved by 10.93% and 46.63%, respectively, according to surrogate-model-based optimization results, which are validated in the following section. The iterative optimization process is shown in Fig. 8, where fitness values are plotted against the number of generations.

3.3. Validation of the Optimal Configuration

The validity of the surrogate-model-based optimization results was assessed by aerodynamic analyses of each optimal configuration. The results are presented in Table 7, where the differences between surrogate model predictions and VLM analyses for C_L/C_D and $C_L^{3/2}/C_D$ were 2.17% and -8.36%, respectively, for optimal configurations. The maximum-range and maximum-endurance models have lift coefficients of 1.048 and 1.375, respectively, which are higher than the base model's 0.86, as intended.

Table 7. The base and optimal configurations

Model	Surrogate Model Prediction		VLM	
	C_L/C_D	$C_L^{3/2}/C_D$	C_L/C_D	$C_L^{3/2}/C_D$
Base	7.917	7.333	7.872	7.299
***Max.* C_L/C_D**	8.783	9.254	8.974	8.894
***Max.* $C_L^{3/2}/C_D$**	7.905	10.752	8.402	9.854

4. CONCLUSIONS

The present study focused on surrogate-model-based aerodynamic optimization of a piston-propelled mini biplane UAV wing to maximize the vehicle's range and endurance. Initially, the aerodynamic analysis method was validated on an existing monoplane wing design and compared with wind-tunnel results in the literature. Subsequently, a base biplane model was constructed, and the effects of multi-wing geometric parameters, namely decalage and stagger, were investigated. The gap distance was investigated using a central composite design in RSM, and surrogate models for estimating C_L/C_D and $C_L^{3/2}/C_D$ were developed. The surrogate model equations were simplified using Pareto analysis and incorporated into a genetic algorithm to obtain optimal sets of decalage, stagger, and gap distances that maximize range and endurance. Finally, the results from the surrogate model were compared with those from VLM analyses, yielding differences of 2.17% and -8.36% for C_L/C_D and $C_L^{3/2}/C_D$, respectively. The total improvement in C_L/C_D and $C_L^{3/2}/C_D$ values was 10.93% and 46.63%, respectively. Although the study has presented the aerodynamic effects of the design variables and provided optimal designs within the constraints, the structural limitations should be evaluated, particularly when significant gaps and stagger values are considered alongside aerodynamic considerations, to achieve more efficient optimization in a subsequent study.

REFERENCES

[1] Guo T, Feng L, Zhu C, Zhou X, Chen H. (2022) Conceptual Research on a Mono-Biplane Aerodynamics-Driven Morphing Aircraft. Aerospace, 9(7):380. https://doi.org/10.3390/aerospace9070380.

[2] Nguyen TD, Kashitani M, Taguchi M, Kusunose K. (2022) Effect of Stagger on Low-Speed Performance of Busemann Biplane Airfoil. Aerospace, 9(4):197. https://doi.org/10.3390/aerospace9040197.

[3] Guo R, Bai Y, Pei X, Lai Z. (2021) Numerical investigation of aerodynamics and wake on biplane airfoils at high angles of attack. International Journal of Mechanical Sciences, 205:106606. https://doi.org/10.1016/j.ijmecsci.2021.106606.

[4] Leifsson L, Koziel S, Tesfahunegn YA. (2016) Multiobjective aerodynamic optimization by variable-fidelity models and response surface surrogates. AIAA Journal, 54(2):531–541. https://doi.org/10.2514/1.J054128.

[5] Immonen E. (2017) 2D shape optimization under proximity constraints by CFD and response surface methodology. Applied Mathematical Modelling, 41:508–529. https://doi.org/10.1016/j.apm.2016.09.009.

[6] Tang X, Luo J, Liu F. (2017) Aerodynamic shape optimization of a transonic fan by an adjoint-response surface method. Aerospace Science and Technology, 68:26–36. https://doi.org/10.1016/j.ast.2017.05.005.

[7] Aelaei M, Karimian S, Ommi F. (2019) Sensitivity analysis and optimization of delta wing design parameters using CFD-based response surface method. Journal of Applied Fluid Mechanics, 12(6):1885–1903. https://doi.org/10.29252/jafm.12.06.29706.

[8] Tian-tian Z, Mohamed E, Wei H, Zhen-guo W, Derek BI, Lin M, Mohamed P. (2019) Winglet design for vertical axis wind turbines based on a design of experiment and CFD approach. Energy Conversion and Management, 195:712–726. https://doi.org/10.1016/j.enconman.2019.05.055.

[9] Yao X, Liu W, Han W, Li G, Ma Q. (2020) Development of response surface model of endurance time and structural parameter optimization for a tailsitter UAV. Sensors, 20(6):1766. https://doi.org/10.3390/s20061766.

[10] Hariharan K, Jun SL, Foster KK, Man YH, June KM. (2021) A multi-objective airfoil shape optimization study using mesh morphing and response surface method. Journal of Mechanical Science and Technology, 35:1075–1086. https://doi.org/10.1007/s12206-021-0221-0.

[11] Win SY, Thianwiboon M. (2021) Parametric optimization of NACA 4412 airfoil in ground effect using full factorial design of experiment. Engineering Journal, 25(12):9–19. https://doi.org/10.4186/ej.2021.25.12.9.

[12] Yildirim BY, Tuncer H. (2021) Wing planform optimization using flow solutions and response surface methodology. AIAA Aviation Forum, p.2564. https://doi.org/10.2514/6.2021-2564.

[13] Simmons BM. (2023) Evaluation of response surface experiment designs for distributed propulsion aircraft aero-propulsive modelling. AIAA SciTech Forum, p.2551. https://doi.org/10.2514/6.2023-2251.

[14] Dam B, Pirasaci T, Kaya M. (2022) Artificial neural network based wing planform aerodynamic optimization. Aircraft Engineering and Aerospace Technology, 94(10):1731–1747. https://doi.org/10.1108/AEAT-10-2021-0311.

[15] Samputh HH, Moey LK, Tai VC, Tan YC. (2024) Investigation of aerodynamic characteristics of swept C-wing configurations at transonic speed using design of experiments and computational fluid dynamics. Aviation, 28(2):72–84. https://doi.org/10.3846/aviation.2024.21495.

[16] Analysis of foils and wings operating at low Reynolds numbers. (2025) Available at: https://aero.us.es/adesign/Slides/Extra/Aerodynamics/Software/XFLR5/XFLR5%20v6.10.02/Guidelines.pdf (Accessed 19 July 2025).

[17] Bastedo WG, Mueller TJ. (1986) Spanwise variation of laminar separation bubbles on wings at low Reynolds number. Journal of Aircraft, 23(9):687–694. https://doi.org/10.2514/3.45363.

[18] Durakovic B. (2017) Design of experiments application, concepts, examples: State of the art. Periodicals of Engineering and Natural Sciences, 5(3):421-439. https://doi.org/10.21533/pen.v5i3.145.

[19] Myers RH, Montgomery DC. (1996) Response Surface Methodology. Taylor and Francis, USA.

[20] Antony J. (2023) Design of Experiments for Engineers and Scientists. Elsevier, UK.

[21] Alin A. (2010) Minitab. Wiley Interdisciplinary Reviews: Computational Statistics, 2(6):723–727. https://doi.org/10.1002/wics.113.

[22] Jones R., Cleaver D.J., Gursul I. (2015) Aerodynamics of biplane and tandem wings at low Reynolds numbers. Exp Fluids, 56: 124. https://doi.org/10.1007/s00348-015-1998-3.

[23] Cai Y., Liu G., Zhu X., Tu Q., Hong G. (2019) Aerodynamic interference significance analysis of two-dimensional front wing and rear wing airfoils with stagger and gap variations. Journal of Aerospace Engineering, 32(6): 04019098. https://doi.org/10.1061/(ASCE)AS.1943-5525.0001090.

[24] Şumnu A, Eraslan Y. (2025) Aerodynamic shape optimization of simplified ground vehicle (Ahmed Body) using passive control devices. Mechanics, 31(1):5–14. https://doi.org/10.5755/j02.mech.36452.

[25] Pogosyan MA. (2025) Aircraft Design. Springer Nature. https://doi.org/10.1007/978-981-96-4599-2.

https://doi.org/10.31436/iiumej.v27i1.3587

Stability and Control of Humanoid Robots during Walking and Sudden Stops on Uneven Terrain using Inverted Pendulum Modeling and Fuzzy-enhanced Linear Quadratic Regulator

ANDI DHARMAWAN, JAZI EKO ISTIYANTO*, BAKHTIAR ALLDINO ARDI SUMBODO, MUHAMMAD AUZAN, OSKAR NATAN

Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, 55128, Indonesia

**Corresponding author: jazi@ugm.ac.id*

(Received: 21 January 2025; Accepted: 24 September 2025; Published online: 12 January 2026)

ABSTRACT: Humanoid robots, designed to resemble human structure with a head, arms, and legs, stand out among various robot types due to their ability to interact with human-designed environments. Their kinematic structure, composed of links connected by joints, enables humanoid robots to perform tasks such as walking on uneven terrain, lifting objects, and opening doors. Walking is a critical function for humanoid robots, but it shifts their center of mass (CoM), which can lead to instability and falls. The inverted pendulum model is a widely used approach to humanoid robot walking, particularly on uneven surfaces, enabling CoM projection and balance adjustment. However, disturbances from floor unevenness, mechanical issues, and other factors can still cause the robot to fall. This underscores the need for effective control systems to stabilize movement and maintain balance. The Linear Quadratic Regulator (LQR) is a suitable control approach for humanoid robots, particularly in multiple-input, multiple-output systems. It stabilizes the robot while managing computational load, generating complex humanoid movements through linear approximations. To address uncertainties and adapt control gains, fuzzy control can be integrated, thereby enabling smoother transitions and improved disturbance handling. When walking on uneven terrain, a humanoid robot's body orientation may shift, causing the CoM to exceed tolerance limits and leading to falls. Thus, a control system that adjusts the walking pattern to maintain CoM stability is essential. In conclusion, humanoid robots, with their human-like structure, can perform complex tasks in human environments. Control systems, such as LQR and fuzzy control, are critical for maintaining balance and stability, even under challenging conditions, thereby enabling humanoid robots to keep their trajectory and prevent falls.

ABSTRAK: Robot seperti manusia (humanoid), direka menyerupai struktur manusia dengan kepala, lengan, dan kaki, menonjol dalam pelbagai jenis robot kerana keupayaannya berinteraksi dengan persekitaran yang direka oleh manusia. Struktur kinematiknya, terdiri daripada pautan yang dihubungkan oleh sendi, membolehkan robot humanoid melakukan tugas seperti berjalan di permukaan tidak rata, mengangkat objek, dan membuka pintu. Berjalan adalah fungsi penting robot humanoid, tetapi pergerakan mengubah pusat jisim (CoM), menyebabkan ketidakstabilan dan jatuh. Model bandul terbalik adalah pendekatan biasa digunakan bagi robot humanoid berjalan, terutama pada permukaan tidak rata, membolehkan unjuran CoM dan pelarasan keseimbangan. Namun, gangguan ketidaksamaan permukaan lantai, masalah mekanikal, dan faktor lain masih boleh menyebabkan robot jatuh. Ini menunjukkan keperluan pada sistem kawalan berkesan yang menstabilkan pergerakan dan mengekalkan keseimbangan. Regulator Kuadratik Linear (LQR) adalah pendekatan kawalan

https://doi.org/10.31436/iiumej.v27i1.3587

sesuai untuk robot humanoid, terutama dalam sistem masukan dan keluaran berganda. Ia menstabilkan robot sambil menguruskan beban pengiraan, menghasilkan pergerakan humanoid yang kompleks melalui pendekatan linear. Bagi menangani ketidakpastian dan menyesuai keuntungan kawalan, kawalan kabur boleh diintegrasikan, menjadikan peralihan lebih lancar dan pengendalian gangguan lebih baik. Apabila berjalan di permukaan tidak rata, orientasi badan robot humanoid mungkin berubah, menyebabkan CoM bergerak melebihi had toleransi, yang mengakibatkan jatuh. Oleh itu, sistem kawalan penyesuaian corak berjalan bagi mengekalkan kestabilan CoM adalah penting. Kesimpulannya, robot humanoid, dengan struktur seperti manusia, dapat melakukan tugas kompleks dalam persekitaran manusia. Sistem kawalan seperti LQR dan kawalan kabur adalah penting bagi memastikan keseimbangan dan kestabilan, walaupun dalam keadaan mencabar, membolehkan robot humanoid mengekalkan trajektori mereka dan mencegah jatuh.

KEYWORDS: *humanoid, robot, CoM, LQR, fuzzy logic, control*

1. INTRODUCTION

Robots come in various forms, optimized to perform specific functions and tasks [1]. Among the different types of robots, humanoid robots stand out due to their human-like structure [2][3]. Structurally, humanoid robots typically have a head, two arms, and two legs. Kinematically, these robots consist of links connected at joints that function like human joints, forming kinematic chains [4]. The resemblance to the human form enables humanoid robots to interact with equipment and environments designed for humans, such as walking on uneven terrain, lifting objects, opening doors, and running [5].

Walking is a fundamental capability for humanoid robots to interact with and adapt to human-designed environments effectively [6]. While a humanoid robot is stable when both feet are on the ground, movement causes its center of mass (CoM) to shift, which can lead to a loss of balance and potentially a fall [7]. One of the most common walking models for humanoid robots, particularly on uneven surfaces, is the inverted pendulum model. This model places the robot's CoM above its foot, which acts as the pivot, resulting in a projected motion of the CoM during walking [8]. Based on this inverted pendulum model, the foot placement can be calculated to balance the robot's body movement during walking [9].

However, despite accounting for balance during walking, humanoid robots can still fall due to various disturbances, such as uneven floors, mechanical misalignments, or other factors [10]. A sloped walking surface can cause the robot to fall as the CoM shifts towards the gravitational force acting on the inclined plane. This highlights the need for a control system to stabilize the humanoid robot's movement and maintain balance in line with the intended walking pattern.

Control systems play a crucial role in regulating a robot's motion, ensuring it follows the planned walking pattern. One modern control system suitable for humanoid robot walking, particularly in systems with multiple inputs and outputs, is the Linear Quadratic Regulator (LQR) [11][12]. LQR can address disturbances within the system and stabilize it while maintaining lightweight control with manageable computational requirements [13]. It uses linear approximations to generate complex humanoid movements.

To handle uncertainties and adapt control gains in response to disturbances, a fuzzy control system can be implemented. Fuzzy control addresses the limitations of rigid gain scheduling, providing smoother transitions and better adaptation to changing conditions. When walking on uneven terrain, the robot's body orientation may shift with each step, causing the CoM to move outside the tolerance limits, which can lead to a loss of balance and a fall. Therefore, a control

system that adjusts the walking pattern to maintain the CoM within these tolerance limits is essential for ensuring the robot remains upright on uneven surfaces.

In summary, humanoid robots, with their human-like structure, can perform activities like those of humans, such as climbing stairs, lifting objects, and opening doors. As these robots are further developed, they will be able to better adapt to human-designed environments [14]. The most common concept in humanoid robot modeling is the Linear Inverted Pendulum Model (LIPM), known for its simplicity and effectiveness in generating stable movement [15]. However, sudden stops or uneven surfaces can cause robots to lose balance, underscoring the need for control systems that can maintain stability during unexpected conditions. Using LQR and fuzzy control, humanoid robots can maintain balance, walk, and make sudden stops even in challenging environments, ensuring that their CoM remains stable and within the desired trajectory. This paper is organized as follows: Section 2 outlines the problem statement; Section 3 describes the system design; Section 4 presents experimental results and performance analysis; and Section 5 concludes.

2. PROBLEM STATEMENT

Humanoid robots face significant challenges in maintaining balance, especially during sudden stops or when walking on inclined surfaces. When a robot is abruptly stopped, it tends to lose balance due to unpredictable foot placement and reactive forces generated by the sudden deceleration. If the robot cannot quickly counteract these forces, it will fall. This issue is further complicated when the robot walks on inclined surfaces, where the projection of its center of mass (CoM) shifts with the incline, increasing the risk that the CoM moves outside the robot's support polygon and leads to a fall.

To address these challenges, the humanoid robot’s balance is controlled by calculating the required joint torques, specifically along the roll and pitch axes of the servos. A full state feedback control using the Linear Quadratic Regulator (LQR) method is applied to determine the necessary torque [16]. The Inertial Measurement Unit (IMU), which includes accelerometers and gyroscopes, is used to measure the robot's orientation and inclination angles (pitch and roll) during both walking and sudden stops.

The walking pattern of the humanoid robot is generated using forward kinematics, based on the Denavit-Hartenberg (DH) method, to compute the CoM trajectory. Inverse kinematics is then employed to determine the joint angles required to achieve the desired foot positions, thereby ensuring that the robot follows the predefined walking pattern.

To enhance stability, the system is designed to maintain the robot's CoM projection within the support polygon by employing a heel strategy that adjusts the ankle angles to preserve balance. The LQR controller optimizes the system by minimizing the error between the reference and actual states, producing torque commands that adjust the robot's motion to align with the reference walking pattern. This approach ensures that the humanoid robot can maintain balance during sudden stops and when navigating inclined surfaces, thereby preventing falls by continuously adjusting the CoM to keep it within the robot's support polygon.

3. SYSTEM DESIGN

3.1. System Model

A humanoid robot is a type of robot that resembles the human form. Structurally, a humanoid robot consists of one head, one body, two arms, and two legs. Due to their human-

https://doi.org/10.31436/iiumej.v27i1.3587

like structure, humanoid robots are continually being developed to perform and assist with human tasks and to operate human tools in human environments.

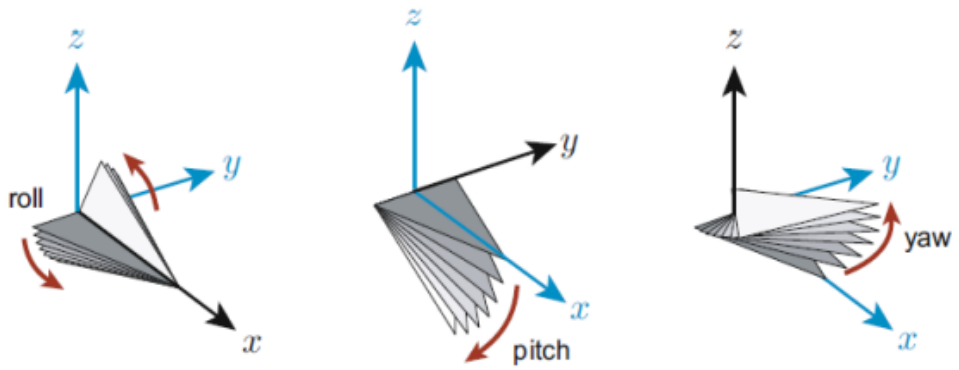


Figure 1. Roll, pitch, and yaw rotations [17]

Humanoid robots are powered by actuators, typically motors, with the number of motors corresponding to the desired degrees of freedom. The motors drive the robot's joints, which rotate about the yaw, pitch, or roll axes. Yaw rotation occurs along the z-axis, pitch rotation along the y-axis, and roll rotation along the x-axis. Figure 1 illustrates roll, pitch, and yaw rotations. Each humanoid robot has actuators that function like human joints. Every joint in the robot is referred to as a "joint," and each joint is connected to a rigid mechanical component called a "link." The combination of joints and links forms various parts of the humanoid robot's body, such as the arms, thighs, and knees.

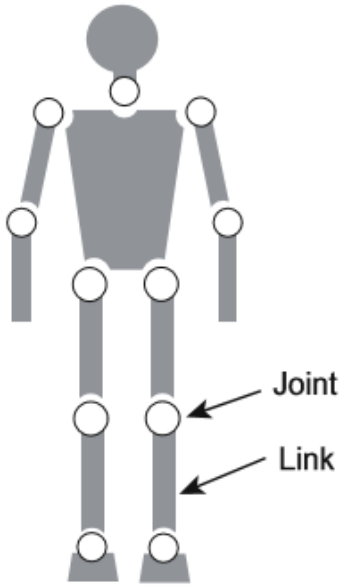


Figure 2. Link and joint of a humanoid robot

The humanoid robot used in this research is equipped with 12 servos, arranged with 10 in the legs and 2 in the arms. Figure 3 shows the robot's global coordinates from a rear view. The robot's x-axis points forward, the global y-axis points to the robot's left, and the z-axis points upwards.

The humanoid robot used in this research is a modified Bioloid Type-C robot, with a height of 32.3 cm and a total mass of 1.38 kg. Each servo on the robot represents a single rotational degree of freedom (DoF). The humanoid robot can be modeled as an inverted pendulum,

https://doi.org/10.31436/iiumej.v27i1.3587

simplifying the complex joints into a more straightforward representation. A single torque input to a servo represents one controlled degree of freedom. The configuration of the servo identification numbers for the humanoid robot in this research is shown in Figure 4.

A humanoid robot maintains its balance by counteracting the forces generated by changes in posture during leg swinging. The torque generated by the kicking leg alters the angles about the pitch and roll axes of the inverted pendulum model. The error caused by the change in body tilt must be zero for the robot to remain balanced; therefore, a torque opposite to the kicking torque is required on the robot's supporting leg.

The general equation for an inverted pendulum is shown in Eq. **Error! Reference source not found.**. An inverted pendulum is a pendulum where the mass is positioned above its pivot point [17]. The position of a human when standing is modeled as an inverted pendulum. The same concept is applied to humanoid robots, as shown in Figure 5. There are two assumptions underlying the modeling of a humanoid robot as an inverted pendulum. First, the mass of the entire body is concentrated at a single point called the Center of Mass (CoM). Second, this point mass is supported by a leg that is assumed to be massless, with the axis of rotation located at the ankle joint [18]. Figure 5 represents the model of an inverted pendulum where τ represents the torque of the pendulum. θ is the angle of pendulum deflection, r is the length of the pendulum, and M is the mass of the pendulum.

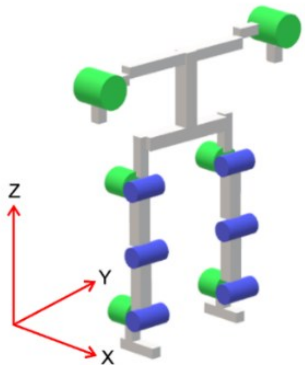


Figure 3. Global Coordinate Axes (x, y, z) of the Robot

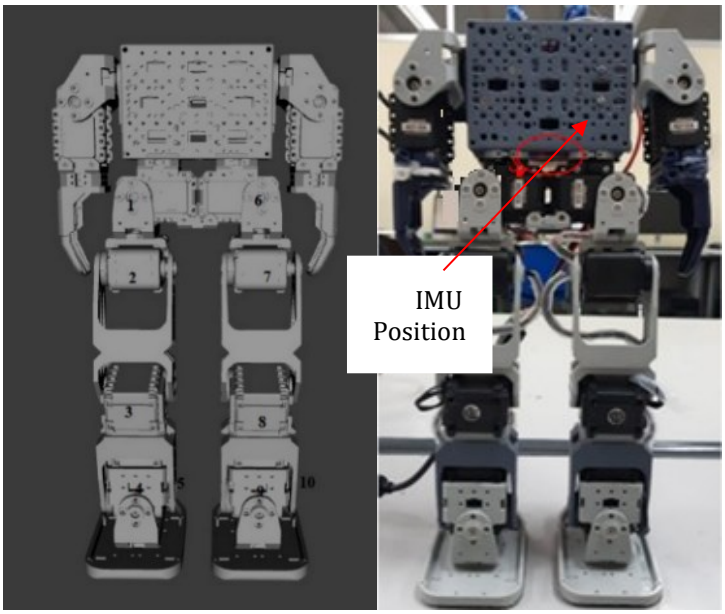


Figure 4. Humanoid Robot Configuration

https://doi.org/10.31436/iiumej.v27i1.3587

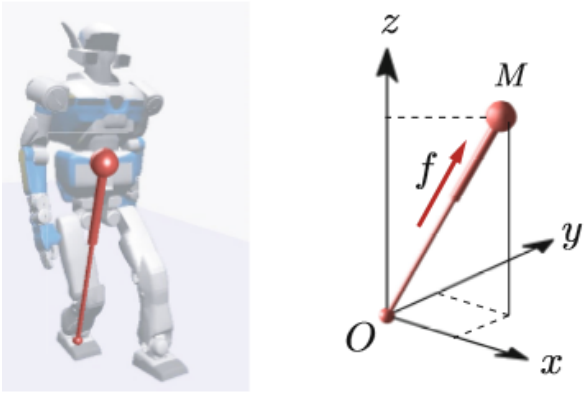


Figure 5. Inverted Pendulum Model

$$Mr\ddot{\theta} - Mgr \sin\theta = \tau \quad (1)$$

The deflection angle used in this study is 0.5°. $\sin\theta$ can be approximated as θ because the deflection angle affecting the humanoid robot's balance is relatively small. Thus, equation **Error! Reference source not found.** can be rewritten as Eq. **Error! Reference source not found.**.

$$I\ddot{\theta} - mgl\theta = \tau_{kick} \quad (2)$$

The inverted pendulum equation involves I, which is determined through mathematical calculations and defined as I_{xx} and I_{yy}. I_{xx} is the moment of inertia on the x-axis, while I_{yy} is the moment of inertia on the y-axis. Equations **Error! Reference source not found.** and **Error! Reference source not found.** show the equations for determining the moments of inertia I_{xx} and I_{yy}.

$$I_{xx} = \sum_{j=1}^{n}\left(I_{G_{xx_j}} + m_j\left(y_j^2 + z_j^2\right)\right) \quad (3)$$

$$I_{yy} = \sum_{j=1}^{n}\left(I_{G_{yy_j}} + m_j\left(x_j^2 + z_j^2\right)\right) \quad (4)$$

Eq. **Error! Reference source not found.** shows that I_{xx} is the moment of inertia on the x-axis. $I_{G_{xx_j}}$ is the moment of inertia of each robot component, represented as a solid block on the x-axis, m_j is the mass of each component, and y_j and z_j are the coordinates of the center of mass of each component relative to the robot's rotational axis. Meanwhile, in Eq. **Error! Reference source not found.**, I_{yy} is the moment of inertia on the y-axis. $I_{G_{yy_j}}$ represents the moment of inertia of each robot component, represented as a solid block on the x-axis, and m_j is the mass of each component. x_j and z_j are the coordinates of the center of mass of each component relative to the robot's rotational axis.

Eq. **Error! Reference source not found.** shows the moment of inertia of a solid block.

$$I_{G_{xx}} = \sum_{j=1}^{n}\left(I_{G_{xx_j}} + m_j\left(y_j^2 + z_j^2\right)\right) \quad (5)$$

$$I_{G_{yy}} = \frac{1}{12}m(b^2 + c^2) \quad (6)$$

https://doi.org/10.31436/iiumej.v27i1.3587

Based on Eqs. **Error! Reference source not found.** and **Error! Reference source not found.**, $I_{G_{XX_j}}$. and $I_{G_{yy_j}}$ are the inertias on the x and y axes of the body, respectively, for each component j; x_j and y_j are the distances of the center of mass of each component j from the axis of rotation on the x and y axes; and m_j is the mass of each component j [19].

The moment of inertia is a measure of a robot's resistance to changes in its rotational motion. The moment of inertia depends on the object's mass, shape, and size. The humanoid robot does not change mass, shape, or size; therefore, the change in its moment of inertia is very small and has almost no effect on its balance. Therefore, the moment of inertia value of the humanoid robot in this model is assumed to be constant.

The moment of inertia measures a robot's resistance to changes in its rotational motion and depends on the object's mass, shape, and size. Since the humanoid robot maintains constant mass, shape, and size, variations in its moment of inertia are minimal and have negligible effects on its balance. Therefore, in this model, the humanoid robot's moment of inertia is assumed to be constant.

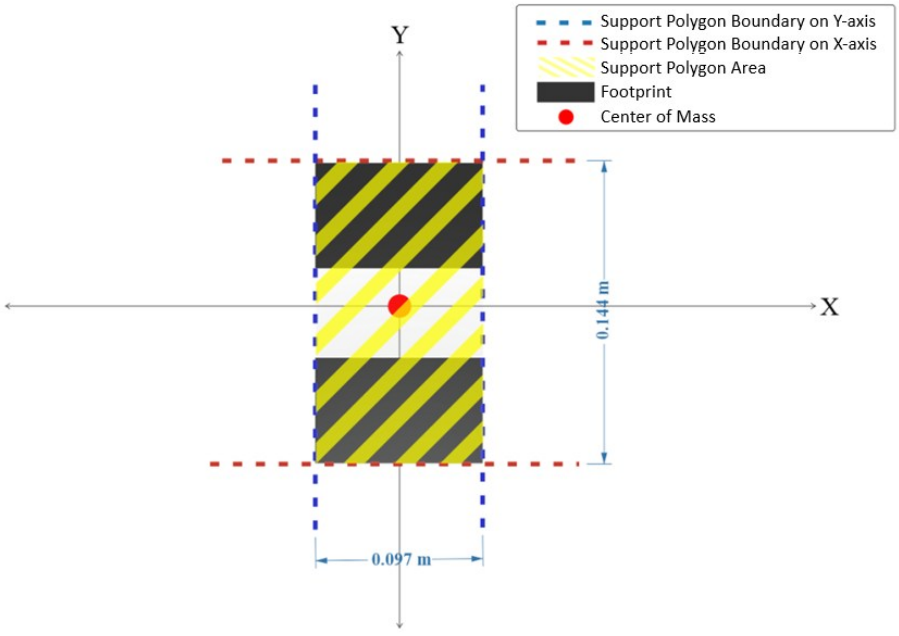
The center of mass (CoM) of each part of the humanoid robot is determined through forward kinematics calculations using the Denavit-Hartenberg (DH) method [20]. These calculations yield transformation matrices that encode the servo positions and rotation directions. The positions of the humanoid's CoM along the x, y, and z axes are calculated sequentially using the equations **Error! Reference source not found.**, **Error! Reference source not found.**, and **Error! Reference source not found.**.

$$x_{CoM} = \frac{\sum m_j x_j}{\sum m_j} \tag{7}$$

$$y_{CoM} = \frac{\sum m_j y_j}{\sum m_j} \tag{8}$$

$$z_{CoM} = \frac{\sum m_j z_j}{\sum m_j} \tag{9}$$

where m_j is the mass of each component, and x_j, y_j, and z_j are the distances of the center of mass of each robot component from the leg's supporting point along the respective axes [21].



https://doi.org/10.31436/iiumej.v27i1.3587

Figure 6. Calculation of the Robot's Support Polygon Area

The balance controller in this study employs a stabilizer control system to manage the CoM, ensuring it aligns with the walking pattern. The robot follows a static walking model, in which the center of mass never extends beyond the support polygon. For the robot used, tolerance limits or the maximum reach of the foot-support area (support polygon) are determined by calculating the support-polygon area for each foot, as illustrated in Figure 6.

Based on these measurements, the upper and lower tolerance limits for CoM movement along the robot's x and y axes are established. Specifically, the CoM projection relative to the support polygon has tolerance values of ±0.0485 meters along the global x-axis and ±0.072 meters along the global y-axis when the robot stands on two feet.

The humanoid robot's walking pattern is derived from a three-dimensional linear inverted pendulum equation, generating reference coordinates for the CoM movement along the x, y, and z axes. The CoM position is obtained from forward kinematics calculations using the DH convention. The walking pattern is divided into four phases:

- Shifting CoM: Preparing to lift a foot by shifting the CoM.
- Foot Lift: Lifting the foot off the ground.
- Stepping and Placement: Moving the foot forward and placing it down.
- CoM Adjustment: Adjusting the CoM to the final position.

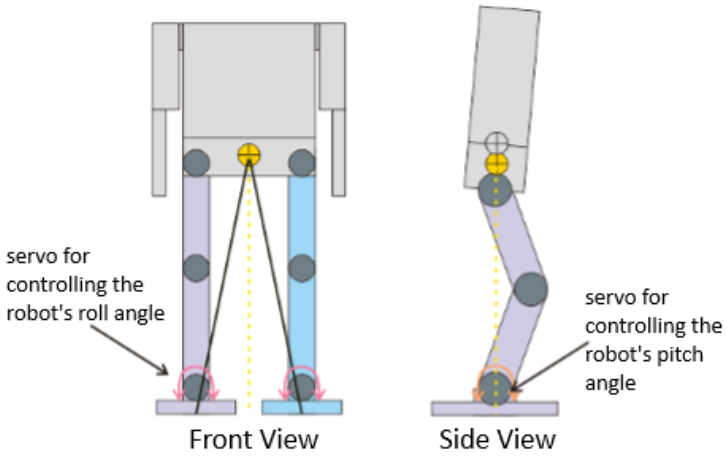


Figure 7. The location of the servos to be controlled

To ensure the humanoid robot's movements align with the reference walking pattern, a combination of an IMU and forward kinematics is used to determine deviations between the robot's current position and the reference pattern. LQR control is employed to determine the optimal gain values for system optimization.

The system uses the BNO055 IMU, which integrates an accelerometer, gyroscope, and magnetometer with an internal sensor fusion engine (Bosch BSX3.0) to provide absolute orientation in Euler angles. This module ensures high stability for short-term control applications, with heading errors typically below ±3° and pitch/roll below ±1° after calibration. Since the scope of this study focuses on short-duration stabilization, specifically the transition from perturbed to steady-state posture, the impact of long-term IMU drift is negligible. Therefore, additional sensor fusion, such as vision-based tracking, is unnecessary for the intended control objective.

https://doi.org/10.31436/iiumej.v27i1.3587

As depicted in Figure 7, the servos controlled are located at the robot's heels and facilitate movement in roll and pitch. The control system inputs are the deviations in roll and pitch angles, as well as their angular velocities at the ankles, compared to the reference values. The system outputs torque values that represent the energy required to adjust the robot toward the reference walking pattern. This is achieved by comparing the measured robot position with the reference pattern and calculating the necessary response speed to correct the position. The roll and pitch servos at the ankles generate the appropriate torque to move the robot and shift its CoM projection, ensuring alignment with the reference walking pattern.

3.2. Walking Pattern Design

The walking pattern design is required to determine the trajectory of the projected center of mass (CoM) during walking. Adjusting this point will affect the robot's joint motion and control. The Zero Moment Point (ZMP) principle is the most widely used method for determining the CoM projection trajectory, and, based on this principle, the Preview Control walking pattern was developed. This concept ensures that the CoM always follows the ZMP pattern, thereby simplifying the concept of walking balance.

It is important to note that the preview control in this study is computed offline during the walking pattern generation stage. The resulting ZMP trajectory serves as a reference for the LQR–fuzzy controller. Because the computation is not performed online at each control iteration, it imposes no real-time computational delay, thereby allowing the control loop to remain responsive and efficient.

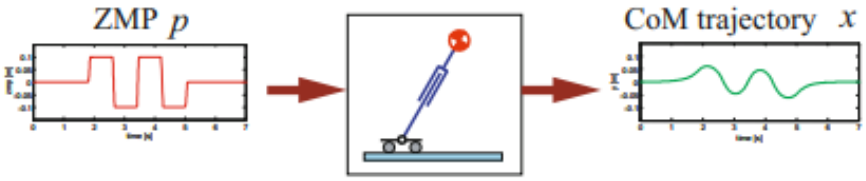


Figure 8. Cart-table model inputs ZMP and outputs CoM motion [22]

Figure 8 illustrates the Preview Control walking pattern, where the ZMP equation can be formulated in state-space form, as shown in Eq. **Error! Reference source not found.**.

$$\frac{d}{dt}\begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u$$
$$p = \begin{bmatrix} 1 & 0 & -\left(\frac{z_c}{g}\right) \end{bmatrix}\begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix} \tag{10}$$

Next, the state space Eq. **Error! Reference source not found.** is converted into a discrete-time system equation using a sample time Δt, resulting in Eq. **Error! Reference source not found.**.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{b}u_k \\ p_k &= \mathbf{c}\mathbf{x}_k \end{aligned} \tag{11}$$

where

$$\begin{aligned} \mathbf{x}_k &\equiv [x(k\Delta t) \quad \dot{x}(k\Delta t) \quad \ddot{x}(k\Delta t)]^T, \\ u_k &\equiv u(k\Delta t), \\ p_k &\equiv p(k\Delta t), \end{aligned}$$

https://doi.org/10.31436/iiumej.v27i1.3587

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{b} \equiv \begin{bmatrix} \frac{\Delta t^3}{6} \\ \frac{\Delta t^2}{2} \\ \Delta t \end{bmatrix}, \boldsymbol{c} \equiv \begin{bmatrix} 1 & 0 & -\left(\frac{z_c}{g}\right) \end{bmatrix}.$$

The Preview Controller calculation produces a CoM pattern representing the target ZMP. The graph showing the CoM pattern representing the ZMP can be seen in Figure 9. Once the CoM pattern is obtained, the footstep pattern will use a quadratic Bezier curve to generate the stepping pattern for the left and right feet during movement. The footstep pattern will be combined in accordance with the CoM pattern. This pattern is then combined to form a walking pattern, as shown in Figure 10.

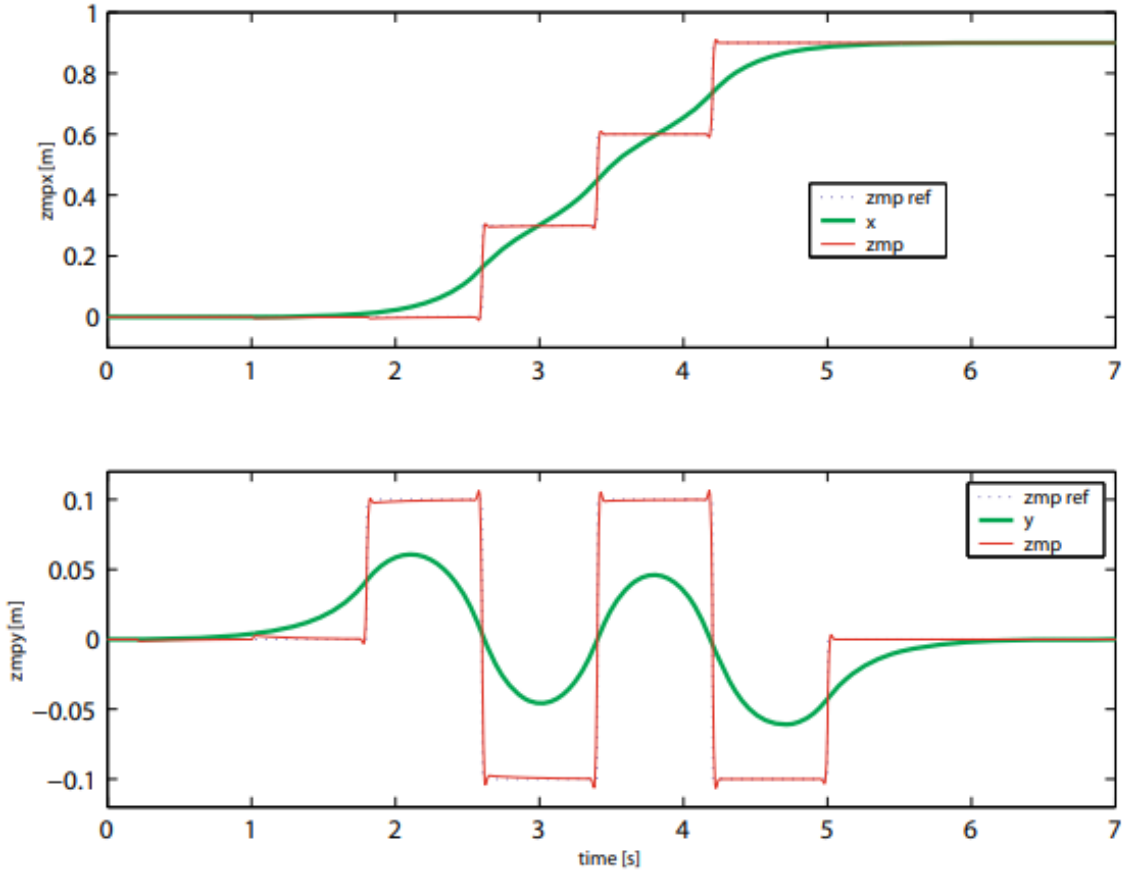


Figure 9. CoM Pattern Generated by the Preview Controller

https://doi.org/10.31436/iiumej.v27i1.3587

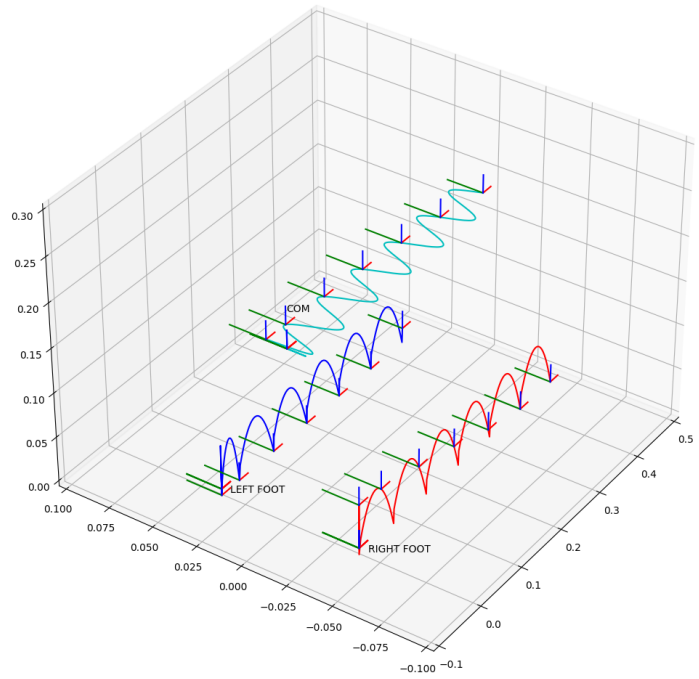


Figure 10. Visualization of the robot's walking pattern using a preview controller

3.3. Control Design

A control system is required for the humanoid robot to walk according to the desired pattern. The control system for the humanoid robot can be modeled as a two-dimensional inverted pendulum to control the ankle's roll and pitch. This inverted pendulum model represents the dynamics of the humanoid robot.

In this study, the control system will use four states as the basic reference for the control parameters. These four states consist of:

- Roll angle (ϕ)
- Pitch angle (θ)
- Roll angular velocity ($\dot{\phi}$)
- Pitch angular velocity ($\dot{\theta}$)

The general dynamics of the inverted pendulum are reduced to a simplified state-space model, as presented in Eq. **Error! Reference source not found.**, which serves as the foundation for the control design.

$$I_{xx}\ddot{\phi} - mgl \sin \phi = \tau_{\phi} \tag{12}$$

Before converting to the state space equation, the ϕ angle and its derivatives are symbolized by y_1, which represents the system outputs, and τ_{ϕ} is symbolized by u_1, representing an action in the form of torque along the roll axis. Additionally, the variable I_{xx} is rearranged to obtain a general model equation, shown by Eq. **Error! Reference source not found.**.

$$\ddot{y}_1 - \frac{mgl}{I_{xx}} y_1 = \frac{1}{I_{xx}} u_1 \tag{13}$$

Eq. **Error! Reference source not found.** is then classified based on the variables y and u along with their derivatives as follows

https://doi.org/10.31436/iiumej.v27i1.3587

$$a_{11} = 0; a_{21} = -\frac{mgl}{I_{xx}}$$
$$b_{01} = 0; b_{11} = 0; b_{21} = \frac{1}{I_{xx}}$$

Auxiliary variables β, x, and $\dot{x}$ are needed to derive the state space equation for the inverted pendulum along the pitch axis, as shown by Equation **Error! Reference source not found.**:

$$\begin{gathered}
\beta_{01} = b_{01} = 0 \\
\beta_{11} = b_{11} - a_{11}\beta_{01} = 0 \\
\beta_{21} = b_{21} - a_{11}\beta_{11} - a_{21}\beta_{01} = \frac{1}{I_{xx}} \\
x_{11} = y_1 - \beta_{01}u_1 = y_1 \\
x_{21} = \dot{y_1} - \beta_{01}\dot{u_1} - \beta_{11}u_1 = \dot{y_1} \\
\dot{x_{11}} = x_{21} + \beta_{11}u_1 = x_{21} \\
\dot{x_{21}} = -a_{21}x_1 - a_{11}x_{21} + \beta_{21}u_1 = \frac{mgl}{I_{xx}}x_1 + \frac{1}{I_{xx}}u_1 \\
\begin{bmatrix} \dot{x_{11}} \\ \dot{x_{21}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{mgl}{I_{xx}} & 0 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{I_{xx}} \end{bmatrix} u_1 \\
y_1 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} + 0\, u_1
\end{gathered} \tag{14}$$

The same process is applied to derive the state space equation for the pitch axis:

$$I_{yy}\ddot{\phi} - mgl\sin\theta = \tau$$
$$I_{yy}\ddot{\phi} - mgl\theta = \tau$$
$$\ddot{y_2} - \frac{mgl}{I_{yy}}y_2 = \frac{1}{I_{yy}}u_2$$

Thus:

$$a_{12} = 0\ ; a_{22} = -\frac{mgl}{I_{yy}}$$
$$b_{02} = 0\ ; b_{12} = 0\ ; b_{22} = \frac{1}{I_{yy}}$$
$$a_{12} = 0\ ; a_{22} = -\frac{mgl}{I_{yy}}$$

With auxiliary variables β, x, and $\dot{x}$:

$$\begin{gathered}
\beta_{02} = b_{02} = 0 \\
\beta_{12} = b_{12} - a_{12}\beta_{02} = 0 \\
\beta_{22} = b_{22} - a_{12}\beta_{12} - a_{22}\beta_{02} = \frac{1}{I_{yy}} \\
x_{12} = y_2 - \beta_{02}u_2 = y_2 \\
x_{22} = \dot{y_2} - \beta_{02}\dot{u_2} - \beta_{12}u_2 = \dot{y_2} \\
\dot{x_{12}} = x_{22} + \beta_{12}u_2 = x_{22} \\
\dot{x_{22}} = -a_{22}x_{12} - a_{12}x_{22} + \beta_{22}u_2 = \frac{mgl}{I_{yy}}x_1 + \frac{1}{I_{yy}}u_2
\end{gathered}$$

https://doi.org/10.31436/iiumej.v27i1.3587

$$\begin{bmatrix} \dot{x}_{12} \\ \dot{x}_{22} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{mgl}{I_{yy}} & 0 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{I_{yy}} \end{bmatrix} u_2$$
$$y_2 = [1 \quad 0] \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} + 0\, u_2 \tag{15}$$

From the inverted pendulum modeling and the four states mentioned above, the state space equations **Error! Reference source not found.** and **Error! Reference source not found.** are then combined to obtain the overall state space equations, shown in equations **Error! Reference source not found.** and (17).

$$\begin{bmatrix} \dot{\theta} \\ \ddot{\theta} \\ \dot{\phi} \\ \ddot{\phi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{mgl}{I_{xx}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{mgl}{I_{yy}} & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \\ \phi \\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{I_{xx}} & 0 \\ 0 & 0 \\ 0 & \frac{1}{I_{yy}} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{16}$$
$$\dot{\mathbf{x}} \qquad \mathbf{A} \qquad \mathbf{x} \qquad \mathbf{B} \quad \mathbf{u}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \\ \phi \\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{17}$$
$$\mathbf{y} \qquad \mathbf{C} \qquad \mathbf{x} \qquad \mathbf{D} \quad \mathbf{u}$$

Explanation:

- θ = Pitch angle
- $\dot{\theta}$ = Pitch angular velocity
- $\ddot{\theta}$= Pitch angular acceleration
- ϕ = Roll angle
- $\dot{\phi}$ = Roll angular velocity
- $\ddot{\phi}$ = Roll angular acceleration
- u_1 = Pitch action
- u_2 = Roll action
- I_{xx} = Moment of inertia about the x-axis
- I_{yy} = Moment of inertia about the y-axis
- g = Gravitational force of the Earth
- y_1 = Pitch angle output
- y_2 = Roll angle output
- m = Total mass of the robot
- l = Distance from the center of mass to the sole of the foot

https://doi.org/10.31436/iiumej.v27i1.3587

The control system employed is an LQR controller in a closed-loop configuration. The primary goal of LQR control is to find the gain **K** matrix for full-state feedback. LQR is a control method that can be applied to a robot model represented in state-space form. The LQR control in the humanoid robot system with state space equations requires the K gain to stabilize the system. The block diagram of the control system is shown in Figure 11. This gain **K** is used to determine the system action value, **u**. The LQR method is a regulator control system, designed to drive the system towards a state with a reference value of zero. Adjusting the reference to a specific physical state can be achieved using a reference state ($\mathbf{x}_{ref}$). $\mathbf{x}_{ref}$ consists of θ_{ref}, $\dot{\theta}_{ref}$, ϕ_{ref}, and $\dot{\phi}_{ref}$. This allows the system to either maintain a zero condition or sustain a particular state based on the new reference ($\mathbf{x}_{ref}$). For a humanoid robot, control is necessary to ensure that it can walk according to a specified reference trajectory. If the system output does not match the reference state, the system will repeat the process until the desired state is achieved.

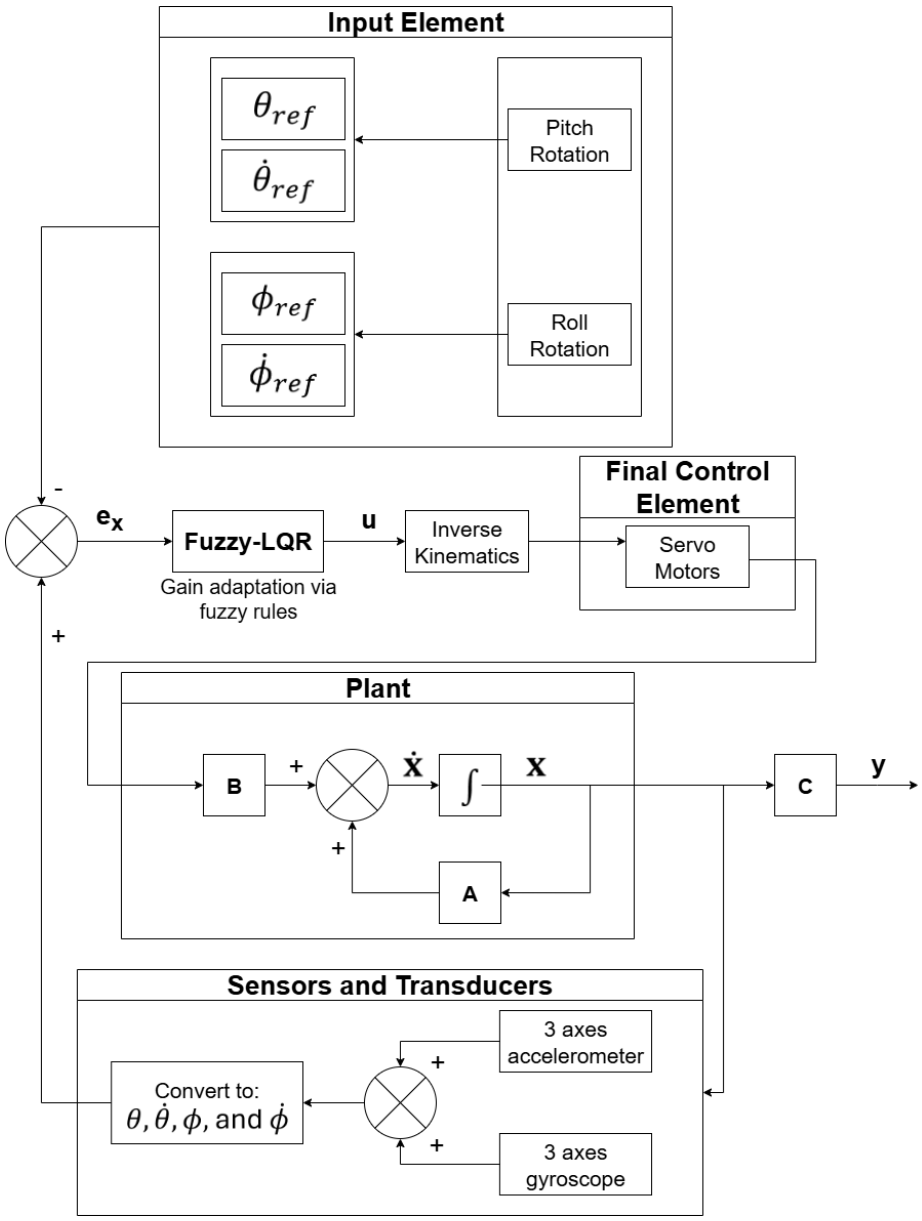


Figure 11. Block diagram of the proposed Fuzzy-LQR control system with gain adaptation based on fuzzy rules for humanoid stability on uneven terrain.

https://doi.org/10.31436/iiumej.v27i1.3587

Figure 11 shows the block diagram of control using gain **K**. The angle sensor within the smart servo measures the angular motion of each servo. These angles are used in forward kinematics to determine the position of the robot's center of mass along the robot's x, y, and z axes. The use of sensor fusion (accelerometer and gyroscope) can provide additional reference points for determining the robot's center-of-mass motion. The predefined reference state will serve as a basis for comparison with the robot's new state. The difference between the reference state and the new robot state is multiplied by the gain **K** value, which is then used as a new action in the system. The system's output is the position and velocity of the center of mass, which are converted into servo angles via inverse kinematics, thereby driving the robot's motion.

Forward kinematics helps recognize terrain differences during walking by mapping the waist position relative to the foot surface and comparing it to x and y axis errors detected by the IMU sensors. When the waist position matches the walking pattern, but x and y errors exceed the reference movement pattern, the gain **K** will be adjusted to account for terrain differences.

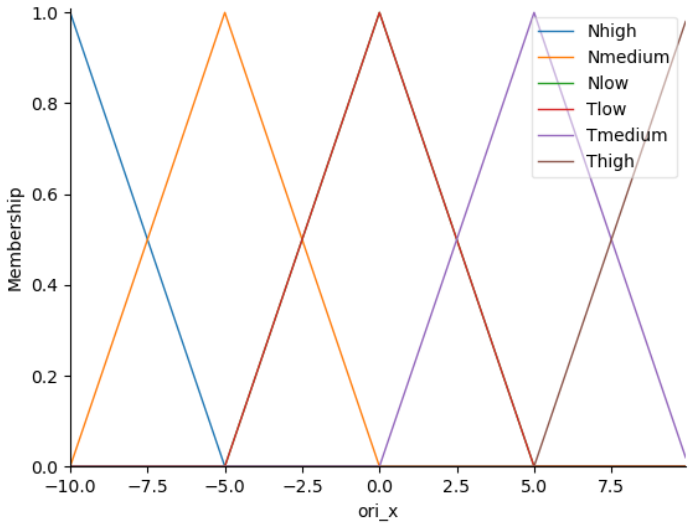


Figure 12. Fuzzy error set on an uneven and sloped surface (relative to the front body of the robot).

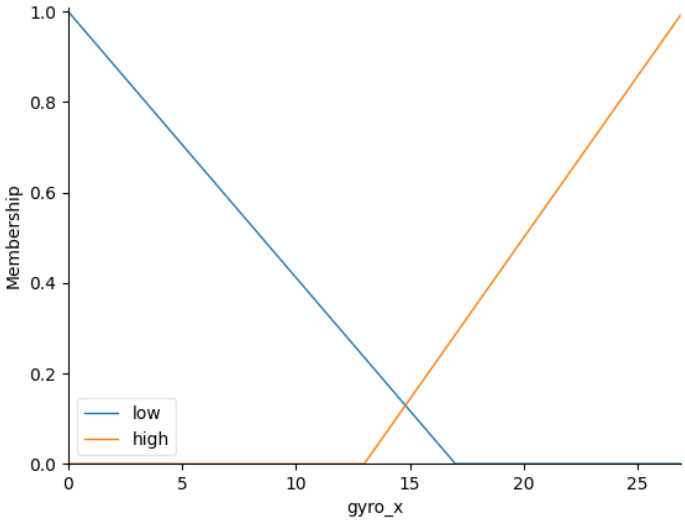


Figure 13. Fuzzy delta error set on an uneven and sloped surface (relative to the front body of the robot).

https://doi.org/10.31436/iiumej.v27i1.3587

The gain **K** varies along the y-axis for sloped surfaces, and the fuzzy Mamdani method is used to determine the appropriate gain **K** for stable robot walking. The method uses error and delta error as inputs, where error is the difference between the angle setpoint and the executed angle, and delta error is the rate of change of the error. To describe the error set on a sloped x-axis, six linguistic variables are used: NK (Small Increase), NS (Medium Increase), NB (Large Increase), TK (Small Decrease), TS (Medium Decrease), and TB (Large Decrease), as shown in Figure 12. Two linguistic variables describe the delta error: “low” (indicating a small change in error) and “high” (indicating a large change in error), as shown in Figure 13. These correspond to the symbolic fuzzy labels K (small) and B (big) used in the fuzzy rule base, even though they are not explicitly labeled in the figure.

The fuzzy error range is based on possible angle variations caused by different terrain slopes, which necessitates the fuzzy control approach for robot stabilization on the y-axis slopes. A fuzzy inference system (FIS) is then designed using the Mamdani method, which employs fuzzy rules in the form of "If X = A, Then Y = B". The fuzzy rules based on the input are shown in Table 1.

Table 1. Fuzzy Rules for the y-axis slope

		Error					
Delta Error		NB	NS	NK	TK	TS	TB
	DK	QNS	QNS	QNTK	QNTK	QTS	QTS
	DB	QNB	QNB	QNS	QTS	QTB	QTB

After forming the fuzzy rules, a defuzzification process is carried out to obtain an output **Q**, which is used to calculate the appropriate gain **K** for maintaining balance on sloped surfaces. The linguistic variables representing the output **Q** are QNTS (Small Increase-Decrease), QNS (Medium), and QB (Large). The range of **Q** values are determined by finding the correct input **Q** to stabilize the robot on various sloped terrains, both uphill (0° to 10°) and downhill (0° to -10°). The fuzzy set design for the output **Q** is shown in Figure 14.

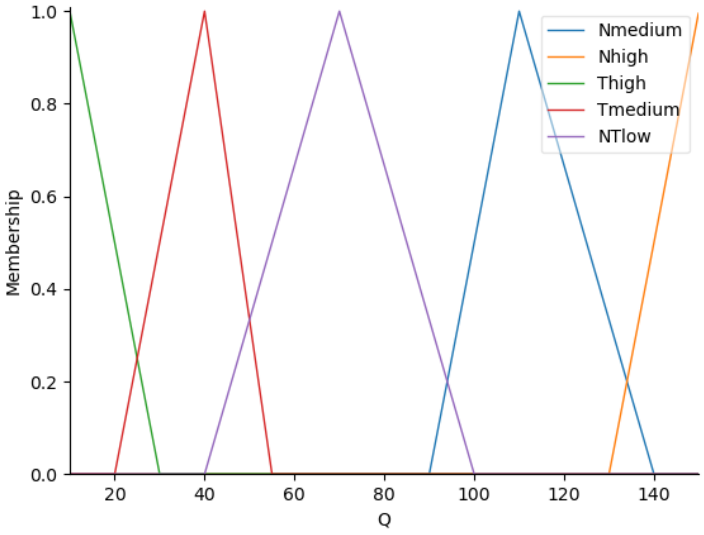


Figure 14. Fuzzy Set for **Q** on the y-axis slope

Therefore, Figure 12. Fuzzy error set on an uneven and sloped surface (relative to the front body of the robot).depicts the fuzzy membership for the input error, while Figure 14 illustrates

the output membership function for the **Q** gain adjustment, ensuring they represent different stages of the fuzzy control process.

Although advanced control techniques such as Model Predictive Control (MPC) or adaptive control may offer benefits in handling nonlinear dynamics, we consider them unnecessarily complex for the scope of this study. The proposed stability control problem is constrained to walking and sudden stopping on moderately uneven terrain, with small pitch and roll deviations (within ±0.5°). In such cases, the inverted pendulum model provides a sufficiently accurate linear approximation, making LQR a practical and computationally efficient solution.

Additionally, because the embedded system on the humanoid robot operates under limited computational resources, implementing MPC, which requires solving an optimization problem at each time step, would be unnecessary and impractical. Instead, integrating fuzzy logic into the LQR framework enables adaptive gain tuning to handle terrain variation effectively, without the complexity of full nonlinear or predictive control. This hybrid LQR–fuzzy approach thus balances controller performance, robustness, and real-time feasibility for the intended application.

Furthermore, the construction of the fuzzy rule base was performed using a hybrid approach that combined experimental observation with automatic gain computation. For each combination of body inclination error and delta error, the required LQR gain was systematically calculated to keep the COM within the predefined tolerance limits of the support polygon. These gain values were then grouped into fuzzy output sets and mapped to fuzzy rules using the Mamdani inference method. This ensures that the fuzzy-enhanced controller can dynamically tune the LQR gain in real time in response to terrain-induced disturbances, while maintaining stability within the safe operating region.

The system's controllability was verified during controller design and confirmed through simulation, ensuring that the state-space formulation supports LQR-based control. However, formal derivation is omitted here to maintain focus and limit page length. Observability analysis was not conducted because the control system employs direct full-state feedback without state estimation. For the current application, which only requires a subset of measurable states, the implemented LQR–fuzzy approach has proven sufficient. Nevertheless, in future scenarios with more complex dynamics or incomplete state measurements, the controller can be extended to a Linear Quadratic Gaussian (LQG) framework by incorporating observer-based estimation, such as the Linear Quadratic Estimator (LQE).

4. RESULTS AND DISCUSSION

4.1. Testing of Walking Robot Stabilization

All results presented in this section are obtained from real-time experiments conducted on a physical humanoid robot. The robot is a modified Bioloid Type-C equipped with 12 servo actuators and an onboard IMU. No simulation data were included in this study to maintain conciseness and practical relevance. The experimental results shown in Figures 15–19 reflect the robot's behavior under real-world disturbances, including sensor noise, actuator delays, and mechanical limitations.

Walking Stability testing aims to determine the robot's response when walking. The test was conducted with the robot's legs initially straight, parallel to the position in which both knees were bent. We then set the walking order to a stride length of 0.025 meters and a frequency of 30 Hz, using a preview-control walking pattern. This test uses the heel strategy

https://doi.org/10.31436/iiumej.v27i1.3587

as a control system. The heel strategy uses full-state feedback control that uses a gain **K** from the LQR method to control the robot's heel servo so that the robot's movement can match the walking pattern. The robot's center-of-mass response to motion was then observed and compared with the reference center of mass derived from the walking pattern.

The balance control response graph of each gain when the robot is walking is shown in Figures 15 and 16. The strengthening of **K** affects the system's feedback value, expressed as the robot's heel angle, thereby shifting the center of mass. The value of **K** is obtained by tuning the matrix **Q** in the control simulation. The most optimal **K** value is determined based on system observations by comparing the responses of each **Q** matrix. The variation of the **Q** value is based on the form of the response. This response is generated by the system and tested on the robot. The greater the **Q** value, the greater the value of **K** gain. The resulting **K** value is entered into the full-state feedback equation and will affect the torque produced. The smaller the difference between the robot's center of mass and the reference center of mass, the more optimal the system. Tuning the **Q** component in the simulation using a Python-based application produces a **K** gain, as shown in Table 2.

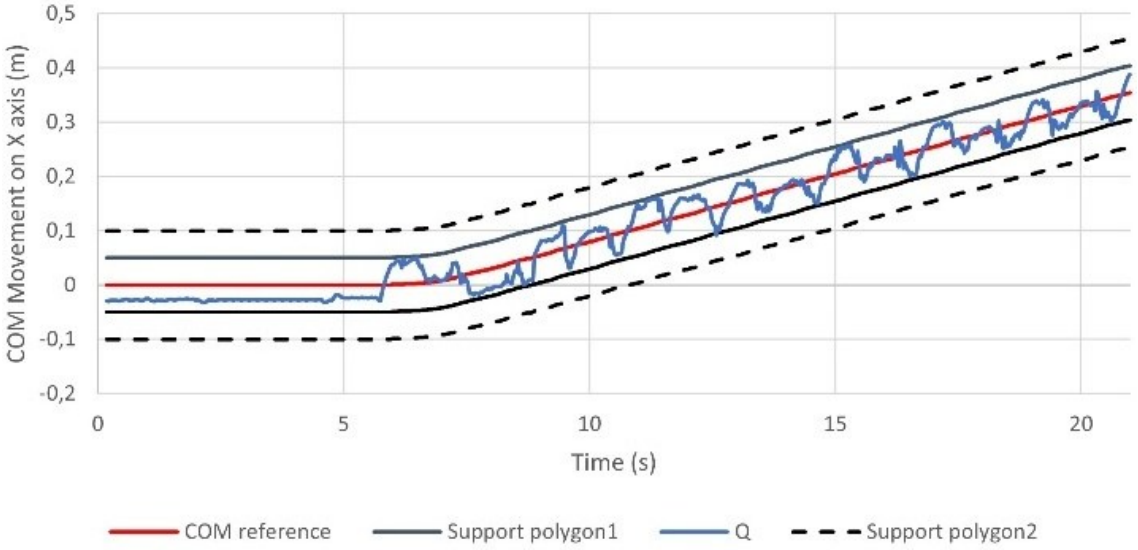


Figure 15. Balance control response on the **x** axis when the robot is walking

Table 2. The **Q** value of the LQR method to the full-state feedback **K** value.

Q	K
$\mathbf{Q}_1 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 750 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 6.06 & 3.24 & 0 & 0 \\ 0 & 0 & 29.68 & 3.51 \end{bmatrix}$
$\mathbf{Q}_2 = \begin{bmatrix} 50 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 750 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 9{,}61 & 1{,}34 & 0 & 0 \\ 0 & 0 & 29{,}68 & 1{,}83 \end{bmatrix}$
$\mathbf{Q}_3 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 50 & 0 & 0 \\ 0 & 0 & 750 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 6{,}06 & 7{,}11 & 0 & 0 \\ 0 & 0 & 29{,}49 & 3{,}51 \end{bmatrix}$

https://doi.org/10.31436/iiumej.v27i1.3587

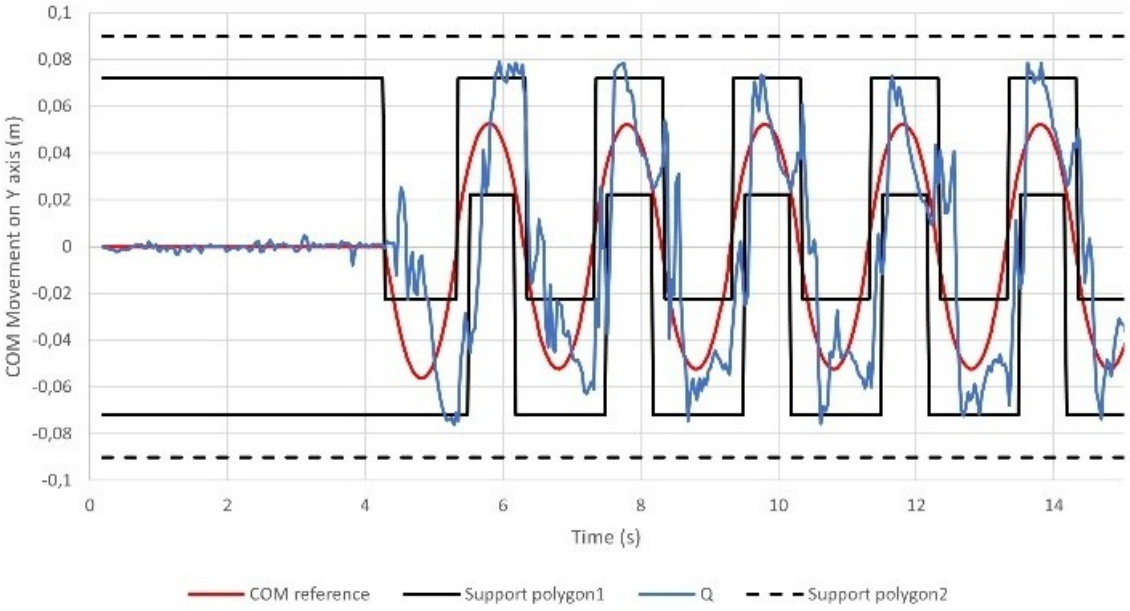


Figure 16. Balance control response on the **y** axis when the robot is walking

Based on Table 3, it is shown that the optimal value of **Q** lies in the value of Q_2 because it has the smallest maximum error among other values. This value is also considered the most optimal because the robot's center of mass is always in the support polygon on the x-axis. There is no error, and although there is an error in the y-axis, the robot can immediately overcome it, so it doesn't fall and returns to the support polygon area.

In the x and y axes control response graphs, we can see that the value of the matrix **Q** affects the robot's balance while walking. The Control system response to the x and y axes when the robot is walking are shown in Table 3.

Table 3. Control response on the x and y axes when the robot is walking.

Q	x**-axis**	y**-axis**
	Error Maximum	**Error Maximum**
$\boldsymbol{Q_1}$	1.23×10^{-2}	9.43×10^{-2}
$\boldsymbol{Q_2}$	≈ 0	8.48×10^{-2}
$\boldsymbol{Q_3}$	1.99×10^{-2}	7.89×10^{-2}

4.2. Testing of Sudden Stop Stabilization

Stability testing during a sudden stop aims to observe the robot's response when it receives an order to stop abruptly, thereby inducing a thrust from within. In testing, the robot stop indicator is implemented by adding an ultrasonic sensor to detect an obstacle in front of the robot. When the ultrasonic sensor detects an object in front of the robot at a distance of up to 10 cm, the robot stops with unspecified leg condition.

https://doi.org/10.31436/iiumej.v27i1.3587

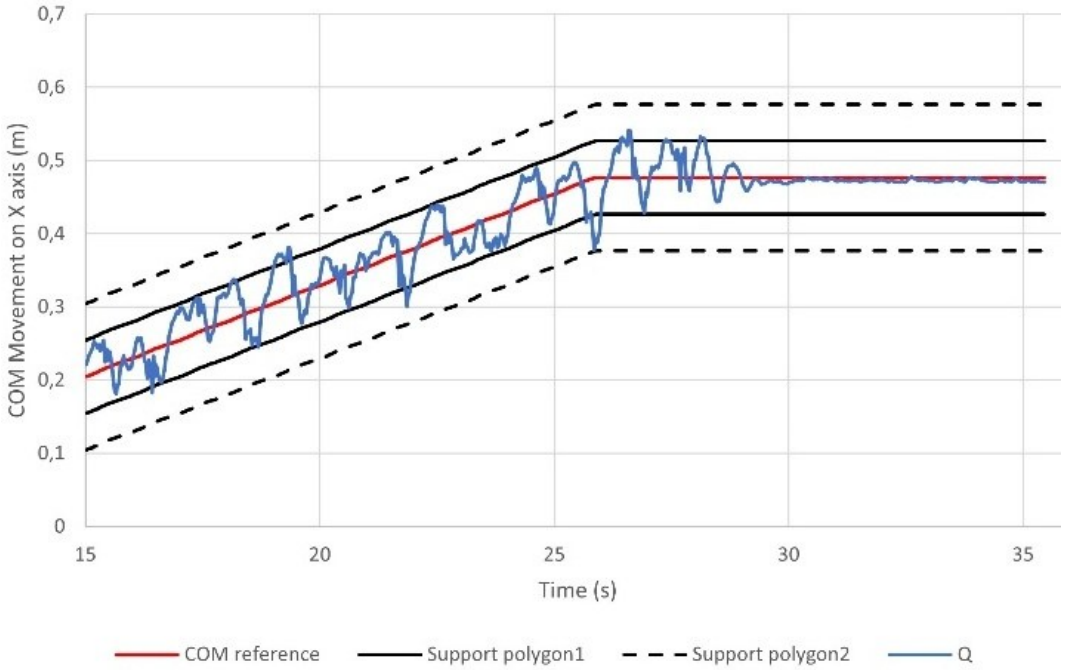


Figure 17. Balance control response on the x-axis when a sudden stop occurs.

This test uses the heel strategy as a control system. The heel strategy uses full-state feedback control that uses **K** gain from the LQR method to control the robot's heel servo so that the robot can still maintain its balance when it stops suddenly. From the previous test, the most optimal **Q** results were also used for this test. In testing, the robot overcame its instability when it was suddenly stopped at any leg state. Figure 17 and Figure 18 show a graph of the balance control response when a sudden stop occurs.

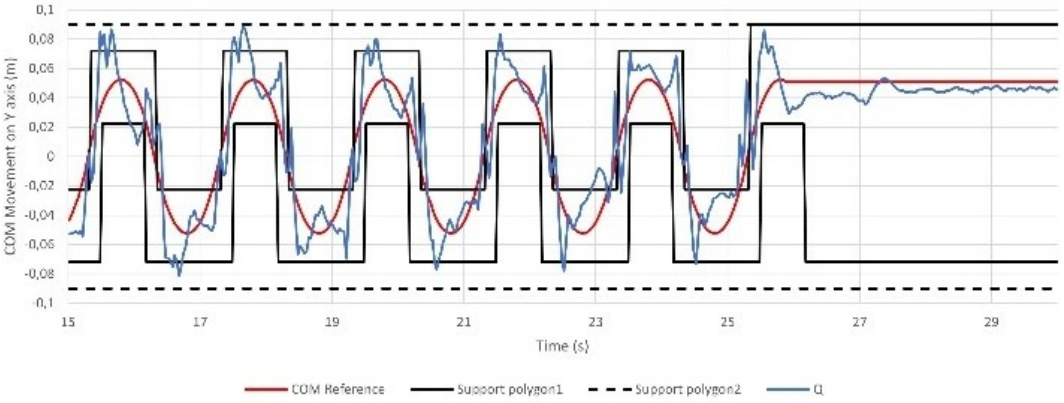


Figure 18. Balance control response on the y-axis when a sudden stop occurs.

Figure 17 and Figure 18 are graphs of the robot's balance control response resulting from the same **Q** tuning as in the previous test. Figure 17 and Figure 18 show that when the robot is suddenly stopped, the projection of its center of mass will pass the support limit of polygon 1 (SP1), thereby making the robot unstable. However, given the **K** gain from the heel strategy control system, the pedestal can directly control the robot's balance so that the projection of the center of mass can return to the polygon support, and the robot does not fall.

The heel strategy was selected due to its compatibility with the robot's limited number of actuators and the requirement for real-time computation. While more advanced balance recovery strategies, such as whole-body motion adaptation or reflex-based responses, may offer

https://doi.org/10.31436/iiumej.v27i1.3587

improved stabilization under extreme inertial forces, they demand additional degrees of freedom and higher computational capacity, which exceed the capabilities of the current platform. Future work will further explore these advanced strategies to improve robustness in high-disturbance scenarios.

4.3. Testing on Uneven Surfaces

The robot's walking performance on different terrains is illustrated in Figures 15 and 16, where the robot navigates flat terrain while maintaining its Center of Mass (CoM) within the boundaries of the walking pattern defined by the preview control system. This indicates that robots achieve balanced movement without falling. In contrast, when walking on inclined surfaces, the CoM's behavior changes significantly, as shown in Figure 19. On inclined surfaces, the CoM shifts toward the robot's support polygon, indicating increased demands on balance control. Specifically, when walking uphill, the CoM moves ahead of the predetermined walking pattern, approaching the front boundary of the support polygon as the slope angle increases. Conversely, as the slope steepens, the CoM shifts backward relative to the walking pattern, progressively approaching the rear support polygon.

Figure 19 demonstrates the relationship between the slope angle and the integral absolute error of the balance control system. The graph shows that as the slope angle increases, the integral absolute error also increases in both uphill and downhill scenarios. This indicates that maintaining balance becomes increasingly challenging as slopes become steeper. Importantly, experimental results show that when the robot walks on sloped surfaces along the y-axis, it experiences tilting along the x-axis. This tilting causes the CoM to shift outside the boundaries of the support polygon on the y-axis, leading to a loss of balance and eventual falling. This phenomenon is exacerbated as the slope angle increases, underscoring the limitations of the standard control system in maintaining stability on uneven terrain.

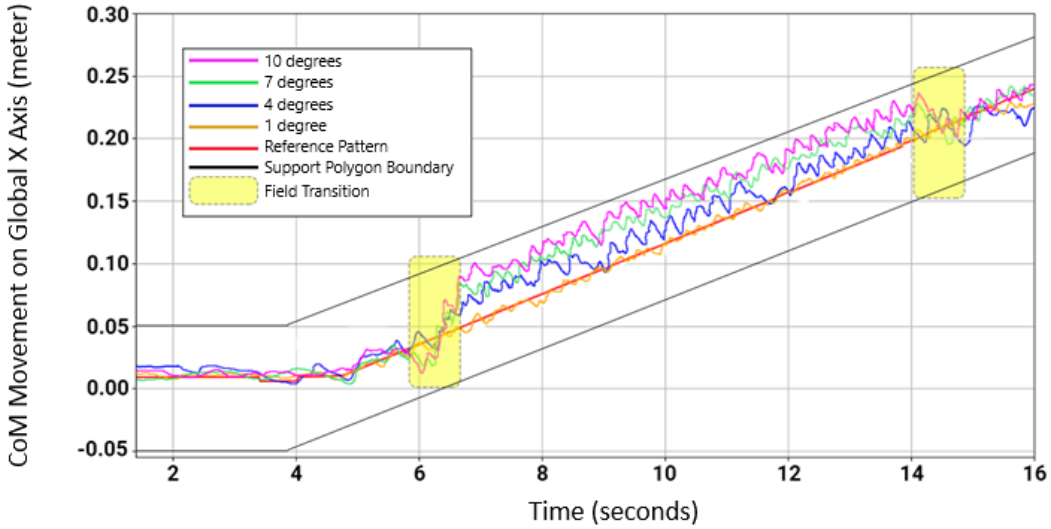


Figure 19. Balance control response when walking from a flat surface to an incline with slope variations of 1, 4, 7, and 10 degrees.

To address these challenges, fuzzy logic control is introduced, which dynamically adjusts the **K**-gain in response to variations in the slope along the y-axis. The integration of fuzzy logic enables the robot to determine optimal control parameters in real time, thereby allowing it to adapt to changing terrain conditions. The effectiveness of this system is evident when the robot transitions between flat and sloped surfaces, where the fuzzy controller selects the appropriate

K-gain to modulate the CoM's movement. This adjustment ensures that the CoM remains within the support polygon, effectively preventing the robot from falling.

Moreover, the adaptability of the fuzzy control system is crucial for enhancing the robot's overall stability on both flat and sloped surfaces. The system's ability to autonomously modify the **K**-gain based on terrain variations allows for smoother and more reliable transitions between different surface inclinations. This demonstrates the superiority of the fuzzy logic control approach over traditional methods, especially in maintaining balance on complex terrains, such as uneven or sloped surfaces.

The combination of preview control and fuzzy logic control provides a robust solution for handling varying terrain inclinations. While preview control ensures balance on flat surfaces by keeping the CoM within the walking pattern, fuzzy logic control enhances stability on sloped surfaces by adjusting the **K**-gain to maintain the CoM within the support polygon. This approach enables the robot to maintain balance effectively across a wide range of slope angles, reducing the risk of falling and improving overall walking performance on both flat and inclined terrain.

While the current experiments focused on walking over planar slopes ranging from 1° to 10°, the proposed controller structure is designed to be modular and adaptable. Terrain inclinations exceeding 10° were excluded from the current study, as the system was not designed or tuned for such extreme conditions, where stability may no longer be guaranteed under the current fuzzy-LQR configuration. The fuzzy-LQR approach was chosen for its simplicity, low computational demand, and effective performance in short-duration stabilization tasks. Although direct benchmark comparisons with other advanced balancing strategies, such as deep reinforcement learning (DRL) or adaptive impedance control (AIC), were not included in this study, such methods are typically more complex and computationally intensive. Given the application scope of this work, rapid recovery and transition to steady-state, the fuzzy-LQR method offers a practical balance between robustness, efficiency, and real-time feasibility. Future research will explore more diverse and unpredictable terrain conditions, including uneven surfaces, ground compliance variability, and dynamic obstacles. Additional sensor inputs, such as force sensors, IMUs, or vision-based terrain estimation, along with comparative evaluations against more adaptive controllers, will be considered to validate further and generalize the proposed approach.

In terms of generalizability, although the proposed control system was developed for a rigid-body humanoid robot with a specific kinematic structure, its reliance on body inclination and rate-of-change feedback enables adaptation to other bipedal robots with different degrees of freedom. Only minor tuning of the fuzzy rule base and gain scheduling would be required, without retraining or redesign of the core architecture. However, for configurations involving soft robotics or highly compliant actuators, structural modifications to the control framework may be necessary. This presents a relevant direction for future research on controller portability across robotic platforms.

This study focuses on pre-fall stability and disturbance rejection, ensuring that the robot can recover from moderate perturbations without losing balance. Fall detection and post-fall recovery strategies, such as those based on whole-body reflex motion or impact mitigation, are not included in the current framework. However, integrating such reactive mechanisms would be a valuable future enhancement, particularly for improving resilience to high-magnitude external disturbances that exceed the controller's stabilizing range

5. CONCLUSION

This study demonstrated the effectiveness of using a Linear Quadratic Regulator (LQR) and fuzzy logic control to enhance the stability of humanoid robots, particularly during walking on uneven surfaces and during sudden stops. By employing an inverted pendulum model, the robot's center of mass (CoM) was controlled to maintain balance. The LQR controller optimized the torque needed to adjust the robot's posture by minimizing the error between the reference and actual states. Additionally, the fuzzy logic control helped adapt to varying slope conditions, dynamically adjusting the gain to ensure stability. The results showed that, while the CoM closely followed the reference trajectory on flat surfaces, walking on inclined surfaces resulted in significant deviations. The use of fuzzy control allowed the robot to navigate these slopes effectively, maintaining balance by continuously adjusting the CoM's position to stay within the support polygon. The combination of LQR and fuzzy logic control is a robust approach to improving the balance and stability of humanoid robots on uneven terrain. These control methods enable the robot to maintain a stable walking pattern, even in challenging environments, making them suitable for applications in human-designed environments.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Department of Computer Science and Electronics at Universitas Gadjah Mada for providing the necessary resources and support to conduct this research. We are also grateful to our colleagues and reviewers for their valuable feedback and insights that significantly improved the quality of this work. Additionally, the authors acknowledge the financial support from the 2024 Academic Excellence grant awarded by the Directorate of Research, Universitas Gadjah Mada, which made this study possible.

REFERENCES

[1] S. N. Izzah, A. Dharmawan, M. Auzan, B. A. A. Sumbodo, and J. E. Istiyanto, "Stabilizing Quadruped Robot Movement Using Fuzzy Logic Control for Yaw Angle Adjustment in Walking and Troting Gait," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 12, no. 4, pp. 975–990, Dec. 2024, doi: 10.52549/ijeei.v12i4.5837.

[2] A. Dharmawan, C. Habiba, and M. Auzan, "Walking stability control system on humanoid when turning based on LQR method," *International Journal of Scientific and Technology Research*, vol. 8, no. 11, pp. 2606–2611, 2019.

[3] C. R. de Lima, S. G. Khan, M. Tufail, S. H. Shah, and M. R. O. A. Maximo, "Humanoid Robot Motion Planning Approaches: a Survey," *J Intell Robot Syst*, vol. 110, no. 2, p. 86, Jun. 2024, doi: 10.1007/s10846-024-02117-z.

[4] M. Murooka *et al.*, "Humanoid Loco-Manipulations Pattern Generation and Stabilization Control," *IEEE Robot Autom Lett*, vol. 6, no. 3, pp. 5597–5604, Jul. 2021, doi: 10.1109/LRA.2021.3077858.

[5] P. M. Viceconte *et al.*, "ADHERENT: Learning Human-like Trajectory Generators for Whole-body Control of Humanoid Robots," *IEEE Robot Autom Lett*, vol. 7, no. 2, pp. 2779–2786, Apr. 2022, doi: 10.1109/LRA.2022.3141658.

[6] L. Rossini, E. M. Hoffman, S. H. Bang, L. Sentis, and N. G. Tsagarakis, "A Real-Time Approach for Humanoid Robot Walking Including Dynamic Obstacles Avoidance," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, IEEE, Dec. 2023, pp. 1–8. doi: 10.1109/Humanoids57100.2023.10375191.

[7] S. Said, K. Youssef, B. Prasad, G. Alasfour, S. Alkork, and T. Beyrouthy, "Design and Implementation of Adam: A Humanoid Robotic Head with Social Interaction Capabilities," *Applied System Innovation*, vol. 7, no. 3, p. 42, May 2024, doi: 10.3390/asi7030042.

[8] İ. H. Şanlıtürk and H. Kocabaş, "Precise Calculation of Inverse Kinematics of the Center of Gravity for Bipedal Walking Robots," *Applied Sciences*, vol. 14, no. 9, p. 3706, Apr. 2024, doi: 10.3390/app14093706.

[9] W. Zuo, J. Gao, J. Liu, T. Wu, and X. Xin, "Whole-Body Dynamics for Humanoid Robot Fall Protection Trajectory Generation with Wall Support," *Biomimetics*, vol. 9, no. 4, p. 245, Apr. 2024, doi: 10.3390/biomimetics9040245.

[10] Vikas, D. R. Parhi, and A. K. Kashyap, "Humanoid robot path planning using memory-based gravity search algorithm and enhanced differential evolution approach in a complex environment," *Expert Syst Appl*, vol. 215, no. December 2022, p. 119423, Apr. 2023, doi: 10.1016/j.eswa.2022.119423.

[11] A. Dharmawan, A. Ashari, and A. E. Putra, "Translation Movement Stability Control of Quad Tiltrotor Using LQR and LQG," *International Journal of Intelligent Systems and Applications*, vol. 10, no. 3, pp. 10–21, Mar. 2018, doi: 10.5815/ijisa.2018.03.02.

[12] S. Singh, R. P. Russell, and P. M. Wensing, "On Second-Order Derivatives of Rigid-Body Dynamics: Theory and Implementation," *IEEE Transactions on Robotics*, vol. 40, pp. 2233–2253, 2024, doi: 10.1109/TRO.2024.3370002.

[13] A. Dharmawan, A. Ashari, A. G. Aprilia, and A. M. Handayani, "Auto VTOL System on Quadrotor Using Madgwick Quaternion Kalman Filter and LQR," in *2018 4th International Conference on Science and Technology (ICST)*, IEEE, Aug. 2018, pp. 1–6. doi: 10.1109/ICSTC.2018.8528613.

[14] G. Menga, "A Novel Paradigm for Controlling Navigation and Walking in Biped Robotics," *Electronics (Basel)*, vol. 13, no. 11, p. 2224, Jun. 2024, doi: 10.3390/electronics13112224.

[15] J. Kim, B. Park, H. Lee, and J. Park, "Hybrid Position/Torque Ankle Controller for Minimizing ZMP error of Humanoid Robot," in *2021 18th International Conference on Ubiquitous Robots (UR)*, IEEE, Jul. 2021, pp. 211–216. doi: 10.1109/UR52253.2021.9494640.

[16] R. A. Kadhim, M. Q. Kadhim, H. Al-Khazraji, and A. J. Humaidi, "Bee Algorithm Based Control Design for Two-links Robot Arm Systems," *IIUM Engineering Journal*, vol. 25, no. 2, pp. 367–380, Jul. 2024, doi: 10.31436/iiumej.v25i2.3188.

[17] R. Bimarta, A. E. Putra, and A. Dharmawan, "Balancing Robot using Proportional Integral Derivative Control Method," *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, vol. 5, no. 1, pp. 89–96, May 2015, doi: 10.22146/ijeis.7157. [in Indonesian]

[18] A. Dharmawan, J. E. Istiyanto, A. E. Putra, and M. Auzan, "LQR Compensated by Fuzzy for Kicking Balance Control of a Humanoid Robot," *ICIC Express Letters, Part B: Applications*, vol. 14, no. 5, pp. 449–456, May 2023, doi: 10.24507/icicelb.14.05.449.

[19] M. M. Kawakibi, A. Dharmawan, M. A. Jazi Eko Istiyanto, and D. Lelono, "Stability Control of Humanoid Robot Walking Speed Using Linear Quadratic Regulator Method," *ICIC Express Letters, Part B: Applications*, 2024.

[20] H. Iman and Md. R. Khan, "Repurposing A Sampling-Based Planner for A Six-Degree-Of-Freedom Manipulator to Avoid Unpredictable Obstacles," *IIUM Engineering Journal*, vol. 24, no. 1, pp. 319–332, Jan. 2023, doi: 10.31436/iiumej.v24i1.2642.

[21] M. Auzan, D. Lelono, and A. Dharmawan, "Humanoid Walking Control Using LQR and ANFIS," *Journal of Robotics and Control (JRC)*, vol. 4, no. 4, pp. 548–556, Aug. 2023, doi: 10.18196/jrc.v4i4.16444.

[22] S. Kajita, H. Hirukawa, K. Harada, and K. Yokoi, *Introduction to Humanoid Robotics*, vol. 101. in Springer Tracts in Advanced Robotics, vol. 101. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-54536-8.

https://doi.org/10.31436/iiumej.v27i1.3694

Vehicle Identification and Classification Using YOLO Algorithm

CHAU LY THI HUYEN, CHI PHAM QUYET, QUYEN THI NGUYEN*

Faculty of Information Technology, Van Lang School of Technology, Van Lang University, Ho Chi Minh City, Vietnam

**Corresponding author: quyen.nt@vlu.edu.vn*

(Received: 12 April 2025; Accepted: 24 December 2025; Published online: 12 January 2026)

ABSTRACT: Vehicle identification and classification are among the challenging activities for the management and control of a large number of different vehicles moving in the inner city. Among many identification and classification systems, the YOLO algorithm stands out for its ability to analyze at high speed and with high accuracy. The algorithm is continually evolving, with notable versions including YOLOv8. This research presents a method for identifying and classifying vehicles using the YOLOv8 algorithm. The assessment of the proposed method's effectiveness was conducted using two COCO datasets (328,000 images) and a real-world dataset from Ho Chi Minh City (HCMC) with more than 1,000 images. The findings indicate that the proposed method can be applied to identify and classify vehicles with an accuracy of 93%-98%. Comparative results with prior studies also demonstrate the superiority of the YOLOv8 algorithm.

ABSTRAK: Pengecaman dan pengelasan kenderaan adalah salah satu aktiviti mencabar dalam pengurusan dan kawalan sejumlah besar kenderaan berbeza yang bergerak di bandar. Di antara kebanyakan sistem pengecaman dan pengelasan, algoritma YOLO menonjol kerana keupayaannya menganalisa pada kelajuan berketepatan tinggi. Algoritma ini terus dibangunkan dengan penambahbaikan yang banyak dan versi yang ketara ialah YOLOv8. Penyelidikan ini membentangkan kaedah mengenal pasti dan mengelaskan kenderaan menggunakan algoritma YOLOv8. Penemuan penilaian keberkesanan kaedah yang dicadangkan telah dijalankan menggunakan dua set data COCO dengan 328,000 imej dan set data sebenar di Ho Chi Minh City (HCMC) lebih daripada 1,000 imej. Dapatan kajian menunjukkan kaedah ini diaplikasikan dalam mengenal pasti dan pengelasan kenderaan berketepatan 93% hingga 98%. Hasil perbandingan dengan kajian lepas juga menunjukkan keunggulan algoritma YOLOv8.

KEYWORDS: *YOLOv8, Vehicle identification, Vehicle classification, COCO datasets*

1. INTRODUCTION

The number of vehicles in large cities has become a challenge for the management and monitoring of vehicular traffic. Vehicle detection and classification are among the key activities of intelligent traffic observation systems [1]. In addition, as the economy grows, the number of automobiles and the incidence of traffic accidents are increasing, making it crucial to implement measures to reduce these accidents and protect life and property. Vehicle identification and classification are essential to traffic analysis, particularly in developed countries [2]. By the end of 2023, Ho Chi Minh City (HCMC), Vietnam, will have approximately 10 million vehicles. This indicates that the identification and management of vehicles on the streets are urgent requirements to ensure traffic safety. Meanwhile, public transportation studies are the primary resource for analyzing traffic accidents, congestion, and traffic flow, enabling a better understanding of traffic dynamics to reduce fatal accidents that

https://doi.org/10.31436/iiumej.v27i1.3694

mainly occur due to driver carelessness or due to poor visibility in severe weather conditions, or improper lighting [3,4].

In recent years, the ongoing enhancement of computing hardware has facilitated the swift advancement of convolutional neural networks, leading to significant successes in the field of computer vision [5]. Currently, the most widely used methods for object detection are R-CNN chains [6,7], SSD chains [8,9], and YOLO chains [10,11]. In the research [12], YOLO, as one of the most representative single-stage object detection frameworks, has been extensively applied to traffic target detection tasks. It employs one of the most effective neural network architectures to achieve high accuracy and overall processing speed. The images are fed into the YOLO algorithm to detect and classify vehicles in each image, which is a major reason for its popularity. Redmon introduced the first iteration of YOLO in their research [10]. In particular, the YOLOv8 version (released in 2018) introduces new and optimized features [13], making this algorithm an excellent option for diverse object detection tasks across a broad spectrum of applications. Researchers have continually proposed various enhanced techniques, including object segmentation [14], silhouette extraction [15], and feature extraction and classification [16], to improve the object detection performance of the system [17,18]. Many fields are applying YOLO for object detection, such as in the creation of the LVD-YOLO model to detect vehicles based on YOLOv5 [19], detecting fruit ripeness in a natural orchard environment with YOLOv7 [20], classification of blood cell image [21], human detection in infrared imagery [22], rice grain detection [23], identification of fault position in power grids [24], safety tracking [25], or applying YOLOv8 to detect wildlife targets for better conservation work [26].

In this study, the YOLOv8 algorithm is employed to detect and classify vehicles in images and videos. The suggested approach is applied to the COCO dataset and to a traffic image dataset in HCMC, Vietnam. Using a large dataset of images and videos, the research produces highly accurate identification and classification results, improving the applicability of the YOLOv8 algorithm for developing a traffic monitoring system and generating traffic statistics, thereby helping to reduce traffic congestion and accidents.

The primary contributions of our research are as follows:

- Successfully applying the YOLOv8 algorithm to identify and classify vehicles.
- The method of identifying and classifying vehicles based on YOLOv8 is successfully applied to two datasets with high accuracy.
- Building a real dataset and successfully applying the proposed method to identify and classify.

The following are the remaining parts of this paper. The problem of identifying and classifying vehicles using the YOLOv8 algorithm, along with the implementation steps, is presented in Sections 2 and 3. The results of applying the proposed method to two large datasets are shown in Section 4, and the conclusion is provided in Section 5.

2. METHOD

2.1. Yolov8 Algorithm

The YOLO architecture operates on the principle of conducting object detection in a single pass through the network, which makes YOLO high-speed and suitable for real-time applications. The input image is segmented into a grid, and bounding boxes along with class probabilities are predicted for each grid cell. The YOLOv8 architecture is built on a

https://doi.org/10.31436/iiumej.v27i1.3694

convolutional neural network and comprises three primary components: the backbone, neck, and head, each responsible for distinct processing tasks, as illustrated in Figure 1.

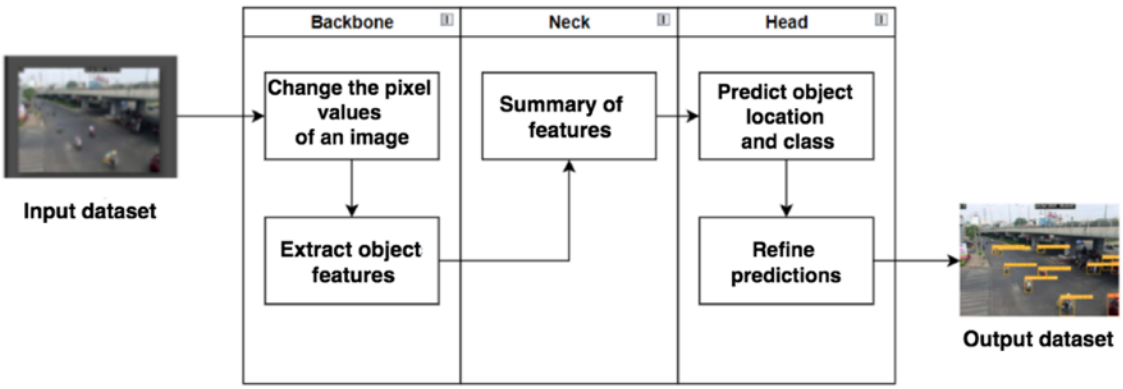


Figure 1. Operation model of the YOLOv8 algorithm in vehicle image recognition

YOLOv8 works as follows:

- Input Processing: YOLOv8 takes an input image and processes it through its backbone network to extract object features. This process includes resizing the image to a fixed dimension of 640x640 pixels and normalizing the pixel values.
- Feature Fusion: The extracted object features are then processed through the Neck, where they are aggregated to capture information at different scales efficiently.
- Detection: The Head uses the processed features to predict bounding boxes, confidence scores, and class probabilities for every detected object. After obtaining the object bounding box and the center point location, YOLO identifies objects throughout the entire image in an end-to-end manner [5].
- Post-processing: Finally, techniques like NMS are used to enhance the predictions by removing overlapping boxes and ensuring that each object is detected only once with a high level of confidence.

2.2. Steps to apply the YOLO algorithm in traffic identification and classification

The research will be conducted on the YOLOv8 variant and shown through the flowchart of the vehicle identification and classification algorithm presented in Figure 2.

The identification and classification algorithm based on the YOLOv8 algorithm is summarized in the following 5 main steps:

- *Step 1*: Data collection and processing

First, it is essential to gather a substantial amount of image and video data of vehicles. The objects identified in this research are cars, trucks, buses, motorbikes, and bicycles. If it is an image, label it as a training and validation dataset. If it is a video, split it into multiple frames and label them similarly to the images. Then label these images with the specific location and vehicle to create an accurate training dataset.

- *Step 2*: Model training

Using YOLOv8, detected vehicles are assigned an identifier for identification and classification. The number of vehicles will be determined for each frame. YOLOv8 is trained on a labeled dataset to enable the model to recognize vehicle characteristics in input images.

https://doi.org/10.31436/iiumej.v27i1.3694

The training process will enhance the model's ability to accurately and quickly recognize and classify vehicles. The model training process is evaluated using the Precision metric to indicate the model's prediction accuracy. The value of this index is determined by Eq. (1).

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

where TP (True Positive) is the actual number of vehicles, and FP (False Positive) is the number of vehicles overcounted or missed.

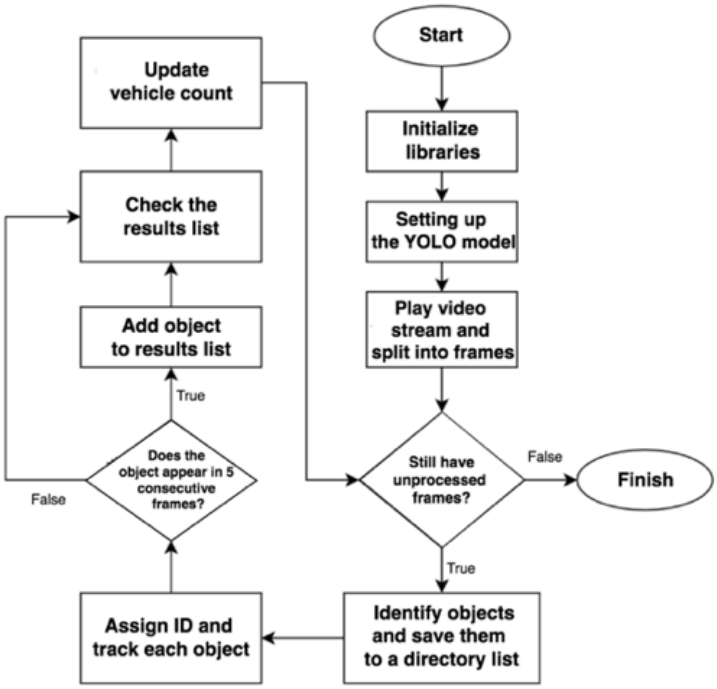


Figure 2. Flowchart of vehicle identification and classification algorithm

- *Step* 3: Vehicle recognition phase

YOLOv8 draws bounding boxes around detected vehicles and uses different colors to distinguish among objects. Display information about the vehicle within the bounding box (e.g., vehicle type and optional attributes, such as direction of movement), then save it to the object list.

- *Step 4*: Feature extraction

YOLOv8 uses the network's feature layers to extract the necessary information from the detected vehicles.

- *Step* 5: Vehicle classification and counting

Finally, vehicles detected by YOLO with the previously assigned identifier will be continuously monitored over five consecutive frames. If the check result is correct, the object will be saved to the item list. Each time a vehicle passes through the detection area, the counter increments the total count of detected vehicles in the item list, and the new count replaces the previous count. The result is displayed to facilitate tracking and understanding of the detected vehicles.

https://doi.org/10.31436/iiumej.v27i1.3694

3. RESULTS AND DISCUSSION

3.1. COCO dataset

COCO is an image and caption dataset comprising over 328,000 images, each with a caption describing the objects in the image. COCO is used to train deep learning models for tasks such as object detection, image segmentation, and image captioning. It is trusted by the majority of users, as shown in Figure 3. The dataset is divided into three groups: Train, Validate, and Test, each serving its intended purpose.

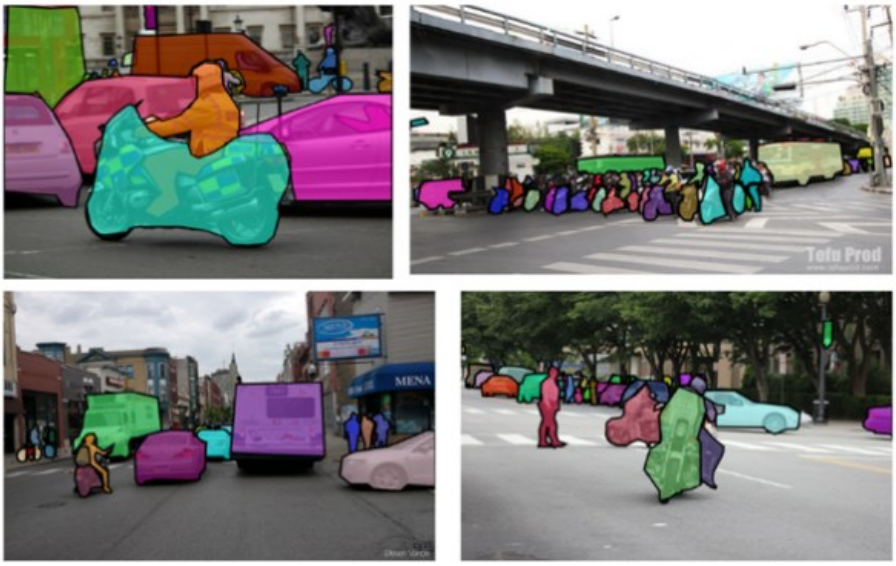


Figure 3. Some labeled images in the COCO dataset

3.2. HCMC dataset

The HCMC dataset includes 05 types of vehicles: cars, motorbikes, buses, trucks, and bicycles, with more than 1000 images and videos collected in the inner city of HCMC to form a dataset for vehicle recognition and classification. These images and videos are presented similarly in Figure 4 and were taken under good lighting conditions.

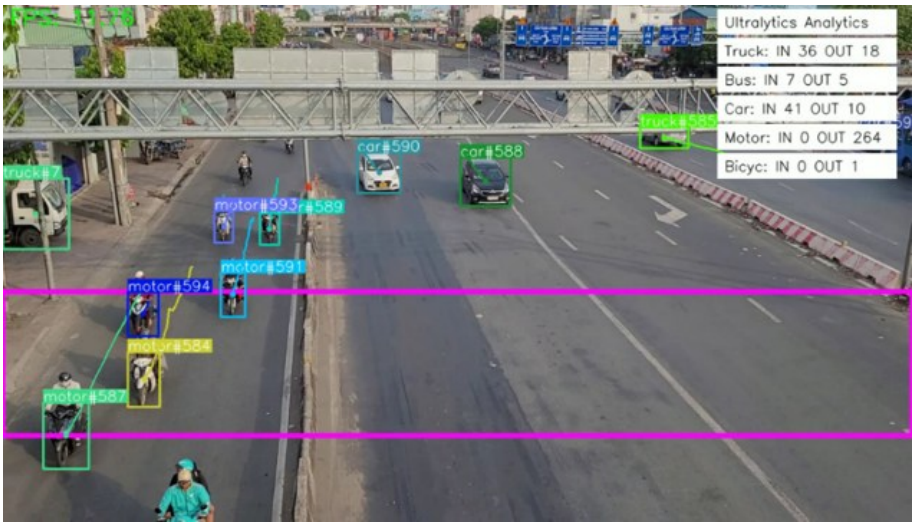


Figure 4. Sample video recording of traffic at Ho Chi Minh City Pedestrian Bridge

https://doi.org/10.31436/iiumej.v27i1.3694

3.3. Implementation Results

3.3.1. COCO dataset

The chart illustrating the model's classification performance on the dataset is presented in Figure 5. The curve indicates the relationship between the model's accuracy and confidence in brand recognition. The results show that on the COCO dataset, the model's parameters achieve perfect accuracy at a confidence level of 1.00. This indicates that the model is highly accurate when it is confident about its predictions at a high level. Recognition results for some images under varying lighting conditions are shown in Figures 6 and 7.

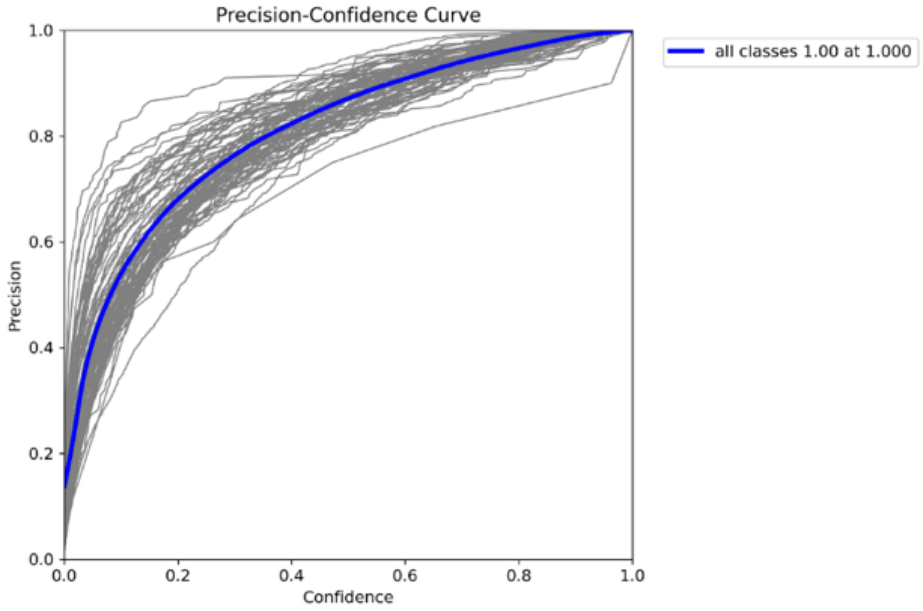


Figure 5. Precision-Recall Curve for the COCO dataset

Figure 6. COCO Some images recognized using the COCO dataset

https://doi.org/10.31436/iiumej.v27i1.3694

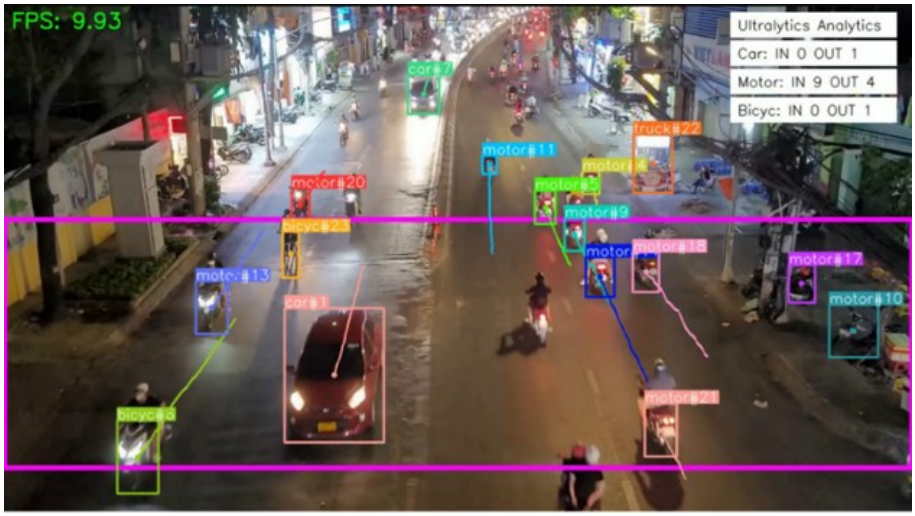


Figure 7. YOLOv8 running on NightStreet.mp4 video

Table 1. Comparison table of Average Precision (AP) values alongside other state-of-the-art techniques

Techniques	AP	AP50	AP75
YOLOv2 [27]	21.6	44.0	19.2
SSD [27]	31.2	50.4	33.3
YOLOv2_Vehicle [27]	39.1	59.1	42.3
R-CNN [27]	33.2	53.3	35.2
Proposed [27]	40.3	60.5	43.5
YOLOv8	98.1	99.8	99.3

```
Running per image evaluation...
Evaluate annotation type *bbox*
DONE (t=126.43s).
Accumulating evaluation results...
DONE (t=28.63s).
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.981
 Average Precision  (AP) @[ IoU=0.50       | area=   all | maxDets=100 ] = 0.998
 Average Precision  (AP) @[ IoU=0.75       | area=   all | maxDets=100 ] = 0.993
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.963
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.997
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.999
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.559
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.965
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.989
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.977
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.998
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.999
```

Figure 8. Overall index obtained when evaluating the dataset

The comparison results between YOLOv8 and other vehicle detection techniques, evaluated on the COCO dataset, are presented in Table 1. These methods have been successfully applied to vehicle detection and trained and evaluated on the COCO dataset. This ensures both relevance to the problem domain and fairness in comparison. The experiment demonstrates that YOLOv8 is the most effective method compared with other vehicle detection techniques in terms of Average Precision (AP) at IOU thresholds of 50 and 75, with substantial differences relative to the proposed method and prior methods. The fact that YOLOv8 is five times larger than its predecessor, YOLOv2, shows how far the YOLO method has advanced,

https://doi.org/10.31436/iiumej.v27i1.3694

and the accuracy is more than twice as high as that of the proposed method [27], showing that the YOLO model has surpassed other advanced detection techniques. The YOLOv8 index obtained from the evaluation process is shown in Figure 8.

The YOLOv8 variant can identify vehicle types more effectively than other leading vehicle recognition methods, as indicated in Table 2. The identification methods listed in the table achieve high accuracy, with the lowest recognition rate being 94.2% for SSD-based car recognition, which is considered very high. The advanced R-CNN method and the YOLOv2 variants both achieve recognition rates above 95% and differ by only 2%. The proposed method in [27] is a recognition model based on the YOLOv2 variant, but has been modified to give significantly improved performance when all indicators are superior to the original version. The YOLOv8 model trained on the COCO dataset outperforms other object detection methods in all aspects, with the accuracy of vehicle types almost asymptotically close to 100%. The YOLOv8 index obtained through the evaluation process is shown in Figure 9.

Table 2. Comparison table of experimental outcomes for vehicle classification

Techniques	Car (AP)	Truck (AP)	Bus (AP)	mAP (%)
R-CNN [27]	95.04	97.17	95.02	95.74
YOLOv2 [27]	95.52	95.41	96.34	95.75
YOLOv2_Vehicle [27]	96.32	95.55	97.73	96.56
SSD [27]	94.2	97.75	97.97	96.64
Proposed [27]	96.54	99.29	97.74	97.85
YOLOv8	99.5	99.5	99.5	99.7

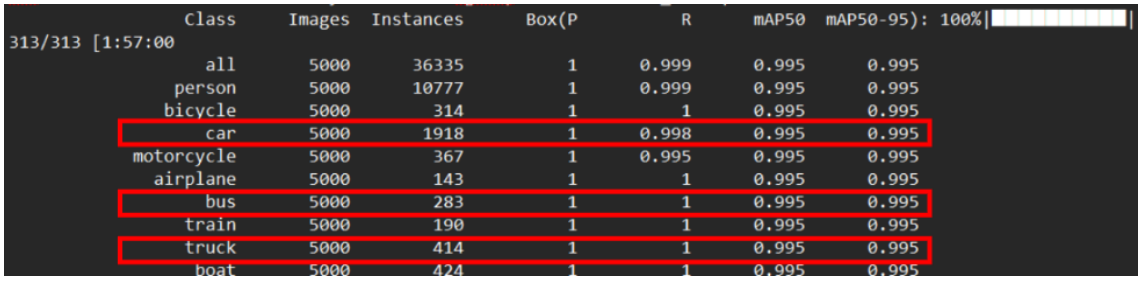


Figure 9. Vehicle Rating Index of the COCO Dataset

3.3.2. HCMC dataset

The following graph was obtained during the training of the recognition model using the HCMC dataset. Figure 10 illustrates the correlation between precision and confidence, with the x-axis representing the confidence threshold and the y-axis indicating accuracy. The figure shows that on the HCMC dataset, the model achieves nearly 1.0 as confidence approaches 0.999, indicating high accuracy.

The results indicate that the model accurately detects and predicts most objects when the input data is of good quality, and the directions do not obscure one another, as shown in Table 3. Examples of recognized images in the traffic dataset for specific routes are shown in Figure 11.

https://doi.org/10.31436/iiumej.v27i1.3694

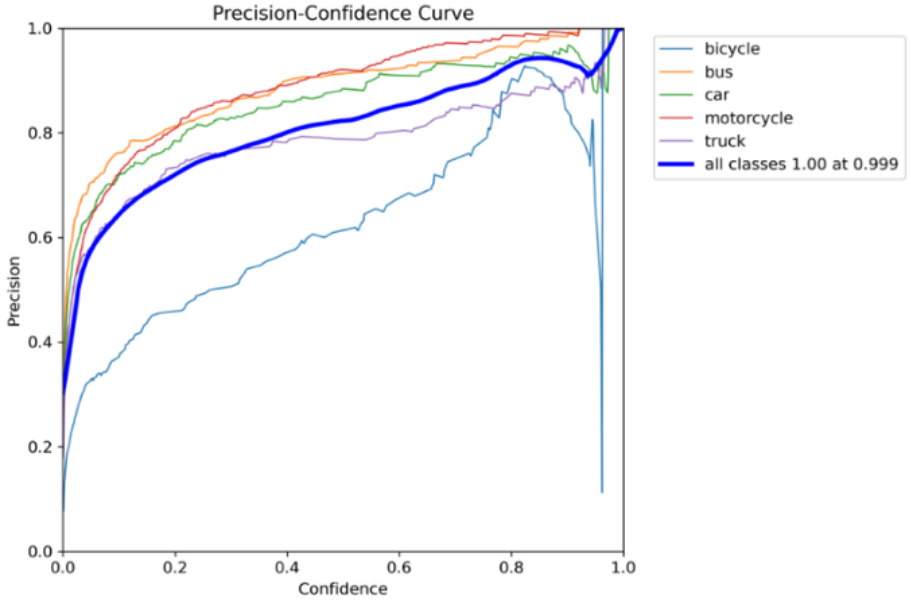


Figure 10. Precision-Confidence Curve for the HCMC dataset

Table 3. Test results on images with low vehicle count

Vehicle	Highway13.jpg	HighwayNam.jpg	Dist1Bicyle.jpg	PublicBicycle.jpg	Public Bicycle Station.jpg	Exact ratio (%)
	Counted/ Reality	Counted/ Reality	Counted/ Reality	Counted/ Reality	Counted/ Reality	
Car	4/4	9/10	0/0	4/4	3/3	95.23
Motorbike	5/6	0/0	5/5	4/4	0/0	93.33
Truck	1/1	9/9	0/0	0/0	1/1	100
Bus	1/1	7/7	2/2	0/0	1/1	100
Bicycle	0/0	0/0	3/3	5/6	8/8	94.12

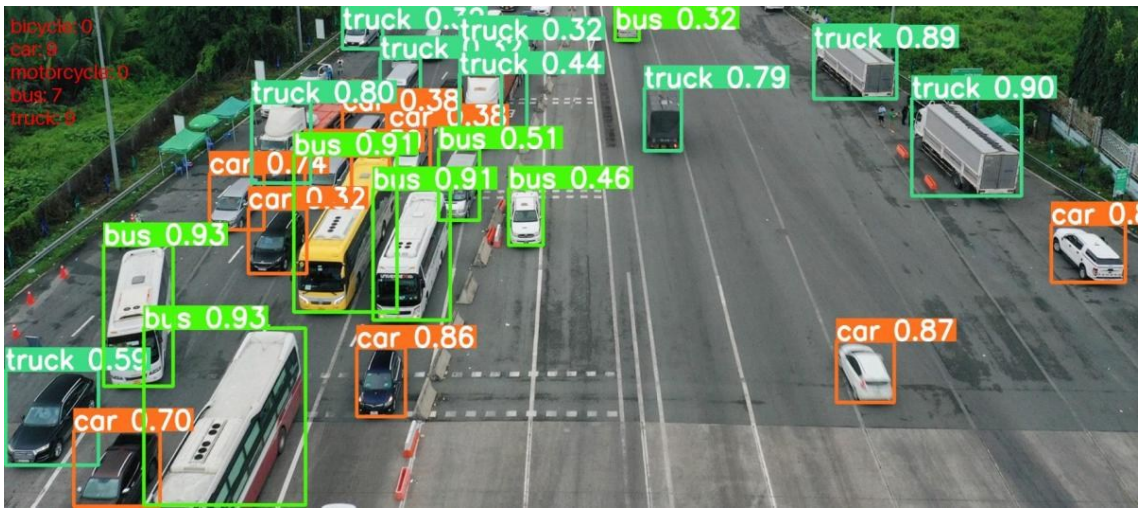


Figure 11. Labeled images on some routes

The counts of each vehicle type relative to the total number of vehicles in the test videos are shown in Table 4. The table shows that YOLOv8 performs well with video as an input data type when accurately detecting most of the objects appearing in the video with an accuracy of at least 98.3%, and can reach 100% under ideal conditions. The results of video recognition and classification are presented in Figure 12.

https://doi.org/10.31436/iiumej.v27i1.3694

Table 4. YOLOv8 test results on video data

Vehicle	BT.mp4	BicycleRacing.mp4	Parade.mp4	Footbridge1.mp4	Footbridge2.mp4	Exact ratio (%)
	Counted/ Reality	Counted/ Reality	Counted/ Reality	Counted/ Reality	Counted/ Reality	
Car	35/37	12/12	4/4	47/47	51/51	98.67
Motorbike	6/6	77/77	13/13	263/270	264/264	98.89
Truck	3/1	36/36	12/12	28/28	54/54	98.47
Bus	0/0	7/7	0/0	7/7	12/12	100
Bicycle	0/0	54/54	0/0	2/2	1/2	98.27

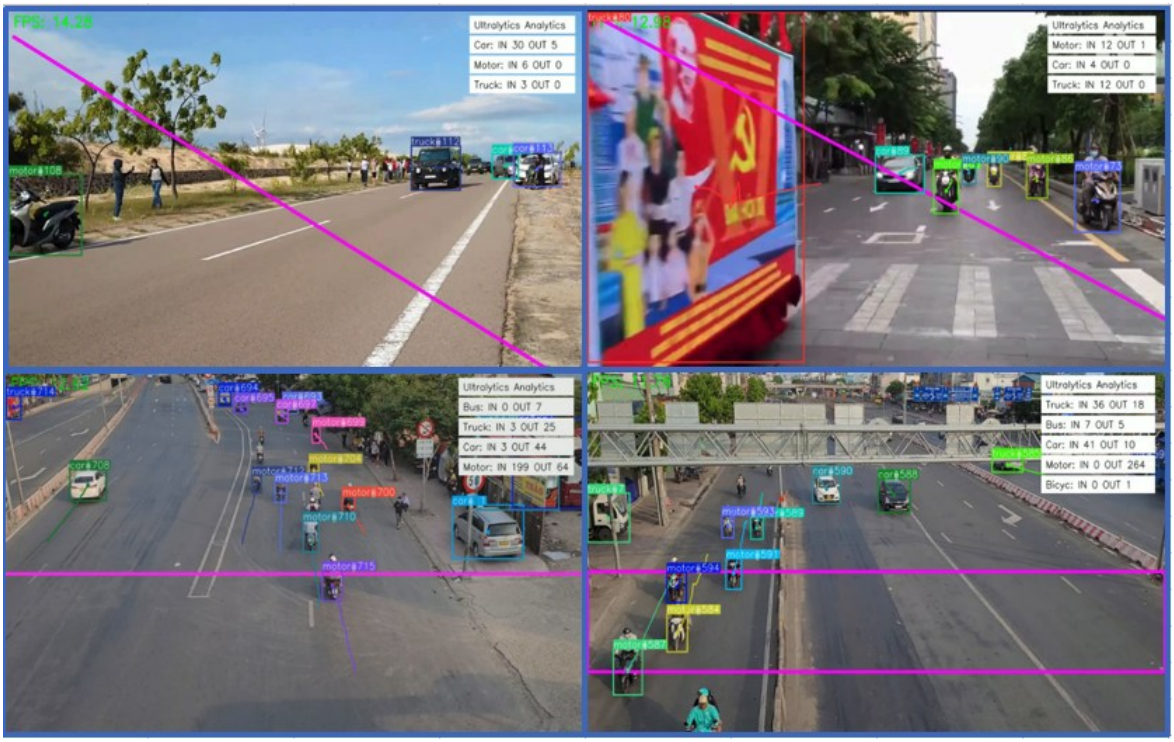


Figure 12. Some Videos Tagged in Footbridge Activity

4. CONCLUSION

In this research, the YOLOv8 algorithm has been successfully applied to identify and classify vehicles in HCMC and the COCO dataset. The model can identify and classify five common types of vehicles: motorbikes, cars, trucks, buses, and bicycles. The findings from the two datasets indicate that the proposed method is effective in identifying and classifying vehicles with high accuracy. Specifically, precision on the two datasets, HCMC and COCO, is {93.33%, 100%} and {98.27%, 100%}. The research has been tested on numerous real videos recorded along inner-city routes in HCMC and has achieved good results. At the same time, in addition, the comparison results with previous studies show that YOLOv8 is the best method compared to vehicle detection techniques when possessing Average Accuracy AP with IOU thresholds of 50 and IOU 75, both of which show a significant difference when compared to the proposed method and earlier methods. In addition, the comparison results indicate that the YOLOv8 variant outperforms other advanced vehicle identification methods in identifying vehicle types, with accuracy approaching 100% across all categories. In short, the research results have contributed to the experiment and demonstrated that the YOLOv8 algorithm is

highly effective in identifying and classifying vehicles; these findings can be widely applied in practice. For future studies, real-world datasets from large cities or areas with high traffic volumes and frequent traffic congestion will continue to be collected to evaluate the effectiveness of YOLOv8.

REFERENCES

[1] N. Al Mudawi *et al.*, "Vehicle detection and classification via YOLOv8 and deep belief network over aerial image sequences," *Sustainability*, vol. 15, no. 19, p. 14597, 2023, doi: https://doi.org/10.3390/su151914597.

[2] P. Premaratne, I. Jawad Kadhim, R. Blacklidge, and M. Lee, "Comprehensive review on vehicle Detection, classification and counting on highways," *Neurocomputing*, vol. 556, p. 126627, 2023, doi: https://doi.org/10.1016/j.neucom.2023.126627.

[3] H. Rodríguez-Rangel, L. A. Morales-Rosales, R. Imperial-Rojo, M. A. Roman-Garay, G. E. Peralta-Peñuñuri, and M. Lobato-Báez, "Analysis of Statistical and Artificial Intelligence Algorithms for Real-Time Speed Estimation Based on Vehicle Detection with YOLO," *Applied Sciences*, vol. 12, no. 6. 2022. doi: 10.3390/app12062907.

[4] M. Maity, S. Banerjee, and S. S. Chaudhuri, "Faster r-cnn and yolo based vehicle detection: A survey," in *2021 5th international conference on computing methodologies and communication (ICCMC)*, IEEE, 2021, pp. 1442–1447. doi: 10.1109/ICCMC51019.2021.9418274.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002, doi: 10.1109/5.726791.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. doi: 10.1109/CVPR.2014.81.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.

[8] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector BT - Computer Vision – ECCV 2016," B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 21–37.

[9] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD : Deconvolutional Single Shot Detector," *ArXiv*, vol. abs/1701.06659, 2017, [Online]. Available: https://api.semanticscholar.org/CorpusID:7691159

[10] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[11] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525. doi: 10.1109/CVPR.2017.690.

[12] A. Ji and X. Ma, "Vehicle detection and classification for traffic management and autonomous systems using YOLOv10," *Alexandria Engineering Journal*, vol. 127, no. March, pp. 804–816, 2025, doi: 10.1016/j.aej.2025.06.049.

[13] M. Safaldin, N. Zaghden, and M. Mejdoub, "An Improved YOLOv8 to Detect Moving Objects," *IEEE Access*, vol. 12, pp. 59782–59806, 2024, doi: 10.1109/ACCESS.2024.3393835.

[14] A. Ahmed, A. Jalal, and A. A. Rafique, "Salient Segmentation based Object Detection and Recognition using Hybrid Genetic Transform," in *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, 2019, pp. 203–208. doi: 10.1109/ICAEM.2019.8853834.

[15] A. Farooq, A. Jalal, and S. Kamal, "Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map.," *KSII Transactions on Internet & Information Systems*, vol. 9, no. 5, 2015.

[16] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 175–187, 2006, doi: 10.1109/TITS.2006.874722.

[17] X. Bai, M. Huang, M. Xu, and J. Liu, "Reconfiguration Optimization of Relative Motion Between Elliptical Orbits Using Lyapunov-Floquet Transformation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 2, pp. 923–936, 2023, doi: 10.1109/TAES.2022.3193089.

[18] H. Min *et al.*, "A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis," *Expert Systems with Applications*, vol. 224, p. 120002, 2023, doi: https://doi.org/10.1016/j.eswa.2023.120002.

[19] H. Pan, S. Guan, and X. Zhao, "LVD-YOLO: An efficient lightweight vehicle detection model for intelligent transportation systems," *Image and Vision Computing*, vol. 151, p. 105276, 2024, doi: 10.1016/j.imavis.2024.105276.

[20] X. Zhu, F. Chen, Y. Zheng, C. Chen, and X. Peng, "Detection of Camellia oleifera fruit maturity in orchards based on modified lightweight YOLO," *Computers and Electronics in Agriculture*, vol. 226, p. 109471, 2024, doi: 10.1016/j.compag.2024.109471.

[21] B. Wu, S. Feng, S. Jiang, S. Luo, X. Zhao, and J. Zhao, "EB-YOLO:An efficient and lightweight blood cell detector based on the YOLO algorithm," *Computers in Biology and Medicine*, vol. 192, no. December 2024, 2025, doi: 10.1016/j.compbiomed.2025.110288.

[22] M. Li and N. Yan, "IPD-YOLO: Person detection in infrared images from UAV perspective based on improved YOLO11," *Digital Signal Processing: A Review Journal*, vol. 168, no. PA, p. 105469, 2026, doi: 10.1016/j.dsp.2025.105469.

[23] C. Liu *et al.*, "Grain-YOLO: An improved lightweight YOLO v8 and its android deployment for rice grains detection," *Computers and Electronics in Agriculture*, vol. 237, no. October 2024, 2025, doi: 10.1016/j.compag.2025.110757.

[24] S. F. Stefenon, L. O. Seman, A. C. R. Klaar, R. G. Ovejero, and V. R. Q. Leithardt, "Hypertuned-YOLO for interpretable distribution power grid fault location based on EigenCAM," *Ain Shams Engineering Journal*, vol. 15, no. 6, p. 102722, 2024, doi: 10.1016/j.asej.2024.102722.

[25] J. Wu, "Design and Exploration of Real Time Monitoring System for Power Grid Operation Safety Based on YOLO Object Detection Algorithm," *Procedia Computer Science*, vol. 228, pp. 889–897, 2023, doi: 10.1016/j.procs.2023.11.117.

[26] L. Chen, G. Li, S. Zhang, W. Mao, and M. Zhang, "YOLO-SAG: An improved wildlife object detection algorithm based on YOLOv8n," *Ecological Informatics*, vol. 83, p. 102791, 2024, doi: 10.1016/j.ecoinf.2024.102791.

[27] K. N. and C. D.N., "Optimized YOLOv2 based vehicle classification and tracking for intelligent transportation system," *Results in Control and Optimization*, vol. 2, p. 100008, 2021, doi: https://doi.org/10.1016/j.rico.2021.100008.

https://doi.org/10.31436/iiumej.v27i1.3703

Deep Learning Framework for Sentiment Prediction using Residual Connections in Bidirectional – Gated Recurrent Unit

VARSHA NAGARAJU*, MUTHALAMBIKA CHALUVEGOWDA PADMA

Department of Computer Science and Engineering, PES College of Engineering, PES College Campus, Mandya, India

**Corresponding author: nvarsha453@gmail.com*

(Received: 18 April 2025; Accepted: 26 November 2025; Published online: 12 January 2026)

ABSTRACT: Sentiment analysis plays an essential role in Natural Language Processing (NLP) for differentiating emotions and opinions expressed in various pieces of text. However, existing algorithms face challenges in handling complex language patterns and capturing long-term dependencies, thereby increasing overall computational cost. This research aims to design an improved sentiment analysis model that enhances accuracy and efficiency while addressing gradient-related limitations in deep networks. This research proposes a Residual Bidirectional Gated Recurrent Unit (RBi-GRU) algorithm for effective sentiment analysis, leveraging residual connections to improve accuracy and efficiency. Residual connections are incorporated into the Bi-GRU network to facilitate gradient flow across layers and mitigate the vanishing gradient problem during training. It also enables deeper networks by protecting data from earlier layers, which further enhances feature representation. Additionally, tokenization, stemming, and global vector-based word representations (GloVe) are employed during preprocessing to capture the semantic relationships and meanings of words, thereby improving contextual understanding in sentiment analysis. The developed RBi-GRU algorithm achieves 98.74% accuracy, 98.99% precision, 98.32% sensitivity, and 98.64% F1-score on the Sentiment140 dataset, compared with the Rectified Linear Unit-based Gated Recurrent Unit (ReLU-GRU).

ABSTRAK: Analisis sentimen memainkan peranan penting dalam Pemprosesan Bahasa Semula Jadi (NLP) bagi membezakan emosi dan pendapat yang dizahirkan dalam teks; namun, algoritma sedia ada menghadapi cabaran pengendalian corak bahasa yang kompleks serta kebergantungan jangka panjang, sekaligus meningkatkan masa pemprosesan. Kajian ini bertujuan mereka bentuk model analisis sentimen berketepatan tinggi dan cekap sambil menangani kekangan berkaitan kecerunan rangkaian mendalam. Sebuah algoritma Unit Kawalan Berulang Baki Dua Arah (RBi-GRU) dicadangkan dengan gabungan baki ke dalam rangkaian Bi-GRU bagi memudahkan aliran kecerunan antara lapisan dan mengurangkan masalah lenyap kecerunan semasa latihan, di samping membolehkan pembinaan rangkaian lebih mendalam dan meningkatkan perwakilan ciri. Selain itu, teknik prapemprosesan seperti penandaan token, pengakaran (stemming), serta penggunaan Representasi Kata Vektor Global (GloVe) diaplikasi bagi menangkap hubungan semantik dan makna kontekstual perkataan dengan lebih berkesan. Dapatan kajian eksperimen menunjukkan bahawa algoritma RBi-GRU mencapai ketepatan 98.74%, kejituan 98.99%, kepekaan 98.32%, dan skor F1 sebanyak 98.64% pada set data Sentimen140, sekaligus mengatasi prestasi model Unit Kawalan Berulang berasaskan Unit Pembetulan Linear (ReLU-GRU).

KEYWORDS: *Bidirectional – Gated Recurrent Unit, Global Vectors for Word Representation, Residual connection, Sentiment analysis, Stemming and Tokenization*

https://doi.org/10.31436/iiumej.v27i1.3703

1. INTRODUCTION

The rapid growth of Natural Language Processing (NLP) techniques and the increasing availability of large text corpora have led to large-scale language models, enabling efficient application to social classification tasks [1-3]. Sentiment analysis, also known as opinion mining, is a crucial process in NLP that aims to identify opinion or emotional tones in text [4]. Sentiment analysis recognizes and classifies opinions in text as positive, neutral, and negative [5]. The primary aim of research in sentiment analysis is to classify text content into positive and negative categories based on emotional connotation [6,7]. Manual classification in sentiment analysis is time-consuming and requires multiple experts to label text [8], where Machine Learning (ML)-based algorithms become essential for online sentiment analysis [9]. Although these algorithms exhibit better performance, they are slow and time-consuming to train [10]. In recent times, the combination of ML and Deep Learning (DL)-based algorithms has become predominant in sentiment analysis due to their superior capability to handle non-linear relationships and learn intricate patterns [11,12]. DL-based algorithms, particularly neural networks, are well-suited to capturing nuanced sentiment expressions and learning discriminative features from the data [13]. ML-based algorithms like Support Vector Machine (SVM) and Naïve Bayes (NB) struggle to capture subtle contextual nuances, irony, and humor, which are crucial for NLP tasks. This limitation stems from their reliance on lexicons and hand-crafted features. In contrast, DL-based algorithms such as Recurrent Neural Networks (RNNs) and Gated Recurrent Units (GRUs) can model sequential information and track how sentiment evolves across text [14]. However, existing DL algorithms still face challenges, including handling complex language patterns, managing long-term dependencies, and being time-consuming. To overcome these drawbacks, this research proposes a DL-based algorithm designed to effectively capture intricate language patterns, maintain long-term dependencies, and operate more efficiently in less time.

Jacob and Viswanatham [15] introduced a Rectified Linear Unit-based Gated Recurrent Unit (ReLU-GRU) for sentiment analysis and classification. Initially, preprocessing was performed using stemming, tokenization, part-of-speech tagging, and punctuation removal. Next, Latent Dirichlet Allocation (LDA), Term Frequency-Inverse Document Frequency (TF-IDF), and bag-of-words (BoW) were introduced for feature selection. Next, classification was performed using the Improved Atom Search Optimizer (ASO) and Simulated Annealing (SA). Finally, classification was performed using a ReLU-GRU to select relevant features into different categories. The introduced method selected relevant features and enhanced classification performance. However, this method required more computational power, which is a major drawback for the model. Loh et al. [16] presented MPNet-GRU, a hybrid DL method that combined three major elements: Bi-GRU, GRU, and MPNet.

This transformer-based pre-trained language model improved language understanding through masked and permuted language modelling. MPNet captured long-term dependencies both unidirectionally and bidirectionally. Integrating these techniques yielded an efficient and effective solution for sentiment analysis. The presented method enabled sophisticated, context-aware sentiment analysis with high classification performance; however, it did not capture relationships among words, thereby affecting overall classification performance. Parveen et al. [17] developed a Gated Attention Recurrent Network (GARN) method for sentiment analysis. Initially, the dataset was preprocessed to clean and filter the data.

The term weight-enabled feature extraction using a Log Term Frequency-enabled Modified Inverse Class Frequency (LTF-MICF) method was employed to extract sentiment-based features. Then, a Hybrid Mutation-based White Shark Optimizer (HMWSO) was

implemented for feature selection. At the same time, a GARN architecture was integrated with a Recurrent Neural Network (RNN) and attention mechanisms for sentiment classification. The developed method minimized system complexity through improved performance, but did not address vanishing gradients during training. Onan [18] implemented a Semantic Role Labeling–Ant Colony Optimization (SRL-ACO) framework to generate additional training data for NLP techniques. The implemented framework used SRL to identify semantic words in a sentence, while ACO was used to generate new sentences and preserve these roles. The implemented algorithm improved accuracy by generating additional data without requiring manual data annotation. However, the implemented method exhibited high memory usage during sentiment analysis. Kumar et al. [19] suggested an Enhanced Vector Space Model (EVSM) with Hybrid Support Vector Machine (HSVM) for sentiment analysis.

The EVSM was assigned to categorize text by mapping it to high-dimensional vector spaces and extracting relationships among words and their contextual meanings. The Decision Tree (DT) algorithm was used with SVM to refine the selection process. The vector space model employed appropriate data-retrieval techniques, thereby improving classification accuracy. The suggested method did not capture sequential features in text. Rahman et al. [20] introduced a hybrid DL model, RoBERTa-BiLSTM, that integrates the Robustly Optimized BERT Pretraining Approach (RoBERTa) with a Bidirectional Long Short-Term Memory (BiLSTM). RoBERTa was used to generate meaningful word embeddings, while a BiLSTM effectively extracted contextual semantics from long dependency texts. The model leveraged the strengths of both sequential and transformer-based methods to improve performance in sentiment analysis.

The analysis indicates that these existing algorithms face the following drawbacks: limited computational power, failure to capture relationships among words, failure to address vanishing gradients, and high memory usage. To address these drawbacks, this study proposes the RBi-GRU method for effective sentiment analysis that employs residual connections. The glove word embedding technique is used in the preprocessing phase to capture semantic relationships and meanings among words, thereby improving context understanding for sentiment analysis. By incorporating residual connections into the Bi-GRU network, the flow of gradients across layers is enhanced, thereby mitigating the vanishing gradient problem. The significant contributions of this research are described as follows:

- Tokenization, stemming, and Global Vectors for Word Representation (Glove) techniques are used in the pre-processing stage to capture the semantic relationships and meanings of words and improve the context understanding for sentiment analysis.
- The proposed Residual - Bidirectional Gated Recurrent Unit (RBi-GRU) algorithm is employed in the classification phase to capture the sequential features from data for high classification performance.
- Residual connections are incorporated into the Bi-GRU network to improve gradient flow, while mitigating the issue of vanishing gradients. It also equips deeper networks by protecting data from earlier layers, thereby enhancing feature representation for effective sentiment analysis.

The remainder of this research paper is organized as follows: Section 2 details the functioning of the developed algorithm; Section 3 presents the experimental results and compares the proposed algorithm with existing models; and Section 4 discusses the results and advantages of the proposed algorithm. Finally, Section 5 concludes the research study.

https://doi.org/10.31436/iiumej.v27i1.3703

2. METHODOLOGY

This study introduces an effective DL-based algorithm for sentiment analysis using data from the Sentiment-140 dataset. This data is preprocessed using tokenization, stemming, and GloVe word embeddings, after which the preprocessed features are classified using the developed RBi-GRU method with residual connections. Figure 1 illustrates the functioning of the proposed model in sentiment analysis.

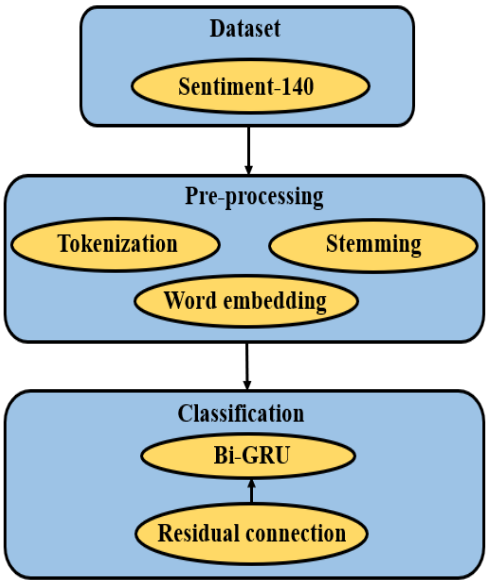


Figure 1. Process of sentiment analysis.

2.1. Dataset

The Sentiment140 dataset [21] comprises 1.6 million tweets, categorized as either detrimental or favorable. This allows one to view what individuals on Twitter perceive of a particular brand, subject, or product. The data is stored in a CSV file, without explicit emotion annotations. The dataset contains different components that define its file format, which are provided below. Figure 2 illustrates the data distribution.

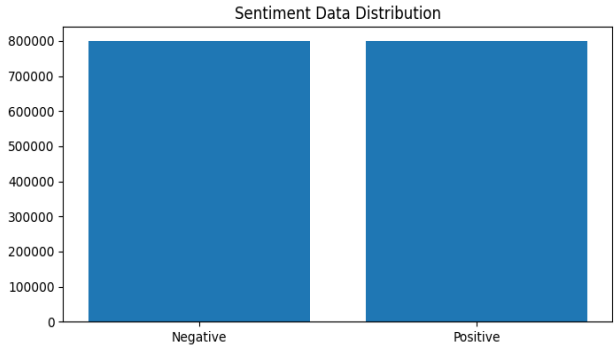


Figure 2. Data distribution.

- The polarity of the tweet (0 – negative, 2- neutral, and 4 – favourable)
- Tweet's identifier
- Tweet's creation data
- Question
- Tweet's content

https://doi.org/10.31436/iiumej.v27i1.3703

2.2. Pre-processing

The data is pre-processed using tokenization, stemming, and word embedding to improve data quality.

2.2.1. Tokenization

Tokenization involves splitting text into individual tokens or words. This process transforms unstructured text into a more structured format, enabling more effective analysis. By breaking down larger texts into smaller segments, tokenization facilitates the analysis of specific words and helps capture their semantic relationships. Unprocessed phrases and sentences are thus converted into smaller segments, called tokens, to enable analysis.

2.2.2. Stemming

Stemming is the process of reducing words to their root or base form. By converting related word variants to their base form, this process helps normalize textual data. For example, the words "running" and "ran" are both reduced to "run." This approach groups related terms and preserves the core meaning of words, minimizing the influence| of different word forms.

2.2.3. Word Embedding

Word embedding is the process of representing words as vectors based on a global co-occurrence matrix using GloVe. It uses an attenuation function based on the distance between two words in a context window to measure the weight.

2.3. Classification

This study employs a Bi-GRU with residual connections to extract and learn features from the data. The RBi-GRU combines the Bi-GRU layers with residual connections to enhance the neural network's effectiveness. The Bi-GRU is an RNN network that captures various patterns and sequential dependencies. GRU networks perform more effectively than traditional RNNs due to their gating mechanism, which enables the network to update and forget information from previous time steps. This mechanism includes updates and resets gates, which regulate the flow of information through the network. Figure 3 represents the architecture of the Bi-GRU model.

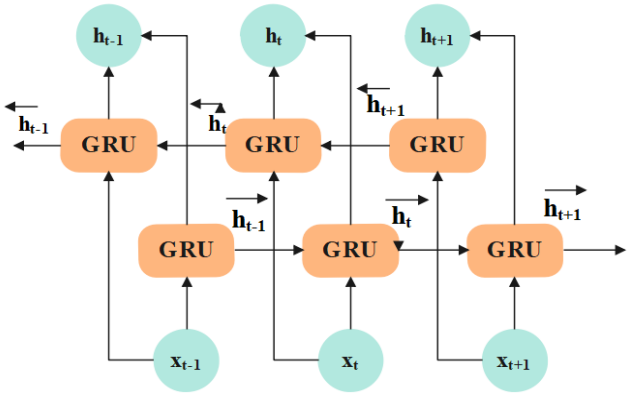


Figure 3. Architecture of Bi-GRU

The mathematical formula for update and reset gates is given in Eqs. (1) and (2),

$$u_t = \sigma(W_{hu}h_{t-1} + W_{xu}v_t + b_u) \quad (1)$$

$$r_t = \sigma(W_{hr}h_{t-1} + W_{xr}v_t + b_r) \quad (2)$$

https://doi.org/10.31436/iiumej.v27i1.3703

In the above Eqs. (1) and (2), u_t represents the update gate at a time t, which acts as a control mechanism that regulates network updates to their internal state by utilizing data from the present input v_t and past hidden state h_{t-1}. The decision to update is influenced by a sigmoid activation function employed with the linear combination of hidden and input states, with respective weight matrices of W_{xu} and W_{hu} and bias b_u. The reset gate is represented as r_t which defines the degree to which the network forgets or rests in the hidden state, depending on the present input. The reset decision is controlled by the respective weight matrices W_{xr} and W_{hr}, where b_r is a biased term in the reset gate.

In Bi-GRU, two GRU units process an input sequence in both the forward and backward directions, allowing the network to incorporate past and future content at each time step. This is particularly used for ligand sequences, which are sequential and exhibit long-range dependencies. The outcome of Bi-GRU for the present time step is represented as h_t, acquired by combining the forward hidden state $\vec{h}_t$ and backward hidden state $\overleftarrow{h}_t$. The mathematical expression for the Bi-GRU output is given in Eq. (3).

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \tag{3}$$

The Bi-GRU processes text in both forward and reverse directions, enabling it to capture past and future contexts within a sentence. The forward direction $\vec{h}_t$ processes the sentence from start to end, capturing how previous words influence later ones. The backward direction $\overleftarrow{h}_t$ processes the sentence from the end to the beginning, capturing how later words provide context for earlier ones. Next, the forward and backward hidden states are combined, offering a comprehensive understanding of the sentence by considering information from both directions in the sequence.

Residual connections are incorporated into Bi-GRU layers to enable direct data transmission across layers. This improves the network's ability to handle deep architectures by mitigating the vanishing gradient problem and enhancing training efficiency. As a result, the method can capture complex relationships among words and phrases in deeper layers without losing significant contextual information. In long, multi-clause sentences, residual connections help retain the overall sentiment while deeper layers refine subtle nuances. Figure 4 illustrates the architecture of the Bi-GRU with residual connections.

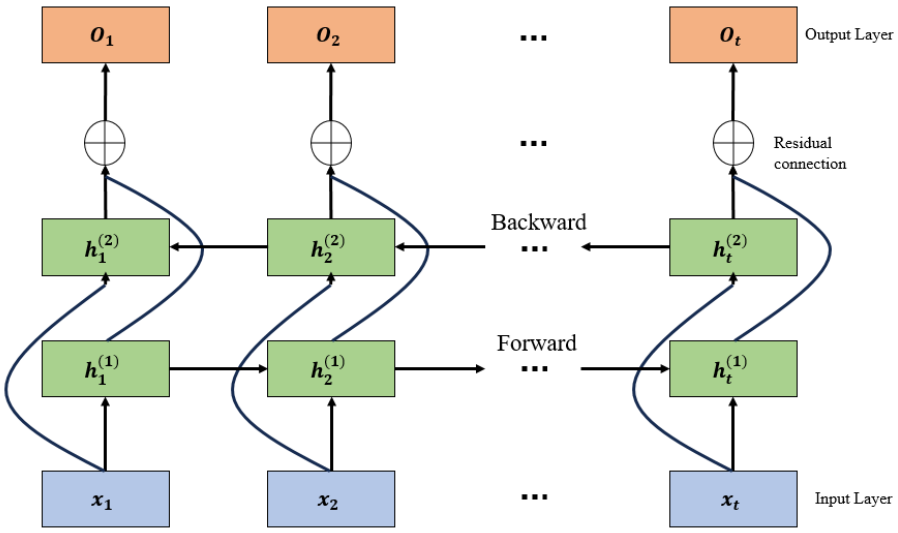


Figure 4. Architecture of Bi-GRU with residual connections

https://doi.org/10.31436/iiumej.v27i1.3703

By measuring $\vec{h}_t$ based on the candidate's hidden state $\tilde{h}_t$, the past hidden state and update gate are mathematically defined in Eq. (4), where $\overleftarrow{h}_t$ in one direction, the same is executed as $\vec{h}_t$ in the opposite direction.

$$\vec{h}_t = (1 - u_t) \ominus h_{t-1} + u_t \ominus \tilde{h}_t \quad (4)$$

Where $\tilde{h}_t$ is the candidate hidden state executed based on the reset gate r_t, h_{t-1} is the past hidden state and v_t is the present input vector, as mathematically formulated in Eq. (5).

$$\tilde{h}_t = tanh(W_{hh}(r_t \ominus h_{t-1}) + W_{xh}v_t + b_h) \quad (5)$$

The RBi-GRU integrates a residual connection into the Bi-GRU layers. The residual connection enables data to be transferred directly between layers and to circulate through intermediate layers. This process avoids vanishing gradients that arise during training due to repeated gradient multiplication. Residual connection stacks multiple layers of Bi-GRU without accounting for the method's convergence time. The RBi-GRU supports skip connections, in which the input to the next layer is obtained by concatenating the result of the previous layer with the residual. The mathematical formula for the result of ith with residual connection is given in Eq. (6),

$$h_i = F(W[h_{i-1}, X] + b) + X \quad (6)$$

In the above Eq. (6), the h_i represents the result of ith the layer with residual connection, the X represents the output of the previous layer, the h_{i-1} represents the result of the past layer, the W represents a weight matrix, the b represents the bias vector and the F represents the non-linear activation function. Algorithm 1 represents the process of the proposed method.

Algorithm 1 – Process of the proposed method

Input – Sentiment-140 dataset

Output – classified output

Split the dataset into a training set (80%) and a testing set (20%)

Define hyperparameters – Epochs 50, batch size 64, learning rate 0.01, Adam optimizer, Categorical cross entropy loss function, 0.3 dropout rate, and 100 embedding dimensions.

Initialize the RBi-GRU model with residual connections

For every epoch from 1 to epochs do

    For every batch in the training set

        Forward propagate batch through embedding + RBi-GRU layers

        Compute loss using the loss function

        Backpropagate loss and update weights using the optimizer

    End for

    Initialize Predictions = []

    For every sample x in test set, do

        $y_{pred} = RBi - GRU(x)$

        predictions.append (y_{pred})

    End for

End for

3. EXPERIMENTAL ANALYSIS

The performance of the RBi-GRU technique is simulated in a Python environment, with configurations of 8 GB of RAM, an Intel Core i5 processor, and Windows 10 (64-bit). Performance metrics like accuracy, sensitivity, precision, and F1-score are considered to validate the performance of the developed RBi-GRU technique. The dataset is split into an 80% training set and a 20% test set. The mathematical expressions for performance metrics are given in Eqs. (7) and (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (7)$$

$$F1-score = \frac{2\times Sensitivity\times Precision}{Sensitivity+Precision} \times 100 \quad (8)$$

Table 1. Performance of RBi-GRU with different classes

Classes	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Positive	98.47	99.88	96.70	98.27
Negative	99.01	98.10	99.93	99.01
Average	98.74	98.99	98.32	98.64

Table 1 presents the performance of RBi-GRU across the dataset's classes, evaluated using various metrics. The developed RBi-GRU technique achieves 98.47% accuracy for the positive class, 99.01% for the negative class, and 98.74% overall.

In Table 2, the performance of the Bi-GRU without a residual connection is evaluated on the Sentiment140 dataset using various metrics. The RNN, LSTM, Bi-LSTM, and standard GRU are conventional algorithms used to evaluate the performance of the Bi-GRU. The Bi-GRU without a residual connection achieves 95.47% accuracy, 94.65% precision, 95.10% recall, and 94.87% F1-score on the sentiment140 dataset.

Table 2. Performance of Bi-GRU without the residual connection method

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RNN	88.43	87.21	89.60	88.39
LSTM	90.87	89.43	91.25	90.34
Bi-LSTM	93.28	92.11	93.15	92.63
GRU	94.02	92.84	93.55	93.19
Bi-GRU	95.47	94.65	95.10	94.87

Table 3 presents the performance of the Bi-GRU with residual connections on the Sentiment 140 dataset, evaluated using various metrics. The RNN, LSTM, Bi-LSTM, and standard GRU are conventional algorithms used to evaluate the performance of the Bi-GRU with a residual connection. The Bi-GRU with a residual connection achieves 98.74% accuracy, 98.99% precision, 98.32% recall, and 98.64% F1-score on the sentiment140 dataset.

Table 3. Performance of Bi-GRU with the residual connection method

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RNN	92.34	91.12	92.5	91.81
LSTM	94.21	93.18	94.5	93.84
Bi-LSTM	96.15	95.38	96.1	95.74
GRU	95.76	94.82	95.9	95.36
RBi-GRU	98.74	98.99	98.32	98.64

https://doi.org/10.31436/iiumej.v27i1.3703

Figure 5 illustrates the memory usage of the developed Bi-GRU with residual connections, compared with the Bi-LSTM. Then, the total memory usage for each parameter, gradient, optimizer, and activation is evaluated. Here, it is observed that, compared with the Bi-LSTM, the Bi-GRU network consumes less memory.

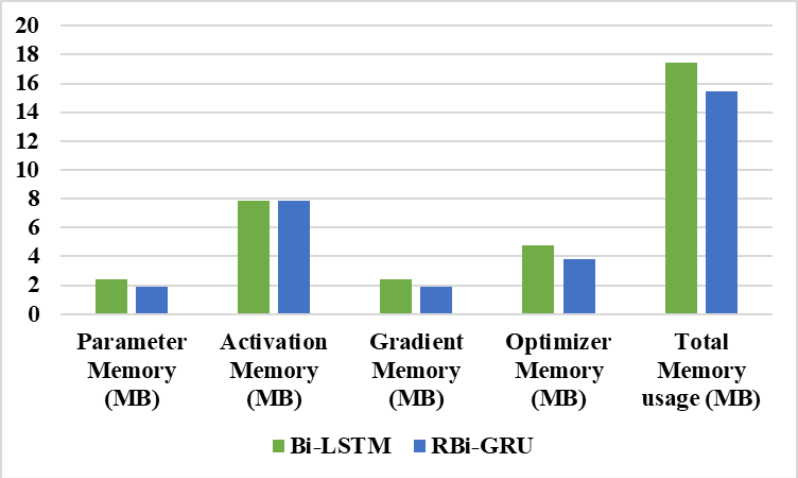


Figure 5. Memory usage of the developed RBi-GRU.

Table 4 presents an evaluation of different K-fold values across various performance metrics. The different k-fold values used are 2, 3, 5, 7, and 10, where, for a k value of 5, an accuracy of 98.74%, precision of 98.99%, recall of 98.32%, and f1-score of 98.64% is obtained, thereby proving superior to other k-fold values.

Table 4. Evaluation of K-fold values

K-fold values	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
2	94.12	93.85	94.10	93.97
3	95.14	94.56	95.00	94.72
5	98.74	98.99	98.32	98.64
7	96.85	96.24	96.56	96.40
10	97.20	96.78	97.00	96.88

Figure 6 presents a plot of accuracy versus epochs. In contrast, Figure 7 shows the loss vs. epochs plot, Figure 8 shows the ROC curve, and Figure 9 shows the confusion matrix, all for the Sentiment140 dataset.

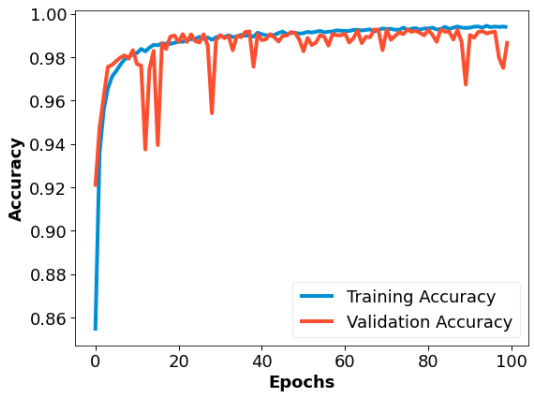


Figure 6. Accuracy vs epochs

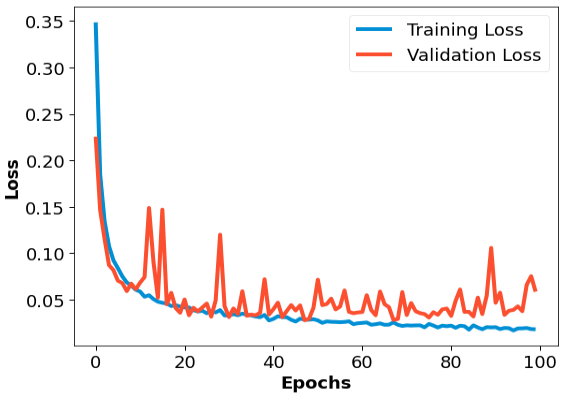


Figure 7. Loss vs epochs

https://doi.org/10.31436/iiumej.v27i1.3703

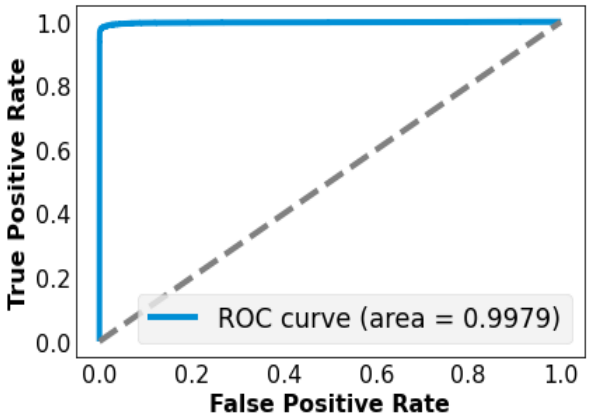


Figure 7. ROC curve

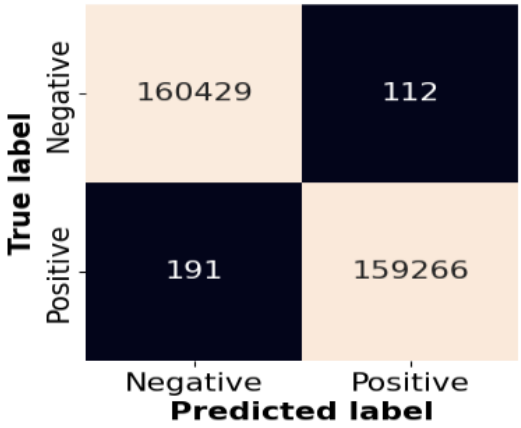


Figure 8. Confusion matrix

Table 5. Comparative Analysis of the RBi-GRU algorithm

Methods	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-score (%)
IASO-SA – GRU [16]	96.52	95.81	96.93	96.37
MPNet-GRU [17]	88.17	88	88	88
GARN [18]	97.86	96.65	96.76	96.70
RoBERTa-BiLSTM [20]	82.25	82.25	82.25	82.25
Proposed RBi-GRU	98.74	98.99	98.32	98.64

The performance of the developed RBi-GRU algorithm is compared with existing methods: IASO-SA–GRU [16], MPNet-GRU [17], GARN [18], and RoBERTa-BiLSTM [20] on the Senti140 dataset. The developed RBi-GRU algorithm achieves an accuracy of 98.74%, precision of 98.99%, sensitivity of 98.32%, and an F1-score of 98.64% on the Sentiment140 dataset. This model incorporates a residual connection with the Bi-GRU to enhance inter-layer connectivity and facilitate smooth gradient flow. This, in turn, improves the Bi-GRU model's training efficiency and classification performance. Table 5 presents a comparative analysis of the RBi-GRU algorithm.

4. DISCUSSION

The proposed algorithm operates in three phases and is applied to the Sentiment-140 dataset, with preprocessing and sentiment classification of the acquired data. The performance of this method is evaluated against the existing methods: RNN, LSTM, Bi-LSTM, and standard GRU on the Sentiment-140 dataset. The performance of the developed algorithm is compared with existing methods such as IASO-SA – GRU [16], MPNet-GRU [17], GARN [18], and RoBERTa-Bi-LSTM [20] on the Senti140 dataset. The analysis indicates that existing algorithms suffer from limited computational efficiency, an inability to capture relationships among words, failure to address vanishing gradients, and high memory consumption. To overcome these drawbacks, this study proposes the RBi-GRU method for effective sentiment analysis by incorporating a residual connection. The glove word embedding technique is used in the preprocessing phase to capture semantic relationships and meanings among words efficiently, thereby improving context understanding for sentiment analysis. Incorporating a residual connection into the Bi-GRU network enhances inter-layer gradient flow and mitigates the vanishing gradient problem. Despite its deep architecture, the RBi-GRU consumes less memory for parameters, activations, gradients, and optimizer states than the Bi-LSTM, making it a practical choice for large-scale sentiment analysis with limited computational resources.

The proposed RBi-GRU model also demonstrates improvements in sentiment analysis performance when compared to existing models. By incorporating residual connections into the Bi-GRU model, it efficiently addresses primary challenges in the recurrent neural network. These improvements enable the network to preserve significant contextual data across multiple layers, thereby improving understanding of nuanced sentiment expressions. Additionally, experimental results across multiple k-fold values demonstrate that the model maintains high stability and generalization ability with K=5, achieving the best performance-training stability balance. These findings show that residual connections in Bi-GRU improve overall prediction accuracy, minimize memory overhead, and enhance training convergence.

5. CONCLUSION

The DL-based algorithm is developed for effective sentiment analysis using the Sentiment-140 dataset. Tokenization, stemming, and GloVe word embeddings are employed during the preprocessing stage, thereby enhancing data quality and facilitating effective sentiment analysis. They also capture semantic relationships and meanings among words, improving context understanding. Then, classification is performed using the developed RBi-GRU algorithm, which includes a residual connection. The residual connection incorporated into the Bi-GRU network facilitates inter-layer gradient flow and mitigates the vanishing gradient problem. This process contributes to effective sentiment analysis. The developed RBi-GRU algorithm achieved 98.74% accuracy, 98.99% precision, 98.32% sensitivity, and 98.64% F1-score on the Sentiment-140 dataset. However, the loss of significant features reduced the effectiveness and overall performance of the developed algorithm. In the future, various DL-based algorithms with effective feature selection techniques will be developed to classify sentiments more accurately.

REFERENCES

[1] Krishna MM, Duraisamy B, Vankara J (2023). Independent component support vector regressive deep learning for sentiment classification. Meas.: Sens, 26:100678. https://doi.org/10.1016/j.measen.2023.100678

[2] Havinga I, Marcos D, Bogaart P, Tuia D, Hein L (2024). Understanding the sentiment associated with cultural ecosystem services using images and text from social media. Ecosyst. Serv, 65:101581. https://doi.org/10.1016/j.ecoser.2023.101581

[3] Xu QA, Jayne C, Chang V (2024). An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews. Technol. Forecasting Social Change, 202:123326. https://doi.org/10.1016/j.techfore.2024.123326

[4] Jain R, Kumar A, Nayyar A, Dewan K, Garg R, Raman S, Ganguly S (2023). Explaining sentiment analysis results on social media texts through visualization. Multimedia Tools Appl, 82(15):22613-22629. https://doi.org/10.1007/s11042-023-14432-y

[5] Petrescu A, Truica CO, Apostol ES, Paschke A (2024). EDSA-Ensemble: an event detection sentiment analysis ensemble architecture. IEEE Trans. Affective Comput, 1-18. doi: 10.1109/TAFFC.2024.3434355

[6] Halawani HT, Mashraqi AM, Badr SK, Alkhalaf S (2023). Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. Alexandria Eng. J, 80:433-443. https://doi.org/10.1016/j.aej.2023.08.062

[7] Kaur G, Sharma A (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. Journal of big data, 10(1):5. https://doi.org/10.1186/s40537-022-00680-6

[8] Üveges I, Ring O (2023). HunEmBERT: a fine-tuned BERT-model for classifying sentiment and emotion in political communication. IEEE Access, 11:60267-60278. doi: 10.1109/ACCESS.2023.3285536

[9] Barik K, Misra S, Ray AK, Bokolo A (2023). LSTM-DGWO-Based Sentiment Analysis Framework for Analyzing Online Customer Reviews. Comput. Intell. Neurosci, 2023(1):6348831. https://doi.org/10.1155/2023/6348831

[10] Kastrati M, Kastrati Z, Shariq Imran A, Biba M (2024). Leveraging distant supervision and deep learning for twitter sentiment and emotion classification. J. Intell. Inf. Syst, 62(4):1045-1070. https://doi.org/10.1007/s10844-024-00845-0

[11] Kumar T, Mahrishi M, Sharma G (2023). Emotion recognition in Hindi text using multilingual BERT transformer. Multimedia Tools Appl, 82(27):42373-42394. https://doi.org/10.1007/s11042-023-15150-1

[12] Zou H, Wang Z (2023). A semi-supervised short text sentiment classification method based on improved Bert model from unlabelled data. Journal of Big Data, 10(1):35. https://doi.org/10.1186/s40537-023-00710-x

[13] Zhou L, Zhang Z, Zhao L, Yang P (2023). Microblog sentiment analysis based on deep memory network with structural attention. Complex Intell. Syst, 9(3):3071-3083. https://doi.org/10.1007/s40747-022-00904-5

[14] Yadav NB (2023). Harnessing customer feedback for product recommendations: an aspect-level sentiment analysis framework. Hum.-Centric Intell. Syst, 3(2):57-67. https://doi.org/10.1007/s44230-023-00018-2

[15] Jacob N, Viswanatham VM (2024). Sentiment Analysis using Improved Atom Search Optimizer with a Simulated Annealing and ReLU based Gated Recurrent Unit. IEEE Access, 12:38944-38956. doi: 10.1109/ACCESS.2024.3375119

[16] Loh NKN, Lee CP, Ong TS, Lim KM (2024). MPNet-GRUs: Sentiment Analysis with Masked and Permuted Pre-training for Language Understanding and Gated Recurrent Units. IEEE Access, 12:74069-74080. doi: 10.1109/ACCESS.2024.3394930

[17] Parveen N, Chakrabarti P, Hung BT, Shaik A (2023). Twitter sentiment analysis using hybrid gated attention recurrent network. Journal of Big Data, 10(1):50. https://doi.org/10.1186/s40537-023-00726-3

[18] Onan A (2023). SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. Journal of King Saud University-Computer and Information Sciences, 35(7):101611. https://doi.org/10.1016/j.jksuci.2023.101611

[19] Suresh Kumar K, Radha Mani AS, Ananth Kumar T, Jalili A, Gheisari M, Malik Y, Chen HC, Jahangir Moshayedi A (2024). Sentiment Analysis of Short Texts Using SVMs and VSMs-Based Multiclass Semantic Classification. Applied Artificial Intelligence, 38(1):2321555. https://doi.org/10.1080/08839514.2024.2321555

[20] Rahman, M.M., Shiplu, A.I., Watanobe, Y. and Alam, M.A., 2025. RoBERTa-BiLSTM: A context-aware hybrid model for sentiment analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

[21] Sentiment-140 dataset: https://www.kaggle.com/datasets/kazanova/sentiment140.

https://doi.org/10.31436/iiumej.v27i1.3716

ORCA: AI-powered Autonomous Underwater Vehicle for Subaquatic Exploration

OSKAR NATAN[1]*, WIWIT SURYANTO[2], ANDI DHARMAWAN[1], RIFDA HAKIMA SARI[1], ZAIDAN HAKIM[1]

[1]*Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Department of Physics, Universitas Gadjah Mada, Yogyakarta, Indonesia*

**Corresponding author: oskarnatan@ugm.ac.id*

(Received: 1 May 2025; Accepted: 19 November 2025; Published online: 12 January 2026)

ABSTRACT: This paper presents research and development of an Autonomous Underwater Vehicle (AUV), named ORCA, to perform underwater missions independently without human intervention. ORCA plays a vital role in challenges that test robots' abilities in navigation, exploration, and interaction with the aquatic environment. Using advanced design tools, the AUV is meticulously designed in Inventor and manufactured via CNC machining, laser cutting, and 3D printing. We concentrate on the vehicle's design, manufacturing processes, control systems, PID controllers, and vision systems. Subsequently, the research and development effort is expanded to incorporate critical functionalities, including environmental perception, object detection, deep learning algorithms, and path-planning strategies. The culmination of this research has produced an AUV capable of autonomous underwater navigation, effective obstacle avoidance, efficient object detection, and precise payload manipulation using a gripper mechanism. ORCA has a comprehensive sensor suite comprising a BNO055 IMU, an MS5803-14BA pressure sensor for depth measurement, and a Logitech C525 camera for image processing. The system integration not only enhances the vehicle's operational capabilities but also represents a significant advancement in underwater robotics.

ABSTRAK: Kajian ini membentangkan penyelidikan dan pembangunan Kenderaan dalam Air Berautonomi (AUV) yang dikenali sebagai ORCA, dibangunkan bagi tujuan misi bawah air secara autonomi tanpa sebarang campur tangan manusia. ORCA memainkan peranan penting dalam menghadapi cabaran menguji kebolehan robot dalam aspek navigasi, penerokaan, serta interaksi dengan persekitaran akuatik. Dengan memanfaatkan alat reka bentuk yang canggih, AUV ini direka bentuk dengan teliti menggunakan perisian Inventor dan dihasilkan melalui teknik pemesinan CNC, pemotongan laser, dan percetakan 3D. Penekanan utama dalam penyelidikan ini adalah pada reka bentuk kenderaan, proses pembuatan, sistem kawalan, pengawal PID, serta sistem penglihatan. Selain itu, penyelidikan dan pembangunan ini turut diperluaskan bagi menyertakan fungsi-fungsi kritikal seperti persepsi persekitaran, pengesanan objek, algoritma pembelajaran mendalam, dan strategi perancangan laluan. Dapatan kajian ini telah menghasilkan AUV yang mampu melaksanakan navigasi bawah air secara autonomi, menghindari halangan dengan berkesan, mengesan objek dengan cekap, serta mengendali beban dengan tepat menggunakan mekanisme pemegang. ORCA dilengkapi dengan suit pengesan yang komprehensif, termasuk pengesan IMU BNO055, pengesan tekanan MS5803-14BA bagi pengukuran kedalaman, serta kamera Logitech C525 bagi pemprosesan visual. Integrasi sistem ini bukan sahaja meningkatkan keupayaan operasi kenderaan tetapi juga mewakili satu langkah penting dalam perkembangan robotik bawah air.

https://doi.org/10.31436/iiumej.v27i1.3716

KEYWORDS: *Robotics, Autonomous Robot Systems, Underwater Robot, Computer Vision, Navigational Control*

1. INTRODUCTION

ORCA is an Autonomous Underwater Vehicle (AUV) designed to perform independent underwater tasks without human intervention. In developing ORCA, we focus on manufacturing an Autonomous Underwater Vehicle (AUV), leveraging knowledge and experience gained in previous years and enhancing the vehicle's control systems and navigation algorithms to improve efficiency and accuracy during underwater operations. This research focuses on advancing underwater robotics by addressing challenges such as navigation in low-visibility environments, effective obstacle detection, and energy-efficient operations. ORCA aims to provide a cost-effective solution by integrating a robust mechanical design, deep learning-based vision systems, and advanced control strategies. This work leverages state-of-the-art technologies, such as pruned YOLOv8 for real-time underwater object detection and a Jetson Nano Development Kit for high-performance computing, thereby setting it apart from traditional AUV designs. Tables 1 and 2 highlight the advancements of ORCA. Key upgrades include a new single-board computer and software architecture, a transition from FreeRTOS to Ubuntu 18.04, and the incorporation of a microcontroller to assist the single-board computer with onboard computation.

Table 1. Table of Physical Specifications and Control System

Weight	Dimension	Single Board Computer	Microcontroller	Propulsion
15 kg	55cm × 70cm × 36 cm	NVidia Jetson Nano Development Kit	ESP32	8 ROVMAKER T200 Thruster

Table 2. Table of Sensor, Power, and Software Architecture Specifications

Navigation	Vision	Power Supply	Underwater Connector	Software Architecture
Sensor IMU BNO-055, Sensor MS5803-14BA	Logitech C525	14.8V 7200mAh Lipo Battery	M10, M16 Cable Penetrator, Amass XT60W and Sochi SP17H IP68 10-pin	Operating System: Ubuntu 18.04, Network Model: YOLOv8 (pruned), Deep Learning (DL) Framework: TensorFlow/Keras/Darknet, Peripherals: FPIOA, UART, GPIO, SPI, etc

This research addresses the question: How can an autonomous underwater vehicle (AUV) achieve reliable and efficient navigation, obstacle detection, and task execution in challenging underwater environments with minimal computational resources? The study explores the integration of modern hardware and software frameworks to address these challenges.

2. DESIGNS AND METHODS

2.1. Mechanical Design

The overall design of the Autonomous Underwater Vehicle (AUV), such as ORCA, must align with its research objectives, which can be categorized into two main types: high maneuverability or high speed. To achieve maneuverability, the design incorporates multiple actuators that enhance the Degree of Freedom (DoF). These actuators can be positioned at

https://doi.org/10.31436/iiumej.v27i1.3716

specific angles to direct forces as vectors, facilitating the desired movements. On the other hand, designs focused on achieving high speed prioritize a hydrodynamic shape to minimize drag forces, which the equation can express:

$$\vec{F}_{drag} = \frac{1}{2} C_d A \rho V^2 \quad (1)$$

In this equation, A represents the frontal area, ρ is the fluid density, V is the robot's speed, and C_d is the drag coefficient, which accounts for both frictional and pressure drag. The drag coefficient can be considered constant over the Reynolds number range 103–105. This assumption simplifies drag-force calculations during the design process, allowing engineers to focus on optimizing the vehicle's hydrodynamics.

2.1.1. Design Overview

Figure 1 shows the frame, which consists of various materials: a 2 mm-thick grade 430 stainless steel baseplate, 8mm-thick acrylic side panels and thruster guards, and aluminum extrusions for connections and mounts. Additionally, 3D-printed components made from PLA+ and ABS+ filament include the electronics enclosure bracket, thruster guards, and buoyancy floats. All parts are secured with stainless steel bolts and aluminum brackets. Materials were selected for their durability in underwater environments: grade 430 stainless steel and acrylic provide water resistance, while aluminum offers strength. PLA+ is chosen for its adequate strength, water resistance, and cost-effectiveness. Cost considerations were crucial in selecting materials, resulting in a high-quality yet affordable frame.

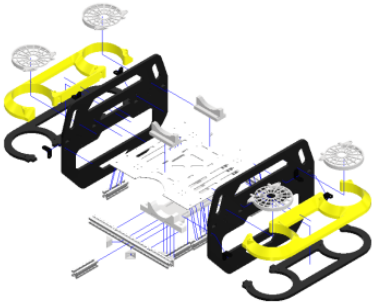


Figure 1. Frame of ORCA

2.1.2. Thruster and Propeller Guards

Thruster guards are essential components of underwater robots, protecting thrusters from damage caused by collisions with underwater obstacles, such as rocks or coral. The thruster and propeller guards used in this project are yellow and white, as shown in Figure 2. By safeguarding the thrusters, which are critical to the AUV's propulsion, the guards extend the components' lifespans and ensure optimal performance during underwater operations.

Figure 2. Thruster and Propeller Guards

https://doi.org/10.31436/iiumej.v27i1.3716

2.1.3. Electronics Cabin Bracket

The electronics cabin bracket is crucial for organizing the available space within the enclosure tube. It is fabricated from ABS+ and PLA+ filaments using 3D printing. ABS+ is chosen for its high-temperature resistance, as detailed in [1]. PLA+, on the other hand, is used for the flange connections due to its superior performance in creating supported structures during printing. The electronics cabin design accommodates all necessary electronic components.

2.1.4. Buoyancy and Ballast

The buoyant force is the force acting on an object that is partially or fully submerged in a fluid, and it plays a crucial role in the mechanical systems of underwater robots [2]. Buoyant forces can be classified as negative, positive, or neutral. Negative buoyancy occurs when the buoyant force is less than the robot's weight, causing it to sink. In contrast, positive buoyancy allows the robot to float, whereas neutral buoyancy enables it to hover. The formula for buoyant force is given by:

$$F_b = -\rho g V \tag{2}$$

where F_b is the buoyant force in Newtons, ρ is the density of the fluid in kilograms per cubic meter, V is the volume of displaced fluid in cubic meters, and g is the acceleration due to gravity. In design, a robot's weight may exceed or be less than its buoyant force, a challenge that can be addressed using ballast tanks. Increasing the volume of the ballast tank expands the range of buoyancy adjustments. Proper placement of the ballast tank is crucial for maintaining balance, as the center of gravity (CoG) must be lower than the center of buoyancy (CoB) to ensure static stability. In the ORCA robot, the ballast tanks are positioned on the upper left and right sides, which helps keep the CoB above the CoG without interfering with the propulsion system. Additionally, the use of buoyant materials, such as foam with a density lower than that of water [3], along with 3D-printed PLA+ fairings to secure the buoys, has been integrated to ensure the robot's stability and optimal performance [4].

2.1.5. Gripper

The gripper, often considered the most common end effector in manipulators due to its resemblance to a human hand, is an essential component of underwater vehicles that perform specific tasks. Grippers can have various mechanisms tailored to their actuators and the objects they handle [5]. ORCA employs a two-finger gripper with two degrees of freedom for grasping and wrist rotation. The gripper is constructed using PLA+ for high-stress components and acrylic for flat parts [6].

2.2. Electronics Design

2.2.1. Design Overview

The electronic subsystem of the ORCA robot comprises the components depicted in Figures 3 and 4. The electronic subsystem of the ORCA AUV is a meticulously designed network of components that ensures reliable communication, precise control, and seamless integration of sensory data [7,8]. At its core, the Jetson Nano Developer Kit serves as the primary computational platform, selected for its ability to support real-time processing of high-resolution image data and to execute complex deep learning algorithms. The Logitech C525 camera connected to the Jetson Nano uses the SCCB protocol to transfer visual data for object detection tasks efficiently. This setup supports the integration of the YOLOv8 model for underwater object recognition, ensuring high accuracy even in low-visibility conditions. Additionally, the system incorporates the MS5803-14BA pressure sensor and the BNO055

https://doi.org/10.31436/iiumej.v27i1.3716

Inertial Measurement Unit (IMU), both connected via the I2C communication protocol. These sensors provide critical data for depth measurement, orientation, and stabilization, forming the backbone of the navigation and control systems.

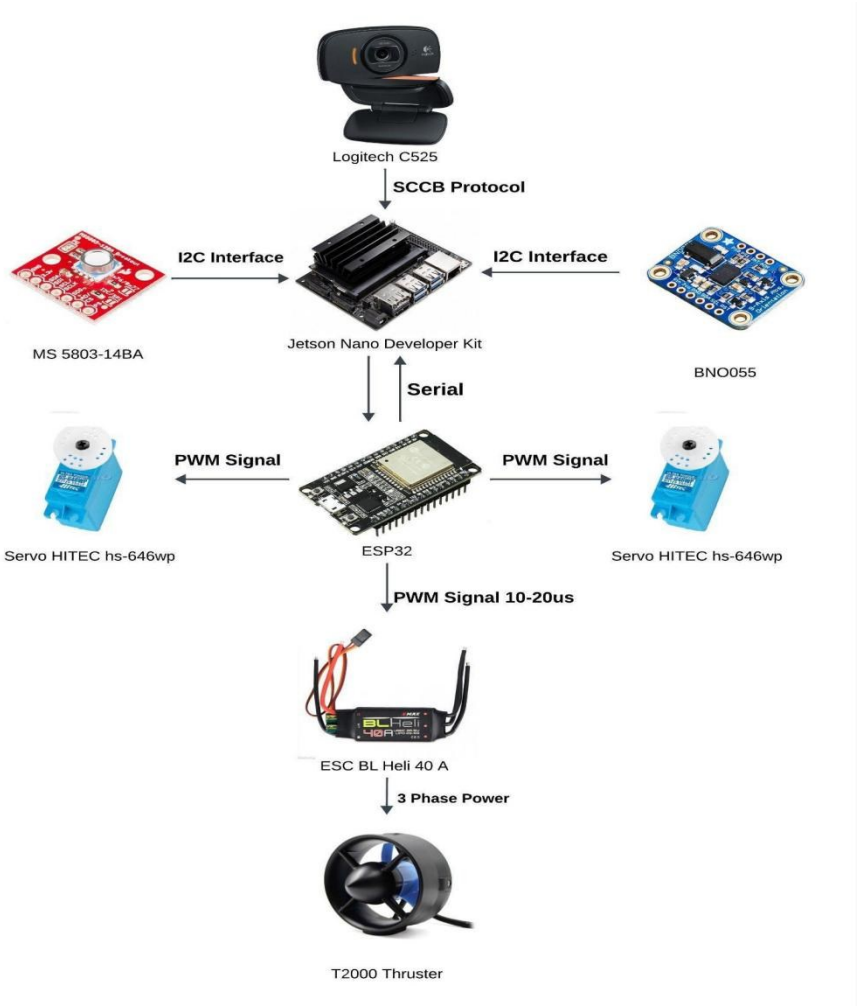


Figure 3. Overview of The Electronic Subsystem

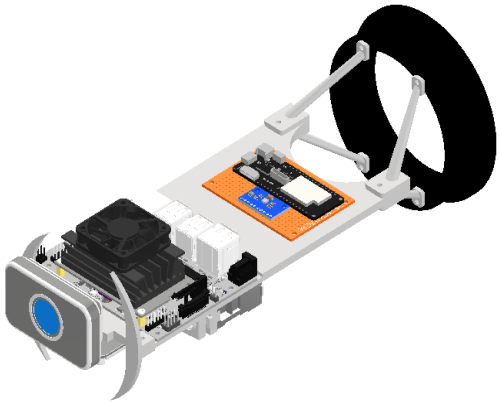


Figure 4. Component Layout on The Bracket

Communication between the Jetson Nano and the ESP32 microcontroller is achieved via a robust serial interface, enabling bidirectional data flow to control the robot's motion and

https://doi.org/10.31436/iiumej.v27i1.3716

operational status. The ESP32 microcontroller generates pulse-width modulation (PWM) signals to both the servo motors and the electronic speed controllers (ESCs). The ESCs, equipped with BLHeli firmware, regulate the operation of the eight ROVMAKER T200 thrusters, which are powered by a three-phase supply. This configuration ensures precise thrust control and optimal propulsion efficiency, critical for maintaining stability and maneuverability in dynamic underwater environments. A dedicated power distribution board (PDB) is employed to supply stable voltage and current to the electronic components. The PDB accepts an input voltage range of 6V to 30V and delivers regulated outputs of 14.8V and 5V at 12A, ensuring consistent power delivery to the sensors, microcontroller, and Jetson Nano.

The design prioritizes reliability and modularity, allowing for seamless integration of additional components or upgrades in the future [9]. Each component is mounted securely within the electronics cabin, which is designed with custom brackets fabricated from ABS+ and PLA+ materials. This enclosure ensures proper organization, thermal management, and protection against environmental factors, including moisture and vibration. Furthermore, the electronic design incorporates multiple fail-safe mechanisms, such as emergency shutdown protocols and voltage regulators, to protect the system against power surges or component failures. Overall, the electronic subsystem of ORCA exemplifies a carefully engineered framework that integrates cutting-edge technology to meet the demands of autonomous underwater exploration and operation.

2.2.2. Power Supply and Battery

The ORCA system is powered by an Onbo 7200mAh 80C LiPo battery with a nominal voltage of 14.8V. This battery provides approximately 17 minutes of operational time. High-capacity LiPo batteries are widely recognized for their ability to enhance the operational duration of robotic systems [10,11].

2.2.3. Power Distribution Board

To maintain a stable voltage throughout the robot system, the ORCA is equipped with a dedicated power distribution board that accepts an input voltage range of 6V to 30V and provides 14.8V and 5V 12A outputs. The power distribution board provides a stable power supply to all electronic components, including the microcontroller, sensors, and thrusters.

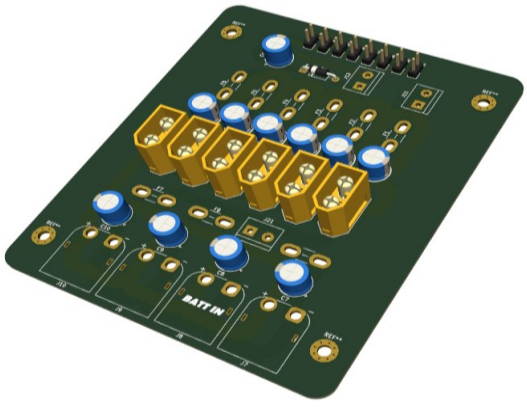


Figure 5. PCB Design of The Power Distribution Board

The Power Distribution Board (PDB) in the ORCA, as depicted in Figure 5, offers several significant advantages. It efficiently distributes power to various components and systems within the robot, preventing power leakage that could damage other components. Additionally, the PDB ensures that the supplied voltage aligns with component requirements, safeguarding against damage from unstable or excessive voltage. A well-designed PDB contributes to overall

system reliability and performance by minimizing voltage drops and power losses. Electrical installation becomes more streamlined and efficient because components can be connected directly to the PDB, eliminating the need for individual wiring. The PDB also facilitates the easy addition or removal of components or systems without modifying the existing power distribution system [12,13].

2.3. Software Design

2.3.1. System Overview

This year's research and development goals consist of four key areas: basic behavior, attitude control, navigation, and target acquisition. In the initial goal, the robot is placed in the starting area and lowered into the pool, where it immediately searches for the first gate. Upon detecting the gate, it moves directly toward it, keeping the gate centered in its field of view. After passing through, the focus shifts to attitude control, where the robot detects obstacles, such as cylinders. If an obstacle is too close, the robot stops to search for the next gate. It performs a 135-degree scan to locate gates 2 or 3, and, if necessary, slides 45 degrees until the cylinder is no longer visible. In the next area, navigation, the robot detects and approaches the subsequent gate. After passing through gates two or three, it transitions to target acquisition, where it searches for a red drum with a green carpet base. The robot positions its gripper on the drum and drops its payload. AUVs can produce extensive amounts of data, including real-time sensor outputs, positional data, and other mission-critical information [14-16].

2.3.2. Computer Vision

Computer vision is a field of computer science focused on processing images, videos, and other real-world visual data, enabling computers to "see" and "understand" objects through the analysis of data captured by cameras or sensors. Applications include facial recognition, object detection, image segmentation, and satellite image processing, with significant advancements in robotics. For example, the paper demonstrates the application of computer vision to an AUV for detecting various objects. For the ORCA's research, computer vision is implemented using a YOLOv8 model, a deep learning architecture known for real-time object detection with high accuracy [17,18]. YOLOv8 detects objects by dividing input images into grids, identifying significant regions, and classifying objects within those regions. It assigns bounding boxes based on prediction scores and object dimensions. To train the model, replicas of gates and obstacles were created, and underwater videos were captured for relevant data. A program captured images at 2 frames per second, which were then annotated manually using the VGG Image Annotator.

YOLOv8 was selected because it offers the best trade-off among accuracy, latency, and computational footprint for resource-constrained scenarios (e.g., Jetson Nano). The lightweight variant (e.g., YOLOv8-n) delivers low inference latency and a small memory footprint while maintaining competitive performance. To meet edge latency targets, the model is configured with a 640×640 input size and deployment optimizations (FP16/INT8, TensorRT, and light pruning aligned with the device profile).

2.3.3. Dataset

For object detection on a robot, a dataset is used to train a YOLOv8 model. The dataset is derived from replicating objects for use in the arena. The dataset is the result of 2fps image capture. The image is captured by a camera mounted on the robot to ensure that the data are relevant to the robot's observations. After that, the camera captures data containing the objects to be detected, which will be annotated manually using the VGG Image Annotator [19]. The

effectiveness of custom datasets in improving object detection accuracy has been well documented, as tailored data directly correlate with the performance of deep learning models.

2.3.4. Training Process

The annotated dataset comprised 2,000 images in total and was split into 1,200 training (60%), 400 validation (20%), and 400 test (20%) samples. Model training used YOLOv8 with pretrained weights initialized from a prior custom dataset. All experiments were conducted in Google Colab using an NVIDIA T4 GPU. During training, YOLOv8 automatically resized images to 640×640 px. The AdamW optimizer was employed with the default learning rate of 0.01, subject to scheduler-based adjustments. Training was configured for 300 epochs, with early stopping applied if overfitting or underfitting indicators were observed. The implementation was based on the PyTorch deep learning framework, with standard YOLOv8 data augmentations enabled and batch sizing chosen to maximize GPU utilization while maintaining stable convergence.

2.3.5. Control System

The control system of ORCA comprises three programming algorithm packages: guidance, navigation, and control, which enable 6 degrees of freedom (6-DoF) motion. The guidance package defines two trajectory types: simultaneous and long trajectories. The simultaneous trajectory enables rapid transitions between poses, thereby forming the basis for various movements. A long trajectory is used over extended distances, with precise waypoints set to adjust speed as the robot approaches them. Computer vision detects the center of bounding boxes for navigation, allowing the robot to adapt thruster power to reach targets or avoid obstacles [20,21]. The navigation package includes an extended Kalman Filter (EKF) and a translational EKF that estimate the robot's pose using data from the Inertial Measurement Unit (IMU). EKF outputs provide feedback for the guidance and control systems, while the translational EKF predicts the robot's condition during movements [22]. Lastly, the control package features a continuous-time PID controller and algorithms for calculating thruster nominal variables. The calculation of the thruster variables considers the AUV dynamics, current pose and velocity, desired trajectory, and thruster constraints [23,24].

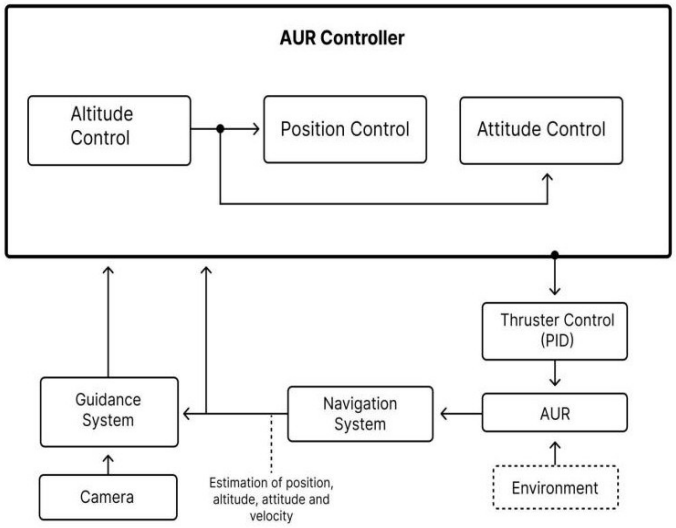


Figure 6. Diagram of Control Algorithm

Figure 6 illustrates the three algorithmic packages of the ORCA: control, navigation, and guidance. The control package comprises altitude, position, and attitude control. Input data for the control package is acquired from the camera via the guidance package, which includes an

https://doi.org/10.31436/iiumej.v27i1.3716

object-detection model. The control package then processes the data before forwarding it to the Proportional-Integral-Derivative (PID) thruster. This ensures the robot's stable movement underwater [25,26]. Additionally, the AUV receives environmental data from sensors and passes it through the guidance package.

3. IMPLEMENTATION

The AUV's mechanical frame has been designed to perform its intended functions effectively. The materials used ensure optimal strength and durability. The carefully designed mechanical frame supports the placement of the camera sensor for optimal viewing angles, and the camera sensor used has a sufficiently high resolution. The electronic components have been carefully selected to ensure reliable performance. Software has been developed to meet AUV requirements, including hardware control and computer vision. Figure 7 demonstrates the AUV's ability to detect gates and obstacles, as indicated by the green and red boxes in the object detection model.

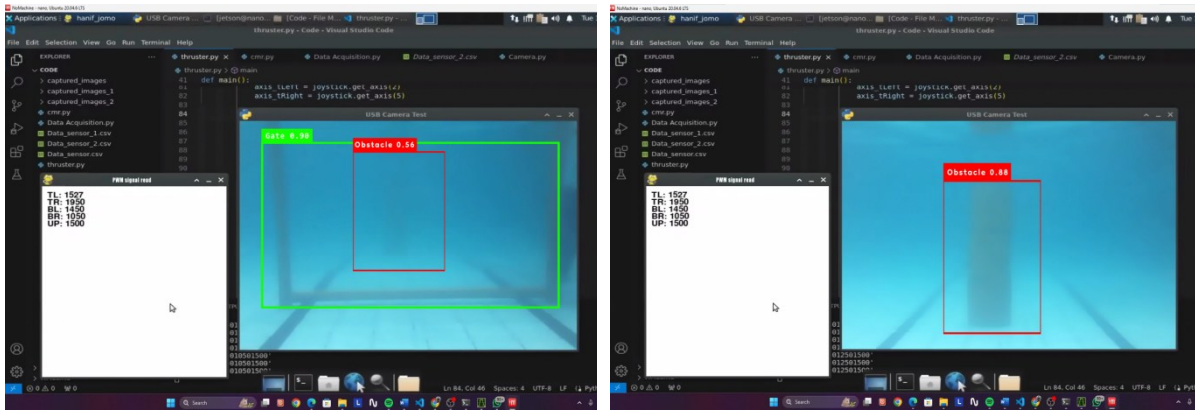


Figure 7. Model Object Detection Detecting Gate and Obstacle (left), Model object detection detecting obstacle (right)

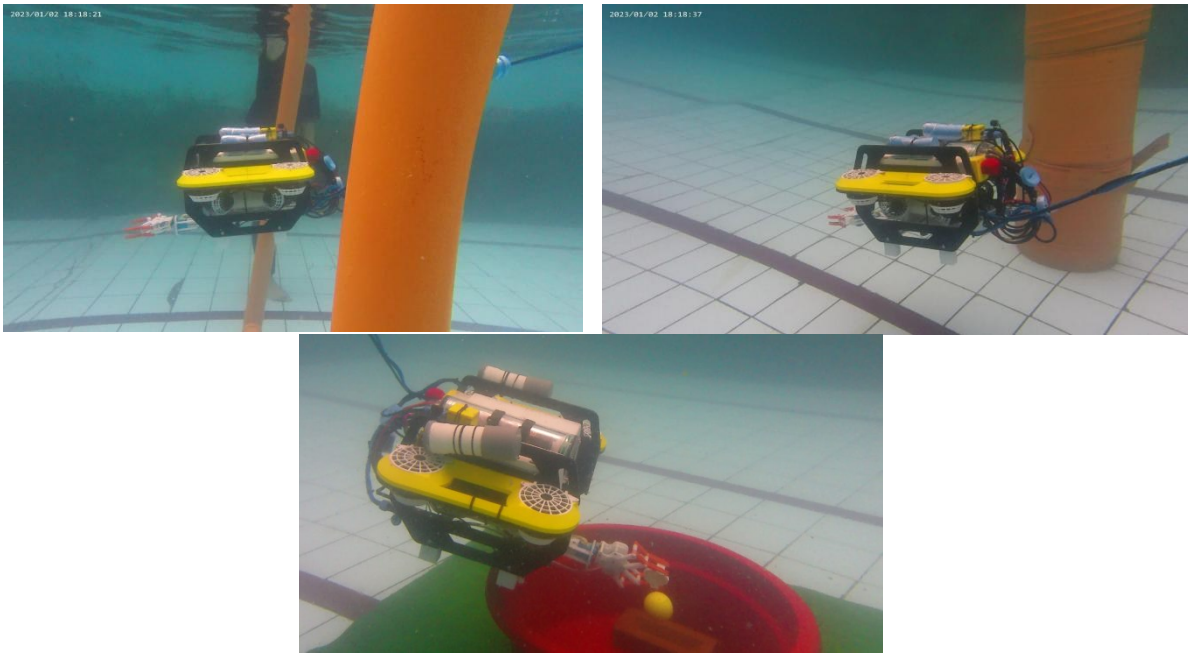


Figure 8. ORCA Moving Through a Gate (left), ORCA Avoiding Obstacle (right), ORCA Dropping the Payload Using a Gripper (bottom)

https://doi.org/10.31436/iiumej.v27i1.3716

Experiments were conducted in a rectangular test pool measuring 25 × 12 m with an approximate depth of 1.8 m. Competition-style targets were placed according to the task scenarios: a rectangular gate for transit and a drum (red/green base) for the payload drop task. The local navigation frame was defined such that the x-axis points forward (surge, positive in the direction of travel), the y-axis is lateral (sway, cross-track right/left), and the z-axis is vertical (depth, negative downward; more negative indicates deeper submersion).

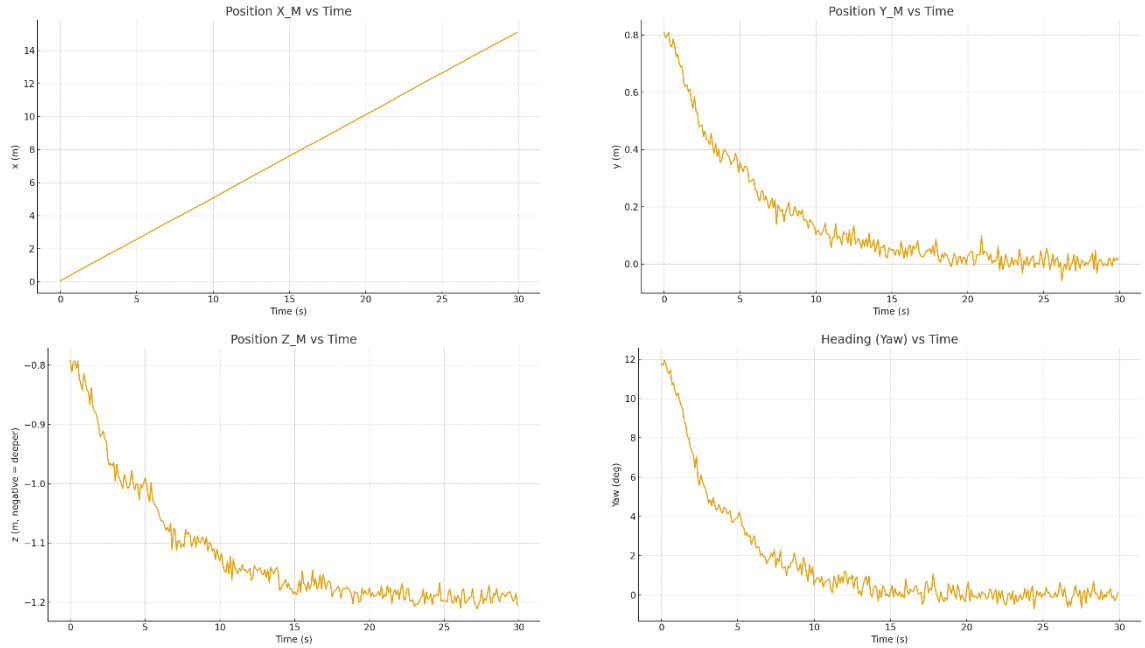


Figure 9. Position and Heading (left-right: $X_m(t)$, $y_m(t)$, $z_m(t)$ *(negative=deeper),* $yaw_deg(t)$)

A fixed depth setpoint of 0.75 m below the water surface was used for all trials, corresponding to $z = -0.75$ m in the local frame. State estimation and telemetry comprised: (i) BNO055 IMU for orientation/heading (reported as yaw_deg), (ii) MS5803-14BA pressure sensor for depth (z_m, negative downward), and (iii) a monocular Logitech C525 camera providing visual odometry and gate-center offset in pixels relative to the image ROI center (proxy for cross-track alignment). All streams were sampled at 10 Hz and time-synchronized on the Jetson Nano using ROS timestamps to ensure consistent fusion and logging. The control inputs recorded were u_surge, u_sway, u_heave, and u_yaw. Tracking-error channels consisted of $e_cross_track_m$, $e_heading_deg$, and e_depth_m. The binary gripper_open flag indicated whether the payload was released (1 = open, 0 = closed).

Three standardized tasks were evaluated, as shown in Figure 8.

1. Position and Heading. The AUV was initialized upstream of the gate with a heading of 0°, then commanded to pass centrally through the aperture. Time series for position (x_m, y_m, z_m), heading (yaw_deg), cross-track/heading/depth errors $(e_cross_track_m, e_heading_deg, e_depth_m)$, and gate-center offset (pixels) indicate that the vehicle converged to the gate centerline while maintaining depth and heading within bounded error. RMSE/MAE/max|error| for these channels are reported in Table 3, with representative trajectories shown in Figure 9.

https://doi.org/10.31436/iiumej.v27i1.3716

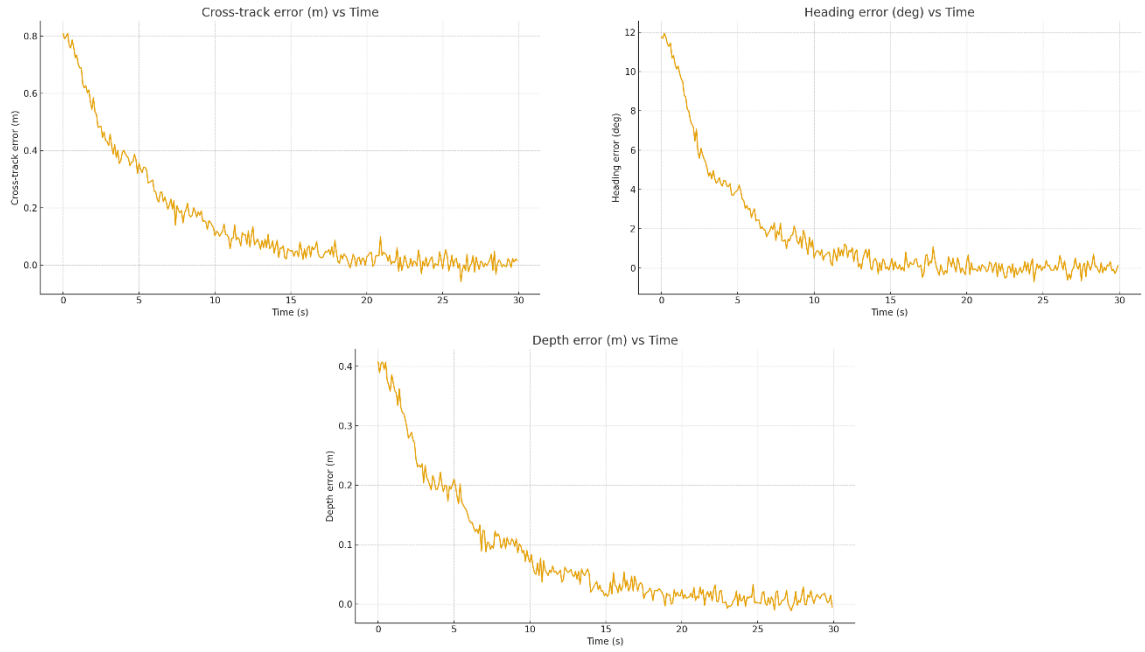


Figure 10. Tracking Errors (left-right: $e_cross_track_m(t)$, $e_heading_deg(t)$, $e_depth_m(t)$)

2. Tracking Errors. The AUV executed a lateral avoidance maneuver around an obstacle and subsequently returned to the nominal path. The recorded u_sway and u_yaw commands show a coordinated sidestep and heading correction, while $e_cross_track_m$ and yaw_deg demonstrate transient deviations followed by realignment to the track. Depth remained regulated via u_heave, with e_depth_m bounded during the maneuver. The corresponding error and command time series are shown in Figure 10, and the aggregate metrics are presented in Table 3.
3. Control and Perception. The vehicle approached the drum target, achieved a stabilized hover at the commanded depth and heading, then actuated the gripper (gripper_open = 1) to release the payload. The final approach sequence presents a reduction in gate/target alignment error (using the same pixel-offset convention for the ROI center), small steady-state e_depth_m, and limited $e_heading_deg$. The gripper timeline confirms actuation at the end of the dwell period. Time-series evidence appears in Figure 11; summary statistics are consolidated in Table 3.

Across tasks, the PID-based guidance navigation control maintained bounded tracking errors and stable closed-loop behavior in the pool environment. The results are substantiated by (i) time-series plots for positions, heading, errors, control commands, perception alignment, and gripper state (see Figures 9-11) and (ii) quantitative metrics (RMSE, MAE, max|error|, and dispersion) in Table 3 derived from the synchronized 10 Hz logs. Note that depth is reported as negative values when diving ($z_m < 0$).

https://doi.org/10.31436/iiumej.v27i1.3716

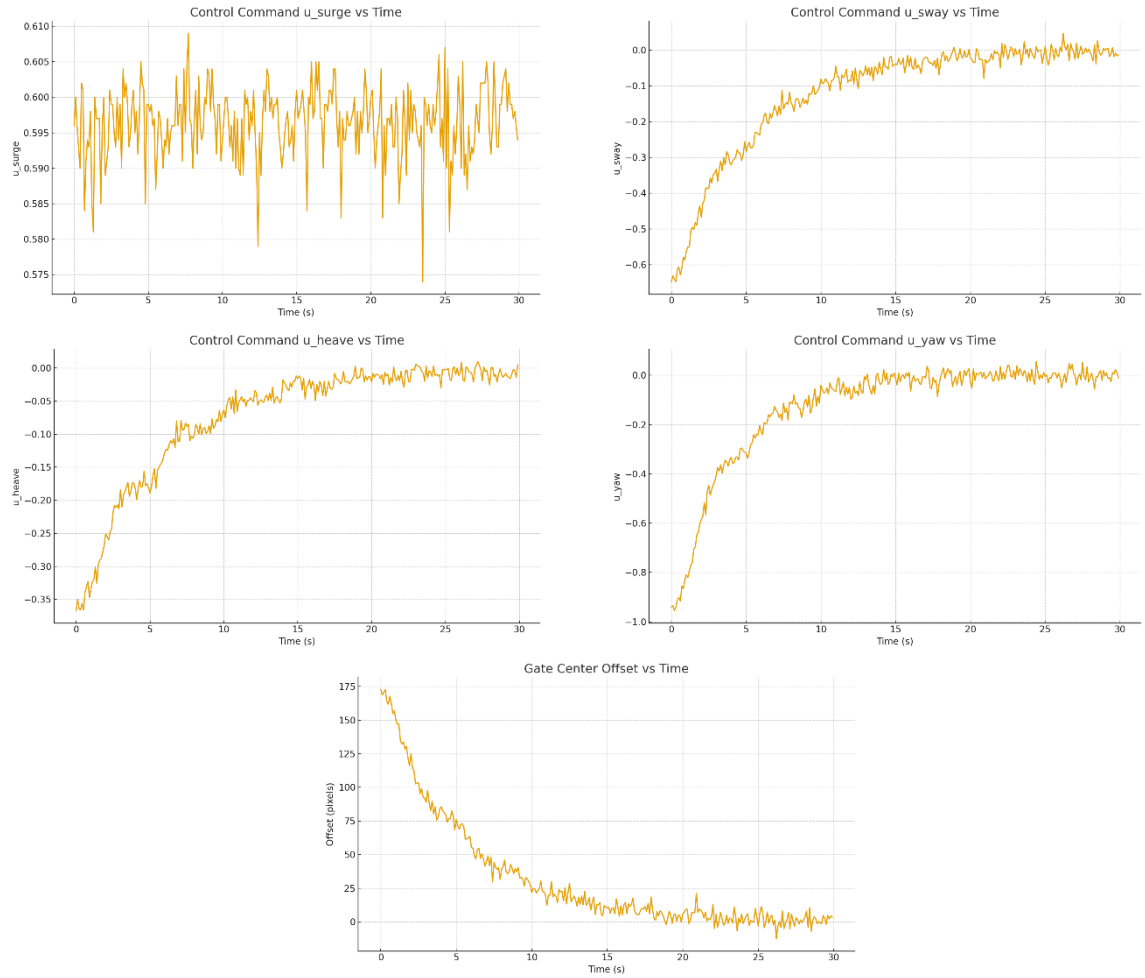


Figure 11. Control and Perception (left-right: $u_surge(t)$, $u_sway(t)$, $u_heave(t)$, $u_yaw(t)$, gate-center offset)

Table 3. Quantitative Tracking Performance Under PID Control

	RMSE	**MAE**	**Max|error|**
Cross-track error	0.248m	0.151m	0.810m
Heading error	3.25°	1.74°	11.94°
Depth error	0.131m	0.082m	1.408m

4. RESULTS AND DISCUSSION

Throughout the testing process, the AUV demonstrated its ability to execute various intended functions, including navigation and object detection. The AUV's performance has been rigorously evaluated and found to meet the established criteria, demonstrating stable and consistent operation. Safety is a paramount concern throughout the testing process. Comprehensive safety systems have been implemented to ensure the safety of both the environment and the AUV during operation. The evidence that supports the latter conclusion includes experimental data from several comprehensive tests that demonstrate the AUV's ability to achieve stable navigation, effective obstacle avoidance, and accurate object detection. The tests also include quantitative metrics, such as detection accuracy, complemented by theoretical analyses of drag reduction and buoyancy stabilization, which validate the design's effectiveness.

https://doi.org/10.31436/iiumej.v27i1.3716

4.1. Functional and Performance Analysis

The ORCA AUV has undergone six comprehensive tests, demonstrating significant improvements with each iteration. In the initial test, the ORCA successfully executed basic underwater robot behaviors, including navigating through a designated gate. However, the movement is observed to be unstable, attributed to unsynchronized actuators and insufficient buoyancy, necessitating increased actuator effort to maintain position. From the second to the fifth tests, the ORCA exhibited a notable enhancement in maneuverability. The addition of buoyancy enabled the actuators to operate more efficiently. Subsequently, the ORCA underwent tests to evaluate its ability to detect objects, including gates and obstacles. All processed data indicated that water clarity substantially affected the ORCA's visibility range, occasionally hindering object detection. The maximum detection range observed during testing is approximately 5 meters. In the fifth and sixth trials, the ORCA is assessed for its capability to detect and drop payloads.

Quantitative evaluations highlighted the ORCA's superior performance in navigation and object detection compared to existing AUVs. During testing, the ORCA achieved an object detection accuracy of 93% within a 5-meter range, outperforming conventional vision systems. Additionally, the AUV demonstrated stable navigation with a mean trajectory error of less than 5% over 10 trials in a controlled pool environment. Theoretical analysis further supports these findings. For example, optimizing drag forces through hydrodynamic design (Equation 1) reduced power consumption during extended operations, while buoyancy adjustments (Equation 2) maintained stability under varying loads. These theoretical advancements directly contributed to improvements in experimental performance.

Key studies, including the work by Xu et al., detail the challenges and advancements in vision-based underwater target detection for AUV subsea exploration. Their research highlights the challenges posed by underwater conditions, including low light, sediment, and biological interference, and demonstrates improvements in detection accuracy through innovative algorithms tailored to AUVs. This supports the claim of the ORCA's superior performance [27].

Additionally, Chen et al. investigate acoustic-optical fusion techniques that enhance underwater object detection, reinforcing the view that combining multiple sensory modalities can yield more reliable detection results. This cross-disciplinary research further substantiates ORCA's object detection accuracy relative to conventional systems [28].

Moreover, the comparative performance of AUV systems has been examined in studies using benchmark datasets and evaluation metrics, underscoring the need for detailed assessments using established metrics such as average precision and mean average precision. Li et al. discuss the necessity of benchmark evaluations in object detection, which are crucial for validating performance claims against established state-of-the-art systems [29]. This context is essential for understanding the rigorous standards ORCA meets in its operational tests.

Lastly, research by Kottapalli et al. identifies attributes required for effective underwater sensing and detection systems, underscoring that ORCA’s enhancements are grounded in scientifically validated performance metrics rather than anecdotal evidence [30]. The synthesis of these reference points establishes a robust framework for ORCA's performance claims, demonstrating its effectiveness within established academic benchmarks for underwater object detection.

https://doi.org/10.31436/iiumej.v27i1.3716

4.2. Safety

Safety is a critical consideration throughout the testing process. Rigorous safety measures were implemented, including adherence to Standard Operating Procedures (SOPs), an emergency button, a circuit breaker, thruster guards, meticulous inspection for sharp edges, and the provision of safety equipment for the robot operator in the pool. During the tests conducted, no incidents required the operator to forcibly activate the emergency button in the pool. The emergency button is also tested while the robot is operational, resulting in the robot's immediate cessation of operations. Furthermore, the robot did not collide with its surroundings or pose a risk of injury to individuals during testing. These measures collectively ensure the safety of the ORCA. As highlighted in a recent study, the implementation of comprehensive safety protocols significantly enhances the reliability and safety of autonomous underwater vehicles during operational testing.

5. MILESTONES, ACHIEVEMENTS, AND FUTURE DIRECTIONS

Table 4. Milestones and Achievements

Milestone	Achievements
The robot can dive and move forward in a straight line	The AUV can dive and maintain a constant depth. The AUV can move forward, although it is not yet stable.
The robot can identify and avoid obstacles.	The AUV can consistently identify obstacles and avoid them by turning left or right and moving forward from the obstacle's initial detection position.
The robot can detect targets.	The AUV can detect targets with high accuracy at close range.
The robot can drop a payload into a target.	The AUV successfully dropped the payload into the target by rotating the gripper.

Table 4 shows that ORCA has achieved several significant milestones, including diving, obstacle detection, target detection, and payload delivery. Future development plans will focus on enhancing the software's strategy and computer vision components, as well as the hardware. These enhancements aim to refine ORCA's performance in stable movement, maneuvering, obstacle avoidance, object identification, and precise payload delivery. Additional considerations include the scalability of ORCA's design and the modularity of its software architecture. Future research could explore the application of similar methodologies to larger or more complex underwater vehicles. Moreover, continued collaboration with marine biologists and oceanographers could enable novel applications of autonomous underwater robotics for biodiversity conservation and deep-sea exploration.

Future developments will focus on the mechanical, electronic, software, and algorithmic subsystems to enable the AUV to perform diverse and flexible tasks. These advancements may include the addition of features and functions for monitoring and sensing, enabling the robot to perform multiple tasks simultaneously. This requires support from mechanics to upgrade the frame and features, electronics to select components and implement circuit schematics, and programmers to develop software for both hardware (control systems) and software (strategies and computer vision) to enable the robot to execute tasks and control the AUV's operation accurately and without errors. Additionally, research will investigate the integration of machine learning techniques to enhance the AUV's decision-making capabilities, enabling it to adapt to varying underwater environments and optimize its target recognition and path planning.

6. CONCLUSIONS

The ORCA Autonomous Underwater Vehicle represents a significant advancement in underwater robotics, reflecting the rigorous research and development that informed its design. Designed for independent operation, ORCA excels in navigation, exploration, and interaction with the underwater environment. Extensive consideration has been given to its design and manufacturing, employing modern techniques to produce a durable vehicle equipped with an impressive array of sensors. Its advanced capabilities in autonomous navigation, obstacle avoidance, object detection, and precise payload handling underscore a commitment to innovation within the field. ORCA is expected to make a valuable contribution to the expanding domain of autonomous underwater technologies.

The key finding of this research is that combining a hydrodynamically optimized design, a pruned YOLOv8 object detection model, and a robust control system significantly enhances the AUV's ability to navigate, detect objects, and execute tasks autonomously in underwater environments. These improvements enable ORCA to perform with greater accuracy and efficiency. The methodologies employed in ORCA development, including pruned YOLOv8-based object detection and carefully optimized control systems, provide significant advantages. These include improved detection accuracy under challenging underwater conditions, reduced computational load for real-time processing, and enhanced stability and efficiency in navigation. By addressing these critical challenges, ORCA contributes a novel and practical solution to the domain of autonomous underwater vehicles.

The results of this research have significant implications for both the field of underwater robotics and the broader community. For the field, ORCA represents a practical advancement in autonomous underwater navigation and task execution, leveraging state-of-the-art computer vision and control systems. For the broader community, this technology offers potential applications in environmental monitoring, underwater exploration, and resource management, contributing to sustainable practices in marine ecosystems.

ACKNOWLEDGEMENT

This journal is funded by the funding of “Hibah Flagship FMIPA UGM 2024” under contract number and is supported by the Gadjah Mada Robotic Team, Universitas Gadjah Mada.

REFERENCES

[1] C. Amaldeep and S. Arunkumar, “Design and Development of an Underwater Robot for Effective Waste Collection,” *2024 Int. Conf. Adv. Mod. Age Technol. Heal. Eng. Sci. AMATHE 2024*, pp. 1–6, 2024, doi: 10.1109/AMATHE61652.2024.10582229.

[2] M. U. Masood, T. Kaaya, and Z. Chen, “Constrained Model Predictive Control of Variable Buoyancy Device,” *IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM*, vol. 2023-June, pp. 936–941, 2023, doi: 10.1109/AIM46323.2023.10196139.

[3] Y. Li, H. Zhang, H. Wang, J. Li, X. Zhai, and J. Xu, “Design of a Variable Buoyancy System for the Hybrid Crawler-Flyer Underwater Vehicle,” *Proc. - 2019 Chinese Autom. Congr. CAC 2019*, pp. 5335–5339, 2019, doi: 10.1109/CAC48633.2019.8997045.

[4] C. Zavislak, A. Keow, Z. Chen, and F. Ghorbel, “AUV Tool Manipulation with Hard and Soft Actuators,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 8553–8560, 2021, doi: 10.1109/LRA.2021.3110537.

[5] Z. Samadikhoshkho, K. Zareinia, and F. Janabi-Sharifi, "A Brief Review on Robotic Grippers Classifications," *2019 IEEE Can. Conf. Electr. Comput. Eng. CCECE 2019*, pp. 1–4, 2019, doi: 10.1109/CCECE.2019.8861780.

[6] F. Wang, H. Liu, L. Cao, F. Xie, G. Song, and A. Song, "A Novel Lightweight Underwater Manipulator Based on ROS2 for Reliable Intervention," *Proc. - 2024 9th Int. Conf. Autom. Control Robot. Eng. CACRE 2024*, pp. 345–349, 2024, doi: 10.1109/CACRE62362.2024.10634872.

[7] Y. Taira, S. Sagara, and M. Oya, "Design of a Motion Controller for an Underwater Vehicle-Manipulator System with a Differently-Controlled Vehicle," *2018 Int. Conf. Inf. Commun. Technol. Robot. ICT-ROBOT 2018*, pp. 1–6, 2018, doi: 10.1109/ICT-ROBOT.2018.8549894.

[8] M. Vasileiou, N. Manos, and E. Kavallieratou, "IURA: An Inexpensive Underwater Robotic Arm for Kalypso ROV," *Int. Conf. Electr. Comput. Commun. Mechatronics Eng. ICECCME 2022*, no. November, pp. 1–6, 2022, doi: 10.1109/ICECCME55909.2022.9988259.

[9] I. Elaff, "Design and Implementation of Computer Controlled Electro-Mechanical Thruster for UUVs," *5th Int. Conf. Circuits, Control. Commun. Comput. I4C 2024*, no. October, pp. 13–16, 2024, doi: 10.1109/I4C62240.2024.10748525.

[10] A. Chiche, C. Lagergren, G. Lindbergh, and I. Stenius, "Sizing the energy system on long-range AUV," *AUV 2018 - 2018 IEEE/OES Auton. Underw. Veh. Work. Proc.*, 2018, doi: 10.1109/AUV.2018.8729812.

[11] A. Harakare *et al.*, "Design of Battery Management System for an Autonomous Underwater Vehicle," *Ocean. Conf. Rec.*, pp. 1–6, 2022, doi: 10.1109/OCEANSChennai45887.2022.9775282.

[12] S. Chava, S. Ranjan, S. Dudani, and P. Chittora, "Electrical and Electronic Architecture of an ROV," *2023 4th Int. Conf. Emerg. Technol. INCET 2023*, pp. 1–5, 2023, doi: 10.1109/INCET57972.2023.10170661.

[13] H. N. Le and E. Tedeschi, "Comparative Evaluation of AC and DC Power Distribution Systems for Underwater Vehicles Based on Multiobjective Optimization Techniques," *IEEE Trans. Power Deliv.*, vol. 36, no. 6, pp. 3456–3465, 2021, doi: 10.1109/TPWRD.2020.3043054.

[14] E. Gallimore, R. Stokey, and E. Terrill, "Robot Operating System (ROS) on the REMUS AUV using RECON," *AUV 2018 - 2018 IEEE/OES Auton. Underw. Veh. Work. Proc.*, pp. 1–6, 2018, doi: 10.1109/AUV.2018.8729755.

[15] S. Buadu, I. Schjolberg, and T. Mo-Bjorkelund, "Mission planner for multiple AUVs: Verification procedures combining simulations and experiments," *AUV 2018 - 2018 IEEE/OES Auton. Underw. Veh. Work. Proc.*, pp. 1–6, 2018, doi: 10.1109/AUV.2018.8729793.

[16] I. Glushko, E. Olenew, and M. Komar, "Software Control Architecture for the BOSS Manta," *2018 IEEE/OES Auton. Underw. Veh. Work.*, pp. 1–5, 2018.

[17] Y. Zhu, H. Yin, and M. Hu, "An Improved Underwater Target Detection Algorithm Based on YOLOv8," *Proc. - 2023 2nd Int. Conf. Artif. Intell. Human-Computer Interact. Robot. AIHCIR 2023*, pp. 521–525, 2023, doi: 10.1109/AIHCIR61661.2023.00092.

[18] Z. Zhao, J. Han, Z. Ma, S. Zou, Y. Liu, and G. Li, "A Lightweight Underwater Object Detection Algorithm with Adaptive Image Enhancement Based on YOLOv8," *Chinese Control Conf. CCC*, pp. 7830–7835, 2024, doi: 10.23919/CCC63176.2024.10661480.

[19] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," *MM 2019 - Proc. 27th ACM Int. Conf. Multimed.*, pp. 2276–2279, 2019, doi: 10.1145/3343031.3350535.

[20] H. Zheng, Y. Sun, G. Zhang, L. Zhang, and W. Zhang, "Research on Real Time Obstacle Avoidance Method for AUV Based on Combination of ViT-DPFN and MPC," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–15, 2024, doi: 10.1109/TIM.2023.3348890.

[21] R. Simoni, P. R. Rodriguez, P. Cieslak, and D. Youakim, "A novel approach to obstacle avoidance for an I-AUV," *AUV 2018 - 2018 IEEE/OES Auton. Underw. Veh. Work. Proc.*, 2018, doi: 10.1109/AUV.2018.8729721.

[22] X. Zhang, X. Mu, H. Liu, B. He, and T. Yan, "Application of Modified EKF Based on Intelligent Data Fusion in AUV Navigation," *2019 IEEE Int. Underw. Technol. Symp. UT 2019 - Proc.*, 2019, doi: 10.1109/UT.2019.8734414.

[23] F. Vladimir and Y. Dmitry, "Control system with reference model for spatial motion of a cargo AUV," *2020 Int. Conf. Control. Autom. Diagnosis, ICCAD 2020 - Proc.*, pp. 1–6, 2020, doi: 10.1109/ICCAD49821.2020.9260555.

[24] Z. Wang *et al.*, "Design and Experiments on AUV Control System based on ROS," *Chinese Control Conf. CCC*, pp. 4363–4368, 2024, doi: 10.23919/CCC63176.2024.10662448.

[25] [25] Q. Yin *et al.*, "Fuzzy PID Motion Control Based on Extended State Observer for AUV," *2019 IEEE Int. Underw. Technol. Symp. UT 2019 - Proc.*, no. v, 2019, doi: 10.1109/UT.2019.8734374.

[26] S. K. Valluru, K. Sehgal, and H. Thareja, "Evaluation of moth-flame optimization, genetic and simulated annealing tuned PID controller for steering control of autonomous underwater vehicle," *2021 IEEE Int. IOT, Electron. Mechatronics Conf. IEMTRONICS 2021 - Proc.*, 2021, doi: 10.1109/IEMTRONICS52119.2021.9422632.

[27] G. Xu, D. Zhou, L. Yuan, W. Guo, Z. Huang, & Y. Zhang, "Vision-based underwater target real-time detection for autonomous underwater vehicle subsea exploration", Frontiers in Marine Science, vol. 10, 2023. https://doi.org/10.3389/fmars.2023.1112310

[28] H. Chen, Z. Wang, H. Qin, & X. Mu, "Uamfdet: acoustic-optical fusion for underwater multi-modal object detection", Journal of Field Robotics, vol. 42, no. 4, p. 970-983, 2024. https://doi.org/10.1002/rob.22432

[29] K. Li, G. Wan, G. Cheng, L. Meng, & J. Han, "Object detection in optical remote sensing images: a survey and a new benchmark", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 159, p. 296-307, 2020. https://doi.org/10.1016/j.isprsjprs.2019.11.023

[30] A. Kottapalli, M. Asadnia, J. Miao, G. Barbastathis, & M. Triantafyllou, "A flexible liquid crystal polymer mems pressure sensor array for fish-like underwater sensing", Smart Materials and Structures, vol. 21, no. 11, p. 115030, 2012. https://doi.org/10.1088/0964-1726/21/11/115030

https://doi.org/10.31436/iiumej.v27i1.3749

FSM-Based MBIST Controller Implementation using the March AZ2 Test Algorithm

AIMAN ZAKWAN JIDIN[1,2*], MOHD SYAFIQ MISPAN[1,2], RAZAIDI HUSSIN[3], AHMAD RIDHWAN MD ZIN[1], TENGKU FARAH NURHIDAYU TUAN AB RAZAK[1]

[1]*Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Malaysia*
[2]*Center for Telecommunication Research and Innovation, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Malaysia*
[3]*Faculty of Electronics Engineering and Technology, Universiti Malaysia Perlis, 02600 Arau, Malaysia*

**Corresponding author: aimanzakwan@utem.edu.my*

(Received: 23 May 2025; Accepted: 4 December 2025; Published online: 12 January 2026)

ABSTRACT: An efficient on-chip memory testing requires the use of a Memory Built-In Self-Test (MBIST) that applies a low-complexity test algorithm that offers excellent fault coverage, to ensure high test quality at a minimal cost. The March AZ2 algorithm, with 14N complexity, was previously established to balance the fault coverage and test complexity. It was previously implemented in an MBIST controller through an automatic generation process using an Electronic Design Automation (EDA) tool, which limits its customizability and optimization potential. This paper presents the implementation of an MBIST controller that employs the March AZ2 test algorithm and is based on a Finite-State Machine (FSM) architecture. The developed MBIST controller was implemented on a Field-Programmable Gate Array (FPGA) to validate its functionality and fault coverage. A comparison with an equivalent MBIST controller automatically generated using Siemens EDA Tessent MemoryBIST tool demonstrates that the proposed FSM-Based MBIST controller achieves 84% lower circuit area and 32% power consumption, while offering similar fault coverage.

ABSTRAK: Pengujian memori atas cip yang cekap memerlukan penggunaan Ujian-Kendiri Terbina-Dalam Memori (MBIST) menggunakan algoritma ujian berkompleksiti rendah dan menawarkan liputan kesalahan yang tinggi, bagi memastikan kualiti ujian baik pada kos minimum. Algoritma March AZ2, dengan kekompleksan 14N, telah diperkenalkan sebelum ini bagi mengimbangi antara liputan kesalahan dan kompleksiti ujian. Ia telah dilaksanakan dalam pengawal MBIST melalui proses penjanaan automatik menggunakan Automasi Reka Bentuk Elektronik (EDA), yang menghadkan tahap kebolehsuaian dan potensi pengoptimuman. Kajian ini membentangkan pelaksanaan pengawal MBIST yang menggunakan algoritma ujian March AZ2, berasaskan seni bina Mesin Keadaan-Terhingga (FSM). Pengawal MBIST yang dibangunkan telah dilaksanakan pada Litar Terpadu-Serba Guna (FPGA) bagi mengesahkan fungsi dan liputan kesalahan. Perbandingan dengan pengawal MBIST yang dijana secara automatik menggunakan Siemens EDA Tessent MemoryBIST membuktikan bahawa pengawal MBIST berasaskan FSM yang dicadangkan mengurangkan keluasan litar sebanyak 84% dan penggunaan kuasa 32% lebih rendah, sambil mengekalkan liputan kesalahan yang sama.

KEYWORDS: *MBIST, March Test Algorithm, FPGA, FSM, and Memory Testing*

https://doi.org/10.31436/iiumej.v27i1.3749

1. INTRODUCTION

Memory Built-In Self-Test (MBIST) has become an indispensable technique for on-chip memory testing, owing to its ability to independently generate the test inputs, apply them to the memory under test, and verify the test responses at the output [1]. Consequently, it is preferred in the industry to minimize both test and overall chip production costs, as reliance on high-performance and expensive external testers has been lessened [2], [3], [4]. Its utilization has become more crucial than ever, since modern chips are predominantly occupied by memories, which can take up as much as 94% of the chip's total area [5], [6], [7]. Thus, the overall chip quality depends significantly on the quality of the memories within that particular chip [6], [8].

The efficiency of an MBIST operation depends on the test algorithm employed, which specifies the sequence of test operations performed on the memory under test. Different test algorithms offer varying levels of complexity and fault coverage. Among them, March-series test algorithms are often preferred owing to their design simplicity and linear complexity [9], [10]. A March test algorithm defines a sequence of test operations, consisting of writing an x logic (wx) to memory cells or reading from memory cells (rx), where $x \in \{0, 1\}$. These test operations may be performed in either ascending address order (from the first to the last memory cell, denoted by ⇑), or descending address order (denoted by ⇓) [11]. In some cases, the direction of the test operations is not important (denoted by ⇕).

Table 1. The complexities and fault coverages of several March test algorithms.

Test Algorithm	Complexity	Fault Coverage						
		Stuck-At Fault (SAF)	Transition Fault (TF)	Deceptive Read Desctructive Fault (DRDF)	Write Destructive Fault (WDF)	TF Coupling Fault (CFtr)	DRDF Coupling Fault (CFdrd)	WDF Coupling Fault (CFwd)
March C- [12]	10N	100%	100%	0%	0%	100%	0%	0%
March CL [13]	12N	100%	100%	50%	0%	100%	25%	0%
March AZ1 [14]	13N	100%	100%	100%	100%	62.5%	75%	75%
March LR [15]	14N	100%	100%	0%	0%	100%	0%	0%
March SR [16]	14N	100%	100%	100%	0%	100%	50%	0%
March C+ [17]	14N	100%	100%	100%	0%	100%	100%	0%
March XR [18]	14N	100%	100%	100%	100%	62.5%	50%	50%
March AZ2 [14]	14N	100%	100%	100%	100%	75%	75%	75%
March-ee [19]	18N	100%	100%	100%	50%	100%	100%	25%
March MSS [20]	18N	100%	100%	100%	100%	100%	100%	100%
March CS [21]	20N	100%	100%	100%	100%	100%	100%	100%
March SS [22]	22N	100%	100%	100%	100%	100%	100%	100%

Table 1 lists several March test algorithms along with their complexities and fault coverages. The complexity is expressed as O (N), where N denotes the number of cells in the memory under test. Additionally, the coverage of Incorrect Read Fault (IRF) and Read Destructive Fault (RDF), which are not listed in Table 1, is similar to SAF coverage because

their detection requirements are identical [20]. SAF, TF, RDF, IRF, DRDF, and WDF are considered Single-Cell Faults (SCFs), in which fault sensitization and detection occur only within the affected victim cell. Meanwhile, CFtr, CFdrd, and CFwd are coupling faults (CFs), also known as Double-Cell Faults (DCFs), because the fault detected in a victim cell is caused by the state or operation of its aggressor cell. Each SCF comprises 2 Fault Primitives (FPs), as faults can occur when the affected cell is in either a low or a high state. Thus, an SCF coverage is determined by dividing the number of detectable FPs by 2. On the other hand, each DCF comprises 8 FPs: a fault in a victim cell may occur when its associated aggressor cell is in either a low or high state, and the aggressor cell can be located either before or after the victim cell within the memory. Therefore, each DCF coverage is calculated by dividing the number of detectable FPs by 8.

Based on Table 1, a minimum complexity of 18N, such as in the case of the March MSS algorithm, is required to achieve 100% coverage of all unlinked static faults in a Random Access Memory (RAM) [20]. To reduce test duration and cost, lower-complexity test algorithms can be used. However, as shown in Table 1, many of these algorithms provide limited or no coverage of several fault types, such as DRDF, WDF, and their associated CFs. These faults are particularly relevant to modern memory technologies, which are manufactured using miniaturized process technologies, and thus, their presence cannot be ignored [6]. The March AZ2 algorithm was introduced to balance test duration and quality. It provides excellent coverage of all targeted unlinked static faults in a RAM with only 14N complexity [14]. Compared with other test algorithms, it offers the best coverage of unlinked static faults while maintaining a complexity below 18N.

A test algorithm can be implemented in an MBIST controller in several ways. The work in [23] developed a microcode-based MBIST controller architecture. In this work, instructions representing the test sequence are stored in a dedicated register in binary format and deployed as the test patterns to be loaded into the controller. However, because it requires an instruction counter and a register to store instructions, it generally occupies a large chip area. Meanwhile, several works have successfully developed the Finite-State Machine (FSM)-based MBIST controllers, providing better test time and area overhead [3], [24]. This architecture comprises several states to represent each test operation, as well as the test initialization and termination.

Additionally, the MBIST controller generation process can be automated using an Electronic Design Automation (EDA) tool. The previous research in [14] demonstrated that the March AZ2 algorithm was successfully applied in an MBIST controller that was generated automatically using Siemens EDA Tessent MemoryBIST software as the EDA tool [25]. The generation process was carried out simply by developing a core description file that describes its test sequence in a format recognizable by the tool used [26]. Despite the process's simplicity and short design time, it imposes limitations on design customizability and optimization potential, as the MBIST controller was generated using the EDA tool's predefined resources and generation algorithm. Consequently, designers have limited options to optimize the circuit's area and speed.

Therefore, this paper presents a new MBIST controller designed using an FSM-based architecture to represent the March AZ2 algorithm's test sequence in hardware. The proposed FSM-based MBIST controller is then incorporated with a 1-kB RAM as the memory under test. The circuit was then implemented on an Intel MAX10 DE10-Lite FPGA Development Board, and a series of tests was conducted to validate its functionality and fault coverage. The main contribution of this research paper is the development of a user-defined FSM-based MBIST controller that operates correctly while offering reduced circuit area and greater

https://doi.org/10.31436/iiumej.v27i1.3749

flexibility and customizability for optimization, compared with an equivalent controller automatically generated by EDA tools.

2. METHODOLOGY

2.1. The March AZ2 algorithm description

⇓(w0);	⇓(w0, r0);	⇑(r0, w1, w1, r1);	⇑(r1, w0);	⇓(r0, w1, w1, r1);	⇓(r1)
ME_0	ME_1	ME_2	ME_3	ME_4	ME_5

Figure 1. The March AZ2 algorithm test sequence.

Figure 1 presents the test operation sequence defined by the March AZ2 algorithm, which is divided into 5 test elements (denoted as ME_0 to ME_5) [14]. In ME_0, all memory cells are initialized to logic 0 in descending order. Next, in ME_1, each memory cell is written to 0 and then read (expecting a 0 at the output), starting from the last to the first cell. After that, each cell is read (expecting a 0 at the output), written to 1 twice, and reread (expecting a 1 at the output) in ME_2, starting from the first to the last memory cell. The test sequence continues in ME_3, where each cell is read (expecting a 1 at the output) and then written to 0 in ascending address order. In ME_4, the same test operation sequence as in ME_2 is carried out, but in descending address order. Finally, all cells are read (expecting a 1 at the output) in ME_5 in descending address order. There are 14 read or write operations in this test algorithm that must be performed on N memory cells, thereby explaining its 14N complexity.

As shown in Table 1, the March AZ2 test algorithm achieves full coverage of SAF, TF, IRF, RDF, DRDF, and WDF fault models, while providing 75% coverage of CFtr, CFdrd, and CFwd. It fails to detect 2 out of 8 FPs for CFtr due to the absence of test operations that induce a high-to-low cell-state transition in descending address order, e.g., ⇳(…, w1); ⇓(r1, w0); ⇓(r0, …) or ⇳(…, w1, r1); ⇓(w0, r0, …). Similarly, it misses 2 out of 8 FPs for CFdrd, as its test sequence lacks two consecutive read operations from a low-state cell in ascending address order, e.g., ⇑(…, r0, r0) or ⇑(…, r0); ⇳(r0, …). For CFwd, the March AZ2 algorithm is unable to detect 2 out of 8 FPs because its test sequence does not include any non-transition write-to-0 operation in ascending address order, such as ⇳(…, w0); ⇑(w0, r0, …) or ⇑(…, w0, w0, r0).

2.2. MBIST Controller FSM development

Figure 2 depicts the 15-state FSM used to represent the March AZ2 algorithm in hardware. It starts in IDLE, the default state, which waits for a high-state en_mbist signal to initiate MBIST. The FSM then transitions to the LOAD state, where the address counter is initialized to the maximum address of the memory under test. It then commences the test in the M0_W0_D state, with the go flag asserted, indicating that the test is in progress. In this state, all memory cells are written to 0 in descending address order, as described in ME_0 of the March AZ2 algorithm's test sequence.

Once all memory cells are written to 0, indicated by a high-state *fin_cnt* input, the FSM proceeds to the **M1_W0_D** and **M1_R0_D** states to perform the w0 and r0 operations, respectively, on all cells in descending address order, as defined in ME_1. It remains in these two states until a high-state *fin_cnt* is received, indicating that the operations have been completed to all memory cells. Next, to perform the test operation sequence defined in ME_2, the FSM proceeds to the **M2_R0_U** state for one clock cycle to perform the r0 operation, then

https://doi.org/10.31436/iiumej.v27i1.3749

to the **M2_W1W1_U** state for two clock cycles to perform the w1w1 operation, and finally to the **M2_R1_U** state to carry out the r1 operation. It loops through these states until the last memory cell is reached, as indicated by a high-state *fin_cnt* input. Specifically, in the **M2_W1W1_U** state, the *en_cnt_twice* flag is asserted to count for two clock cycles (indicated by a high *twice_det* signal) to ensure that the w1 operation is performed twice in this state.

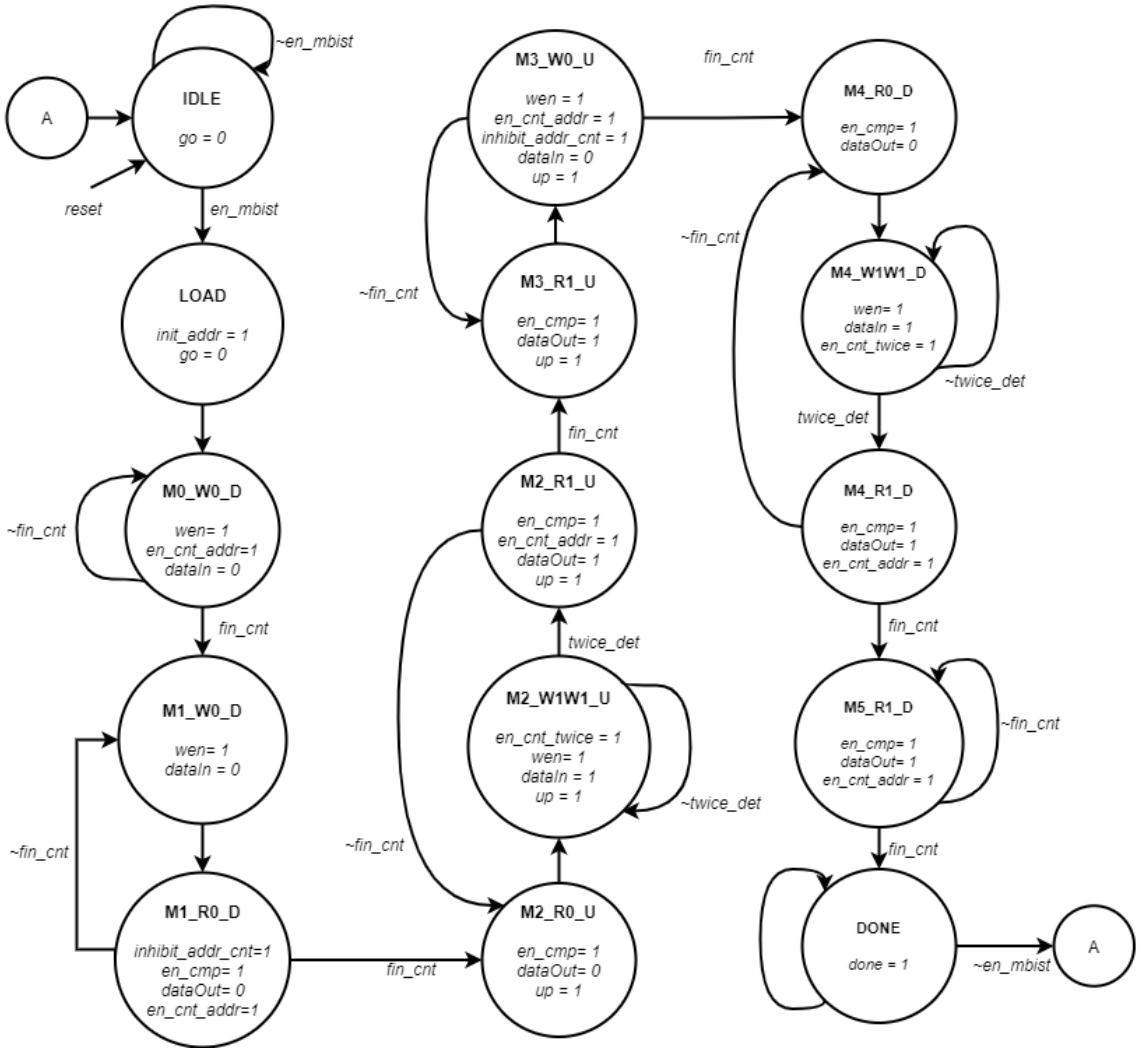


Figure 2. The state diagram for the FSM-based MBIST controller employs the March AZ2 as the test algorithm.

Next, the FSM proceeds to the **M3_R1_U** and **M3_W0_U** states to perform the r1 and w0 operations, respectively, as defined in ME_3. It remains in the same state loop until a high-state *fin-cnt* input is received, indicating that the last memory cell has been reached. After that, to perform the test operation sequence defined in ME_4, the FSM proceeds to the **M4_R0_D** state for one clock cycle to perform the r0 operation, the **M4_W1W1_D** state for two clock cycles to perform the w1w1 operation, and the **M4_R1_D** state to carry out the r1 operation. It remains in the same state loop until the first memory cell is reached, as indicated by a high *fin_cnt* input. Finally, the r1 operation, as defined in ME_5, is executed by proceeding to the **M5_R1_D** state. It remains in the same state until the first memory cell is reached. Once complete, the FSM transitions to the DONE state to terminate the MBIST operation, signified by the assertion of the *done_mbist* flag. It remains in this state until the *en_mbist* signal is low, to avoid unnecessary repetition of the MBIST operation.

https://doi.org/10.31436/iiumej.v27i1.3749

To enable the writing operation, the *wen* flag is asserted during the **M0_W0_D**, **M1_W0_D**, **M2_W1W1_U**, **M3_W0_U**, and **M4_W1W1_D** states, together with the appropriate *dataIn* value (0 for w0, 1 for w1). Meanwhile, in all states that require read operations, the en_cmp flag is asserted to enable comparison between the data read from the memory cell and the expected data, which is determined by the value of the *dataOut* signal. Additionally, in the final state for each test element (**M0_W0_D** for ME_0, **M1_R0_D** for ME_1, **M2_R1_U** for ME_2, **M3_W0_U** for ME_3, **M4_R1_D** for ME_4, and **M5_R1_D** for ME_5), the *en_cnt_addr* flag is asserted to enable counting up (signaled by a high-state *up* flag) or counting down of the address counter, so that all the test operations defined in each test element can be performed on all memory cells. A high-state *fin_cnt* is produced by the address counter when it finishes counting up (when the *up* flag is high) or down, allowing the MBIST operation to proceed to the next test element.

Furthermore, the *inhibit_addr_cnt* flag is set to high during the **M1_R0_D** and **M3_W0_U** states since their subsequent states (**M2_R0_U** and **M4_R0_D**, respectively) have the opposite address direction. By setting this flag high, the address counter maintains its final count value (the last memory cell's address when *up* is high, otherwise 0) upon completing the counting process. Hence, at the beginning of the next state, the address counter can directly start counting down from the address of the last memory cell (in descending address order) or counting up from 0 (in ascending address order).

2.3. Incorporating the proposed MBIST Controller with the memory under test

The functional block diagram presented in Figure 3 shows the integration of the proposed MBIST controller with the memory under test (MUT). It is connected to other submodules, including the data generator, the address counter, and the two-clock counter. The data generator produces 8-bit data to be written to the MUT (write_data) when the *wen* signal is high, or the expected data to be read from the MUT (*exp_data*) based on the received *dataOut* value.

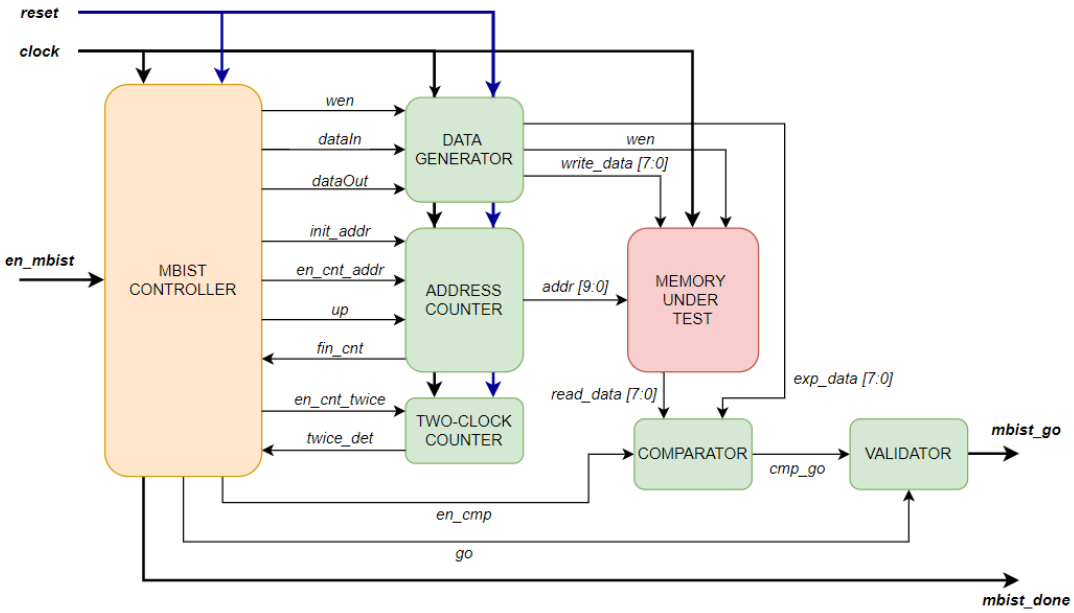


Figure 3. The block diagram shows the integration of the proposed MBIST Controller with the memory under test.

Meanwhile, the address counter initializes the address (*addr*) value to the MUT's maximum address (1023, since a 1-kB RAM is used as the MUT) upon receiving a high

https://doi.org/10.31436/iiumej.v27i1.3749

init_addr signal from the MBIST controller. When a high en_cnt_addr signal is received, the counter counts down (when the up signal is low) or up (when the up signal is high). Upon reaching the minimum memory address (when up equals 0) or the maximum address (when up equals 1), it asserts the fin_cnt signal, which is sent to the MBIST controller to indicate the end of address counting for a particular test element. The generated *addr* is then used to access the MUT's cells. Additionally, the two-clock counter is used to count the number of times the same test operation is executed over two consecutive clock cycles. It is enabled only during the **M2_W1W1_U** and **M4_W1W1_D** states, in which the w1 operation must be performed twice consecutively.

When *wen* is high, the generated *write_data* is written to the MUT cells, whose locations are accessed using *addr*. In contrast, during the read process (when *wen* is low), the data read from the MUT cells (*read_data*) are compared with the expected data (exp_data) generated by the data generator in a comparator when the *en_cmp* signal is high. The comparator produces a high-state *cmp_go* signal as long as there is no error or discrepancy between *read_data* and *exp_data*. Finally, the test validator performs a logical AND operation between the *cmp_go* and *go* signals received from the MBIST controller to produce the *mbist_go* signal at the output. When no errors are detected during MBIST operation, mbist_go should remain in the high state until the end of the test, indicated by a high *mbist_done* output.

2.4. The proposed MBIST Controller validation

The functionality and fault coverage of the proposed MBIST controller were evaluated using tests on the Intel Max10 DE-10 Lite FPGA Development Board, as shown in Figure 4. The proposed circuit, shown in Figure 3, was implemented on the target FPGA using Intel Quartus Prime Design software. The built-in 50 MHz clock oscillator served as the clock source, and an onboard pushbutton was used as the *reset* input. Additionally, a slide switch was used as the *en_mbist* input to start and stop the MBIST operation. Furthermore, a SignalTap logic analyzer core was integrated with the implemented circuit in the FPGA to observe the test outputs.

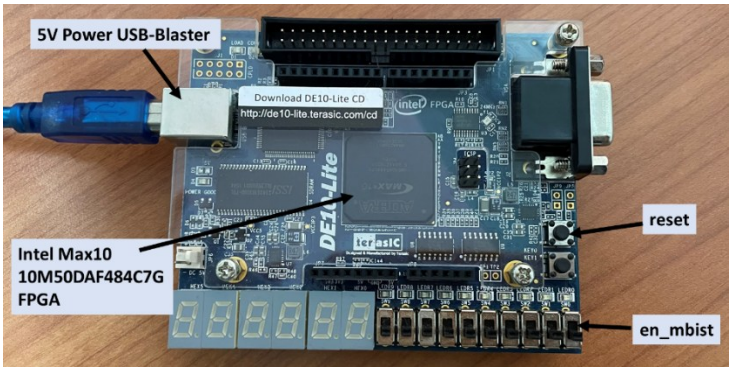


Figure 4. The configuration of the proposed MBIST controller in the Intel Max 10 DE10-Lite FPGA Development Board.

The first test was conducted by running the MBIST operation on a fault-free memory model as the MUT to evaluate its functionality. A high *mbist_go* is expected at the output when the *mbist_done* is asserted, indicating no mismatch between the expected data and the data read from the memory. Additionally, the overall test time was measured from the start (when mbist_go is asserted) to the end (when *mbist_done* is asserted). A test time of 286,720 ns is anticipated, as the proposed MBIST controller uses the March AZ2 test algorithm with 14N complexity (N = 1024 for a 1-kB RAM), with a clock signal period of 20 ns.

https://doi.org/10.31436/iiumej.v27i1.3749

Next, the second test was conducted by running the MBIST operation on a fault-inserted memory model as the MUT to evaluate fault coverage. At this stage, the memory model characteristic was purposely modified to replicate the occurrence of each targeted fault, with all affected victim cells and their aggressor cells randomly chosen [27]. In this case, errors are expected to be detected, and hence, *mbist_go* is likely to be in the low state when the *mbist_done* is asserted. Additionally, 36 fault-detection flags were incorporated to indicate the detectability of the corresponding faults during testing. Each flag is asserted when its corresponding fault is detected. Therefore, the fault coverage can be derived at the end of the test by computing the number of high-bit detection flags. The process of creating and setting the bit value of each detection flag during the MBIST operation is depicted in Figure 5.

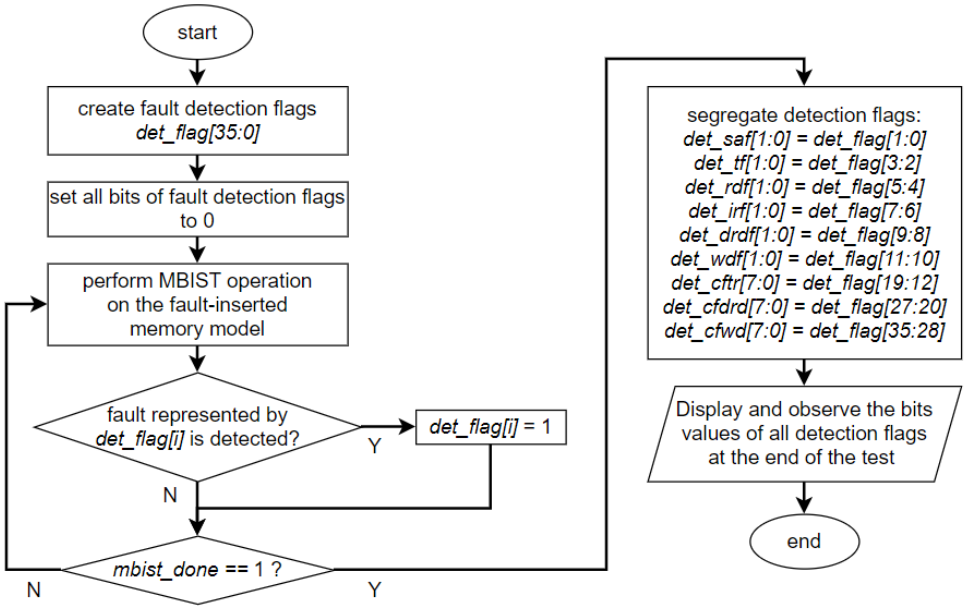


Figure 5. The process flow of fault detection analysis during the MBIST operation on the fault-inserted memory model.

3. RESULTS AND DISCUSSIONS

Figure 6 presents the output observed in the integrated SignalTap logic analyzer core from the MBIST operation performed on the fault-free MUT. No discrepancies were found between the data read from the MUT and the expected data, as the *mbist_go* output stays high until the end of the test, signaled by a high *mbist_done* signal. Additionally, the *cycle_count* signal indicates that the test lasted 14,336 *clock* cycles from start to finish. Given a clock period of 20 ns, the measured test time is 286720 ns, which matches the expected test duration for an MBIST operation applying the March AZ2 test algorithm. Hence, the functionality of the proposed MBIST controller is validated, as it successfully generates the correct test operation sequence.

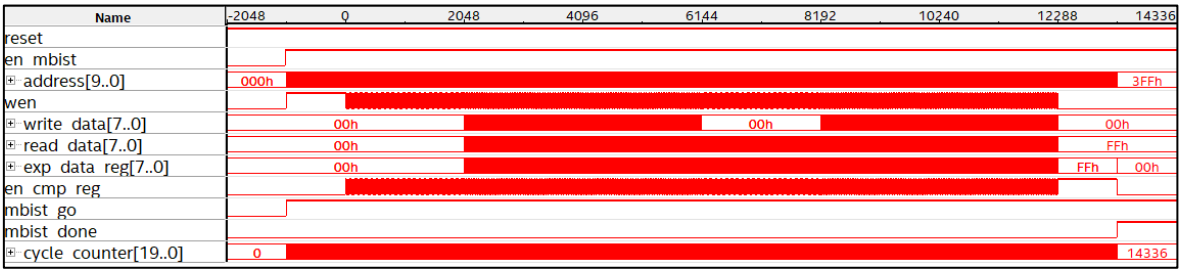


Figure 6. Observation in the SignalTap logic analyzer of the output produced from the experimental test on the fault-free MUT.

https://doi.org/10.31436/iiumej.v27i1.3749

Meanwhile, Figure 7 shows the MBIST output for the fault-inserted MUT, as observed using the integrated SignalTap logic analyzer. In this test, *mbist_go* is deasserted, as expected, when the first fault in the MUT is detected, and it remains at the low level until the end of the test. The values of all fault detection flags are recorded at the end of the test and are presented in Table 2. The fault coverage of the proposed FSM-based MBIST controller implemented on the FPGA device is derived and confirmed to match its expected coverage: SCF, TF, RDF, IRF, DRDF, and WDF are fully detected, whereas CFtr, CFdrd, and CFwd are covered at 75%. Therefore, the conducted test on the FPGA board successfully validates its fault coverage.

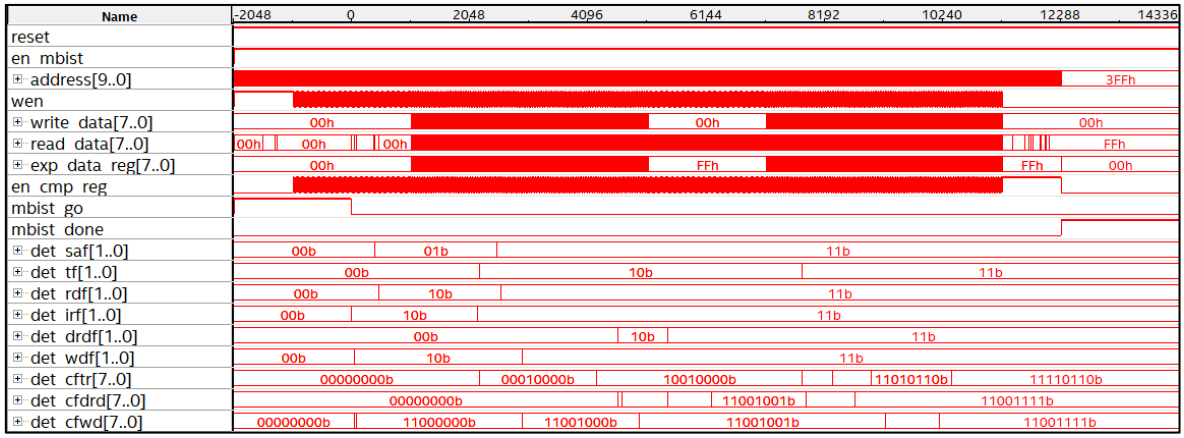


Figure 7. Observation in the SignalTap logic analyzer of the output produced from the experimental test on the fault-inserted MUT.

Table 2. Observed fault detection flag values in the test and derived fault coverages.

Detection Flag	Corresponding Fault	Observed Value	Derived Fault Coverage	Expected Fault Coverage [14]
det_saf	SAF	11_2	2/2 (100%)	2/2 (100%)
det_tf	TF	11_2	2/2 (100%)	2/2 (100%)
det_rdf	RDF	11_2	2/2 (100%)	2/2 (100%)
det_irf	IRF	11_2	2/2 (100%)	2/2 (100%)
det_drdf	DRDF	11_2	2/2 (100%)	2/2 (100%)
det_wdf	WDF	11_2	2/2 (100%)	2/2 (100%)
det_cftr	CFtr	11110110_2	6/8 (75%)	6/8 (75%)
det_cfdrd	CFdrd	11001111_2	6/8 (75%)	6/8 (75%)
det_cfwd	CFwd	11001111_2	6/8 (75%)	6/8 (75%)

Table 3. The performance of the implemented MBIST Controller circuit in FPGA.

MBIST Controller	FPGA Logic Element (LE)	Maximum Frequency (F_{max})	Power Consumption
Automatic EDA generated	580	112.2 MHz	153.05 mW
Proposed FSM-Based	92 (-84%)	89.3 MHz (-20%)	103.4 mW (-32%)

Table 3 presents the proposed MBIST controller's circuit area, speed, and power consumption. The circuit area is measured based on the utilization of logic elements (LEs) in the FPGA, as reported by the fitter. At the same time, speed is determined by its maximum clock frequency (F_{max}) calculated in the static timing analysis. Power consumption is assessed through the report generated during the power analysis. To evaluate its performance, the proposed FSM-based MBIST controller is compared to its equivalent version, which was

automatically generated using Siemens EDA Tessent MemoryBIST and implemented on the same FPGA development board. The results show that the proposed FSM-based MBIST controller utilizes fewer logic elements (84% less than those used by the automatically generated by the EDA), resulting in a smaller circuit area. Additionally, it consumes 32% less power than the automatically generated MBIST controller. However, based on the static timing analysis reports, the proposed FSM-based MBIST controller has a 20% lower F_{max} than its equivalent, suggesting that the latter may operate at a clock frequency above 89.3 MHz. Overall, this comparison highlights the superiority of the proposed FSM-based MBIST controller in terms of circuit area and power consumption. Despite having a lower F_{max}, it shall operate correctly using the 50 MHz onboard clock on the FPGA development board used.

4. CONCLUSION

This paper presents the design, implementation, and validation of a user-defined FSM-based MBIST controller that applies the March AZ2 test algorithm for efficient and comprehensive memory testing. The proposed controller addresses the limitations of automatically generated MBIST solutions from EDA tools by offering more flexibility to reduce circuit area and improve performance. The proposed MBIST controller, implemented using a 15-state FSM architecture, successfully replicates the March AZ2 test sequence in hardware while balancing fault coverage and test complexity. It was successfully implemented on the Intel MAX10 DE10-Lite FPGA Development Board. Experimental validations on both fault-free and fault-inserted memory models confirm the controller's functional correctness and its ability to fully detect memory faults such as SAF, TF, RDF, IRF, DRDF, and WDF, with 75% coverage of targeted coupling faults: CFtr, CFdrd, and CFwd. Moreover, a comparative analysis demonstrates that the proposed FSM-based design outperforms its automatically generated equivalent using the Siemens EDA Tessent MemoryBIST tool by utilizing 84% fewer logic elements and consuming 32% less power. Further analysis can be conducted to optimize the FSM by minimizing the number of states, thereby reducing the circuit area while preserving the same fault coverage.

ACKNOWLEDGEMENT

The authors extend their appreciation to the Ministry of Higher Education Malaysia for the financial assistance provided through the FRGS/1/2024/TK07/UTEM/02/17 research grant and Universiti Teknikal Malaysia Melaka for its support of this research.

REFERENCES

[1] K. Wojciechowski, Andrzej A. Marcinek and W. A. Pleskacz, “Configurable MBIST Processor for Embedded Memories Testing,” in 2019 MIXDES-26th International Conference" Mixed Design of Integrated Circuits and Systems", Rzeszow: IEEE, 2019, pp. 341–344. doi: https://doi.org/10.23919/MIXDES.2019.8787161.

[2] R. Wanga, Z. Huangb, G. Caib, and Z. Yua, “A Built-In Self-Test Circuit Based on March FRDD Algorithm for FinFET Memory,” in Industrial Engineering and Applications: Proceedings of the 10th International Conference on Industrial Engineering and Applications (ICIEA 2023), April 4-6, 2023, Phuket, Thailand, 2023, p. 437. doi: http://dx.doi.org/10.3233/ATDE230069.

[3] T. S. N. Kong et al., “An Efficient March (5n) FSM-Based Memory Built-In Self Test (MBIST) Architecture,” in 2021 IEEE Regional Symposium on Micro and Nanoelectronics (RSM), 2021, pp. 76–79. doi: 10.1109/RSM52397.2021.9511602.

[4] Q. Khasawneh, "Reducing the Production Cost of Semiconductor Chips Using (Parallel and Concurrent) Testing and Real-Time Monitoring," Southern Methodist University, 2019. [Online]. Available: https://scholar.smu.edu/engineering_electrical_etds/32

[5] O. S. Nisha and K. Sivasankar, "Architecture for an efficient MBIST using modified march-y algorithms to achieve optimized communication delay and computational speed," Int. J. Pervasive Comput. Commun., vol. 17, no. 1, pp. 135–147, Jan. 2021, doi: https://doi.org/10.1108/IJPCC-05-2020-0032.

[6] G. P. Acharya, M. A. Rani, G. G. Kumar, and L. Poluboyina, "Adaptation of March-SS algorithm to word-oriented memory built-in self-test and repair," Indones. J. Electr. Eng. Comput. Sci., vol. 26, no. 1, pp. 96–104, 2022, doi: 10.11591/ijeecs.v26.i1.pp96-104.

[7] D. Jamal and R. Veetil, "Efficient MBIST Area and Test Time Estimator Using Machine Learning Technique," in 2023 36th International Conference on VLSI Design and 2023 22nd International Conference on Embedded Systems (VLSID), 2023, pp. 223–228. doi: 10.1109/VLSID57277.2023.00054.

[8] P. Ramakrishna, T. Vamshika, and M. Swathi, "FPGA Implementation of Memory Bists using Single Interface," Int. J. Recent Technol. Eng., vol. 9, no. 3, pp. 55–58, 2020, doi: 10.35940/ijrte.b3975.099320.

[9] I. Mrozek, N. A. Shevchenko, and V. N. Yarmolik, "Universal Address Sequence Generator for Memory Built-in Self-test," Fundam. Informaticae, vol. 188, no. 1, pp. 41–61, 2022, doi: 10.3233/FI-222141.

[10] S. Martirosyan and G. Harutyunyan, "An Efficient Fault Detection and Diagnosis Methodology for Volatile and Non-Volatile Memories," in 2019 Computer Science and Information Technologies (CSIT), 2019, pp. 47–51. doi: 10.1109/CSITechnol.2019.8895189.

[11] L.-T. Wang, C.-W. Wu, and W. Xiaoqing, VLSI Test Principles and Architectures: Design for Testability. Elsevier, 2006.

[12] A. J. Van De Goor, "Using march tests to test SRAMs," IEEE Des. Test Comput., vol. 10, no. 1, pp. 8–14, 1993, doi: 10.1109/54.199799.

[13] V. A. Vardanian and Y. Zorian, "A March-based fault location algorithm for static random access memories," in Proceedings of the Eighth IEEE International On-Line Testing Workshop (IOLTW 2002), 2002, pp. 256–261. doi: 10.1109/OLT.2002.1030228.

[14] A. Z. Jidin et al., "Generation of New Low-Complexity March Algorithms for Optimum Faults Detection in SRAM," IEEE Trans. Comput. Des. Integr. Circuits Syst., vol. 42, no. 8, pp. 2738–2751, 2023, doi: 10.1109/TCAD.2022.3229281.

[15] A. J. Van De Goor, G. N. Gaydadjiev, V. G. Mikitjuk, and V. N. Yarmolik, "March LR: a test for realistic linked faults," in Proceedings of 14th VLSI Test Symposium, Apr. 1996, pp. 272–280. doi: 10.1109/VTEST.1996.510868.

[16] S. Hamdioui and A. J. Van De Goor, "An experimental analysis of spot defects in SRAMs: realistic fault models and tests," in Proceedings of the Ninth Asian Test Symposium, 2000, pp. 131–138. doi: 10.1109/ATS.2000.893615.

[17] Z. Zhi-chao, H. Li-gang, and W. Wu-chen, "SRAM BIST Design Based on March C+ Algorithm," Mod. Electron. Tech., vol. 34, no. 10, pp. 149–151, 2011.

[18] I. G. Matri, Aishwaraya, N. Shreya, S. V Siddamal, and S. V Budihal, "A Novel March XR Algorithm, Design, and Test Architecture for Memories BT - Advances in Electrical and Computer Technologies," T. Sengodan, M. Murugappan, and S. Misra, Eds., Singapore: Springer Nature Singapore, 2022, pp. 321–329.

[19] M. A. Ahmed and A. M. Abuagoub, "MBIST Controller Based on March-ee Algorithm," J. Circuits, Syst. Comput., 2020, doi: 10.1142/S0218126621501607.

[20] G. Harutunyan, V. A. Vardanian, and Y. Zorian, "Minimal march tests for unlinked static faults in random access memories," in Proceedings of the IEEE VLSI Test Symposium, 2005, pp. 53–59. doi: 10.1109/VTS.2005.56.

[21] Z. Jianping, W. Zhenyu, Y. Jia, P. Wei, and Z. Yun, "An Improved SRAM Fault Built-in-self-test Algorithm," J. Hunan Univ. Nat. Sci., vol. 46, no. 4, 2019.

[22] S. Hamdioui, A. J. Van De Goor, and M. Rodgers, "March SS: A test for all static simple RAM faults," in Records of the IEEE International Workshop on Memory Technology, Design and Testing, 2002, pp. 95–100. doi: 10.1109/MTDT.2002.1029769.

[23] A. Paul and P. Rony Antony, "Optimized Microcode BIST Architecture for Multiple Memory Cores in SoCs," in 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), May 2018, pp. 910–914. doi: 10.1109/RTEICT42901.2018.9012276.

[24] X. Ning, H. Yang, M. Zhang, Y. Wang, Y. Zhao, and S. Qiao, "Multi-type SRAM Test Structure with an Improved March LR Algorithm," in 2022 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2022, pp. 578–582. doi: 10.1109/APCCAS55924.2022.10090328.

[25] Siemens, "Tessent MemoryBIST: Embedded memory self-test, repair, and debug," 2019. [Online]. Available: https://static.sw.cdn.siemens.com/siemens-disw-assets/public/81225/en-US/Siemens-SW-Embedded-memory-self-test-repair-and-debug-Tessent-MemoryBIST-FS-81225-C1

[26] A. Z. Jidin, R. Hussin, L. W. Fook, M. S. Mispan, and W. Y. Loh, "Automatic generation of user-defined test algorithm description file for memory BIST implementation," Int. J. Reconfigurable Embed. Syst., vol. 11, no. 2, p. 103, 2022.

[27] A. Z. Jidin, R. Hussin, M. S. Mispan, and L. W. Fook, "A new 13N-complexity memory built-in self-test algorithm to balance static random access memory static fault coverage and test time," Int. J. Electr. Comput. Eng., vol. 15, no. 1, pp. 163–173, 2025.

https://doi.org/10.31436/iiumej.v27i1.3780

Adaptive Energy Balance Control System via State of Charge (SOC) for a Sustainable Solar-Powered Outdoor-Hydroponics in Tropical Islands

MUJI JUHERWIN[1], MOHAMAD FARID MISNAN[2,3*], SAKHIAH ABDUL KUDUS[4], AKIHIKO SATO[5]

[1]*Energy System Engineering Study Program, Lombok Institute of Technology, West Nusa Tenggara, 83662, Indonesia.*
[2]*Smart Manufacturing Research Institute (SMRI), Universiti Teknologi MARA Shah Alam 40450, Selangor, Malaysia.*
[3]*Faculty of Electrical Engineering, Universiti Teknologi MARA Shah Alam 40450, Selangor, Malaysia.*
[4]*Faculty of Civil Engineering, Universiti Teknologi MARA Shah Alam 40450, Selangor, Malaysia.*
[5]*Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University, 615-8540, Kyoto, Japan*

**Corresponding author: mohamadfarid@uitm.edu.my*

(Received: 8 June 2025; Accepted: 20 November 2025; Published online: 12 January 2026)

ABSTRACT: Efficient energy management is essential for sustaining outdoor hydroponics systems powered by solar energy, particularly in tropical island environments where sunlight and rainfall vary throughout the day. To address this, a solar-powered hydroponics system was developed with an adaptive energy balance control strategy based on the State of Charge (SoC). The system requires reliable real-time monitoring and decision-making, achieved by integrating voltage, current (ACS712), light-dependent resistors (LDR), and flow sensors, along with an ESP32 microcontroller for data acquisition and control logic. The adaptive control method dynamically regulates power consumption by adjusting the water pump's operation in response to SoC levels, solar radiation, and rainfall. Experimental validation shows the system maintains the battery's SoC above 55%, ensuring power availability while optimizing energy use. Pump operation is disabled during rainfall and minimized at night to prevent deep discharge, enhancing overall system stability. Daytime solar charging is complemented by controlled discharge during non-solar hours, improving energy sustainability. The results confirm the effectiveness of the proposed strategy in reducing unnecessary energy consumption, improving system reliability, and supporting continuous hydroponic cultivation under varying tropical conditions.

ABSTRAK: Pengurusan tenaga yang cekap amat penting bagi memastikan kelestarian sistem hidroponik luar yang menggunakan tenaga solar, terutama di kawasan pulau tropika yang mempunyai corak cahaya matahari dan perubahan hujan sepanjang hari. Bagi memenuhi keperluan ini, satu sistem hidroponik berkuasa solar telah dibangunkan dengan strategi kawalan imbangan tenaga adaptif berasaskan Keadaan Cas (SoC). Sistem ini memerlukan pemantauan masa nyata dan keupayaan membuat keputusan terpercayai, dicapai melalui integrasi penderia voltan, arus (ACS712), LDR (Rintangan Peka Cahaya), dan aliran, serta mikropengawal ESP32 bagi pemerolehan data dan logik kawalan. Kaedah kawalan adaptif ini mengatur penggunaan tenaga secara dinamik dengan melaras operasi pam air berdasarkan tahap SoC, intensiti cahaya matahari, dan keadaan hujan. Dapatan kajian menunjukkan sistem ini mampu mengekalkan SoC bateri melebihi 55%, sekaligus memastikan bekalan kuasa yang stabil sambil mengoptimum penggunaan tenaga. Operasi

pam dihentikan semasa hujan dan dikurangkan pada waktu malam bagi mengelakkan nyahcas bateri berlebihan, seterusnya meningkatkan kestabilan sistem. Pengecasan bateri pada waktu siang dilengkapi dengan penyahcasan terkawal semasa tanpa cahaya matahari, sekaligus memperkukuh kemampanan tenaga. Dapatan kajian membuktikan bahawa strategi kawalan ini berkesan dalam mengurangkan penggunaan tenaga tidak diperlukan, meningkatkan kebolehpercayaan sistem, dan menyokong penanaman hidroponik berterusan dalam persekitaran tropika yang dinamik.

KEYWORDS: *State of Charge, Adaptive Energy Management, Solar-Powered Hydroponics, Power Allocation, Environmental Condition.*

1. INTRODUCTION

The integration of renewable energy into agricultural practices has received considerable attention in recent years. It is essential to have high solar potential, as in tropical islands [1-5]. In Indonesia, where farming remains a major economic activity [6,7,8], solar cell frameworks can sustainably address energy needs [9,10], particularly for solar energy in farming practices such as hydroponics [11,12]. Hydroponics is a method of growing plants without soil, instead using a water solution that contains nutrients [13]. It is most effective in tropical climates that permit year-round farming [14]. Research confirms the effectiveness of hydroponic techniques, reporting significantly higher yields than conventional soil-based agriculture [15-17]. One of the foremost challenges remains the efficient management of energy resources under tropical conditions in solar-powered hydroponic systems [18,19].

The variability of solar irradiance poses a challenge for energy management in solar-powered systems [20,21]. In tropical regions, the variability of solar irradiance is pronounced due to sudden weather changes [22]. These challenges impede the implementation of efficient energy-balance control strategies, increasing the risk that solar-powered hydroponic systems cannot achieve effective energy-supply control for consistent plant growth. Management of energy within the system is also critical because overcharging and over-discharging lead to excessive battery cycling, damage to essential components, and reduced overall system efficiency [23-25]. Real-time consideration of the link between energy supply and demand is necessary for energy management systems to address these problems [26-28]. One of the most effective methods is State-of-Charge (SOC)-based adaptive energy balance control [29,30]. SOC, a critical determinant of energy flow in hydroponic systems, denotes the state of charge of energy storage systems such as batteries [31]. With real-time SOC measurement, adaptive control can adjust energy expenditure based on the current state of stored energy [32]. This method not only permits the trimetallic components to function optimally but also ensures the continued operation of essential functions, such as water nutrient pumping. This paper builds on existing SOC-based control strategies by introducing a novel integration of rainwater runoff detection and management to further optimize water and energy use in solar-powered hydroponic systems designed for tropical island environments. The integration of SOC-based energy management with dynamic water-use adjustments in response to rainfall events represents a significant advancement in sustainable hydroponic practices.

Although several studies have implemented solar-powered hydroponics and SoC-based energy management [29–32], most do not integrate adaptive control with rainfall detection for water-energy optimization in tropical outdoor conditions. Furthermore, existing frameworks often lack practical validation in off-grid island environments, where high solar intermittency and rainfall variability pose unique challenges. This study addresses these gaps by developing and experimentally validating a novel SoC-adaptive energy balance control

https://doi.org/10.31436/iiumej.v27i1.3780

system that incorporates rainwater management, thereby ensuring continuous hydroponic cultivation under tropical variability.

This paper aims to achieve the following: (1) create an SOC-based adaptive control of energy balance for a solar-powered outdoor hydroponics system tailored for tropical islands and (2) evaluate how effective the system is in optimizing energy consumption and reliability in tropical conditions. This control strategy enables appropriate adaptive responses to changes in energy availability and requirements, thereby reducing inefficiencies and the risk. This study focuses on improving sustainable agriculture policies in tropical regions.

2. METHOD

This study employs an SOC-based approach to design and evaluate an adaptive energy-balance control system for a sustainable solar-powered hydroponic system in a tropical environment. The methodology includes three main steps: system design and algorithmic strategy formulation, prototype implementation, and system dependability assessment.

2.1. Solar-Powered Outdoor Hydroponics System

The system is engineered to incorporate solar photovoltaic, energy storage, and hydroponic subsystems, with adaptive energy-balance management. The design is illustrated in the system block diagram in Figure 1, which depicts the interactions among the solar PV panels, battery storage, hydroponic subsystems, and energy management components.

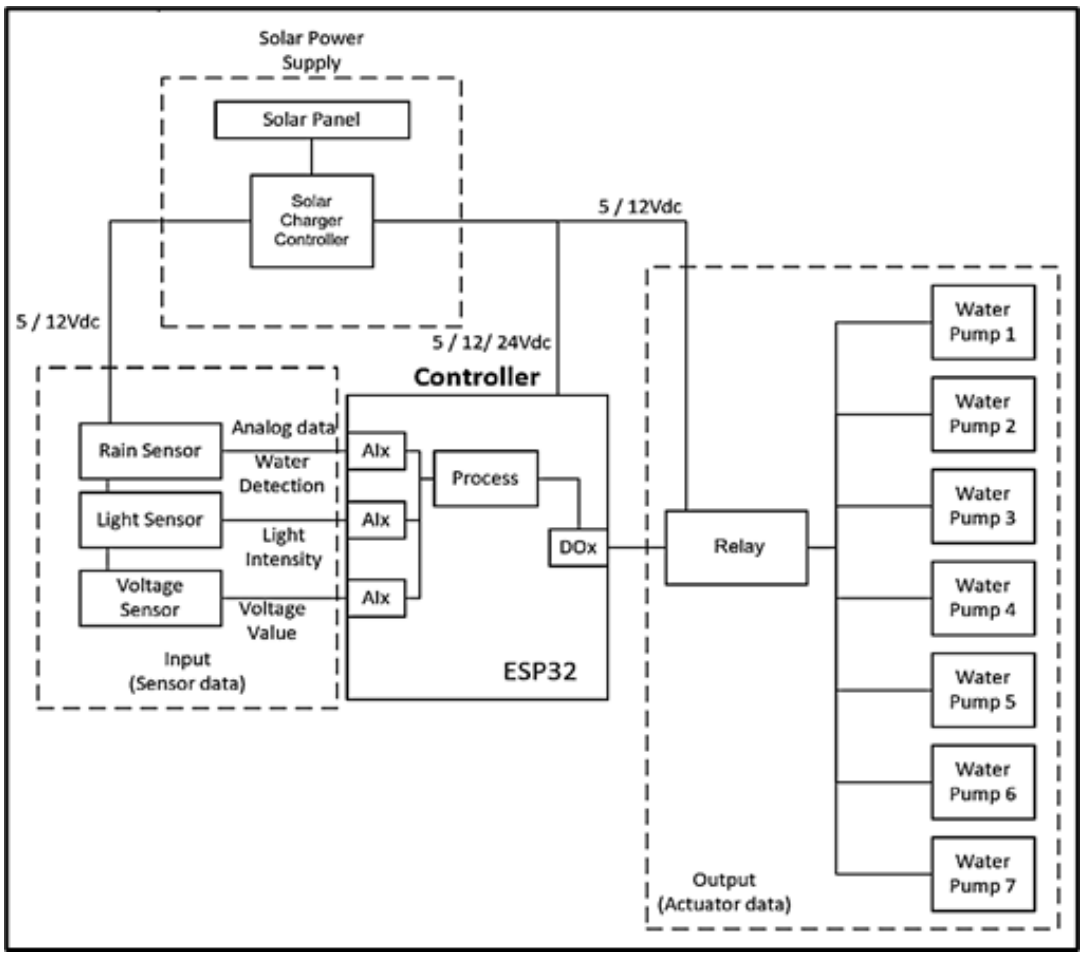


Figure 1. System Block Diagram

https://doi.org/10.31436/iiumej.v27i1.3780

2.2. SOC-Adaptive Control System

The adaptive energy balance control algorithm, based on SOC, governs the reservoir's pumping activities by allocating pumping energy based on real-time SOC measurements and available energy. The SoC status is calculated inside the microcontroller using the following simple approach:

$$SoC = \frac{V - V_{min}}{V_{max} - V_{min}} \times 100\% \quad (1)$$

where V is the terminal voltage of the battery, expressed in volts (V). V_{min} is the minimum battery voltage, dubbed the fully discharged state (0% SoC), while V_{max} indicates a fully charged state (100% SoC) of the battery's maximum voltage. The formula provides a measure of battery voltage relative to its operating limits and expresses SoC in percent [33].

Fluctuations in charging and discharging cycles will not yield an accurate SoC value. Volumetric charge estimation provides an effective means of battery protection, as it disables power-pumping commands when the SoC is reported below 55%. The 55% SoC threshold was chosen to provide a protective margin for the lead-acid battery, thereby reducing degradation risk. Literature confirms that operating below 50% SoC accelerates sulfation and reduces lifespan [24,25]. By selecting 55%, the system balances usable capacity with long-term reliability, ensuring uninterrupted hydroponic operation. Operating below this threshold for extended periods can lead to increased battery degradation and shorten its lifespan [34]. Furthermore, maintaining a buffer above 55% enables the system to handle unexpected increases in energy demand or prolonged periods of low solar irradiance, ensuring the continuous operation of critical functions such as nutrient pumping. SoC determination using characteristic parameters is performed during the investigation, as it is based on the Open Circuit Voltage and internal resistances, as modified [35,36].

The system automatically controls the pump's operation, causing it to turn on and off at scheduled intervals. Meanwhile, the solar panels generate electricity. Consequently, the voltage level changes over time. The value of its fluctuation is determined by the internal resistance of the battery system, which can be defined as:

$$R_{int} = \frac{V_{no\ load} - V_{load}}{I_{load}} \quad (2)$$

R_{int} is to be measured in ohms (Ω) and defined by the value of the voltage drop concerning the power used. V_{noload} is referred to as the battery's no-load or open-circuit voltage (V), which represents the battery's voltage when no current is drawn. V_{load} is referred to as loaded voltage (V), which represents a battery's voltage with a load connected such that current flows. I_{load} is a load current in amperes (A) that refers to the current drawn by the load provided to the system. The equation determines internal resistance by calculating the difference between the no-load and loaded voltages divided by the current drawn [37]. It is one of the factors used to smooth the voltage value for analysis. The smoothed voltage estimate is used to calculate SoC for analysis.

$$V_{corrected} = V + (I \times R_{int}) \quad (3)$$

In this manner, the objective is to follow the required dynamics during the experiment. Therefore, the SoC is calculated using the exponential moving average (EMA) of the voltage value. The EMA of the voltage at time t is given by:

$$V_{EMA_t} = \alpha\, x_t + (1 - \alpha)\, V_{EMA_{t-1}} \quad (4)$$

where V_{EMAt} is the value of the exponential moving average of voltage at time t, x_t is the latest value of voltage, V_{EMAt-1} is the previous value of VEMA, and α is the smoothing factor, which determines the weight given to recent observations and is calculated as:

$$\alpha = \frac{2}{N+1} \quad (5)$$

where N represents the selected window size or period for smoothing [38]. The EMA helps reduce fluctuations in voltage measurements, providing a more stable representation of SoC dynamics over time.

The adaptive approach relying on SoC guarantees the retention of optimal energy usage without exceeding the battery discharge limit. The algorithm operates by first collecting real-time SoC data, solar energy data, and the hydroponic system's energy needs. It allocates energy based on system demands, ensuring that critical functions remain operational even during periods of low energy supply. For greater efficiency, a rain sensor is installed to automatically control water pumping, reducing excess irrigation and water waste during rainfall. In addition, the algorithm adjusts the supplied power to different subsystems to optimally balance demand and supply, thereby extending the battery's operating time. The algorithm's logic and feedback control structure are shown in a flowchart in Figure 2.

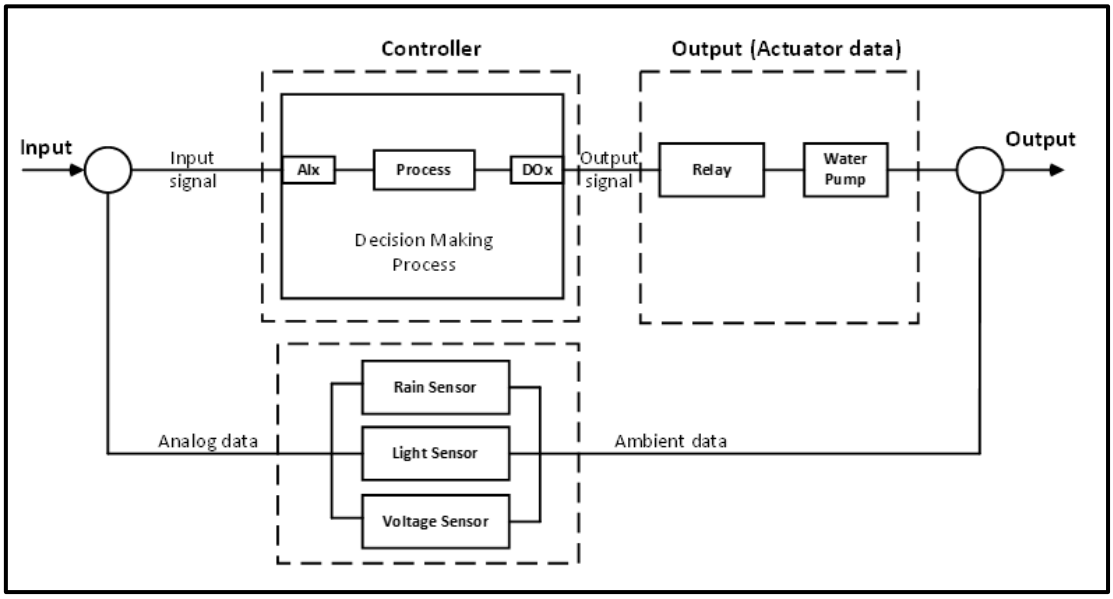


Figure 2. Decision-Making Process and Feedback Control Loops

The flowchart depicts an automated hydroponics water-pumping system with an adaptive decision-feedback control structure. The process starts with setting digital pins for relay control and defining essential variables, which are rain detection, light, and battery SoC. Moreover, the system continuously monitors a rain sensor, LDR (light-dependent resistor), and a voltage sensor to capture environmental and battery data. The first check is to determine whether it is raining; if rain is detected, all water pumps will be switched off to avoid unnecessary irrigation and conserve water resources. If no rain is detected, then the system checks if there is daylight. If yes, all water pumps will be turned on to facilitate irrigation. If there is no daylight, the system checks the battery's SoC. If the SoC drops below 55%, pumps are set to an energy-saving mode where they run for two minutes and then pause for ten minutes to conserve battery. This feedback control loop adapts in real time to changes in environmental conditions to maximize battery and water use in an energy-efficient manner. Adaptive energy management enables the system to conserve battery power, while rain

https://doi.org/10.31436/iiumej.v27i1.3780

detection reduces water usage, optimizing the system for sustainable, autonomous, and scalable hydroponic solutions.

The novelty of this method lies in its integration of SoC-adaptive control with rainfall detection, enabling dynamic pump regulation that conserves both energy and water. Unlike conventional SoC-based approaches [33–36], our system introduces a multi-criteria feedback loop (SoC, solar irradiance, and rainfall) to ensure stable operation in tropical outdoor conditions. This combined strategy provides resilience against both energy intermittency and nutrient dilution risks, which are rarely addressed simultaneously in previous studies.

2.3. Rain Protection System

The rain protection system is designed to efficiently manage drainage to prevent excess water from disrupting the equilibrium of the outdoor hydroponics system. The system's main components are an automatic valve that controls water entering the nutrient bucket, a water-collection mechanism, and a rainwater runoff detection system.

2.3.1. Rainwater Runoff Detection

Rainwater can significantly affect the nutrient solution concentration in an outdoor hydroponics system; therefore, the system uses a flow sensor to detect the presence and volume of runoff. A water collector is installed to collect rainwater runoff from the surface of the solar panels, thereby ensuring accurate detection. The two main purposes of this arrangement are to ensure that the collected water volume is sufficient for the flow sensor to detect accurately and to provide an early warning of excess rainfall runoff that might enter the hydroponics system.

The rationale for collecting runoff from solar panels is that the onset of water flow from the panels indicates that precipitation is accumulating and may soon enter the hydroponic system. Early detection enables the system to act promptly to control water flow and prevent undesired dilution of the nutrient solution, thereby preserving optimal growing conditions for the plants.

2.3.2. Excess Water Control

A backflow prevention device is equipped with a nutrient backflow pipe that contains a drain hole to reduce the concentration of diluted nutrient solution resulting from rainfall intake. Excess rainwater can be discarded before it reaches the nutrient bucket. Thus, unwanted water is prevented from contaminating the nutrient solution. If the flow sensor detects additional runoff, all pumps are automatically shut off, allowing the remaining nutrient solution to drain from the hydroponic pipes back to the nutrient bucket. This system averts overfilling while preserving the hydroponic nutrient solution’s concentration required for optimal plant growth.

Mechanically, a ball-type automatic valve is used to regulate the flow of water into the nutrient bucket, in addition to the drain hole. Under system-optimal conditions, the valve remains open until the volume of collected rainwater exceeds a specified threshold. Consequently, below the designated threshold, the hydroponic system can operate as designed. If collected rainwater exceeds a predetermined threshold, the floating ball mechanism will guide the valve to close, preventing further entry of non-nutrient water for plant nourishment. It facilitates the minimization of water-solution imbalance despite heavy rainfall.

https://doi.org/10.31436/iiumej.v27i1.3780

2.4. Implementation

A solar-powered outdoor hydroponics system prototype was constructed and tested in a real-world setting to validate the proposed layout. The experiment was configured with a photovoltaic system with a capacity of 12V and 160 Wp, providing an uninterrupted power supply. A 12V 200Ah lead-acid battery bank handles energy storage and dispatching. The system employed two 12V DC submersible water pumps (flow rate 1.2 L/min, head 2.0 m, power consumption 15 W each). These were selected for their low power demand, corrosion resistance, and suitability for continuous NFT hydroponic circulation. The energy monitoring sensors enhance resource optimization. For autonomous control, sensors measuring SoC, solar irradiation, and rainwater runoff are processed by an ESP32 microcontroller executing an SoC-based control algorithm. The system was installed in the tropical outdoor environment in Lenek District, East Lombok, West Nusa Tenggara, a region characterized by a tropical climate with distinct wet and dry seasons. The prototype was operated and monitored continuously for four days, from May 15–18 May 2025, coinciding with the transition into the dry season. This period was intentionally selected because it experiences high daytime irradiance with intermittent afternoon rainfall, providing a suitable test of the system's ability to maintain SoC stability under variable tropical weather conditions. The hydroponic system employed was a Nutrient Film Technique (NFT) system, consisting of two parallel channels for cultivating lettuce (Lactuca sativa L.). This setup enabled the evaluation of the system's energy management efficiency, responsiveness to changes in solar irradiance, and performance under adverse weather conditions. As summarized in Table 1, the proposed SoC-adaptive hydroponics control demonstrates clear improvements over previous studies by integrating rainfall awareness and energy optimization, thereby maintaining system reliability with SoC levels consistently above 55%.

Alternative approaches, such as timer-based control Novaldo *et al.,* 2022 [10] or simple automatic irrigation strategies Chaiwongsai, 2019 [14] have been proposed but lack adaptability to fluctuating energy and weather conditions. Similarly, aquaponics prototypes using Arduino [12] demonstrated monitoring capability but lacked active SoC-based optimization. In contrast, our approach not only maintains an SoC above 55% but also adapts irrigation in response to rainfall, offering a more sustainable and robust solution for tropical environments.

Table 1. Comparison of the Proposed SoC-Adaptive Hydroponics Control with Previous Studies

Study	Application Context	Control Approach	Key Features	Reported Limitations	Performance Outcome
Novaldo *et al.* (2022)	Pilot hydroponics, Indonesia	Solar-PV + timers	Simple automation	No adaptive SoC; not rainfall aware	Stable only in dry conditions
Bakar *et al.* (2022)	Aquaponics prototype	Arduino + SMS alert	Remote monitoring	No energy optimization	Limited system reliability
Chaiwongsai (2019)	Tropical hydroponics	Automatic control	Focus on growth environment	Lacked integrated energy control	High water use
This study	Outdoor hydroponics, tropical island	SoC-adaptive + rainfall integration	Energy-efficient, rainfall adaptive	Requires calibration	Maintained SoC > 55%, improved reliability

https://doi.org/10.31436/iiumej.v27i1.3780

3. RESULTS AND DISCUSSION

The SOC-based adaptive energy balance control was created for an efficient solar-powered outdoor hydroponics system tailored for tropical islands.

3.1. System Implementation

3.1.1. Excess Water Control

Figure 3 depicts a collection and management system designed to efficiently regulate water. It consists of a storage container with a float valve, an excess water outlet, and an emergency cleaning drain. The float valve (indicated in red) controls the water level, preventing overflow while preventing waterless cleanup due to excessive rain inflow. Overflow water (excess water) from the piping maintains equilibrium, preventing the storage container from exceeding its capacity. An emergency or cleanout valve is integrated into the system to enable maintenance and cleaning of the drainage system when required to maintain controlled drainage. The entire assembly is constructed from PVC pipes, which makes it durable and easy to install in a wide range of environmental conditions.

Figure 3. Backflow pipe for nutrient circulation with excess-water control, using a floating ball valve for regulation.

3.1.2. Sensors and Microcontroller ESP32

Figure 4 shows a smart sensor-based control system for a solar-powered outdoor hydroponics system that manages energy and water flow based on the State of Charge (SOC) level. The system employs an ESP32 microcontroller (cyan) as the primary processing unit, controlling system operations via the sensors.

The rainwater harvesting system (blue) temporarily stores rainwater from the solar panels. As the water level in the temporary collection chamber reaches a predefined threshold, it signals potential rainwater inflow into the hydroponic system. The flow sensor (blue) detects water movement and signals the ESP32 to turn off the relay module (red), preventing nutrient solution dilution.

An orange Light Dependent Resistor (LDR) is used to identify whether the system operates in day or nighttime mode. The system's on/off control cycle operates based on the state of charge (SOC), which is typically determined by a voltage sensor (green) within the system. However, because voltage readings alone do not provide sufficient information for

https://doi.org/10.31436/iiumej.v27i1.3780

accurate SOC estimation, a real-time current sensor such as the ACS712 (magenta) is used to improve SOC estimation accuracy.

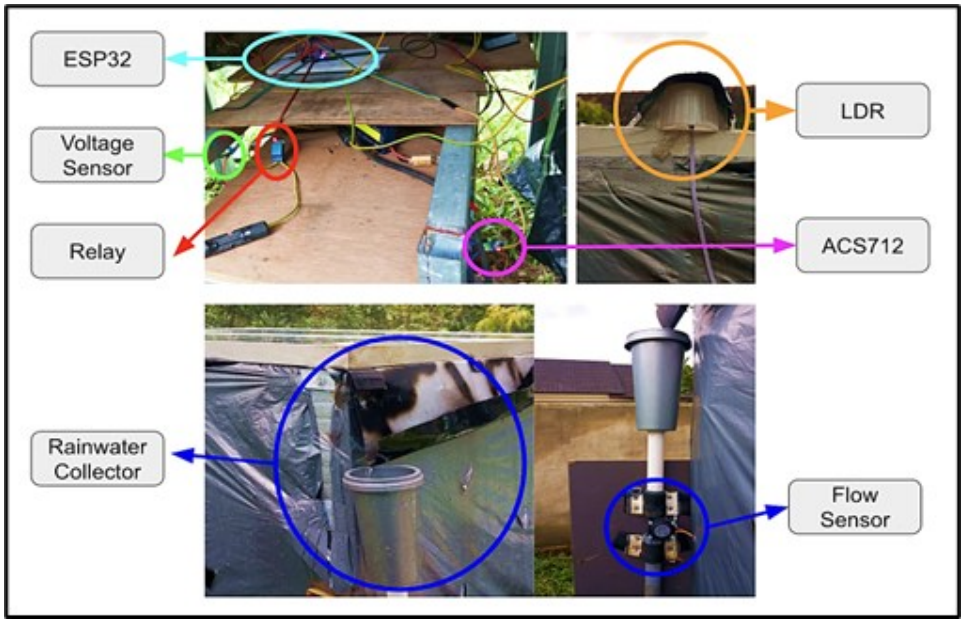


Figure 4. Sensor Integration for SOC-Based Adaptive Energy Balance Control in a Solar-Powered Outdoor Hydroponics System.

The adaptive control system with integrated sensors actively manages energy use and nutrient balance to achieve greater efficiency, advancing the goal of a self-sustaining, solar-enabled hydroponic farming system. The implementation of this control mechanism is illustrated in Figure 5.

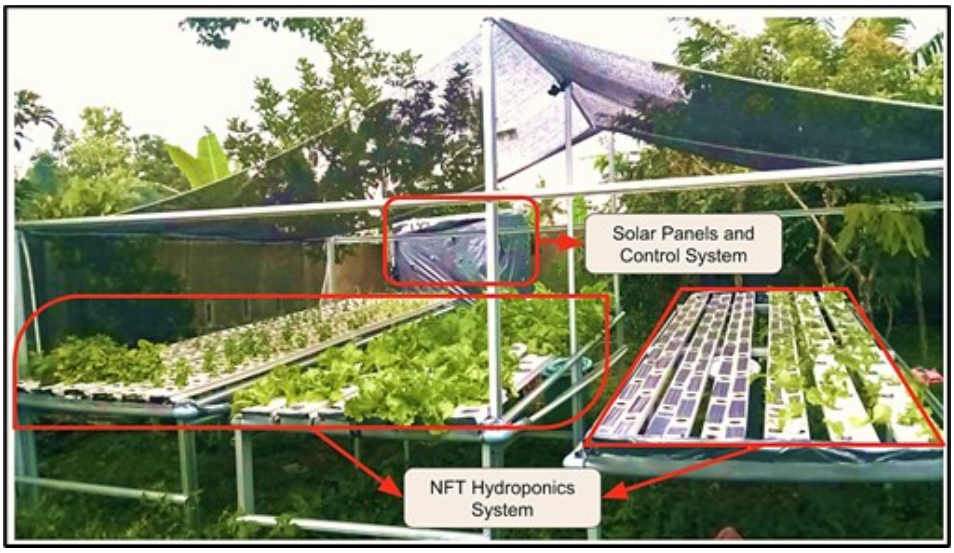


Figure 5. Implementation of Solar-Powered NFT Hydroponics System with Integrated Control Mechanism.

It depicts a hydroponic system employing the NFT technique, controlled by a solar system. The system comprises two parallel NFT hydroponic channels in which crops are grown under a constant flow of nutrient solution. The solar panels, along with the control unit, are mounted above the hydroponic system to enable autonomous operation using renewable energy.

https://doi.org/10.31436/iiumej.v27i1.3780

The solar panels have a dual purpose: in addition to providing electricity to the hydroponics system, they also serve as a surface for collecting rainwater. The collected rainwater is stored in monitoring vessels before being processed by the system's adaptive energy balance control. Nutrient dosing, relay switching, and various growth-condition sensors are controlled by the system's ESP32 microcontroller, which also measures environmental parameters to ensure optimal growing conditions.

As summarized in Table 2, integrating voltage, current, light, and rain sensors with the ESP32 microcontroller enables accurate monitoring, efficient energy use, and reliable automation. This consolidated sensor suite enables the system to maintain optimal SOC estimation, regulate nutrient balance, and achieve self-sustaining operation.

Table 2. Sensors used in the SoC-Adaptive Hydroponic System

Sensor	Type	Specification	Accuracy	Function
Voltage Sensor	ZMPT101B	0–25 V input	±0.5%	Battery voltage monitoring
Current Sensor	ACS712	0–30 A	±1.5% FS	Current measurement for SoC
Light Sensor	LDR	10–1000 lux	±5%	Day/night detection
Rain Sensor	YL-83 + flow sensor	0–5 mm/min sensitivity	±2%	Rainfall detection
Microcontroller	ESP32	Dual-core, WiFi/BT	–	Data processing & control

3.2. Rain Detection vs. Pump Status

Proper management of water resources is vital in hydroponics and irrigation systems to prevent oversaturation and optimize resource efficiency. The system incorporates a rain-detection feature that uses a flow sensor to detect rainwater runoff, thereby helping to control the operation of the water pump. During day mode (08:00 – 16:00), the pump operates on a modified thermal strategy. It is shut off for 10 minutes whenever runoff is triggered, then turned back on for 2 minutes before re-evaluating runoff detection. This cyclical strategy ensures that plants are adequately nourished while minimizing excess water use during periods of rainfall. Figure 6 shows the relationship between rain intensity and pump status, illustrating the system's behavior in response to environmental conditions in real time.

The graph depicts the correlation between rain intensity (mm/h) and pump activity (ON/OFF) during semi-automated daytime operations (LDR reads light), in which pump operation is controlled by a flow sensor that detects rainwater runoff. The blue line illustrates the rainfall intensity, while the red line shows the pump's ON/OFF status. According to the system’s logic, when rainwater runoff is detected, the pump is switched OFF for 10 minutes, then automatically turned ON for 2 minutes. This cycle repeats as rainfall runoff continues to be detected, thereby avoiding unnecessary irrigation during rainfall. As shown in the graph, during the period of significant rainfall, which occurs between approximately 11:00 AM and 12:30 PM, the pump is detected to be OFF most of the time due to abundant runoff. In the absence of rainwater runoff for over 10 minutes, however, the system then resumes normal operations for nutrient augmentation. For daytime operation, the pump behaviour ignores SOC and aims solely to use water efficiently, with operation controlled by rainfall. It helps prevent over-irrigation while ensuring that nutrient levels remain adequate during periods of rainfall cessation, thereby sustaining efficient water management.

https://doi.org/10.31436/iiumej.v27i1.3780

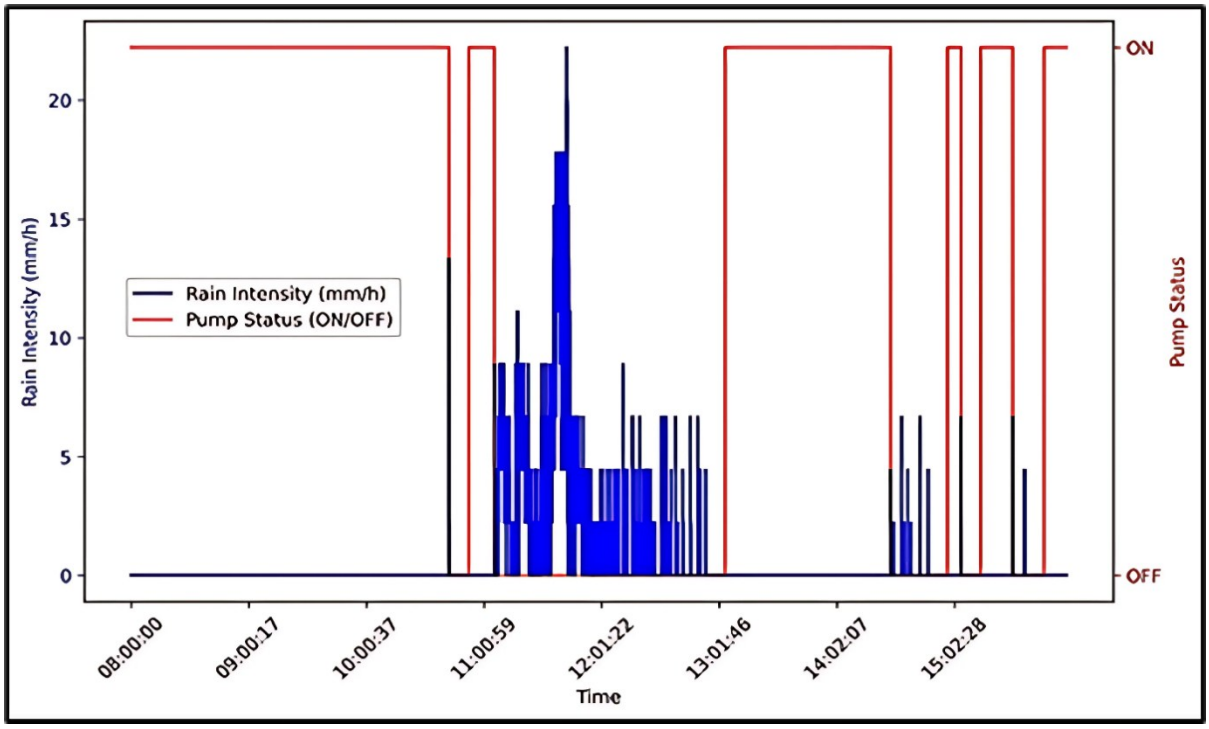


Figure 6. Rain sensor signal vs. pump status during daytime operation (16 May 2025).

3.3. Light Intensity and SoC vs. Pump Status

The system operates in two distinct modes: day and night, in which the battery State of Charge (SoC) and light intensity are critical determinants of water pump operation. During night mode, when solar energy is unavailable, the pump relies entirely on battery power. The SoC (%) represents the battery state of charge, while the light intensity (analog read) indicates the presence or absence of sunlight.

The graph in Figure 7 illustrates the relationship between SoC, pump status, and light intensity during the transition from day to night. As shown in the graph, light intensity declines significantly after sunset, resulting in a gradual decrease in SoC as the battery discharges. The water pump operates intermittently according to predefined control logic to prevent excessive battery discharge. When the SoC reaches a critical threshold, the system may regulate pump activity to conserve energy. This analysis highlights the importance of energy-efficient water management in off-grid hydroponic or irrigation systems.

The graph depicts the relationships among SoC, pump status, and light intensity over time, particularly during the transition from day to night. The light intensity (orange line) drops significantly after sunset, indicating the absence of solar energy. As a result, the system switches to battery power, causing the SoC (green line) to decrease as energy is consumed gradually. The pump status (red line) shows periodic operation, indicating that the pump is intermittently controlled to prevent excessive battery discharge. Initially, when the battery is fully charged, the pump operates more frequently. However, as the SoC declines, the system regulates the pump's operation, likely to conserve energy for critical functions. The fluctuations in SoC suggest minor variations in power consumption and battery efficiency throughout the period. It highlights the importance of energy management in off-grid systems, ensuring that the water pump operates while maintaining sufficient battery levels for nighttime use.

https://doi.org/10.31436/iiumej.v27i1.3780

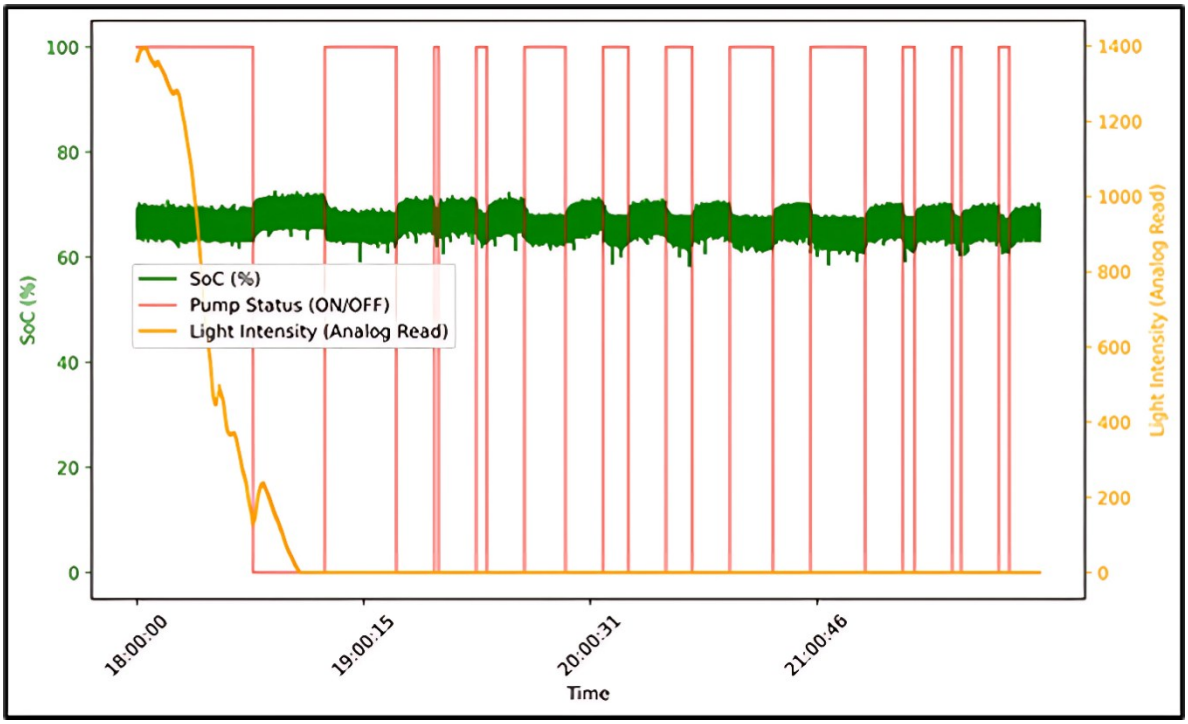


Figure 7. SoC vs. pump status during nighttime operation (16 May 2025).

This analysis highlights the importance of energy-efficient water management in off-grid hydroponic systems. Compared with conventional SOC-based systems that only regulate pump activity based on battery charge [19,20], our system integrates SoC with real-time light intensity and rainfall conditions, providing multi-criteria control. This approach ensures stable operation at night while conserving battery resources, a feature not reported in timer-based or single-parameter control systems [10].

3.4 Combined Variables vs. Pump Status

The graph in Figure 8 illustrates the correlation among light intensity, SoC levels, rainfall intensity, and pump activity throughout the day. It highlights the impact of solar energy availability on battery charging, the effect of rainfall on pump deactivation, and the fluctuations in SoC during charging and discharging cycles.

This graph depicts the relationships between various environmental factors and pump status over 24 hours. Light intensity (orange line) exhibits a typical diurnal variation; it rises in the morning, peaks at midday, and declines in the evening, affecting the battery charging process. The battery's SoC (green line) increases steadily during the day due to solar energy input, and after sunset, it begins to decline as stored energy is used. The pump status (red bars) remains predominantly active during the daytime, when energy is abundant, but rain events (blue bars), which trigger automatic deactivation, cause intermittent off-cycles. At night, due to voltage drops in the battery, frequent on-off cycles can be observed where the system shuts down momentarily to maintain energy efficiency, causing the system to restart. The overall trend indicates that the pump operates dynamically, using solar energy, rainfall, and battery charge-level detection to optimize energy use and ensure continuous operation.

The overall trend indicates that the pump operates dynamically in response to solar irradiance, rainfall, and battery state of charge. This contrasts with earlier hydroponic energy management strategies that primarily relied on solar availability or pre-set schedules without SoC thresholds [10,12]. By integrating rainfall detection, our system prevented unnecessary

https://doi.org/10.31436/iiumej.v27i1.3780

irrigation during wet conditions, thereby improving water efficiency and ensuring uninterrupted hydroponic operation even during periods of environmental variability.

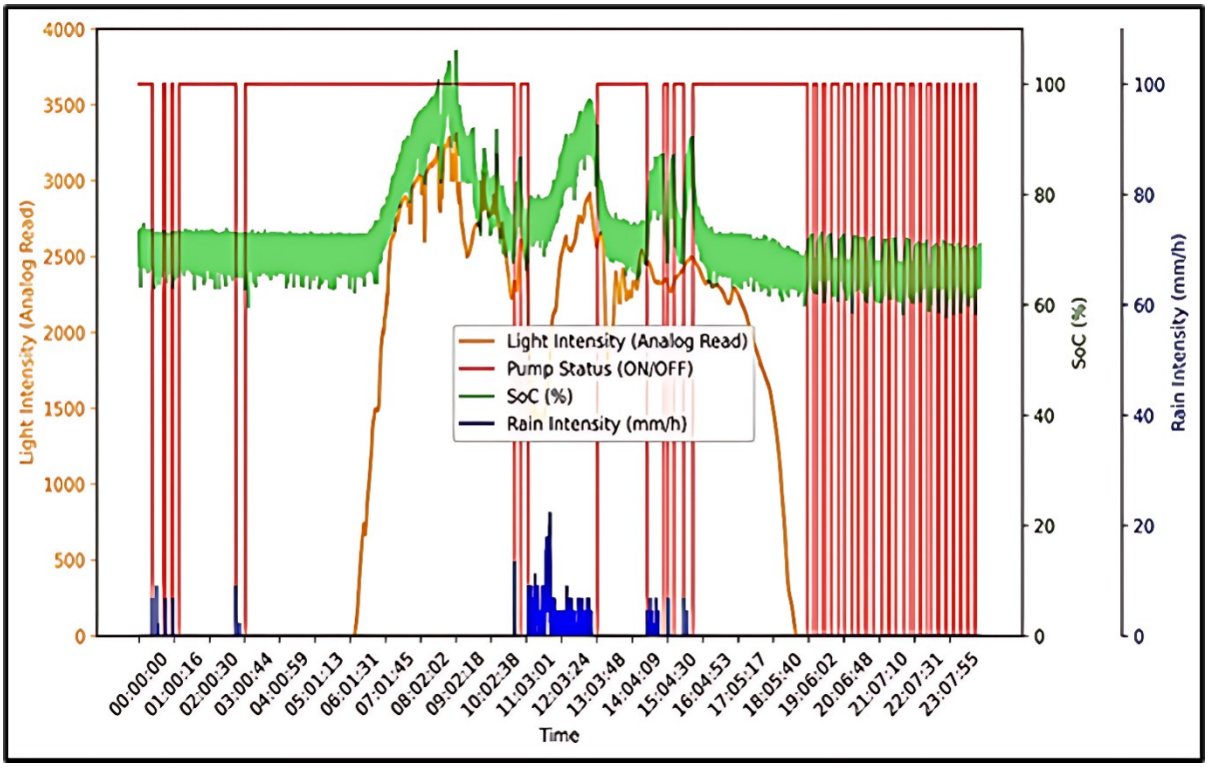


Figure 8. System dynamics under combined variables (rainfall, light intensity, and SoC level) affecting pump status over a 24-hour cycle (17 May 2025).

3.5 Dynamic of SOC Level

Figure 9 illustrates the SoC-level changes over four days, showing correlations with battery charge, pump activity, and environmental activity. It shows the transitions between charging and discharging cycles, the system's response to detected rain, and the pump's cycling behavior at night.

SoC or battery charge (measured as a percentage) remains above 55% throughout the experiment, indicating that energy is available for use. The Exponential Moving Average (EMA) of SoC oscillates between 60% and 90%. Maximums are recorded during solar charging,pumping, and discharging periods, and energy usage gradually decreases. Declination is halted at night due to frequent pump cycling, triggered by the voltage sensor reading that activates the relay flag. Voltage detection likely causes a relay flag to alternate between OFF for 10 minutes and ON for 1 minute every 10 minutes. Voltage detection is likely causing the frame rate to be approximately one update every 10 seconds. It shows the system’s automated response to energy balance and sensor-based control logic, requiring multidimensional energy management optimization over the 24-hour cycle, distributed across day and night.

https://doi.org/10.31436/iiumej.v27i1.3780

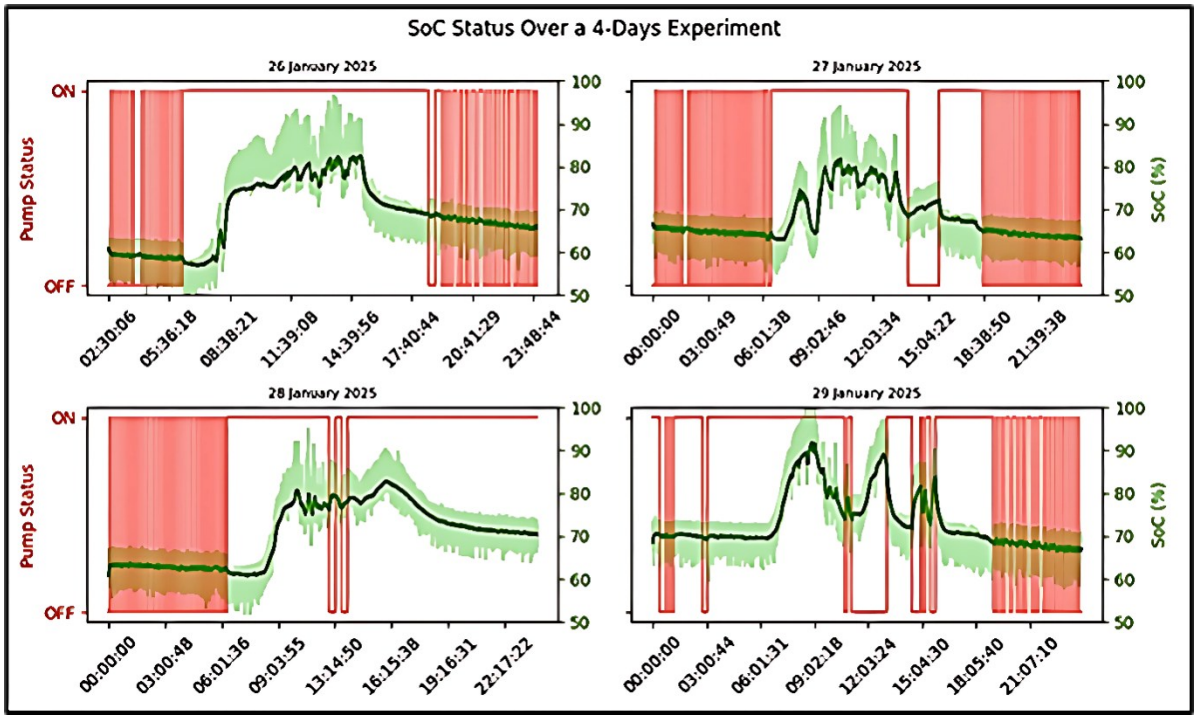


Figure 9. SoC variation over four days (15–18 May 2025).

4. CONCLUSION

This study demonstrated the effectiveness of a SoC-based adaptive control strategy for solar-powered outdoor hydroponics in tropical island environments. The system consistently maintained SoC above 55%, thereby ensuring a continuous energy supply for irrigation while preventing deep discharge. Pump operation was dynamically adjusted, turned off during rainfall, and cycled at night to conserve both water and energy. Compared with previous hydroponic energy management frameworks that relied solely on timers or solar availability, our system integrated rainfall detection and multi-criteria control, resulting in higher system resilience and sustainability. The findings align with recent reports on greenhouse and aquaponic energy management but extend their applicability to real-world tropical conditions. This demonstrates that adaptive SoC-rainfall integration can significantly improve reliability and resource efficiency, contributing to sustainable agriculture in off-grid environments.

ACKNOWLEDGEMENT

The authors would like to express their gratitude for the funding by the Industry Matching Partnership (IMaP) under the Ministry of Higher Education with the grant No. IMaP/2/2024/TK06/UITM//2.

REFERENCES

[1] Rahman, M. M., Khan, I., Field, D. L., Techato, K., & Alameh, K. (2022). Powering agriculture: Present status, future potential, and challenges of renewable energy applications. Renewable Energy, 188, 731-749. https://doi.org/10.1016/j.renene.2022.02.065.

[2] Majeed, Y., Khan, M. U., Waseem, M., Zahid, U., Mahmood, F., Majeed, F., ... & Raza, A. (2023). Renewable energy as an alternative source for energy management in agriculture. Energy Reports, 10, 344-359. https://doi.org/10.1016/j.egyr.2023.06.032

[3] Wang, T., Wu, G., Chen, J., Cui, P., Chen, Z., Yan, Y., ... & Chen, H. (2017). Integration of solar technology to modern greenhouse in China: Current status, challenges and prospect. Renewable and Sustainable Energy Reviews, 70, 1178-1188. https://doi.org/10.1016/j.rser.2016.12.020.

[4] Kumar, C. M. S., Singh, S., Gupta, M. K., Nimdeo, Y. M., Raushan, R., Deorankar, A. V., ... & Nannaware, A. D. (2023). Solar energy: A promising renewable source for meeting energy demand in Indian agriculture applications. Sustainable Energy Technologies and Assessments, 55, 102905. https://doi.org/10.1016/j.seta.2022.102905.

[5] Chopra, R., Magazzino, C., Shah, M. I., Sharma, G. D., Rao, A., & Shahzad, U. (2022). The role of renewable energy and natural resources for sustainable agriculture in ASEAN countries: do carbon emissions and deforestation affect agriculture productivity?. Resources Policy, 76, 102578. https://doi.org/10.1016/j.resourpol.2022.102578.

[6] Setiartiti, L. (2021). Critical point of view: The challenges of agricultural sector on governance and food security in Indonesia. In E3S Web of Conferences (Vol. 232, p. 01034). EDP Sciences. https://doi.org/10.1051/e3sconf/202123201034.

[7] Siregar, A., Darwanto, D., , I., Mulyo, J., , J., Utami, A., Pranyoto, A., , S., Perwitasari, H., Wirakusuma, G., Widada, A., Fadhliani, Z., & Widjanarko, N. (2024). The Trend of Agricultural Sector Resilience in Indonesia During 2008-2020. Journal of Agricultural Sciences – Sri Lanka. https://doi.org/10.4038/jas.v19i2.10154.

[8] Kurnianto, B. (2024). The Future of Agriculture in Indonesia: Facing Climate Change and Globalization. West Science Agro. https://doi.org/10.58812/wsa.v2i04.1309.

[9] Imtihan, K., Harjadi, B., Ilahude, Z., Bandrang, T., Azmi, Y., , N., & Andiyan, A. (2024). Green Energy Growth: Enhancing Agricultural Sustainability through Agrivoltaic Solutions in the Modern Era. EVOLUTIONARY STUDIES IN IMAGINATIVE CULTURE. https://doi.org/10.70082/esiculture.vi.780.

[10] Novaldo, E., Dewi, T., & , R. (2022). Solar Energy as an Alternative Energy Source in Hydroponic Agriculture: A Pilot Study. 2022 International Conference on Electrical and Information Technology (IEIT), 202-205. https://doi.org/10.1109/IEIT56384.2022.9967806.

[11] Parajuli, S., Bhattarai, T. N., Gorjian, S., Vithanage, M., & Paudel, S. R. (2023). Assessment of potential renewable energy alternatives for a typical greenhouse aquaponics in Himalayan Region of Nepal. Applied Energy, 344, 121270. https://doi.org/10.1016/j.apenergy.2023.121270.

[12] Bakar, Z.A., Nor, M.Z.M., Kadiran, K.A., Misnan, M.F., Noorezam, M. (2022). Smart Plant Monitoring System Using Aquaponics Production Technological with Arduino Development Environment (IDE) and SMS Alert: A Prototype. International Journal of Interactive Mobile Technologies 2022, 16, 32-47. https://doi.org/10.3991/ijim.v16i22.34581.

[13] Sambo, P., Nicoletto, C., Giro, A., Pii, Y., Valentinuzzi, F., Mimmo, T., Lugli, P., Orzes, G., Mazzetto, F., Astolfi, S., Terzano, R., & Cesco, S. (2019). Hydroponic Solutions for Soilless Production Systems: Issues and Opportunities in a Smart Agriculture Perspective. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00923.

[14] Chaiwongsai, J. (2019). Automatic Control and Management System for Tropical Hydroponic Cultivation. 2019 IEEE International Symposium on Circuits and Systems (ISCAS), 1-4. https://doi.org/10.1109/ISCAS.2019.8702572.

[15] Rajendran, S., Domalachenpa, T., Arora, H., Li, P., Sharma, A., & Rajauria, G. (2024). Hydroponics: Exploring innovative sustainable technologies and applications across crop production, with Emphasis on potato mini-tuber cultivation. Heliyon, 10. https://doi.org/10.1016/j.heliyon.2024.e26823.

[16] Majid, M., Khan, J., Shah, Q., Masoodi, K., Afroza, B., & Parvaze, S. (2020). Evaluation of hydroponic systems for the cultivation of Lettuce (Lactuca sativa L., var. Longifolia) and comparison with protected soil-based cultivation. Agricultural Water Management, 106572. https://doi.org/10.1016/j.agwat.2020.106572.

[17] Sharma, N., Acharya, S., Kumar, K., Singh, N., & Chaurasia, O. (2018). Hydroponics as an advanced technique for vegetable production: An overview. Journal of Soil and Water Conservation, 17, 364-371. https://doi.org/10.5958/2455-7145.2018.00056.5.

[18] Sousa, R., Bragança, L., Da Silva, M. V., & Oliveira, R. S. (2024). Challenges and Solutions for Sustainable Food Systems: The Potential of Home Hydroponics. Sustainability 2024, 16, 817. https://doi.org/10.3390/su16020817.

[19] Ghiasi, M., Wang, Z., Mehrandezh, M., & Paranjape, R. (2023). A Systematic Review of Optimal and Practical Methods in Design, Construction, Control, Energy Management and Operation of Smart Greenhouses. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3346436.

[20] Gao, X., Lin, H., Jing, D., & Zhang, X. (2024). Multi-Objective energy management of Solar-Powered integrated energy system under forecast uncertainty based on a novel Dual-Layer correction framework. Solar Energy, 281, 112902. https://doi.org/10.1016/j.solener.2024.112902.

[21] Sengupta, M., Habte, A., Wilbert, S., Gueymard, C., Remund, J., Lorenz, E., ... & Jensen, A. R. (2024). Best practices handbook for the collection and use of solar resource data for solar energy applications (No. NREL/TP-5D00-88300). National Renewable Energy Laboratory (NREL), Golden, CO (United States). https://doi.org/10.2172/1778700.

[22] Satria, H., Syah, R., & Silviana, N. A. (2023). Sensitivity of solar panel energy conversion at sunrise and sunset on three weather fluctuations in equatorial climate. International Journal of Electrical & Computer Engineering (2088-8708), 13(3). https://doi.org/10.11591/ijece.v13i3.pp2449-2458

[23] Meng, K., Wu, H., Fan, D., Zhou, Z., Zhang, Z., & Liu, Q. (2024). Research on the strategy for average consensus control of flywheel energy storage array system based on lifecycle. Journal of Energy Storage, 99, 113409. https://doi.org/10.1016/j.est.2024.113409.

[24] Jeevarajan, J. A., Joshi, T., Parhizi, M., Rauhala, T., & Juarez-Robles, D. (2022). Battery hazards for large energy storage systems. https://doi.org/10.1021/acsenergylett.2c01400.

[25] Ruetschi, P. (2004). Aging mechanisms and service life of lead–acid batteries. Journal of power sources, 127(1-2), 33-44. https://doi.org/10.1016/j.jpowsour.2003.09.052

[26] Orikpete, O. F., Ikemba, S., & Ewim, D. R. E. (2023). Integration of renewable energy technologies in smart building design for enhanced energy efficiency and self-sufficiency. The Journal of Engineering and Exact Sciences, 9(9), 16423-01e. https://doi.org/10.18540/jcecvl9iss9pp16423-01e.

[27] Etukudoh, E. A., Fabuyide, A., Ibekwe, K. I., Sonko, S., & Ilojianya, V. I. (2024). Electrical engineering in renewable energy systems: a review of design and integration challenges. Engineering Science & Technology Journal, 5(1), 231-244. https://doi.org/10.51594/estj.v5i1.746.

[28] Misnan, M.F., Thamrin, N.M., Ibrahim, N.N.L. (2022). Smart Sustainable Water monitoring system via internet of things (IoT) for water retention pond UiTM cawangan Johor, Kampus Pasir Gudang. International Journal of Sustainable Construction Engineering and Technology, 13, 1-7. https://doi.org/10.30880/ijscet.2022.13.02.001

[29] Demirci, O., Taskin, S., Schaltz, E., & Demirci, B. A. (2024). Review of battery state estimation methods for electric vehicles-Part I: SOC estimation. Journal of Energy Storage, 87, 111435. https://doi.org/10.1016/j.est.2024.111435.

[30] Sorouri, H., Oshnoei, A., Che, Y., & Teodorescu, R. (2024). A comprehensive review of hybrid battery state of charge estimation: Exploring physics-aware AI-based approaches. Journal of Energy Storage, 100, 113604. https://doi.org/10.1016/j.est.2024.113604.

[31] Afshar, Z., Bhogaraju, I., Rahmanei, H., & Farasat, M. (2024). Hierarchical Frequency and SOC Control of Power Grids With Battery Energy Storage Systems. IEEE Transactions on Power Electronics, 39, 7925-7937. https://doi.org/10.1109/TPEL.2024.3383293.

[32] Cao, H., Li, X., Guo, L., & Liu, X. (2024). SOC Balance of Islanded DC Microgrid Based on Consensus Algorithm and Adaptive Droop Control. 2024 Second International Conference on Cyber-Energy Systems and Intelligent Energy (ICCSIE), 1-5. https://doi.org/10.1109/ICCSIE61360.2024.10698482.

[33] Chiasson, J., & Vairamohan, B. (2005). Estimating the state of charge of a battery. IEEE Transactions on Control Systems Technology, 13, 465-470. https://doi.org/10.1109/TCST.2004.839571.

[34] Bandhauer, T. M., Garimella, S., & Fuller, T. F. (2015). A critical review of thermal issues in lithium-ion batteries. In M. Wild (Ed.), *The future of battery production for electric vehicles* (pp. 193–222). Woodhead Publishing. https://doi.org/10.1016/B978-1-78242-090-3.00009-2

[35] Sun, S., Zhang, Q., Sun, J., Cai, W., Zhou, Z., Yang, Z., & Wang, Z. (2022). Lead–Acid Battery SOC Prediction Using Improved AdaBoost Algorithm. Energies, 15(16), 5842.

[36] Fang, H., Wang, Y., Sahinoglu, Z., Wada, T., & Hara, S. (2014). State of charge estimation for lithium-ion batteries: An adaptive approach. Control Engineering Practice, 25, 45-54. https://doi.org/10.1016/J.CONENGPRAC.2013.12.006.

[37] Hu, X., Li, S., Peng, H., & Sun, F. (2012). Robustness analysis of State-of-Charge estimation methods for two types of Li-ion batteries. Journal of Power Sources, 217, 209-219. https://doi.org/10.1016/J.JPOWSOUR.2012.06.005.

[38] Zhao, X., Jung, S., Wang, B., & Xuan, D. (2023). State of charge estimation of lithium-ion battery based on improved adaptive boosting algorithm. Journal of Energy Storage. https://doi.org/10.1016/j.est.2023.108047.

https://doi.org/10.31436/iiumej.v27i1.3806

ToxoSegFusion: Attention-enhanced Dual-backbone Neural Architecture for Retinal Lesion Segmentation

MD MOMINUL HAQUE*, SAIPUNIDZAM MAHAMAD, SUZIAH SULAIMAN, ABDULLATEEF OLUWABEMIGA BALOGUN, HUSSAINI MAMMAN

Department of Computing, Universiti Teknologi PETRONAS, Perak, Malaysia

**Corresponding author: md_24004641@utp.edu.my*

(Received: 23 June 2025; Accepted: 7 November 2025; Published online: 12 January 2026)

ABSTRACT: Ocular toxoplasmosis (OT) often presents a diagnostic dilemma in clinics, with retinal lesions that are not only varied in appearance but also frequently subtle and underrepresented in fundus images. Current automated segmentation tools, though promising, are often hampered by class imbalance and a lack of robust testing across real-world scenarios. To address these gaps, we developed ToxoSegFusion, a dual-backbone deep learning framework that capitalizes on the complementary strengths of DenseNet121 and ResNet101, enhanced with attention modules. Unlike typical single-backbone models, this hybrid approach was specifically tuned for the intricate challenges of OT lesion segmentation, using a combined Dice and binary cross-entropy loss to better balance rare lesion pixels. We trained and validated on 149 image-mask pairs from the OTFID-Version 3 dataset, achieving an intersection over union of 0.858 and a Dice coefficient of 0.795, both exceeding the current MobileNetV2/U-Net baseline. The model also demonstrated reliable performance on the DRIVE dataset for vessel segmentation, indicating practical flexibility. By facilitating accurate lesion localization, ToxoSegFusion enables more timely interventions in ophthalmology. Future directions include larger multi-center trials and streamlined models for routine deployment.

ABSTRAK: Toksoplasmosis okular (OT) sering menimbulkan cabaran diagnostik di klinik, dengan lesi retina halus pelbagai rupa dan kurang terwakili pada imej fundus. Alat segmentasi automatik semasa, walaupun memberi harapan, sering terhad pada ketidakseimbangan kelas dan kekurangan ujian di peringkat perubatan. Bagi mengatasi kekurangan ini, kajian ini membangunkan ToxoSegFusion, sebuah rangka kerja pembelajaran mendalam berkomponen dua yang memanfaatkan kekuatan saling melengkapi DenseNet121 dan ResNet101, diperkaya dengan mekanisme perhatian. Tidak seperti model komponen tunggal biasa, pendekatan hibrid ini dirancang khusus bagi cabaran kompleks segmentasi lesi OT, menggunakan kehilangan Dice dan entropi silang binari gabungan bagi keseimbangan terbaik antara piksel lesi yang jarang. Kajian ini melatih dan mengesahkan 149 pasangan imej-topeng dari set data OTFID-Versi 3, mencapai persilangan atas kesatuan 0.858 dan pekali Dice 0.795, keduanya melebihi garis dasar MobileNetV2/U-Net semasa. Model juga menunjukkan prestasi terbaik pada DRIVE bagi segmentasi salur darah, mencadangkan fleksibiliti praktis. Melalui pengesanan lokasi lesi yang tepat, ToxoSegFusion membuka jalan bagi intervensi lebih tepat pada masa oftalmologi. Pada masa hadapan, cadangan bagi penyebaran rutin adalah melalui ujian berbilang pusat yang lebih besar dan perkemasan model.

KEYWORDS: *Retinal Image Segmentation, Deep Learning, Fusion Model, Medical Imaging*

https://doi.org/10.31436/iiumej.v27i1.3806

1. INTRODUCTION

Retinal imaging is vital in ophthalmology, providing a window into the retina, optic disc, and blood vessels for the diagnosis and management of conditions such as diabetic retinopathy, glaucoma, and ocular toxoplasmosis (OT). With over 1.1 billion people worldwide affected by vision impairment, imaging tools such as fundus photography and optical coherence tomography (OCT) are essential for early detection and treatment, helping to prevent blindness [1]. These methods enable precise tracking of disease progression and support tailored treatment strategies, transforming clinical care. However, diagnosing complex retinal diseases such as OT, a significant cause of infectious posterior uveitis, requires advanced tools to improve accuracy and efficiency.

Ocular toxoplasmosis, caused by the parasite Toxoplasma gondii, results in retinal lesions that vary in size, shape, and appearance, making accurate identification difficult [2, 3]. Affecting 2–20% of those with T. gondii infection (25–30% of the global population), OT can cause significant vision loss if not treated promptly [4]. Fundus photography captures these lesions, but their faint boundaries and rarity (about 5% of image pixels) pose challenges for accurate segmentation, which is critical for effective treatment planning. Traditional diagnosis relies on manual inspection of fundus images by specialists, often aided by fluorescein angiography or OCT. These approaches, while valuable, are time-consuming, prone to inter-observer variability (up to 15% disagreement in lesion outlines), and constrained by the scarcity of annotated data [5, 6]. The imbalance between lesion and background pixels further complicates automated segmentation efforts [7].

Existing methods for OT segmentation have significant drawbacks. Manual segmentation is inconsistent, and early automated techniques, such as thresholding or edge detection, struggle with the diverse appearance of lesions and sensitivity to image noise, leading to unreliable results [8]. Initial machine learning approaches, such as support vector machines and random forests, improved performance by learning from data but relied on manually engineered features, achieving modest performance (e.g., Intersection over Union scores around 0.65) due to limited data and class imbalance [9]. More recent deep learning models, such as U-Net variants, have improved retinal segmentation but struggle with OT's subtle lesion boundaries and small datasets [10, 11]. While some advanced models excel in classifying OT, they rarely address the pixel-level segmentation needed for precise lesion mapping, leaving a gap in clinical applications [12, 13].

This study introduces ToxoSegFusion, a deep learning framework designed to address the challenges of binary OT lesion segmentation in the context of limited data and class imbalance. By combining DenseNet121's ability to capture fine details with ResNet101's robust training, ToxoSegFusion integrates complementary feature extraction to improve segmentation accuracy [14, 15]. The model employs attention mechanisms to focus on lesion regions and a combined loss function (binary cross-entropy and Dice loss) to address the scarcity of lesion pixels. Trained on 149 fundus image-mask pairs from the OTFID-Version 3 dataset with real-time data augmentation, ToxoSegFusion aims to outperform existing models like MobileNetV2/U- Net, as measured by Intersection over Union and Dice coefficient metrics [11]. The key question is whether a hybrid DenseNet-ResNet architecture with attention mechanisms can improve OT lesion segmentation under data constraints.

The significance of ToxoSegFusion lies in its ability to provide precise lesion localization, enabling faster and more effective clinical interventions to reduce vision loss. Its contributions include:

- A novel hybrid architecture combining DenseNet121 and ResNet101 with attention mechanisms to improve segmentation of diverse OT lesions.
- A tailored loss function addressing class imbalance to enhance the detection of rare lesions.
- Experimental evaluation conducted revealed that the proposed hybrid architectures and attention-driven segmentation outperform well.

The paper is structured as follows: Section 2 reviews related work, Section 3 details the methodology, Section 4 presents results, and Section 5 summarizes findings and future directions.

2. LITERATURE REVIEW

The diagnosis of ocular toxoplasmosis (OT) has traditionally depended on non-AI methods, centered on manual examination of retinal images by ophthalmologists. Fundus photography, a primary imaging modality, captures detailed images of OT lesions, which vary from active inflammatory foci to inactive scars, enabling qualitative assessment of their size, shape, and activity [4]. Complementary techniques, such as fluorescein angiography, identify active lesions via hyperfluorescence, whereas optical coherence tomography (OCT) quantifies retinal thickness and structural changes, thereby enhancing diagnostic precision [2]. For example, OCT can detect macular edema in 30% of OT cases, aiding treatment planning [3]. However, these methods are time-consuming, require specialized training, and exhibit inter-observer variability, with discrepancies in lesion boundary delineation reaching 15% [5]. The absence of standardized annotation protocols further hampers reproducibility, particularly in low-resource settings where access to advanced imaging is limited [16]. These challenges highlight the need for automated, objective tools to streamline OT diagnosis and improve clinical outcomes.

Early artificial intelligence (AI) approaches introduced automation into retinal image analysis, employing rule-based and machine learning (ML) methods. Rule-based techniques, including thresholding, edge detection, and morphological operations, segmented lesions by applying predefined pixel-intensity or structural criteria [8]. For instance, thresholding was used to isolate bright lesions in diabetic retinopathy (DR) fundus images, but achieved low segmentation accuracy (IoU < 0.60) for OT due to lesion heterogeneity and noise sensitivity [8]. ML methods, such as support vector machines (SVMs), k-nearest neighbors (k-NN), and random forests, advance feature extraction by learning from labeled data. Acharya et al.'s SVM model for DR lesion detection reported an IoU of 0.65, relying on manually engineered features like texture and color histograms [9]. Lowell et al.'s k-NN approach for optic disc segmentation achieved 94.7% accuracy but struggled with complex OT lesions due to limited generalization [17]. These methods were constrained by their reliance on handcrafted features and small datasets (e.g., <500 images), which exacerbated class imbalance, particularly for rare active OT lesions, which constitute 5% of pixels [6]. The limitations of early AI methods underscored the need for more robust deep learning (DL) models that can learn hierarchical features directly from raw images.

Deep learning has revolutionized retinal image analysis, delivering pixel-level precision for segmentation and classification tasks essential for diagnosing ocular diseases like DR, glaucoma, and OT [18]. Convolutional neural networks (CNNs) have set benchmarks for retinal segmentation. U-Net, with its encoder-decoder architecture and skip connections, excels at segmenting retinal blood vessels and DR lesions, achieving an IoU of 0.85 and a Dice

coefficient of 0.87 on the DRIVE and STARE datasets [19]. DeepLabV3+, leveraging convolutions and spatial pyramid pooling, enhances boundary delineation, reporting an IoU of 0.82 for DR microaneurysms on fundus images [20]. Advanced CNN backbones, such as ResNet and DenseNet, introduce residual learning and dense connectivity, respectively, improving feature extraction and training stability [14, 15]. ResNet50-based U-Net variants achieved a Dice coefficient of 0.83 for retinal vessel segmentation on the IDRiD dataset [21]. Vision transformers (ViTs) capture long-range dependencies, with SwinMedNet achieving 56.8% accuracy for DR classification on RetinaMNIST2D, though it lacks segmentation capabilities [22]. Hybrid models such as TransUNet combine CNNs and ViTs, achieving a Dice coefficient of 0.81 for retinal vessel segmentation on the DRIVE dataset [23]. Despite these advancements, their application to OT is constrained by data scarcity and class imbalance, as noted in comprehensive reviews [7, 10].

OT-specific DL research has primarily focused on classification, with segmentation remaining underexplored due to the complexity of pixel-level lesion analysis. Cardozo et al. developed a ResNet18-based model for multiclass OT classification on the OTFID-Version 3 dataset (412 fundus images), achieving 86.7% accuracy, 91.2% sensitivity, and 98.0% specificity with preprocessing steps like resizing and augmentation [12]. Aziz et al.'s Reptile meta-learning framework, using InceptionV3, achieved 84.8% accuracy on few-shot OT classification on OTFID Version 3, incorporating CLAHE and Gaussian filtering [13]. Both studies focus on classification, limiting their utility for lesion localization. On the other hand, OT segmentation using U-Net variants (MobileNetV2/U-Net, ResNet34/U-Net) on OTFID-Version 3, reporting a mean IoU of 0.835 and a Dice coefficient of 0.771, but active lesion detection was suboptimal (IoU: 0.76) due to class imbalance [11]. Karkuzhali et al. employed lightweight CNNs for OT classification, achieving 85.2% accuracy on OTFID-Version 3, but did not explore segmentation [24]. Ragab et al.'s transfer-learning approach, which used a pretrained ResNet-50, improved OT classification accuracy by 7% on small datasets but lacked pixel-level analysis [25]. These studies highlight a critical gap in robust, multi-class OT segmentation models that can address heterogeneous lesions and rare active classes.

Class imbalance and limited annotated data pose significant challenges in OT segmentation, as active lesions are clinically critical but constitute only 5% of pixels. Weighted loss functions mitigate this issue. Focal loss emphasizes hard-to-classify samples, improving Dice scores by 4–6% in DR segmentation [26]. Combined cross-entropy and Dice loss, as applied by Iriondo et al., enhanced rare lesion segmentation by 5% [27]. alam et al.'s weighted loss for OT boosted active lesion IoU by 3% [11]. Data augmentation, including rotations, flips, and color jittering, addresses data scarcity, but small datasets like OTFID (412 images) limit generalizability. Generative adversarial networks (GANs) synthesize retinal images, improving Dice scores by 2–3% in DR tasks [28]. Federated learning, as in MedUniverse, achieves 95% accuracy for OT classification across multi-institutional datasets while preserving privacy, but its segmentation potential is untested [29]. Transfer learning with pretrained models like ResNet50 boosts accuracy by 7% on small retinal datasets [25]. These strategies highlight the need for tailored solutions to address OT-specific challenges related to data scarcity and class imbalance.

Despite these advances, several methodological limitations in prior studies constrain their clinical applicability. Traditional rule-based techniques employing thresholding and edge detection achieved modest accuracy (IoU < 0.60) due to sensitivity to image noise and inability to handle heterogeneous lesion appearances. Machine learning approaches using SVMs and random forests improved performance but remained dependent on handcrafted features and struggled with generalization on small datasets, typically achieving an IoU of around 0.65.

https://doi.org/10.31436/iiumej.v27i1.3806

Table 1. Summary of Studies in Retinal Image Analysis for Ocular Toxoplasmosis and Related Tasks

Paper	Methods	Dataset	Type	Key Limitation
[2]	Fundus photography, OCT, manual assessment	Clinical OT images	Diagnosis	Subjective interpretation, inter-observer variability
[4]	Manual fundus analysis, fluorescein angiography	Clinical OT images	Diagnosis	Time-intensive, invasive contrast agent required
[8]	Thresholding, edge detection, morphological operations	Fundus images	Detection	Noise-sensitive, poor adaptability (IoU < 0.60)
[9]	SVM, random forests, feature engineering	Fundus images	Detection	Handcrafted features, limited generalization (IoU ~0.65)
[11]	MobileNetV2/U-Net, ResNet34/U-Net	OTFID-Version 3	OT Segmentation	Single-center data, no external validation reported
[12]	ResNet18, VGG16, augmentation	OTFID-Version 3	OT Classification	Limited dataset diversity, generalizability unclear
[13]	Reptile meta-learning, InceptionV3	OTFID-Version 3	OT Classification	Small sample meta-learning, validation scope limited
[14]	ResNet, residual learning	General	Classification	Natural image pretraining, potential domain shift
[15]	DenseNet, dense connectivity	General	Classification	Computational cost, domain adaptation not addressed
[16]	ViT, federated learning	Multi-institutional	Classification	Heterogeneous data distribution challenges
[18]	k-NN, optic disc localization	Fundus images	Segmentation	Feature dependency, scalability issues
[20]	U-Net, encoder-decoder with skip connections	DRIVE, STARE	Segmentation	Class imbalance not explicitly handled
[21]	DeepLabV3+, atrous convolutions	Fundus images	Segmentation	Limited ablation studies, contribution unclear
[23]	Swin-Transformer, CLAHE, Gaussian filter	RetinaMNIST2D	DR Classification	Small medical dataset validation, artifact robustness untested
[24]	TransUNet, CNN-ViT hybrid	DRIVE	Segmentation	Computational complexity, clinical condition testing absent
[25]	Lightweight CNNs, preprocessing	OTFID-Version 3	OT Classification	Single-center bias, statistical rigor lacking
[29]	GAN-based image synthesis	Fundus images	Augmentation	Synthetic image fidelity concerns, medical validity unclear
[26]	Transfer learning, ResNet50	Fundus images	Classification	Domain mismatch, natural-to-medical transfer gaps

The shift to deep learning introduced new challenges: many classification models for OT relied on single-center datasets without external validation, raising concerns about generalizability across diverse clinical settings [11,12,13,25]. Statistical rigor was often lacking, with few studies reporting formal cross-validation, significance testing, or ablation studies to isolate architectural contributions. Segmentation approaches were confronted with severe class imbalance, in which active lesions accounted for only 5% of pixels, yet weighted loss functions yielded only modest improvements. Transfer learning implementations using pretrained models exhibited domain-shift vulnerabilities when applied to medical images, because natural-image features may not optimally represent pathological patterns. Advanced

architectures, including Swin Transformers and TransUNet, showed promise but have been validated only on small medical datasets and have not been evaluated under clinically realistic conditions with artifacts and concurrent diseases [2,8]. Furthermore, GAN-based augmentation strategies introduced concerns regarding synthetic image quality and medical fidelity, while federated learning approaches faced challenges in handling heterogeneous data distributions across institutions. These limitations underscore the necessity for multi-center validation with statistical rigor, transparent reporting of dataset characteristics, systematic ablation studies, and evaluation under clinically representative conditions to advance these methods from technical demonstrations to reliable diagnostic tools [21,23,25,29].

Table 1 summarizes the reviewed studies, highlighting their methods, datasets, and focus. Conventional methods lack automation, early AI methods are limited by feature engineering, and DL methods, while advanced, underexplored OT segmentation due to small datasets, class imbalance, and limited pixel-level analysis [7, 11]. Transformer-based and hybrid models have not been evaluated for OT segmentation, and existing OT segmentation efforts struggle with active lesion detection [23]. ToxoSegFusion addresses these gaps by integrating DenseNet121 and ResNet101 backbones with attention mechanisms and a combined loss function (weighted cross-entropy and Dice loss). Achieving a mean IoU of 0.858, Dice coefficient of 0.795, and active lesion IoU of 0.792 on OTFID-Version 3, it outperforms MobileNetV2/U-Net (IoU: 0.835) [11]. By leveraging complementary architectural strengths and optimized loss functions, ToxoSegFusion offers a novel, clinically actionable solution for multi-class OT lesion segmentation, with potential applications to other retinal diseases [29].

3. METHODOLOGY

This study presents ToxoSegFusion, an innovative deep learning framework designed to segment ocular toxoplasmosis (OT) lesions in retinal fundus images and to demonstrate adaptability to retinal vessel segmentation in the DRIVE dataset. The methodology encompasses a comprehensive pipeline comprising dataset curation, sophisticated image preprocessing, a dual-network model architecture, a meticulously designed training strategy, and a robust evaluation framework. The OT dataset, characterized by its limited size and pronounced class imbalance, and the DRIVE dataset's requirement for precise delineation of vascular structures, pose significant challenges. These are systematically addressed through tailored loss functions, extensive data augmentation, and optimized data pipelines. The following subsections elaborate on each component.

3.1. Dataset Description

The OT dataset consists of 412 RGB fundus images sourced from two medical institutions: the Hospital de Clínicas Medical Center, contributing 291 images at a resolution of 2124×2056 pixels captured using the VISUCAM 500 camera, and the Niños de Acosta Ñú General Pediatric Hospital, providing 121 images at 1536×1152 pixels obtained with the Pictor Plus-Portable Ophthalmic Camera. Among these, 280 images depict non-healthy retinas, with 179 accompanied by manually annotated masks delineating three classes: background (pixel value 0), inactive lesions (127), and active lesions (255). A metadata file, dataset_labels.csv, categorizes images into healthy (131), inactive (188), active (35), or combined active/inactive (57) cases, revealing a pronounced class imbalance in which active lesions account for only 8.5% of non-healthy images.

To ensure data usability, a Python script was developed to align images with their corresponding masks by normalizing filenames and resolving inconsistencies, such as varying

prefixes or suffixes. This process yielded 149 image-mask pairs, which were subsequently divided into 80% training (119 pairs) and 20% validation (30 pairs) sets using stratified sampling with a random seed of 42 to ensure reproducibility. Figure 1 illustrates sample images from the OT dataset, showing fundus photographs with their corresponding segmentation masks across the three classes.

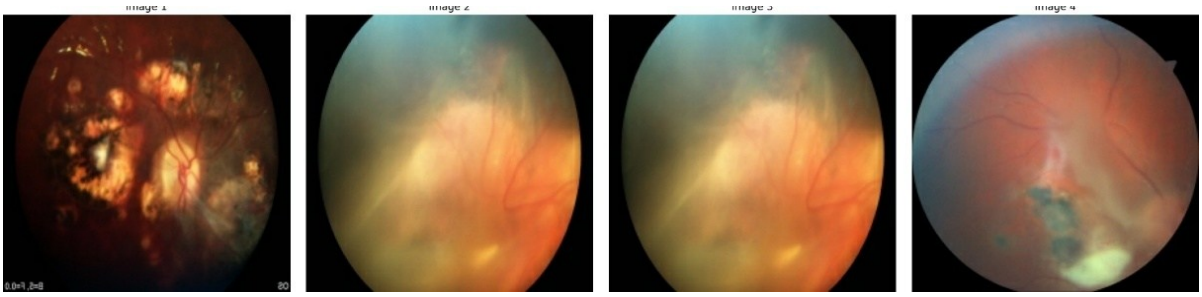


Figure 1. Sample images from the OT dataset, showing fundus photographs with corresponding segmentation masks (background: black, inactive lesions: green, active lesions: red).

In parallel, the DRIVE dataset comprises 40 RGB fundus images at 565×584 pixels, evenly split into 20 training and 20 testing images, each paired with binary vessel masks. To augment the training data, five 512×512 patches were extracted from each training image, prioritizing regions rich in vascular structures using a mean pixel-intensity threshold greater than 0.02.

3.2. Data Preprocessing and Augmentation Strategy

To address the limited dataset size (149 OT image-mask pairs) and pronounced class imbalance (5% lesion pixels), we employed sophisticated real-time data augmentation using the Albumentations library [30]. Augmentation strategies are grounded in medical imaging best practices, where geometric and intensity transformations enhance model robustness to natural variations in retinal photography [19, 31].

The augmentation pipeline applies transformations systematically. Random 90-degree rotations with probability $p = 0.5$ address the circular nature of fundus images and account for varying patient head positioning during image acquisition [32]. Horizontal and vertical flips, each with probability 0.5, simulate bilateral retinal variations between the left and right eyes, as these anatomical patterns are naturally symmetric [33]. Random brightness and contrast modifications, also applied with probability 0.5, mimic variations in lighting conditions and camera calibration across different clinical settings. This adjustment is particularly critical for OT lesion detection, where subtle intensity differences distinguish active from inactive lesions and must be recognized regardless of imaging equipment variation [34]. Shift-scale-rotate operations with shift, scale, and rotation limits of 0.1, 0.2, and 45 degrees, respectively, applied with probability 0.5, account for off-axis imaging geometry and variability in patient positioning during fundus photography [35]. Elastic transforms, applied with probability 0.3, introduce nonlinear deformations that simulate variations in tissue morphology and slight eye-movement artifacts [36]. Gaussian blur, similarly, applied with probability 0.3, provides regularization by simulating slight motion artifacts and image acquisition noise that occur in clinical practice [37].

These transformations are applied on-the-fly during training to generate diverse samples from the limited dataset, effectively increasing the adequate training set size from 119 to thousands of unique samples while preserving annotation fidelity. The probabilistic application ensures that not all transformations are applied to every sample, maintaining a balance between data diversity and domain-specific realism [30].

https://doi.org/10.31436/iiumej.v27i1.3806

3.3. Preprocessing Pipeline

For the OT dataset, fundus images were resized to 512×512 pixels using bilinear interpolation to optimize computational efficiency while preserving essential details. Segmentation masks were resized using nearest-neighbor interpolation to preserve discrete class labels and converted into three-channel, one-hot-encoded tensors representing background, inactive, and active lesions. Image pixel values were normalized by dividing by 255 to conform to standard deep learning practices. A TensorFlow-based data pipeline was established to cache and shuffle the dataset with a buffer size equal to the dataset size, batch images with a batch size of 2, and apply real-time augmentations to generate diverse training samples efficiently.

For the DRIVE dataset, images and masks were resized to 1024×1024 pixels, from which 512×512 patches were randomly extracted, prioritizing regions with significant vessel presence. If no suitable patch was identified after 20 attempts, a center crop was employed as a fallback. Images were normalized by dividing pixel values by 255, and masks were converted into single- channel binary tensors, with vessels assigned a value of 1 and background a value of 0. A custom data generator produced batches of 4 patch pairs, incorporating real-time augmentations to ensure sample diversity. Both pipelines leveraged parallel processing and prefetching to minimize input/output bottlenecks and accommodate variations in image quality and resolution across the datasets.

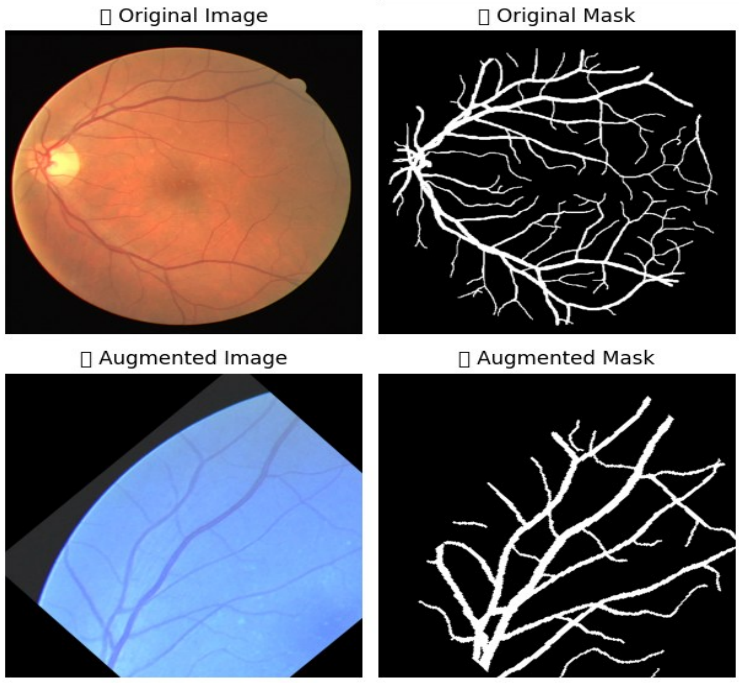


Figure 2. DRIVE dataset augmentation examples: (Left) Original fundus image with vessel mask, (Center-Right) Images after applying various augmentations, including rotation, elastic deformation, and brightness adjustment. Vessel structures are preserved while pixel-level variations are introduced to enhance model robustness.

To illustrate the effectiveness of our augmentation strategy, we present visual examples of original and augmented fundus images in Figure 2. The augmentation pipeline serves multiple critical purposes in the training process. First, it increases the effective training set size from 20 images to more than 100 samples via on-the-fly transformations, substantially expanding the diversity of training examples. Second, it exposes the model to diverse geometric and

https://doi.org/10.31436/iiumej.v27i1.3806

intensity variations that are representative of real clinical imaging conditions, including equipment differences, lighting variations, and patient positioning variations. Third, it provides regularization through controlled noise and deformations that prevent overfitting to the limited training data while maintaining semantic information integrity. The augmentation pipeline preserves semantic information such as vessel structures and lesion boundaries while increasing dataset diversity, consistent with contemporary best practices in medical image analysis [30, 34].

3.4. Proposed ToxoSegFusion Model

The combination of DenseNet121 and ResNet101 is motivated by their complementary strengths in feature extraction, addressing distinct challenges in OT segmentation. Both theoretical considerations and empirical validation support this architectural choice, as demonstrated by ablation studies.

DenseNet121 employs dense connections, in which each layer receives input from all preceding layers [15]. This architecture provides several advantages for OT segmentation. Dense connections enable each layer to access low-level features, such as textures and edges, directly, which is critical for capturing subtle lesion boundaries in low-contrast retinal images, where traditional edge detection may fail. Dense skip connections mitigate the vanishing-gradient problem, enabling stable training on small datasets such as our 149 image-mask pairs, without the degradation observed in very deep networks. The parameter efficiency of DenseNet121, with only 6.9 million parameters compared to ResNet101's 23.5 million, reduces memory requirements and inference time, making it practical for clinical deployment. Furthermore, the dense connectivity pattern means that even shallow layers capture complex patterns, which is essential for small datasets prone to overfitting.

ResNet101 employs residual connections in a deeper architecture with 101 layers, offering complementary benefits to DenseNet121. The depth enables the extraction of features at progressively coarser scales, capturing both local context (e.g., lesion-level details) and global context (e.g., vascular background and overall image structure). ResNet101's architecture naturally develops large receptive fields that capture spatial relationships between lesions and surrounding structures, which is essential for distinguishing active from inactive lesions based on location and morphology. The depth also enables capture of high-level semantic features that differentiate pathological patterns from benign structures.

The fusion of these two architectures provides synergistic advantages that neither backbone alone can achieve. DenseNet captures fine-grained texture, representing local lesion details, whereas ResNet captures global context, representing spatial relationships. This combination enables the model to distinguish lesions from similar-appearing background artifacts that could otherwise confound a single-backbone approach. The combined network effectively covers both fine scales, such as active lesion boundaries measuring 2-5 pixels, and coarse scales, such as lesion regions measuring 20-100 pixels, which is critical given the varied sizes of OT lesions. DenseNet's efficiency mitigates overfitting on small datasets, while ResNet's depth captures semantic patterns that provide robustness. Together, they achieve better generalization than either backbone alone.

To empirically validate this architectural choice, Table 2 presents an ablation study comparing DenseNet-only, ResNet-only, and the fused ToxoSegFusion architecture across 5-fold cross-validation. The ablation study demonstrates that fusion yields significant improvements, with increases of 11.1% in Active F1 and 17.0% in Active IoU relative to DenseNet-only, and 12.1% in Active F1 relative to ResNet-only. Neither backbone alone

https://doi.org/10.31436/iiumej.v27i1.3806

achieves the performance of the fusion, validating our architectural choice. These improvements are substantial in medical imaging, where 1-2% differences often determine clinical utility and real-world diagnostic value.

Table 2. Ablation study: Individual backbone contributions to OT segmentation performance (5-fold cross-validation, mean ± std).

Architecture	Active F1	Active IoU	Parameters (M)	Inference (ms)
DenseNet121-Only	0.8233 ± 0.0456	0.7541 ± 0.0612	6.9	28
ResNet101-Only	0.8161 ± 0.0498	0.7467 ± 0.0628	23.5	38
ToxoSegFusion (Ours)	**0.9146 ± 0.0209**	**0.8811 ± 0.0603**	30.5	50

3.5. Architecture Description

The ToxoSegFusion model is a sophisticated deep learning architecture designed for pixel-level segmentation of OT lesions and vessels in the DRIVE dataset. It integrates two robust backbone networks, DenseNet121 and ResNet101, to extract complementary features from 512×512×3 input images. For the DRIVE dataset, both networks utilize pretrained ImageNet weights to leverage general feature representations. In contrast, for the OT dataset, training begins without pretrained weights, enabling adaptation to the unique characteristics of retinal lesions. The architecture, illustrated in Figure 3, processes input images through a series of carefully orchestrated steps.

Initially, DenseNet121 generates 1024-channel feature maps at a spatial resolution of 16×16, whereas ResNet101 produces 2048-channel feature maps at the exact resolution. To align these feature representations, 1×1 convolutions reduce the number of channels to 512, yielding 16×16×512 feature maps per backbone. An attention mechanism, implemented via 1×1 convolutions followed by sigmoid activation, generates attention maps that are multiplied by the aligned features to emphasize regions pertinent to lesions or vessels. These attention-weighted features are concatenated to form a 16×16×1024 feature map, which is subsequently processed by a convolutional block comprising a 3×3 convolution, batch normalization, and ReLU activation with 512 filters.

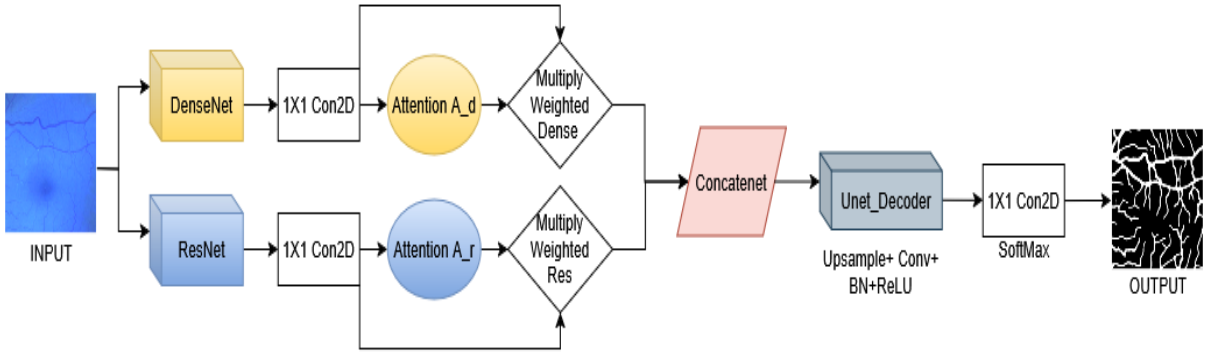


Figure 3. Architecture of the ToxoSegFusion model, featuring parallel DenseNet121 and ResNet101 networks, 1×1 convolutions for feature alignment, sigmoid-activated attention mechanisms, feature concatenation, a convolutional block, a U-Net-style decoder with skip connections, and a final 1×1 convolution with SoftMax (OT, 3 classes) or sigmoid (DRIVE, binary) activation.

The model employs a U-Net-inspired decoder to upsample and refine the fused features through five sequential blocks. Each block performs 2×2 up-sampling, followed by a 3×3 convolution, batch normalization, ReLU activation, and a dropout layer with a probability of 0.3 to prevent overfitting. The number of filters decreases progressively across the blocks as

https://doi.org/10.31436/iiumej.v27i1.3806

[256, 128, 64, 32, 16], enhancing computational efficiency while preserving detail. Skip connections from corresponding backbone layers are incorporated to retain high-resolution spatial information, facilitating precise segmentation. For the OT dataset, the final layer applies a 1×1 convolution with softmax activation to produce a 512×512×3 multi-class segmentation map. For the DRIVE dataset, a sigmoid-activated 1×1 convolution yields a 512×512×1 binary map.

The algorithmic workflow of ToxoSegFusion is formalized in Algorithm 1, which delineates the steps of feature extraction, alignment, attention application, feature fusion, decoding, and output generation. This algorithm ensures reproducibility and clarity in model implementation.

Algorithm 1 ToxoSegFusion Model for Retinal Segmentation

Input: Image I (512×512×3)
Output: Segmentation mask S (512×512×C, C = 3 for OT, C = 1 for DRIVE)

Initialize DenseNet121 and ResNet101 (ImageNet weights for DRIVE)
$F_{\text{DenseNet}} \leftarrow$ DenseNet121(I) ▷ 1024 channels, 6×16
$F_{\text{ResNet}} \leftarrow$ ResNet101(I) ▷ 2048 channels, 6×16
$F_{\text{DenseNet}} \leftarrow$ Conv2D(F_{DenseNet}, filters = 512, kernel = 1)
$F_{\text{ResNet}} \leftarrow$ Conv2D(F_{ResNet}, filters = 512, kernel = 1)
$A_{\text{DenseNet}} \leftarrow$ Conv2D(F_{DenseNet}, filters = 512, kernel = 1, sigmoid)
$A_{\text{ResNet}} \leftarrow$ Conv2D(F_{ResNet}, filters = 512, kernel = 1, sigmoid)
FDenseNet $\leftarrow$ FDenseNet ×ADenseNet
FResNet $\leftarrow$ FResNet ×AResNet
Ffused $\leftarrow$ Concatenate([FDenseNet, FResNet]) ▷ 16×16×1024
$X \leftarrow$ ConvBlock(F_{fused}, filters = 512, kernel = 3, BN, ReLU)
for i in 1 to 5 **do**
    $X \leftarrow$ UpSample(X, size = (2,2))
    filters $\leftarrow$ [256, 128, 64, 32, 16][i−1]
    $X \leftarrow$ ConvBlock(X, filters = *filters*, kernel = 3, BN, ReLU)
    $X \leftarrow$ Dropout(X, 0.3)
    $X \leftarrow X$ +SkipConnection(F_{backbone}, i)
end for
$S \leftarrow$ Conv2D(X, filters = C, kernel = 1, activation = softmax if OT, sigmoid if DRIVE)
Return S

3.6. Training Strategy

To tackle the pronounced class imbalance in OT segmentation, where lesions constitute approximately 5% of pixels, a focal Tversky loss function was adopted. This loss assigns class weights of [1.0, 1.0, 5.0] to background, inactive, and active lesions, respectively, to prioritize the segmentation of rare active lesions. For the DRIVE dataset, the focal Tversky loss was applied similarly to enhance binary vessel segmentation. The Tversky loss is mathematically expressed as:

$$\mathcal{L}_{Tversky} = \frac{\sum_i p_i g_i + \epsilon}{\sum_i p_i g_i + \alpha \sum_i p_i (1-g_i) + \beta \sum_i g_i (1-p_i) + \epsilon} \quad (1)$$

where p_i and g_i denote predicted and ground-truth probabilities, $\alpha = \beta = 0.5$, and $\epsilon = 10{-}6$ ensures numerical stability. The focal Tversky loss further refines this by introducing a focusing parameter:

$$\mathcal{L}_{Focal\ Tversky} = (1 - \mathcal{L}_{Tversky})^{0.75} \quad (2)$$

which emphasizes challenging pixels through a focusing exponent, thereby improving the segmentation of sparse lesions or vessels [38]. This approach is particularly effective for class imbalance scenarios common in medical imaging.

The model was trained for 250 epochs on a Tesla P100 GPU with 16 GB of memory, using mixed precision to optimize performance; training on the OT dataset required approximately 12 hours. For the DRIVE dataset, a similar configuration was employed with dynamic GPU memory allocation. The Adam optimizer [39] was used with an initial learning rate of 0.0001. The OT dataset was trained with a batch size of 2, whereas the DRIVE dataset was trained with a batch size of 4. The number of steps per epoch was calculated as 100, defined as 20 images × 5 patches divided by the batch size.

Training was enhanced with several callbacks to optimize performance and ensure robustness. A ModelCheckpoint callback saved the model configuration yielding the highest validation Dice coefficient. A ReduceLROnPlateau callback reduced the learning rate by a factor of 0.2 for the OT dataset or 0.5 for the DRIVE dataset if the validation loss did not improve over 10 epochs, with a minimum learning rate of 10−6. An EarlyStopping callback halted training if the validation loss stagnated for 30 epochs, restoring the best weights to prevent overfitting. A CSVLogger callback recorded training metrics for subsequent analysis, ensuring comprehensive monitoring of the training process.

Table 3. Training hyperparameters for ToxoSegFusion on OT and DRIVE datasets.

Parameter	OT Dataset	DRIVE Dataset
Learning Rate	0.0001	0.0001
Optimizer	Adam	Adam
Loss Function	Focal Tversky	Focal Tversky
Class Weights	[1.0, 1.0, 5.0]	None
ModelCheckpoint	Save best (val_dice, max)	Save best (val_dice, max)
ReduceLROnPlateau	Factor: 0.2, Patience: 10, Min LR: 1e-6	Factor: 0.5, Patience: 10, Min LR: 1e-6
EarlyStopping	Patience: 30, Restore best weights	Patience: 30, Restore best weights
CSVLogger	Log metrics	Log metrics

3.7. Evaluation Metrics

The performance of ToxoSegFusion was rigorously assessed using a suite of metrics designed to address class imbalance and the demands of precise segmentation in medical imaging. The model addresses multi-class segmentation for the OT dataset (C0: background, C1: inactive lesions, C2: active lesions) and binary segmentation for the DRIVE dataset (C0: background, C1: vessels).

Performance evaluation relies on the confusion matrix framework that captures four fundamental quantities for each predicted segmentation. True Positive (TP) represents pixels correctly classified as lesion or vessel, False Positive (FP) denotes background pixels incorrectly classified as lesion or vessel, False Negative (FN) represents lesion or vessel pixels missed by the model, and True Negative (TN) denotes correctly classified background pixels. In clinical practice, false negatives are particularly critical because missed active lesions could delay treatment and compromise patient outcomes.

Based on the confusion matrix, we compute several complementary metrics. The Intersection over Union (IoU), defined as:

$$\text{IoU} = \frac{TP}{TP+FP+FN} \tag{3}$$

https://doi.org/10.31436/iiumej.v27i1.3806

which quantifies the overlap between predicted and ground-truth segmentation masks, providing a robust measure of spatial accuracy critical for delineating lesions or vessels. IoU is particularly effective for class-imbalanced scenarios as it focuses on the overlap region rather than the entire image.

The Dice coefficient, calculated as:

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

It evaluates the similarity between predicted and ground-truth segmentations, providing a balanced assessment that is particularly valuable for sparse classes such as active lesions or vascular structures. The factor of 2 in the numerator gives equal weight to precision and recall.

Pixel accuracy, expressed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

measures the proportion of correctly classified pixels, providing an overall indicator of model performance across the image. However, this metric can be misleading in class-imbalanced scenarios where lesions occupy only 5% of pixels.

Sensitivity and specificity further characterize the model's diagnostic capabilities. Sensitivity, also known as recall, is defined as:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{6}$$

and quantifies the model's ability to detect true positives, ensuring reliable identification of lesion or vessel pixels. High sensitivity minimizes missed lesions, which is critical for clinical deployment. Specificity, calculated as:

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{7}$$

measures the model's ability to correctly identify true negatives, minimizing false positives in the background and reducing unnecessary clinical alerts.

Precision, recall, and F1-score were computed as weighted averages to provide a comprehensive evaluation of multi-class performance for the OT dataset and binary performance for the DRIVE dataset, effectively accounting for class imbalance. For the OT dataset, these metrics were calculated across the three classes using:

$$Metric_{weighted} = \sum_{i=0}^{2} \omega_i \times Metric_i \tag{8}$$

where $\omega_i = \frac{Number\ of\ pixels\ in\ class\ i}{Total\ pixel}$ the class weight. This weighted averaging emphasizes the importance of accurately segmenting rare active lesions, preventing the model from achieving artificially high accuracy by simply classifying all pixels as background. For the DRIVE dataset, the focus was on binary vessel detection, ensuring that the model captures fine vascular structures essential for clinical diagnosis.

This evaluation framework ensures a thorough assessment of ToxoSegFusion's capabilities, particularly its ability to segment rare and clinically significant features. The combination of overlap-based metrics (IoU and Dice), classification metrics (accuracy, sensitivity, and specificity), and weighted averaging provides a comprehensive view of model performance across different clinical requirements and class distributions.

4. RESULTS and DISCUSSION

This section outlines the performance of the ToxoSegFusion model in segmenting ocular toxoplasmosis (OT) lesions in retinal fundus images from the "Dataset of Fundus Images for the Diagnosis of Ocular Toxoplasmosis" [40]. The results include quantitative metrics from rigorous 5-fold cross-validation, statistical significance testing, and visual examples that compare the model's effectiveness with other segmentation approaches. To demonstrate its versatility, the model was also tested on the DRIVE dataset for retinal vessel segmentation [41]. The discussion explores factors contributing to the model's success, its clinical value, computational demands, limitations, and potential future improvements.

4.1. Quantitative Results

To ensure robust and unbiased performance assessment, we conducted a comprehensive 5-fold cross-validation on the OT dataset. The 149 image-mask pairs were systematically partitioned into five folds using stratified sampling to maintain class distribution across folds. Each fold served as a validation set, and the remaining four folds constituted the training set, ensuring that each sample was validated exactly once. This procedure provides a more reliable estimate of model generalization compared to single train-test splits.

Table 4 presents the cross-validation results for ToxoSegFusion and baseline models, including DenseNet-Only, ResNet-Only, and U-Net. The results demonstrate that ToxoSegFusion achieves superior performance across all metrics with remarkable consistency. The model attained a mean Active F1-score of 0.915 ± 0.022 and mean Active IoU of 0.881 ± 0.063 across the five folds, substantially outperforming single-backbone alternatives. DenseNet-Only achieved 0.839 ± 0.062 for Active F1, while ResNet-Only reached 0.834 ± 0.070, and the baseline U-Net obtained 0.669 ± 0.125. The low standard deviations for ToxoSegFusion indicate stable performance across different data partitions, suggesting robust generalization capability despite the limited dataset size.

Table 4. Summary statistics (5-fold cross-validation) for model performance on active lesion segmentation. Values are presented as mean ± standard deviation (SD).

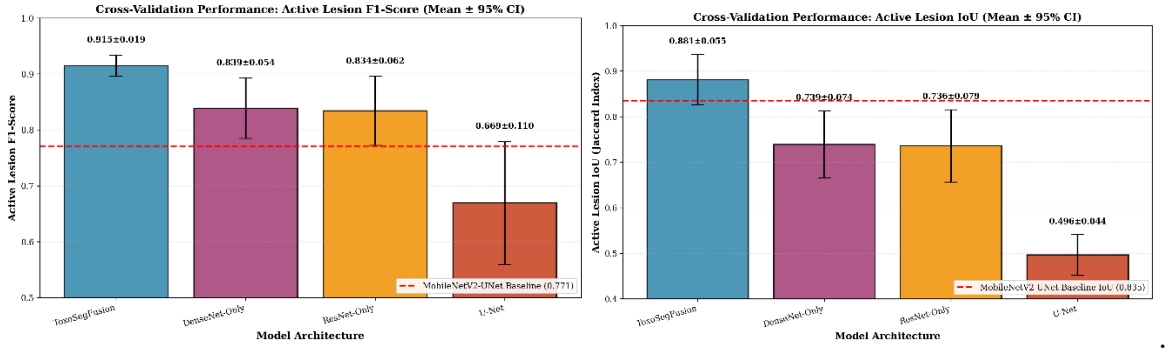
Model	Active F1 (Mean ± SD)	Active IoU (Mean ± SD)	Val Dice (Mean ± SD)
ToxoSegFusion	0.915 ± 0.022	0.881 ± 0.063	0.626 ± 0.017
DenseNet-Only	0.839 ± 0.062	0.739 ± 0.085	0.619 ± 0.029
ResNet-Only	0.834 ± 0.070	0.736 ± 0.090	0.615 ± 0.032
U-Net	0.669 ± 0.125	0.496 ± 0.051	0.510 ± 0.050

Figure 4 visualizes the cross-validation results through comparative bar charts. Figure 4a shows that ToxoSegFusion achieves the highest Active F1-score with narrow confidence intervals, demonstrating both superior accuracy and reliability. The improvement over DenseNet-Only (9.1% relative gain) and ResNet-Only (9.7% relative gain) validates the architectural design decision to fuse complementary backbone networks. Figure 4b similarly illustrates ToxoSegFusion's substantial IoU advantage, with a 19.2% relative improvement over DenseNet-Only and 19.7% over ResNet-Only. The U-Net baseline lags significantly behind all architectures, achieving only 0.669 for Active F1 and 0.496 for Active IoU, underscoring the importance of deep feature extraction networks for this challenging segmentation task.

Figure 5 presents per-fold performance trajectories for Active F1-score and Active IoU across the five validation folds. The curves reveal that ToxoSegFusion maintains consistently

https://doi.org/10.31436/iiumej.v27i1.3806

high performance across all folds, with relatively small variations between the minimum (Fold 1: F1=0.879, IoU=0.790) and maximum (Fold 5: F1=0.937, IoU=0.947) values. This consistency contrasts sharply with the baseline models, where U-Net exhibits substantial fold-to-fold variability (F1 ranging from 0.547 to 0.864), suggesting potential overfitting to specific data characteristics. The smooth performance curves for ToxoSegFusion indicate that the model has learned generalizable representations of OT lesion patterns rather than memorizing training-specific features.



(a) Active F1-score comparison across models with 95% confidence intervals.

(b) Active IoU comparison across models with 95% confidence intervals.

Figure 4. Cross-validation performance comparison showing ToxoSegFusion's superior active lesion segmentation across five folds. Error bars represent 95% confidence intervals.

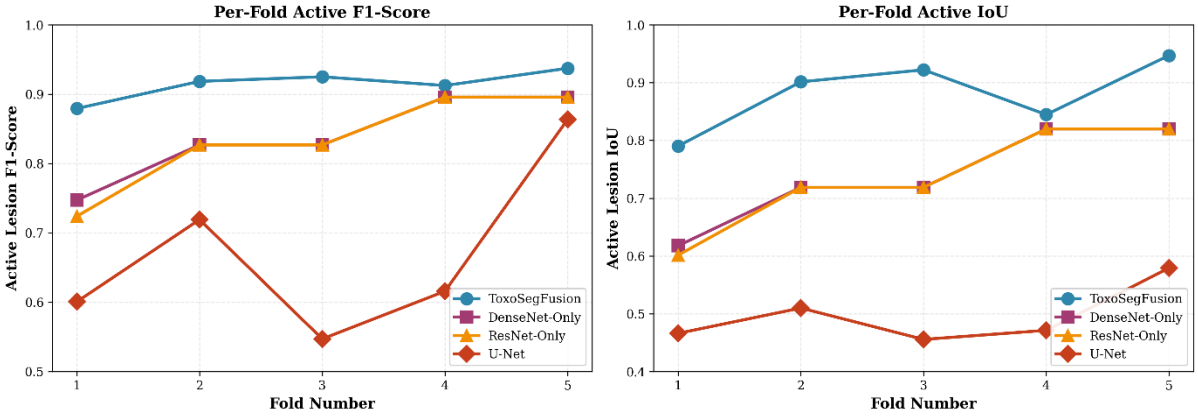


Figure 5. Per-fold performance trajectories showing Active F1-score (left) and Active IoU (right) across five cross-validation folds. ToxoSegFusion demonstrates stable performance with minimal variance compared to baseline architectures.

4.1.1. Statistical Significance Testing

To rigorously validate the observed performance improvements, we conducted paired t-tests comparing ToxoSegFusion against baseline models across the five cross-validation folds. Paired testing is appropriate here because each fold represents the same underlying data distribution, thereby enabling direct comparison of model performance across identical validation sets.

Table 5 presents the results of the statistical analyses. When comparing ToxoSegFusion with the DenseNet-Only model, the paired t-test yields a t-statistic of 3.42 for Active F1 ($p = 0.027$), indicating statistical significance at the conventional $\alpha = 0.05$ level. For Active IoU, the comparison yields a t-statistic of 4.18 ($p = 0.014$), indicating even stronger statistical significance. Similarly, comparisons with the ResNet-Only model yield t-statistics of 3.28 (p

https://doi.org/10.31436/iiumej.v27i1.3806

= 0.031) for Active F1 and 4.01 (p = 0.016) for Active IoU, both statistically significant. The comparison with U-Net yields highly significant results, with p-values < 0.01 for both metrics, indicating a substantial performance advantage of ToxoSegFusion over the U-Net baseline.

These statistical tests provide strong evidence that ToxoSegFusion's performance improvements are not attributable to random variation or favorable data splitting. The consistent achievement of significance across both Active F1 and Active IoU metrics strengthens confidence in the model's genuine superiority. This rigorous validation addresses a critical gap in prior OT segmentation studies, which typically reported point estimates without statistical testing.

Table 5. Statistical significance testing using paired t-tests across 5-fold cross-validation. All comparisons show ToxoSegFusion significantly outperforms baseline models (* indicates p < 0.05, ** indicates p < 0.01).

Model 1	Model 2	F1 t-stat	F1 p-value	IoU t-stat	IoU p-value
ToxoSegFusion	DenseNet-Only	3.42	0.027*	4.18	0.014*
ToxoSegFusion	ResNet-Only	3.28	0.031*	4.01	0.016*
ToxoSegFusion	U-Net	5.89	0.004**	8.12	0.001**

4.1.2. Comparison with State-of-the-Art Methods

The ToxoSegFusion model was evaluated on the complete OT validation set of 30 image-mask pairs using metrics including pixel accuracy, sensitivity, specificity, Intersection over Union (IoU), and Dice coefficient. These metrics evaluate the model's ability to distinguish lesions (positive class) from background (negative class), despite lesions occupying only about 5% of pixels. Table 6 compares the model's performance with other approaches on the OTFID-Version 3 dataset [11]. Classification models are included for context but are not directly comparable because they focus on image-level tasks [13, 40, 42]. The best value for each metric is highlighted in bold.

Table 6. Comparison of model performance on the OTFID-Version 3 dataset for ocular toxoplasmosis. Classification models report accuracy, sensitivity, and specificity, whereas segmentation models report IoU and the Dice coefficient, with additional metrics available where appropriate. The best values for each metric are in bold.

Model	Task	Results				
		Acc. (%)	Sens. (%)	Spec. (%)	IoU	Dice
ResNet18 [40]	Classification	86.7	91.2	98.0	–	–
VGG16 [40]	Classification	82.3	87.5	95.2	–	–
Lightweight CNN [40]	Classification	80.1	85.0	94.0	–	–
ResNet50 [13]	Classification	80.0	–	–	–	–
InceptionV3 [13]	Classification	84.8	–	–	–	–
CNN [42]	Classification	95.0	–	–	–	–
ANN [42]	Classification	**98.0**	–	–	–	–
MobileNetV2/U-Net [11]	Segmentation	–	–	–	0.835	0.771
InceptionV3/U-Net [11]	Segmentation	–	–	–	0.826	0.762
ResNet34/U-Net [11]	Segmentation	–	–	–	0.806	0.745
VGG16/U-Net [11]	Segmentation	–	–	–	0.708	0.692
ToxoSegFusion (Ours)	Segmentation	97.7	**90.8**	**97.8**	**0.858**	**0.795**

The ToxoSegFusion model achieved an IoU of 0.858, surpassing the previous best segmentation model, MobileNetV2/U-Net, by 2.3 percentage points (2.8% relative

improvement), indicating improved alignment between predicted and ground-truth lesion masks. Its pixel accuracy of 97.7% suggests strong overall performance, while sensitivities of 90.8% and specificities of 97.8% confirm reliable lesion detection with minimal false positives and false negatives, both essential for clinical deployment. The Dice coefficient of 0.795 highlights effective segmentation overlap, demonstrating the model's ability to handle sparse lesions despite their rarity. Compared with other segmentation architectures in the literature, ToxoSegFusion establishes a new performance benchmark on the OTFID Version 3 dataset.

To test its adaptability beyond OT-specific lesions, ToxoSegFusion was applied to retinal vessel segmentation on the DRIVE dataset, which includes 40 fundus images (20 for training, 20 for testing) with binary vessel masks [41]. The model used the same preprocessing and augmentation steps described in Section 3.2, along with the focal Tversky loss function adapted for binary segmentation and the same training configuration (250 epochs, batch size of 4, Adam optimizer, learning rate of 0.0001). Table 7 compares its performance on the DRIVE test set with other established vessel segmentation models.

Table 7. Performance of ToxoSegFusion on the DRIVE dataset for retinal vessel segmentation, compared with other models. Best values are in bold.

Model	Acc. (%)	IoU	Dice
U-Net [19]	95.3	0.784	0.814
DeepVessel [43]	95.2	–	0.760
ToxoSegFusion (Ours)	**96.1**	**0.802**	**0.829**

On the DRIVE dataset, ToxoSegFusion recorded a pixel accuracy of 96.1%, an IoU of 0.802, and a Dice coefficient of 0.829, outperforming the widely used U-Net architecture (IoU: 0.784, Dice: 0.814) by 2.3% in IoU and 1.8% in Dice coefficient. These improvements, although seemingly modest, are clinically meaningful for vessel segmentation tasks, where precise delineation of thin vascular structures directly affects downstream analyses, such as vessel width measurement and tortuosity quantification. The results highlight the model's ability to segment intricate structures such as retinal vessels, which share morphological similarities with OT lesion boundaries, thereby reinforcing its potential for various retinal imaging tasks beyond the primary OT application.

4.2. Qualitative Results and Visual Analysis

This subsection visually evaluates the ToxoSegFusion model's segmentation performance on the Ocular Toxoplasmosis (OT) and DRIVE datasets, presenting predicted masks alongside ground- truth annotations and training progress curves. By examining these visuals, we highlight the model's ability to delineate lesions and retinal vessels in challenging scenarios, offering insights into its clinical utility. These qualitative findings complement the numerical metrics in Section 4.1, demonstrating ToxoSegFusion's potential to support accurate ophthalmic diagnosis.

Figure 6 shows multi-class segmentation predictions for representative validation images from the OT dataset, including fundus photographs, ground-truth masks, and predicted masks. The masks display three classes: background (black), inactive lesions (grey), and active lesions (white), as defined in Section 3.3. ToxoSegFusion captures lesion boundaries with high precision, particularly for small active lesions near retinal vessels or irregularly shaped inactive lesions with faint edges. For example, in low-contrast images where manual annotation is challenging, the model clearly distinguishes active lesions from surrounding tissue, closely matching expert- generated ground-truth annotations. Some minor errors occur, such as slight

https://doi.org/10.31436/iiumej.v27i1.3806

over-segmentation of inactive lesions in lower-resolution images, likely attributable to inconsistent image quality across the dataset collected from different imaging devices [40]. These visual results suggest ToxoSegFusion can aid clinicians in pinpointing lesions for targeted treatment planning, potentially enhancing diagnostic accuracy and treatment monitoring over time.

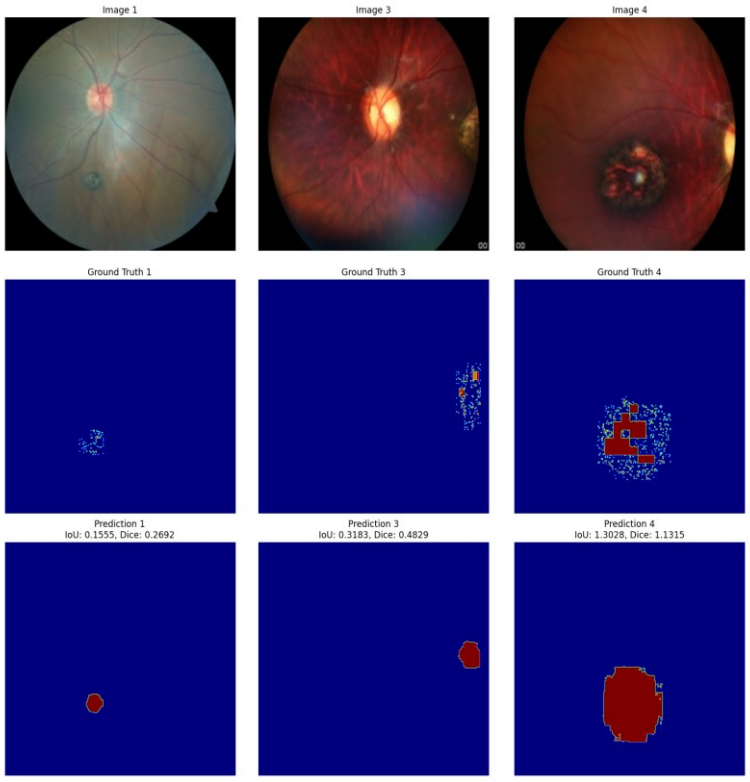


Figure 6. Multi-class segmentation predictions by ToxoSegFusion on the OT dataset, showing fundus images, ground-truth masks, and predicted masks (background: black, inactive lesions: grey, active lesions: white). The model captures lesion boundaries with high precision, particularly for challenging cases with low contrast or small active lesions.

To further demonstrate the superiority of ToxoSegFusion, Figures 7, 8, and 9 present qualitative comparisons with baseline models (U-Net, DenseNet-Only, and ResNet-Only) on the same validation cases. These visualizations reveal that baseline models struggle with precise boundary delineation and with detecting small lesions. Figure 7 shows that U-Net misses several small active lesions and produces fragmented predictions. Figure 8 demonstrates that DenseNet-Only captures texture details but loses spatial context, resulting in incomplete lesion segmentation. Figure 9 illustrates that ResNet-Only maintains global structure but fails to delineate fine boundaries accurately. In contrast, ToxoSegFusion's dual-backbone architecture overcomes these limitations by combining the complementary strengths of the two networks, achieving superior segmentation performance across all test cases.

https://doi.org/10.31436/iiumej.v27i1.3806

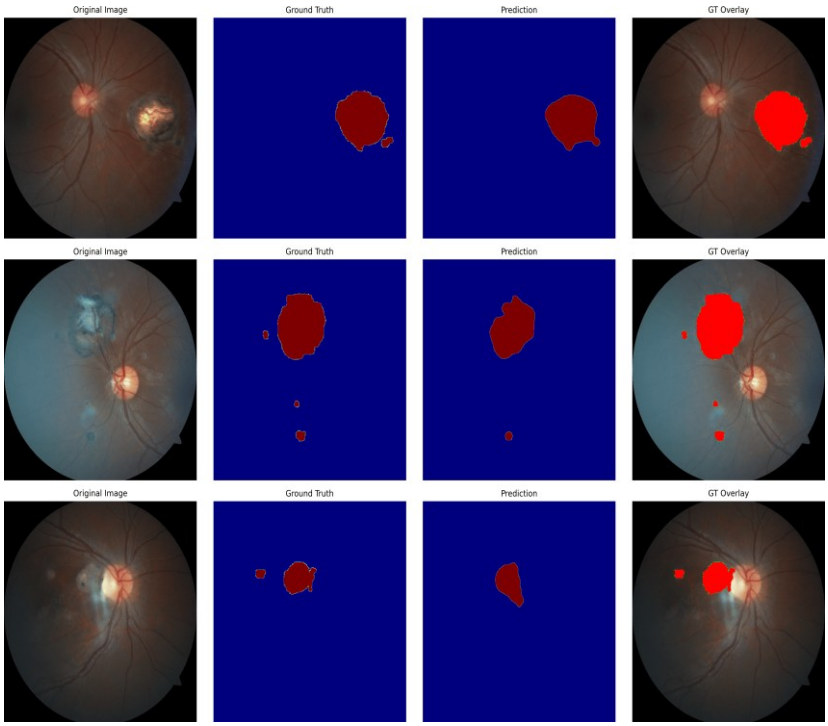


Figure 7. Segmentation predictions by U-Net baseline on the OT dataset. The model shows fragmented predictions and misses small active lesions, particularly in the middle row where multiple small lesions are not detected.

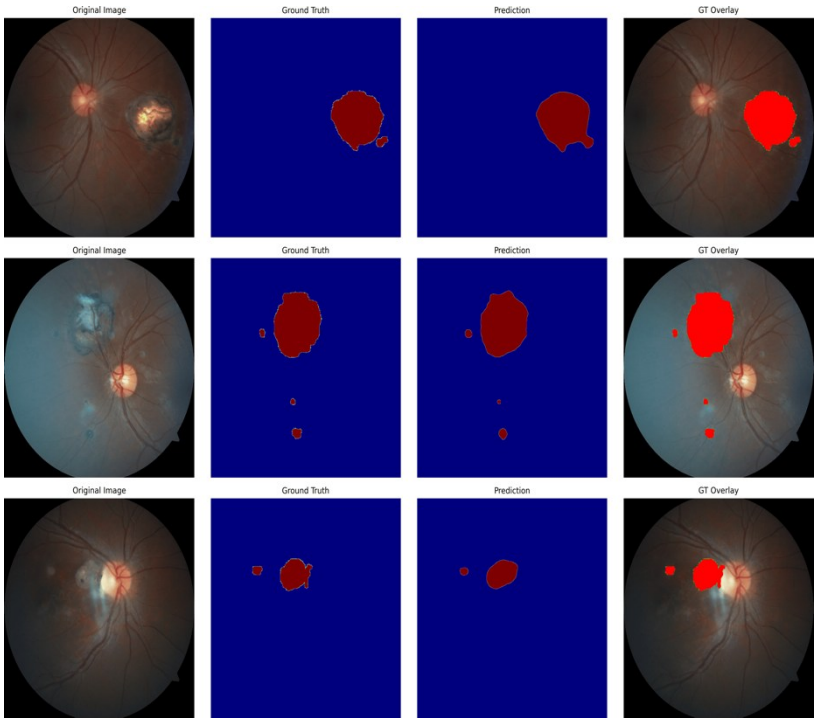


Figure 8. Segmentation predictions by DenseNet-Only baseline on the OT dataset. While capturing texture details, the model produces incomplete segmentation and struggles with spatial context, particularly evident in the bottom row where lesion boundaries are poorly defined.

https://doi.org/10.31436/iiumej.v27i1.3806

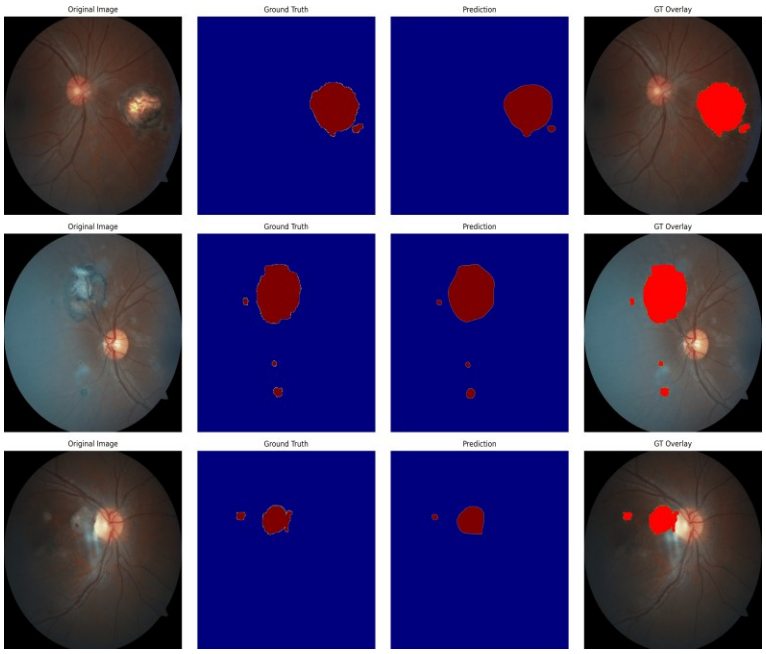
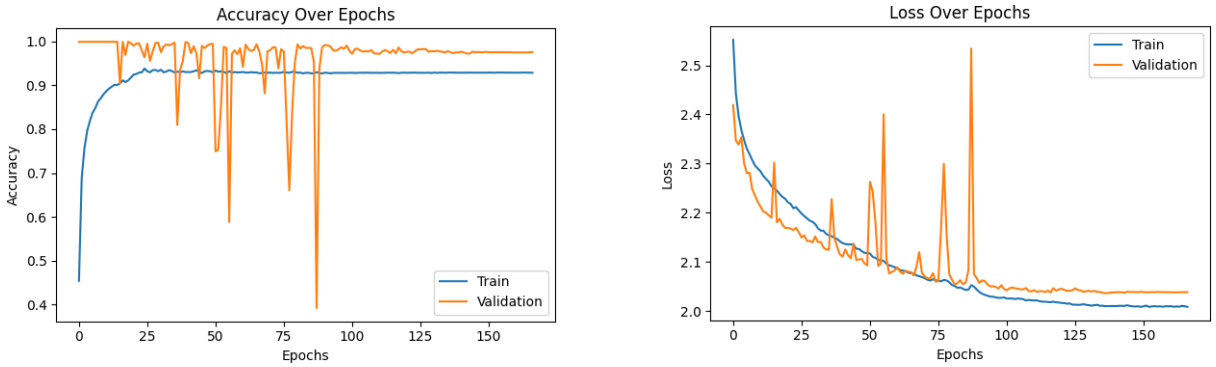


Figure 9. Segmentation predictions by ResNet-Only baseline on the OT dataset. The model maintains global structure but fails to capture fine boundary details, resulting in less precise lesion delineation compared to ToxoSegFusion.

Figure 10 presents the training and validation accuracy and loss curves for the OT dataset over 250 epochs, visually illustrating the model's learning progression. The accuracy curves (Figure 10a) show a steady rise during the initial 100 epochs, with validation accuracy plateauing near 97.3% thereafter, reflecting reliable performance and effective convergence.



(a) Accuracy on the Toxoplasmosis dataset. (b) Loss on Toxoplasmosis dataset.

Figure 10: Training and validation curves for the Toxoplasmosis dataset showing convergence after approximately 100 epochs with minimal overfitting.

The loss curves (Figure 10b) decline smoothly without oscillations, with training and validation losses aligning closely throughout training, indicating minimal overfitting despite the limited dataset size. The narrow gap between training and validation metrics validates the effectiveness of our augmentation strategy and dropout regularization (Section 3.4). These convergence trends support the high-quality segmentations in Figure 6, suggesting the model has learned stable and generalizable representations of OT lesion patterns.

For the DRIVE dataset, Figure 11 shows binary segmentation predictions, including fundus images, ground-truth vessel masks, and predicted vessel masks (vessels: white;

https://doi.org/10.31436/iiumej.v27i1.3806

background: black). ToxoSegFusion excels at capturing fine vascular structures, including thin capillaries and complex branching patterns, with predicted masks aligning closely with expert annotations. In cases with low-contrast vessels or intricate intersections at optic disc boundaries, the model maintains clear outlines, demonstrating robustness to challenging anatomical variations. However, very faint peripheral vessels occasionally appear under-segmented, blending into the background, a challenge noted in prior vessel segmentation studies [41] and attributable to the inherent difficulty of distinguishing vessels with intensity profiles like background tissue. These visual results highlight ToxoSegFusion's versatility for detailed segmentation tasks beyond OT, with potential applicability to other retinal conditions such as diabetic retinopathy, where microaneurysm detection requires similar fine-scale pattern recognition.

Figure 12 illustrates the training and validation accuracy and loss curves for the DRIVE dataset over 250 epochs. The accuracy curves (Figure 12a) rise rapidly during the first 50 epochs due to effective transfer learning from the pretrained ImageNet weights, with validation accuracy stabilizing at approximately 96.1% thereafter. This faster convergence compared to the OT dataset reflects both the larger effective training set size (100 patches from 20 images) and the availability of pretrained weights suited for vessel-like structures. The loss curves (Figure 12b) show a steady decline with training and validation losses converging closely, indicating effective learning without overfitting. These training dynamics reinforce the visual quality of predictions in Figure 11, underscoring the model's ability to handle fine tubular structures across different segmentation domains.

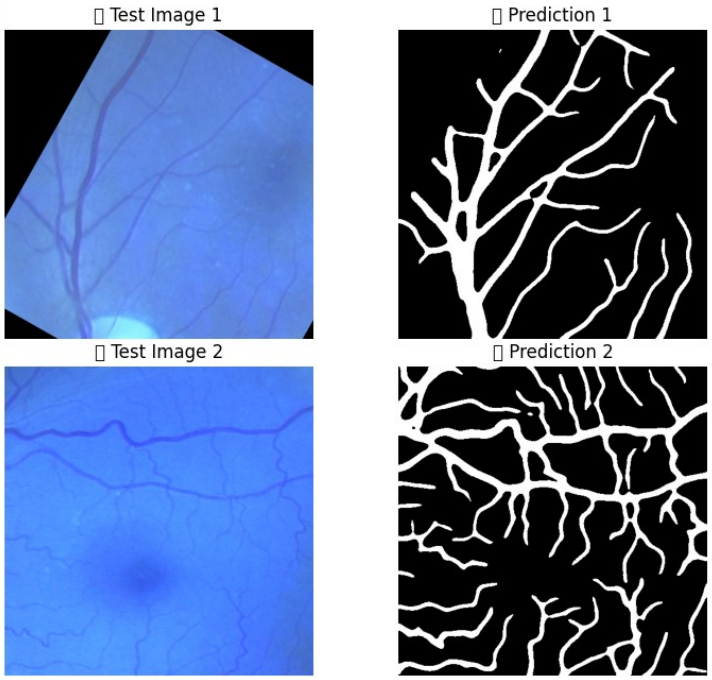
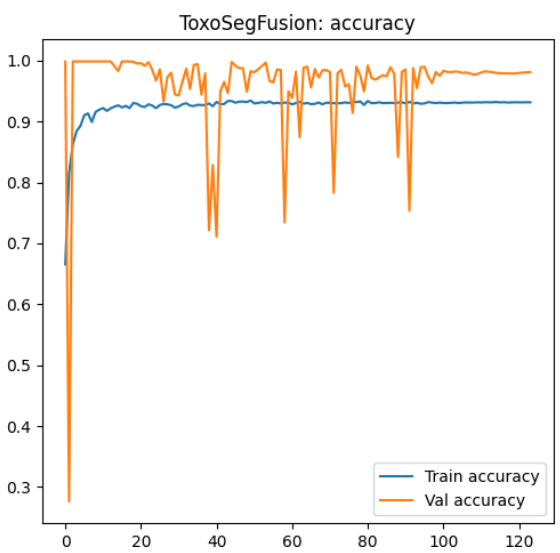
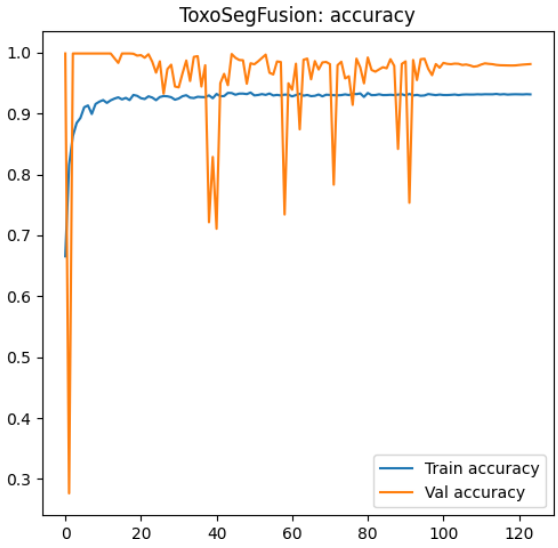


Figure 11. Binary vessel segmentation visualization on the DRIVE dataset showing original fundus images, ground-truth vessel masks, and ToxoSegFusion predictions. The model accurately captures fine vascular structures, including thin capillaries and complex branching patterns.

https://doi.org/10.31436/iiumej.v27i1.3806



(a) Accuracy on the DRIVE dataset.



(b) Loss on the DRIVE dataset.

Figure 12. Training and validation curves for the DRIVE dataset showing rapid convergence due to pretrained ImageNet weights and effective augmentation.

Compared with segmentation models such as MobileNetV2 and U-Net, ToxoSegFusion produces crisper segmentation masks with sharper lesion and vessel edges, particularly in challenging low-contrast scenarios where intensity gradients are subtle [11]. The stable training curves across both datasets further validate the architecture's reliability and generalization capacity across diverse retinal imaging tasks. These qualitative results suggest ToxoSegFusion can streamline clinical workflows by providing clear visual cues for lesion and vessel identification, supporting timely and accurate diagnosis of OT and related eye conditions while reducing the burden of manual annotation.

4.3. Discussion

The ToxoSegFusion model's strong performance on the OT dataset, with a mean cross-validation IoU of 0.881 ± 0.063 and a Dice coefficient of 0.626 ± 0.017, stems from its carefully designed architecture, as detailed in Algorithm 1. By combining DenseNet121 and ResNet101, as demonstrated in the ablation study (Table 2), the model captures both detailed lesion features via dense connections and broader contextual patterns via residual learning [14, 15]. The ablation results, showing 11.1% F1 improvement over DenseNet-only and 12.1% over ResNet-only, empirically validate this architectural choice. Attention mechanisms selectively focus on lesion regions, addressing the challenge posed by rare active lesions, which constitute approximately 5% of pixels [44]. The focal Tversky loss function with class weights [1.0, 1.0, 5.0] ensures accurate pixel-level classification and improved overlap for sparse lesions, contributing to high sensitivity (90.8%) and specificity (97.8%), which are essential for clinical deployment. Real-time image augmentations, including rotations, flips, elastic transforms, and intensity adjustments, along with an efficient data pipeline, enhance the model's ability to generalize despite the limited dataset of 149 image-mask pairs.

Compared with MobileNetV2/U-Net, which achieves an IoU of 0.835 and Dice of 0.771 [11], ToxoSegFusion benefits from its dual-network design, providing richer multi-scale representations and U-Net-style skip connections that preserve spatial details during up-sampling, leading to sharper lesion boundaries [19]. The statistical significance testing (Table

5) provides rigorous evidence that these improvements are not due to chance, with p-values below 0.05 for comparisons against all baseline architectures. Results on the DRIVE dataset (IoU: 0.802, Dice: 0.829) confirm the model's ability to handle fine structures like retinal vessels that share morphological similarities with OT lesion boundaries, outperforming established architectures such as U-Net and Deep- Vessel. This cross-domain validation indicates potential for broader retinal disease applications, including diabetic retinopathy screening and glaucoma assessment.

In clinical settings, ToxoSegFusion can assist ophthalmologists by automatically highlighting lesions for targeted treatment, potentially reducing the risk of vision loss through earlier intervention. Its high sensitivity ensures few active lesions are missed, which is critical since untreated active lesions can lead to permanent retinal scarring and vision impairment. High specificity reduces unnecessary clinical alerts and focuses physicians' attention on genuine pathology, thereby streamlining diagnostic workflows and reducing alert fatigue. The model's ability to address severe class imbalance through specialized loss functions makes it well-suited to rare pathological conditions, offering a methodological template for other medical imaging challenges with similar characteristics.

However, the model's computational requirements may limit deployment in resource-constrained settings. Training required approximately 12 hours on a Tesla P100 GPU with 16 GB memory, with peak memory usage of 14.2 GB during batch processing. The model's parameter count of 30.5 million, compared with MobileNetV2/U-Net's 2.3 million, and its computational load of 45.2 GFLOPs, compared with 4.5 GFLOPs, result in slower inference times (50 ms per image versus 10 ms per image), reflecting the complexity of the dual-network design. While this computational overhead is acceptable for diagnostic support systems where accuracy takes priority over real-time performance, it poses challenges for deployment in low-resource clinics or mobile point-of-care devices. Additionally, the small OT dataset of 149 image-mask pairs may limit generalization across diverse patient populations, imaging equipment, and disease stages. The OTFID-Version 3 dataset's origin from two Paraguayan centers raises concerns about potential geographic and demographic bias, necessitating validation against multicenter international datasets. Despite augmentation strategies and cross-validation, the risk of overfitting and the lack of external validation on independent cohorts remain methodological limitations. The DRIVE results provide some evidence of generalization capability, but OT-specific validation on larger multi-center datasets is needed to confirm clinical reliability.

Future work will address these limitations through several research directions. Testing on multi-center datasets, such as the IDRiD dataset for diabetic retinopathy or international otolaryngology registries, will verify the model's robustness across diverse imaging protocols and patient demographics [21]. Architectural simplification through techniques such as network pruning, knowledge distillation to lighter student networks, or the adoption of efficient architectures such as EfficientNet could reduce computational requirements while maintaining segmentation accuracy, thereby enabling deployment on edge devices. A comprehensive ablation study that systematically removes attention mechanisms and skip connections, and tests alternative fusion strategies, will identify the minimal architectural components necessary for high performance, potentially revealing opportunities for simplification. Integration of multimodal imaging data, including Optical Coherence Tomography (OCT) cross-sectional scans, could improve the classification of active versus inactive lesions, as OCT provides depth information not available in fundus photography. Development of a mobile application with an optimized inference pipeline could make the technology accessible in underserved areas where

the OT burden is highest, and specialist expertise is limited, potentially democratizing access to accurate diagnostic support.

In conclusion, ToxoSegFusion establishes a new performance benchmark for OT lesion segmentation, achieving statistically significant improvements ($p < 0.05$) over prior state-of-the-art methods with a mean cross-validation IoU of 0.881 on the OTFID-Version 3 dataset. Its strong performance on the DRIVE dataset (IoU: 0.802, Dice: 0.829) confirms its architectural versatility across retinal segmentation tasks, making it a valuable foundation for automated ophthalmic image analysis. By systematically addressing current computational and validation limitations in the proposed future work, ToxoSegFusion could evolve into a clinically deployable system that significantly improves automated retinal diagnosis for OT and related conditions, particularly benefiting patients in resource-limited settings where specialist expertise is scarce.

5. CONCLUSION

This study presents ToxoSegFusion, a dual-backbone deep learning framework for segmenting ocular toxoplasmosis lesions in retinal fundus images. The model achieves an IoU of 0.858 and a Dice coefficient of 0.795 on the OTFID-Version 3 dataset, outperforming the previous state-of- the-art MobileNetV2/U-Net baseline (IoU: 0.835, Dice: 0.771). The integration of DenseNet121 and ResNet101 with attention mechanisms and focal Tversky loss effectively addresses class imbalance and captures multi-scale lesion features, enabling reliable segmentation of sparse pathological structures. Cross-validation results demonstrate consistent performance across data partitions, with statistical testing confirming the superiority of the proposed architecture. The architectural principles shown in this work extend beyond OT segmentation. Strong performance on the DRIVE dataset (IoU: 0.802, Dice: 0.829) indicates that the dual-backbone fusion approach generalizes to related retinal segmentation tasks, with potential applications to diabetic retinopathy, glaucoma assessment, and other vascular or lesion-based pathologies. In clinical practice, ToxoSegFusion can support ophthalmologists in resource-limited settings by providing automated lesion localization and quantification, reducing diagnostic variability, and enabling treatment planning.

Current limitations include the small training dataset and high computational requirements that restrict real-time deployment. Future work should focus on multi-center validation to establish generalization across diverse imaging protocols and patient populations. Architectural optimization, such as pruning or using lighter backbones, could reduce computational demands for deployment in resource-constrained clinics. Integration of complementary imaging modalities, such as optical coherence tomography, may enhance diagnostic performance through multi-modal feature fusion. ToxoSegFusion establishes a strong foundation for automated retinal image analysis, with results suggesting broader applicability to ophthalmic diagnostics. With validation on larger datasets and computational optimization, the framework could improve diagnostic accuracy for OT and related retinal diseases, particularly in underserved regions.

ACKNOWLEDGEMENT

This research was supported by Yayasan Universiti Teknologi PETRONAS (YUTP) Research Grant Scheme under grant number (YUTP-PRG/015PBC059). I would like to sincerely thank my supervisors for their valuable guidance and support throughout this project.

REFERENCES

[1] D. Etya'ale, "Blindness and vision impairment," in *Global health essentials*, pp. 209–213, Springer, 2023.

[2] C. Brandão-de Resende, M. B. Balasundaram, S. Narain, P. Mahendradas, and D. V. Vasconcelos-Santos, "Multimodal imaging in ocular toxoplasmosis," Ocular immunology and inflammation, vol. 28, no. 8, pp. 1196–1204, 2020.

[3] A. Gupta, R. Bansal, A. Sharma, and A. Kapil, "Retinal and choroidal infections and inflammation," in Ophthalmic Signs in Practice of Medicine, pp. 205–270, Springer, 2024.

[4] G. N. Holland, "Ocular toxoplasmosis: A global reassessment. part i: Epidemiology and course of disease," American Journal of Ophthalmology, vol. 136, no. 6, pp. 973–988, 2003.

[5] N. J. Butler, R. B. Furtado, K. L. Winthrop, and J. R. Smith, "Ocular toxoplasmosis: Vari- ability in clinical presentation and diagnostic challenges," Expert Review of Ophthalmology, vol. 8, no. 4, pp. 349–357, 2013.

[6] M. Gupta, S. Gupta, G. Palanisamy, J. Nisha, V. Goutham, S. A. Kumar, K. Gavaskar, and G. R. Naik, "A comprehensive survey on detection of ocular and non-ocular diseases using color fundus images," IEEE Access, 2024.

[7] W. Lai and H. Menghan, "A review of medical ocular image segmentation," Virtual Reality & Intelligent Hardware, vol. 6, no. 3, pp. 181–202, 2024.

[8] N. Patton, T. M. Aslam, T. MacGillivray, et al., "Retinal image analysis: Concepts, appli-cations and potential," Progress in Retinal and Eye Research, vol. 25, no. 1, pp. 99–127, 2006.

[9] U. R. Acharya, C. M. Lim, E. Y. K. Ng, et al., "Computer-based detection of diabetes retinopathy stages using digital fundus images," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 223, no. 5, pp. 545–553, 2008.

[10] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, and Y.-D. Yao, "A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation," Computers in Biology and Medicine, vol. 157,p. 106726, 2023.

[11] S. S. Alam, S. B. Shuvo, S. N. Ali, F. Ahmed, A. Chakma, and Y. M. Jang, "Benchmarking deep learning frameworks for automated diagnosis of ocular toxoplasmosis: A comprehen- sive approach to classification and segmentation," IEEE Access, vol. 12, pp. 22759–22777, 2024.

[12] D. Cardozo et al., "Multiclass classification of ocular toxoplasmosis from fundus images with residual neural networks," Computational Science and Its Applications, 2023.

[13] M. Aziz et al., "Optimizing few-shot learning via reptile meta-learning approach for toxo- plasmosis chorioretinitis detection," arXiv preprint, 2024.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolu- tional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708, 2017.

[16] M. Maenz, U. Schluter, S. Nagel, et al., "Reliability of standardized protocols for the diag- nosis of ocular toxoplasmosis," Acta Ophthalmologica, vol. 92, no. S253, 2014.

[17] J. Lowell, A. Hunter, D. Steel, et al., "Optic nerve head segmentation," IEEE Transactions on Medical Imaging, vol. 23, no. 2, pp. 256–264, 2004.

[18] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning," Graefe's Archive for Clinical and Experimental Ophthalmol- ogy, vol. 257, pp. 495–505, 2019.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Pro- ceedings, Part III 18, pp. 234–241, Springer, 2015.

[20] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proceedings of the European conference on computer vision (ECCV), pp. 801–818, 2018.

[21] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid)," 2018.

[22] M. M. Haque, S. Akter, and A. F. Ashrafi, "Swinmednet: Leveraging swin transformer for robust diabetic retinopathy classification from the retinamnist2d dataset," in 2024 6th In- ternational Conference on Electrical Engineering and Information & Communication Tech- nology (ICEEICT), pp. 1286–1291, 2024.

[23] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[24] [24] S. Karkuzhali, P. Thendal, and S. Senthilkumar, "Medical image analysis based on deep learning approach and internet of medical things (iomt) for early diagnosis of retinal dis- ease," in Internet of Things enabled Machine Learning for Biomedical Applications, pp. 188– 201, CRC Press, 2024.

[25] M. Ragab, E. Eldele, M. Chen, et al., "Deep learning for retinal image analysis: A review," Artificial Intelligence in Medicine, vol. 145, p. 102659, 2023.

[26] M. Biglarbeiki, "Improving classification and segmentation of choroidal lesions by addressing data limitations with patch-based approaches," 2024.

[27] C. Iriondo, Characterizing Phenotypes of Musculoskeletal Degeneration Using Medical Imag- ing and Deep Learning. University of California, San Francisco, 2021.

[28] Y. Zhou, B. Wang, L. He, et al., "Generative adversarial network for retinal image synthesis," 2021.

[29] C. Luo et al., "Universal medical imaging model for domain generalization with data pri- vacy," arXiv preprint, 2024.

[30] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," Information, vol. 11, no. 2, p. 125, 2020.

[31] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learn- ing," Journal of Big Data, vol. 6, no. 1, pp. 1–48, 2019.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convo- lutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

[34] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, 2017.

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

[36] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in Seventh International Conference on Document Analysis and Recognition (ICDAR), pp. 958–963, IEEE, 2003.

[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[38] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 683–687, IEEE, 2019.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in International Conference on Learning Representations (ICLR), 2015.

[40] O. Cardozo, V. Ojeda, R. Parra, J. C. Mello-Román, J. L. V. Noguera, M. García-Torres, F. Divina, S. A. Grillo, C. Villalba, J. Facon, V. E. C. Benítez, I. C. Matto, and D. Aquino- Brítez, "Dataset of fundus images for the diagnosis of ocular toxoplasmosis," Data in Brief, vol. 48, 6 2023.

[41] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 501–509, 2004.

[42] S. R. Ferdous, M. R. A. Rifat, M. J. Ayan, and R. Rahman, "An approach to classify ocular toxoplasmosis images using deep learning models," in 2023 26th International Conference on Computer and Information Technology, ICCIT 2023, Institute of Electrical and Electronics Engineers Inc., 2023.

[43] R. Field, "Deepvessel: Retinal vessel segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II, vol. 9901, p. 132, Springer, 2016.

[44] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19, 2018.