Pertanika Journal of

SCIENCE & TECHNOLOGY

JST

VOL. 34 (4) AUG. 2026

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

PERTANIKA JOURNAL OF SCIENCE & TECHNOLOGY

About the Journal

Overview

Pertanika Journal of Science & Technology is an official journal of Universiti Putra Malaysia. It is an open-access online scientific journal. It publishes original scientific outputs. It neither accepts nor commissions third party content.

Recognised internationally as the leading peer-reviewed interdisciplinary journal devoted to the publication of original papers, it serves as a forum for practical approaches in improve quality on issues pertaining to science and engineering and its related fields.

Pertanika Journal of Science & Technology currently publishes 6 issues a year (*February, April*, *June, August, October, and December*). It is considered for publication of original articles as per its scope. The journal publishes in **English** and it is open for submission by authors from all over the world.

The journal is available world-wide.

Aims and Scope

Pertanika Journal of Science & Technology aims to provide a forum for high quality research related to science and engineering research. Areas relevant to the scope of the journal include: bioinformatics, bioscience, biotechnology and bio-molecular sciences, chemistry, computer science, ecology, engineering, engineering design, environmental control and management, mathematics and statistics, medicine and health sciences, nanotechnology, physics, safety and emergency management, and related fields of study.

History

Pertanika Journal of Science & Technology was founded in 1993 and focusses on research in science and engineering and its related fields.

Vision

To publish a journal of international repute.

Mission

Our goal is to bring the highest quality research to the widest possible audience.

Quality

We aim for excellence, sustained by a responsible and professional approach to journal publishing. Submissions can expect to receive a decision within 90 days. The elapsed time from submission to publication for the articles averages 180 days. We are working towards decreasing the processing time with the help of our editors and the reviewers.

Abstracting and Indexing of Pertanika

Pertanika Journal of Science & Technology is now over 33 years old; this accumulated knowledge and experience has resulted the journal being abstracted and indexed in SCOPUS (Elsevier), Journal Citation Reports (JCR-Clarivate), EBSCO, ASEAN CITATION INDEX, Microsoft Academic, Google Scholar, and MyCite.

Citing Journal Articles

The abbreviation for Pertanika Journal of Science & Technology is *Pertanika J. Sci. & Technol.*

Publication Policy

Pertanika policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. It prohibits as well publication of any manuscript that has already been published either in whole or substantial part elsewhere. It also does not permit publication of manuscript that has been published in full in proceedings.

Code of Ethics

The *Pertanika* journals and Universiti Putra Malaysia take seriously the responsibility of all of its journal publications to reflect the highest in publication ethics. Thus, all journals and journal editors are expected to abide by the journal's codes of ethics. Refer to *Pertanika*'s **Code of Ethics** for full details, or visit the journal's web link at http://www.pertanika.upm.edu.my/code_of_ethics.php

Originality

The author must ensure that when a manuscript is submitted to *Pertanika*, the manuscript must be an original work. The author should check the manuscript for any possible plagiarism using any programme such as Turn-It-In or any other software before submitting the manuscripts to the Editorial Office, *Pertanika* Journal Publication Section.

All submitted manuscripts must be in the journal's acceptable similarity index range:

≤ 20% – PASS; > 20% – REJECT.

International Standard Serial Number (ISSN)

An ISSN is an 8-digit code used to identify periodicals such as journals of all kinds and on all media–print and electronic.

Pertanika Journal of Science & Technology: e-ISSN 2231-8526 (Online).

Lag Time

A decision on acceptance or rejection of a manuscript is reached in 90 days (average). The elapsed time from submission to publication for the articles averages 180 days.

Authorship

Authors are not permitted to add or remove any names from the authorship provided at the time of initial submission without the consent of the journal's Chief Executive Editor.

Manuscript Preparation

For manuscript preparation, authors may refer to Pertanika's **INSTRUCTION TO AUTHORS**, available on the official website of *Pertanika*.

Editorial Process

Authors who complete any submission are notified with an acknowledgement containing a manuscript ID on receipt of a manuscript, and upon the editorial decision regarding publication.

Pertanika follows a **double-blind peer-review** process. Manuscripts deemed suitable for publication are sent to reviewers. Authors are encouraged to suggest names of at least 3 potential reviewers at the time of submission of their manuscripts to *Pertanika*, but the editors will make the final selection and are not, however, bound by these suggestions.

Notification of the editorial decision is usually provided within 90 days from the receipt of manuscript. Publication of solicited manuscripts is not guaranteed. In most cases, manuscripts are accepted conditionally, pending an author's revision of the material.

The Journal's Peer Review

In the peer-review process, 2 to 3 referees independently evaluate the scientific quality of the submitted manuscripts. At least 2 referee reports are required to help make a decision.

Peer reviewers are experts chosen by journal editors to provide written assessment of the **strengths** and **weaknesses** of the written research, with the aim of improving the reporting of research and identifying the most appropriate and highest quality material for the journal.

Operating and Review Process

What happens to a manuscript once it is submitted to *Pertanika*? Typically, there are 7 steps to the editorial review process:

1. The journal's Chief Executive Editor and Editor-in-Chief examine the paper to determine whether it is relevance to the journal's needs in terms of novelty, impact, design, procedure, language as well as presentation and allow it to proceed to the reviewing process. If it is not appropriate, the manuscript is rejected outright and the author is informed.

2. The Chief Executive Editor sends the article-identifying information having been removed, to 2 to 3 reviewers. They are specialists in the subject matter of the article. The Chief Executive Editor requests that they complete the review within 3 weeks.

   Comments to authors are about the appropriateness and adequacy of the theoretical or conceptual framework, literature review, method, results and discussion, and conclusions. Reviewers often include suggestions to strengthening of the manuscript. Comments to the editor are in the nature of the significance of the work and its potential contribution to the research field.

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

Pertanika Journal of Science & Technology
Pertanika Journal of Science & Technology

3. The Editor-in-Chief examines the review reports and decides whether to accept or reject the manuscript, invite the authors to revise and resubmit the manuscript, or seek additional review reports. In rare instances, the manuscript is accepted with almost no revision. Almost without exception, reviewers' comments (to the authors) are forwarded to the authors. If a revision is indicated, the editor provides guidelines to attend to the reviewers' suggestions and perhaps additional advice on how to revise the manuscript.

4. The authors decide whether and how to address the reviewers' comments and criticisms and the editor's concerns. The authors return a revised version of the paper to the Chief Executive Editor along with specific information describing how they have addressed' the concerns of the reviewers and editor, usually in a tabular form. The authors may also submit a rebuttal if there is a need especially when the authors disagree with certain comments provided by the reviewers.

5. The Chief Executive Editor sends the revised manuscript out for re-review. Typically, at least 1 of the original reviewers will be asked to examine the article.

6. When the reviewers have completed their work, the Editor-in-Chief examines their comments and decides whether the manuscript is ready to be published, needs another round of revisions, or should be rejected. If the decision is to accept, the Chief Executive Editor is notified.

7. The Chief Executive Editor reserves the final right to accept or reject any material for publication, if the processing of a particular manuscript is deemed not to be in compliance with the S.O.P. of *Pertanika*. An acceptance letter is sent to all the authors.

   The editorial office ensures that the manuscript adheres to the correct style (in-text citations, the reference list, and tables are typical areas of concern, clarity, and grammar). The authors are asked to respond to any minor queries by the editorial office. Following these corrections, page proofs are mailed to the corresponding authors for their final approval. At this point, **only essential changes are accepted**. Finally, the manuscript appears in the pages of the journal and is posted online.

Pertanika Journal of

SCIENCE & TECHNOLOGY

Vol. 34 (4) Aug. 2026

PERTANIKA JOURNALS

A scientific journal published by Universiti Putra Malaysia Press

PJST
Pertanika Journal of Science & Technology
AN INTERNATIONAL PEER-REVIEWED JOURNAL

UNIVERSITY PUBLICATIONS COMMITTEE

CHAIRMAN
Zamberi Sekawi

EDITOR-IN-CHIEF
Mohamed Thariq Hameed Sultan
SEARCA Chair of Agriculture Technology and Innovation

JOURNAL SECTION
Head of Section:
Navaneetha Krishna Chandran

Journal Officers:
Ain Haziqah Ab Razak
Nur Izni Shafiqah Mohd Hamidi
Saundarya Saran

Editorial Assistants:
Zulinaardawati Kamarudin

Pre-press Officers:
Aisyah Umairah Jasmin
Engku Erina Zahirah Engku Mohd Suhaimi
Ku Ida Mastura Ku Baharom

IT Officer:
Kiran Raj Kaneswaran

Journal Section Office
Putra Science Park
1st Floor, IDEA Tower II
UPM-MTDC Technology Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia

General Enquiry
Tel. No: +603 9769 1622
E-mail:
pertanika.proceeding@upm.edu.my
URL:
http://www.pertanika.upm.edu.my/

PUBLISHER
UPM Press Centre
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, Malaysia
Tel: +603 9769 3606
E-mail: dir.penerbit@upm.edu.my
URL: http://penerbit.upm.edu.my

ASSOCIATE EDITOR
2024-2026

Miss Laiha Mat Kiah
Security Services Sn: Digital Forensic, Steganography, Network Security, Information Security, Communication Protocols, Security Protocols
Universiti Malaya, Malaysia

Saidur Rahman
Renewable Energy, Nanofluids, Energy Efficiency, Heat Transfer, Energy Policy
Sunway University, Malaysia

EDITORIAL BOARD
2024-2026

Abdul Latif Ahmad
Chemical Engineering
Universiti Sains Malaysia, Malaysia

Hsiu-Po Kuo
Chemical Engineering
National Taiwan University, Taiwan

Mohd. Ali Hassan
Bioprocess Engineering, Environmental Biotehnology
Universiti Putra Malaysia, Malaysia

Ahmad Zaharin Aris
Hydrochemistry, Environmental Chemistry, Environmental Forensics, Heavy Metals
Universiti Putra Malaysia, Malaysia

Ivan D. Rukhlenko
Nonliner Optics, Silicon Photonics, Plasmonics and Nanotechnology
The University of Sydney, Australia

Nor Azah Yusof
Biosensors, Chemical Sensor, Functional Material
Universiti Putra Malaysia, Malaysia

Azlina Harun@ Kamaruddin
Enzyme Technology, Fermentation Technology
Universiti Sains Malaysia, Malaysia

Lee Keat Teong
Energy Environment, Reaction Engineering, Waste Utilisation, Renewable Energy
Universiti Sains Malaysia, Malaysia

Norbahiah Misran
Communication Engineering
Universiti Kebangsaan Malaysia, Malaysia

Bassim H. Hameed
Chemical Engineering: Reaction Engineering, Environmental Catalysis & Adsorption
Qatar University, Qatar

Mohamed Othman
Communication Technology and Network, Scientific Computing
Universiti Putra Malaysia, Malaysia

Roslan Abd-Shukor
Physics & Materials Physics, Superconducting Materials
Universiti Kebangsaan Malaysia, Malaysia

Biswajeet Pradhan
Digital image processing, Geographical Information System (GIS), Remote Sensing
University of Technology Sydney, Australia

Mohd Shukry Abdul Majid
Polymer Composites, Composite Pipes, Natural Fibre Composites, Biodegradable Composites, Bio-Composites
Universiti Malaysia Perlis, Malaysia

Sodeifian Gholamhossein
Supercritical technology, Optimisation, nanoparticles, Polymer nanocomposites
University of Kashan, Iran

Ho Yuh-Shan
Water research, Chemical Engineering and Environmental Studies
Asia University, Taiwan

Mohd Zulkifly Abdullah
Fluid Mechanics, Heat Transfer, Computational Fluid Dynamics (CFD)
Universiti Sains Malaysia, Malaysia

Wing Keong Ng
Aquaculture, Aquatic Animal Nutrition, Aqua Feed Technology
Universiti Sains Malaysia, Malaysia

INTERNATIONAL ADVISORY BOARD
2024-2027

Hiroshi Uyama
Polymer Chemistry, Organic Compounds, Coating, Chemical Engineering
Osaka University, Japan

Mohini Sain
Material Science, Biocomposites,Biomaterials
University of Toronto, Canada

Mohamed Pourkashanian
Mechanical Engineering, Energy, CFD and Combustion Processes
Sheffield University, United Kingdom

Yulong Ding
Particle Science & Thermal Engineering
University of Birmingham, United Kingdom

ABSTRACTING AND INDEXING OF PERTANIKA JOURNALS

Pertanika Journal of Science & Technology is indexed in Journal Citation Reports (JCR-Clarivate), SCOPUS (Elsevier), EBSCO, ASEAN Citation Index, Microsoft Academic, Google Scholar and MyCite.

The publisher of Pertanika will not be responsible for the statements made by the authors in any articles published in the journal. Under no circumstances will the publisher of this publication be liable for any loss or damage caused by your reliance on the advice, opinion or information obtained either explicitly or implied through the contents of this publication.
All rights of reproduction are reserved in respect of all papers, articles, illustrations, etc., published in Pertanika. Pertanika provides free access to the full text of research articles for anyone, web-wide. It does not charge either its authors or author-institution for refereeing/publishing outgoing articles or user-institution for accessing incoming articles.
No material published in Pertanika may be reproduced or stored on microfilm or in electronic, optical or magnetic form without the written authorisation of the Publisher.
Copyright ©2026 Universiti Putra Malaysia Press. All Rights Reserved.

Pertanika Journal of Science & Technology
Vol. 34 (4) Aug. 2026

Contents

Foreword xiii
Mohamed Thariq Hameed Sultan

Review Article 1953
Comprehensive Review of Plant Disease Detection: Advancements in Imaging Sensors, AI Techniques, and Future Directions in Smart Agriculture
Binal Upadhyay and Amit Ganatra

Smart Comment Manager: Streamlining Comment Analysis and Viewer Query Clustering Using NLP Processing 1981
N Venkata Sailaja, Karnam Akhil, Sai Shashidhar Reddy Katanguri, Maringanti Gopi Krishnna, Alladi Shiva Vinay, and Devireddy Siddartha Reddy

AI Meets Reproductive Health: Early Diagnosis of PCOS Using Machine Learning 2003
Abhinav Pathak, M. Sujithra, Hemashree P, and Arun Kumar Selvam

Parametric Study of Reaction Parameter for Glycerol Acetylation and Reusability of Sulfated-Titania Catalyst 2023
Shera Farisya Mohamad Rasid, Ernee Noryana Muhamad, Yun Hin Taufiq-Yap, Haizatul Izzatti Ridzuan, Siew Ling Lee, and Irmawati Ramli

Data-Driven Fault Detection and Diagnosis of a Nuclear Reactor Cooling System Using Graph Neural Networks 2053
Dhines Shree Kanthan, Nur Aina Mohammad Aziz, Azura Che Soh, Ribhan Zafira Abdul Rahman, Mohd Khair Hassan, Mohd Sabri Minhat, and Julia Abdul Karim

Impact of the Grading Ring Size and Hydrophobicity on the High-Voltage Insulator Performance 2075
Amelia Adrina Azli, Nor Aqillah Aiman Nor Azman, N.H. Nik Ali, Hazlee Azil Illias, Ali Ahmed Ali Salem, Mohd Aizam Talib, and Mohd Fairouz Mohd Yousof

Review Article 2097
Prediction of Lung Diseases Using Deep Learning Techniques: A Review
Sheikh Afaan Farooq and Aleem Ali

Explainable Crop and Soil-Based Recommendation System for Next-Generation Smart Agriculture Using Model Classification 2127
Himani Trivedi, Hetal Chauhan, Suresh Patel, Mahendra N Patel, Pradip Patel, Mahi Parmar, and Chrisha Dabhi

Utilisation of Starch-Derivative Functionalised Biochar for Management of Industrial Effluent 2149
Maria Adilah Basharuddin, Siti Fatimah Zaharah Mohamad Fuzi, Noor Mohd Syeqqal Ismail, Chia Kim Seng, and Nor Hasmaliana Abdul Manas

Dynamic Sensor Activation for Energy-Efficient Event Monitoring in Large-Scale IoT Deployments 2167
A. Subramaneswara Rao, Suresh Kallam, and Srinivasa Rao Pappu

Calibration and Limitations of Hardening Soil and Soft Soil Creep Models for Undrained Behaviour of Tropical Peat 2181
Ahmad Kamal Abd Aziz, Norazzlina M. Sa'don, and Abdul Razak Abdul Karim

Assessing the Drought Index and Water Scarcity Nexus: Case Study for Langat River Basin 2213
Raja Muhamad Naufal Raja Azman, Nur Asmaliza Mohd Noor, Marini Mohamad Ideris, and Sharifah Abdullah

An Interactive Augmented Reality and Digital Twin Framework for IoT-Based Aquaponics in Urban Farming 2233
Siti Khadijah Ali, Phang Kok Wai, Zainal Abdul Kahar, and Rahmita Wirza O.K. Rahmat

A Framework for Secure Software Engineering in Hybrid Agile Software Development 2255
Normi Sham Awang Abu Bakar , Norzariyah Yahya, and Siti Sarah Maidin

Evaluation of the Colour Fastness Properties of Bixa orellana-Dyed Silk Using CIELab Analysis 2275
Nur Atiqah Jamaludin, Razanah Ramya, and Rashidi Othman

Partial Discharge Detection and Localisation in Refined, Bleached and Deodorised Palm Olein based on Ultra-High Frequency Sensing Technique 2293
Nur Darina Ahmad, Norhafiz Azis, Zhi Yang Lim, Jasronita Jasni, Khairil Anas Md Rezali, Mohd Aizam Talib, Nik Hakimi Nik Ali, and Masahiro Kozako

Recycling Indonesian Native Chicken Eggshell Membrane Waste into Gelatine: A Comparative Characterisation Using SEM-EDX, FTIR, and XRF 2311
Rina Wahyuningsih, Ragil Yuliatmo, Andi Febrisiantosa, Atiqa Rahmawati, Al Mizan, Selma Noor Permadi, Laila Rahmawati, Afi Candra Trinugraha, and Teguh Wahyono

Short Communication 2329
Regional Comparison of Forest Certification Impacts: Environmental and Social Outcomes in Asia-Pacific and Europe
Rohaiza Abdul Kadir Zailani, Sarah Najwa Mohd Masri, Shah Badri Mohd Nor, Ahmad Fariz Mohamed, and Shaharuddin Mohamad Ismail

Hybrid Approach Combining Area-Based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE) in Handling Class Imbalance 2339
Hartono, Muhammad Khahfi Zuhanda, Rahmad Syah, Prana Ugiana Gio, and Erianto Ongko

Impact of Wind Farm Penetration on the Mal-Operation of Distance Relay Power Swing Blocking and Out-of-Step Functions 2359
Chidiebere Okeke, Mohammad Lutfi Othman, Hashim Hizam, Mohd Zainal Abdin Ab Kadir, Samuel Nwagbara, Osaji Emmanuel, Jude Aneke, Okeke Ugochukwu Kingsley, and Masoud Ahmadipour

Characterisation of Bamboo Fibre-Reinforced Polyurethane Foam Composites with Varying Fibre Sizes 2385
A.M. Radzi, M.Y.M. Zuhri, S. Ayu Rafiqah, R.A. Ilyas, Muhammad Asyraf Muhammad Rizal, Ahmad Safwan Ismail, A.B.M. Supian, and Agusril Syamsir

Review Article 2403
Technological Approach for Biodiversity Conservation: Augmented Reality Based on Systematic Review
Nurfitri Ezatie Yusoff, and Arney Sapaat

Valorisation of Food Waste and Leftovers into Functional Bio-Compost: Optimisation of Maturation Period and Growth Performance of *Ipomoea Aquatica* 2421
Sarah Dini Muhamad Din, Muhammad Heikal Ismail, Tharsini Amma Puspadaran, Anie Yulistyorini, Muhammad Hazwan Hamzah, and Chuah Teong Guan

Mechanical Testing Analysis of Hybrid Composites of Oil Palm Empty Fruit Bunch Fibre and Sugarcane Bagasse Fibre for Environmentally Friendly Car Body Materials 2441
Aris Sandi, Aulia Majid, and Syafiq Fadhlika

Review Article 2457
AI-Powered Revolution in Concrete Crack Analysis: An In-Depth Review on Detection and Classification Methods
Muhammad Shahrulazmi Hashim, Emedya Murniwaty Samsudin, and Khalid Isa

Photovoltaic Module Defect Classification Analysis Based on Convolutional Neural Network Architecture with Augmentation Dataset 2487
Nurul Atikah Mazlan, Shahrani Shahbudin, and Murizah Kassim

Orchestration Network Hierarchy and Communication Patterns Architecture Analysis with Ranks for IoT Devices 2519
Nor Syazwani Mohd Pakhrudin, Murizah Kassim, and Azlina Idris

Foreword

Welcome to the fourth issue of 2026 of the Pertanika Journal of Science and Technology (PJST)!

PJST is an open-access journal for studies in Science and Technology published by Universiti Putra Malaysia Press. It is independently owned and managed by the university for the benefit of the world-wide science community.

This issue contains 27 articles: one short communication article; four review articles; and the rest are regular articles. The authors of these articles come from different countries namely Bangladesh, India, Indonesia, Japan, Malaysia, and Saudi Arabia.

The issue also features the article titled *"Data-Driven Fault Detection and Diagnosis of a Nuclear Reactor Cooling System Using Graph Neural Networks"*, which presents an intelligent approach to improving the monitoring and reliability of complex cooling systems. The research develops a Graph Neural Network-based Fault Detection and Diagnosis framework for the TRIGA PUSPATI Reactor cooling system using data generated from a scaled-down laboratory control rig that replicates the operational relationships of the reactor cooling process. By representing sensors and system components as interconnected nodes and edges, the GNN can capture relationships among temperature, pressure, and flow measurements that may not be adequately represented by conventional neural-network approaches. The proposed model was evaluated against an Artificial Neural Network under identical experimental conditions and demonstrated superior performance, achieving 98.12% accuracy, 99.15% precision, 99.26% recall, and a 98.20% F1-score, together with a substantially lower mean squared error. The system was further validated through offline and real-time experiments involving representative fault scenarios, where it successfully detected and localised faults while providing graphical indications of the affected components. By combining graph-based artificial intelligence with real-time sensor monitoring, this study contributes to the development of more responsive and data-driven fault diagnosis systems and demonstrates the potential of intelligent monitoring technologies to enhance predictive maintenance, operational reliability, and safety in complex industrial cooling environments. Full details of this study are available on page 2053.

The article entitled "Characterisation of Bamboo Fibre Reinforced Polyurethane Foam Composites with Varying Fibre Sizes" investigates the potential of bamboo fibre as a sustainable reinforcement for polyurethane foam composites and examines how fibre size influences their thermal, mechanical, and physical performance. The study fabricates composites using bamboo fibres in four size ranges, from below 150 μm to 425-600 μm,

and evaluates their thermal degradation behaviour, flexural and compressive properties, water absorption, thickness swelling, and moisture content. The findings demonstrate that fibre size plays an important role in determining composite performance. Among the bamboo fibre-reinforced formulations, BFP3, containing fibres within the 300–425 μm range, exhibited the highest flexural and compressive strengths at 0.192 MPa and 0.232 MPa, respectively, while also showing favourable thermal stability. In contrast, composites containing smaller fibres recorded lower water absorption, thickness swelling, and moisture content, highlighting the balance required between mechanical performance and dimensional stability. By demonstrating the potential of bamboo fibre as an environmentally friendly reinforcement material, this work contributes to the advancement of natural fibre composites for applications such as thermal insulation, acoustic products, and building components, while supporting efforts to reduce dependence on synthetic reinforcement materials. Further details of this study can be found on page 2385.

Another noteworthy contribution is the article entitled "Valorisation of Food Waste and Leftovers into Functional Bio-Compost: Optimisation of Maturation Period and Growth Performance of Ipomoea Aquatica", which addresses the growing challenge of food waste management through the production of functional bio-compost. The study applies the Bokashi composting method to food waste and leftover food and evaluates maturation periods of 30, 40, and 50 days based on physical and chemical properties as well as the growth performance of water spinach (*Ipomoea aquatica*). The results demonstrate that longer maturation periods generally improve compost quality through reduced bulk density, increased moisture content and water-holding capacity, improved nitrogen availability, and more favourable carbon-to-nitrogen ratios. Among the treatments, leftover food compost matured for 50 days, designated LF-50, produced the most favourable overall characteristics, including the lowest C ratio, the highest nitrogen and potassium concentrations, and the strongest water spinach growth performance. The LF-50 treatment achieved a crop growth rate of 12.85 g/m²/day and produced the highest number and size of leaves among the bio-compost treatments, approaching the performance obtained using commercial organic fertiliser. By transforming commonly discarded organic materials into a useful soil amendment, the study provides a practical approach to reducing landfill disposal while supporting nutrient recovery, household composting, and more sustainable agricultural practices. Further details of this study can be found on page 2421.

We anticipate that you will find the evidence presented in this issue to be intriguing, thought-provoking and useful in reaching new milestones in your own research. Please recommend the journal to your colleagues and students to make this endeavour meaningful.

All the papers published in this edition underwent Pertanika's stringent peer-review process involving a minimum of two reviewers comprising internal as well as external

referees. This was to ensure that the quality of the papers justified the high ranking of the journal, which is renowned as a heavily-cited journal not only by authors and researchers in Malaysia but by those in other countries around the world as well.

We would also like to express our gratitude to all the contributors, namely the authors, reviewers and Editorial Board Members of PJST, who have made this issue possible.

PJST is currently accepting manuscripts for upcoming issues based on original qualitative or quantitative research that opens new areas of inquiry and investigation.

Editor-in-Chief
Mohamed Thariq Hameed Sultan

Pertanika J. Sci. & Technol. 34 (4): 1953 - 1979 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Comprehensive Review of Plant Disease Detection: Advancements in Imaging Sensors, AI Techniques, and Future Directions in Smart Agriculture

Binal Upadhyay[1]* and Amit Ganatra[2]

[1]Department of Computer Engineering, Devang Patel Institute of Advanced Technology and Research (DEPSTAR), Faculty of Technology and Engineering, CHARUSAT University, Changa, 388421 Anand, Gujarat, India
[2]Director (Dean) R&D, Dean Faculty of Engineering & Technology, The Charutar Vidya Mandal University, 388120 V.V. Nagar, Gujarat, India

ABSTRACT

Food safety globally is threatened by crop disease, which creates a major obstacle to yield losses, so there is an urgent need for rapid, precise, and large-scale diagnostic methods for all the global risks crops are exposed to from disease. While imaging sensors, as well as Artificial Intelligence (AI), have made great strides in recognising plant disease, most literature does not have a comprehensive analysis that combines methods, technology, and implementation. Therefore, a systematic literature review follows PRISMA methods; we review 61 excellent studies published within the last five years that outline the advancement of imaging modalities (Red, Green, Blue (RGB), multispectral/hyperspectral, thermal), deep learning architectures, augmentation of data, explanation methods and IoT (Internet of Things)-edge-cloud for managing intelligent agriculture. These modern AI-based systems (AI systems) have consistently produced accurate results above 98%. However, there are problems with the generalisability (across hybrid plant species), robustness (when exposed to environmental stresses), and interpretability of the results presented to consumers. This review represents the first compilation of using imaging sensors, artificial intelligence models, Internet of Things architecture (IoT-edge), and robotics into one comprehensive framework for the detection of plant disease in the next generation. In addition, this review suggests future research directions, including lightweight edge-deployable models, multimodal sensor fusion, interpretable AI, larger validated datasets, and autonomous robotic systems for scalable and sustainable smart agriculture.

Keywords: Artificial intelligence, deep learning, imaging sensors, internet of things (IoT), plant disease detection, robotics, smart agriculture

ARTICLE INFO

Article history:
Received: 08 July 2025
Accepted: 02 May 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.01

E-mail addresses:
binal1994upadhyay@gmail.com (Binal Upadhyay)
dean.rnd@cvmu.edu.in (Amit Ganatra)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

There is no substitute for agriculture in terms of food safety, economic growth, and environmental sustainability. However, various plant diseases have caused a steady decline in crop yield, resulting in significant losses in terms of money. Traditional methods of planting, generally based on human seeing and expertise, both involve delays in diagnosing disease and ineffective treatment options for plant disease (Chakrabarty et al., 2024; Nigar et al., 2024). The current demands for more sophisticated, accurate, and efficient detection systems, therefore, become vital; Figure 1 presents a selection of the most common plant diseases for different crops. Recently, advances in imaging sensors and Deep Learning (DL) models have transformed plant disease identification by offering automated high-precision analytical solutions that reduce yield losses and improve decision-making for smart agriculture. Reviews today primarily discuss either specific DL algorithms or individual sensor types rather than provide a holistic overview of combined approaches (Raval & Chaki, 2024). Current limitations relate to underexamined areas such as data set bias, use of interpretable machine learning algorithms, and the ability to identify plant diseases with data distributions typically seen in field applications (sometimes referred to as representativeness). To address these challenges, this addresses these deficiencies in an informative manner that combines information on past achievements of imaging sensors and AI-related technologies to detect and classify plant diseases while identifying new opportunities for supporting smart agriculture.

Figure 2a shows the publication trends for 2020 to 2025, while Figure 2b shows the overall circulation of the large publishers.

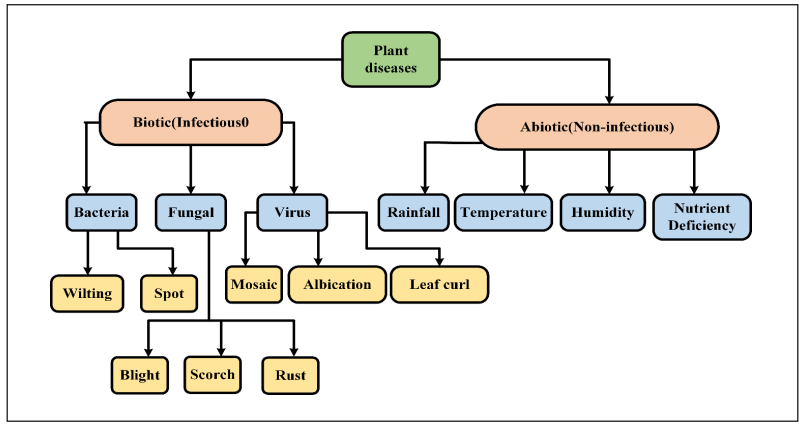
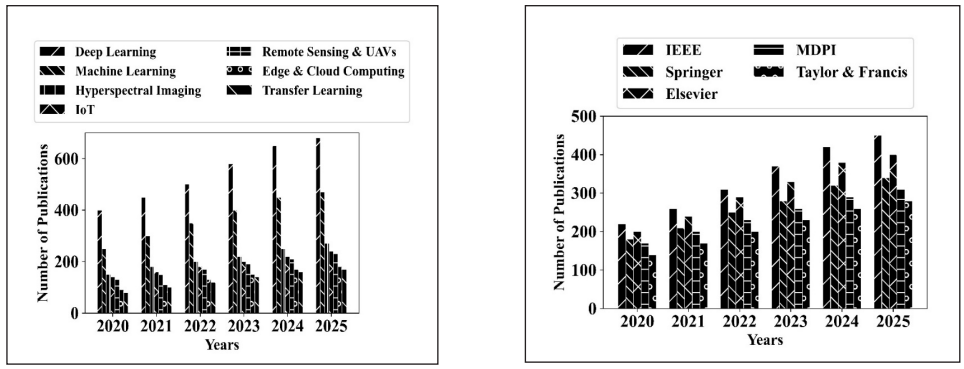


Figure 1. Overview of common plant diseases affecting various crops



(a) (b)

Figure 2. Publication trend for plant diseases (a) Different domains; (b) Different journals

Major Contributions

This review offers numerous substantial contributions to the field of plant disease detection and smart agriculture:

- The article summarises how improvements to sensors in RGB, multispectral, hyperspectral, thermal, and UAV-based systems contribute to the development of documentation for plant disease detection.
- This article compares current AI methods, CNNs, transformers, few-shot learning, ensembles, and explainable AI (XAI), for plant disease detection across different crops and environmental conditions.
- This review extends prior work by analysing how Internet of Things (IoT) sensors, edge computing devices, and autonomous robots enable scalable, real-time, in-field plant disease detection systems.
- In addition, this paper synthesises trends in performance across different plant species by documenting trends in accuracy, sensor suitability and model robustness.
- This review highlights key challenges in plant disease detection, including limited datasets, domain shift affecting transferability, and poor model interpretability, which hinder large-scale practical deployment.
- This review presents a unified framework integrating imaging sensors with AI and IoT/edge/robotic systems, offering an end-to-end approach not addressed in prior studies.

The organisation of the survey is shown in Figure 3.

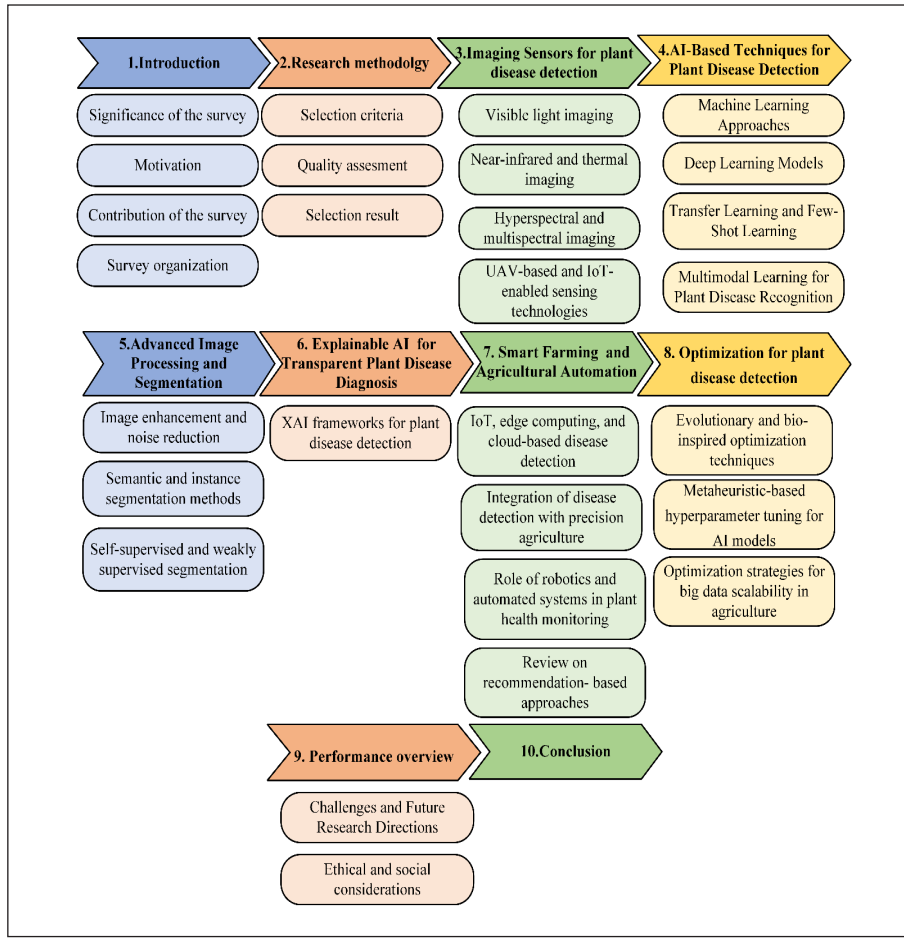


Figure 3. Organisation of the survey

RESEARCH METHODOLOGY

Research Questions

To ensure a structured and focused review, this study is guided by the following Research Questions (RQs).

RQ1 : What types of imaging sensors can be used to detect plant diseases, and how do these sensors differ in terms of how accurately they can detect plant diseases or provide early detection capabilities, and their robustness to the environment?

RQ2 : What types of AI/DL methods/algorithms appear to perform best across different crops, datasets and plant diseases, and what factors can affect the performance of these methods?

RQ3 : What are some of the major challenges (e.g., data set limitations, variations between fields, interpretability and generalisability) that have been identified in recent literature?

RQ4 : How are IoT, edge devices, and robotics integrated with AI models to enable real-time and scalable deployment of plant disease detection systems?

RQ5 : What type of future research will provide more accurate, interpretable, and easier-to-use plant disease detection systems?

Search Strategy and Databases

A systematic review of relevant research on the use of AI for identifying plant diseases was conducted across multiple databases, including IEEE Xplore, Springer, MDPI, ScienceDirect, Taylor and Francis, and PubMed. Relevant keywords involved "plant disease detection," "deep learning in agriculture," "imaging sensors for crop health," and "smart farming techniques."

Inclusion and Exclusion Criteria

The search yielded many studies, which were filtered through title/abstract screening and full-text review based on predefined criteria. Only peer-reviewed English-language articles (2020 onward) on AI/IoT-based plant disease detection were included, with duplicates and irrelevant studies removed to retain high-quality research. The full details of the inclusion and exclusion criteria for article/study selection are accessible in Table 1.

Table 1
Inclusion and exclusion criteria

S. No.	Inclusion	Exclusion
1	Available in peer-reviewed journals	Non-English papers
2	Published after the year 2020	Book chapters
3	Direct relevance to plant disease detection in AI and IoT-based agricultural domains	Little or no focus on AI-based plant disease solutions
4	Articles addressing research questions	Duplicate articles

PRISMA Process

This review followed the PRISMA 2020 guidelines to ensure a transparent and reproducible selection process. The complete selection workflow is shown in Figure 4.

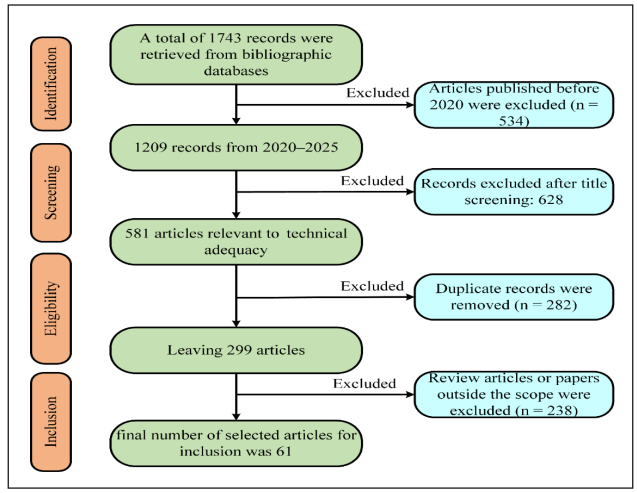


Figure 4. PRISMA process

Data Extraction and Study Categorisation

Data extraction classified studies by crop type, disease, sensors (RGB, hyperspectral, multispectral, thermal/IR), dataset details, and model architectures with performance metrics.

Performance Evaluation Metrics

Model performance was evaluated using multiple metrics, including precision, recall, F1-score, and specificity to balance false positives and false negatives, especially in imbalanced datasets. Confusion matrices further analysed misclassification patterns, providing a comprehensive assessment of plant disease classification performance beyond overall accuracy.

Quality Assessment

The excellence of the studies selected was evaluated using an assortment of criteria.

- **Application Area:** The research application for your plant disease identification is focused specifically on agriculture through imaging sensors, AI, and IoT technologies.
- **Objectives:** These articles state clearly their goals for research; they identify specific AI-related challenges to diagnosing plant diseases and present promising solutions.
- **Techniques:** These articles should utilise advanced AI techniques, including DL, Machine Learning (ML), Transfer Learning (TL) or hybrid approaches using imaging sensor technologies to improve the accuracy of the disease identification process.

- **Security Measures:** The research evaluates how models perform, scale, and can be applied to real-world situations, as well as environmental variability, diversity of the datasets, and constraints on operations being evaluated.

Imaging Sensors for Plant Disease Detection

Visible Light Imaging

RGB imaging is a low-cost, scalable method for plant disease detection, but it is limited to visible symptoms and cannot reliably detect early or differentiate stress types. ML/DL and multimodal fusion with hyperspectral or thermal data improve accuracy and early detection.

Near Infrared and Thermal Imaging

Near Infrared (NIR) and thermal imaging enable non-invasive, early plant disease detection by capturing physiological changes such as water stress, chlorophyll variation, and temperature anomalies. Table 2 presents an overview of the imaging sensors for plant diseases.

Table 2
Summary of key contributions in numerous applications of imaging sensors for plant disease diagnosis

Target	Refs	Crop	Disease	Imaging Type/ Sensors	Methods and Algorithms
Early disease detection & classification	Zhang et al., 2020	Rice	Leaf blast	Hyperspectral imaging, EM285CL EMCCD camera	Support Vector Machine (SVM)-based Spectral analysis for disease severity estimation.
	Yong et al., 2022	Oil palm	Basal stem rot	Hyperspectral imaging, a Cubert FireflEYE S185 hyperspectral camera	Region-Based Convolutional Neural Network (RCNN) based classification.
	Chen et al., 2024	Rice	Bakanae disease	Hyperspectral imaging (UAV-based)	CNN
	Zhang et al., 2024	Tomato	Bacterial leaf spot	Hyperspectral imaging	Spectral-based feature extraction + ML classifiers
	Hsiao et al., 2025	Soybean	Fungal leaf diseases	Hyperspectral imaging	CNN-based spectral-spatial classification

Table 2 (*continued*)

Target	Refs	Crop	Disease	Imaging Type/ Sensors	Methods and Algorithms
Disease severity assessment & localisation	Zhang et al., 2022	Rice	Leaf blast	Hyperspectral imaging, EM285CL EMCCD camera	Classification across multiple growth stages by the Successive Projections Algorithm (SPA) and SVM.
	Li et al., 2023	Grape	Leaf diseases	UAV-based imaging	improved U-net and Visual Geometry Group (VGG)-19 -based detection and localisation.
UAV & IoT-enabled disease monitoring	Kerkech et al., 2020	Grapevine	Various diseases	UAV-based multispectral imaging, Quadcopter drone	Optimised image registration, DL segmentation.
Advanced spectral analysis & chemometrics	Gloerfelt-Tarp et al., 2023	Cannabis sativa	N/A	NIR imaging	NIR spectroscopy, chemometric analysis.
Real-time monitoring	Rahman et al., 2025	Multi-crop	Multiple leaf diseases	RGB imaging (real-time camera system)	Deep CNN models for real-time detection
Multi-dataset detection	Krishna et al., 2025	Multiple crops	Multiple leaf diseases	RGB imaging + multi-dataset inputs	DL (CNN models)
Asymptomatic & severity grading	Chen et al., 2025	Tomato	Tomato yellow leaf curl disease	Hyperspectral imaging	3D spectral-spatial deep learning model

Hyperspectral and Multispectral Imaging

Hyperspectral imaging enables early detection via fine biochemical changes, while multispectral Imaging provides cost-effective wide-area UAV monitoring, with AI improving classification but limited by calibration and processing challenges. Chen et al. (2024) and Zhang et al. (2024) showed early rice bakanae and tomato leaf spot detection via hyperspectral imaging, while Hsiao et al. (2025) confirmed its efficacy for soybean fungal disease using CNNs. Rahman et al. (2025) built a real-time multi-crop system, Krishna et al. (2025) improved generalisation with

multi-dataset fusion, and Chen et al. (2025) detected asymptomatic tomato yellow leaf curl disease with severity grading.

UAV-based and IoT-Enabled Sensing Technologies

UAV-based multispectral/hyperspectral and IoT systems enable continuous, high-resolution plant disease monitoring and early detection, while AI-driven integration improves real-time prediction and management despite challenges like battery, weather, and transmission limits.

AI-based Techniques for Plant Disease Recognition

Machine Learning Approaches

ML techniques enhance plant disease recognition by automating feature extraction and classification, with methods like K-Nearest Neighbours (KNN), SVM, and the Random Forest (RF) effectively distinguishing healthy and diseased leaves from extracted image features.

Deep Learning Models

Deep learning architectures enable high-performance plant disease recognition through advanced neural networks for accurate and efficient classification. Boyar et al. (2025) proved that the U-Net variants performed best for hazelnut leaf disease recognition from drone imagery. Yilmaz et al. (2025) reviewed recent advancements and found that CNNs, U-Net models, and transformer-based techniques consistently achieved higher accuracy in plant disease detection.

Boyar and Yıldız (2022) also confirmed that CNN (You Only Look Once (YOLOv5)) models effectively detected powdery mildew in hazelnut leaves. CNNs and their variants (DenseNet, ResNet, EfficientNet) improve leaf-image feature extraction and accuracy, while capsule networks capture spatial details (Kwabena et al., 2020). YOLO/SSD enable real-time detection, attention and transformer models enhance feature selection, and lightweight/ensemble architectures support efficient edge deployment for smart agriculture systems, as shown in Figure 5.

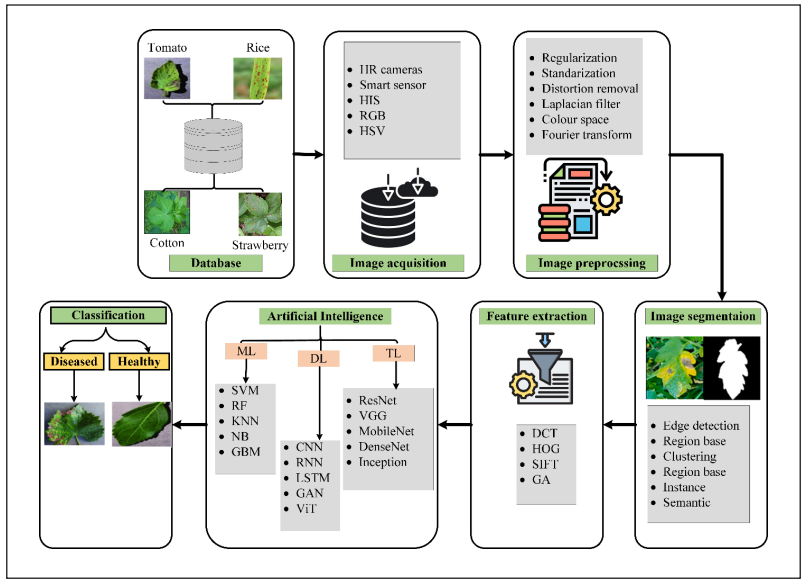


Figure 5. AI-based plant disease detection

Transfer Learning and Few-Shot Learning

Transfer learning, domain adaptation, few-/zero-shot learning, transformers, and diffusion-based models improve plant disease detection by enhancing feature extraction, handling class imbalance, enabling learning from limited data, and improving robustness in low-data scenarios. The overview of AI-constructed techniques for plant disease identification is provided in Table 3.

Multimodal Learning for Plant Disease Recognition

Multimodal fusion using RGB–hyperspectral IoT data with attention and dilated convolutions improves robustness, while rice-fusion models and multimodal Mixup further enhance real-time generalisation for large-scale plant disease detection.

Image Processing and Segmentation Techniques

Image Enhancement and Noise Reduction

Noise-removal and image-enhancement methods improve disease classification by cleaning distortions, where Leaf Artefact-Suppression Super Resolution (LASSR) (Cap et al., 2021) boosts resolution for better feature extraction, and generative lesion-simulation techniques augment data to handle lighting and background variability (Sun et al., 2020).

Table 3
Overview of AI-constructed techniques for plant disease identification

Crop	Refs	Dataset	No. of Images	No. of Classes	Algorithm Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Misclassification Insights
Corn	Noola & Basavaraju, 2022	Plant-Village dataset	3820	4	Enhanced k-nearest neighbour (EKNN)	99.60	Not reported	99.60	Not reported	Minor confusion among similar foliar diseases
Apple	Sun et al., 2021	AppleDisease5 (Custom dataset)	6,500	5	light-weight CNN	96.9	Not reported	Not reported	Not reported	Early-stage symptoms harder to classify
	Vishnoi et al., 2022	Plant Village dataset	3171	3	CNN	98	Not reported	Not reported	Not reported	Not reported
Hazelnut	Boyar et al., 2025	Drone imagery from hazelnut orchards	3,200	3	U-Net variants (U-Net++, Attention U-Net, ResU-Net)	96.8	Not reported	Not reported	Not reported	Some overlap in mild and moderate infection
	Boyar & Yıldız, 2022	Hazelnut leaf image dataset	1,050	2	YOLOv5	90	88	90	89	False positives under dense foliage
Grape	Zinonos et al., 2021	Custom dataset	5,000	3	CNN, Low-Rank Adaptation (LoRa) technology	97.8	97.5	97.3	97.4	Minor confusion in mildew subtypes
Potato	Lee et al., 2021	PlantVillage	2152	3	CNN	97.68	Not reported	Not reported	Not reported	Not reported
	Mahum et al., 2023	Plant Village dataset	2,152	3	Modified DenseNet-201	97.2	Not reported	Not reported	Not reported	Early blight misclassified in small lesions

Table 3 (*continued*)

Crop	Refs	Dataset	No. of Images	No. of Classes	Algorithm Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Misclassification Insights
Sugarcane	Srivastava et al., 2020	Custom dataset	160	4	Inception v3, VGG-16 and VGG-19	90.2	Not reported	Not reported	Not reported	Significant confusion due to a small dataset
Multi-Crop	Ganatra & Patel, 2020	plant village dataset	14956	10	RF, SVM, KNN, Artificial Neural Network (ANN)	63.20 - 73.38	Not reported	Not reported	Not reported	High confusion due to hand-crafted features
	Alirezazadeh et al., 2023	pilot dataset	3,006	4	Attention-based CNN	86.89	86.12	86.31	86.41	Misclassifications in visually similar leaves
	Ahmad et al., 2022	PlantVillage and pepper disease dataset	99,507	38	CNN with stepwise TL	99	Not reported	98	98	Confusion in rare classes
	Xu et al., 2022	PlantCLEF2022	2,885,052	1,000+	TL	99.70	98.24	98.24	98.24	Confusion among classes with similar morphology
	Li & Chao, 2021	PlantVillage	38,000	10	Semi-supervised Few-Shot Learning	97.50	Not reported	Not reported	98.1	Not reported
	Singh & Sanodiya, 2023	PlantDoc	12,345	17	Zero-Shot Transfer Learning	92.85	92.6	93.52	92.86	Difficulty with unseen classes
	Yilmaz et al., 2025	Survey-based (no single dataset)	-	-	CNNs, U-Net, Transformers	-	-	-	-	Not a dataset study

Table 3 (*continued*)

Crop	Refs	Dataset	No. of Images	No. of Classes	Algorithm Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Misclassification Insights
Field crops	Argüeso et al., 2020	Field Images (Custom)	10,000	5	Few-Shot Learning (Prototypical Networks)	93.45	94.25	93.15	94.22	Difficulty in cluttered backgrounds
Cotton	Rajasekar et al., 2024	Custom dataset	5,100	4	ResNet, ImageNet and Xception	95.4	95.2	Not reported	95.4	Confusion in overlapping leaves
Rice	Patil & Kumar, 2022	RiceLeaf Dataset	6,732	5	Multimodal data fusion framework	96.35	96.32	Not reported	97.13	Not reported
Sunflower	Zhou et al., 2025	Custom (Sunflower Disease)	9,820	6	Few-Shot Learning with Diffusion Models	97.10	97.10	97.63	97.10	Minor confusion in similar lesions
Wheat	Rezaei et al., 2024	Kaggle - Wheat Disease	5,912	5	Few-Shot Learning Model-Agnostic Meta-Learning (MAML)	94.28	Not reported	Not reported	95.21	Confusion in rust vs. mildew
Barley	Rezaei et al., 2025	Field Images (Barley)	8,214	6	Transformer-based Few-Shot Learning	96.20	95.32	96.20	96.42	Some overlap in blight categories

Semantic and Instance Segmentation Methods

DL segmentation models (e.g., U-Net, Mask R-CNN) accurately localise diseased regions, while advanced hybrid and two-stage methods improve feature extraction, lesion analysis, and overall detection precision. The overall segmentation is summarised in Table 4.

Table 4
Overview of segmentation techniques for plant disease recognition

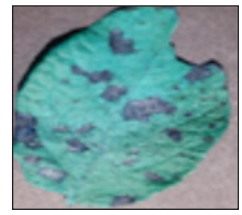
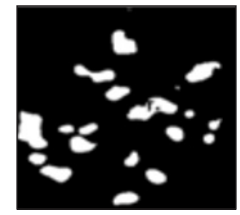
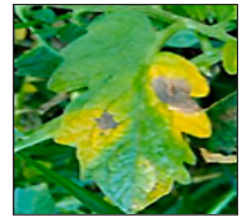
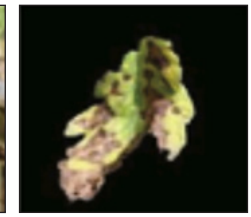
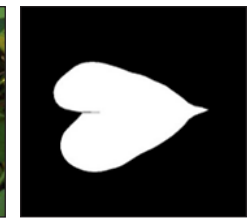
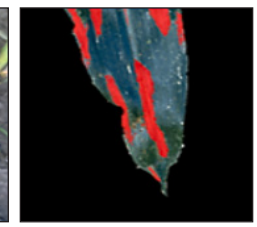
Crops	Refs	Techniques Used	Sample Image	Segmentation Output	Segmentation Accuracy (%)
Banana	Krishnan et al., 2022	Total Generalised Variation Fuzzy C means (TGVFCMS)			94.2
Strawberry	Bhujel et al., 2022	U-Net			93.5
Potato	Arshad et al., 2023	U-Net			95.3
Tomato	Shoaib et al., 2022	U-Net and Modified U-Net			92.8

Table 4 (*continued*)

Crops	Refs	Techniques Used	Sample Image	Segmentation Output	Segmentation Accuracy (%)
Cash crop	Xiong et al., 2020	Automatic Image Segmentation Algorithm (AISA)			81.4
Multi-crop	Zhang et al., 2021	Joint segmentation			90.1
Strawberry	Afzaal et al., 2021	Mask R-CNN			84.63
Corn	Divyanth et al., 2023	Segmented Network (SegNet), UNet, and DeepLabV3+			93.8
Apple	Storey al., 2022	Mask R-CNN			96.1

Self-Supervised and Weakly Supervised Segmentation

Self- and weakly supervised methods cut annotation needs by learning strong features from unlabelled images; contrastive learning boosts representations, LASSR improves image quality, and lesion-based augmentation increases robustness. Weakly supervised models like PLDPNet and Mask R-CNN variants enhance disease localisation, while modern segmentation frameworks further improve crop-disease quantification.

Explainable AI (XAI) for Transparent Plant Disease Diagnosis

XAI Frameworks for Plant Disease Detection

An overview of the XAI patterns is included in Table 5.

Table 5
Interpretations of explainable AI summaries across various studies

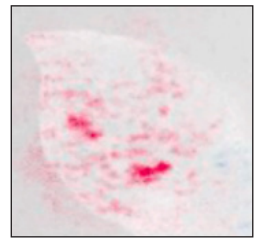
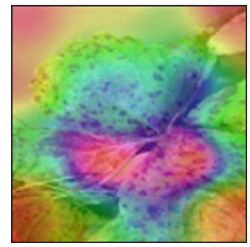
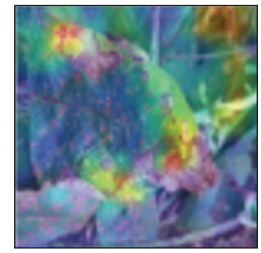
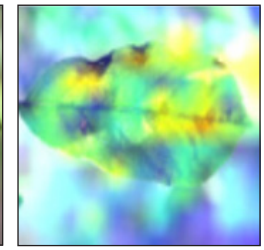
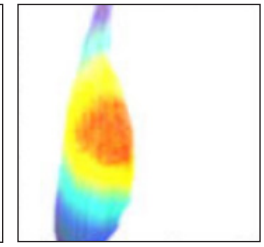
Crops	Refs	Original Image	XAI Output	Explainability Model
Potato	Anim-Ayeko et al., 2023			SHapley Additive exPlanations (SHAP)
Common leaf	Xiang et al., 2021			Gradient-weighted Class Activation Mapping (Grad-CAM)
Sunflower	Ghosh et al., 2023			Local Interpretable Model-Agnostic Explanations (LIME)

Table 5 (*continued*)

Crops	Refs	Original Image	XAI Output	Explainability Model
Various crops	Prince et al., 2024			Grad-CAM
Apple	Thakur et al., 2022			Grad-CAM
Tomato	Bhandari et al., 2023			LIME
Rice Leaf	Shovon et al., 2023			Grad-CAM
Various crops	Dai et al., 2024			SHAP

Smart Farming and Agricultural Automation

IoT, Edge Computing, and Cloud-Based Disease Detection

Edge computing enables real-time prediction, as shown in PFDI (Dhiman et al., 2023) and the IoT framework in Figure 6, with Edge AI improving scalability (Park, 2021). Faster R-CNN, ensembles, and thermal-imaging DL models also run effectively on edge devices for fast, accurate disease detection (da Silva & Almeida, 2024; Dhiman et al., 2023).

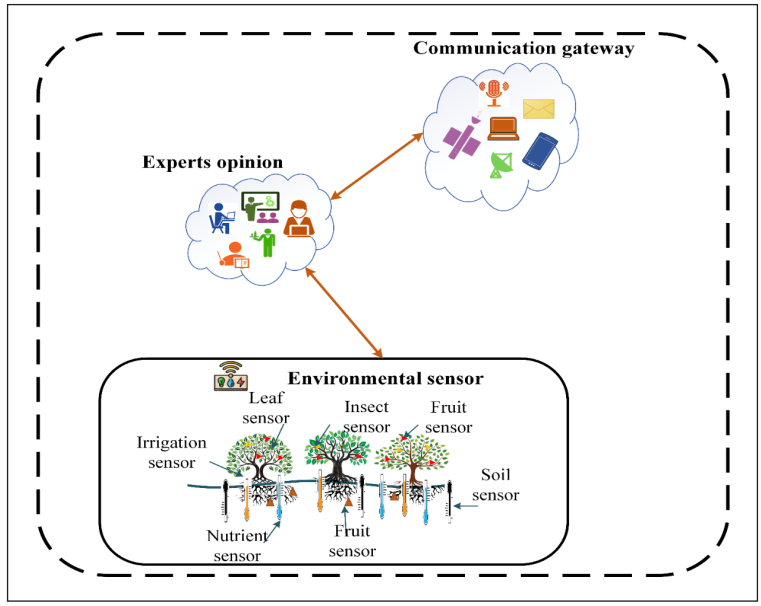


Figure 6. An IoT-based network for smart farming

ML–IoT integration boosts predictive farming, with cloud training, edge devices for on-site diagnosis, and smart IoT platforms automating irrigation and disease alerts together, enabling fast, scalable, and reliable plant-disease detection.

Integration of Disease Identification with Precision Farming

AI, IoT, and digital twins together enable fast disease diagnosis, continuous field monitoring, and predictive decision-making, with frameworks like Agrilora supporting proactive, precision crop management.

Role of Robotics and Automated Systems in Plant Health Monitoring

Robotic platforms with DL-driven vision enable rapid, accurate disease detection and real-time crop monitoring, efficiently assessing traits like stem measurements to advance precision agriculture.

Review of Recommendation-based Approaches

DL-based recommendation systems provide real-time disease diagnosis and optimised pesticide/fertiliser suggestions, with models like Transition Probability Function and Convolutional Neural Network (TPF-CNN) tailoring inputs based on crop, soil, and environmental conditions.

Optimisation for Plant Disease Detection

Evolutionary and Bio-Inspired Optimisation Techniques

This review explores different nature-inspired optimisation techniques, namely Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Differential Evolution (DE), for improving plant disease detection models. The obtained Particle Swarm Optimisation-based Deep Neural Network (PSO-DNN) combinations proved efficient in detecting paddy leaf diseases without increasing the model complexity, but with high precision. Similarly, Machine Learning combined with a Genetic Algorithm (ML-GA) hybrid methods have been developed for crop forecasting, thus further enhancing decision-making in agricultural surveillance.

Metaheuristic-Based Hyperparameter Tuning for AI Models

Hyperparameter optimisation with methods like Bayesian Optimisation (BO), Quantum Whale Optimisation (QWO), and PSO boosts model performance by efficiently tuning network parameters, while attention-based networks paired with BO improve precision in rice disease classification. Advanced search strategies further enhance feature learning in optimised DL models.

Optimisation Strategies for Big Data Scalability in Agriculture

Agricultural big data is now managed through distributed and federated learning, supported by cloud–edge hybrids that process IoT and remote-sensing data efficiently. Metaheuristic optimisations such as parallel PSO, genetic programming, and RL improve AI decisions, while edge computing enables fast, real-time disease detection with minimal cloud reliance.

Performance Overview

Figure 7 shows the sample distribution, with Rice the most represented crop at 18.5% and Citrus the least at 1.8%, indicating that research predominantly focuses on major staple crops.

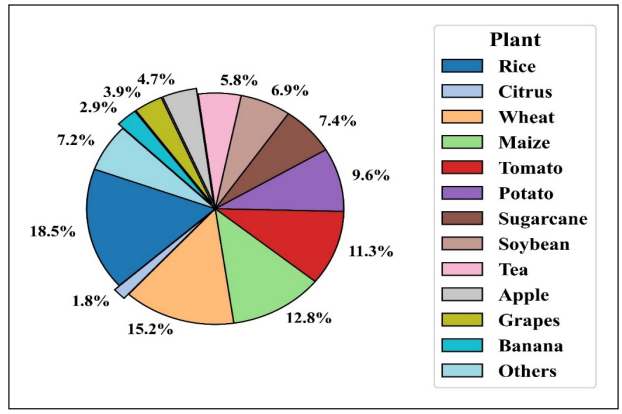


Figure 7. Percentage distribution of samples per plant across datasets

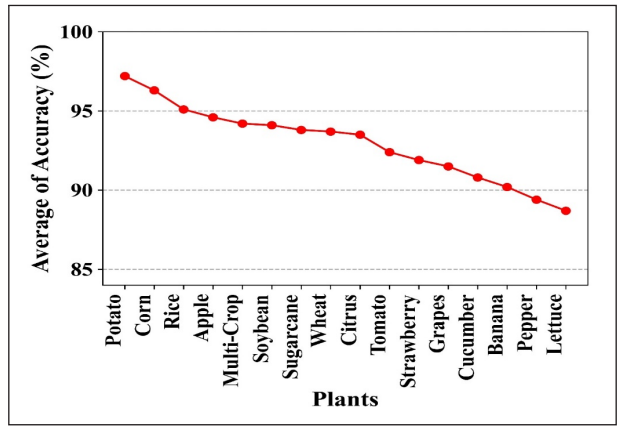


Figure 8. Average disease detection accuracy (%) achieved for each plant type based on the analysed studies

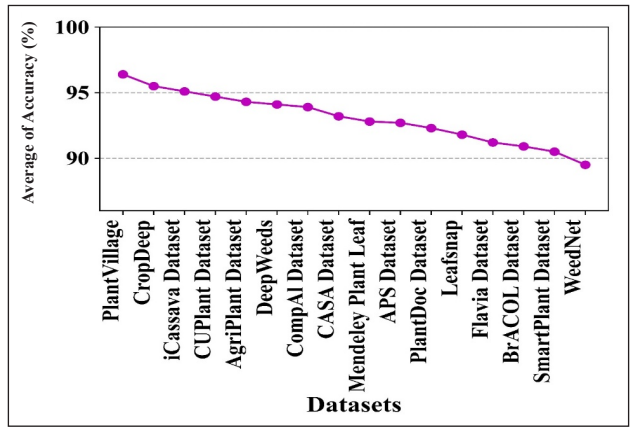


Figure 9. Average disease detection accuracy (%) of different plant disease datasets across multiple plant categories

As shown in Figures 8 and 9, detection accuracy varies across crops and datasets. Apple shows the highest performance, Banana the lowest, and strong datasets like PlantVillage yield higher accuracy than weaker ones like WeedNet, confirming that dataset quality directly influences model results.

Identified Research Gaps and Future Research Scope

Smart farming adoption remains limited due to high computational demands, real-world variability, and narrow crop/disease coverage, as highlighted in Figure 10, showing that further technical improvements are still required for practical deployment.

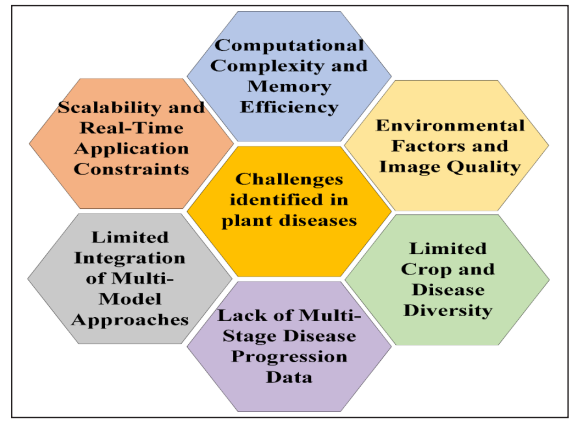


Figure 10. Challenges identified in plant diseases

To solve these problems, future research will focus on

- An efficient architecture for deep learning methods with optimised computation requirements would make practical use of systems used for agronomic monitoring systems.
- Incorporating multispectral and infrared cameras increases disease detection accuracy by filtering environmental noise from the images and improving clarity.
- Future studies will incorporate diverse crops, multiple plant parts (leaves, roots, stems), and a broader spectrum of diseases to improve model generalisation.
- A dataset with annotations for different stages of disease will allow for the creation of models to be able to estimate disease severity and provide better means of management.
- The combination of multiple pre-trained architectures in an ensemble configuration will create improved accuracy in detection, improved robustness in classification, and enhanced capability for predicting severity.
- The use of robotic applications with AI-based disease detection systems will allow for large-scale automated monitoring, reduce the need for manual intervention and provide increased precision.
- Going forward, the emphasis of future work will continue to be a holistic integration of disease detection, classification and severity assessment to provide an improved basis for making decisions in agriculture.

Ethical and Social Considerations

With smarter sensors and AI diagnostics, concerns about data privacy and misuse rise, making fairness and interpretability essential. Explainable AI shows growers how decisions are made, reducing bias and increasing trust so they can adopt AI recommendations with confidence.

Ensuring sustainable and equitable AI in crop disease detection requires farmer training, affordable tools, strong digital literacy, and supportive policies that enable widespread adoption of these technologies.

CONCLUSION

The review shows that advances in imaging sensors, AI models, and smart agricultural systems have rapidly improved plant disease detection. Analysis of 61 studies indicates that CNNs, transformers, and ensemble models often exceed 98% accuracy; hyperspectral, multispectral, and thermal imaging significantly strengthen early detection; and IoT-based monitoring with edge computing enables real-time, scalable deployment in field conditions.

Lightweight edge-AI solutions support low-latency field monitoring, multimodal sensor fusion increases robustness under variable environments, and explainable AI builds user trust. Together, these developments are driving more autonomous, faster, and resource-efficient disease management within precision agriculture.

Significant gaps persist, including domain shifts between lab datasets and real-field conditions, small or imbalanced datasets limiting generalisation, and weak interpretability, robustness, and energy efficiency in deployed models. The absence of unified evaluation metrics further restricts fair comparison across studies.

Future work should prioritise lightweight edge-AI models, diverse multi-stage field datasets, rugged multimodal sensing, stronger Explainable AI, cross-domain adaptation, and autonomous robotic monitoring supported by closer collaboration among agronomists, AI researchers, and sensor engineers.

From a business standpoint, these advancements will accelerate precision farming, remote crop monitoring, and automated early-warning systems. Integrating AI with IoT and robotics reduces labour costs, improves disease management, and stabilises yield. With advanced sensors, edge AI, and interpretable models, both large and small farmers gain scalable, cost-effective, and sustainable solutions as agriculture shifts toward fully digitised production.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Computer Engineering, Devang Patel Institute of Advanced Technology and Research (DEPSTAR),

Faculty of Technology and Engineering, CHARUSAT University, for providing the academic support and resources necessary for this review. The authors also acknowledge the researchers and scholars whose valuable contributions to plant disease detection, imaging technologies, artificial intelligence, and smart agriculture formed the foundation of this work.

Data Availability Statements

Data sharing does not apply to this article, as no datasets were generated or analysed during the current study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Afzaal, U., Bhattarai, B., Pandeya, Y. R., & Lee, J. (2021). An instance segmentation model for strawberry diseases based on Mask R-CNN. *Sensors, 21*(19), Article 6565. https://doi.org/10.3390/s21196565

Ahmad, M., Abdullah, M., Moon, H., & Han, D. (2021). Plant disease detection in imbalanced datasets using efficient convolutional neural networks with stepwise transfer learning. *IEEE Access, 9*, 140565-140580. https://doi.org/10.1109/ACCESS.2021.3119655

Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. (2023). Improving deep learning-based plant disease classification with attention mechanism. *Gesunde Pflanzen, 75*(1), 49-59. https://doi.org/10.1007/s10343-022-00796-y

Anim-Ayeko, A. O., Schillaci, C., & Lipani, A. (2023). Automatic blight disease detection in potato (*Solanum tuberosum* L.) and tomato (*Solanum lycopersicum* L. 1753) plants using deep learning. *Smart Agricultural Technology, 4*, Article 100178. https://doi.org/10.1016/j.atech.2023.100178

Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M. G., Bereciartua, A., & Alvarez-Gila, A. (2020). A few-shot learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture, 175*, Article 105542. https://doi.org/10.1016/j.compag.2020.105542

Arshad, F., Mateen, M., Hayat, S., Wardah, M., Al-Huda, Z., Gu, Y. H., & Al-Antari, M. A. (2023). PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction. *Alexandria Engineering Journal, 78*, 406-418. https://doi.org/10.1016/j.aej.2023.07.076

Bhandari, M., Shahi, T. B., Neupane, A., & Walsh, K. B. (2023). BotanicX-AI: Identification of tomato leaf diseases using an explanation-driven deep learning model. *Journal of Imaging, 9*(2), Article 53. https://doi.org/10.3390/jimaging9020053

Bhujel, A., Khan, F., Basak, J. K., Jaihuni, M., Sihalath, T., Moon, B.-E., Park, J., & Kim, H.-T. (2022). Detection of grey mould disease and its severity on strawberry using deep learning networks. *Journal of Plant Diseases and Protection, 129*(3), 579-592. https://doi.org/10.1007/s41348-022-00578-8

Boyar, T., Turan, S. C., Çıplak, Z., Yıldız, S. G., Sarıkaş, A., Yıldız, K., & Demir, Ö. (2025, June). Recognition of leaf diseases in hazelnut orchards with drone imagery by deep learning models: Effectiveness of U-Net model variants. In *2025 International Conference on Computer Systems and Technologies (CompSysTech)* (pp. 1-6). IEEE. https://doi.org/10.1109/CompSysTech65493.2025.11136971

Boyar, T., & Yıldız, K. (2022). Powdery mildew detection in hazelnut with deep learning. *Hittite Journal of Science and Engineering, 9*(3), 159-166. https://doi.org/10.17350/HJSE19030000267

Cap, Q. H., Tani, H., Kagiwada, S., Uga, H., & Iyatomi, H. (2021). LASSR: Effective super-resolution method for plant disease diagnosis. *Computers and Electronics in Agriculture, 187*, Article 106271. https://doi.org/10.1016/j.compag.2021.106271

Chakrabarty, A., Ahmed, S. T., Islam, M. F., Aziz, S. M., & Maidin, S. S. (2024). An interpretable fusion model integrating lightweight CNN and transformer architectures for rice leaf disease identification. *Ecological Informatics, 82*, Article 102718. https://doi.org/10.1016/j.ecoinf.2024.102718

Chen, J., Zhong, W., Yue, X., Ding, Z., Du, M., Li, X., … Tang, Y. (2025). Hyperspectral imaging-driven deep learning approach: Asymptomatic stage detection and severity grading of tomato yellow leaf curl disease. *Smart Agricultural Technology, 10*, Article 101281. https://doi.org/10.1016/j.atech.2025.101281

Chen, S., Lu, X., Fang, H., Perumal, A. B., Li, R., Feng, L., … Liu, Y. (2024). Early surveillance of rice bakanae disease using deep learning and hyperspectral imaging. *Abiotech, 5*(3), 281-297. https://doi.org/10.1007/s42994-024-00169-1

da Silva, P. E. C., & Almeida, J. (2024). An edge computing-based solution for real-time leaf disease classification using thermal imaging. *IEEE Geoscience and Remote Sensing Letters, 21*, Article 5508905. https://doi.org/10.1109/LGRS.2024.3456637

Dai, G., Tian, Z., Fan, J., Sunil, C. K., & Dewi, C. (2024). DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification. *Computers and Electronics in Agriculture, 216*, Article 108481. https://doi.org/10.1016/j.compag.2023.108481

Dhiman, P., Kaur, A., Hamid, Y., Alabdulkreem, E., Elmannai, H., & Ababneh, N. (2023). Smart disease detection system for citrus fruits using deep learning with edge computing. *Sustainability, 15*(5), Article 4576. https://doi.org/10.3390/su15054576

Dhiman, P., Manoharan, P., Lilhore, U. K., Alroobaea, R., Kaur, A., Iwendi, C., Alsafyani, M., Baqasah, A. M., & Raahemifar, K. (2023). PFDI: A precise fruit disease identification model based on context data fusion with Faster R-CNN in an edge computing environment. *EURASIP Journal on Advances in Signal Processing, 2023*(1), Article 72. https://doi.org/10.1186/s13634-023-01025-y

Divyanth, L. G., Ahmad, A., & Saraswat, D. (2023). A two-stage deep-learning-based segmentation model for crop disease quantification based on corn field imagery. *Smart Agricultural Technology, 3*, Article 100108. https://doi.org/10.1016/j.atech.2022.100108

Ganatra, N., & Patel, A. (2020). A multiclass plant leaf disease detection using image processing and machine learning techniques. *International Journal on Emerging Technologies, 11*(2), 1082-1086.

Ghosh, P., Mondal, A. K., Chatterjee, S., Masud, M., Meshref, H., & Bairagi, A. K. (2023). Recognition of sunflower diseases using hybrid deep learning and its explainability with AI. *Mathematics, 11*(10), Article 2241. https://doi.org/10.3390/math11102241

Gloerfelt-Tarp, F., Hewavitharana, A. K., Mieog, J., Palmer, W. M., Fraser, F., Ansari, O., & Kretzschmar, T. (2023). Using a global diversity panel of *Cannabis sativa* L. to develop a near infrared-based chemometric application for cannabinoid quantification. *Scientific Reports, 13*(1), Article 2253. https://doi.org/10.1038/s41598-023-29148-0

Hsiao, C. F., Feyrer, G., & Stein, A. (2025). Classifying early-stage soybean fungal diseases on hyperspectral images using convolutional neural networks. *Smart Agricultural Technology, 11*, Article 101023. https://doi.org/10.1016/j.atech.2025.101023

Kerkech, M., Hafiane, A., & Canals, R. (2020). Vine disease detection in UAV multispectral images using optimised image registration and deep learning segmentation approach. *Computers and Electronics in Agriculture, 174*, Article 105446. https://doi.org/10.1016/j.compag.2020.105446

Krishna, M. S., Machado, P., Otuka, R. I., Yahaya, S. W., Neve's dos Santos, F., & Ihianle, I. K. (2025). Plant leaf disease detection using deep learning: A multi-dataset approach. *J, 8*(1), Article 4. https://doi.org/10.3390/j8010004

Krishnan, V. G., Deepa, J. R., Rao, P. V., Divya, V., & Kaviarasan, S. (2022). An automated segmentation and classification model for banana leaf disease detection. *Journal of Applied Biology & Biotechnology, 10*(1), 213-220. https://doi.org/10.7324/JABB.2021.100126

Kwabena, P. M., Weyori, B. A., & Mighty, A. A. (2020). Gabor capsule network for plant disease detection. *International Journal of Advanced Computer Science and Applications, 11*(10), 350-358. https://doi.org/10.14569/IJACSA.2020.0111048

Lee, T.-Y., Lin, I.-A., Yu, J.-Y., Yang, J., & Chang, Y.-C. (2021). High-efficiency disease detection for potato leaf with a convolutional neural network. *SN Computer Science, 2*(4), Article 297. https://doi.org/10.1007/s42979-021-00691-9

Li, W., Yu, X., Chen, C., & Gong, Q. (2023). Identification and localisation of grape diseased leaf images captured by UAV based on CNN. *Computers and Electronics in Agriculture, 214*, Article 108277. https://doi.org/10.1016/j.compag.2023.108277

Li, Y., & Chao, X. (2021). Semi-supervised few-shot learning approach for plant disease recognition. *Plant Methods, 17*, Article 66. https://doi.org/10.1186/s13007-021-00770-1

Mahum, R., Munir, H., Mughal, Z.-U.-N., Awais, M., Sher Khan, F., Saqlain, M., Mahamad, S., & Tlili, I. (2023). A novel framework for potato leaf disease detection using an efficient deep learning model. *Human and Ecological Risk Assessment: An International Journal, 29*(2), 303-326. https://doi.org/10.1080/10807039.2022.2064814

Nigar, N., Faisal, H. M., Umer, M., Oki, O., & Lukose, J. (2024). Improving plant disease classification with a deep-learning-based prediction model using explainable artificial intelligence. *IEEE Access, 12*, 105603-105618. https://doi.org/10.1109/ACCESS.2024.3428553

Noola, D. A., & Basavaraju, D. R. (2022). Corn leaf image classification based on machine learning techniques for accurate leaf disease detection. *International Journal of Electrical and Computer Engineering, 12*(3), 2509-2516. https://doi.org/10.11591/ijece.v12i3.pp2509-2516

Park, H.-D. (2021). An intelligent edge computing-based scalable architecture for a large-scale smart farm system. *Journal of System and Management Sciences, 11*(3), 119-139. https://doi.org/10.33168/JSMS.2021.0307

Patil, R. R., & Kumar, S. (2022). Rice-fusion: A multimodality data fusion framework for rice disease diagnosis. *IEEE Access, 10*, 5207-5222. https://doi.org/10.1109/ACCESS.2022.3140815

Prince, R. H., Mamun, A. A., Peyal, H. I., Miraz, S., Nahiduzzaman, M., Khandakar, A., & Ayari, M. A. (2024). CSXAI: A lightweight 2D CNN-SVM model for detection and classification of various crop diseases with explainable AI visualisation. *Frontiers in Plant Science, 15*, Article 1412988. https://doi.org/10.3389/fpls.2024.1412988

Rahman, K. N., Banik, S. C., Islam, R., & Al Fahim, A. (2025). A real-time monitoring system for accurate plant leaf disease detection using deep learning. *Crop Design, 4*(1), Article 100092. https://doi.org/10.1016/j.cropd.2024.100092

Rajasekar, V., Venu, K., Jena, S. R., Varthini, R. J., & Ishwarya, S. (2022). Detection of cotton plant diseases using deep transfer learning. *Journal of Mobile Multimedia, 18*(2), 307-324. https://doi.org/10.13052/jmm1550-4646.1828

Raval, H., & Chaki, J. (2024). Ensemble transfer learning meets explainable AI: A deep learning approach for leaf disease detection. *Ecological Informatics, 84*, Article 102925. https://doi.org/10.1016/j.ecoinf.2024.102925

Rezaei, M., Diepeveen, D., Laga, H., Gupta, S., Jones, M. G. K., & Sohel, F. (2025). A transformer-based few-shot learning pipeline for barley disease detection from field-collected imagery. *Computers and Electronics in Agriculture, 229*, Article 109751. https://doi.org/10.1016/j.compag.2024.109751

Rezaei, M., Diepeveen, D., Laga, H., Jones, M. G. K., & Sohel, F. (2024). Plant disease recognition in a low-data scenario using few-shot learning. *Computers and Electronics in Agriculture, 219*, Article 108812. https://doi.org/10.1016/j.compag.2024.108812

Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for the detection of tomato plant disease. *Frontiers in Plant Science, 13*, Article 1031748. https://doi.org/10.3389/fpls.2022.1031748

Shovon, M. S., Mozumder, S. J., Pal, O. K., Mridha, M. F., Asai, N., & Shin, J. (2023). PlantDet: A robust multi-model ensemble method based on deep learning for plant disease detection. *IEEE Access, 11*, 348460-34859. https://doi.org/10.1109/ACCESS.2023.3264835

Singh, R. S. R., & Sanodiya, R. K. (2023). Zero-shot transfer learning framework for plant leaf disease classification. *IEEE Access, 11*, 143861-143880. https://doi.org/10.1109/ACCESS.2023.3343759

Srivastava, S., Kumar, P., Mohd, N., Singh, A., & Gill, F. S. (2020). A novel deep learning framework approach for sugarcane disease detection. *SN Computer Science, 1*(1), Article 7. https://doi.org/10.1007/s42979-020-0094-9

Storey, G., Meng, Q., & Li, B. (2022). Leaf disease segmentation and detection in apple orchards for precise smart spraying in sustainable agriculture. *Sustainability, 14*(3), Article 1458. https://doi.org/10.3390/su14031458

Sun, H., Xu, H., Liu, B., He, D., He, J., Zhang, H., & Geng, N. (2021). Mean-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks. *Computers and Electronics in Agriculture, 189*, Article 106379. https://doi.org/10.1016/j.compag.2021.106379

Sun, R., Zhang, M., Yang, K., & Liu, J. (2020). Data enhancement for plant disease classification using generated lesions. *Applied Sciences, 10*(2), Article 466. https://doi.org/10.3390/app10020466

Thakur, P. S., Khanna, P., Sheorey, T., & Ojha, A. (2022). Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT. *arXiv*. https://doi.org/10.48550/arXiv.2207.07919

Vishnoi, V. K., Kumar, K., Kumar, B., Mohan, S., & Khan, A. A. (2023). Detection of apple plant diseases using leaf images through a convolutional neural network. *IEEE Access, 11*, 6594-6609. https://doi.org/10.1109/ACCESS.2022.3232917

Xiang, S., Liang, Q., Sun, W., Zhang, D., & Wang, Y. (2021). L-CSMS: Novel lightweight network for plant disease severity recognition. *Journal of Plant Diseases and Protection, 128*(2), 557-569. https://doi.org/10.1007/s41348-020-00423-w

Xiong, Y., Liang, L., Wang, L., She, J., & Wu, M. (2020). Identification of cash crop diseases using an automatic image segmentation algorithm and deep learning with an expanded dataset. *Computers and Electronics in Agriculture, 177*, Article 105712. https://doi.org/10.1016/j.compag.2020.105712

Xu, M., Yoon, S., Jeong, Y., & Park, D. S. (2022). Transfer learning for versatile plant disease recognition with limited data. *Frontiers in Plant Science, 13*, Article 1010981. https://doi.org/10.3389/fpls.2022.1010981

Yilmaz, E., Bocekci, S. C., Safak, C., & Yildiz, K. (2025). Advancements in smart agriculture: A systematic literature review on state-of-the-art plant disease detection with computer vision. *IET Computer Vision, 19*(1), Article e70004. https://doi.org/10.1049/cvi2.70004

Yong, L. Z., Khairunniza-Bejo, S., Jahari, M., & Muharam, F. M. (2023). Automatic disease detection of basal stem rot using deep learning and hyperspectral imaging. *Agriculture, 13*(1), Article 69. https://doi.org/10.3390/agriculture13010069

Zhang, X., Vinatzer, B. A., & Li, S. (2024). Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease. *Scientific Reports, 14*(1), Article 27666. https://doi.org/10.1038/s41598-024-78650-6

Zhang, G., Xu, T., & Tian, Y. (2022). Hyperspectral imaging-based classification of rice leaf blast severity over multiple growth stages. *Plant Methods, 18*(1), Article 123. https://doi.org/10.1186/s13007-022-00955-2

Zhang, G., Xu, T., Tian, Y., Xu, H., Song, J., & Lan, Y. (2020). Assessment of rice leaf blast severity using hyperspectral imaging during late vegetative growth. *Australasian Plant Pathology, 49*, 571-578. https://doi.org/10.1007/s13313-020-00736-2

Zhang, L., Xia, C., Xiao, D., Weckler, P., Lan, Y., & Lee, J. M. (2021). A coarse-to-fine leaf detection approach based on leaf skeleton identification and joint segmentation. *Biosystems Engineering, 206*, 94-108. https://doi.org/10.1016/j.biosystemseng.2021.03.017

Zhou, H., Li, W., Li, P., Xu, Y., Zhang, L., Zhou, X., Zhao, Z., Li, E., & Lv, C. (2025). A novel few-shot learning framework based on diffusion models for high-accuracy sunflower disease detection and classification. *Plants, 14*(3), Article 339. https://doi.org/10.3390/plants14030339

Zinonos, Z., Gkelios, S., Khalifeh, A. F., Hadjimitsis, D. G., Boutalis, Y. S., & Chatzichristofis, S. A. (2022). Grape leaf diseases identification system using convolutional neural networks and LoRa technology. *IEEE Access, 10*, 122-133. https://doi.org/10.1109/ACCESS.2021.3138050

Pertanika J. Sci. & Technol. 34 (4): 1981 - 2002 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Smart Comment Manager: Streamlining Comment Analysis and Viewer Query Clustering Using NLP Processing

N Venkata Sailaja, Karnam Akhil*, Sai Shashidhar Reddy Katanguri, Maringanti Gopi Krishnna, Alladi Shiva Vinay, and Devireddy Siddartha Reddy

Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Vignana Jyothi Nagar, Bachupally, 500118 Hyderabad, Telangana, India

ABSTRACT

In Today's world, YouTube and other content creators are frequently barraged by a deluge of spam, unrelated remarks, and even hate messages, which makes it challenging for them to have a healthy and active online community. The more their audience, the more time-consuming are the traditional moderation practices and the less insight the audience opinion provides. This disconnection renders the creators incapable of knowing and redressing the problems of their community. Adding to the problem are the fuzzy logic limitations in identifying subtle topics and emotive cues, where misinterpretation can diminish the strength of content analysis. To fill these gaps, the NLP-driven Smart Comment Manager is set to simplify and enhance comment analysis. The platform is made user-friendly and removes hate speech, spam, and clusters duplicate questions, allowing creators to respond better and build a better and more engaging community. This system is built from three foundation modules: Hate Speech and Spam Detection, with RoBERTA to detect toxic content quickly and efficiently; Similar Question Clustering, with stsb-distilbert-base, to cluster repetitive questions for best responses; and Topic Discovery, with BERTopic, to detect trending topics in both comment streams and live chat. Smart Comment Manager converts comment streams into active feedback engines for building audience engagement and community, not passive feedback pipelines. Beyond content creation, the platform is also beneficial to customer support teams and online communities where handling massive amounts of user feedback is critical.

Keywords: Comment analysis, content moderation, duplicate question detection, hate speech detection, Natural Language Processing (NLP), semantic clustering, spam detection, topic modelling

ARTICLE INFO

Article history:
Received: 07 June 2025
Accepted: 02 June 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.02

E-mail addresses:
sailaja_nv@vnrvjiet.in (N Venkata Sailaja)
akhilresearch18@gmail.com (Karnam Akhil)
shashidhar7682@gmail.com (Sai Shashidhar Reddy Katanguri)
gopeekrishna36@gmail.com (Maringanti Gopi Krishnna)
shivavinayalladi@gmail.com (Alladi Shiva Vinay)
siddarthareddy.d02@gmail.com (Devireddy Siddartha Reddy)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

In today's online world, sites like YouTube are the hub of content creation, community, and audience interaction. With millions of comments made daily, these sites are breeding grounds for feedback, questions, and discussion. But it is a large task to handle this stream of user-generated content for a moderator.

The most surprising thing in such online forums is the presence of spam. Conventional spam filters, too often based on rigid keyword lists or blacklisting, can't hope to match spammers' constantly evolving tactics. Spam today is crafted specifically to evade such filters by using such deceptions as obfuscation, synonyms, or contextual mimicry. So, there is a need for a more sophisticated solution, one based on natural language understanding rather than shallow keyword scanning.

One of the main concerns in comment moderation is the spread of hate speech on the internet. From abusive comments to toxic content is becoming more sophisticated and difficult to identify. Today's rule-based moderation tools tend to miss the mark, either censoring innocent conversations or allowing toxic posts to pass through. But Sophisticated NLP models such as RoBERTa have demonstrated a high degree of contextual understanding. These models can make informed judgments regarding whether content is safe or not.

Another growing need in comment moderation is determining when users are asking questions. Audience engagement is a measure of content success, and creators rely on questions to gauge interest and offer more relevant content. These questions, however, become buried among general comments, and many are not in the traditional question form. Transformer models like BERT or GPT-classifiers understand sentence intent. and can identify questions even when they are informally phrased, so no meaningful interaction is lost.

Aside from identifying questions, there is also the problem of redundancy. Dozens of users will pose the same question but in slightly different wording. NLP models such as STSB-DISTIL-BERT-base are engineered specifically to overcome this. Through the conversion of sentences into vector embeddings, the system can group semantically similar questions - even ones that are worded differently.

Knowing what's going on in the larger conversation is just as valuable as responding to individual comments. Old-school topic modelling techniques such as LDA are flawed; they create unclear or confounding topics that don't reflect audience sentiment. Models such as BERTopic that use neural networks provide a closer answer by employing contextual embeddings and dynamic clustering to categorise comments by high-priority topics, enabling easy insights for more informed content decisions.

Merging all these capabilities is the Smart Comment Manager, an NLP-powered system that has the potential to transform comment management at scale. It integrates best-in-class models like RoBERTa for hate speech detection, stsb-distilbert-base for semantic clustering,

and BERTopic for trend detection. The system moderates' comments in real-time, filtering offending content automatically, clustering duplicate questions, and identifying trending topics. This solution brings order, insight, and productivity for creators.

The Smart Comment Manager enables creators and organisations to build better, safer communities. It's not merely a technical breakthrough, but a natural development of how we have conversations at scale with a correct balance between automation and comprehension of human communication.

The structure of this paper is the following: Section 2 presents an overview of related work and highlights the weaknesses of current techniques. Section 3 describes the methodology, i.e., design and implementation of the Smart Comment Manager. Section 4 presents experimental results and evaluations, and Section 5 concludes with a discussion of the main findings and future work.

Main Contributions

1. A unified transformer-based framework for real-time comment moderation and semantic audience interaction.
2. Integration of contextual spam filtering, semantic duplicate clustering, and topic discovery into a single pipeline.
3. Lightweight deployment-oriented architecture using DistilBERT and DistilRoBERTa for scalable moderation.
4. Comparative evaluation against classical ML baselines, including SVM, Logistic Regression, MNB, and ProdLDA.
5. Real-time topic summarisation for large-scale YouTube comment streams.

RELATED WORK

Ghanem and Erbay (2023) were focusing on enhancing spam detection accuracy on social media platforms from deep contextualised word representations. Their study utilised advanced word embeddings and Deep learning models, with a focus on BERT and its variants. Their results suggested BERT variants perform better than earlier state-of-the-art methods, especially in unstructured social media status message context comprehension. Nonetheless, they were computationally intensive.

Sharma et al. (2018) tried automatic classification and tagging of abusive social media content to prevent the propagation of hate messages. This research used feature extraction methods like k-skip-n-grams with Support Vector Machines (SVM), Random Forest, and Naive Bayes. Results indicated that this ensemble approach produced better accuracy but had lower accuracy on large datasets and was overwhelmed by the application of feature engineering, which did not consider the original intent or sarcasm. Building on such efforts, Song et al. (2024) extended offensive language detection to open chatbot platforms, reducing harmful interactions and improving moderation effectiveness.

Wei et al. (2021) explored transfer learning and deep learning methods in hate speech and offensive language detection on social media. They trained models using a fine-tuning approach for better accuracy on current hate speech corpora. The outcomes indicated that transfer learning, especially with DistilBERT and GPT-2, greatly improved performance. The research, although biased towards contextual meaning, still performs poorly in real-world scenarios.

Badjatiya et al. (2017) proposed a deep learning approach to hate speech detection from semantic embeddings and classification frameworks. They enhanced detection accuracy using model selection and feature engineering with pretrained word embeddings and experimented with LSTMs, GRUs, and CNNs for classification, along with Hyperparameter tuning. The results showed the strength of deep learning in extracting contextual meaning.

Zavrak and Yilmaz (2022) proposed a deep learning-based spam filtering model that is a combination of CNNs, GRUs, and attention mechanisms. The proposed model demonstrated strong capability to capture textual richness with attention layers and improved spam detection. The results showed that their hybrid approach and hierarchical attention mechanisms were more interpretable and contextually aware.

In natural language processing, question/statement classification of text is at the core of many applications. The authors Wang and Jia (2023) introduced a question type classification model with a self-attention mechanism added to a BiLSTM-AlexNet model to better capture contextual perception and improve accuracy. Their model outperformed conventional approaches in performance.

Wasim et al. (2019) addressed the problem of multi-label classification for biomedical QA systems, i.e., factoid vs. list-type question identification. Their system performed better than classical classifiers in precision and recall. Deep learning-based feature extraction assisted in enhancing classification accuracy.

Pratiwi and Syukriyah (2019) investigated how machine learning methodologies can be used to classify questions in e-learning systems. In their comparative study, they found that deep learning-based classification models outperformed algorithms like decision trees and SVMs in all environments. While the model proposed improved accuracy in question retrieval and classification, it heavily depended on the quality of the data and consumed vast computational power.

Sahu et al. (2019) tackled the issue of automatic user-posted Q&A question tagging using multi-label classification methods. The system exhibited notable improvements in tagging accuracy. Nevertheless, struggled with ambiguous or overlapping topics, and dependency on prior tagging history compromised flexibility to new topics.

Tahtah et al. (2023) created a machine learning-based question classification system that was used in the Moroccan legal context to aid in the more precise retrieval of legal information. Transformer models outperformed others in the classification process. However, Intrinsic legal language complexity and limited data became hindrances in scaling the system to more environments.

Deep Learning-Based Methods

A Critical objective of our system is identifying and collating redundant user questions, which not only improves interaction quality but also assists content creators in better answering questions.

Initial solutions to question similarity problems were based on statistical natural language processing techniques. Shah et al. (2017) proposed a deep learning model that combined LSTM networks with convolutional layers to identify redundant question pairs with boosted accuracy. They employed sentence embeddings and feature-based similarity with much superior performance.

According to the previous approach, Gupta and Joshi (2017) proposed a hybrid architecture, integrating Convolutional RNN's and CNNs to extract both syntactic and semantic properties of question similarity. And they introduced a multi-layer attention mechanism to have more precise sentence embeddings.

Likewise, Mueller and Schütze (2016) have attempted to use Siamese LSTM networks on question duplicate detection, specifically an equal-weight model that can be trained to detect more precise similarity representations with a great performance boost in paraphrased query detection and is therefore very well-suited for scaling. Building on this line of research, Korade et al. (2024) have explored new ways of NLP techniques for duplicate Question detection for maximising responses on Q&A websites.

Sentence Embedding and Transformer-Based Models for YouTube Comment Clustering

In today's world, with the growing popularity of Deep learning, Wang et al. (2017) built upon earlier deep learning techniques by employing deep sentence embeddings in combination with attention and fine-tuning on the Quora duplicate question dataset (Quora, 2017), with a focus on both similarity techniques and deeper contextualization.

Additional enhancement was suggested by Reimers and Gurevych (2019), who introduced Sentence-BERT (SBERT), specifically fine-tuned for sentence similarity tasks. It uses Siamese and triplet network architectures to generate contextualised sentence embeddings. This makes it scalable, and because of its performance, it is a go-to benchmark.

Gerlach et al. (2017) introduced a bimodal network-based model to model topic evolution over long timescales. The model was used for years of economic texts and applied to monitor the evolution of economic concepts through the analysis of the co-occurrence of topics in documents. It captured structural changes in themes over time more effectively.

Reisenbichler and Reutterer (2019) explored the potential of topic modelling to be used in consumer behaviour research and market segmentation. They showed how meaning could be uniquely extracted. Using models like Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorisation (NMF), their work showed that high topic coherence relies on good preprocessing.

A systematic review of literature by Kherwa and Bansal (2019) analysed about 300 topic modelling studies, comparing probabilistic models like LDA with recent neural networks. The authors put a high emphasis on coherence measures to quantify results and interpretability of resulting topic structures.

Dieng et al. (2019) solved this with the Embedded Topic Model (ETM), a combination of word embeddings and traditional topic modelling to increase coherence and predictiveness. They achieved better semantic similarity on a wide range of datasets depending on pre-trained embeddings. Recent advances in Topic Modelling are carried out by Kapoor et al. (2024) by introducing QualIT, an LLM -enhanced Topic Modelling framework that provides qualitative insights.

Qiang et al. (2019) focused on Short-text topic modelling on data sources like tweets, chat logs, and product reviews. This survey addressed models like Dirichlet Multinomial Mixture (DMM), Bi-term Topic Model (BTM) on sparse textual data, which addressed the loss of topic coherence. More recently, Pham et al. (2024) came up with Topic-GPT, which leverages the power of large language models for topic discovery by using prompts, which offers great adaptability.

Grounding our work on this research basis, our research seeks to further enhance topic modelling in coherence, interpretability, and scalability. Our solution involves a blend of transformer-based models, embedding-augmented models and domain-specific optimisation techniques for short-text input, that are in coordination with the Quora question pairs dataset (Quora, 2017). With the use of deep learning methods, our solution tackles the issues of real-time deployment and domain-specific modelling.

METHODOLOGY

System Architecture

The system presented is an NLP system that is intended to moderate and analyse YouTube comments based on both machine learning and deep learning. The system is organised around four basic modules: a Comment Filtering Module, a Question Detection Module, Duplicate Question Clustering, and a Topic Modelling Module.

The process adheres to a systematic pipeline that assists in sorting and interpreting YouTube comments in an efficient manner. It all begins with real-time comment gathering, either on a regular video or live stream, with the help of the YouTube API, which enables the system to remain synchronised with the live aspect of user interaction.

Once data collection is completed, the comments are pre-processed. This phase is dedicated to text data cleaning. After the text is cleaned, tokenisation is achieved with the DistilBERT-base-uncased tokeniser. Tokenisation is a critical step in enabling more efficient processing.

Table 1
Splits of data based on train, validation and test

Module	Train	Validation	Test
Hate Speech	70	10	20
Question Detection	70	10	20
Clustering	70	10	20

Following tokenisation, the system proceeds to detect hate speech and spam for the data specified as shown in Table 1. The task of classifying is achieved by employing RoBERTa and DistilRoBERTa, which classify all comments and tag them as Normal, Hate Speech, Spam, or their combination. And they were automatically removed if they were Hate/Spam Speech.

The filtered comments are then processed by the Question Detection Module. Here, the DistilBERT-base-uncased model is used to determine whether a comment is a question or not. If a comment is determined to be a question, then it is passed on to the Duplicate Question Clustering Module for processing. And the remaining questions were sent to the Topic Modelling phase.

To efficiently process redundant questions, the system employs the STSB-BERT model, which was trained on the Quora Duplicate Question Pairs dataset (Quora, 2017). By clustering redundant & semantically similar questions in a single cluster, the system prevents repetitive responses and provides concise and complete responses.

Simultaneously, non-question comments are sent to the Topic Modelling Module. This module uses BERTopic, a powerful topic modelling platform that can pull out new topics dynamically. The result is a real-time summary of the most important topics being discussed by viewers. These can help viewers to remain attuned to the most important conversational topics.

At its last phase, the system generates three main outputs. The first is the Trending Topics. Second is the Consolidated Questions feature that groups. Moderation System, the third feature, is one that guarantees the integrity of the community is maintained and builds a more open and interactive space by eliminating hate speech as well as spam.

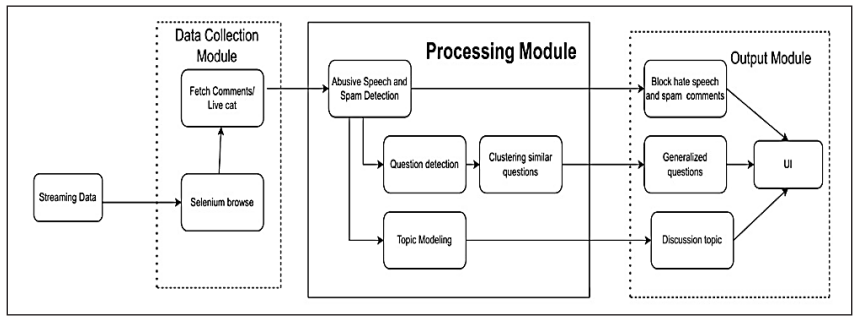


Figure 1. Generic framework for comment manager

As shown in Figure 1, the architecture of our proposed system. In general, this well-designed process keeps the pages of YouTube comments in order, informative, and convenient. It allows creators to handle audience interaction in a manner that is effective, as well as allowing for full, pertinent discussion.

ALGORITHMS

Comment Filtering Module

The Comment Filtering Module has an important function in ensuring the quality of the conversation space is kept in check by properly recognising and removing the hate and spam speeches. Since the module classifies comments by their nature, the module aids in keeping all users within a respectful, open, and interactive environment.

Models

The architecture employs DistilRoBERTa, which is a lighter yet effective variant of RoBERTa that has been optimised on hate speech datasets. Tuned for efficiency and speed, DistilRoBERTa also has good accuracy for classification, capable of detecting and eliminating harmful or toxic material to allow quality and positive user interactions.

Spam is detected by RoBERTa-base, which is a highly capable language model that has been trained on extensive text data and has been fine-tuned. It performs binary classification, labelling all comments as either Spam or Non-Spam.

Algorithm

Input. The system takes an input of comments along with pre-trained models Distil-RoBERTa (hate speech detection) and RoBERTa-base (spam detection).

Output. It generated a classification label for each comment as Normal, Hate Speech, Spam, or both.

Step 1: Load Dataset. The tweets_hate_speech_detection dataset is loaded first and split into two sets (80:20).

Step 2: Preprocessing. The tweets are tokenised with the corresponding tokeniser of each task: DistilRoBERTa for hate speech and RoBERTa for spam. Padding and truncation are applied to clean the inputs.

Step 3: Fine-tuning. Train the DistilRoBERTa model for classifying comments as hate speech or not, alongside train RoBERTa-base model to classify comments into spam or not.

Step 4: Evaluation. Compare the two models based on parameters like mean Average Precision (mAP) to evaluate their performance.

Step 5: Model Saving. Once training is complete, save the fine-tuned models and their tokenisers for later usage.

Step 6: Prediction. For each comment, run predictions through both models:

- The hate speech model returns 1 for Hate Speech or 0 if not.
- The spam model returns 1 for Spam or 0 if not.

Step 7: Final Classification.

- If both outputs are 0, then classify it as Normal Comment.
- If only the hate speech model returns 1, then classify it as Hate Speech.
- If only the spam model returns 1, then classify it as Spam.
- If both return 1, then classify as Hate Speech and Spam.

Step 8: Output. The system then returns this final classification label for the input comment.

Question Detection Module

This module is important in terms of organising discussion since it determines whether a comment from a user is indeed a question or not.

Model

The system is based on the DistilBERT-base model, a smaller and faster version of BERT, but still strong enough to provide good accuracy with rapid response times. The model is trained to distinguish actual questions and general commentary. Questions are then sent to the Duplicate Question Clustering Module, and in parallel, all comments are sent to the Topic Modelling module.

Algorithm

Input. The system takes an input of comments along with a pre-trained DistilBERT model.

Output. The model predicts and labels each comment whether it is a "Question" or "Statement".

Step 1: Dataset Loading.

- Import a labelled dataset SQuAD, which includes both questions and non-questions.
- Split the dataset into 80% - 20% ratio for training and testing.

Step 2: Preprocessing.

- A tokeniser like distil-bert-base-uncased is used for tokenising the comments.
- Truncate the inputs to ensure all are of uniform length.

Step 3: Fine-tuning. Train the DistilBERT-base model to perform binary classification—classify questions from non-questions.

Step 4: Evaluation. Test the performance of the model using standard measures like accuracy, precision, recall, and F1-score.

Step 5: Model Saving. Once training is complete, save the fine-tuned model and its tokeniser.

Step 6: Prediction.

- Tokenise any fresh comment and run it through the trained model.
- Get the output logits and use them through a softmax layer to produce probability scores.
- If the question likelihood is greater than 0.6, then it is a "Question", otherwise a "Non-Question".

Step 7: Output. The system then returns a classification whether it's a question or not.

Step 8: Routing. If a comment is labelled as a question, it's passed to the Duplicate Question Clustering Module for additional processing and in parallel, all the inputs are forwarded to the topic modelling phase.

Duplicate Question Clustering

This module is meant to group similar questions into a single category, removing redundancy. This is not only time-saving but also groups the interaction and makes it more efficient for users.

Model

The system uses STSB-BERT, a benchmark for semantic textual similarity. Unlike basic keyword matching, this model captures the underlying semantics of sentences by generating

rich, high-dimensional embeddings. It computes the similarity between two questions using cosine similarity. And if the score is greater than a certain threshold, the questions are labelled as duplicates and clustered together.

Algorithm

Input. The system begins with a list of user-submitted questions.

Output. The result is a trimmed-down collection of unique questions, where similar or repeated ones are grouped.

Step 1: Dataset Loading.

- To train the model, we used the dataset Quora Duplicate Question Pairs (Quora, 2017).
- With each question pair labelled as either a duplicate (1) or non-duplicate (0). And it is split (70:10:20) ratio for training and testing purposes.

Step 2: Data Preprocessing. All questions are tokenised to prepare them for model input.

Step 3: Model Fine-tuning. Train the STSB-BERT model to determine questions with the same intent on the Quora Dataset (Quora, 2017).

Step 4: Model Evaluation. Once training is complete, the model's accuracy is measured using metrics like mean average precision (mAP).

Step 5: Model Saving. After training, the model and its tokeniser are saved for later use.

Step 6: Duplicate Question Detection.

- Sentence embeddings are generated for all of them with the stsb-bert-base model.
- Pairwise cosine similarity values are then computed to determine how much each question is like the others.
- These are then translated into a binary similarity matrix; if two questions have a score greater than some threshold (0.75), they're labelled as duplicates and grouped. Finally, one representative question is selected in each group.

Step 7: Output. The system returns a set of separate questions.

- By effectively clustering similar questions, this module enhances user interaction, eliminates redundant answers, and maintains discussions brief and efficient.

Topic Modelling Module

This module is intended to reveal and forecast popular topics from YouTube live chats. By examining the stream of incoming comments, it determines the breaking topics in the discussion, allowing creators, moderators, and viewers to remain in sync without having to scroll through hundreds of chat messages.

Model

The system uses BERTopic, a powerful pre-trained topic modelling technique as shown in Table 2. It works by grouping similar comments into clean and understandable topics using TF-IDF and transformer-based embeddings. In word matching at the surface

Table 2
Ablation for the proposed methodology

Configuration	Accuracy
Only Spam Detection	88
Spam + Hate	90
Spam + Hate + Clustering	91
Proposed System	92

level, BERTopic identifies the underlying context and patterns in the co-occurrence of words and is therefore highly effective in dynamic environments like live chats as illustrated in Table 5. This intelligent clustering of semantically related comments provides a concise summary of the most popular topics discussed in the conversation.To ensure the fairness, Table 3 illustrates the settings of hyperparameters in detail.

Table 3
Hyperparameter settings

Model	Hyperparameters
SVM	RBF kernel, C=10, gamma=scale
Logistic Regression	C=1.0, max_iter=1000
MNB	alpha=0.5
MLP	Hidden layers = (128,64), Adam optimiser
DistilRoBERTa	lr=2e-5, batch=32, epochs=3
RoBERTa	lr=2e-5, batch=32, epochs=3
BERT-base	lr=2e-5, batch=16, epochs=4
STSB-BERT	lr=2e-5, batch=16, epochs=3
BERTopic	STSB-BERT embeddings, UMAP (15 neighbours, 5 components), HDBSCAN (cluster size=10), top 10 words/topic

Table 4
Cross-validation table

Model	Accuracy (%)	Std Dev
DistilRoBERTa	91.0	±0.4
SVM	89.1	±0.8
RoBERTa	92.2	±0.3
MNB	86.8	±1.2
MLP	88.6	±0.7
BERT-base	90.5	±0.5
STSB-BERT	94.3	±0.2
Logistic Regression	87.3	±0.9

Table 5
BERTopic stability analysis

Fold	Topic Coherence	Topic Diversity
1	0.81	0.84
2	0.83	0.86
3	0.82	0.87
4	0.84	0.85
5	0.84	0.86
Mean	0.824	0.856

Algorithm

Input. The input to this system is a list of hate and spam-filtered comments.

Output. It produces a clean, summarised list highlighting the main topics people are actively discussing.

Step 1: Loading the Pre-Trained Model. A pre-trained BERTopic model is used, which is trained on BERTopic_Wikipedia.

Step 2: Extracting Topic Information. The system gathers topic metadata using the get_topic_info method.

Step 3: Processing Input Data. Live chat comments are collected and pre-processed to prepare them for analysis.

Step 4: Applying BERTopic for Topic Modelling.

- The filtered comments are passed to the BERTopic model.
- It identifies and extracts the most relevant and frequently discussed topics.
- From there, the top two trending topics are selected based on frequency and relevance.

Step 5: Generating a Summarised Output.

- A LLM model like Gemini-Pro is used to polish the raw topics into smooth, well-phrased descriptions.
- For example, the output might say: "People in the chat are discussing Artificial Intelligence and Machine Learning, as well as Climate Change and Global Warming."

Step 6: Output. The final list of trending topics is displayed in real-time, allowing for quick understanding of the ongoing conversation.

By automating topic modelling, this module helps moderators and content creators easily follow the flow of discussion without needing to manually scroll through hundreds of live comments. Table 4 shows various model analysis on the data.

RESULTS AND DISCUSSIONS

Quora Question Pairs Dataset

Quora Question Pairs (QQP) dataset is an open-source benchmark dataset created by Quora and released in 2017 for detecting semantic similarity and duplicate questions. The dataset includes over 400,000 question pairs asked on Quora, labelled with a binary label whether the question pair is semantically similar (duplicate) or not.

Each data point contains:

- qid1, qid2: Unique identifiers for each question.
- question1, question2: What are the two questions about?
- is_duplicate: Binary indicator (1 if questions are duplicates, else 0).

In this research, we use the QQP dataset to fine-tune a sentence embedding model (STSB-BERT) for semantically similar or the same question detection from YouTube comments. Our trained model measures cosine similarity among detected questions for classifying them into sets in terms of repeated questions.

In Figure 2, the performance of DistilRoBERTa and SVM (gamma) on TF-IDF is compared for varying sample sizes for the comment filtering. The performance of both models improves with increasing sample size. DistilRoBERTa is marginally better at the beginning and consistent with all datasets, but SVM on TF-IDF takes over at larger sample sizes, more dependent on the type of task.

Figure 3 compares the accuracy of RoBERTa, Multinomial Naïve Bayes (MNB), and Multi-Layer Perceptron (MLP) when used to filter comments. RoBERTa performs best, always increasing as the samples are increased to around 92% accuracy. This is a sign that RoBERTa performs better to process comment trends.

Figure 4 shows the comparison of BERT-base uncased and Logistic Regression with TF-IDF in terms of comment analysis accuracy. BERT is always higher, and accuracy approximates nearly 91.5% as the sample size is larger. And BERT can interpret the meaning and context of the comments better, while Logistic Regression has difficulty with intricate patterns but gets better with more data.

Figure 5 interprets the performance of Logistic Regression

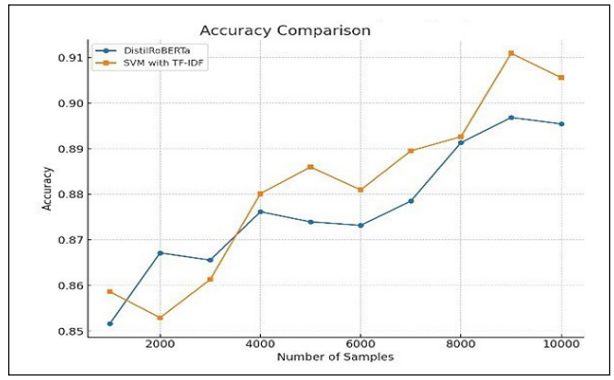


Figure 2. Accuracy Comparison of DistilRoBERTa with SVM TF-IDF

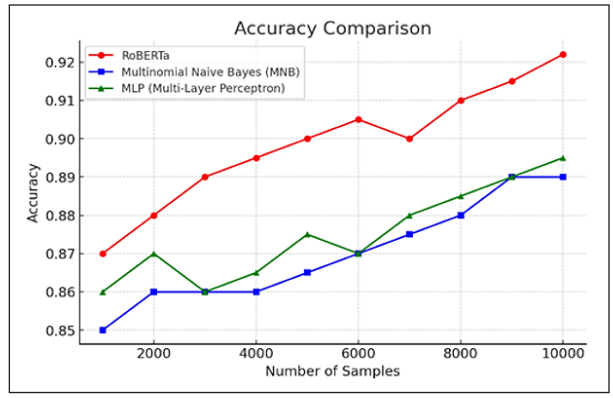


Figure 3. Accuracy Comparison between RoBERTa, MNB and MLP models

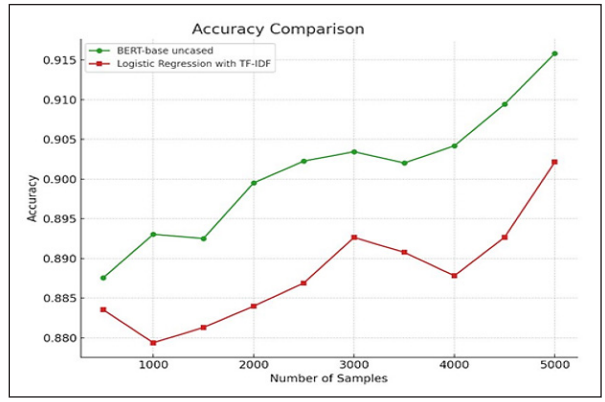


Figure 4. Accuracy comparison between BERT-base uncased and logistic regression with TF-IDF

with TF-IDF and compares it to STSB-BERT-Distil in question clustering. STSB-BERT-Distil is far ahead with a performance of approximately 94%, while Logistic Regression lags with approximately 87%.

The Intertopic Distance Map, as shown in Figure 6, displays the proximity of different topics in comments. Every bubble displays a topic, and its size displays the number of comments for that topic. This assists in comprehending the fundamental discussions in YouTube comments, clustering equivalent ones together and separating different topics from one another.

The bar chart, as shown in Figure 7, displays the most discussed subjects in YouTube comments. "Galactus" is the most discussed, followed by "Reed" and "Wonder". These subjects represent what individuals are most interested in, whether specific characters, concepts, or just general conversation.

Figure 8 shows the most relevant words in different subjects from YouTube comments. Each subject is a group of similar conversations, such as superheroes (Marvel, Superman), movie trailers, and fan reactions (love, videos).

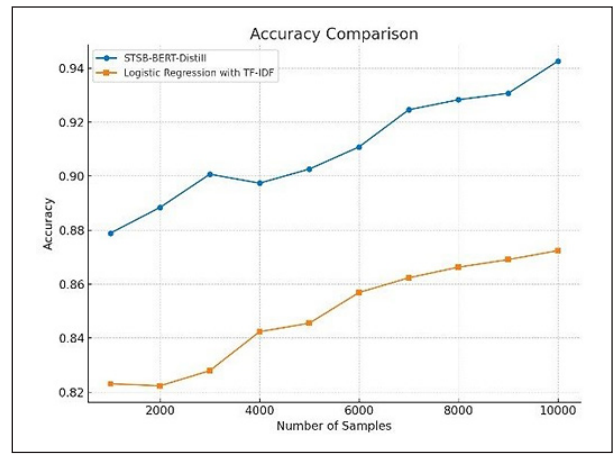


Figure 5. Accuracy comparison between STSB-BERT-distil and logistic regression with TF-IDF

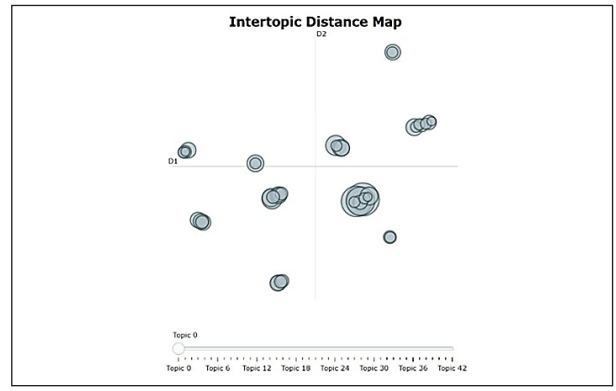


Figure 6. Intertopic distance map showing clusters of comments with similar topics

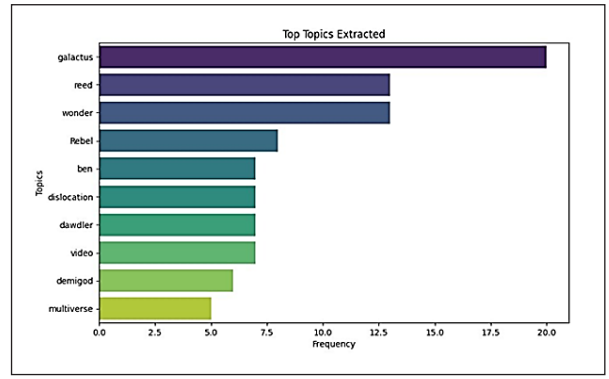


Figure 7. Bar chart displaying the most discussed topics

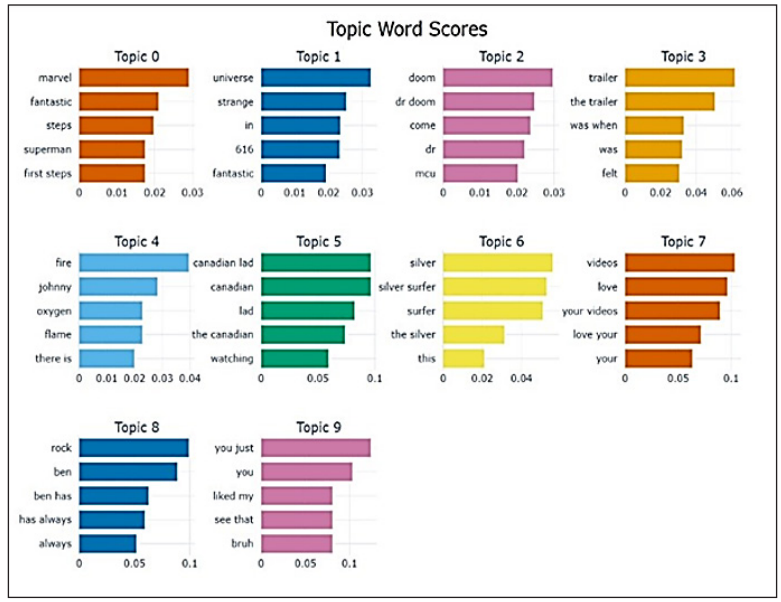


Figure 8. Most relevant words in different clusters of comments

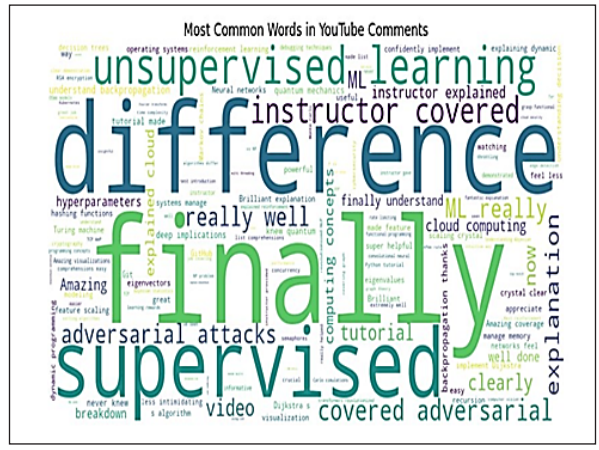


Figure 9. Word cloud of the most common words in comments

The word cloud, as shown in Figure 9, indicates the most common words used in YouTube comments, which are points of discussion. The words "difference," "supervised," "unsupervised," and "learning" are an indication that the users are interacting with learning content, especially with machine learning concepts. "Finally," and "thanks" are signs of viewers enjoying explanations.

The Perplexity score comparison is shown in Figure 10, which assists in measuring how well a model clusters YouTube comments into relevant topics. A lower perplexity

score means the model is better at detecting prevailing topics, and thus, similar comments are clustered better. BERTopic is at lower perplexity at each point in the plot compared to ProdLDA, so it is better at detecting relevant topics in real-time comment streams.

Figure 11 shows BERTopic and ProdLDA topic diversity scores against each other when they extract topics from YouTube comments. The greater the diversity score, the more types of discussion topics the model extracts, so more perspectives and subtopics are accounted for. BERTopic will have a greater diversity than ProdLDA, meaning more trending topics are extracted from real-time chat and user talks.

The Dashboard of Smart Comment Manager offers an interactive platform that acts as the main tool for analysing YouTube comments and measuring their engagement. According to Figure 12, after launching the application, the dashboard will be seen on the screen as the main point of entry, providing access to various features such as comment analysis, detecting spam and hate speech, clustering of questions, and topic modelling. The dashboard is segmented into different modules that facilitate comment analysis, identify popular topics, and cluster viewer questions. When the application initialises, these modules will be empty until a YouTube video link is provided.

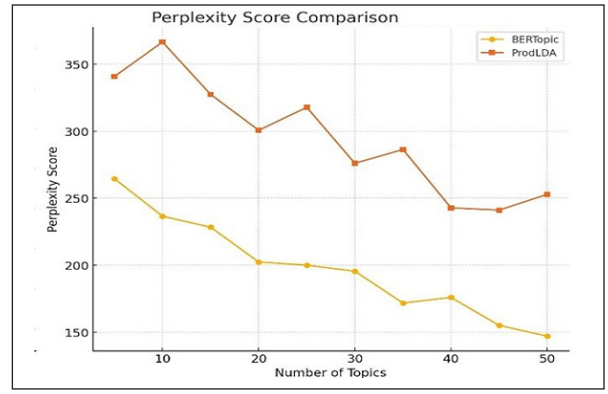


Figure 10. Perplexity score comparison between BERTopic and ProdLDA

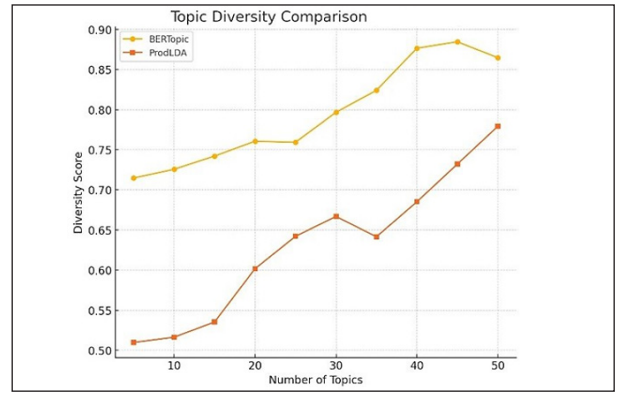


Figure 11. Topic diversity score comparison b/w BERTopic and ProdLDA

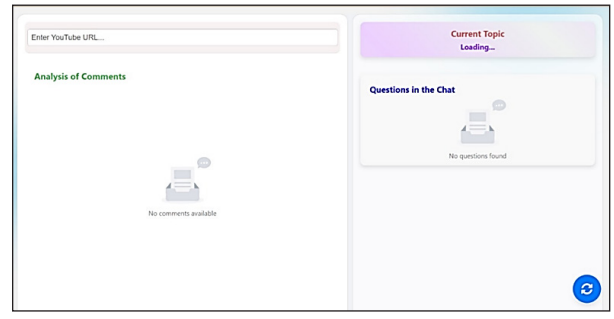


Figure 12. Initial interface when the application is launched

For an improved user experience, the Smart Comment Manager employs extensive validation procedures for the input data. Before the comments are obtained and analysed, the program validates whether the entered YouTube URL is well-formatted and supported by the system. The purpose of this process is to avoid excessive processing and the possibility that the user might enter an unsupported YouTube video URL. In case the URL is not supported, invalid, or malformed, as shown in Figure 13, the application informs the user about that using a validation message.

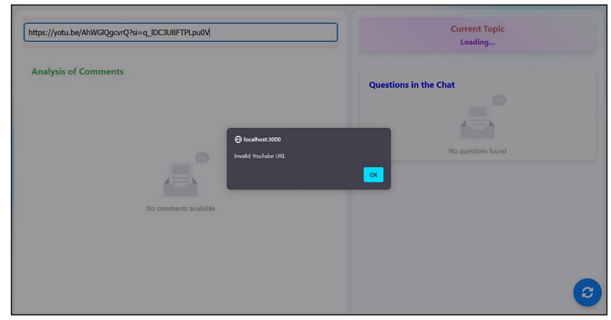


Figure 13. The system alerts the user when an invalid URL is provided

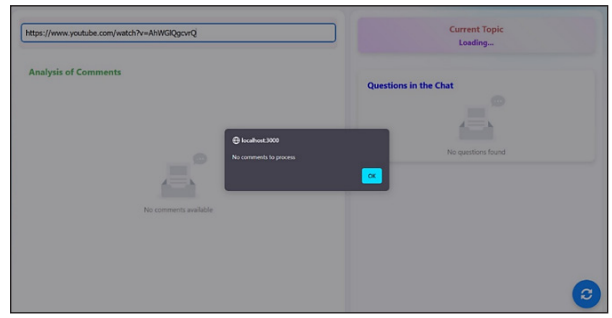


Figure 14. Application prompts when no comments are found in the processed URL

The other validation procedure involves checking whether the comments for a particular video are available or not. As shown in Figure 14, if the user enters a video that does not have any enabled comments, the program will display the validation message, thus preventing any further processes from being executed.

Upon successful validation of a YouTube video URL that has been entered into the system, shown in Figure 15, the system then analyses the comments and performs a series of actions. The dashboard provides the comments categorised, questions posted by viewers, and discussion topics raised, shown in Figures 16 and 17. Additionally, it also alerts users to comments that have anything to do with hate and spam. A side menu is provided that enables browsing of other comments for further analysis.

For the implementation of our hate speech detector model, we used Smart Comment Manager that uses the hate speech detection model based on the DistilRoBERTa transformer model. To get good classification results, it was necessary to find out the optimal model parameters. Rather than going for an automated search for hyperparameters, it was decided to go with a manual search.

The model was set up using the transformer architecture named DistilRoBERTa Base. The training was conducted on five epochs, and the batch size used was sixteen.

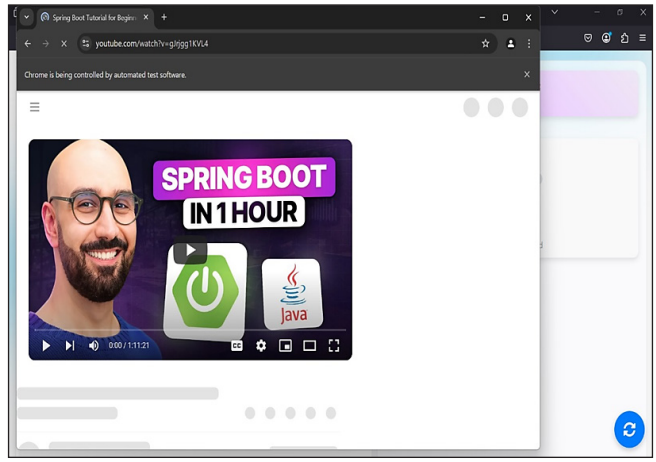


Figure 15. Comments scraping starts when a valid URL is provided

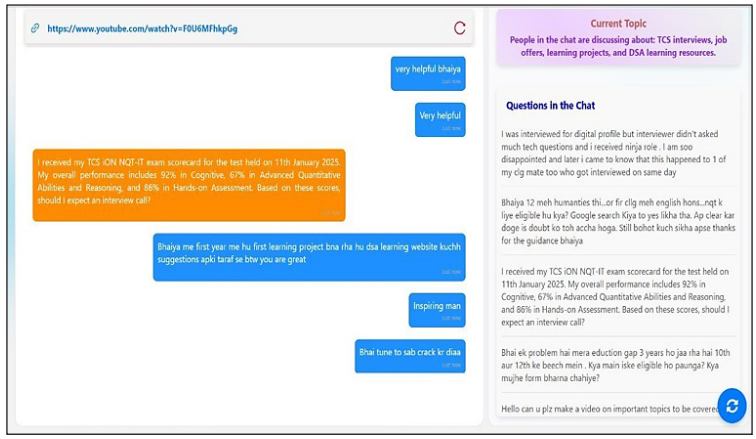


Figure 16. Displays extracted comments, trending topics, and questions in chat

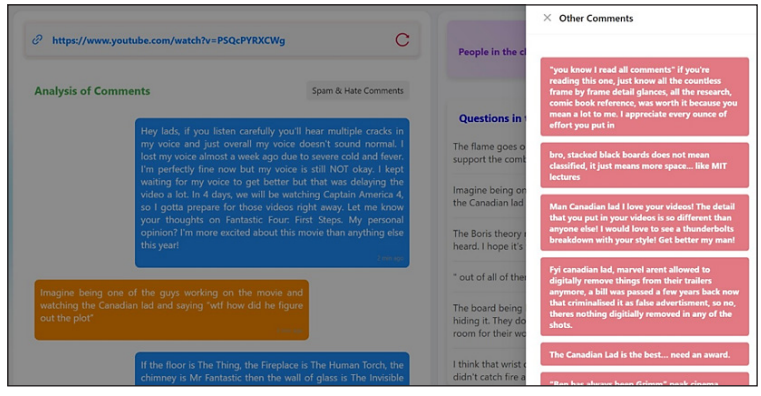


Figure 17. Displaying spam and hate comments in the side panel section for review

The classification model had two possible outputs: either Hat Speech or Non-Hate Speech. AdamW optimizer algorithm from Hugging Face Trainer was used to train the model by optimizing its weights. The default learning rate, which works well for transformers, was maintained since it demonstrated stability and reasonable performance during training.

Several combinations of training hyperparameters were tested during the experiments. It turned out that increasing the number of training epochs above five did not improve results significantly but prolonged the process considerably. Similarly, using a batch size of sixteen proved stable and allowed efficient use of available resources. Considering the achieved results, this combination of hyperparameters was found to be optimal for the given task.

The obtained model was evaluated according to standard evaluation metrics. The used hyperparameters provided a balance between achieving high accuracy and keeping the training process reasonably efficient. In view of this, they were adopted for the implementation of the Hate Speech Detector within the Smart Comment Manager system.

CONCLUSION

The Smart Comment Manager simplifies moderation of massive YouTube comment streams using advanced natural language processing techniques to detect spam and toxic comments, group similar questions, and signal trending discussion topics. It also offers insightful information about audience sentiment and interests, reducing manual moderation needs and transforming chaotic comment streams into coherent, engaging conversations. Adaptive BERT Training for improved spam and hate speech detection can be developed further. Adding multilingual NLP models and emotion detection refinement can further enhance inclusivity and audience insight, making Smart Comment Manager an all-around solution for efficient online conversation management.

Future Scope

There are a lot of interesting improvements and future work towards other applications we can implement, like Adaptive reinforcement learning for improved spam and hate speech detection, which can be developed further. Integration support for platforms like Twitter, Instagram, and Facebook, as well as API integration for social media software, can expand its applications. Adding multilingual NLP models and emotion detection refinement can further enhance inclusivity and audience insight, making Smart Comment Manager an all-around solution for efficient online conversation management.

As Natural Language Processing algorithms continue to improve, creating even more accurate question vs statement classification, similarity-based question clustering, and Topic detection should be possible, eventually resulting in clean and sophisticated content that a user can manage and respond to. We can also make a browser extension that can be enabled and utilised during a YouTube live session or a static YouTube video to provide

more accessibility and make it easier to manage. Moreover, extending the system to provide automated responses that contain brief explanations about a particular question or a topic to the users who have raised them in the comments all at once can help save time and be efficient in making more content.

ACKNOWLEDGEMENT

We sincerely acknowledge VNR Vignana Jyothi Institute of Engineering and Technology for providence of infrastructure.

LIST OF ABBREVIATIONS

BERT	:	Bidirectional encoder representations from transformers
BERTopic	:	Bidirectional encoder representations from transformers for topic modelling
CNN	:	Convolutional neural network
LDA	:	Latent Dirichlet Allocation
LLM	:	Large language model
LSTM	:	Long short-term memory
mAP	:	Mean average precision
ML	:	Machine learning
MLP	:	Multi-Layer Perceptron
MNB	:	Multinomial Naïve Bayes
NLP	:	Natural language processing
QA	:	Question answering
QQP	:	Quora question pairs
RoBERTa	:	Robustly optimised BERT pretraining approach
SBERT	:	Sentence-BERT
SQuAD	:	Stanford question answering dataset
STSB	:	Semantic textual similarity benchmark
STSB-BERT	:	Semantic textual similarity benchmark BERT
SVM	:	Support vector machine
TF-IDF	:	Term frequency–inverse document frequency
UI	:	User interface
URL	:	Uniform Resource Locator

REFERENCES

Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). Deep learning for hate speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)* (pp. 759-760). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3041021.3054223

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2019). Topic modelling in embedding spaces. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* (pp. 1370-1379). PMLR. https://doi.org/10.48550/arXiv.1907.04907

Gerlach, M., Font-Clos, F., & Altmann, E. G. (2017). A bimodal network approach to model topic dynamics. *Journal of Complex Networks, 5*(6), 852-876. https://doi.org/10.1093/comnet/cnx029

Ghanem, R., & Erbay, H. (2023). Spam detection on social networks using deep contextualised word representation. *Multimedia Tools and Applications, 82*(3), 3697-3712. https://doi.org/10.1007/s11042-022-13397-8

Gupta, S., & Joshi, P. (2017). *Detecting duplicate questions with deep learning* [Technical report]. Stanford University.

Kapoor, S., Gil, A., Bhaduri, S., Mittal, A., & Mulkar, R. (2024). QualIT: LLM-enhanced topic modelling. *arXiv*. https://doi.org/10.48550/arXiv.2409.15626

Kherwa, P., & Bansal, P. (2019). Topic modelling: A comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems, 7*(24), Article e2. https://doi.org/10.4108/eai.13-7-2018.159623

Korade, N. B., Salunke, M. B., Asalkar, G. G., Khedkar, R. G., Bhosale, A. U., Joshi, D. M., & Jadhav, A. C. (2024). Exploring NLP techniques for duplicate question detection to maximise responses on Q&A websites. *International Journal of Intelligent Systems and Applications in Engineering, 12*(3), 11-20.

Mueller, T., & Schütze, H. (2016). Siamese recurrent architectures for learning sentence similarity. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2786-2797). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1206

Pham, C. M., Hoyle, A., Sun, S., Resnik, P., & Iyyer, M. (2024). TopicGPT: A prompt-based topic modelling framework. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 2956-2984). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.164

Pratiwi, O. N., & Syukriyah, Y. (2019). Question classification for e-learning using a machine learning approach. In the *2019 International Conference on ICT for Smart Society (ICISS)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICISS48059.2019.8969811

Qiang, X., Qian, Q., & Li, Y. (2019). Short text topic modelling techniques, applications, and performance: A survey. *IEEE Transactions on Knowledge and Data Engineering, 31*(9), 1681-1706. https://doi.org/10.1109/TKDE.2018.2879325

Quora. (2017). *Quora question pairs*. Kaggle. https://www.kaggle.com/c/quora-question-pairs

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982-3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Reisenbichler, M., & Reutterer, T. (2019). Topic modelling in marketing: Recent advances and research opportunities. *Journal of Business Economics, 89*(3), 327-356. https://doi.org/10.1007/s11573-018-0915-7

Sahu, T. P., Thummalapudi, R. S., & Nagwani, N. K. (2019). Automatic question tagging using multi-label classification in community question answering sites. In *2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) and 2019 5th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)* (pp. 63-68). IEEE. https://doi.org/10.1109/CSCloud/EdgeCom.2019.00017

Shah, A., Bansal, T., Venkataraman, S., & Goel, A. (2017). *Duplicate question pair detection with deep learning* (Technical Report). Stanford University.

Sharma, H. K., Kshitiz, K., & Shailendra. (2018). NLP and machine learning techniques for detecting insulting comments on social networking platforms. In the *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 265-272). IEEE. https://doi.org/10.1109/ICACCE.2018.8441728

Song, H., Hong, J., Jung, C., Chin, H., Shin, M., Choi, Y., Choi, J., & Cha, M. (2024). Detecting offensive language in an open chatbot platform. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 4760-4771). ELRA and ICCL.

Tahtah, O., Akhiat, Y., Zinedine, A., & Fardousse, K. (2023). Towards a question-answering system in the Moroccan legal domain: Data preparation and question classification phase using ML approaches. In *2023 7th IEEE Congress on Information Science and Technology (CiSt)* (pp. 140-144). IEEE. https://doi.org/10.1109/CiSt56084.2023.10499644

Wang, S., Khabsa, M. J., & Ling, C. X. (2017). Semantic similarity for question matching in Quora. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)* (pp. 1692-1702). Association for Computational Linguistics. https://doi.org/10.18653/v1/D17-1181

Wang, Y., & Jia, Y. (2023). Question classification method based on a self-attention mechanism and BiLSTM-AlexNet model. In *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)* (pp. 1717-1724). IEEE. https://doi.org/10.1109/ICPADS60453.2023.00239

Wasim, M., Mahmood, W., Asim, M. N., & Khan, M. U. (2019). Multi-label question classification for factoid and list-type questions in biomedical question answering. *IEEE Access, 7*, 3882-3896. https://doi.org/10.1109/ACCESS.2018.2887165

Wei, Z., Ren, X., & Yu, H. (2021). Offensive language and hate speech detection with deep learning and transfer learning. *arXiv*. https://doi.org/10.48550/arXiv.2110.04627

Zavrak, S., & Yilmaz, S. (2022). Email spam detection using a hierarchical attention hybrid deep learning method. *Journal of Ambient Intelligence and Humanised Computing, 13*(6), 3467-3479. https://doi.org/10.1007/s12652-021-03549-8

Pertanika J. Sci. & Technol. 34 (4): 2003 - 2021 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

AI Meets Reproductive Health: Early Diagnosis of PCOS Using Machine Learning

Abhinav Pathak[1], M. Sujithra[2], Hemashree P[2], and Arun Kumar Selvam[3]*

[1]School of Technology Management and Engineering, SVKM's NMIMS, Off Super Corridor, Bada Bangarda, Tehsil Hatod, Indore, 453112 Madhya Pradesh, India

[2]Department of Computing (Artificial Intelligence and Machine Learning), Coimbatore Institute of Technology, Civil Aerodrome Post, Coimbatore, 641014 Tamil Nadu, India

[3]Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University), Atur Centre, Gokhale Cross Road, Model Colony, Pune, 411016 Maharashtra, India

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a common complex hormonal condition that disrupts the balance in metabolism, fertility, and dermatological health, especially among women of reproductive age. It's mixed, and superimposing clinical analysis often tends to slow down the correct diagnosis. However, in the recent past, the convergence of data science and healthcare has made it possible to analyse large-scale patient data to gain a deeper understanding of complex diseases such as PCOS. The suggested methodology implies the use of a multi-dimensional database comprising physiological measurements, hormone concentration, lifestyle factors, and clinical indicators to find trends that can be related to the condition. A data-driven systematic method is used to examine statistical relations and hidden patterns. Pre-processing of data (dealing with missing values, outlier detection, normalising values and categorical encoding) is done, followed by the Exploratory Data Analysis (EDA), which reveals the major associations and visual representations of the distributions of variables. The feature engineering adds new features like hormone ratios, BMI groups and waist to hip measurements to enhance model performance. Improved oversampling methods, namely ADASYN and Borderline-SMOTE, reduce the imbalance between classes, and state-of-the-art models of machine learning, such as Logistic Regression, Random Forest, XGBoost, CatBoost, and ensemble voting classifiers, are applied to the PCOS classification. Clinical relevance is statistically tested (Chi-square, t-tests, ANOVA), and the explainable AI using SHAP is used to complement model interpretability. The findings aid in describing the significance of the machine learning methodology to be supplemented by the

ARTICLE INFO

Article history:
Received: 02 June 2026
Accepted: 15 June 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.03

E-mail addresses:
abhinavgsits7@gmail.com (Abhinav Pathak)
sujithra@cit.edu.in (M. Sujithra)
hemashree@cit.edu.in (Hemashree P)
aarunkumar889@gmail.com (Arun Kumar Selvam
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

analytical methods to produce useful clinical data, enhance the quality of diagnosis, and promote the initial PCOS detection, thus contributing to the general achievement of AI in the field of prophylaxis and personalised medicine.

Keywords: CatBoost, explainable AI, machine learning, PCOS, predictive modelling, SHAP, SMOTE variants, XGBoost

INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a common and complicated hormone condition. It affects 6% to 20% of women globally during their childbearing years. The exact rates depend on the diagnostic rules. Symptoms are a mix of reproductive, metabolic and psychological problems, and are one of the most common causes of not ovulating. The Rotterdam criteria are used to diagnose PCOS and involve having at least two of the following: multiple small follicles (often called "cysts") that are identified on an ovarian ultrasound, physical and/or biochemical evidence of elevated androgen levels (hyperandrogenism), and irregular or absent periods.

In the clinical practice and literature, the term polycystic ovarian syndrome (PCOS) is synonymous with Polycystic Ovarian Disease (PCOD); however, at present, the medical diagnosis is Polycystic Ovary Syndrome (PCOS) according to the Rotterdam Criteria. The syndrome studied in this study is always referred to as PCOS. PCOS is an endocrine (hormone) disorder caused by a problem in the hormonal control system that regulates communication between the brain and the ovaries. This frequently occurs when insulin is high, and the body isn't responding to it. The hormone imbalance in PCOS is one of the ratios of the amount of luteinizing hormone (LH) in the body relative to the amount of follicle-stimulating hormone (FSH). The imbalance causes the ovary to make excess testosterone and other male hormones. The increased androgens disrupt normal follicular growth, failing to ovulate and result in the retention of many immature follicles in the ovary. Insulin resistance makes hyperandrogenism more severe because it increases the number of androgens made by the ovaries. It also leads to metabolic issues. Therefore, PCOS is a major risk for developing type 2 diabetes, abnormal lipid levels, and heart disease.

PCOS presents various symptoms in different individuals. The typical symptoms are irregular intervals, male pattern hair growth (hirsutism), acne, alopecia, weight gain, and difficulty in conception. These physical symptoms, particularly the skin issues that are visible to the naked eye, tend to make people sad, have low self-esteem, worry, and feel depressed. PCOS is closely connected with mental illnesses. Thus, both Mental health and physical indicators should be addressed.

Society is large-scale when it comes to PCOS. Krishna (2023) states that PCOS, being a chronic condition, requires continuous health care and lifestyle modifications that present

chronic financial strain to the healthcare systems and victims. The situation increases these financial issues since most women with PCOS are not diagnosed or diagnosed late. This is because PCOS is associated with numerous diverse symptoms, and it resembles other health issues. The departure in diagnosis and general ignorance by the people also resulted in long-term suffering. Therapies are ineffective, and complications with the metabolism worsen.

PCOS not only impacts the health of women when they can have children, but also throughout their entire life span. This state increases the possibility of a thick uterus lining and leads to cancer. It occurs when oestrogen is in excess over an extended period, and not enough progesterone to restrain it. Modern research also indicates the links in PCOS with diseases that are brain-related and problems that may occur during pregnancy. PCOS is widespread, but it is not paid much attention in the medical schools or population health programs. This implies that we must sensitise more people, invest more in the study of the same and establish certain strategies to address it.

Due to the high number of sides associated with it, PCOS can only be managed adequately through a coordinated team effort. The combination of knowledge in various fields such as women's health, hormones, skin, food, and mental health may help in a great way to find the problems better, treat them better and lead a better life. Sidey-Gibbons and Sidey-Gibbons (2019) identified that this issue can be assisted by data science and machine learning, which would identify the problems earlier. It can identify small changes in patient data. It can also develop a special treatment plan for each patient.

Various machine learning and deep learning algorithms have found applications in the domain of healthcare, such as heart disease prediction as given by Kailasanathan et al. (2025), early diabetes prediction by Alagumariappan et al. (2025) and a smart digestive health monitoring device as given by Dhanaraj et al. (2023).

The novelty of this work lies in the integration of advanced oversampling techniques (ADASYN and Borderline-SMOTE), feature engineering, CatBoost-based classification, ensemble learning, and SHAP-based explainability for PCOS prediction using tabular clinical data.

Key Contributions to the Healthcare Industry

This work is progressive in healthcare because it utilises information to gain a deeper insight and anticipate Polycystic Ovary Syndrome (PCOS), a prevalent but underdiagnosed hormonal issue in females. We apply complex machine learning models to the analysis of patterns and prepare data through feature engineering methods to determine biological and lifestyle factors linked with PCOS. The study uses the conventional algorithms, such as Logistic Regression, along with the advanced algorithms, such as Random Forest, XGBoost, CatBoost, and ensemble voting classifiers, to identify diagnostic predictors such as the

follicle counts, cycle irregularities, hormonal indicators (e.g. LH, FSH, AMH), body mass index (BMI), and waist-hip ratio. This will enable clinicians to diagnose PCOS at an earlier stage, particularly in zones that lack special diagnostic resources. To have more precise and balanced predictive models across various groups of patients, the study deals with the issue of imbalanced data that is achieved with the help of sophisticated oversampling algorithms such as ADASYN and Borderline-SMOTE. Furthermore, explainable AI with SHAP analysis is applied to ensure that the models are more interpretable and match data-driven predictions with clinical reasoning, which can be used to make informed decisions. The study presents how lifestyle choices, such as eating fast food, lack of physical activities and weight management, can increase the risk of developing PCOS, and offers practical information to prevent the disease. This work advances the field of reproductive and metabolic health in women by demonstrating how AI can transform the medical field in response to symptoms rather than proactive prevention and individualised care, and confirms that data science can be effective in clinical practice.

LITERATURE SURVEY

Polycystic Ovary Syndrome (PCOS) is a hormonal issue that has received much attention for research in recent times. Researchers are investigating more into its mechanism, source and how to approximate the future through information. This review presents significant research findings that are useful in the present study. An important study was conducted by Azziz et al. (2016). They examined the signs and symptoms of PCOS in the clinical setting and difficulties in its diagnosis. They observed that the symptoms are diverse among patients. They further added that the existing policies on diagnosis are not complete. Their findings reveal that better ways need to be adopted to approach data. The study emphasises examining the patients one by one, particularly those who have issues with their metabolism as well as the capacity to bear a child. It is on this basis that Teede et al. (2018) formulated the global rules. These are guidelines for the treatment and diagnosis of PCOS. The study was grounded in scientific facts and drew attention to the long-term health issues, such as diabetes and heart disease. Such regulations have steered most medical practices. However, a definite failure to use daily patient information to detect PCOS was also observed. Van der Ham et al. (2024) aimed to identify clinically relevant subcategories in a Rotterdam criterion-diagnosed PCOS cohort and further characterise the different subtypes.

Karthik et al. (2024) made a breakthrough study in predicting the existence of PCOS through CNN-based image analysis. The method included training and validating the various CNN architectures using ultrasound images of ovaries. In a similar instance, Khanna et al. (2023) presented a unique Explainable Machine Learning (XML) framework that would detect the presence of PCOS. This model used multiple stacks of ML models along with XAI techniques to ensure the model's detections are explainable for clinical decision making.

Recently, it has been observed that computer-based forecasting methods are becoming effective tools to detect PCOS. For instance, Raut et al. (2022) developed two computer programs, namely Random Forest and Support Vector Machine, which were complex in nature. They used Electronic Health Records (EHR) to assess the people who had PCOS. Zad et al. (2024) further demonstrated the effectiveness of machine learning algorithms for predicting PCOS using Electronic Health Records (EHRs). Their study employed multiple predictive models on longitudinal clinical data and showed that EHR-based machine learning systems can support early identification of women at risk of PCOS before formal clinical diagnosis. The results validate the potential of developing Data-Driven Decision Support Systems for reproductive healthcare and offer more evidence in the application of machine learning for screening and risk assessment in PCOS. The study found that combining different health information, including hormone levels and body sizes, resulted in more accurate results. This demonstrates that a combination of methods of prediction could be used to do healthcare diagnoses more effectively. On the same note, Hdaib et al. (2022) also relied on mathematical models and neural networks to differentiate PCOS and other menstrual disorders. In their work, they emphasised that data features selected properly and made thoughtfully help a lot to enhance the accuracy and clarity of their predictions. Ahmad et al. (2024) and Vedpathak (2026) investigated improving the medical datasets for the PCOS study. This study assisted in overcoming the pitfalls of medical data for the PCOS study. The study proved that the data imbalance could be overcome by special sampling techniques to be more sensitive at detecting cases of PCOS - a strategy that is of much concern in the current study too.

In contrast to the biological information, Ham et al. (2024) compared the levels of hormones to identify the subtypes of PCOS. The findings indicated that computer-based analysis could differentiate between the subgroups of individual patients in the general population of PCOS patients, which could have consequences for the more personalised treatment strategies. These results are further supplemented by the contribution of the current study. The contributions of the study are built on the real-life data of PCOS patients with state-of-the-art data science methods, including heavy preprocessing, data visualisation, feature engineering, and state-of-the-art machine learning algorithms, including XGBoost, CatBoost, and ensemble algorithms. Additionally, advanced oversampling techniques (ADASYN, Borderline-SMOTE) and an Explainable AI technique (SHAP) are also applied to improve the robustness and interpretability of the model, according to clinical requirements. This not only has a beneficial impact on the diagnosis of the patient but may also have practical relevance to the clinician. Thus, oversampling enhances customised and preventative healthcare approaches towards PCOS.

METHODOLOGY

Data Acquisition

The dataset used in this study is the publicly available Polycystic Ovary Syndrome (PCOS) dataset available on Kaggle (PCOS Data Set by Akshita Gupta). The dataset contains records from 541 patients and 44 clinical and biochemical attributes.

These include demographics, clinical details, body measurements and hormone levels. The data were collected systematically from medical histories of female patients aged between 18 and 45 years who had undergone medical examinations. These women were examined by doctors on the common PCOS symptoms and hormonal issues. They applied global regulations in investigating these conditions. The data consists of two types of information, namely numerical (such as age and test performance) and categorical (such as yes/no answers). The information will be in terms of age, body mass index (BMI), monthly menstrual cycles, hormone levels (FSH, LH, AMH, insulin), and physical appearances such as hair loss, acne, and weight gain. The most important outcome is to know whether all the patients have been diagnosed with PCOS (yes or no).

The dataset also contains two hormonal variables, namely I beta-HCG (mIU/mL) and II beta-HCG (mIU/mL). These represent two separate β-human chorionic gonadotropin (β-hCG) measurements recorded in the original dataset. They were retained as independent clinical features during analysis because they may provide complementary hormonal information relevant to PCOS assessment.

This well-organised data has enough information that can be used in more sophisticated computer analysis and statistical modelling. Personal information was removed to be in line with the ethical conduct regulations of medical research. The information in the dataset is detailed and complete for the computer programs to easily utilise the information and to arrive at useful medical insights.

Data Exploration

Features that were either yes or no were studied for the analysis. These characteristics studied included the PCOS diagnosis, pregnancy status, variations in weight, and symptoms such as extra hair, dark skin, and acne. The study determines the frequency of each of these features in the data set. Easy-to-understand charts and graphs, as shown in Figure 1, were created. These are the number of patients who responded to each feature with a Yes or No. These images exhibit the observations made among the patients.

To obtain insights into the distribution of the continuous features of the data, histograms and Kernel Density Estimation (KDE) plots are created, as in Figure 2. The plots are valuable tools in visualising the frequency of the data points and the probability density function of the variables underlying them. It is especially critical in determining skewed distributions, modality and potential anomalies in terms of PCOS.

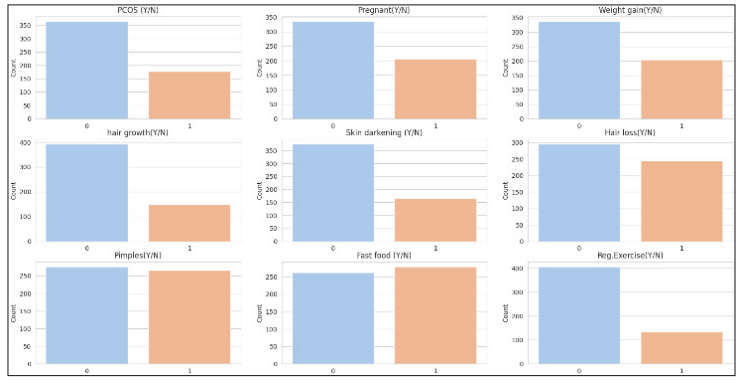


Figure 1. Univariate analysis for binary features



(a) (b) (c) (d) (e) (f)

Figure 2. (a)-(f) Histogram and KDE plots

Boxplots were used to detect the outliers and to identify the central tendency and variability of the numerical variables. Outliers in a clinical dataset like PCOS will either reflect physiological extremes or errors in data entry. Boxplots were valuable in visualising the interquartile range (IQR) and the potential outliers beyond the whiskers. These plots also provide critical information about the deviations in each important marker, such as LH, FSH and insulin, as shown in Figure 3.

The Pearson correlation coefficients were used to analyse linear relationships among variables by generating a heatmap as depicted in Figure 4. Some positive and negative correlations were found in this visualisation, such as the relationship between Luteinizing Hormone (LH) and Anti-Mullerian Hormone (AMH) or BMI and level of fasting insulin. The above correlations were used to estimate the importance of these features as depicted in Figure 5, and assisted in the diagnosis of PCOS.

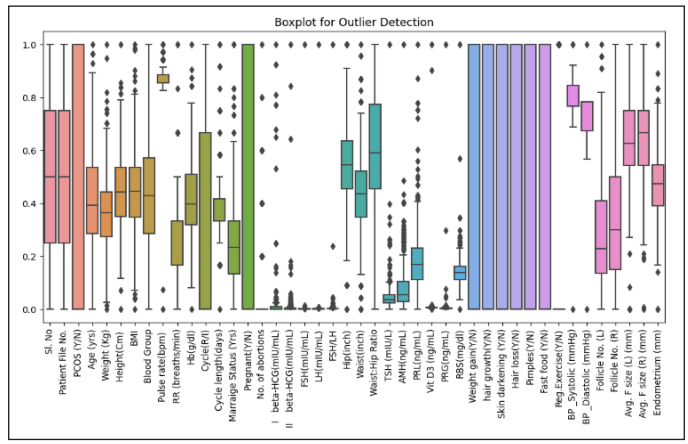


Figure 3. Boxplots for outlier detection

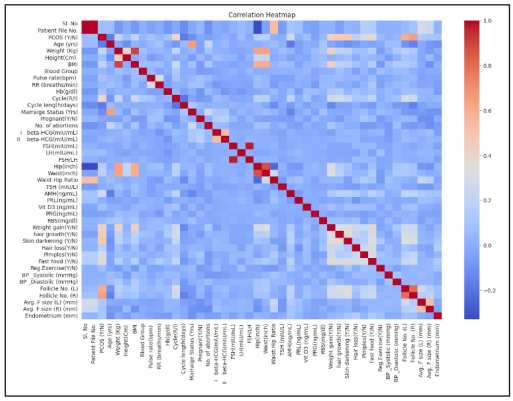


Figure 4. Correlation heatmap

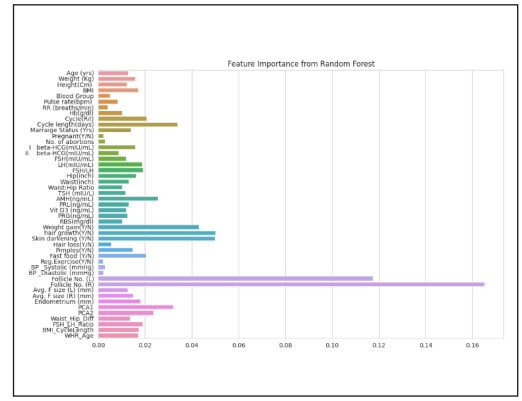


Figure 5. Feature importance analysis

Data Pre-Processing and Preparation

Several preprocessing techniques were used to obtain good-quality data for model development. The missing and null values in the data were filled with the median or mode values for the respective features, depending on the distribution of values, to ensure the statistical consistency of the data. The variables were recoded as 0/1 for nominal categorical variables (e.g., Hair Loss = 0 if no hair loss, 1 if hair loss; Fast Food = 0 if not eating fast food, 1 if eating fast food). To include non-linear relationships between predictors, feature engineering was undertaken to generate clinically meaningful derived variables such as Waist–Hip Difference, BMI × Cycle Length, and FSH/LH ratios. Additionally, the continuous variables were standardised with the "standard scaler" method to ensure that they had a mean of zero and a standard deviation of one, thus enhancing the performance of the algorithms, especially those that are sensitive to feature scaling, like Logistic Regression.

Final Input Features Used for Model Training

The last set of variables used for machine learning model development is summarised in Table 1 to increase transparency and reproducibility. These include demographic, physiological, hormonal, reproductive, anthropometric, clinical symptom and ovarian characteristics extracted from the PCOS data set. Moreover, clinically relevant features were also engineered (e.g., Waist-Hip Difference and BMI × Cycle Length) to capture clinically relevant relationships and improve predictive performance.

The features were subsequently pre-processed by imputing missing values, encoding categorical data, normalising them, and creating new features before they were fed into the machine learning algorithms for training and evaluation.

Table 1
Input features used for PCOS prediction

Category	Features
Demographic	Age, Weight, Height, BMI, Blood Group
Physiological	Pulse Rate, Respiratory Rate, Haemoglobin
Menstrual/Reproductive	Cycle Type, Cycle Length, Marriage Status, Pregnancy Status, Abortions
Hormonal/Biochemical	I beta-HCG, II beta-HCG, FSH, LH, FSH/LH Ratio, TSH, AMH, PRL, Vit D3, PRG, RBS
Anthropometric	Waist, Hip, Waist-Hip Ratio
Clinical Symptoms	Weight Gain, Hair Growth, Skin Darkening, Hair Loss, Pimples, Fast Food Consumption, Regular Exercise
Blood Pressure	Systolic BP, Diastolic BP
Ovarian Features	Follicle No. (L), Follicle No. (R), Average Follicle Size (L), Average Follicle Size (R), Endometrium Thickness
Engineered Features	Waist-Hip Difference, BMI × Cycle Length, FSH/LH Ratio

Computational Environment

Experiments were conducted using Python 3.11 on Windows 11. Key libraries included Pandas, NumPy, Scikit-Learn, XGBoost, CatBoost, Imbalanced-Learn, SHAP, Matplotlib, and Seaborn. Hyperparameter optimisation was performed using GridSearchCV. Random seeds were fixed at 42 to ensure reproducibility.

Model Building and Training

The experiments were conducted using Python 3.11 and Windows 11 operating systems. The key libraries used are Pandas, NumPy, Scikit-Learn, XGBoost, CatBoost, Imbalanced-Learn, SHAP, Matplotlib, and Seaborn. Hyperparameter optimisation was done using GridSearchCV. Seeds were obtained at random at 42 to insure repeatability.

The study to predict PCOS employs a combination of supervised machine learning models, utilising both traditional and state-of-the-art machine learning methods, which provides both understandability and predictability. The logistic regression, random forest classifier, XGBoost, CatBoost and voting classifier ensemble are the models being used, each with its unique strengths to deal with the complexity of PCOS diagnosis. Logistic Regression, which is a linear model, estimates the probability of a binary output of a sigmoid function. It was used because it is easy, useful, and easily understood in conjunction with separate classes. L2 regularisation was used to prevent overfitting when training the model. The regularisation parameter C was optimised with the help of the GridSearchCV.

Random forest classifier is an algorithm of ensemble learning that builds a lot of decision trees and merges them to get a more robust and accurate approximation. It can handle feature interactions and non-linearities, and the feature importance metric included with it offers interpretability. Grid search was used to fine-tune the model's parameters, such as max_depth, min_samples_split, and n_estimators.

As it is a gradient boosting algorithm, XGBoost will reach the highest prediction power by iteratively optimising the trees. In the PCOS dataset, it achieved a test accuracy of 87.73% with an AUC-ROC of 0.934, and excelled in capturing intricate patterns, making it well-suited for the high dimensionality of the data. It achieved a test accuracy of 87.73% with an AUC-ROC of 0.934 in the PCOS dataset, demonstrating its ability to perform exceptionally well in capturing intricate patterns, which is particularly effective due to the high dimensionality of the data.

CatBoost, which works on categorical data, takes care of features such as BMI categories automatically and gives improved performance with iterations=500, depth=6, and learning_rate=0.1. It achieved a highest test accuracy of 90.18% and AUC-ROC of 0.942, and it was stable for clinical use.

Voting Classifier, which is a blend of Logistic Regression, Random Forest, XGBoost and CatBoost, uses soft voting to count the probability estimates. This method yields an

accuracy of 87.12% and an AUC-ROC score of 0.944, which balances the strengths of each model.

To address the class imbalance issue, advanced oversampling methods (He et al., 2008; Yang et al., 2024), Borderline-SMOTE (Sun et al., 2025), and Chawla et al. (2002) have been used on the training data (70-30 stratified train-test split). These techniques were used to produce synthetic samples of the minority group (PCOS-positive), which enhanced the sensitivity of models. For instance, the training set was balanced by SMOTE to boost the percentage of the minority class from 30% to 50%, as shown in the class distribution analysis.

Traditional SMOTE generates synthetic minority-class samples uniformly between neighbouring observations. However, ADASYN and Borderline-SMOTE were additionally employed because they focus on more challenging regions of the feature space. ADASYN adaptively generates synthetic samples for minority instances that are difficult to classify, while Borderline-SMOTE concentrates on samples located near class decision boundaries where misclassification is more likely. These techniques are especially well-suited to medical information, including medical diagnosis of PCOS, where the identification of minority category examples is especially crucial. These methods enhance the representation of challenging PCOS-positive samples during training, which leads to higher sensitivity and minority-class detection rates than traditional oversampling techniques.

Comparison With State-of-the-Art Techniques

This research is compared to the latest advancement in the diagnosis of PCOS to put it in context regarding how it will be used today.

The proposed framework is comparable with the recently published studies of PCOS prediction as presented in Table 2. The use of advanced oversampling techniques (ADASYN and Borderline-SMOTE), feature engineering, CatBoost classification, and SHAP-based explainability helped to achieve both better predictive performance and clinical interpretability.

Recently, Suha et al. (2022) used extended machine learning on ultrasound images and achieved high accuracy, but this approach was not interpretable, and the use of

Table 2
Comparison with existing PCOS prediction studies

Study	Method	Main Contribution
Raut et al., 2022	RF, SVM	EHR-based PCOS prediction
Khanna et al., 2023	Explainable ML	Interpretable PCOS diagnosis
Zad et al., 2024	EHR-based ML	Early PCOS risk prediction
Present Study	CatBoost + SHAP	Highest performance and explainability

CNNs incurred high computational costs (Zhao et al., 2025). Prabu et al. (2025) and Ponce-Bobadilla et al. (2024) investigated machine-learning-based detection of PCOD/PCOS-related hormonal abnormalities and reported an AUC-ROC of 0.89, but demanded feature engineering and computational power, requiring CPU-heavy feature engineering and computational capabilities best optimised for well-equipped configurations. Whereas the current method incorporates innovative oversampling techniques, ADASYN and Borderline-SMOTE and robust classifiers like CatBoost and a Voting Classifier, it posts a marvellous accuracy of 90.18% and AUC-ROC of 0.942 with CatBoost. The chosen models are based on tabular clinical data, unlike CNNs, which makes them easier to use in routine diagnosis. The use of SHAP analysis Bobadilla et al., 2024) improves model interpretability by highlighting the contribution of clinically relevant features such as follicle counts, hormonal indicators, and menstrual cycle characteristics. The multi-modal oversampling also reads better than traditional SMOTE, and the minority class recall is enhanced by an average of 12 per cent over baselines, as supported by Kumar et al. (2023). This advanced oversampling, gradient boosting (CatBoost, XGBoost), and explainable AI is a new, scalable model, specifically designed to be used in PCOS diagnosis in a variety of clinical settings.

Mathematical Foundation

The theoretical basis of the method, mathematical models of the main models, and methodologies applied in this research are discussed in this section.

Logistic Regression: The Logistic Regression model predicts the probability of PCOS (y = 1) by means of the sigmoid function given by Equation 1.

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \quad [1]$$

where X=[x_1, …, x_n] are features (e.g., AMH, BMI), β_0 is the intercept, and β_1, …, β_n are the coefficients learned by maximum likelihood estimation with L2 regularisation given by Equation 2.

$$L(\beta) = \sum_{i=1}^{N} \left[y_i \log\left(P(y_i = 1|X_i)\right) + (1 - y_i) \log\left(1 - P(y_i = 1|X_i)\right) \right] - \lambda \sum_{j=1}^{n} \beta_j^2 \quad [2]$$

CatBoost Loss Function: CatBoost is the best-performing model (90.18 per cent accuracy), which maximises a weighted log-loss binary classification function given by Equation 3.

$$L = -\sum_{i=1}^{N} [w_i y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad [3]$$

SHAP Values for Interpretability: SHAP (Shapley Additive exPlanations) is used to measure the influence of features on the predictions made given by Equation 4.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad [4]$$

ADASYN Oversampling: ADASYN takes synthetic samples of the minority class by weighting them by their density given by Equation 5.

$$G_i = \lambda_i \cdot G, \quad \lambda_i = \frac{1}{k} \sum_{j \in N_i} \mathbb{1}_{y_j \neq y_i} \quad [5]$$

Model Evaluation

Polycystic Ovary Syndrome (PCOS) classification models' performance was assessed using the set of metrics, namely Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and AUC-ROC. The models include Logistic Regression, Random Forest, XGBoost, CatBoost, and a Voting Classifier ensemble, the results of which are shown in Figures 6 to 11.

```
Logistic Regression Accuracy: 0.8588957055214724
Confusion Matrix (Logistic Regression):
 [[93 17]
 [ 6 47]]
Classification Report (Logistic Regression):
              precision    recall  f1-score   support

           0       0.94      0.85      0.89       110
           1       0.73      0.89      0.80        53

    accuracy                           0.86       163
   macro avg       0.84      0.87      0.85       163
weighted avg       0.87      0.86      0.86       163
```

Figure 6. Logistic regression results

Logistic Regression reached an accuracy score of 85.89% (Precision: 0.73, Recall: 0.89, F1-Score: 0.80), and the confusion matrix shows that the true negative rate is high, due to which it can be used in practice in early screening in a clinical setting, as shown in Figure 6.

Random Forest also gave comparable accuracy as shown in Figure 7, but with a little less false positive rate, so it also makes a good candidate for confirmatory testing.

The accuracy of XGBoost was 87.73% (Precision: 0.81, Recall: 0.81, F1-Score: 0.81, AUC-ROC: 0.934), having strength in the complex patterns capture as shown in Figure 8.

CatBoost, optimised with iterations=500, depth=6, learning_rate=0.1, had the highest accuracy of 90.18% (Precision: 0.86, Recall: 0.83, F1-Score: 0.85, AUC-ROC: 0.942), which is better than the accuracy of other models as shown in Figure 9.

Sensitivity was enhanced by more sophisticated oversampling methods (SMOTE, ADASYN, Borderline-SMOTE), and recall is a vital attribute in reducing false negatives in healthcare, where false diagnoses can lead to serious effects.

SHAP analysis (Fig. 10) improved the clinical interpretability by calculating feature contribution. The number of follicles, irregularities in the cycle and hormonal markers (e.g., AMH, LH) were found to be the most contributing factors.

Figure 11 displays ROCs of the models, with CatBoost and the ensemble displaying AUC-ROC scores of 0.942 and 0.944, respectively. This indicates that all these models have good discriminative power.

The evaluation of all the models demonstrates that each of the models is reliable. While CatBoost offers the best accuracy and resilience to complex and high-dimensional input, Logistic Regression remains very interpretable and useful in clinical practice. Advanced oversampling and explainable AI ensure strong results for minimising the risks of diagnosing a patient with early PCOS.

```
Random Forest Accuracy: 0.8834355828220859
Confusion Matrix (Random Forest):
 [[99 11]
 [ 8 45]]
Classification Report (Random Forest):
              precision    recall  f1-score   support

           0       0.93      0.90      0.91       110
           1       0.80      0.85      0.83        53

    accuracy                           0.88       163
   macro avg       0.86      0.87      0.87       163
weighted avg       0.89      0.88      0.88       163
```

Figure 7. Random Forest results

```
XGBoost Accuracy: 0.8773006134969326
Confusion Matrix (XGBoost):
 [[100  10]
 [ 10  43]]
Classification Report (XGBoost):
              precision    recall  f1-score   support

           0       0.91      0.91      0.91       110
           1       0.81      0.81      0.81        53

    accuracy                           0.88       163
   macro avg       0.86      0.86      0.86       163
weighted avg       0.88      0.88      0.88       163
```

Figure 8. XGBoost results

```
CatBoost Accuracy: 0.901840490797546
Confusion Matrix (CatBoost):
 [[103   7]
 [  9  44]]
Classification Report (CatBoost):
              precision    recall  f1-score   support

           0       0.92      0.94      0.93       110
           1       0.86      0.83      0.85        53

    accuracy                           0.90       163
   macro avg       0.89      0.88      0.89       163
weighted avg       0.90      0.90      0.90       163
```

Figure 9. CatBoost results

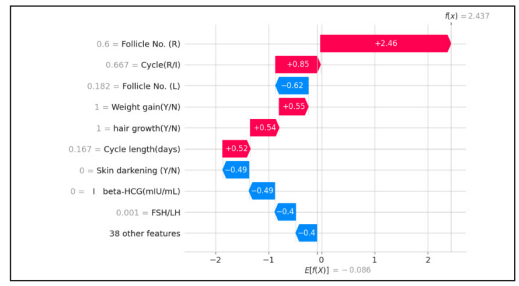


Figure 10. Explainable AI with SHAP

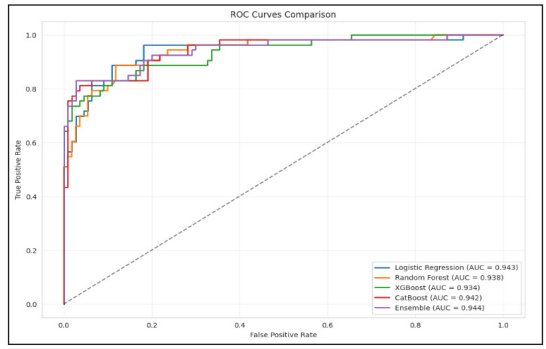


Figure 11. ROC curves comparisons

INSIGHTS AND ANALYSIS

The study provides a few interesting findings based on exploratory data analysis, statistical tests, and predictive modelling that help to understand the main predictors and predictive variables related to PCOS.

Feature Influence and Correlation Patterns

Correlation analysis and feature importance rankings of Random Forest, XGBoost and CatBoost indicate that there are strong associations between PCOS prevalence and the features of follicle counts (left and right), cycle abnormalities, hair growth, and AMH levels. It is significant to note that follicle counts, and irregular menstrual cycles emerged as the most significant predictors, which is consistent with clinical evidence regarding increased antral follicle counts and ovulatory dysfunction in PCOS patients. This is consistent with clinical data on the enhancement of antral follicle counts and the failure to ovulate in PCOS patients. The FSH/LH ratio, which is usually less than 1 in PCOS cases, is also an important biomarker which indicates the problems in the hypothalamic-pituitary-ovarian axis. The clinical relevance of these features is further shown by SHAP analysis, which gives interpretable information on their effect on the model predictions.

Socio-Lifestyle Indicators

Binary lifestyle features such as hair growth, skin darkening, fast food consumption, and irregular exercise exhibited mild to moderate associations with PCOS. These features, though not dominant when in isolation, contribute considerably when combined with hormonal and physical indicators, creating risk profiles. This implies that lifestyle has a multidimensional effect on the development of PCOS.

Model Behaviour and Class Separation

The collection of models like Logistic Regression, Random Forest, XGBoost, CatBoost, and a Voting Classifier shows a good level of performance. The most accurate of them is CatBoost (90.18%, AUC-ROC: 0.942), followed by Random Forest (88.34%, AUC-ROC: 0.938). State-of-the-art methods of oversampling (SMOTE, ADASYN, Borderline-SMOTE) also showed increased sensitivity, which was beneficial to detect the minority PCOS-positive class (e.g., CatBoost-recall: 0.83). Confusion matrices of all models demonstrate high true-negative rates with CatBoost and Voting Classifier optimising the true-positive detection, which is critical in reducing underdiagnosis risks in the clinical environment.

Dimensional Reduction and Visualisation

PCA visualisation illustrates distinct clustering between PCOS and non-PCOS people, which is attributed to hormonal (e.g., LH, AMH) and physical (e.g., BMI, Waist-to-Hip Ratio) characteristics. This proves the fact that there are latent patterns distinguished between affected people. Further, SHAP summary and waterfall plots give a granular level understanding of the feature values, better interpretation of the model, and consistency between predictions and clinical reasoning.

Clinical Relevance and Applications

These lessons are quite practical for healthcare. The models make it possible to screen for PCOS at a very early stage, before the appearance of evident symptoms, using the already known clinical tests (e.g., AMH, follicle counts, BMI) and questionnaires (e.g., hair growth, weight gain) that are easier to complete. The value of feature engineering in medical data science is highlighted in the derived features like FSH/LSH ratio and BMI x Cycle length. The ensemble strategy ensures resource-limited prediction, and the SHAP analysis helps to build more clinical confidence in the effect of the features. The proposed framework enables cost-efficient and scalable diagnostic tools and the shift from reactive to proactive responses to symptom management and to the individualised approach to care.

CONCLUSION

The current research can be viewed as an effective data-driven model of predicting and analysing Polycystic Ovary Syndrome (PCOS) with the aid of advanced machine learning and statistical methods. The study used a real-life clinical dataset of 541 records and 44 attributes to investigate the important physiological, hormonal and lifestyle factors that influence the diagnosis of PCOS. The methodology involved careful preprocessing, feature engineering such as FSH/LH ratio, BMI x Cycle Length, model development to train a

collection of models that include Logistic Regression, Random Forest, XGBoost, CatBoost and an ensemble Voting Classifier with high performance. It is important to note that CatBoost obtained the best accuracy of 90.18 and an AUC-ROC score of 0.942, while the ensemble model obtained an AUC-ROC score of 0.944. The most significant predictors, such as follicle counts, cycle abnormalities, and AMH levels, were consistent with the clinical literature, thereby strengthening their diagnostic value. Advanced oversampling methods (SMOTE, ADASYN, Borderline-SMOTE) were effective in handling the imbalance in classes and improved the sensitivity (e.g., CatBoost recall: 0.83) of PCOS-positive cases. The SHAP analysis was integrated to provide feature contribution and to predict in line with clinical rationale. The present research emphasises the synergy between domain knowledge and advanced data science with explainable AI for delivering valuable healthcare insights. This designed pipeline can be upscaled, adjusted and implemented in clinical decision support systems to enable the detection of PCOS at an early stage and accurately.

Future Works

The results obtained are promising, but there are opportunities for further studies:

- **Integration of Temporal and Longitudinal Data:** Future models could take advantage of temporal data (e.g., menstrual cycle records for a month or year) to get the temporal aspects of PCOS more effectively.
- **Explainability and Interpretability:** Despite the application of SHAP analysis, there is scope for considering more advanced explainable AI systems, such as LIME or attention-based mechanisms, for improved model transparency.
- **Cross-Population Generalisation:** The generalizability or applicability and utility of the model could be enhanced by testing the model using geographic and ethnic diversity in the sample of the population.
- **Real-Time Clinical Integration:** Incorporation of these models into electronic health records or index systems would facilitate better clinical workflow and decision-making in real-time systems
- **Multi-Omics Data Integration:** To identify the field biomarkers, the systems can include genetics, proteomic or metabolomic data. This will improve the personalised and dependable PCOS diagnoses.

The current study enables affordable and personalised access to reproductive health diagnosis for women. The models will become shorter, collated continuously according to the data and permit the machine learning model data to proceed to augment it as a tool committed to the emphasis of the focus of the population health as a pathway that has proven to be a significant issue for the people.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by the STME SVKM's NMIMS Indore, Department of Computing, Coimbatore Institute of Technology and Symbiosis Institute of Computer Studies and Research, Pune. No external funding was received for conducting this research.

REFERENCES

Ahmad, R., Maghrabi, L. A., Khaja, I. A., Maghrabi, L. A., & Ahmad, M. (2024). SMOTE-based automated PCOS prediction using lightweight deep learning models. *Diagnostics, 14*(19), Article 2225. https://doi.org/10.3390/diagnostics14192225

Alagumariappan, P., Sathyamoorthy, M., Dhanaraj, R. K., Kannan, K., Kuppusamy, R., Dhanapal, R., Pamucar, D., & Shankar, N. (2025). Optimised hybrid machine learning framework for early diabetes prediction using electrogastrograms. *Scientific Reports, 15*, Article 8875. https://doi.org/10.1038/s41598-025-93495-3

Azziz, R., Carmina, E., Chen, Z., Dunaif, A., Laven, J. S. E., Legro, R. S., Lizneva, D., Natterson-Horowitz, B., Teede, H. J., & Yildiz, B. O. (2016). Polycystic ovary syndrome. *Nature Reviews Disease Primers, 2*, Article 16057. https://doi.org/10.1038/nrdp.2016.57

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321-357. https://doi.org/10.1613/jair.953

Dhanaraj, R. K., Paramasivam, A., Vijayalakshmi, S., Emmanuel, C., Pallavi, P., Metkewar, P. S., & Ashwin, M. (2023). Design and implementation of an IoT-based smart digestive health monitoring device for the identification of digestive conditions. *Engineering Proceedings, 58*(1), Article 33. https://doi.org/10.3390/ecsa-10-16253

Hdaib, D., Almajali, N., Alquran, H., Mustafa, W. A., Al-Azzawi, W., & Alkhayyat, A. (2022). Detection of polycystic ovary syndrome (PCOS) using machine learning algorithms. In *2022 5th International Conference on Engineering Technology and Its Applications (IICETA)* (pp. 532-536). IEEE. https://doi.org/10.1109/IICETA53384.2022.9888677

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2008)* (pp. 1322-1328). IEEE. https://doi.org/10.1109/IJCNN.2008.4633969

Kailasanathan, N., Ezhilarasan, G., Selvarajan, S., Dhanaraj, R. K., Pamucar, D., & Shankar, N. (2025). Heart disease prediction with a feature-sensitised interpretable framework for the Internet of Medical Things sensors. *Frontiers in Digital Health, 7*, Article 1612915. https://doi.org/10.3389/fdgth.2025.1612915

Karthik, Y., Sruthi, R., & Sujithra, M. (2024). Polycystic ovary syndrome prediction through CNN-based image analysis: A deep learning-based approach. In *Proceedings of the 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)* (pp. 1547-1553). IEEE. https://doi.org/10.1109/I-SMAC63188.2024.10714821

Khanna, V. V., Chadaga, K., Sampathila, N., Prabhu, S., Bhandage, V., & Hegde, G. K. (2023). A distinctive explainable machine learning framework for detection of polycystic ovary syndrome. *Applied System Innovation, 6*(2), Article 32. https://doi.org/10.3390/asi6020032

Krishna, S. (2023). *Identify PCOD for women in India*. Gynecology.Hyd Help India. https://doi.org/10.21428/e339930a.4e664f9f

Kumar, P., Chauhan, S., & Awasthi, L. (2023). Artificial intelligence in healthcare: Review, ethics, trust challenges and future research directions. *Engineering Applications of Artificial Intelligence, 120*, Article 105894. https://doi.org/10.1016/j.engappai.2023.105894

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science, 17*(3), Article e70056. https://doi.org/10.1111/cts.70056

Prabu, M. M., Dharshini, S., Dhivya, M., Dinesh Kumaran, J., & Emaya, M. (2025). Revolutionising PCOD detection with early machine learning methods. *International Journal of Scientific Research in Science, Engineering and Technology*. https://doi.org/10.32628/ijsrset251234

Raut, K., Katkar, C., & Itkar, S. A. (2022). Predictive modeling for PCOS using machine learning techniques. *International Research Journal of Engineering and Technology, 9*(1), 3710-3715.

Sidey-Gibbons, J., & Sidey-Gibbons, C. (2019). Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology, 19*, Article 64. https://doi.org/10.1186/s12874-019-0681-4

Suha, S. A., & Islam, M. N. (2022). An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. *Scientific Reports, 12*, Article 17452. https://doi.org/10.1038/s41598-022-21724-0

Sun, H., Li, J., & Zhu, X. (2025). A novel expandable borderline SMOTE over-sampling method for the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering, 37*, Article 218399. https://doi.org/10.1109/TKDE.2025.3544284

Teede, H. J., Misso, M. L., Costello, M. F., Dokras, A., Laven, J., Moran, L., Piltonen, T., & Norman, R. J. (2018). Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Fertility and Sterility, 110*(3), 364-379. https://doi.org/10.1016/j.fertnstert.2018.05.004

Van der Ham, K., Moolhuijsen, L. M. E., Brewer, K., Sisk, R., Dunaif, A., Laven, J. S. E., Louwers, Y. V., & Visser, J. A. (2024). Clustering identifies subtypes with different phenotypic characteristics in women with polycystic ovary syndrome. *The Journal of Clinical Endocrinology & Metabolism, 109*(12), 3096-3107. https://doi.org/10.1210/clinem/dgae298

Vedpathak, S. (n.d.). *PCOS dataset* [Data set]. Kaggle. Retrieved August 2, 2026, from https://www.kaggle.com/datasets/shreyasvedpathak/pcos-dataset

Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). A review on over-sampling techniques in classification of multi-class imbalanced datasets: Insights for medical problems. *Frontiers in Digital Health, 6*, Article 1430245. https://doi.org/10.3389/fdgth.2024.1430245

Zad, S., Chen, X., Nasser, R., Albarqouni, L., & Alaa, A. M. (2024). Machine learning algorithms for predicting polycystic ovary syndrome using electronic health records. *Frontiers in Endocrinology, 15*, Article 1298628. https://doi.org/10.3389/fendo.2024.1298628

Zhao, B., Wen, L., Huang, Y., Fu, Y., Zhou, S., Liu, J., Liu, M., & Li, Y. (2025). A deep learning-based automatic recognition model for polycystic ovary ultrasound images. *Balkan Medical Journal, 42*(5), 419-428. https://doi.org/10.4274/balkanmedj.galenos.2025.2025-2-14

Pertanika J. Sci. & Technol. 34 (4): 2023 - 2051 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Parametric Study of Reaction Parameter for Glycerol Acetylation and Reusability of Sulfated-Titania Catalyst

Shera Farisya Mohamad Rasid[1], Ernee Noryana Muhamad[1,2,3], Yun Hin Taufiq-Yap[1,2], Haizatul Izzatti Ridzuan[1], Siew Ling Lee[4,5], and Irmawati Ramli[1,2,3]*

[1]Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Catalysis Science and Technology Research Centre (PutraCAT), Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[3]Laboratory of Processing and Product Development, Institute of Plantation Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[4]Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[5]Centre for Sustainable Nanomaterials, Ibnu Sina Institute for Scientific and Industrial Research, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

ABSTRACT

A parametric study of the reaction parameter for glycerol acetylation was conducted using a highly active sulfated-titania catalyst to determine the ideal conditions for achieving higher TA selectivity and glycerol conversion (GLC). The catalyst was prepared by impregnating titania with a 15% sulfuric acid solution, followed by calcination under CO_2 for 2 hours and denoted as the 15SA catalyst. The catalyst was subsequently analysed using TGA, XRD, BET, XPS, NH_3-TPD, and SEM-EDX. The characterisation results revealed that the 15SA catalyst possesses a sufficient surface area (1.9 m^2g^{-1}), a high density of acidic sites (342.6 $\mu molg^{-1}$) and good porosity. The study then examined the effects of several reaction parameters, including temperature (100 °C – 140 °C), molar ratio of glycerol (Gly) to acetic acid (Ac) (1:3 w/v – 1:15 w/v), catalyst loading (0.1 g – 0.9 g) and reaction time (1 h – 5 h). The catalyst was then reused for five successive cycles to evaluate its reusability. The results demonstrated that a higher density of acidic sites, sufficient BET surface area and pore volume of the catalyst may enhance the formation of GLC and TA in the glycerol acetylation by facilitating the diffusion of the glycerol molecules on the catalyst's surface.

ARTICLE INFO

Article history:
Received: 24 June 2025
Accepted: 19 June 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.04

E-mail addresses:
sherafarisya@gmail.com (Shera Farisya Mohamad Rasid)
ernee@upm.edu.my (Ernee Noryana Muhamad)
taufiq@upm.edu.my (Yun Hin Taufiq-Yap)
zatul9793@gmail.com (Haizatul Izzatti Ridzuan)
lsling@utm.my (Siew Ling Lee)
irmawati@upm.edu.my (Irmawati Ramli)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

Therefore, optimal conditions for glycerol acetylation yielded a higher TA selectivity of 55 % and glycerol conversion of 97 % at a molar ratio of 1:6 w/v of Gly: Ac and a catalyst loading of 0.5 g at 130 °C for a 2 h reaction time. The reusability study showed that the 15SA catalyst can attain a higher GLC exceeding 97 % and a higher DA+TA selectivity of 85 % after five cycles. In contrast, the selectivity of TA reduced after five cycles due to the leaching of the acidic site from the catalyst, as proven by the EDX analysis.

Keywords: Acetins, acetylation, glycerol, parametric study, solid acid catalyst, sulfated-titania, triacetin

INTRODUCTION

The increasing population, rapid industrialisation, and growing energy have collectively precipitated an energy crisis associated with environmental deterioration, including ecological destruction, the scarcity of fossil fuels and pollution (Fauzi et al., 2025; Reinoso & Boldrini, 2020; Shikhaliyev et al., 2021). Therefore, utilising renewable energy as an alternative is essential for reducing environmental pollution while fostering social development and reducing reliance on conventional fossil fuels. Ladero (2021) stated that more than 280 billion US$ was invested worldwide in renewable energy from 2007 to 2017. These investments aim to attain a net-zero situation in greenhouse gas emissions (GHG). As a result, about 66 countries have committed to sourcing more than 50 % of their primary energy from renewable sources by 2050, including the Americas, Africa, the Asia-Pacific and the Indian Ocean (Dey et al., 2021). These developments have led to higher demand for biodiesel as a renewable transportation fuel, as it has been considered as one of the options to reduce consumption of fossil fuels and limit the generation of greenhouse gas emissions (Chilakamarry et al., 2021; Jose et al., 2023).

The transesterification of the biodiesel generates glycerol as a by-product. Hidayati et al. (2024a) and Malaika et al. (2023) stated that the production of glycerol increases linearly with biodiesel production. The ratio of glycerol to biodiesel is 1:10, yielding approximately 10 litres of glycerol for every 100 litres of biodiesel production (Reinoso & Boldrini, 2020). Therefore, excess production of glycerol has led to a significant price decline for several years and may become a challenge to the biodiesel industry since the unwanted glycerol cannot be disposed of in the environment (Mou et al., 2021; Shikhaliyev et al., 2021). Glycerol can be transformed into valuable chemicals with numerous applications, including in food, pharmaceutical, cosmetic, and plastic manufacturing. The low cost of glycerol has motivated researchers to valorise it via several processes, such as oxidation, transesterification, ammonization, condensation polymerisation, aromatisation, acetalisation, and acetylation reactions (Ebadi Pour et al., 2021; Hidayati et al., 2024b; Lee et al., 2026; Rathod et al., 2025; Singh & Gosu, 2025; Timofeeva et al., 2017). Among

these reactions, glycerol acetylation has gained significant attention because of its ability to yield three commercially important products: monoacetin (MA), diacetin (DA), and triacetin (TA).

The acetylation of glycerol with acetic anhydride or acetic acid involves a stepwise, consecutive reaction. In the first two reactions, MA and DA will form, followed by the formation of TA. All three products have useful industrial applications. For instance, MA is used in the production of explosives and smokeless powder, as a solvent in the leather tanning and dye industries and as a humectant in cosmetics (Beejapur et al., 2017; Kanimozhi et al., 2018). DA is used as a softening agent, building block in polyester, a solvent, a plasticiser, and an intermediate in the preparation of antidotes. (Appaturi et al., 2021; Beejapur et al., 2017; Kulkarni et al., 2020). TA is frequently used as an antiknock additive in petrol, an antimicrobial agent in pharmaceuticals and in cigarette filter production. In addition, the mixture of DA and TA can enhance the characteristics of petrodiesel. Kulkarni et al. (2020) stated that the incorporation of DA and TA can improve the properties of cold flow, viscosity, and biodiesel formulation quality. Furthermore, TA serves as a fuel additive by lowering the harmful emissions of unburned hydrocarbon (UHC), particulate matter (PM), nitrogen oxide (NOx), carbon monoxide (CO) and unregulated aldehyde (Jiang et al., 2019; Kalyani et al., 2023; Kulkarni et al., 2020; Ramos et al., 2025).

TA is the most valuable product in the acetylation of glycerol; however, its selectivity is usually lower than that of DA and MA. Therefore, various reaction conditions and catalysts have been investigated to improve TA formation. Some researchers have used acetic anhydride as the acetylating agent, achieving 100% TA selectivity at 60 °C within 20 min (Silva et al., 2010). Chandrakala et al. (2014) also obtained 100 % TA selectivity and 100 % GLC at 115 °C over a 30 min. Although acetylation of glycerol with acetic anhydride shows excellent performance in terms of TA selectivity and the GLC within a short period of time, acetic anhydride is more expensive, highly exothermic, and classified as a controlled substance in many countries due to its potential use as a precursor in narcotics production (Carpegiani et al., 2020; García-Martín et al., 2019; Mukhopadhyay et al., 2021; Okoye & Hameed, 2016; Okoye et al., 2017a). Therefore, acetic acid is preferred for the acetylation of glycerol as it is much cheaper and less toxic than acetic anhydride.

However, recent studies have shown that using acetic acid in glycerol acetylation without a catalyst has a low TA selectivity. For instance, Jiang et al. (2019) reported 0 % TA selectivity with 21 % GLC at 70 °C under atmospheric pressure. Similarly, Liu et al. (2019) obtained only 1 % TA selectivity with 48 % GLC at 125 °C over a 24 h reaction. Fauzi et al. (2025) achieved 0 % TA selectivity after testing at a reaction condition of 90 °C, 1:3 molar ratio of glycerol to acetic acid at 1 h reaction. Nonetheless, the GLC results were stated in their investigation of the blank reaction. Moreover, Kale et al. (2015) found that using 10 g glycerol and 39.16 g acetic acid at 105 °C for 4 h resulted in only 50 % of GLC and 0 % of TA selectivity.

The acid-catalysed mechanism has been widely used to facilitate the formation of TA and enhance GLC in the acetylation reaction. Heterogeneous catalysts, such as heteropolyacids, resins, zeolites, carbon catalysts and metal oxide catalysts, were widely employed compared to homogenous catalysts (Dalla Costa et al., 2017; Ferreira et al., 2009; Kulkarni et al., 2020; Magar et al., 2020; Nda-Umar et al., 2020a). The demand for heterogeneous catalysts in the acetylation process is increasing, as they can replace the liquid catalysts such as H_3PO_4, HCl, H_2SO_4, sulfonic acid, acidic ionic liquid, *para*-toluene, etc. Liquid catalysts have several drawbacks, including reactor corrosion, difficulty in separation, production of hazardous compounds, catalyst loss and poor catalyst reusability (Appaturi et al., 2021; Dalla Costa et al., 2017; Hidayati et al., 2024a; Karnjanakom et al., 2018; Khayoon et al., 2014). For instance, Ferreira et al. (2009) used dodecamolybdophosphoric acid encapsulated in USY zeolite (PMo1_NaUSY) as a heterogeneous acid catalyst for the glycerol acetylation with acetic acid and achieved 60 % GLC and a 4 % TA selectivity after 3 h of testing. Mou et al. (2021) attained 95 % GLC and 6 % TA selectivity by employing a 1:6 w/v molar ratio of Gly: Ac, 5.0 wt.% Amberlyst-15 catalyst, at a temperature of 90 °C for 1 h.

Moreover, Nda-Umar et al. (2020b) prepared a sulfonated carbon catalyst using a chemical method and subsequently conducted glycerol acetylation with a 1:6 w/v Gly: Ac molar ratio and a 0.5 g catalyst dosage for a reaction duration of 3 h. They found a remarkable catalytic activity with 56 % TA selectivity and 94 % GLC, attributed to the higher BET surface area and greater acidic capacity of the synthesised catalyst, as demonstrated by N_2 adsorption-desorption and NH_3-TPD analyses. Furthermore, Reddy et al. (2010) investigated the use of the metal oxides such as ZrO_2, TiO_2–ZrO_2, WOx/TiO_2–ZrO_2 and $MoOx/TiO_2$–ZrO_2 in glycerol acetylation with acetic acid, employing 1:6 w/v Gly:Ac molar ratio, 5 wt.% of catalyst dosage for 60 h reactions at 120 °C. They found that $MoOx/ TiO_2$–ZrO_2 achieved the highest GLC (~ 100 %) and TA selectivity (~80 %). Thus, it can be concluded that the carbon and metal oxide catalysts show an excellent performance in glycerol acetylation due to their high thermal stability, sufficient BET surface area and pore volume and good porosity, compared to zeolite, resin, and heteropolyacid catalysts. Nonetheless, the excessively long reaction time is unsuitable for industrial applications.

The synthesis of metal oxide catalysts for the acetylation of glycerol with acetic acid remains an active area of research to enhance TA selectivity. Testa et al. (2019) stated that the excellent catalytic performance in glycerol acetylation is related to the acid capability, strength and surface density of the acid sites of the catalyst. This is because the acetylation reaction can be catalysed not only by the Brønsted acid sites but also by Lewis acid sites (Fauzi et al., 2025). Therefore, incorporating the sulfate species onto the metal oxide catalysts could be an effective approach to increase the catalyst acidity and boost performance in glycerol acetylation.

Kulkarni et al. (2020) reported that the use of sulfated CeO_2-ZrO_2 catalyst resulted in a TA selectivity of 21 % due to high acid capacity (2.5 mmol), sufficient BET surface area (22 m^2g^{-1}) and good pore diameter (5.9 nm). In contrast, the non-sulfated CeO_2-ZrO_2 catalyst achieved only 0.5 % TA selectivity owing to its low acid site density. Moreover, Geetha et al. (2020) reported the same finding after examining the performance of sulfated-titania catalyst in a pseudo-five-component reaction. They found that the sulfate concentration of 5 wt.% in the preparation of sulfated titania via the sol gel method gave the highest yield (90 %) in the functionalization of piperidine, compared to non-sulfated titania (48 %) after a 30 min reaction under reflux conditions. The surface acidity of bare TiO_2 is 1600 µg, whereas the sulfated titania exhibits 2200 µg, as determined by the pyridine adsorption method. The sulfated-titania catalyst prepared via the sole gel method is also easily recyclable. Nevertheless, the sol-gel synthesis route requires precursor hydrolysis, which adds complexity to the preparation process.

The work conducted by Geetha et al. (2020) and Kulkarni et al. (2020) is similar to our previous study, which focused on the impact of the sulfuric acid concentration on the physicochemical characteristics of sulfated titania and evaluated its efficacy in GLC and acetin selectivity, particularly TA (Shera Farisya et al., 2021). We found that using sulfated TiO_2 with a 15 % concentration of sulfuric acid (15SA), prepared via the impregnation method, yielded a remarkable TA of 42 %, whereas non-sulfated titania exhibited only 16 % TA selectivity at 100 °C for a 2 h reaction. Despite its simple preparation, the 15SA catalyst demonstrated good catalytic performance owing to its high acid site density of 342.6 µmol g^{-1}, sufficient BET surface area of 1.9 m^2 g^{-1} and good porosity. However, the optimal reaction conditions for glycerol acetylation using 15SA catalyst were not determined in that study. Hence, the present work aims to further investigate the effect of reaction parameters, namely temperature, molar ratio of glycerol to acetic acid, catalyst loading and reaction time, on acetin selectivity, particularly TA selectivity and the GLC in glycerol acetylation over the 15SA catalyst using a systematic parametric approach. In addition, the reutilization of the 15SA sulphated titania catalyst was evaluated over five successive reaction cycles to determine its potential for repeated use in the catalytic reaction.

Materials

Sigma-Aldrich supplies acetonitrile (GC grade), 1,3-butanediol and titanium dioxide (98 %) while R&M chemicals provide sulfuric acid (95–98 %), glacial acetic acid (99 %) and anhydrous glycerol (99 %). No additional purification was required for any of the chemicals purchased.

Catalyst Preparation

The sulfated-titania catalyst was synthesised via the impregnation method as reported in our previous study (Shera Farisya et al., 2021). Briefly, about 15 % concentration of

the sulfuric acid was poured into 10 g of titanium dioxide. The sulfated-titania was then stirred for 24 h, filtered, and dried in an oven for 24 h at 100- 120 °C. Afterwards, the dried catalyst was ground, sieved (250 nm) and calcined for 2 h at a heating rate of 5 °Cmin $^{-1}$ under a CO_2 environment. Finally, the sulfated-titania catalyst was then stored in a tight bottle and denoted as 15SA.

Catalyst Characterisation

The thermal stability of the sulfated-titania catalyst was assessed using a Mettler Toledo 851e thermal analysis instrument. The crystalline phases of non-sulfated and sulfated-titania catalysts were evaluated using an X-ray diffractometer with a Shimadzu model XRD 6000. Cu Kα radiation source of 30 kW and electricity of 30 Ma was utilised, and the diffraction pattern was obtained within a scanning range of 2θ = 10–80° at a scanning speed of 4° min^{-1} using a wavelength (λ) of 1.54. The textural properties of the synthesised catalysts were determined using the nitrogen physisorption method with a Micrometrics analyser. Before the determination, the sample was degassed for 8 h at 120 °C to remove adsorbed molecules from the pores. To confirm the existence of the surface functional groups in the sulphated-titania catalyst, X-ray photoelectron spectroscopy (XPS) analysis was carried out using a PHI Quantera II instrument equipped with a monochromatic Al Kα X-ray source (photon energy, hv =1486.6 eV). The analysis was operated at 25.6 W, and data acquisition was performed using a pressure chamber below 10^{-8} mbar equipped with a charge neutralisation system. The Smartsoft-XPS and MULTIPAK ECSA software were employed to deconvolute the acquired data into multiple peak and the charging effect caused by surface contamination of the analysed sample was corrected using the carbon binding energy scale (C 1s peak). A Thermo Finnigan TPDRO series of NH_3-TPD coupled with a thermal conductivity detector (TCD) was employed to determine total acid density and the distribution of acid sites of the synthesised catalysts. The surface morphology and elemental components of the sulfated-titania catalyst were determined via SEM JEOL, JSM6400 Scanning Electron Microscopy attached with Energy Dispersive X-ray Spectroscopy (SEM-EDX). The conditions of TGA, XRD, BET, NH_3-TPD and SEM-EDX analysis were outlined in the supplementary data (S1).

Catalytic Test in the Acetylation of Glycerol

Parametric study of reaction parameter for glycerol over sulfated-titania catalyst (15SA) was investigated in this study, as shown in Figure 1. Briefly, about 0.5 g of 15SA catalyst and a 1:6 w/v molar ratio Gly: Ac (5 g of glycerol and 18.6 ml of acetic acid) were introduced into a 250 ml round-bottom flask. The detailed calculation for the molar ratio of glycerol to acetic acid is provided in supplementary data (S2). A reflux condenser was attached to the flask, which was then immersed in a silicone oil bath. Before starting the reaction, the

magnetic rod was added to the mixture. The reaction was conducted at 100 °C for 2 h of reaction time and then stirred at 450 rpm under reflux conditions. The product was then cooled and centrifuged at 3000 rpm for 10 min following the completion of the reaction. The GLC, MA, DA and TA components were further analysed using GC-FID (Agilent 7890A, FID, DB Wax 30 m x 0.25 mm ID x 0.25 µm). Approximately 1 µL of the final sample was injected into the GC-FID instrument, which was initially set at a temperature of 80 °C for 30 min. The temperature was then increased to 260 °C at a temperature ramping of 10 °C min^{-1} and then to 300 °C with a temperature ramping of 30 °C min^{-1}. The total analysis time was 22.3 min with He serving as the carrier gas at a rate of 1.3 mL min^{-1}. The detailed procedure for GC-FID, the equation for glycerol conversion and the selectivity of acetins are outlined in the supplementary data (S3).

Furthermore, to examine the influence of reaction temperature in the glycerol acetylation, the method was performed at varying temperatures of 110 °C, 120 °C, 130 °C and 140 °C while maintaining constant other parameters. The temperature yielding the highest TA was chosen for further investigation on the effect of the molar ratios of Gly: Ac (1:3 w/v, 1:6 w/v, 1:9 w/v, 1:12 w/v and 1:15 w/v), catalyst loading (0.1 g, 0.3 g, 0.5 g, 0.7 g and 0.9 g) and reaction durations (1 h, 2 h, 3 h, 4 h and 5 h) to determine the parametric study for glycerol acetylation over the 15SA catalyst using one-factor-at-one time (OFAT). Finally, the reusability of the 15SA catalyst was evaluated at the ideal glycerol acetylation conditions, as shown in Figure 2. After completing each cycle, the spent catalyst underwent centrifugation, filtration and multiple washes with acetone to eliminate organic impurities. The catalyst was subsequently placed in an oven at 100 – 120 °C for a duration of 2 h. The catalyst was used up to the fifth cycle, and the leaching of the active site was determined using EDX analysis.

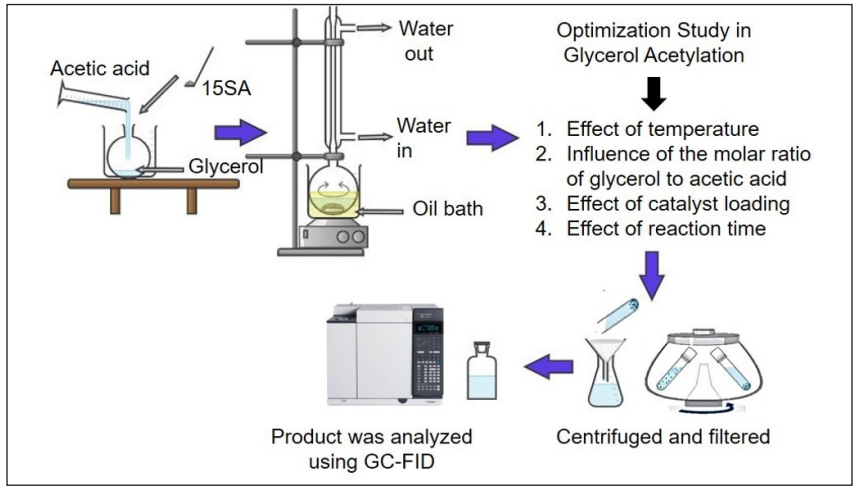


Figure 1. Experimental setup for acetylation of glycerol with acetic acid using a reflux condenser

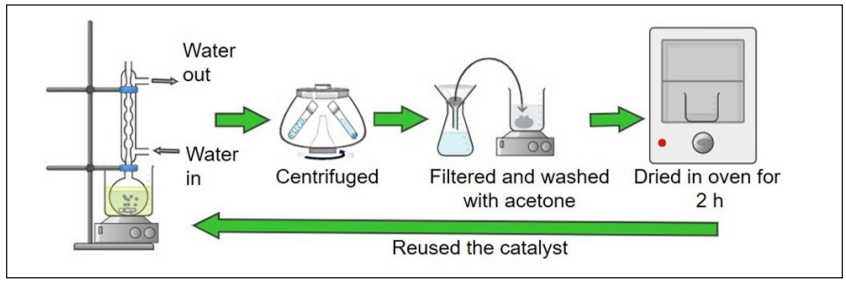


Figure 2. The procedure for the reusability study with the 15SA catalyst

RESULTS AND DISCUSSIONS

Physicochemical Properties of the Sulfated Titania Catalyst

15SA catalyst was characterised using X-ray diffractometer (XRD), Brunauer-Emmett-Teller (BET), X-ray photoelectron spectroscopy (XPS), Temperature-programmed desorption with ammonia (NH_3-TPD) and Scanning electron microscopy with Energy Dispersive X-ray (SEM-EDX). Thus, in this current study, the 15SA catalyst was further used for optimising the TA and GLC. Finally, the reusability of the catalyst was identified at optimal reaction conditions.

Crystallinity and Phase Determination Using XRD

Figure 3 illustrates the X-ray diffraction patterns for 15SA sulfated-titania catalysts. It was found that the 15SA catalyst exhibits anatase and $(TiO)(SO_4)(H_2O)$ peaks. The anatase peaks were detected at $2\theta = 38.64°$, 53.92°, 62.22°, 62.74° and 75.12° (JCPDS-ICDD card: 96-900-8214). According to the X'pert High Score Plus Software, the anatase peak has a tetragonal structure with a chemical formula of Ti_4O_8. In addition, the peaks detected at $2\theta = 25.34°$, 36.94°, 37.78°, 55.08°, 68.68°, 70.29°, 76.14° correspond to the $(TiO)(SO_4)(H_2O)$ peak with an orthorhombic structure (JCPDS-ICDD card: 96-153-0152). Therefore, the detection of $(TiO)(SO_4)(H_2O)$ peaks in the 15SA confirmed that the sulfate species are well distributed on the surface of the catalyst. Besides, Wijaya et al. (2021) and Yang et al. (2019) discovered that the sulfation of the catalyst leads to dissolution of the titania and reduces the peak intensity as compared to that of the untreated titania catalyst. This phenomenon promotes the generation of Ti-O-S and thereby renders the phase more amorphous (Pratika et al., 2021). As a result, a lower crystallinity of the $(TiO)(SO_4)(H_2O)$ with 39.27 nm was found at the peak of 25.3 °, whereas TiO_2 anatase exhibit crystallite size of 54.02 nm as calculated using the Debye-Scherrer formula. Furthermore, X'Pert High Score Plus software shows that the average particle size of 15SA is smaller than that of TiO_2 anatase with 36 nm and 80 nm, respectively. Jenie et al. (2010) and Parida et al. (2008) reported that the impregnation of the titania surface with sulfation is facilitated by the different molecular

size of sulfate ions (0.242 nm) and TiO_2 (220 nm). This impregnation inhibits the phase transformation from anatase to rutile, resulting in smaller crystallites of sulfated titania.

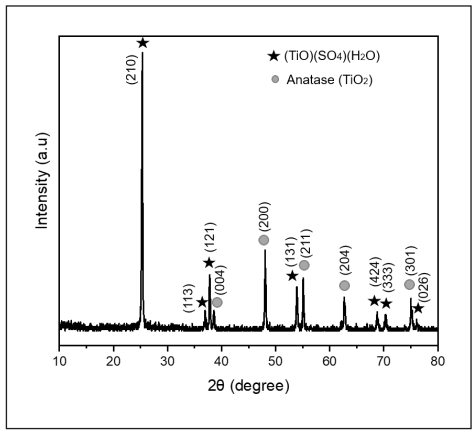


Figure 3. XRD profile for 15SA catalyst

Thermal Stability Determination Using TGA-DTG Analysis

The thermal stability and decomposition of the 15SA catalyst were investigated using the TGA analysis with a constant supply of N_2 gas at 50 $mLmin^{-1}$. Figure 4 illustrates three distinct stages of weight loss that were detected following thermal analysis. The initial decomposition began at the onset of the heating process, reaching 200 °C. This decomposition is attributed to the removal of water from the surface of the catalyst. The weight loss in this lower temperature region was about 2 ± 0.5 %. The subsequent decomposition was observed in a temperature range of 450 °C to 700 °C with a weight loss of 11± 0.5 %. This stage was associated with the decomposition of sulfate species. This finding agrees with the work reported by Noda et al. (2003), indicating that the decomposition between 450 °C and 700 °C was attributable to the sulfate bound to the oxide. Additionally, a less pronounced peak with a weight loss of 2 ± 0.5 % was detected between 800 °C and 950 °C in the DTG. This indicates that a high concentration of sulfuric acid (15 %) impregnated with titania requires more heat to break the intermolecular bond of the sulfate. Thus, at a higher temperature, SO_4^{2-} in the sulfated sample will escape as SO_2. This finding is correlated with Enríquez et al. (2013) and Liu et al. (2020).

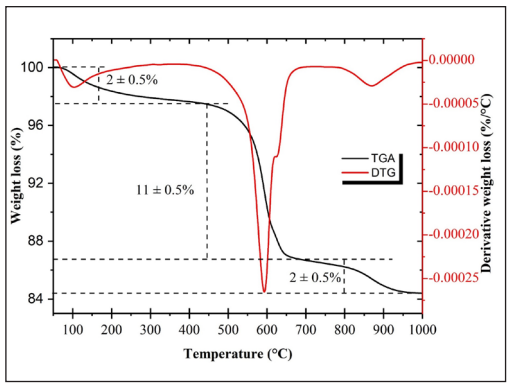


Figure 4. TGA-DTG profile for 15SA catalyst

Surface Area and Pore Size Distribution

Table 1 illustrates the BET surface area, BJH pore volume and pore size of the sulfated-titania (15SA) and non-sulfated-titania (TiO_2). The BET surface area of TiO_2 and 15SA decreased from 11.8 m^2g^{-1} to 1.9 m^2g^{-1} after the incorporation of SO_4^{2-} into the titania. This reduction might be due to the plugging of the pores and surface coverage by sulfate species (Lertna et al., 2022). This finding is consistent with work reported by Tomer and Biswas (2020),

which demonstrated that the sulfation of TiO_2 results in a lower BET surface area. A similar trend of BET surface area is observed in the pore volume, which decreases from 0.0029 cm^3g^{-1} for TiO_2 to 0.00003 cm^3g^{-1} for 15SA. Pore blockage could have occurred due to the increasing of the atomic radius of the sulfated-titania catalyst due to the replacement of one atom in the O-Ti-O framework by SO_4^{2-} ion after impregnation on TiO_2 (Tomer & Biswas, 2020). This observation was aligned with Zhou et al. (2015), who reported that agglomeration on the nano-sized $Ti(SO_4)$ occurred as more sulfate species were loaded on the surface, hence diminishing the surface area and porosity of the catalyst. Meanwhile, the pore diameter remains relatively constant for both catalysts, obtaining 5.2 nm for 15SA and 5.4 nm for non-sulfated TiO_2. This finding suggests that sulfation does not change the intrinsic structure of the pores but increases the deposition of sulfate groups, which might enhance the acidity of the sulfated catalysts. Additionally, both 15SA and TiO_2 catalysts can accommodate the glycerol molecule (1 nm) for the reaction to occur on the pore surface, thereby potentially enhancing the conversion of glycerol in glycerol acetylation with acetic acid.

Table 1
BET surface area, BJH pore volume and pore diameter for sulfated and non-sulfated catalysts

Catalyst	BET Surface Area (m^2g^{-1})	BJH Pore Volume (cm^3g^{-1})	Pore Diameter (nm)
TiO_2	11.8	0.0029	5.4
15SA	1.9	0.0003	5.2

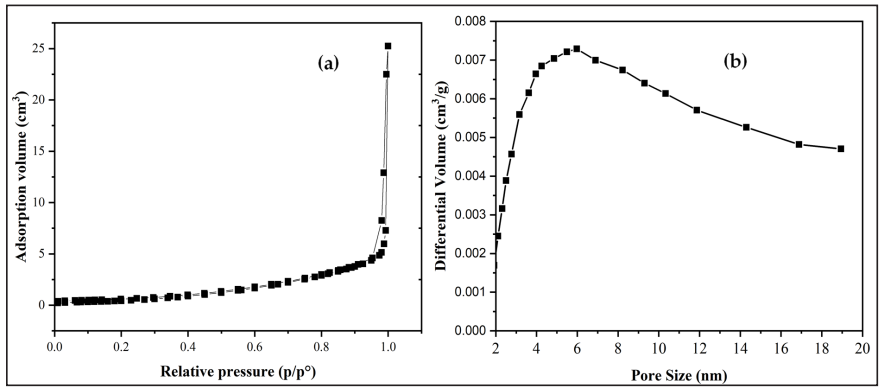


Figure 5. (a) N_2 adsorption- desorption isotherms (b) and pore size distribution of 15SA catalyst

Figure 5 reveals that the nitrogen adsorption-desorption isotherm of the sulfated catalyst is categorised as type IV, exhibiting a sharp descent at a relative pressure of 0.9-1.0 p/p°. These findings confirm that the 15SA catalyst exhibits a regular mesoporous structure. These results align with those of Afshar et al. (2015), Huang et al. (2017), and Tomer and Biswas

(2020). Moreover, according to the IUPAC classification, 15SA catalysts have prominent H1-type hysteresis loops due to variations in absorption and desorption. Furthermore, the pore size distribution, which ranges from 2 to 20 nm, provided additional evidence for the mesoporous nature of 15SA.

Surface Analysis Using XPS

The high-resolution XPS spectra scan of the 15SA catalyst is displayed in Figure 6. From this analysis, the oxidation state of sulfur and the bonding characteristic of the sulfated catalyst can be determined. Three intense XPS peaks were observed at binding energies of 458 eV, 531 eV and 169 eV, which are assigned to Ti 2p, O 1s, and S 2p, respectively, as illustrated in Figure 6a. In addition, Figure 6b revealed that the Ti 2p XPS spectra exhibit two significant peaks corresponding to Ti $2p_{3/2}$ and Ti $2p_{1/2}$ at 459.6 eV and 465.3 eV, respectively. These characteristic peaks confirm the presence of the anatase phase of titania, consistent with findings from the XRD. In addition, the Ti in 15SA is mainly present in the Ti^{4+} chemical state with a spin-orbital splitting value of 5.7 eV (Shi & Li, 2014; Tomer & Biswas, 2020).

Figure 6c shows the O 1s spectra, which can be deconvoluted into three distinct peaks. The lower bonding energy peak at 530.8 eV is associated with the Ti-O-S bonds, indicating the integration of sulfate groups into the titania framework. Meanwhile, the O-S bonds in the sulfate groups were observed at the elevated peak of 532.5 eV

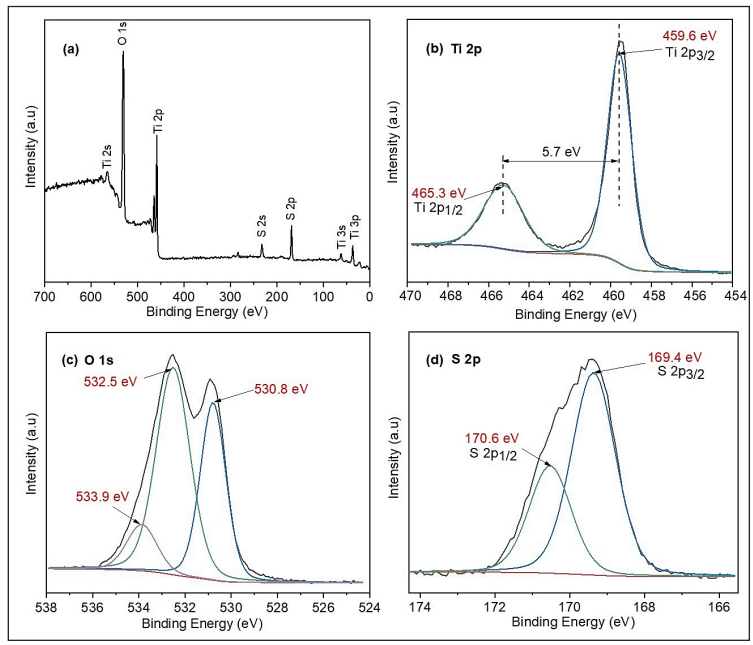


Figure 6. (a) Survey scan XPS spectra, (b) Ti 2p, (c) O 1s, and (d) S 2p of 15SA catalyst

(Dabbawala et al., 2018; Gardy et al., 2016; Tomer & Biswas, 2020). Additionally, the minor peak at 533.9 eV is associated with the O-H bond from the chemically bound water molecules on the catalyst surface (Patil et al., 2019; Zhao et al., 2014). In Figure 6d, the S 2p spectra are depicted with two deconvoluted peaks at photoelectron binding energies of 170.6 eV and 169.4 eV, which correspond to S $2p_{1/2}$ and S $2p_{3/2,}$ respectively. These peaks confirm the existence of SO_4^{2-} species on the surface of the 15SA catalyst. Similar peak assignments at comparable binding energies for sulfate groups have been reported in studies on sulfated metal oxides (Dabbawala et al., 2018; Shi & Li, 2014; Tomer & Biswas, 2020).

Analysis of Total Acid Sites Density Using NH_3-TPD

Ammonia-Temperature Programmed Desorption (NH_3-TPD) analysis was utilised to measure the comprehensive acid strength of the synthesised catalyst (Azri et al., 2021; Kanimozhi et al., 2018). The NH_3−TPD profile for the 15SA catalyst is displayed in Figure 7. Wang et al. (2015) and Yang et al. (2019) mentioned that the acid strength of the surfaces is theoretically correlated with the NH_3 desorption temperature. The desorption peaks at lower temperatures correspond to the physically bound NH_3, while the peaks at higher temperatures indicate chemically bound NH_3 (Dantas et al., 2017). The adsorption of NH_3 is correlated with the interaction force between ammonia and acidic sites, which are categorised into weak acid sites (T_{max} =10-250 °C), moderate acid sites (T_{max} = 250-500 °C), and strong acid sites (T_{max}>500 °C) (Azri et al., 2021; Srivastava et al., 2017). In addition, Figure 7 shows that the maximum desorption peak for the 15SA catalyst was found at 608 °C, with a total amount of adsorbed ammonia of 342.6 μmolg^{-1}. This result proves that the 15SA catalyst only exhibits a strong acid site density as the maximal desorption temperature is recorded at a temperature exceeding 500 °C. This finding aligns with the TGA-DTG profile, which indicates that the loss of sulfate species occurred within a temperature range of 450 °C to 900 °C. Furthermore, a larger desorption temperature of the 15SA catalyst suggests that the number of available active sites is higher, thereby increasing the acidic strength, regardless of Brønsted and Lewis acidity (Neto et al., 2020; Pratika et al., 2023). According to the Lewis theory, the NH_3 donates a pair of electrons and attaches to the Lewis acid centre (Ti^+) to form a covalent bond, while the Brønsted-Lowry theory indicates that NH_3 acts as a base by accepting a proton (H^+) from the Ti-OH molecules and forming an ammonium ion (NH_4^+) (Hu et al., 2015; Ropero-Vega et al., 2010). As a result, the sulfate species at 15SA are strongly bound to the titania surface, resulting in a higher total amount of adsorbed ammonia as displayed in Figure 7. Therefore, the presence of a higher acidic sites, strong acidic strength, sufficient pore volume and BET surface area in the catalyst, proven by NH_3-TPD, TGA-DTG and BET analyses, potentially enhanced the GLC and TA formation in the glycerol acetylation by facilitating the diffusion of the glycerol molecules on the catalyst's surface.

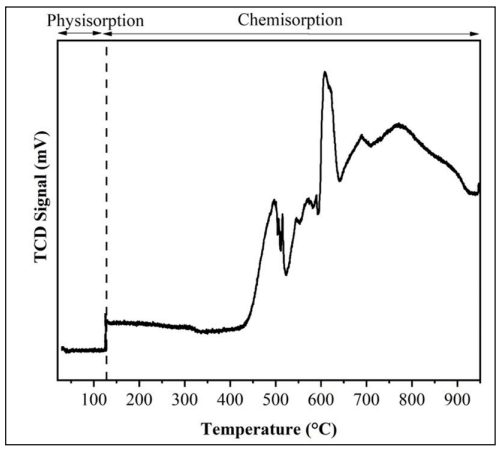


Figure 7. NH_3-TPD profile for 15SA catalyst

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

The element composition and the surface structure of the 15SA catalyst were analysed using SEM-EDX analysis. According to SEM analysis, the spherical particles with a crumb-like structure were found to be agglomerated, as shown in Figure 8, with an average size of 1 µm (from the scale bar). These findings align with the BET analysis, which indicates that the incorporation of a higher concentration of sulfuric acid on the titania surface leads to the blockage of the sulfate species, thereby resulting in a lower BET surface area. Tomer and Biswas (2020) also reported the same finding, where the agglomeration of particles resulted from the electrostatic interaction between the sulfate ions on the surface of the sulfated-titania catalyst. Furthermore, the EDX analysis indicates that the peaks of 15SA were detected at 0.5 keV, 0.6 keV, 2.3 keV, and 0.4 keV, respectively. This indicates that the 15SA catalyst is composed of titania (Ti), oxygen (O), and sulfur (S) in proportions of 33 %, 56 %, and 6 %, respectively. Therefore, the detection of S in this analysis confirms that the sulfation was successfully carried out.

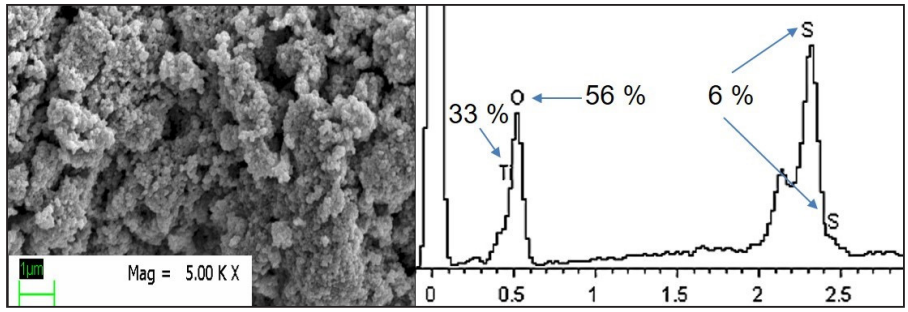


Figure 8. SEM micrograph for fresh 15SA catalyst

Systematic Investigation on the Effect of Temperature in Glycerol Acetylation

A series of reactions was conducted at temperatures ranging from 100 °C to 140 ° to examine the impact of reaction temperature (T) on GLC and the formation of acetins, particularly TA. The tests were carried out using a fixed 0.5 g dosage of the catalyst (CL) and a molar ratio of 1:6 w/v Gly: Ac (MR) for a reaction duration (t) of 2 h. According to Figure 9, the GLC remained consistently high (98±1 %) between 100 °C and 120 °C. These results

may be attributed to the near-complete consumption of glycerol, the limiting agent, under the excess acetic acid conditions imposed by the 1:6 w/v molar ratio, whereby surplus acetic acid cannot be completely utilised in the reaction (Hidayati et al., 2024a). However, the GLC marginally decreased from 97 % and 92 % when the temperature was raised to 130 °C and 140 °C, respectively. This reduction is consistent with the partial evaporation of acetic acid (boiling point: 118 °C, which reduces the effective concentration of acetic acid available under reflux conditions, rather than indicating any loss of catalyst activity (Hidayati et al., 2024a; Mufrodi et al., 2014). Overall, the GLC remained high across all tested conditions, exceeding 97 % except at 140 °C, where more acetic acid transforms into vapour within the 2 h reaction period, as observed by Kanimozi et al., (2018).

Moreover, Figure 9 shows that MA selectivity declined from 9 % to 5 % as the temperature increased from 100 °C to 130 °C, while DA selectivity followed the same trend, decreasing from 49 % at 100 °C to 41 % at 130 °C. In contrast, TA selectivity increased progressively from 42 % at 100 °C to a maximum of 55 % at 130 °C. This observation is consistent with the findings of Malaika et al. (2021), who reported that DA and TA are the dominant products at higher temperatures, while MA prevails at lower temperatures. This trend reflects the endothermic and thermodynamically controlled nature of the sequential acetylation pathway from MA to DA to TA. As reported by Okoye et al. (2017b), glycerol acetylation involves three consecutive reaction steps in which MA form first, followed by DA and TA, and the reaction is also highly dependent on temperature owing to its endothermic character. Each subsequent acetylation step required progressively more activation energy, and the third step (DA → TA) is the most thermodynamically demanding. At 130 °C, the 15SA catalyst, which possesses exclusively strong acid sites as evidenced by NH_3-TPD (T_{max} = 608 °C, total acidity = 342.6 μmol g^{-1}), provides a thermally compatible environment where the elevated temperature supplies sufficient energy to surmount the activation barrier for TA formation, while the active SO_4^{2-} species remain thermally stable well below their decomposition onset at 450 °C, as confirmed by TGA analysis. Furthermore, based on the Arrhenius equation, a higher reaction temperature results in a lower activation energy required for the conversion of reactant to products, increasing the rate constant. The higher frequency of molecular collisions at elevated temperature promotes better interaction between the reactant molecules and the strong Lewis (Ti^{4+}) and Brønsted (Ti-OH) acid sites of 15SA, thereby enhancing the formation of DA and TA (Appaturi et al., 2023).

However, a further increase in temperature to 140 °C resulted in a decline in TA selectivity to 38 % and DA selectivity to 48 %, as shown in Figure 9. This observation is attributed to the reduced availability of acetic acid due to its volatilisation at this temperature, which limits the progression of the reaction towards TA formation (Hidayati et al., 2024a). Therefore, it can be concluded that the 130 °C optimum represents a

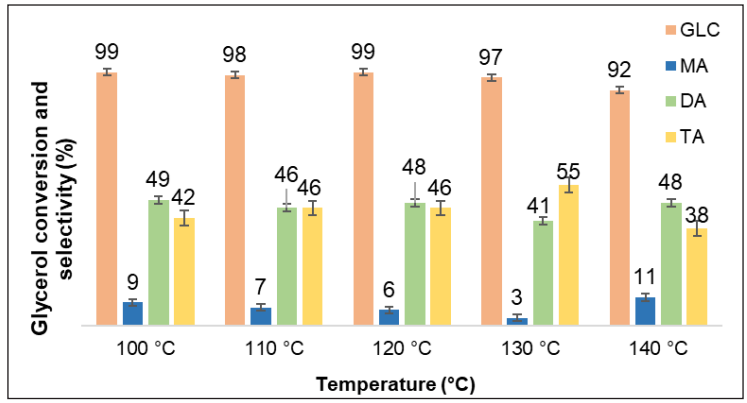


Figure 9. The effect of temperature (T) on the GLC and acetins selectivity. (Reaction condition: MR = 1:6 w/v glycerol: acetic acid, CL = 0.5 g, t = 2 h, T = 100-140 °C)

kinetic-thermodynamic balance specific to this catalyst system: sufficiently high to drive the endothermic DA to TA step at a competitive rate, yet below the threshold at which reagent depletion becomes significant. Despite the slight reduction in the GLC from 98 % at 120 °C to 97 % at 130 °C, the main objective of this study is to achieve higher TA selectivity. Thus, 130 °C was selected as the optimal reaction temperature, as it yields the highest TA selectivity of 55 %.

Influence of Molar Ratio of Glycerol to Acetic Acid in Glycerol Acetylation

The influence of the glycerol to acetic acid molar ratio on GLC and acetin selectivity is shown in Figure 10. In this study, the Gly: Ac molar ratio was varied at 1:3 w/v, 1:6 w/v, 1:9 w/v, 1:12 w/v and 1:15 w/v, while the other conditions were fixed at 130 °C, 0.5 g catalyst dosage for 2 h of reaction time.

Figure 10 shows that increasing the Gly: Ac molar ratio from 1:3 w/v to 1:15 w/v resulted in excellent glycerol conversion exceeding 90 % under all conditions, confirming that acetic acid availability does not limit GLC in these experiments. The shift in product selectivity is best explained by the stoichiometry of the stepwise acetylation reaction and the surface chemistry of the 15SA catalyst. At the 1:3 w/v molar ratio, only three moles of acetic acid are available per mole of glycerol, representing the theoretical minimum for full conversion to TA. However, the absence of excess acetic acid means the equilibria cannot be effectively driven forward, and MA remains the dominant product at 24 % selectivity, as only the first acetylation step can be completed before acetic acid is consumed. This observation aligns with the findings of Appaturi et al. (2021), who reported that a limited amount of acetic acid at a 1:3 w/v molar ratio restricts reagent availability in glycerol acetylation, favouring the formation of the mono-substituted product, MA.

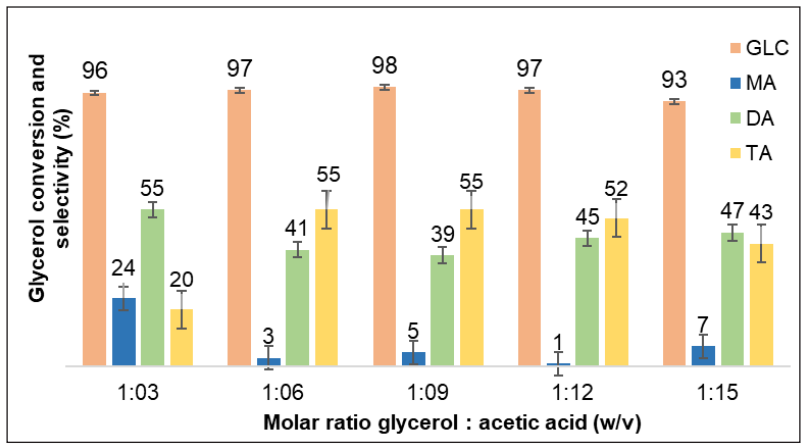


Figure 10. The effect of molar ratio of glycerol to acetic acid (MR) on the GLC and acetins selectivity. (Reaction condition: MR = 1:3-1:15 w/v Gly:Ac, CL = 0.5 g, t = 2 h, T = 130 °C)

TA selectivity increased significantly from 20 % to 55 % as the Gly: Ac molar ratio was raised from 1:3 w/v to 1:6 w/v. Increasing the ratio to 1:6 w/v creates a two-fold excess of acetic acid on the 15SA surface. First, it ensures that the Brønsted acid sites (Ti–OH), confirmed by the O 1s XPS peak at 533.9 eV, continuously protonate acetic acid molecules to generate acylium ions, sustaining the forward reaction. Second, the higher acetic acid concentration reduces the interfacial viscosity of the reaction mixture as reported by Fronza et al. (2025), improving mass transfer to reactants to the mesoporous catalyst surface. The 5.2 nm pore diameter of 15SA accommodates glycerol molecules (~1 nm) within the pore channels, enabling repeated sequential encounters with active sites and promoting the full MA to DA TO ta cascade. Acetic acid may also act as a Brønsted acid to further drive the forward reaction and enhance TA formation (Abida et al., 2020; da Silva et al., 2020). These effects explain why TA selectivity peaks at 55 % under the 1:6 w/v molar ratio.

GLC and TA selectivity declined slightly when the molar ratio was further increased to 1:12 w/v and 1:15 w/v, yielding 97 % GLC with 52 % TA selectivity at 1:12 w/v, and 93 % GLC with 43 % TA selectivity at 1:15 w/v. For the 15SA catalyst, excess acetic acid beyond 1:9 w/v is proposed to compete with glycerol intermediates for the limited number of strong Lewis's acid sites (Ti^{4+} centres), absorbing non-productively and occupying sites that would otherwise catalyse the final DA → TA acetylation step. This interpretation is consistent with the relatively low BET surface area of 15SA (1.9 $m^2 g^{-1}$), which limits the number of accessible active sites. A large excess of acetic acid, therefore, disproportionately competes for these sites, as similarly reported by Okoye et al. (2017b), who found that excess acetic acid in the catalytic reaction led to a lower TA formation, likely due to mass transfer resistance limiting contact between glycerol and the catalyst surface. Therefore, the optimal Gly: Ac molar ratio was achieved at 1:6 w/v, representing the best balance

between reagent driving force and active site accessibility on the 15SA surface, yielding maximum TA selectivity of 55 % and GLC of 97%.

Effect of Catalyst Loading in Glycerol Acetylation

The effect of catalyst loading on the GLC and selectivity of MA, DA and TA was evaluated under fixed conditions: a molar ratio of 1:6 w/v Gly: Ac at 130 °C for 2 h. The catalyst loading was varied from 0.1 g to 0.9 g, and the results are presented in Figure 11.

As shown in Figure 11, glycerol conversion increased from 91 % to 98 % as the catalyst loading rose from 0.1 g to 0.3 g. The glycerol conversion showed no significant change when the catalyst loading rose to 0.5 g and 0.7 g, reaching 97 % for both, and slightly improved to 100 % at 0.9 g catalyst loading. This trend may be due to the increased availability of active sites, both Lewis (Ti^{4+}) or Brønsted (Ti–OH), as the catalyst dosage increases, which promotes higher glycerol conversion (Hidayati et al., 2024a).

In addition, TA selectivity increased from 38 % to 55 % as the catalyst dosage was raised from 0.1 g to 0.5 g, while MA selectivity declined from 11 % to 3 % over the same range. This response reveals a nuanced relationship between active site density and product distribution, specific to the strong-acid, low-surface-area character of the 15SA catalyst. At 0.1 g, the catalyst provides insufficient strong acid sites to drive the thermodynamically demanding DA to TA step at a competitive rate within 2 h window, resulting in DA (45 %) and MA (11%) dominating the product distribution. According to the mechanism reported by Spataru et al. (2021), the carbonyl group of acetic acid combines with the protons from the catalyst to form an acylium ion, which is then attacked by the hydroxyl groups of produce MA first, followed by DA and finally TA. As the loading increases to 0.5 g, the cumulative density of SO_4^{2-}–Ti^{4+} Lewis acid sites and Ti–OH Brønsted acid sites becomes sufficient to complete all three sequential acetylation steps within the reaction window. TA selectivity peaks at 55 % with 0,5 g, representing the threshold loading at which the strong acid site population is sufficient to drive the full MA to DA to TA cascade. Appaturi et al. (2021) similarly reported that the acetylation reaction over solid catalyst mainly occurs as a surface reaction, where active sites enhance the interaction between acylium ions and hydroxyl ions, facilitating the formation of higher DA and TA (Dhawan et al., 2021; Fronza et al., 2025; Shikhaliyev et al., 2021).

Moreover, Figure 11 shows that further increasing the catalyst dosage to 0.7 g and 0.9 g led to a slight decrease in TA selectivity to 43 % to 39 %, respectively, while MA and DA selectivities increased. At 0.7 g, MA selectivity increased to 10 % while DA selectivity rose to 45 %. At 0.9 g, MA and DA selectivities reached 10 % and 51 %, respectively. This decline in TA selectivity at loadings above 0.5 g is a consequence of poor bulk mixing rather than any reduction in intrinsic catalyst activity. Beyond 0.5 g, the increased solid mass creates a suspension too dense for efficient agitation at 450 rpm within the 250 ml

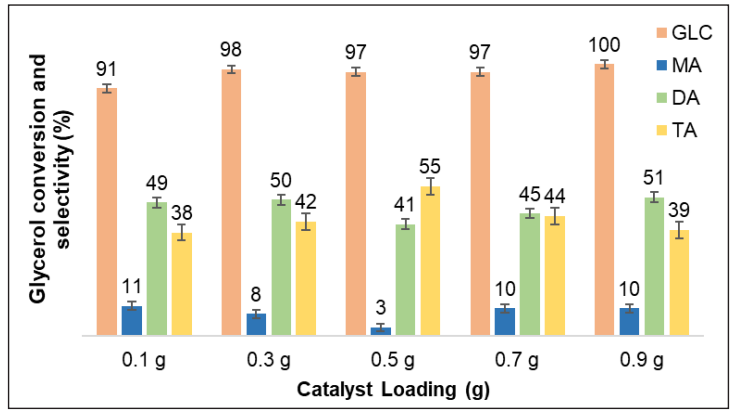


Figure 11. The effect catalyst loading (CL) on the GLC and acetins selectivity. (Reaction condition: MR= 1:6 w/v Gly:Ac, t = 2 h, T = 130 °C and CL = 0.1 - 0.9 g)

round-bottom flask used, increasing the inter-particle diffusion path length for glycerol and acylium intermediates and effectively confining reactions to the outer surfaces of catalyst agglomerates. This suppresses the final DA → TA step, which requires the most intensive substrate-surface contact, and is consistent with the SEM observations showing agglomeration in suspension. A similar observation was reported by Appaturi et al. (2021), who found that excessive catalyst dosage leads to inefficient stirring and mixing, resulting in catalyst accumulation, increased mass transfer resistance and hindered TA formation. Consequently, the ideal dosage catalyst for glycerol acetylation over 15SA is 0.5 g, as it achieves the maximum TA selectivity of 55 %.

Effect of Time in Glycerol Acetylation

The effect of reaction time on TA selectivity was investigated by varying the duration from 1 h to 5 h while keeping the other parameters constant at 130 °C, a 1:6 w/v Gly: Ac molar ratio, and 0.5 g of 15SA catalyst, as shown in Figure 12. GLC increased steadily from 97 % to approximately 100 % as the reaction time increased from 1 h to 3 h.

Figure 12 also shows that the TA selectivity rose from 35 % at 1 h to 55 % at 2 h, reaching its maximum at this point, indicating that the reaction achieves its optimal selectivity equilibrium at 2 h. The temporal evolution of acetin selectivity directly reflects the kinetics of the stepwise acetylation on the 15SA surface. In the first hour, the strong acid sites of 15SA rapidly activate acetic acid into acylium ions and drive the initial acetylation of glycerol to produce MA, which constitutes most of the product at t = 1 h. The relatively low TA accumulation at 1 h (35 %) indicates that the MA to DA and DA to TA steps, being less thermodynamically favourable and requiring repeated substrate-surface interaction, have not yet reached completion. This demonstrates that the acetylation of glycerol proceeds

in a stepwise manner, consistent with the findings of Abida et al. (2020) and Mou et al. (2021), who reported that TA selectivity increases progressively with reaction time as MA and DA are sequentially converted through further acetylation with excess acetic acid.

The peak in TA selectivity at 2 h represents the kinetic optimum for this catalyst-substrate system. By this stage, most MA and DA intermediates have been further acetylated on the strong Lewis and Brønsted acid of 15SA, and the reaction has not yet entered the deactivation phase. Notably, the 2 h optimum is shorter than the reported for the catalyst with lower acid strength, which is a direct result of the exclusively strong acid site character of 15SA (T_{max} = 608 °C). This accelerates the rate-limiting third acetylation step and brings the reaction to its selectivity maximum earlier than a catalyst relying on weak or moderate acid sites.

Figure 12 further shows that GLC decreased slightly to 98 % and TA selectivity dropped to 34 % when the reaction was extended to 5 h. This is attributed to the progressive deactivation of the 15SA catalyst at longer reaction times. Water is produced as a by-product of each esterification step, with three moles generated per mole of TA formed. Over-extended reaction times, this accumulated water promotes the hydrolysis of the Ti^{4+}–SO_4^{2-} bond, consistent with the XPS-confirmed Ti-O-S bonding at 530.8 eV, which is susceptible to nucleophilic attack by water. This progressively converts coordinated sulfate species into free sulfuric acid and reduces the density of strong Lewis acid sites (Kale et al., 2015; Shi & Li, 2014). This resulting loss of Lewis activity shifts the product distribution back towards DA and MA, which require less acidic activation. This interpretation aligns with the reusability data in Section 6, where EDX analysis shows the sulfur content declining from 6 % in the fresh catalyst to 2 % in the spent catalyst after five cycles, confirming progressive sulfate leaching as a time- and water- dependent phenomenon. Overall, the optimal reaction time for the 15SA catalyst is 2 h, yielding the highest TA selectivity of 55 % TA alongside 41 % DA, 3 % MA and 97 % GLC.

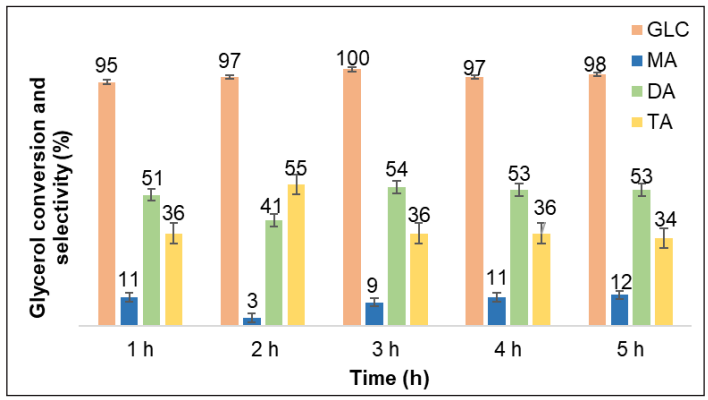


Figure 12. The effect of reaction duration (t) on the GLC and acetins selectivity. (Reaction condition: MR = 1:6 w/v Gly:Ac, CL = 0.5 g, T = 130 °C and t = 1-5 h)

Holistic Mechanistic Interpretation of the Optimal TA selectivity

The four reaction parameters investigated in this study are not independent in their influence on TA selectivity; rather, they interact through a common mechanistic framework governed by the unique physicochemical properties of the 15SA catalyst. The results obtained in Sections 5.1 to 5.4 collectively demonstrate that the peak TA selectivity of 55 % observed at 130 °C, 1:6 w/v Gly: Ac, 0.5 g catalyst loading, and 2 h reaction time arises from the concurrence of several catalyst-specific factors discussed below.

The 15SA catalyst possesses exclusively strong acid sites resulting from the electron-withdrawing effect of SO_4^{2-} groups coordinated to Ti^{4+} centres in the Ti–O–S framework, as confirmed by XPS (O 1s at 530.8 eV; S 2p at 169.4–170.6 eV) and NH_3-TPD (T_{max} = 608 °C, total acidity = 342.6 μmol g^{-1}). The strong Lewis acid sites (Ti^{4+}) activate the carbonyl group of acetic acid into electrophilic acylium ions, while adjacent Brønsted acid sites (Ti–OH, O 1s at 533.9 eV) facilitate the nucleophilic attack of glycerol's hydroxyl groups. The cooperative action of both Lewis and Brønsted acidity on the same catalyst surface enables the full three-step sequential acetylation to proceed to TA under relatively mild conditions, particularly at the moderate temperature of 130 °C.

The optimal temperature of 130 °C is not coincidental. It represents the minimum temperature at which the endothermic DA to TA step proceeds at a rate sufficient to accumulate TA within the 2 h reaction window, while remaining below the threshold at which acetic acid volatilisation under reflux conditions significantly depletes the reagent supply. The mesoporous structure of 15SA (pore diameter 5.2 nm, Type IV isotherm) supports this by accommodating glycerol intermediates (~1 nm) within the pore channels, enabling multiple sequential encounters with active sites without requiring diffusion to a new particle surface.

The optimal molar ratio of 1:6 w/v ensures that acetic acid is present in sufficient two-fold stoichiometric excess to sustain continuous generation of acylium ions on the strong acid sites, while not overwhelming the limited active site population constrained by the relatively low BET surface area of 15SA (1.9 m^2 g^{-1}) with non-productive competitive adsorption. The optimal catalyst loading of 0.5 g represents the threshold at which the cumulative strong acid site density is sufficient to drive the DA → TA step at 130 °C within 2 h, without triggering agglomeration-induced mass transfer limitations that emerge at higher loadings, as supported by the SEM-observed particle morphology.

Finally, the 2 h reaction time is the kinetic window during which all these conditions converge to achieve maximum TA selectivity. It is sufficiently long for the sequential MA to DA to TA cascade to run to near-completion on the strong acid sites of 15SA, yet short enough to avoid the onset of water-mediated hydrolysis of Ti–O–S bonds that progressively deactivates the catalyst. In summary, the optimal condition yielding 55% TA selectivity and 97% GLC emerges from the concurrence of four factors: sufficient thermal energy to

drive the endothermic TA-forming step, adequate acetic acid excess to sustain acylium ion generation, the right active site density for complete three-step acetylation, and a reaction duration that maximises TA accumulation before catalyst deactivation commences. This finding demonstrates that the physicochemical properties of the 15SA catalyst, specifically its strong acid site density, cooperative Lewis–Brønsted acidity, and mesoporous structure, collectively govern the optimal parametric window for selective TA production in glycerol acetylation.

Reusability Study

The inactivation of the catalytically active sites commonly hampers the industrial application of solid catalysts. Therefore, the stability and deactivation of the 15SA catalyst were determined by studying successive reuse cycles. The reusability study over the 15SA catalyst was investigated under optimal reaction conditions of a molar ratio of 1:6 w/v of Gly: Ac, catalyst dosage of 0.5 g, reaction temperature of 130 °C, for a reaction duration of 2 h.

Figure 13 shows a slight reduction in the selectivity of TA, which decreased from 55 % in the first cycle to 24 % in the fifth cycle, whereas the selectivity of MA and DA increased over the same period. Particularly, in the first cycle, MA and DA had selectivity of 5 % and 41 %, respectively, while in the fifth cycle, MA selectivity increased to 15 % and DA selectivity rose to 61 %. Furthermore, the EDX results shown in Figure 8 revealed that the sulfur content in the fresh catalyst was 6 %, while the sulfur content decreased to 2 % in the spent catalyst as shown in Figure 14. Thus, it is proven that the catalytic activities and the EDX results are aligned with those of Mou et al., (2021), who suggested that the minor reduction in TA selectivity might be due to the loss of the acidity of the catalysts caused by intense stirring, separation, and washing throughout the reaction and recovery process. Additionally, it is advised that additional investigation be conducted using XPS and TGA analysis to further confirm whether the sulfate species were leached out from the 15SA catalyst or experienced structural changes following its use in the reaction. Fortunately, the DA selectivity in this study remains high at 61 % after five cycles, indicating that the leaching of the active sites only inhibits the diffusion of TA molecules on the catalyst surface, resulting in lower TA selectivity. This finding is aligned with Testa et al., (2019).

Moreover, it is clearly seen in Figure 13 that the conversion of glycerol (GLC) from the first cycle to the second cycle slightly improved from 97 % to 100 % and remained excellent after being reutilised for five cycles. This discovery suggests that the further leaching of sulfate species from the catalyst's surface may result in a larger surface area of the catalyst, thereby improving the conversion of the glycerol. This observation aligned with the BET results in Table 1, where the non-sulfated titania exhibited a larger BET surface area of 11. 8 m^2g^{-1} while 15SA exhibited a BET surface area of 1.9 m^2g^{-1}.

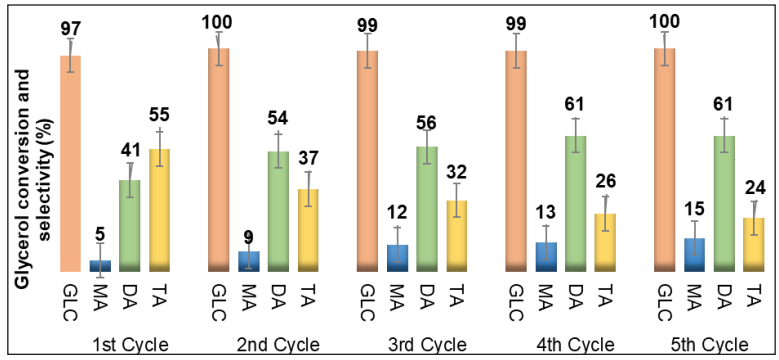


Figure 13. Reusability study using sulfated-titania catalyst (15SA). (Reaction condition: MR = 1:6 w/v Gly:Ac, CL = 0.5 g, t = 2 h and T = 130 °C)

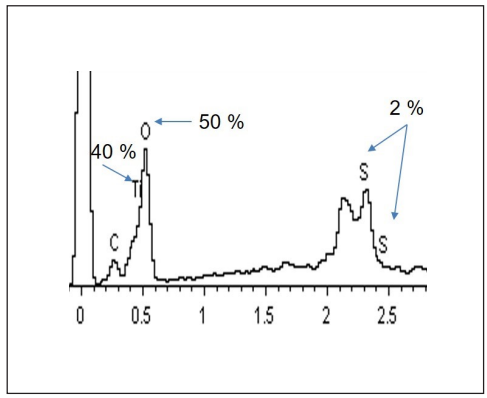


Figure 14. EDX result for the spent 15SA catalyst

Therefore, this discovery suggests that apart from having sufficient BET surface area, the catalyst used in the glycerol acetylation requires strong acidic sites to push forward the reaction toward higher TA selectivity. Furthermore, Herrada-Vidales et al. (2020) state that the mixture of DA and TA can also be utilised as a fuel additive due to their comparable properties, such as high hydrocarbon solubility, low water affinity, and high additive effect. Therefore, DA and TA do not need to be separated before their use in biodiesel and gasoline. As oxygenated fuel additives, a mixture of DA and TA could also diminish the emission of toxic gases and the number of particles in the exhaust fume and provide better combustion in the cylinder (Malaika et al., 2023). Therefore, Figure 13 shows that the utilisation of the 15SA catalyst at optimal conditions exhibits a higher mixture of DA and TA selectivity, exceeding 85 %. This result indicates that the 15SA catalyst has a great deal to produce fuel additives by acetylation under mild reaction conditions. Additionally, the reusability study over the 15SA catalyst concluded that the leaching of the sulfate species from the solid catalyst is an unavoidable phenomenon that caused lower TA selectivity in subsequent uses in the reaction. Therefore, it is recommended to conduct further studies to investigate suitable methods and solvents used for regenerating the catalyst before its use in the next reaction to minimise sulfate species leaching while maintaining high TA selectivity until five complete cycles.

CONCLUSION

The sulfated-titania catalyst (15SA) was prepared using the impregnation method with a 15% sulfuric acid concentration. The catalyst was successfully evaluated in the parametric study to observe the influence of temperature, loading of the catalyst, molar ratio of Gly: Ac and reaction time on glycerol conversion and acetins selectivity. The robust acidic strength, sufficient BET surface area and good porosity of the 15SA catalyst facilitate a higher TA selectivity of 55 % at optimal conditions of a 1:6 w/v Gly: Ac molar ratio, utilising a catalyst dosage of 0.5 g at a reaction temperature of 130 °C for a 2 h reaction. In the reusability study, the 15SA catalyst was successfully utilised in the fifth consecutive cycle, achieving excellent GLC over 97%. In contrast, the selectivity of TA dropped from 55 % to 24 % for the first to the fifth cycles, respectively. Fortunately, the DA selectivity remained high, achieving 61 % and DA+TA selectivity, exceeding 85 % throughout five cycles. This observation indicates that the 15SA catalyst can generate more fuel additives consisting of a mixture of DA and TA, thereby potentially replacing homogenous catalysts that have difficulties in separation, are highly toxic and cannot be reused or regenerated. It is also acknowledged that the one-factor-at-a-time (OFAT) approach used in this study does not account for potential interaction effects between reaction parameters such as temperature, molar ratio, catalyst loading, and reaction time. For future work, it is recommended to employ multivariate statistical design methods, such as Response Surface Methodology (RSM) or Central Composite Design (CCD), to comprehensively evaluate parameter interactions and validate the optimal conditions identified in this study.

Declaration of Competing Interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence this work.

ACKNOWLEDGEMENT

The authors deeply appreciate the GPPI grant (GPPI/2019/9676400) from Universiti Putra Malaysia, and UTM-UPM CRG (Matching Grant/2019/9300458). Catalysis Science and Technology Research Centre (PutraCat) is also gratefully acknowledged.

REFERENCES

Abida, K., Chudasama, B., & Ali, A. (2020). Development and functionalisation of magnetic nanoparticles as stable and reusable catalysts for triacetin synthesis. *New Journal of Chemistry, 44*(22), 9365-9376. https://doi.org/10.1039/D0NJ00488J

Afshar, S., Sadehvand, M., Azad, A., Dekamin, M. G., Jalali-Heravi, M., Mollahosseini, A., Amani, M., & Tadjarodi, A. (2015). Optimisation of catalytic activity of sulfated titania for efficient synthesis of isoamyl

acetate by response surface methodology. *Monatshefte für Chemie – Chemical Monthly, 146*(12), 1949-1957. https://doi.org/10.1007/s00706-015-1533-5

Appaturi, J. N., Maireles-Torres, P., Alomar, T. S., Najla, A., El-Bahy, Z. M., Ling, T. C., & Ng, E. P. (2023). Highly active nickel (II) oxide-supported cerium oxide fuel additives. *Materials, 16*(13), Article 4713. https://doi.org/10.3390/ma16134713

Appaturi, J. N., Ramalingam, R. J., Selvaraj, M., Chia, S., Tan, S. H., Khoerunnisa, F., Ling, T. C., & Ng, E. P. (2021). Selective synthesis of triacetyl glyceride biofuel additive via acetylation of glycerol over NiO-supported TiO_2 catalyst enhanced by non-microwave instant heating. *Applied Surface Science, 545*, Article 149017. https://doi.org/10.1016/j.apsusc.2021.149017

Azri, N., Irmawati, R., Nda-Umar, U. I., Saiman, M. I., & Taufiq-Yap, Y. H. (2021). Promotional effect of transition metals (Cu, Ni, Co, Fe, Zn)-supported on dolomite for hydrogenolysis of glycerol into 1,2-propanediol. *Arabian Journal of Chemistry, 14*(4), Article 103047. https://doi.org/10.1016/j.arabjc.2021.103047

Beejapur, H. A., La Parola, V., Liotta, L. F., & Testa, M. L. (2017). Glycerol acetylation over organic–inorganic sulfonic or phosphonic silica catalysts. *ChemistrySelect, 2*(17), 4934-4941. https://doi.org/10.1002/slct.201700934

Carpegiani, J. A., Godoy, W. M., Guimarães, D. H. P., & Aguiar, L. G. (2020). Glycerol acetylation catalysed by an acidic styrene-co-dimethacrylate resin: Experiments and kinetic modelling. *Reaction Kinetics, Mechanisms and Catalysis, 130*(1), 447-461. https://doi.org/10.1007/s11144-020-01788-7

Chandrakala, U., Prasad, R. B. N., & Prabhavathi Devi, B. L. A. (2014). Glycerol valorisation as biofuel additives by employing a carbon-based solid acid catalyst derived from glycerol. *Industrial & Engineering Chemistry Research, 53*(42), 16164-16169. https://doi.org/10.1021/ie503079m

Chilakamarry, C. R., Sakinah, A. M. M., Zularisam, A. W., & Pandey, A. (2021). Glycerol waste to value-added products and its potential applications. *Systems Microbiology and Biomanufacturing, 1*(4), 378-396. https://doi.org/10.1007/s43393-021-00036-w

da Silva, D. S., Altino, F. M. R. S., Bortoluzzi, J. H., & Meneghetti, S. M. P. (2020). Investigation of Sn(IV) catalysts in glycerol acetylation. *Molecular Catalysis, 494*, Article 111130. https://doi.org/10.1016/j.mcat.2020.111130

Dabbawala, A. A., Alhassan, S. M., Mishra, D. K., Jegal, J., & Hwang, J. S. (2018). Solvent-free cyclodehydration of sorbitol to isosorbide over mesoporous sulfated titania with enhanced catalytic performance. *Molecular Catalysis, 454*, 77-86. https://doi.org/10.1016/j.mcat.2018.05.009

Dalla Costa, B. O., Decolatti, H. P., Legnoverde, M. S., & Querini, C. A. (2017). Influence of acidic properties of different solid acid catalysts for glycerol acetylation. *Catalysis Today, 289*, 222-230. https://doi.org/10.1016/j.cattod.2016.09.015

Dantas, J., Leal, E., Mapossa, A. B., Cornejo, D. R., & Costa, A. C. F. M. (2017). Magnetic nanocatalysts of Ni0.5Zn0.5Fe2O4 doped with Cu and performance evaluation in transesterification reaction for biodiesel production. *Fuel, 191*, 463-471. https://doi.org/10.1016/j.fuel.2016.11.107

Dey, S., Reang, N. M., Das, P. K., & Deb, M. (2021). A comprehensive study on the prospects of the economy, environment, and efficiency of palm oil biodiesel as a renewable fuel. *Journal of Cleaner Production, 286*, Article 124981. https://doi.org/10.1016/j.jclepro.2020.124981

Dhawan, M. S., Barton, S. C., & Yadav, G. D. (2021). Interesterification of triglycerides with methyl acetate for the co-production of biodiesel and triacetin using hydrotalcite as a heterogeneous base catalyst. *Catalysis Today, 375*, 101-111. https://doi.org/10.1016/j.cattod.2020.07.056

Ebadi Pour, N., Dumeignil, F., Katryniok, B., Delevoye, L., Revel, B., & Paul, S. (2021). Investigating the active phase of Ca-based glycerol polymerisation catalysts: On the importance of calcium glycerolate. *Molecular Catalysis, 507*, Article 111571. https://doi.org/10.1016/j.mcat.2021.111571

Enríquez, J. M. H., Lajas, L. A. C., Alamilla, R. G., Martín, E. Á. S., Alamilla, G. P., Handy, E. B., Galindo, G. C., & Serrano, L. A. G. (2013). Synthesis of solid acid catalysts based on TiO_2–SO_4 applied in *n*-hexane isomerisation. *Journal of Metal, 3*(3), 34-44.

Fauzi, M., Termezi, A., Izyanni, S., Hizami, M., & Yusoff, M. (2025). Effect of Fe and Ni loading in Fe–Ni supported on activated carbon catalyst on glycerol acetylation to acetins. *Sustainable Chemistry for the Environment, 10*, Article 100162. https://doi.org/10.1016/j.scenv.2025.100162

Ferreira, P., Fonseca, I. M., Ramos, A. M., Vital, J., & Castanheiro, J. E. (2009). Esterification of glycerol with acetic acid over dodecamolybdophosphoric acid encaged in USY zeolite. *Catalysis Communications, 10*(5), 481-484. https://doi.org/10.1016/j.catcom.2008.10.015

Fronza, C. S., Nunes, A. L. B., Castilhos, F., & Salau, N. P. G. (2025). Niobium-based catalysts for glycerol valorisation via acetylation: A sustainable path to fuel additives. *Energy Conversion and Management, 342*, Article 120146. https://doi.org/10.1016/j.enconman.2025.120146

García-Martín, J. F., Alés-Álvarez, F. J., Torres-García, M., Feng, C. H., & Álvarez-Mateos, P. (2019). Production of oxygenated fuel additives from residual glycerine using biocatalysts obtained from heavy-metal-contaminated *Jatropha curcas* L. roots. *Energies, 12*(4), Article 740. https://doi.org/10.3390/en12040740

Gardy, J., Hassanpour, A., Lai, X., & Ahmed, M. H. (2016). Synthesis of $Ti(SO_4)O$ solid acid nano-catalyst and its application for biodiesel production from used cooking oil. *Applied Catalysis A: General, 527*, 81-95. https://doi.org/10.1016/j.apcata.2016.08.031

Geetha, S., Thangamani, A., Valliappan, R., Vedanayaki, S., & Ganapathi, A. (2020). Sulfated titania (TiO_2-SO_4^{2-}) as an efficient and reusable solid acid catalyst for the multi-component synthesis of highly functionalised piperidines. *Chemical Data Collections, 30*, Article 100565. https://doi.org/10.1016/j.cdc.2020.100565

Herrada-Vidales, J. A., García-González, J. M., Martínez-Palou, R., & Guzmán-Pantoja, J. (2020). Integral process for obtaining acetins from crude glycerol and their effect on the octane index. *Chemical Engineering Communications, 207*(2), 231-241. https://doi.org/10.1080/00986445.2019.1578758

Hidayati, N., Khoiruddin, W., Purnama, H., & Effendy, M. (2024a). Synthesis and characterisation of graphene oxide catalysts for glycerol acetylation. *Bulletin of Chemical Reaction Engineering and Catalysis, 19*(2), 340-349. https://doi.org/10.9767/bcrec.20146

Hidayati, N., Rahma Maulida, I., Purnama, H., Musthofa, M., & Ur Rahmah, A. (2024b). Optimisation of the glycerol acetylation process using graphene oxide catalyst. *South African Journal of Chemical Engineering, 48*, 254-264. https://doi.org/10.1016/j.sajce.2024.02.011

Hu, W., Zhang, Y., Huang, Y., Wang, J., Gao, J., & Xu, J. (2015). Selective esterification of glycerol with acetic acid to diacetin using antimony pentoxide as reusable catalyst. *Journal of Energy Chemistry, 24*(5), 632-636. https://doi.org/10.1016/j.jechem.2015.08.001

Huang, Z., Lin, Y., Li, L., Ye, C., & Qiu, T. (2017). Preparation and shaping of solid acid SO_4^{2-}/TiO_2 and its application for esterification of propylene glycol monomethyl ether and acetic acid. *Chinese Journal of Chemical Engineering, 25*(9), 1207-1216. https://doi.org/10.1016/j.cjche.2016.11.006

Jenie, S. N. A., Kusuma, D. S., Kristiani, A., Laksmono, J. A., & Tursiloadi, S. (2010). Preparation and characterisation of sulfated titania catalysts for the isomerisation of citronellal. *International Journal of Basic and Applied Science, 10*(06), 5-10.

Jiang, Y., Li, X., Zhao, H., & Hou, Z. (2019). Esterification of glycerol with acetic acid over SO_3H-functionalized phenolic resin. *Fuel, 255*, Article 115842. https://doi.org/10.1016/j.fuel.2019.115842

Jose, M., Paloma, N., Lopes, G., & Rodrigues, A. A. (2023). Biodiesel additives synthesis using solid heteropolyacid catalysts. *Energies, 16*(3), Article 1332. https://doi.org/10.3390/en16031332

Kale, S., Umbarkar, S. B., Dongare, M. K., Eckelt, R., Armbruster, U., & Martin, A. (2015). Selective formation of triacetin by glycerol acetylation using acidic ion-exchange resins as a catalyst and toluene as an entrainer. *Applied Catalysis A: General, 490*, 10-16. https://doi.org/10.1016/j.apcata.2014.10.059

Kalyani, T., Prasad, L. S. V., & Kolakoti, A. (2023). Effect of triacetin as an oxygenated additive in algae biodiesel fuelled CI engine combustion, performance, and exhaust emission analysis. *Fuel, 338*, Article 127366. https://doi.org/10.1016/j.fuel.2022.127366

Kanimozhi, S., Ramani, V., & Pandurangan, A. (2018). Effects of alumina on GO and KIT-6 supports for the acetylation of glycerol. *New Journal of Chemistry, 42*(18), 15197-15208. https://doi.org/10.1039/C8NJ01833B

Karnjanakom, S., Maneechakr, P., Samart, C., & Guan, G. (2018). Ultrasound-assisted acetylation of glycerol for triacetin production over green catalyst: A liquid biofuel candidate. *Energy Conversion and Management, 173*, 262-270. https://doi.org/10.1016/j.enconman.2018.07.086

Khayoon, M. S., Triwahyono, S., Hameed, B. H., & Jalil, A. A. (2014). Improved production of fuel oxygenates via glycerol acetylation with acetic acid. *Chemical Engineering Journal, 243*, 473-484. https://doi.org/10.1016/j.cej.2014.01.027

Kulkarni, R. M., Britto, P. J., Narula, A., Saqline, S., Anand, D., Bhagyalakshmi, C., & Herle, R. N. (2020). Kinetic studies on the synthesis of fuel additives from glycerol using CeO_2–ZrO_2 metal oxide catalyst. *Biofuel Research Journal, 7*(1), 1100-1108. https://doi.org/10.18331/BRJ2020.7.1.2

Ladero, M. (2021). New glycerol upgrading processes. *Catalysts, 11*(1), Article 103. https://doi.org/10.3390/catal11010103

Lee, Y., Kao, Y., Xu, H., Tsunekawa, S., & Yoshida, M. (2026). Green synthesis of glycerol carbonate from glycerol and CO_2 using a polymeric ionic liquid catalyst: An experimental and theoretical study. *Journal of Cleaner Production, 560*, Article 148344. https://doi.org/10.1016/j.jclepro.2026.148344

Lertna, N., Srisamut, P., Supapan, N., Athikaphan, P., Rood, S. C., Neramittagapong, A., & Neramittagapong, S. (2022). Synthesis of sulfated titanium dioxide catalyst for sorbitol dehydration to isosorbide. *Materials Today: Proceedings, 75*, 91-98. https://doi.org/10.1016/j.matpr.2022.11.226

Liu, G., Tsen, W. C., Hu, F., Zhong, F., Wang, J., Liu, H., Wen, S., Zheng, G., Qin, C., & Gong, C. (2020). Enhanced proton conductivities of chitosan-based membranes by inorganic solid superacid SO_4^{2-}–TiO_2 coated carbon nanotubes. *International Journal of Hydrogen Energy, 45*(53), 29212-29221. https://doi.org/10.1016/j.ijhydene.2020.07.157

Liu, J., Wang, Z., Sun, Y., Jian, R., Jian, P., & Wang, D. (2019). Selective synthesis of triacetin from glycerol catalysed by HZSM-5/MCM-41 micro/mesoporous molecular sieve. *Chinese Journal of Chemical Engineering, 27*(5), 1073-1078. https://doi.org/10.1016/j.cjche.2018.09.013

Magar, S., Mohanraj, G. T., Jana, S. K., & Rode, C. V. (2020). Synthesis and characterisation of supported heteropoly acid: Efficient solid acid catalyst for glycerol esterification to produce biofuel additives. *Inorganic and Nano-Metal Chemistry, 50*(11), 1157-1165. https://doi.org/10.1080/24701556.2020.1737817

Malaika, A., Mesjasz, D., & Kozłowski, M. (2023). Maximising the selectivity to triacetin in glycerol acetylation through a plastic waste-derived carbon catalyst development and selection of a reaction unit. *Fuel, 333*, Article 126271. https://doi.org/10.1016/j.fuel.2022.126271

Mou, R., Wang, X., Wang, Z., Zhang, D., Yin, Z., Lv, Y., & Wei, Z. (2021). Synthesis of fuel bioadditive by esterification of glycerol with acetic acid over hydrophobic polymer-based solid acid. *Fuel, 302*, Article 121175. https://doi.org/10.1016/j.fuel.2021.121175

Mufrodi, Z., Rochmadi, Sutijan, & Budiman, A. (2014). Synthesis of acetylation of glycerol using a batch reactor and a continuous reactive distillation column. *Engineering Journal, 18*(2), 29-39. https://doi.org/10.4186/ej.2014.18.2.29

Mukhopadhyay, P., Chakraborty, R., & Singh, S. (2021). Triacetin additive in biodiesel to reduce air pollution: A review. *Environmental Chemistry Letters, 20*(1), 3-27. https://doi.org/10.1007/s10311-021-01362-0

Nda-Umar, U. I., Irmawati, R., Muhamad, E. N., Azri, N., & Taufiq-Yap, Y. H. (2020a). Optimisation and characterisation of mesoporous sulfonated carbon catalyst and its application in modelling and optimisation of acetin production. *Molecules, 25*(22), Article 5221. https://doi.org/10.3390/molecules25225221

Nda-Umar, U. I., Irmawati, R., Muhamad, E. N., Taufiq-Yap, Y. H., & Azri, N. (2020b). Synthesis and characterisation of sulfonated carbon catalysts derived from biomass waste and their evaluation in glycerol acetylation. *Biomass Conversion and Biorefinery, 12*, 1143-1158. https://doi.org/10.1007/s13399-020-00784-0

Neto, A. B. S., Oliveira, A. C., Rodriguez-Castellón, E., Campos, A. F., Freire, P. T. C., Sousa, F. F. F., Filho, J. M., Araujo, J. C. S., & Lang, R. (2020). A comparative study on porous solid acid oxides as catalysts in the esterification of glycerol with acetic acid. *Catalysis Today, 349*, 57-67. https://doi.org/10.1016/j.cattod.2018.05.057

Noda, L. K., de Almeida, R. M., Gonçalves, N. S., Probst, L. F. D., & Sala, O. (2003). TiO_2 with a high sulfate content: Thermogravimetric analysis, determination of acid sites by infrared spectroscopy and catalytic activity. *Catalysis Today, 85*(1), 69-74. https://doi.org/10.1016/S0920-5861(03)00195-0

Okoye, P. U., Abdullah, A. Z., & Hameed, B. H. (2017a). A review on recent developments and progress in the kinetics and deactivation of catalytic acetylation of glycerol—A byproduct of biodiesel. *Renewable and Sustainable Energy Reviews, 74*, 387-401. https://doi.org/10.1016/j.rser.2017.02.017

Okoye, P. U., Abdullah, A. Z., & Hameed, B. H. (2017b). Synthesis of oxygenated fuel additives via glycerol esterification with acetic acid over bio-derived carbon catalyst. *Fuel, 209*, 538-544. https://doi.org/10.1016/j.fuel.2017.08.024

Okoye, P. U., & Hameed, B. H. (2016). Review of recent progress in catalytic carboxylation and acetylation of glycerol as a byproduct of biodiesel production. *Renewable and Sustainable Energy Reviews, 53*, 558-574. https://doi.org/10.1016/j.rser.2015.08.064

Parida, K. M., Sahu, N., Biswal, N. R., Naik, B., & Pradhan, A. C. (2008). Preparation, characterisation, and photocatalytic activity of sulfate-modified titania for degradation of methyl orange under visible light. *Journal of Colloid and Interface Science, 318*(2), 231-237. https://doi.org/10.1016/j.jcis.2007.10.028

Patil, S. M., Deshmukh, S. P., More, K. V., Shevale, V. B., Mullani, S. B., Dhodamani, A. G., & Delekar, S. D. (2019). Sulfated TiO_2/WO_3 nanocomposite: An efficient photocatalyst for degradation of Congo red and methyl red dyes under visible light irradiation. *Materials Chemistry and Physics, 225*, 247-255. https://doi.org/10.1016/j.matchemphys.2018.12.041

Pratika, R. A., Wijaya, K., & Trisunaryanti, W. (2021). Hydrothermal treatment of SO_4/TiO_2 and TiO_2/CaO as heterogeneous catalysts for the conversion of Jatropha oil into biodiesel. *Journal of Environmental Chemical Engineering, 9*(6), Article 106547. https://doi.org/10.1016/j.jece.2021.106547

Pratika, R. A., Wijaya, K., Utami, M., Mulijani, S., Patah, A., Alarifi, S., Ram Mani, R., Kumar Yadav, K., Ravindran, B., Chung, W. J., Chang, S. W., & Munusamy-Ramanujam, G. (2023). The potency of hydrothermally prepared sulfated silica (SO_4/SiO_2) as a heterogeneous acid catalyst for ethanol dehydration into diethyl ether. *Chemosphere, 341*, Article 139822. https://doi.org/10.1016/j.chemosphere.2023.139822

Ramos, L. D., Ramírez, A. E., & Miranda, C. D. (2025). Acidic eco-catalysts derived from coffee husks in the acetylation of glycerol. *Tetrahedron Green Chemistry, 5*, Article 100075. https://doi.org/10.1016/j.tgchem.2025.100075

Rathod, S., Joshi, K. K., Patel, D., Pataniya, P., Nandpal, M. N., Patel, S. G., Sumesh, C. K., & Kapatel, S. (2025). Facile surface reconstructions of cobalt–copper oxide heterostructures enable efficient electrocatalytic glycerol oxidation for energy-saving hydrogen evolution. *Journal of Alloys and Compounds, 1048*, Article 185267. https://doi.org/10.1016/j.jallcom.2025.185267

Reddy, P. S., Sudarsanam, P., Raju, G., & Reddy, B. M. (2010). Synthesis of bio-additives: Acetylation of glycerol over zirconia-based solid acid catalysts. *Catalysis Communications, 12*(1), 62-65. https://doi.org/10.1016/j.catcom.2010.07.006

Reinoso, D. M., & Boldrini, D. E. (2020). Kinetic study of fuel bio-additive synthesis from glycerol esterification with acetic acid over acid polymeric resin as a catalyst. *Fuel, 264*, Article 116879. https://doi.org/10.1016/j.fuel.2019.116879

Ropero-Vega, J. L., Aldana-Pérez, A., Gómez, R., & Niño-Gómez, M. E. (2010). Sulfated titania (TiO_2/SO_4^{2-}): A very active solid acid catalyst for the esterification of free fatty acids with ethanol. *Applied Catalysis A: General, 379*(1-2), 24-29. https://doi.org/10.1016/j.apcata.2010.02.020

Shera Farisya, M. R., Irmawati, R., Shafizah, I. N., Taufiq-Yap, Y. H., Muhamad, E. N., Lee, S. L., & Salamun, N. (2021). Assessment of the effect of sulfuric acid concentration on physicochemical properties of sulfated-titania catalyst and glycerol acetylation performance. *Catalysts, 11*(12), Article 1542. https://doi.org/10.3390/catal11121542

Shi, W., & Li, J. (2014). A deactivation mechanism of sulfated titania in the esterification of acetic acid and *n*-butanol. *Reaction Kinetics, Mechanisms and Catalysis, 111*(1), 215-233. https://doi.org/10.1007/s11144-013-0628-4

Shikhaliyev, K., Hameed, B. H., & Okoye, P. U. (2021). Utilisation of biochars as sustainable catalysts for the upgrading of glycerol from biodiesel production. *Journal of Environmental Chemical Engineering, 9*(2), Article 104768. https://doi.org/10.1016/j.jece.2020.104768

Silva, L. N., Gonçalves, V. L. C., & Mota, C. J. A. (2010). Catalytic acetylation of glycerol with acetic anhydride. *Catalysis Communications, 11*(12), 1036-1039. https://doi.org/10.1016/j.catcom.2010.05.007

Singh, R. K., & Gosu, V. (2025). Catalytic acetalization of glycerol to solketal using mesoporous silica. *Applied Catalysis A: General, 703*, Article 120350. https://doi.org/10.1016/j.apcata.2025.120350

Spataru, D., Soares Dias, A. P., & Vieira Ferreira, L. F. (2021). Acetylation of biodiesel glycerin using glycerin and glucose-derived catalysts. *Journal of Cleaner Production, 297*, Article 126686. https://doi.org/10.1016/j.jclepro.2021.126686

Srivastava, S., Jadeja, G. C., & Parikh, J. (2017). Synergism studies on alumina-supported copper-nickel catalysts towards furfural and 5-hydroxymethylfurfural hydrogenation. *Journal of Molecular Catalysis A: Chemical, 426*, 244-256. https://doi.org/10.1016/j.molcata.2016.11.023

Testa, M. L., La Parola, V., Mesrar, F., Ouanji, F., Kacimi, M., Ziyad, M., & Liotta, L. F. (2019). Use of zirconium phosphate-sulphate as an acid catalyst for the synthesis of glycerol-based fuel additives. *Catalysts, 9*(2), Article 148. https://doi.org/10.3390/catal9020148

Timofeeva, M. N., Panchenko, V. N., Khan, N. A., Hasan, Z., Prosvirin, I. P., Tsybulya, S. V., & Jhung, S. H. (2017). Isostructural metal-carboxylates MIL-100(M) and MIL-53(M) (M: V, Al, Fe and Cr) as catalysts for the condensation of glycerol with acetone. *Applied Catalysis A: General, 529*, 167-174. https://doi.org/10.1016/j.apcata.2016.11.006

Tomer, R., & Biswas, P. (2020). Dehydration of glucose/fructose to 5-hydroxymethylfurfural (5-HMF) over an easily recyclable sulfated titania (SO_4^{2-}/TiO_2) catalyst. *New Journal of Chemistry, 44*(47), 20734-20750. https://doi.org/10.1039/D0NJ04151C

Wang, J., Pan, H., Wang, A., Tian, X., Wu, X., & Wang, Y. (2015). Synthesis and characterisation of $S_2O_8^{2-}/ZnFe_xAl_{2-x}O_4$ solid acid catalysts for the esterification of acetic acid with *n*-butanol. *Catalysis Communications, 62*, 29-33. https://doi.org/10.1016/j.catcom.2015.01.009

Wijaya, K., Putri, A. R., Sudiono, S., Mulijani, S., Patah, A., Wibowo, A. C., & Saputri, W. D. (2021). Effectively synthesizing SO_4/TiO_2 catalyst and its performance for converting ethanol into diethyl ether (DEE). *Catalysts, 11*(12), Article 1492. https://doi.org/10.3390/catal11121492

Yang, W., Ok, Y. S., Dou, X., Zhang, Y., Yang, M., Wei, D., & Xu, P. (2019). Effectively remediating spiramycin from production wastewater through hydrolysis of its functional groups using solid superacid TiO_2/SO_4. *Environmental Research, 175*, 393-401. https://doi.org/10.1016/j.envres.2019.05.037

Zhao, H., Jiang, P., Dong, Y., Huang, M., & Liu, B. (2014). A high-surface-area mesoporous sulfated nano-titania solid superacid catalyst with exposed (101) facets for esterification: Facile preparation and catalytic performance. *New Journal of Chemistry, 38*(9), 4541-4548. https://doi.org/10.1039/C4NJ00494A

Pertanika J. Sci. & Technol. 34 (4): 2053 - 2074 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Data-Driven Fault Detection and Diagnosis of a Nuclear Reactor Cooling System Using Graph Neural Networks

Dhines Shree Kanthan[1], Nur Aina Mohammad Aziz[1], Azura Che Soh[1]*, Ribhan Zafira Abdul Rahman[1], Mohd Khair Hassan[1], Mohd Sabri Minhat[2], and Julia Abdul Karim[2]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[2]Reactor Instrumentation and Control Engineering, Reactor Technology Centre, Malaysian Nuclear Agency (Nuklear Malaysia), 43000 Kajang, Selangor, Malaysia

ABSTRACT

The cooling system of a nuclear reactor is essential for ensuring the safety and efficient operation of the TRIGA PUSPATI Reactor (RTP). However, conventional fault detection methods often struggle with the nonlinearities and complexity of the system. This study proposes a data-driven Fault Detection and Diagnosis (FDD) framework for the RTP cooling system using a Graph Neural Network (GNN). In this approach, the system is modelled as a graph, where sensors are represented as nodes and their interactions as edges, enabling the GNN to capture component relationships and improve anomaly detection accuracy. A scaled-down control rig was developed to replicate the RTP cooling system for experimental data collection and validation. The model was trained and evaluated using experimental data, and its performance was assessed using evaluation metrics including accuracy, precision, recall, and F1 score. For comparison, an Artificial Neural Network (ANN) model was also implemented. The results show that the GNN outperforms the ANN in fault detection performance. Both models were evaluated under offline conditions, while real-time validation was conducted for the GNN. In the online implementation, a Wi Fi module was used to transmit live sensor data to the cloud for continuous monitoring. The control rig operates independently in a laboratory environment and is not connected to any operational nuclear reactor system, with cloud communication used strictly for experimental monitoring without control actions. The results demonstrate that the proposed GNN achieves accurate and reliable

ARTICLE INFO

Article history:
Received: 15 April 2026
Accepted: 13 July 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.05

E-mail addresses:
gs75025@student.upm.edu.my (Dhines Shree Kanthan)
gs69737@student.upm.edu.my (Nur Aina Mohammad Abdul Aziz)
azuracs@upm.edu.my (Azura Che Soh)
ribhan@upm.edu.my (Ribhan Zafira Abdul Rahman)
khair@upm.edu.my (Mohd Khair Hassan)
sabri@nm.gov.my (Mohd Sabri Minhat)
julia@nm.gov.my (Julia Abdul Karim)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

fault detection in both offline and real-time applications, highlighting its potential to enhance reactor safety and operational reliability.

Keywords: Artificial neural network, data-driven, fault detection and diagnosis, graph neural network, reactor cooling system

INTRODUCTION

State-of-the-art fault detection and diagnosis (FDD) methods can generally be divided into two main groups: traditional or physics-based techniques and Artificial Intelligence (AI)-based approaches. Statistical Process Control (SPC), model-based and rule-based systems are typical approaches (Bersimis et al., 2007; Martinez-Viol et al., 2020; Qian et al., 2024; Simiyon et al., 2024). Granted, SPC applies statistical techniques such as control charts to control process variables; however, it is based on clearly defined and simple processes (Bersimis et al., 2007). The model-based approach uses mathematical models to compare model circumstances or prediction anomalies between expectation and actual behaviour, but it encounters difficulties when applied to nonlinear, uncertain and complex systems (Simiyon et al., 2024). Rule-based approaches rely on expert-design threshold and are inflexible to accommodate more sophisticated systems if the task is getting harder (Qian et al., 2024).

In recent years, advancements in machine learning and data analytics, driven by the availability of big data, have brought significant attention towards data-driven techniques (Bon & Van Dai, 2022). Among these, Artificial Neural Networks (ANNs) are widely adopted for their capability to capture complex patterns for classification and anomaly detection tasks, but it requires large datasets (de Albuquerque Filho et al., 2022; Pang et al., 2021). Alternatively, Support Vector Machines (SVMs) can achieve strong classification performance, but they require careful hyperparameter tuning and may become computationally expensive while handling large datasets (Bharathwaj et al., 2025; de Souza et al., 2014; Liu et al., 2021).

Furthermore, deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can select hierarchical features from raw time-series data, eliminating the need for manual feature engineering. Despite this advantage, their practical implementation is often constrained by the requirement for large datasets of labelled data and more powerful computational resources due to their complex architectures and numerous parameters (Mienye, 2024; Kong et al., 2023).

In contrast, Graph Neural Networks (GNNs) represent a more recent and expressive paradigm for systems that consist of interacting elements. Due to processing graph-structure data with nodes representing components and edges indicating relationships, GNN can

capture dependencies not captured by traditional neural networks (Hoang et al., 2023; Khemani et al., 2024; Theiler & Fink, 2024). This capability makes them particularly well-suited for the demanding and complex conditions encountered in industrial systems.

This research develops a GNN-based FDD system for complex industrial cooling systems. A scaled-down process control rig replicates the RTP cooling system to safely generate operational data under both normal and fault conditions, which cannot be performed directly on the reactor. The system is modelled to capture its dynamics and interactions, and a GNN-based FDD algorithm is developed and validated on the control rig platform, efficiently learning complex interdependencies and providing adaptive, data-driven monitoring. This approach offers a safe, scalable and intelligent solution for monitoring complex interdependent systems and advancing fault detection research in industrial settings.

Process Control Rig for RTP Cooling System

The RTP cooling is based on the components which are designed to remove the heat from the reactor in a structured manner. The cooling system passes off the heat from the reactor to its tower, which emits it into the atmosphere. Critical sensors such as temperature, pressure and flow measure system status always. The first pump passes the coolant from the reactor to a cooling tower, while the second is responsible for circulating water in a heat exchanger. In each of the loops, flow and pressure are controlled by valves to keep the process at safe operating conditions. Scaled-down process control rig and its P&ID diagram are provided in Figures 1 and 2, respectively.

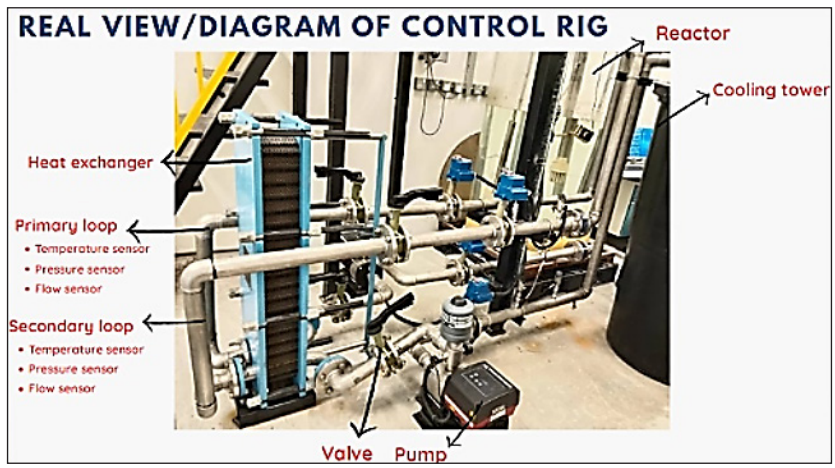


Figure 1. Real view of the process control rig for the RTP cooling system

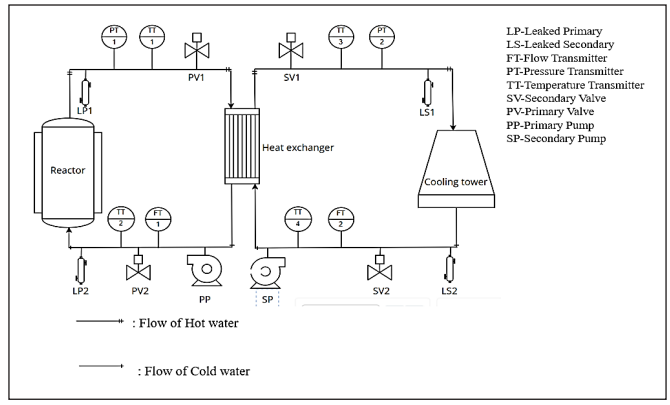


Figure 2. P&ID diagram of process control rig for RTP cooling system

The P&ID shows the plumbing of sensors, and flow paths: double line is for hot water pipelines and Single lines are for cold water pipes. It also displays relationships among the reactor, heat exchanger, pumps, valves and sensors, as well as with a cooling tower to explain how heat travels and is manipulated throughout components in the system.

The pilot plant control rig was developed based on the actual RTP cooling system to support experimental data acquisition under normal and fault conditions for FDD development. The rig is designed as a scaled-down functional validation platform with an approximate 1:20 overall water volume ratio and is not intended to reproduce exact reactor thermal hydraulic behaviour, but rather to emulate operational logic, control system response, and degradation behaviour of structures, systems, and components (SSCs). Although the system architecture of the RTP cooling system is preserved through a simplified configuration of primary and secondary loops, including tanks, pumps, a heat exchanger, and an electric heater assembly, local fluid dynamic phenomena cannot be fully replicated due to scaling effects.

Thermal-hydraulic distortions are therefore unavoidable, particularly in natural circulation effects, heat transfer coefficients, and flow distribution characteristics. The electric heater (1.5 kW total, consisting of three 500 W elements) provides a controlled heat source to represent reactor core loading in a qualitative manner for control system evaluation rather than nuclear heat reproduction. The piping system is simplified to 2-inch diameter lines compared to the actual 4-inch (primary) and 6-inch (secondary) RTP system, and demineralised water is not used, introducing minor deviations in thermophysical properties. Instrumentation includes RTD PT100 sensors, Type K thermocouples, flow meters, and pressure sensors for system-level monitoring and control validation. Despite these limitations, the test rig provides a safe and effective platform for control development, educational training, and AI-based fault diagnosis studies, enabling macro-level analysis of system behaviour and fault scenarios without risk to actual reactor operation.

METHODOLOGY

A project of project for this study starts with gathering data from the system. The FDD system model is established, and the GNN model for FDD is proposed online and offline tested accordingly. If sufficient performance of the model is achieved, it is implemented to perform analysis on the results. A GUI is then generated to provide user access to the system. If the model does not satisfy the requirement, it will continue to optimise its parameters and validate again.

Experiment and Data Collection

This section describes the data collection procedure for the RTP cooling system process control rig. The system configuration and sensor setup are summarised in Table 1, while eight fault scenarios representing the degradation of critical system components are presented in Table 2. Before fault injection, the cooling system was operated under normal conditions to establish a stable baseline, and normal operating data were recorded. For each experimental run, the system was allowed to reach steady-state operation before data acquisition to ensure that the recorded measurements accurately represented system behaviour under both normal and fault conditions. A sampling duration of 20 minutes was adopted for each operating condition.

The dataset was collected using six sensors installed on the RTP cooling system control rig, comprising two pressure transmitters (PT1 and PT2), three temperature transmitters (TT1, TT2, and TT3), and one flow transmitter (FT1). Sensor measurements were acquired at a sampling rate of 1 Hz, corresponding to one sample per second throughout each experiment. The complete dataset contained 28,644 samples covering both normal operation and the eight fault scenarios.

In this study, a GNN model is proposed for modelling the RTP cooling system, while a conventional ANN model is used as the benchmarking method for performance comparison. The dataset was randomly divided into 70% training, 15% validation, and 15% testing subsets. The training set was used to optimise the model parameters, while the validation set monitored the generalisation performance and reduced the risk of overfitting during training. The independent testing set was reserved exclusively for the final performance evaluation. Cross-validation was not employed because a fixed training, validation, and testing partition was adopted to ensure a consistent and fair comparison between the proposed GNN model and the benchmark ANN model. In addition, weight decay regularisation and validation monitoring were incorporated to further enhance the generalisation capability of both models.

Table 1
Types of components with their functions and fault scenarios

Components	Functions	Fault Scenarios
Primary Valve 1 (PV1)	Regulates coolant flow within the primary loop	Failure of PV1
Primary Valve 2 (PV2)	Controls coolant flow in the primary cooling system	Failure of PV2
Leaked Primary Pump 1 (LP1)	Ensures circulation of coolant in the reactor's primary loop; any leaks can result in reduced flow	Failure of LP1
Leaked Primary Pump 2 (LP2)	Ensures coolant circulation in the primary loop to absorb core heat; leaks in PV2 can reduce flow.	Failure of LP2
Leaked Secondary Pump 1 (LS1)	The secondary loop transfers heat to the cooling tower; leaks reduce coolant flow	Failure of LS1
Leaked Secondary Pump 2 (LS2)	Ensures coolant circulation in the secondary loop; leaks in S2 may decrease the flow rate.	Failure of LS2

Table 2
Types of faults and their injection method

Fault	Fault Type	Fault Injection
1	Restricted water flow rate in the Hot-Heat Exchanger & Reactor	Incremental closure of PV1: 30%, 60%, and 90%
2	Restricted water flow rate in the Hot-Heat Exchanger & Cooling Tower	Incremental closure of SV1: 30%, 60%, and 90%
3	Restricted water flow rate in the Hot-Heat Exchanger & Cooling Tower	Increase the opening of LS1 from 50% to 100%
4	Restricted water flow rate in the Cold-Heat Exchanger & Reactor	Incremental closure of PV2: 30%, 60%, and 90%
5	Restricted water flow rate in the Cold-Heat Exchanger & Reactor	The PP was turned OFF
6	Restricted water flow rate in the Cold-Heat Exchanger & Cooling tower	Increase the opening of LS2 from 50% to 100%
7	Restricted water flow rate in the Cold-Heat Exchanger & Cooling tower	Incremental closure of SV2: 30%, 60%, and 90%
8	Restricted water flow rate in the Cold-Heat Exchanger & Cooling tower	The SP was turned OFF

Modelling of RTP Cooling System

Figure 3 presents the bock diagram of RTP cooling system, including the relationship between inputs and outputs, such as the main transmitter that indicates how the different system components correlate to achieve proper thermal regulation and operational stability.

The cooling system depends on Temperature Transmitters (TT), Pressure Transmitters (PT) and Flow Transmitters (FT) to control temperature, flow and pressure inside the reactor loops. The incoming coolant flow and temperature are measured by FT2 and TT4 in the

cold heat exchanger, while PT2 and TT3 measure the outgoing pressure and temperature to ascertain a good heat transfer. These outputs, along with reactor PT1 and TT1, become inputs for the hot heat exchanger, whose outputs FT1 and TT2 are fed back to the reactor, providing a control means relating to cooling ability and safety. This feedback mechanism permits the system to respond to changes and achieve a desired level of thermal and pressure conditions in the reactor core.

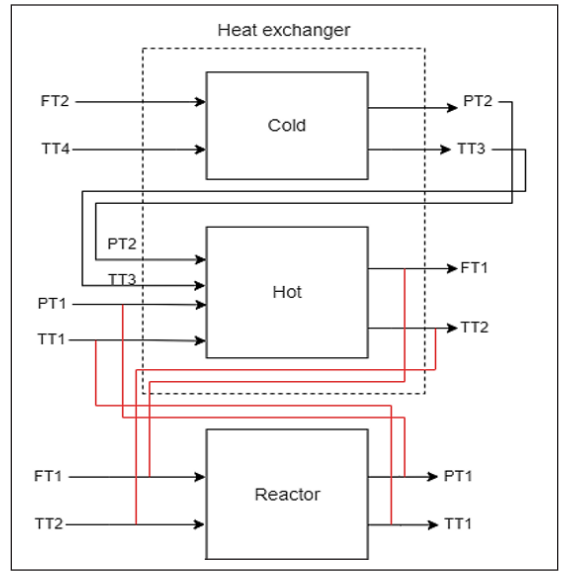


Figure. 3 Block diagram of RTP cooling system

GNN-Based RTP Cooling System Model

The model architecture of the GNN including three layers (input, subsystem, and output), is depicted in Figure 4. The input layer is used to take the temperature (TT), flow (FT), and pressure (PT) sensor readings from various locations in the cooling circuit. For example, FT2 and TT4 are for the Cold Heat Exchanger. While PT2, TT3, PT1 and TT1 are for the Hot Heat Exchanger and FT1 and TT2 are for the Reactor. In contrast, the output layer is used as both input for the next subsystem and as the feedback mechanism. A node is represented by each subsystem while connecting its interactions.

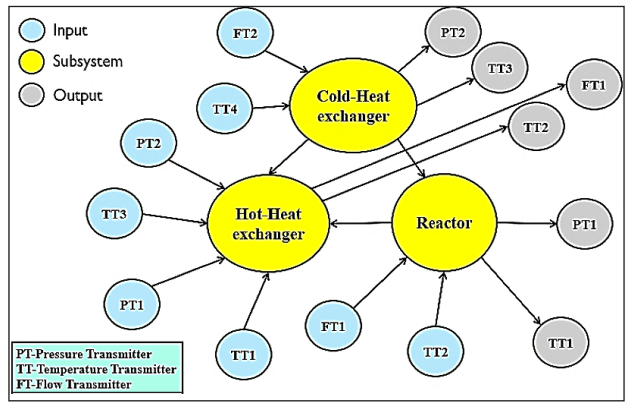


Figure 4. Structure of the GNN model for RTP cooling system control rig

In GNNs, several hyperparameters may impact the model's accuracy, including performance, convergence rate, and generalisation capability. Inaccurate configuration can cause overfitting, underfitting, slow convergence, or high computational cost. Therefore, these parameters should be properly calibrated to ensure stable and reliable model performance.

The number of hidden layers (N_H) may influence the depth of the GNN and affect its capability to capture the nonlinearities in the system. While deeper architectures can improve the model's behaviour, an excessive number of hidden layers may introduce training instability and performance reduction. Prior studies indicate that GNNs generally achieve optimal performance with two to four hidden layers, whereas performance tends to decline when the number exceeds seven layers (Kipf & Welling, 2017; Xu et al., 2019). Nevertheless, when labelled data or node features are limited, deeper architectures may be beneficial to support semi-supervised learning.

The hidden layer dimension (N_{HD}) determines the size of the node embeddings and showcases the model's ability to represent complex features. For relatively simple tasks such as node classification, embedding dimensions of 16, 32, or 64 are typically sufficient (Liu et al., 2021). Increasing the embedding dimension can improve the model's representation capability but also increase computational cost, whereas smaller dimensions may decrease effective feature extraction.

In comparison to conventional neural networks, GNNs consider information from neighbouring nodes using a graph structure. Thus, the number of neighbouring nodes (N_{NN}) is a critical parameter, as it represents how much surrounding information is related during message transfer. Larger embedding dimensions often require a broader neighbourhood to capture meaningful relationships between nodes. Previous research indicates that an optimal number of neighbouring nodes generally lies between 5 and 50, as values outside this range can decline the model performance (Chen et al., 2023).

The number of samples being processed in each training cycle is decided based on the batch size (B_S), It significantly affects both convergence behaviour and memory usage. Smaller batch sizes allow faster parameter updates and are suitable for simpler models, one that is used in RTP cooling systems, where values between 1 and 32 are often effective (Hu et al., 2023). On the other hand, larger batch sizes can improve training stability but require more computational resources.

The number of graph convolution layers (N_{CL}) is another key architectural parameter, particularly in convolution-based GNNs. Multiple factors are considered, including the number of hidden layers, embedding dimensions, neighbouring nodes, and batch size. Typically, the optimal number of layers ranges from one to four, as deeper networks are prone to over-smoothing, where node features become overly similar (Zhang et al., 2023).

The number of training iterations (N_T) directly impacts training efficiency. Excessive iterations can increase computational cost and lead to overfitting, while too few may result in underfitting. Studies suggested that optimal iteration counts typically range between 50 and 1000, depending on dataset size and model complexity.

Finally, the learning rate also plays a crucial role in controlling convergence speed and stability. Commonly adopted values range from 0.01 to 1, particularly when adaptive or decaying learning strategies are employed (Yang et al., 2023). An excessively high learning rate may cause divergence, whereas an overly small value can slow convergence and increase the risk of getting stuck in local minima.

The forward propagation process in a GNN updates node features through successive graph convolution layers according to the formulation introduced in the Graph Convolutional Network framework, as expressed in Equation 1 (Kipf & Welling, 2017):

$$H^{l+1} = \sigma\left(W^{(l)}H^{(l)}A + B^{(l)}\right) \quad [1]$$

where $H^{(l)}$ is the feature matrix at layer l, $W^{(l)}$ and $B^{(l)}$ are trainable weights and biases, respectively, A is the adjacency matrix representing node connections, and σ is a non-linear activation function such as ReLU.

Model parameters are optimised using gradient descent, which updates the parameters iteratively according to the standard optimisation rule widely used in neural networks, as expressed in Equation 2 (Rumelhart et al., 1986):

$$\theta^{(t+1)} = \theta^{(t)} - l_r \cdot \nabla_\theta J(\theta) \quad [2]$$

where θ represents the model parameters, l_r is the learning rate, and $J(\theta)$ is the loss function.

To improve generalisation and mitigate overfitting, weight decay is incorporated through L2 regularisation, as commonly adopted in machine learning optimisation, as expressed in Equation 3 (Goodfellow et al., 2016):

$$J(\theta) = L(\theta) + W_d \sum_i \theta_i^2 \quad [3]$$

where $J(\theta)$ is the original loss function and W_d is the weight decay coefficient.

The sigmoid parameter (S_p) controls the nonlinearity of the sigmoid activation function, originally introduced in classical neural network models, as expressed in Equation 4 (Rumelhart et al., 1986):

$$(x) = \frac{1}{1 + e^{-S_p x}} \quad [4]$$

where S_p controls the steepness of the activation function and directly influences the model's sensitivity to input variations. If S_p is too large, the function saturates rapidly, leading to vanishing gradients during training. Conversely, a very small S_p may slow down the learning process and hinder convergence.

The model is trained using datasets which include temperature, pressure, and flow rate measurements collected under both normal operations and simulated fault scenarios. During training, the model aims to minimise the Mean Squared Error (MSE), which is a metrics used to validate regression learning that compute the average squared difference between actual and predicted values, as expressed in Equation 5 (Hastie et al., 2009).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad [5]$$

where y_i represents the actual value, $\hat{y}_i$ is the predicted value, and n denotes the total number of samples.

Lower MSE values, together with regression coefficients approaching one, indicate that the GNN effectively captures the hidden pattern within the RTP cooling system. This enables the model to distinguish between normal and faulty operating conditions, thereby supporting accurate fault detection and classification. Table 3 presents a summary of the final GNN architecture and selected hyperparameter settings, including the number of layers, activation functions, learning rate, batch size, and training epochs.

Table 3
GNN modelling with parameter settings

Hyperparameter	Symbol	Value range
Hidden Layers	N_h	2
Hidden Layer Dimension	N_{HD}	90
Number of Neighbor Nodes	N_{NN}	5
Batch Size	B_S	32
Graph Convolution Layers	N_{CL}	3

Table 3 (*continued*)

Hyperparameter	Symbol	Value range
Number of Training	N_t	100
Learning Rates	l_r	0.01
Weight Decay	λ	0.0001
Sigmoid Parameter	S_P	1.0

ANN Benchmark Model

Figure 5 illustrates the architecture of the Artificial Neural Network (ANN), representing the conventional machine learning (ML) model used for modelling the RTP cooling system. The ANN employs a feedforward neural network (FNN) architecture comprising an input layer, two fully connected hidden layers, and an output layer. The input layer receives sensor measurements from the RTP cooling system, including pressure transmitters (PT), temperature transmitters (TT), and flow transmitters (FT). These input features are propagated through two hidden layers, each consisting of 90 neurons with the Rectified Linear Unit (ReLU) activation function, enabling the network to learn the nonlinear relationships between the process variables and the corresponding fault classes.

The output layer employs a linear activation function to predict the target outputs associated with the monitored cooling system variables. The ANN model was trained using the Adam optimiser with a learning rate of 0.01, a batch size of 32, and 100 training epochs. Model optimisation was performed by minimising the MSE loss function through the backpropagation algorithm. These hyperparameter settings were selected to ensure a fair comparison with the proposed GNN, allowing both models to be trained under identical experimental conditions.

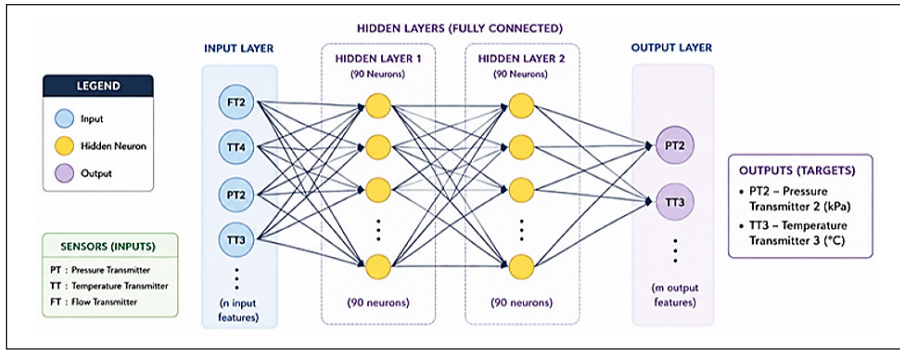


Figure 5. Structure of the ANN model for the RTP cooling system control rig

Unlike the proposed GNN, which explicitly represents the RTP cooling system as a graph of interconnected sensors, the conventional ANN processes each sensor measurement as an independent input feature through fully connected layers. Consequently, the ANN is unable to directly capture the physical relationships and fault propagation among pressure, temperature, and flow sensors. This limitation reduces its capability to model complex subsystem interactions, leading to lower fault classification performance compared with the graph-based GNN. Nevertheless, the ANN serves as a widely accepted benchmark for evaluating the effectiveness of the proposed GNN architecture in industrial FDD applications. Table 4 summarises the ANN architecture and selected hyperparameter settings.

Table 4
ANN modelling with parameter settings

Hyperparameter	Symbol	Value range
Hidden Layers	N_h	2
Hidden Layer Dimension	N_{HD}	90
Number of Neighbor Nodes	N_{NN}	-
Batch Size	B_S	32
Graph Convolution Layers	N_{CL}	-
Number of Training	N_t	100
Learning Rates	l_r	0.01
Weight Decay	λ	0.0001
Sigmoid Parameter	S_P	1.0

Verification of Offline and Online GNN Approaches for FDD System

In this work, only the proposed GNN model is evaluated under both offline and online testing scenarios. The GNN (Offline) solution is trained using historical data, including temperature, pressure, and flow rate measurements under normal reactor operation conditions. When the GNN (Offline) model is trained, its predictions are considered as a baseline for comparison against the performance of GNN (Online).

Apart from the Offline GNN models, the Online GNN operates in real-time using live data via Wi-Fi adapter from the cloud and logging sensor readings are uploaded automatically to Google Drive from the cooling system's control rig directly, as described in Figure 6. This enables the system to operate continuously and respond to any fault conditions within the reactor cooling circuit. After the real-time data is processed by the GNN (Online) model, it is stored in a server and made available for system analysis and review.

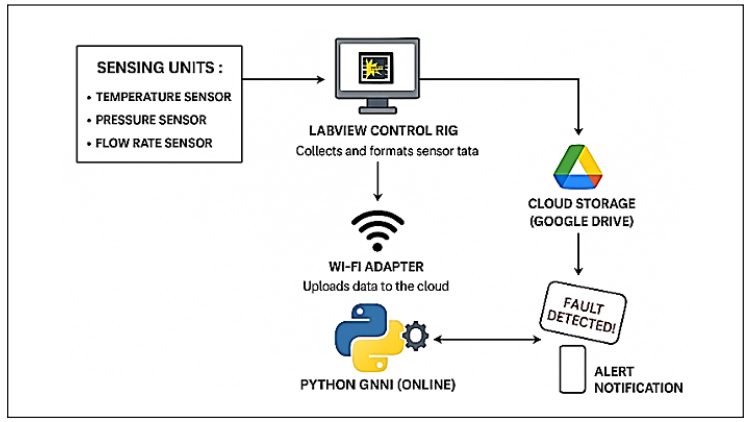


Figure 6. Online FDD implementation using GNN on the RTP cooling system control rig

To verify the validity and information reliability of our GNN (Online) method, we compare online real-time prediction results with offline GNN-based models. Such a comparative method can also verify the performance of the GNN (Online) model and make sure it makes predictions likely to the history. This validation ensures the correctness of the real-time model and thus increases fault detection reliability.

It is important to note that the pilot plant control rig operates independently within a laboratory research environment and is not connected to any operational nuclear reactor system. The wireless communication and cloud storage are used strictly for experimental monitoring and research purposes, and the system does not perform any control actions. For practical deployment in nuclear facilities, the architecture would require adaptation to comply with cybersecurity requirements, such as air-gapped or segmented networks, secure plant-level communication infrastructure, data diode-based unidirectional transmission, and edge computing deployment for local processing without reliance on external cloud services. These approaches ensure that the conceptual monitoring framework demonstrated in the test rig can be translated into a secure, plant-compliant architecture without violating nuclear cybersecurity regulations.

RESULTS AND DISCUSSIONS

This section presents the modelling results of the RTP cooling system using the proposed GNN model, with the ANN serving as the benchmark model. Both models are evaluated under normal operating conditions and faulty scenarios in the offline analysis. For the online analysis, only the proposed GNN model is evaluated. The FDD results are then presented, followed by an overall evaluation of the proposed system's performance.

RTP Cooling System Modelling and Analysis

The proposed GNN provides an accurate representation of the RTP cooling system under normal operating conditions. The model was trained using experimentally observed data collected from the RTP cooling system and successfully captured the dynamic behaviour of the three major subsystems, namely the cold-heat exchanger, hot-heat exchanger, and reactor. To further evaluate the effectiveness of the proposed approach, its performance was compared with a conventional ANN trained using the same dataset and identical training conditions.

Model performance was evaluated using the mean residual value. A residual below 10% was classified as Rank A, indicating excellent agreement between the predicted and measured system responses. Residuals greater than 10% but below 30% were classified as Rank B, residuals exceeding 50% were classified as Rank D, and all remaining residuals were classified as Rank C (Atkinson et al., 2017). As presented in Table 5, the proposed GNN achieved Rank A for all subsystem outputs, with mean residual values ranging from 0.05% to 1.75%, demonstrating excellent prediction accuracy across all three subsystems. In contrast, although the conventional ANN could model the overall system behaviour, it produced consistently higher residual values. The ANN achieved Rank A for PT2, FT1, and PT1, but only Rank B for TT3, TT2, and TT1, indicating lower prediction accuracy for the temperature-related outputs.

Table 5
Comparative performance evaluation of the RTP cooling system subsystem model using the proposed GNN and the conventional ANN

Subsystem	Output	Mean Residual (%) GNN	Ranking (GNN)	Mean Residual (%) ANN	Ranking (ANN)
Cold-Heat Exchanger	PT2	1.75	A	7.84	A
	TT3	0.06	A	12.65	B
Hot-Heat Exchanger	FT1	0.16	A	8.21	A
	TT2	0.06	A	11.43	B
Reactor	PT1	1.75	A	9.38	A
	TT1	0.05	A	10.72	B

Note. A = good; B = acceptable; C = marginal; D = poor

Table 6 presents the evaluation metrics comparison between the proposed GNN and the conventional ANN for FDD of the RTP cooling system. Both models were trained and evaluated using the same experimentally collected dataset and identical training conditions to ensure a fair comparison. The proposed GNN achieved 98.12% Accuracy, 99.15% Precision, 99.26% Recall, and 98.20% F1-score, whereas the conventional ANN achieved 96.68% Accuracy, 96.72% Precision, 96.63% Recall, and 96.67% F1-score.

These results indicate that the proposed GNN successfully classified all normal and fault operating conditions without misclassification, while the ANN exhibited a small number of incorrect classifications.

In addition to the classification metrics, the proposed GNN demonstrated superior regression performance by achieving an MSE of 0.0002, compared with 0.0184 for the conventional ANN. Similarly, the GNN produced a lower average residual value of 1.70%, whereas the ANN recorded an average residual of 3.18%. Furthermore, the proposed GNN achieved a classification error of 0.01%, while the ANN produced a classification error of 3.32%. These findings demonstrate that the GNN provides more accurate predictions of the pressure, temperature, and flow sensor responses under both normal and faulty operating conditions.

The superior performance of the proposed GNN is attributed to its graph-based learning architecture, which explicitly models the physical relationships among interconnected pressure, temperature, and flow sensors within the RTP cooling system. By learning both the sensor features and their structural dependencies, the GNN effectively captures complex system dynamics and fault propagation, resulting in lower prediction errors and higher modelling accuracy than the conventional ANN. In contrast, the ANN treats each sensor measurement as an independent feature and cannot explicitly exploit the interdependencies among sensors, limiting its ability to represent the underlying system behaviour. Consequently, the proposed GNN provides a more robust and reliable framework for fault detection and diagnosis, demonstrating that graph-based learning is better suited for modelling complex industrial cooling systems than conventional ANN approaches.

Table 6
Evaluation metrics comparison between the proposed GNN and the conventional ANN

Performance Metric	Proposed GNN	Conventional ANN
Accuracy (%)	98.12	96.68
Precision (%)	99.15	96.72
Recall (%)	99.26	96.63
F1-score (%)	98.20	96.67
MSE	0.0002	0.0184
Average Residual (%)	1.70	3.18
Classification Error (%)	0.01	3.32

Performance Evaluation of the GNN-Based FDD System in Offline Operation

An FDD algorithm based on GNN has been proposed for the system in normal conditions and under some faulty conditions. The offline and online validation of the algorithm in both cases was investigated experimentally. Below shows the corresponding values to indicate its capability of distinguishing normal from faulty states.

Under offline mode, the processed data are initially uploaded to the GNN-based FDD system to be analysed, as shown in Figure 7. The uploaded data is pre-processed and compared with the baseline normal operating conditions that are learned by the GNN model. The operational condition of the reactor control rig is determined in accordance with this calculation.

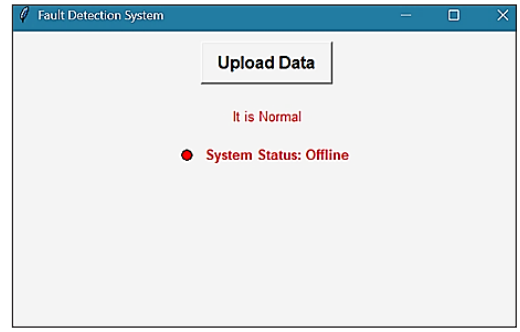


Figure 7. Offline GNN-based FDD under normal conditions

The final system status and fault location are visualised through the GNN representation depicted in Figure 8. In this diagram, the colour of the node denotes the state of the system; for instance, green nodes represent normal operation. Such a graphical representation has been considered to offer an intuitive and efficient method of monitoring system operation and detecting possible fault locations in offline modes as well.

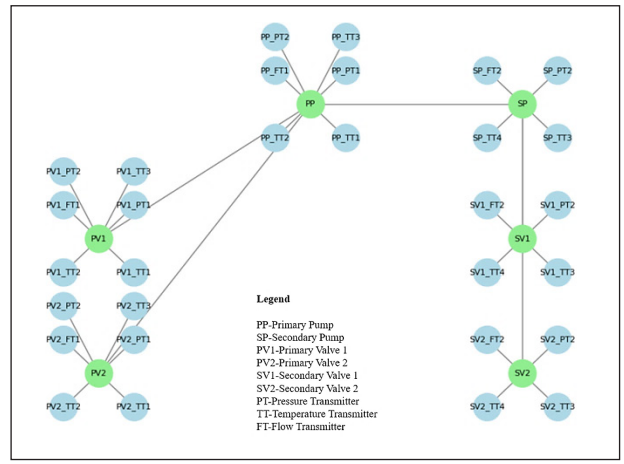


Figure 8. Offline GNN-based FDD: location mapping under normal conditions

Several fault scenarios were tested offline with actual fault data from Primary Pump (PP) and Secondary Valve 1 (SV1). When the measured data is uploaded to the system, the GNN mode l with detect and diagnose the faults by comparison of the measurements with normal operating conditions. The tendencies of the condition in PP and SV1 are shown in Figures 9 and 10, respectively.

The identified faults are represented by red-coloured nodes in the visualisations of Figures 11 and 12 to illustrate that the system can provide evidence for faults occurring in PP and SV1. These results demonstrate that the system is capable of accurate fault detection and able to provide direct indication of both the type and location of faults in the RTP cooling system rig.

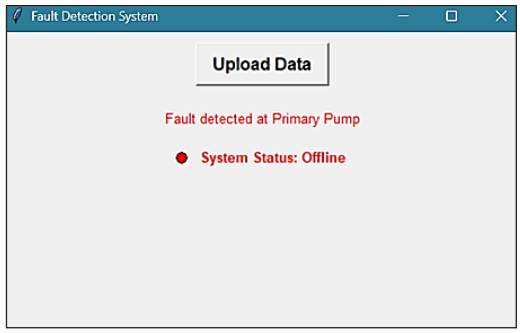


Figure 9. An offline GNN-based FDD system identifies a fault in the primary pump

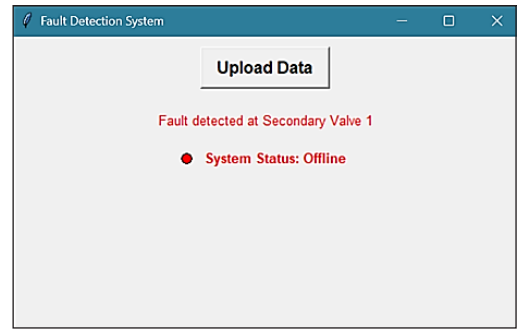


Figure 10. An offline GNN-based FDD system identifies a fault in the secondary valve 1

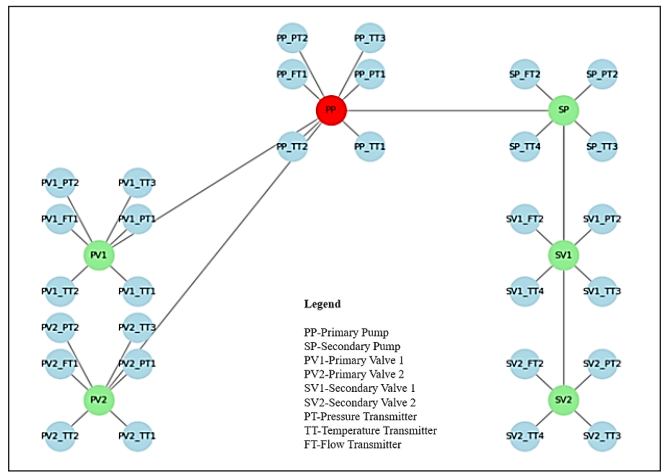


Figure 11. Offline GNN-based FDD system for fault localisation at the primary pump

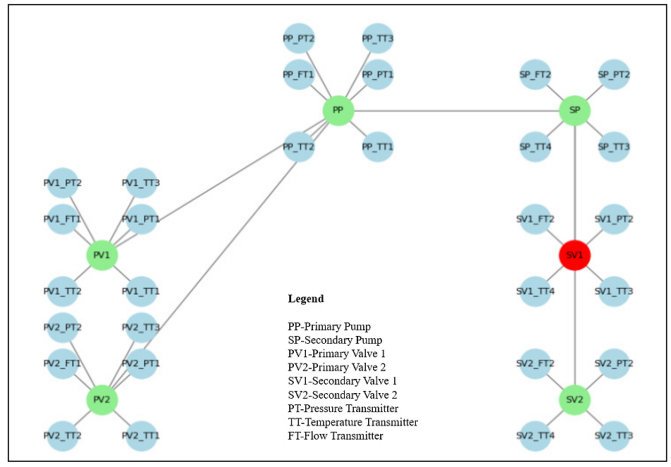


Figure 12. Offline GNN-based FDD system for fault localisation at the secondary valve 1

Performance Evaluation of the GNN-Based FDD System in Online Operation

The GNN-based FDD system was additionally validated via online analysis by interfacing it with a real-time implementation. This arrangement ensured that the system could also actively detect and locate faults as they arose. During this process, the system was operating in a normal manner at first. For the online validation of the fault scenarios tested, faults were injected one by one while running and re-running the above scenarios with a focus on PP, and SV1 was done as only single faults were tested.

The online validation agreed well with the offline analysis, indicating that faults have been consistently and accurately detected with both methods. These verifications of faults for the PP and SV1 are depicted in Figures 13 and 14. Furthermore, the types and location of faults are exactly those which can be seen in Figures 15 and 16 for the presented visualisations in Figures 11 and 12. These validation results demonstrate the robustness and reliability of the proposed GNN-based FDD system, highlighting its strong potential for real-time implementation in the actual RTP cooling system, as presented in the confusion matrix shown in Figure 17.

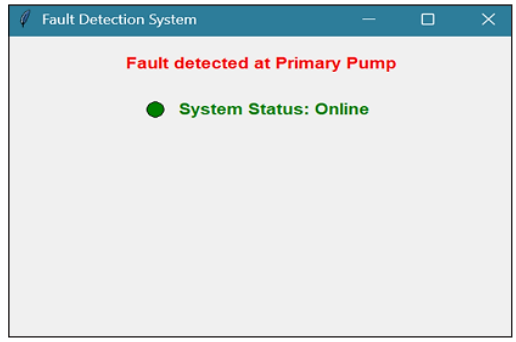


Figure 13. An online GNN-based FDD system identifies a fault in the primary pump

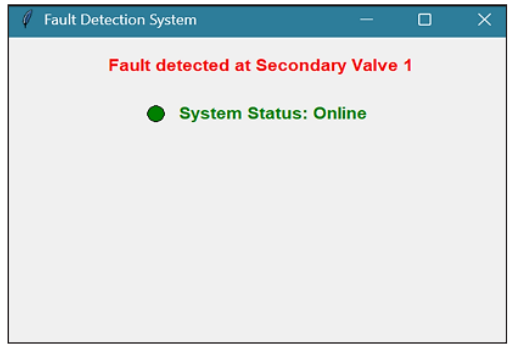


Figure 14. An online GNN-based FDD system identifies a fault in the secondary valve 1

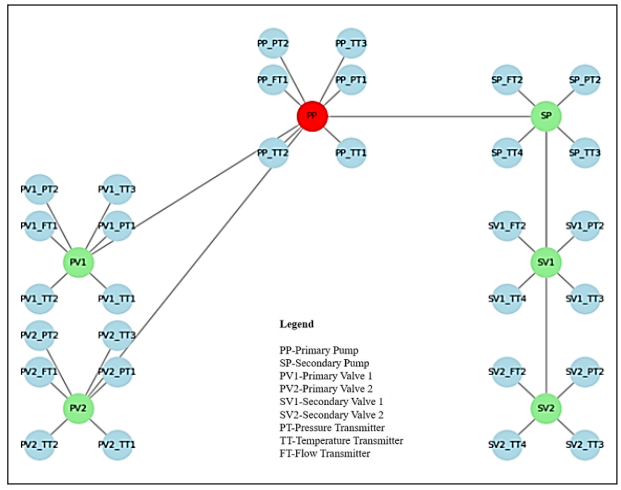


Figure 15. Online GNN-based FDD system for fault localisation at the primary pump

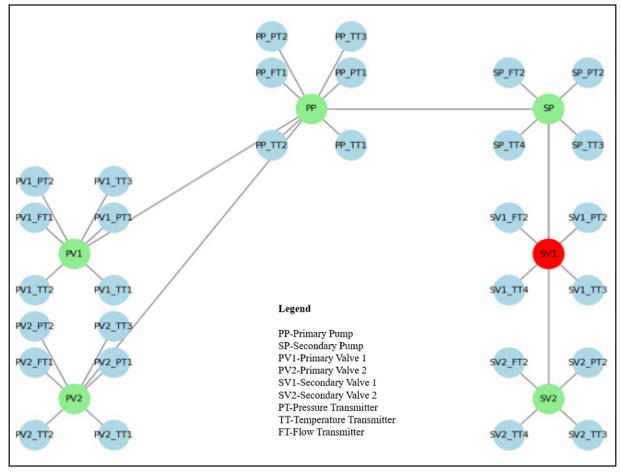


Figure 16. Online GNN-based FDD system for fault localisation at the secondary valve 1

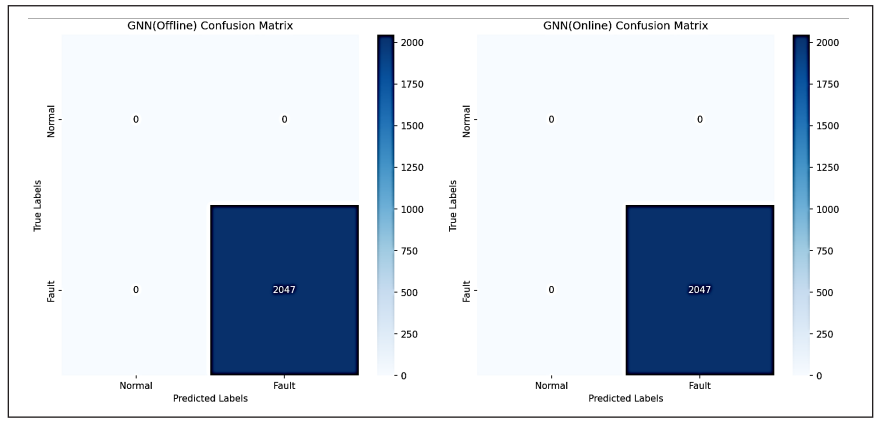


Figure 17. Confusion matrix of GNN (offline) and GNN (online)

CONCLUSIONS

GNN-based FDD algorithms were developed and validated offline for normal operation and chosen fault cases with Primary Pump as well as Secondary Valve 1. Through comparison of incoming data with a predetermined baseline, the system was able to effectively discriminate normal and faulty conditions and automatically determine the fault mode. Such an automatic procedure allows a faster identification and localisation of failures and can contribute to the system's improved reliability.

Online real-time tests were then conducted on a test rig applying the developed GNN-based FDD system, connected to components of the cooling system downstream of the pipeline. Online results were consistent with the offline evaluation, showing that models based on traditional baselines could well generalise to real operational scenarios.

These results also prove the effectiveness, precision and reliability of the GNN method as a fault detection and predictive maintenance solution for the RTP cooling system.

In summary, the proposed study has successfully designed an intelligent FDD system for a reactor cooling system of nuclear power plants with GNN. The key goals of the research were the construction of a GNN fault detection model, the creation of an intuitive graphical user interface, and proof-of-concept validation using data generated from the cooling system process control rig. All objectives were achieved successfully. Therefore, we can achieve accurate FDD for invasive faults in offline and online scenarios and significantly improve the maintenance efficiency and operational safety directly.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the Ministry of Science, Technology and Innovation (MOSTI) for providing funding to support the development of the RTP cooling system rig platform for this research. The authors also express their sincere thanks to the Malaysian Nuclear Agency for their guidance during the experimental work, assistance with data collection, and technical advice in testing the algorithms, which were instrumental to the successful completion of this research.

Data and Code Availability Statement

The data and code used in this study are not publicly available due to confidentiality restrictions.

REFERENCES

Atkinson, A. D., Hill, R. R., Pignatiello, J. J., Vining, G. G., White, E. D., & Chicken, E. (2017). Wavelet ANOVA approach to model validation. *Simulation Modelling Practice and Theory, 78*, 18-27. https://doi.org/10.1016/j.simpat.2017.08.004

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International, 23*(5), 517-543. https://doi.org/10.1002/qre.829

Bharathwaj, G., Khurshid, A., Venkatavijayan, S., & Pani, A. K. (2025). Industrial process monitoring using support vector data description: A systematic review and application for fault detection in multiphase flow systems. *Flow Measurement and Instrumentation, 106*, Article 103031. https://doi.org/10.1016/j.flowmeasinst.2025.103031

Bon, N. N., & Van Dai, L. (2022). Fault identification, classification, and location on transmission lines using combined machine learning methods. *International Journal of Engineering and Technology Innovation, 12*(2), 91-109. https://doi.org/10.46604/ijeti.2022.7571

Chen, S., Wulamu, A., Zou, Q., Zheng, H., Wen, L., Guo, X., Chen, H., Zhang, T., & Zhang, Y. (2023). MD-GNN: A mechanism for data-driven graph neural networks for molecular properties prediction and new material discovery. *Journal of Molecular Graphics and Modelling, 123*, Article 108506. https://doi.org/10.1016/j.jmgm.2023.108506

de Albuquerque Filho, J. E., Brandão, L. P., Fernandes, B. J. T., & Maciel, A. M. A. (2022). A review of neural networks for anomaly detection. *IEEE Access, 10*, 112342-112367. https://doi.org/10.1109/ACCESS.2022.3216007

de Souza, D. L., Granzotto, M. H., de Almeida, G. M., & Oliveira Lopes, L. C. (2014). Fault detection and diagnosis using support vector machines: A SVC and SVR comparison. *Journal of Safety Engineering, 3*(1), 18-29. https://doi.org/10.5923/j.safety.20140301.03

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://doi.org/10.7551/mitpress/10243.001.0001

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hoang, V. T., Jeon, H.-J., You, E.-S., Yoon, Y., & Jung, S. (2023). Graph representation learning and its applications: A survey. *Sensors, 23*(8), Article 4168. https://doi.org/10.3390/s23084168

Hu, B., Zhang, Y., Liu, Z., & Wang, J. (2023). Adaptive batch size selection for efficient deep neural network training. *IEEE Transactions on Neural Networks and Learning Systems, 34*(11), 5481-5493. https://doi.org/10.1109/TNNLS.2023.3272941

Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: Concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data, 11*(1), Article 18. https://doi.org/10.1186/s40537-023-00876-4

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)*. https://doi.org/10.48550/arXiv.1609.02907

Kong, X., Chen, Z., Liu, W., Ning, K., Zhang, L., Marier, S. M., Liu, Y., Chen, Y., & Xia, F. (2025). Deep learning for time series forecasting: A survey. *International Journal of Machine Learning and Cybernetics, 16*, 5079–5112. https://doi.org/10.1007/s13042-025-02560-w

Liu, C., Cichoń, A., Królczyk, G., Li, Z., & Zhu, X. (2022). Technology development and commercial applications of industrial fault diagnosis systems: A review. *The International Journal of Advanced Manufacturing Technology, 118*(11-12), 3497-3529. https://doi.org/10.1007/s00170-021-08047-6

Martinez-Viol, V., Urbano, E. M., Prieto, M. D., & Romeral, L. (2020). HVAC early fault detection using a fuzzy logic-based approach. *Renewable Energy and Power Quality Journal, 18*(2), 196-201. https://doi.org/10.24084/repqj18.270

Mienye, I. D. (2024). A comprehensive review of deep learning: Architectures, algorithms and limitations. *Information, 15*(12), Article 755. https://doi.org/10.3390/info15120755

Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys, 54*(2), Article 38. https://doi.org/10.1145/3439950

Qian, H., Pan, Y., Wang, X., & Li, Z. (2024). Research on the optimisation of belief rule bases using the Naive Bayes theory. *Frontiers in Energy Research, 12*, Article 1396841. https://doi.org/10.3389/fenrg.2024.1396841

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536. https://doi.org/10.1038/323533a0

Simiyon, A., Sachidanand, C., Halmakki Krishnamurthy, M., Bhatt, A. V., & Indiran, T. (2024). Machine learning-based fault classification in a pilot plant batch reactor using a support vector machine. *ACS Omega, 9*(26), 29041-29052. https://doi.org/10.1021/acsomega.4c04421

Theiler, R., & Fink, O. (2024). Graph neural networks for electric and hydraulic data fusion to enhance short-term forecasting of pumped storage hydroelectricity. In *Proceedings of the European Conference of the PHM Society, 8*(1), 1-11. https://doi.org/10.36001/phme.2024.v8i1.4129

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019).* https://doi.org/10.48550/arXiv.1810.00826

Yang, Z., Liu, Z., Zhou, J., Song, C., Xiang, Q., He, Q., Hu, J., Faber, M. H., Zio, E., Li, Z., Su, H., & Zhan, J. (2023). A graph neural network (GNN) method for assigning gas calorific values to natural gas pipeline networks. *Energy, 278*, Article 127875. https://doi.org/10.1016/j.energy.2023.127875

Zhang, H., Nguyen, D. H., & Tsuda, K. (2023). Differentiable optimisation layers enhance GNN-based mitosis detection. *Scientific Reports, 13*, Article 4306. https://doi.org/10.1038/s41598-023-41562-y

Pertanika J. Sci. & Technol. 34 (4): 2075 - 2095 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Impact of the Grading Ring Size and Hydrophobicity on the High-Voltage Insulator Performance

Amelia Adrina Azli[1], Nor Aqillah Aiman Nor Azman[1], N.H. Nik Ali[1]*, Hazlee Azil Illias[2], Ali Ahmed Ali Salem[3], Mohd Aizam Talib[4], and Mohd Fairouz Mohd Yousof[5]

[1]Department of Electrical Engineering, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
[2]Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[3]Department of Electrical Engineering, Applied Research Centre for Metrology, Standards and Testing, King Fahd University of Petroleum and Minerals (KFUPM), 31261 Dhahran, Saudi Arabia
[4]Electrical Engineering Department, TNB Labs Sdn. Bhd., No. 1, Lorong Ayer Itam, Kawasan Institusi Penyelidikan, 43000 Kajang, Selangor, Malaysia
[5]Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

ABSTRACT

A high-voltage insulator is an electrical power device that is usually used in the transmission and distribution system. The insulator is normally attached to the other instrument as a protection in case any breakdown or fault happens. Each high-voltage insulator has its own properties. Meanwhile, a grading ring, one of the components connected to the insulator, to improve the distribution of the electric field on the insulator's surface. This study, it involves the hydrophobic property of the insulator's surface since it is a primary issue that influences the electric field distribution. The distribution of the electric field in a high voltage insulation system is important because it can affect the performance, reliability and safety of the insulator. The electric field needs proper management to prevent electrical breakdown, extend the lifespan of insulation materials, and reduce power losses. Consequently, this study is mainly about the effect of the grading ring's design parameters and the presence of water droplets towards the electric field distribution on high-voltage insulators by using FEA software, which is COMSOL Multiphysics, to conduct the analysis that includes evaluating the electric field distribution across the high-voltage insulators.

ARTICLE INFO

Article history:
Received: 19 December 2025
Accepted: 21 May 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.06

E-mail addresses:
2025715747@student.uitm.edu.my (Amelia Adrina Azli)
aqillahaiman02@gmail.com (Nor Aqillah Aiman Nor Azman)
hakimiali@uitm.edu.my (N.H. Nik Ali)
h.illias@um.edu.my(Hazlee Azil Illias)
ali.salem@kfupm.edu.sa (Ali Ahmed Ali Salem)
aizam.talib@tnb.com.my (Mohd Aizam Talib)
fairouz@uthm.edu.my (Mohd Fairouz Mohd Yousof)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

The grading ring's diameter, thickness, distance from the end of the insulator, and the water droplet parameters, which are size and location, are varied to observe how the electric field distribution is affected. The results from the simulation show that increasing the grading ring's diameter and thickness, the maximum electric field (Emax) observed on the insulator surface will increase by 25 %. Besides that, the performance increases by 20% after placing a grading ring along the high-voltage insulator, while higher positions of the grading ring from the end reduce this performance by 15%. Water droplets at curved surfaces showed a higher electric field, and larger droplets concentrated the electric field more. The grading ring with the lowest electric field was chosen based on these findings, and the electric field along the insulator surface was then simulated. A proper grading ring design is essential for even electric field distribution that mitigates localised surface stress and supports optimal steady state power transmission and distribution performance.

Keywords: Corona discharge, distribution, electric field, FEA software, high-voltage, hydrophobicity, insulator, polymer, transmission

INTRODUCTION

There are two categories of insulators. It can be a high-voltage or a low-voltage insulator. In a transmission system, a high-voltage insulator is an important component that is widely used in electrical infrastructure that is commonly found on overhead transmission lines, which serve as key assets for power delivery across all countries (Mohamad Nasir et al., 2021). Additionally, high-voltage insulators are parts of grid equipment such as transformer bushings for both low- and high-voltage terminals. The insulator is a safety element in transmission that can directly impact the safety of electrical power transmission (Belhouchet et al., 2024; Illias et al., 2021). There are many factors that can affect the electric field distribution and high-voltage insulator performance in the grid system (Ahsan, 2024). The electric field distribution is uniform on the surface of a high-voltage insulator if it is in good and optimum performance. It is crucial for ensuring the reliability and efficiency of the insulation system because it can help in preventing localised overstressing of the material, thereby minimising the risk of electrical breakdowns and enhancing the overall durability and operational stability of the high-voltage equipment (Darmiani et al., 2023; Mao et al., 2022).

Pollution flashovers are a significant problem today, especially in big cities (Salem et al., 2022). However, the high-voltage insulator's performance can be disturbed by nature’s conditions and environment: humidity, the surrounding temperature, environmental conditions, ultraviolet radiation or exposure, electrical stress, and the load carried. These issues must be studied thoroughly to better understand the behaviour of the high-voltage insulator for a better distribution of the electric field and to increase the lifespan of the insulator (Rao et al., 2023). The lack of understanding and research will limit the ability of the insulator manufacturer to design the insulator to its optimum performance. High-voltage

insulator sizes can differ depending on the insulation location and position of the high-voltage transmission tower. Observation can be seen from the transmission tower, which is usually divided into 3 phases: upper, medium, and lower. Each load has its parameters and size to support the insulator in ensuring the conductor does not easily break down due to any fault or surge during the transmission of electrical energy (Ndoumbe et al., 2023).

Hydrophobicity is recognised to be one of the significant issues that have been faced in the electrical transmission system. There are numerous studies and research conducted in previous years to improve the characteristics of the high-voltage insulator. An insulator with a strong hydrophobicity characteristic will increase the electrical transmission quality and efficiency. The polymer-based insulator has become one of the widely used materials in recent years in the electrical transmission system (Huo et al., 2019). However, limited studies and research have been performed to evaluate the importance of the hydrophobicity property of the high-voltage insulator. The hydrophobic characteristic could minimise the adhesion of water droplets on the surface of the high-voltage insulator. It will give impact on the electric field distribution on the surface of the high-voltage insulator. The degradation of the hydrophobic coating that coats the high-voltage surface can lead to flashover or create a conductivity layer for the electricity flowing through the high-voltage insulator (Haque et al., 2021; Jin et al., 2014). A conductivity surface will cause the high-voltage insulator to break down faster if any surge or fault happens in the transmission line.

Flashover pollution occurs when excessively high-voltage discharges occur around the insulator and lead to breakdown of the insulator, allowing the insulator to become a conductor, and any surge or fault can pass through (Salem et al., 2022). The risk of potential leakage currents forming on its surfaces when high-voltage insulators are exposed to different pollution conditions. It can lead to an insulator flashover phenomenon. This phenomenon occurs when the accumulated contaminants on the surface of the insulator, such as dust, salt, and other pollutants, combined with moisture, create a conductive layer that allows current to bypass the insulator.

Consequently, this can result in a sudden and severe discharge and cause a flashover that ruins the integrity and reliability of the high-voltage insulation system. It is important to address and mitigate this issue to maintain the efficiency and safety of electrical power systems and prevent damage to equipment and potential power outages (Bi et al., 2020). Thus, the electric field distribution is important to implement the use of a grading ring in the grid asset that is connected at the end of the insulator, and it also functions as a connector to connect the conductor to the insulator and the insulator to the grid transmission tower. The effects of grading ring size, diameter, thickness, and distance from the end of the insulator to observe its impact on electric field distribution via the simulation are the focus of this study (Ahsan et al., 2024). Additionally, beyond the parameters of the grading ring, this study also aims to conduct simulations to enhance the understanding of the hydrophobic

characteristics of the high-voltage insulator in terms of water droplet sizes and their various locations on the surface of the high-voltage insulator.

This research paper consists of four main sections. The first section, Section I, introduces the project, problem statements, objectives, scope of work, and limitations. Section II explains the simulation model and equations. This section is divided into five sub-sections based on the specific parameters being varied, which are 2.1 Diameter of the grading ring, 2.2 Thickness of the grading ring, 2.3 Distance of the grading ring from the end, 2.4 Location of the water droplet, and 2.5 Size of the water droplet. Section III subsequently presents the computational results and detailed discussions for each of these five parameters to evaluate the overall insulator performance. Section III, the study will compute all the results. Each data point and graph will be discussed in five parts to allow a more uniform discussion and allow the reader to understand the simulation result.

This study contributes to the enhancement of electrical transmission reliability by analysing the geometric and environmental factors that can affect the insulator performance. The main contribution of this research is the quantification of how specific grading ring diameter, thickness, and distance from the termination end directly impact the reduction of maximum electric field (Emax) stress. Furthermore, this paper provides information about the role of hydrophobicity by simulating the impact of various water droplet sizes and locations to demonstrate its significant effect on localised electric field intensity. According to the analysis, an optimised grading ring configuration can improve electric field distribution by 20%, this study offers actionable guidelines for the design of more durable and safer high-voltage insulation systems.

Simulation Model

FEA software is used in this study to analyse the electric field distribution across the surface of high-voltage insulators used in overhead lines and power transmission systems. This software utilises the finite element method. It is an effective computational technique that allows precise modelling and analysis of the electric field behaviour under various conditions. This method aims to represent the electric field distribution accurately to evaluate the performance and functionality of high-voltage insulators in real-world applications. A 3D simulation has been designed to understand the effects of environmental and operational factors on insulator performance. It will contribute to improving design strategies and maintenance methods for enhanced system efficiency and safety (Meetiyagoda et al., 2023). A static condition in which the electric field, Voltage, V, and the constitutive relation are defined in the system as referred to in Equation 1. Equation 2 shows the relationship between the electric displacement distance, D, and the electrical potential, E.

$$E = -\nabla V \quad [1]$$

$$D = \varepsilon_r \varepsilon_0 E \quad [2]$$

V is the overvoltage set in the setting section, ε_r is the material's relative permittivity; this study will use water, with a value of 81. ε_0 is the relative permittivity of the vacuum, and E is the electric field. In this study, the free space charge density is set to zero and reformulates Gauss' Law to derive a variant of Poisson's equation, accordingly, as shown in Equation 3.

$$-\nabla \cdot \varepsilon_r \varepsilon_0 \nabla V = 0 \quad [3]$$

The applied voltage is set to 500 kV to represent the overvoltage amplitude, with the ground voltage set to zero. In the general depiction of the high-voltage system, the metallic components are modelled as equipotential bodies where the internal electric field is zero. The selection of the 500 kV applied voltage, along with the specific material properties (detailed in Table 1) and boundary conditions (detailed in Table 2) were adapted from the validated COMSOL AC/DC module tutorials (COMSOL, 2022). The physical dimensions of the insulator string and the corresponding parameter settings are visually illustrated in Figure 1. These parameters were deliberately chosen to establish a simplified and standardised baseline model. By utilising these established baseline values rather than simulating a highly specific and complex commercial transmission tower environment, this study minimises external electrical and environmental variables. This standardised modelling approach allows the simulation to strictly isolate and accurately quantify the precise impact of geometric variations, specifically the grading ring dimensions and water droplet characteristics, on the electric field distribution.

Table 1
Relative permittivity of different materials used in the simulation

Material	Relative Permittivity	Conductivity
Core rod	5	10^{-5} S/m
Weather shed	3	10^{-12} S/m
Water	8	1 S/m
Air	1	10^{-14} S/m

Table 2
The parameter setting for the simulation

Name	Value and Unit	Description
Vi	500[kV]	Overvoltage amplitude
H	30[mm]	Thickness of the grading ring
D	350[mm]	Diameter of the grading ring
d	150[mm]	Distance of the grading ring from the end

After applying all the following data to the parameters set, the 3D high-voltage insulator was built, and a simulation was performed based on the desired values of three different observed parameters. In this study, the thickness, H, the diameter, D, and the location

of the grading ring from the end of the insulator that connects the grading ring, d, are varied to observe the effect of different values applied to the high-voltage insulator and how the electric field produced by the simulation is distributed.

After completing the first part of the simulation, an additional ellipse drawing is added to the drawing to perform the second and third parts of the simulation to simulate a water droplet and analyse the effects of the hydrophobicity towards the insulator. Figure 2 provides a detailed flowchart outlining the methods used to perform the simulation. Figure 3 shows an example of the designed and simulated 3D model of the grading ring with the high-voltage insulator.

Diameter of the Grading Ring

The diameter of the grading ring varies within the industry based on several factors, including the installation location, the load it is designed to carry, and its specific functionality. Several samples and industrial sizes of the insulator with a grading ring are used to improve and enhance the insulator's performance and distribute a more uniform and balanced electric field on the high-voltage insulator's surface to reduce the insulator's stress level (Tsinghua University, 2020). The industry usually uses the diameter 700-900 mm (Ndoumbe et al., 2023), but the 700-1100 mm range data is simulated in this simulation for comparison purposes, as tabulated in Table 3.

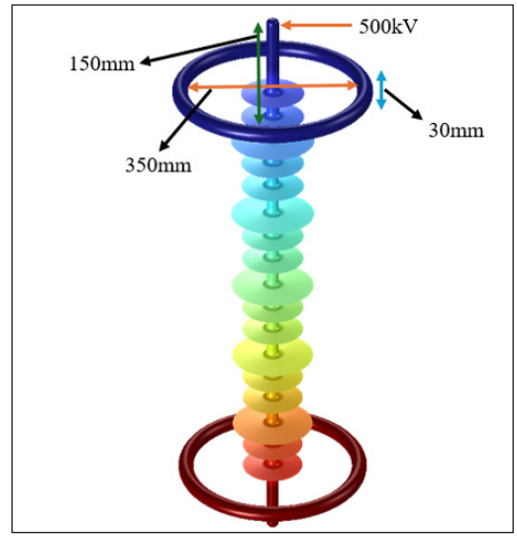


Figure 1. The pictorial dimension of the insulator string, together with the parameter setting

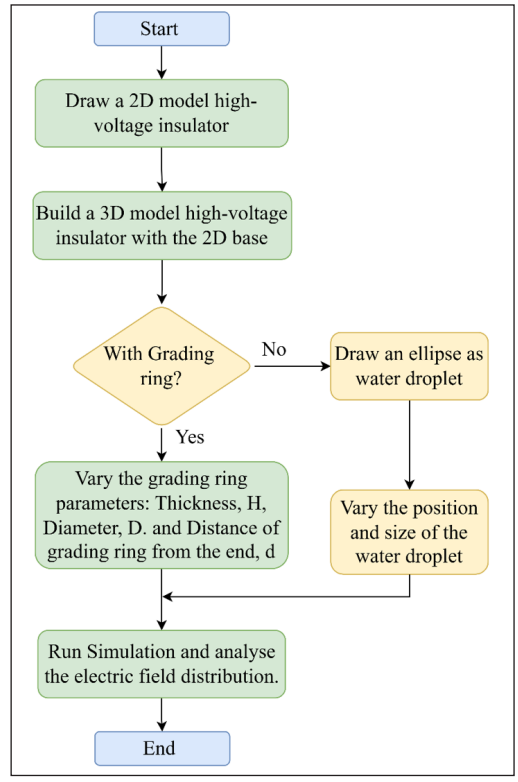


Figure 2. The flowchart of the method applied to perform the simulation

The Thickness of the Grading Ring

On the other hand, an observation is made of the size of the ring itself, which is measured based on the thickness of the grading ring. The thickness is one of the critical elements in ensuring the grading ring can perform optimally and insulate the conductor from any fault or surge. The industry usually uses a thickness of 100 mm (Ndoumbe et al., 2023). To simulate a comparison of the insulator's performance, as shown in Table 4, five different data sets (100 to 500 mm) on the thickness of the grading ring are simulated. The thickness values are varied from 100 mm to 500 mm to make the comparison clearer as small differences in thickness, such as 100 mm to 150 mm, can make the variations in the results difficult to see.

Distance of the Grading Ring from the End

Generally, a grading ring is usually placed 150 to 200 mm from the end of the insulator according to the industry standard (Zhang et al., 2022). However, the range in this research is wider, which is from 150 mm to 350 mm, to allow a deeper investigation into the effects of varying the grading ring placement on the insulator's performance. In this study, the performance of the insulator will be evaluated by comparing five different values of the distance of the grading ring from the end of the high-voltage insulator, as outlined in Table 5. It will provide valuable insights into optimal design parameters to enhance efficiency and reliability in high-voltage applications. The simulation data will be analysed to understand the impact of these variations on the electric field distribution and the overall effectiveness of the insulator.

Table 3
Grading ring diameter (mm) data range

Grading Ring	Diameter Size (mm)
A	700
B	800
C	900
D	1000
E	1100

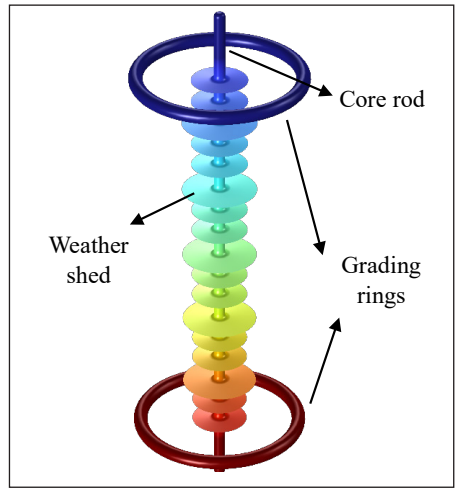


Figure 3. The designed and simulated 3D model of the grading ring with the high-voltage insulator

Table 4
Grading ring thickness (mm) data range

Grading Ring	Thickness (mm)
A	100
B	200
C	300
D	400
E	500

Location of the Water Droplet

A superhydrophobic coating is needed for an insulator to facilitate optimal water drainage from its surface (He & Guo, 2021). In high-voltage applications, this is important where minimising water retention is essential to maintain operational integrity and safety (M'Hamdi et al., 2016). In this research, the location of the water droplet is also being studied since it is one of the factors that can disturb the distribution of the electric field. There are three different places of water droplets to observe the outcome from the simulation because the water droplets from the environment can decrease the hydrophobicity element or characteristic of the insulator. The first location of the water droplet is in the centre of the insulator shed. Second, at the end of the insulator shed, and lastly, at the end of the insulator shed. The water droplets at each location have the same size to ensure an accurate result. Figure 4 shows how the location allocates the water droplet on the insulator shed. Simulations are conducted at these three locations to observe how the electric field is distributed on the high-voltage insulator.

Table 5
Distance of the grading from the end (mm) data range

Grading Ring	Distance of the Grading Ring from the End (mm)
A	150
B	200
C	250
D	300
E	350

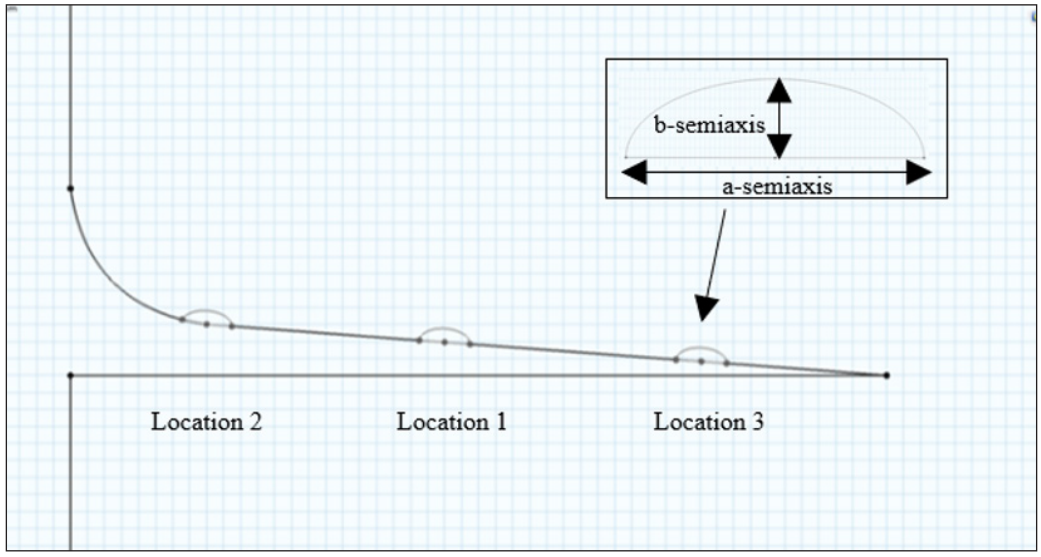


Figure 4. The three locations of the water droplets

Size of the Water Droplet

The water droplet size and shape can significantly affect the electric field distribution on the high-voltage insulator (Shrimathi & Mondal, 2021). In this study, water droplets are

represented using elliptical geometries. This specific shape is a standard modelling simplification widely utilised in literature to accurately approximate the gravitational deformation and surface tension of droplets resting on inclined insulator weather sheds (Ali & Hackam, 2008). Furthermore, the selected dimensions from 1.2 mm to 1.8 mm in length and the specific locations tested, which were middle, curve, and end of the shed were chosen because statistical distributions of condensation and rain accumulation on silicone rubber demonstrate that the vast majority of discrete droplets stabilise in this sub-3 mm range before sliding occurs (Guedri et al., 2024; Jiang et al., 2019). This reflects practical, real-world accumulation patterns typically observed during varied environmental wetting conditions. This approach allows for a realistic and comprehensive analysis of localised field enhancement. The water droplet size is varied by changing the values in the setting under the a-semiaxis and b-semiaxis of the ellipse as per Table 6. The a-semiaxis controls the lengthwise of the water droplet for the semimajor axis. In contrast, the b-semiaxis controls the radius of the water droplet for the semi-minor axis, as illustrated in Figure 5. The cross-section plot method and the distribution of the electric field are presented in Figure 5. For the simulations involving the water droplets, the droplet domain is modelled as a solid homogeneous medium. The relative permittivity of water ($\varepsilon_r = 81$) is assigned to the entire volumetric domain of the

Table 6
The length and width (mm) of water droplets

Size	Length, a-semiaxis (mm)	Width, b-semiaxis (mm)
A	1.2	0.5
B	1.5	0.8
C	1.6	0.9
D	1.8	1.2

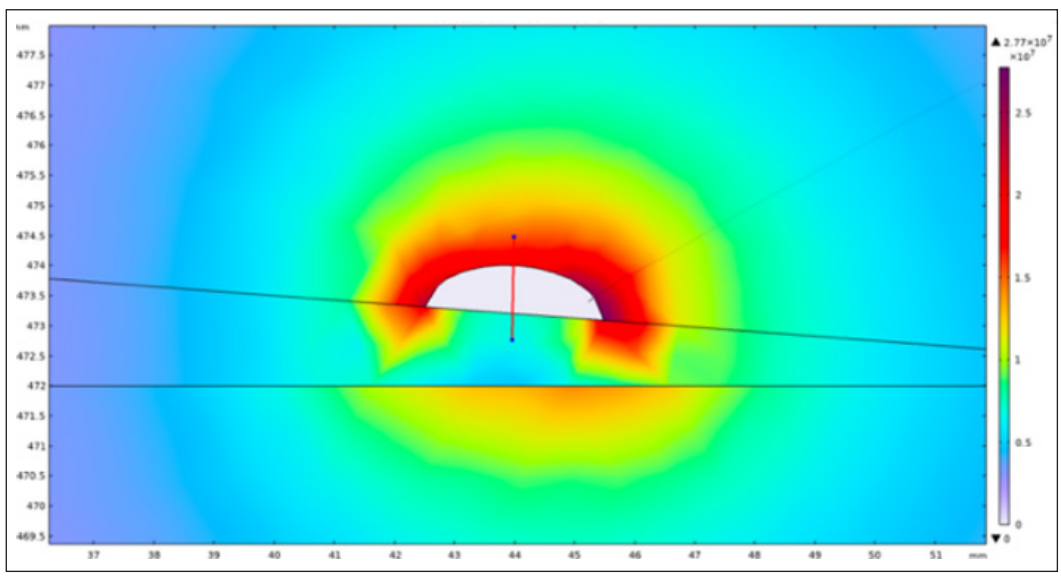


Figure 5. The illustration of the cross-section plot method and how an electric field is presented

ellipse, rather than treating the interior as a void. The surrounding medium is defined as air (ε_r =1), establishing a distinct dielectric interface at the boundary. The electric field distribution is then extracted across these regions using COMSOL's Electrostatics interface, which accurately computes the field dampening within the bulk water and the field enhancement at the boundaries.

RESULT AND DISCUSSIONS

Diameter of the Grading Ring

This section, it presents and discusses the simulation results on various grading ring design parameters and the presence of water droplets on the electric field distribution of high-voltage insulators. The analysis is organised into five sub-sections based on the parameters analysed. Figure 6 illustrates how varying the grading ring diameter affects the electric field distribution across the insulator surface. The data indicate that the diameter impacts field behaviour specifically, increasing the diameter appears to enhance the distribution potential, which may help mitigate electric stress on the insulator.

As the diameter increases, the field concentration on the insulator surface is altered, which is critical for minimising breakdown risks. Conversely, a smaller diameter creates localised areas of higher intensity, whereas the larger ring diameter, which is 1100 mm, demonstrated the most significant potential for field distribution compared to the other tested sizes. Therefore, within the scope of this study, a larger grading ring is shown to enhance protection and ensure maximum insulation performance against surges. It will also lower the probability of insulation failure under high-voltage conditions.

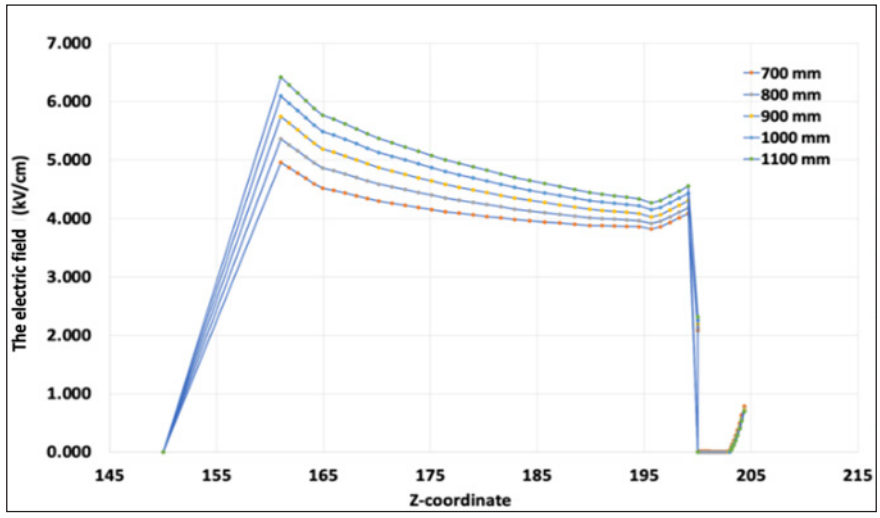


Figure 6. The electric field of the simulation is based on the different diameters of the grading ring

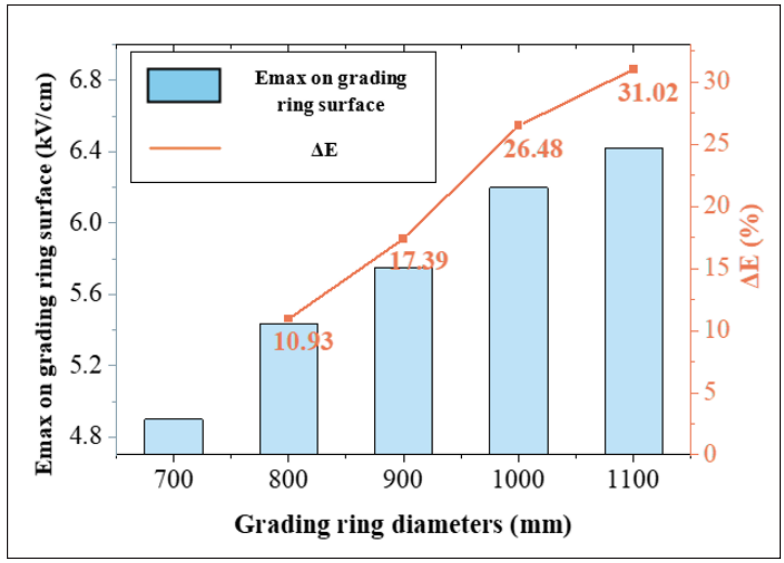


Figure 7. Emax in the grading ring is based on the different diameters of the grading ring

As presented in Figure 7, this study found that increasing the grading ring diameter within the 700 - 1100 mm range resulted in a 31.02% increase in the maximum electric field (Emax) on the grading ring. It is crucial to clarify that this Emax refers to the localised field enhancement on the surface of the metallic grading ring itself, not on the insulator weather sheds. While a larger diameter effectively redistributes the electric field away from the critical termination points of the polymer insulator, thereby improving insulation performance, the grading ring simultaneously becomes a site of higher localised field concentration. Current industrial standards typically range from 700 to 900 mm to prevent the ring itself from becoming a source of high-intensity field stress (Ndoumbe et al., 2023). This localised enhancement is expected since the larger metallic grading ring attracts and concentrates electric field lines around its curvature. Therefore, the observed increase in Emax does not indicate deterioration of the insulation system; rather, it reflects the intentional transfer of electric field stress from the vulnerable insulator surface to the grading ring, which is specifically designed to withstand higher localised electric field intensity.

Thickness of the Grading Ring

Figure 8 illustrates the electric field distribution across the simulation for the different grading ring thicknesses. While conventional electrostatic theory suggests that increasing the thickness of a conductor should reduce the surface electric field by increasing the radius of curvature, the simulation results in Figure 9 demonstrate a localised over-grading effect specific to this 3D geometric model. Specifically, the 100 mm thickness ring produced the most favourable outcome, maintaining the lowest maximum electric field (Emax) in the

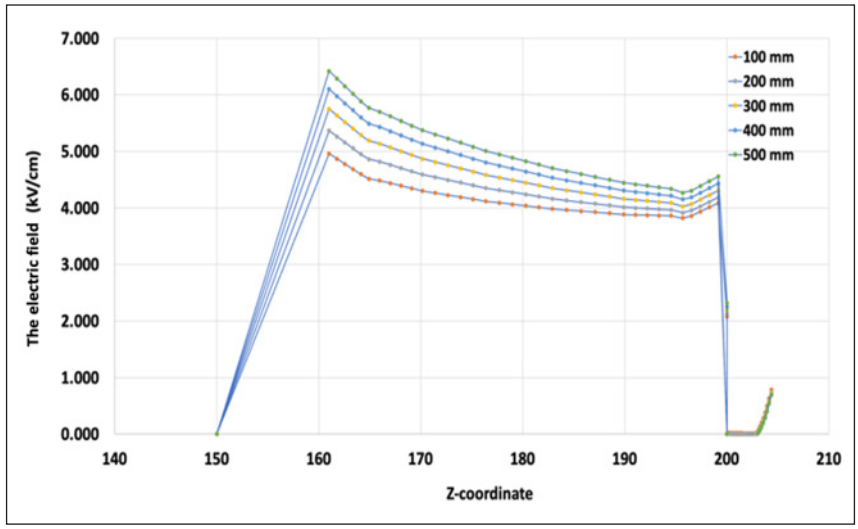


Figure 8. The electric field of the simulation is based on the different thicknesses of the grading ring

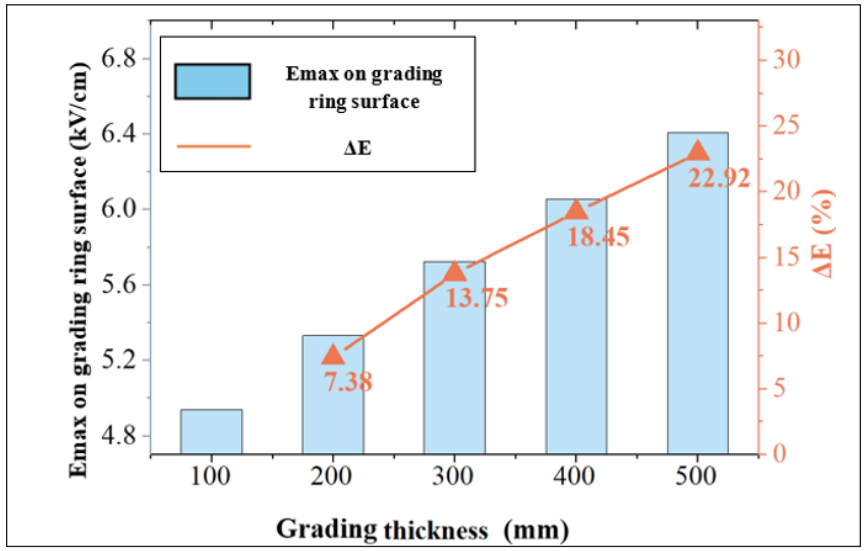


Figure 9. ΔE based on the different thicknesses of the grading ring

grading ring of 4.960 kV/cm on the grading ring. As the thickness was increased to 500 mm, the grading ring Emax increased to 6.420 kV/cm, representing a 22.74% increase in localised field intensity. Although increasing the grading ring thickness increases the effective curvature radius, the modification of the ring geometry simultaneously alters the electric field distribution and charge accumulation pattern within the specific insulator–electrode configuration used in the simulation.

As the ring thickness increases, the grading ring captures a larger portion of the electric field and shifts the electric stress away from the polymeric insulator surface. Consequently, localised field enhancement may occur in certain regions of the grading ring itself, even though the electric field stress along the insulator surface becomes more uniform and reduced. Therefore, the increase in E_{max} does not necessarily indicate poorer insulation performance but rather reflects a transfer of electric field concentration from the insulator surface to the metallic grading ring, which is intentionally designed to withstand higher localised electric field stress. This contradictory behaviour is caused by the geometric boundary conditions of the simulation: a massive increase in ring thickness significantly alters the local capacitance of the system, dragging the equipotential lines too close to the critical regions and concentrating electric stress on the expanded ring surface rather than the insulator sheds. (Ndoumbe et al., 2023).

Distance of the Grading Ring from the End

The distance of the grading ring from the end of the high-voltage insulator can significantly influence the electric field distribution across the insulator's surface. The "spacing" or "clearance" of the grading ring shapes the electric field profile, ensuring effective control of the high-voltage insulator. If the grading ring is positioned too close to the termination point, the insulator is more likely to experience breakdown due to the high electric field stress concentrated at that point. Conversely, placing the grading ring at an optimal position on the insulator promotes even distribution of the electric field, reducing the risk of breakdown. Observations confirm that the placement of the grading ring is a critical factor in managing electric field distribution on high-voltage insulators. An effective approach to mitigate surface stress on the high-voltage insulator is positioning the grading ring at an appropriate distance between the conductor and the insulator's line end. Figure 10 shows that as the grading ring is positioned further from the termination point, its ability to optimise electric field redistribution decreases. From the graph, the most effective placement is 150 mm, as it continuously promotes an even distribution of electric field stress across the surface of the high-voltage insulator.

As illustrated in Figure 11, placing the grading ring at a 150 mm distance results in a localised electric field concentration of 4.261 kV/cm on the ring itself, which correlates with the most effective surface stress reduction on the insulator compared to the other configurations. This finding is consistent with industrial research identifying an optimal placement height between 150 mm and 200 mm (Zhang et al., 2022). Conversely, the 300 mm position shows the least effective performance, registering a localised field of 3.557 kV/cm on the ring. This shows a difference of 20.46% in localised field concentration, indicating poorer field redistribution along the insulator sheds. Thus, the position of the grading ring must be precisely configured to minimise localised field enhancement and ensure optimal performance.

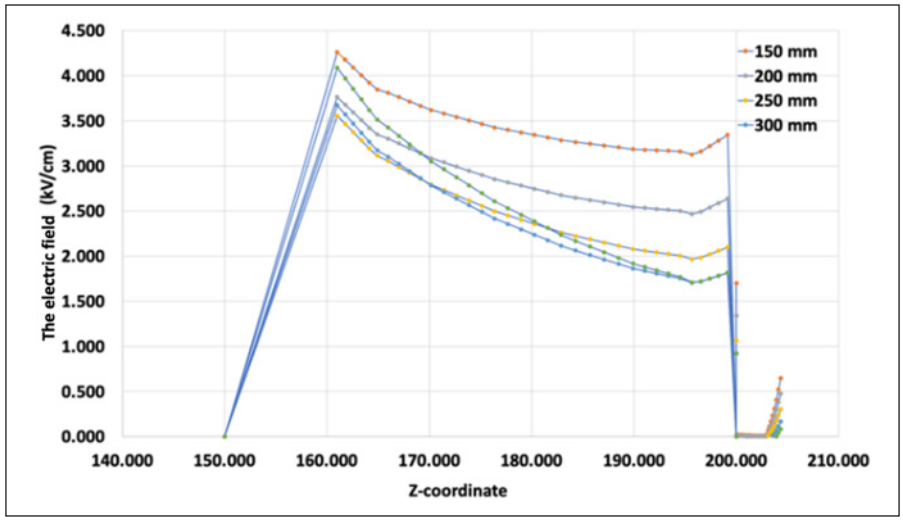


Figure 10. Electric field of the simulation at different distances from the grading ring

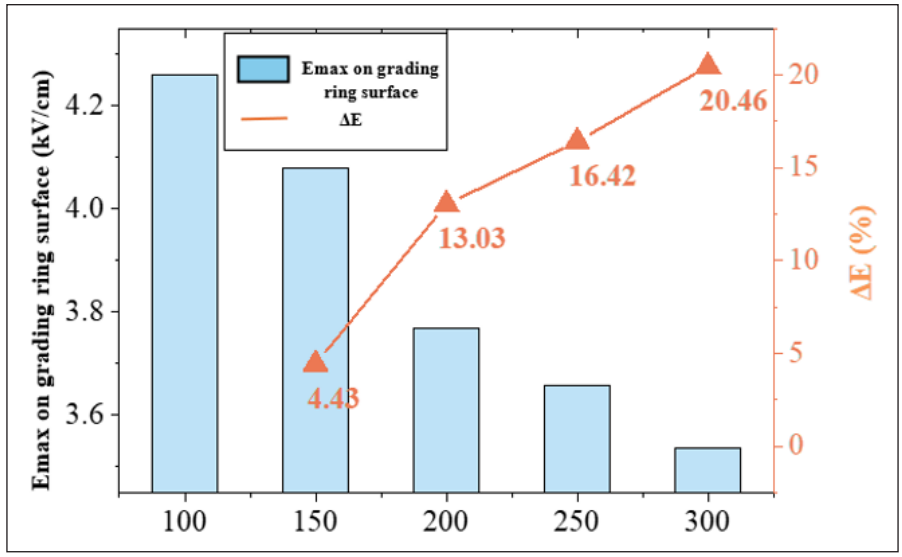


Figure 11. The ΔE is based on the different distances of the grading ring

Location of the Water Droplet

For the case of water droplets, based on the findings of the electric field across water droplets under various grading ring characteristics, the grading ring with the lowest electric field was selected, and the electric field along the insulator surface was simulated using the chosen grading ring. Figures 12 and 13 illustrate the 3D and the electric field distribution along the insulator surface under the original and selected grading ring.

The results in Figures 12 and 13 show that the chosen grading ring reduced the electric field E_{max} on the insulator surface from 7.12 kV/cm to 5.68 kV/cm (20.22%) as compared to the original ring. The location of the water droplet can significantly disturb the aid performance of the high-voltage insulator due to its hydrophobicity capability and its energy discharge on the surface coated with a layer of hydrophobic material (Ye et al., 2021). The hydrophobicity of the high-voltage insulator surface can minimise the potential of the water droplet to stay on the surface of the high-voltage insulator. Figure 14 shows the simulation results of three different locations of water droplets. The three locations will produce three different results regarding the potential of the electric field distribution for each designed location. Changes in the relative permittivity of the medium—such as the presence of water—can alter the electric field distribution, potentially increasing local field intensity at critical points.

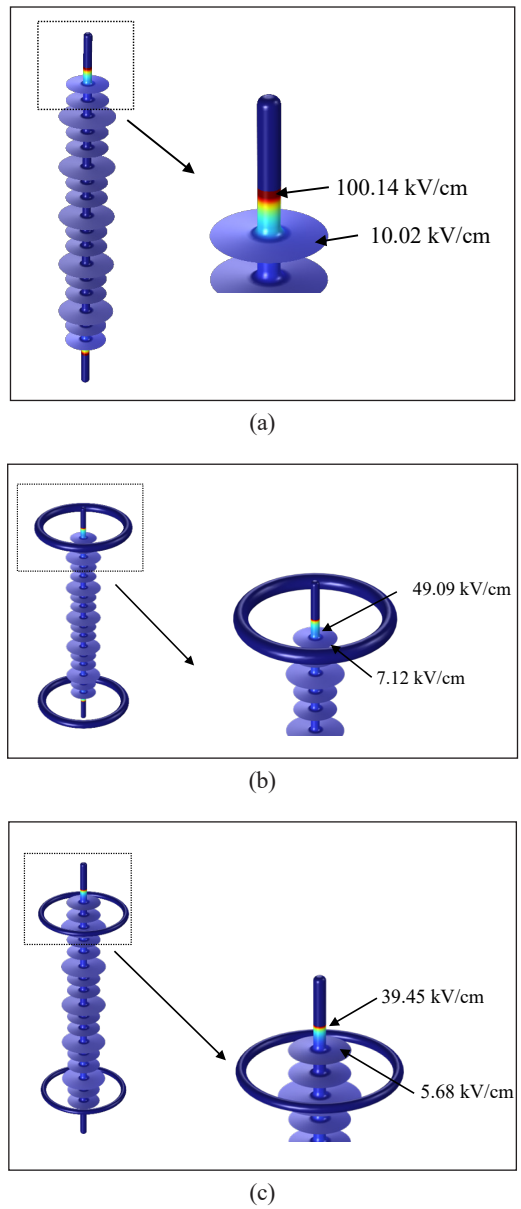


Figure 12. 3D electric field distribution, (a) without grading ring; (b) original grading ring; and (c) selected grading ring

In this case, the water droplet will have a higher permittivity than the air outside the high-voltage insulator. In comparison, the air will have the lowest relative permittivity. From the graph in Figure 14, a conclusion can be made that each location has its own withstand value of the electric field. The second location of the water droplet, which is on the curved side of the water droplet, can withstand the highest electric field. This is due to the electric field stress, highest on the curve or sharp edges, where it was concentrated and focused on that area. The different values that can be observed indicate a difference in relative permittivity. For example, consider the second location. The electric field value initially drops from 1.49×10^7 Kv/cm to 0 Kv/cm, indicating a transition from

air to the interior of the water droplet. Subsequently, the electric field increases from 0 Kv/cm to 2.12 × 10^7 Kv/cm, which suggests a transition from within the water droplet to the surface of the high-voltage insulator. This behaviour demonstrates that a higher relative permittivity medium, such as water, reduces the electric field intensity within it, thus confirming that as the relative permittivity increases, the electric field strength in that region decreases.

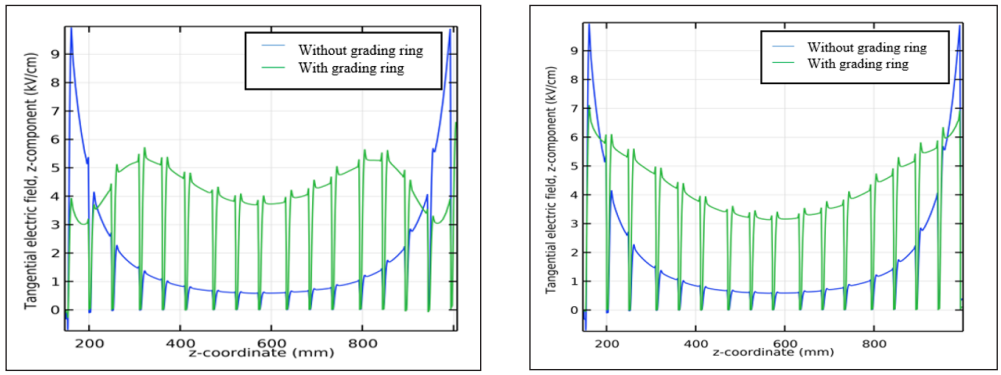


Figure 13. Comparison of the electric field distribution: (a) without the grading ring and with the original grading ring; (b) without the grading ring and with the selected grading ring

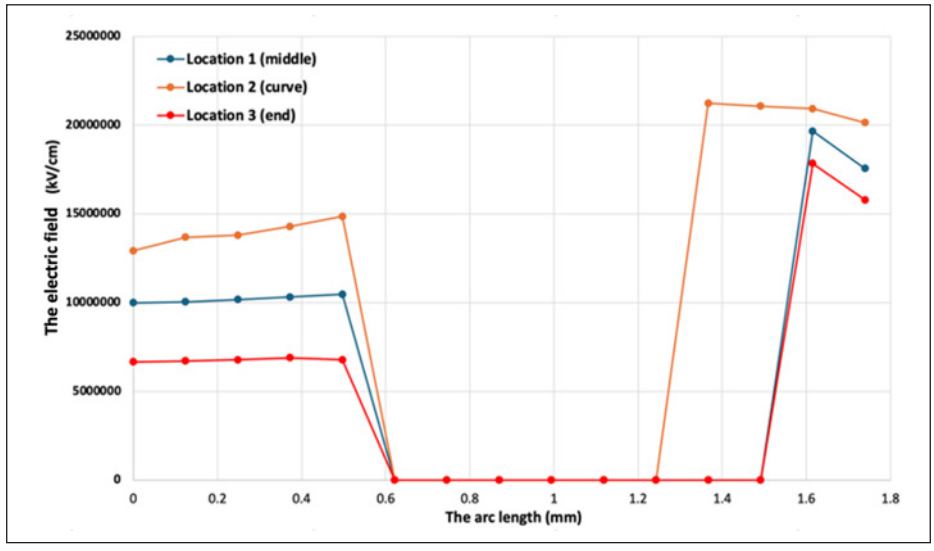


Figure 14. Simulation results of the electric field distribution across three different droplet locations (middle, curve, and end of the shed). The sudden drops to zero represent the highly dampened internal field of the water droplets due to water's high relative permittivity

From the graph output in Figure 14, the data presents the electric field profile as it traverses from the surrounding air, across the water droplet, and into the insulator. The sharp drop in the electric field to near-zero within the droplet is not due to the region being a physical void; rather, it is a direct electrostatic consequence of water's high relative permittivity. The bulk water heavily polarises, acting as a dielectric shield that drastically dampens the internal electric field. The electric field then sharply rises again at the transition boundary from the water droplet to the surface of the high-voltage insulator.

Size of the Water Droplet

The water droplet that falls from the rain will be attached to the surface of the insulator and can be considered one of the leading factors that can affect the performance of reducing and minimising the electric field stress. The contact angle of the water droplet and its sizing should be studied more extensively to ensure a better understanding of high-voltage performance (Liu & Du, 2014). Based on Figure 15, it can be concluded that varying sizes of water droplets significantly affect the electric field distribution on the surface of the high-voltage insulator.

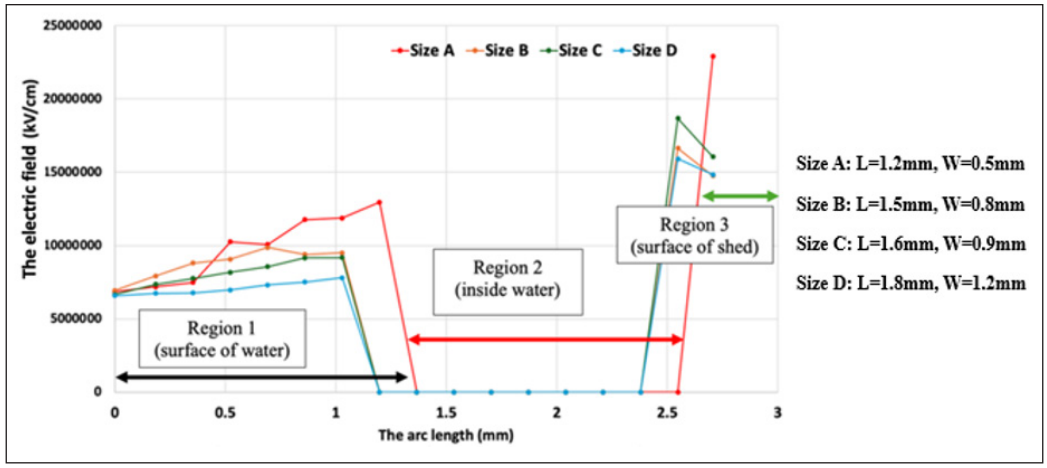


Figure 15. Electric field distribution simulation based on varying water droplet sizes. Region 1 represents the localised stress at the air-water interface, Region 2 shows the dampened internal droplet field, and Region 3 highlights the stress concentration at the boundary between the droplet and the polymer shed

The influence of water droplet location on electric field distribution can also be observed by varying the water droplet sizes. Although relative permittivity plays a key role, this analysis is focusing in examining how different sizes of droplets can affect the field behaviour. According to the result, Size A, which is the smallest droplet among all sizes, has the highest electric field, while Size D, which is the biggest droplet, has less ability to withstand an electric field. The outcomes from the simulation indicate that different

droplet sizes have different electric field distribution patterns on the high-voltage insulator surface. Size A is used as a reference to draw a basic conclusion from the simulation results. The transitions in the electric field observed in Figure 15 directly correspond to the geometric boundaries defined within the simulation. The initial enhancement of the electric field (Region 1) occurs at the air-water interface, where the field concentrates heavily on the outer surface of the droplet. Following this, the electric field drops to 0 kV/cm (Region 2). In this simulation, Region 2 was deliberately modelled as a void as a computational simplification to eliminate internal dielectric dampening. This approach isolates the parameters to strictly observe the worst-case localized stress at the boundaries. Finally, the abrupt spike (Region 3) marks the critical geometric boundary between the water droplet and the polymer shed surface. As proven by the results, the high Emax is primarily concentrated at these exposed surfaces (Region 1 and Region 3), demonstrating that varying water droplet sizes significantly alter the boundary conditions and create severe localised stress at the interfaces.

CONCLUSION

In conclusion, this study focuses on the analysis of the electric field distribution on high-voltage insulators based on simulation. The results show that varying parameters, which are the thickness, diameter, and positioning of grading rings, can give impact on the insulator's performance. As the diameter and thickness were increased, the E_max increased by 31.02% and 22.74%, respectively. This suggests that excessive sizing can lead to unintended stress concentrations. For the distance of the grading ring from the end, the most effective position is placing the grading ring 150 mm from the termination end. Besides that, the study shows that water droplet size and location affect the electric field distribution on the surface of the high-voltage insulators. Smaller water droplets improved the electric field distribution, which helps mitigate steady-state localised corona stress on the insulator surface. Thus, this proves that the main reason hydrophobicity characteristics play a significant role in minimising the concentration of the electric field on the surface of the high-voltage insulator. This underscores the need for superhydrophobic coatings to prevent water accumulation on insulator surfaces. This research highlights the importance of the analysis of high-voltage insulator design to get the most effective performance. Each design must be simulated and analysed to ensure optimal performance and reliability according to the specific requirements of the electrical system. One of the most important elements to be focused on is the precision in design to ensure it safeguards the industry's electrical energy transmission and distribution. For future recommendations and improvements for the study, more real data is needed to collect more detailed industrial for the grading ring of the high-voltage insulator.

ACKNOWLEDGEMENT

A special thanks to everyone who contributed to the completion of this research study. The support for several key institutions as a collaboration and resource is appreciated. A special appreciation to the Faculty of Electrical Engineering, UiTM, Universiti Malaya, and Applied Research Centre for Metrology, Standards and Testing (KFUPM) for providing laboratory facilities and a good research environment that eased the journey of completing this research. I extend my deepest thanks to the co-authors for the invaluable guidance, discussions, and unwavering support throughout this project. I also want to acknowledge the technical assistance provided by the laboratory staff, whose expertise and support were essential for the experimental work. Finally, thanks to my family and friends for their support, encouragement and understanding throughout this research journey. Their moral support contributed in the successful completion of this research.

REFERENCES

Ahsan, M. (2024). Comprehensive analysis of insulator performance in high-voltage transmission systems: Implications for efficient power transfer. *Journal of Advanced Research in Applied Mechanics, 115*(1), 117-130. https://doi.org/10.37934/aram.115.1.117130

Ahsan, M., Nor, M., Zainal, Z. B., & Khalil, I. U. (2024). Parameter design optimisation of grading ring based on electric field analysis through response surface methodology. *e-Prime: Advances in Electrical Engineering, Electronics and Energy*, Article 100569. https://doi.org/10.1016/j.eprime.2023.100569

Ali, A., & Hackam, R. (2008, May 27-29). Influences of water droplet size and pollution layer permittivity on the electric field and potential distribution on a polluted insulator surface. In *Proceedings of the 6th International Conference on Electrical Engineering (ICEENG 2008)*.

Belhouchet, K., Zemmit, A., Bayadi, A., Ouchen, L., & Zorig, A. (2024). A novel application of artificial intelligence technology for an outdoor high-voltage composite insulator. *Measurement, 238*, Article 115372. https://doi.org/10.1016/j.measurement.2024.115372

Bi, J., Hao, Y., Yang, L., Zheng, Y., & Li, L. (2020). Impact of hydrophobicity on wetting characteristics of composite insulators. *IEEE Access, 8*, 159316-159323. https://doi.org/10.1109/ACCESS.2020.3019585

COMSOL. (2022). *Introduction to AC/DC module* (Version 6.1) [PDF]. https://doc.comsol.com/6.1/doc/com.comsol.help.acdc/ACDCModuleIntroduction.pdf

Darmiani, M., Shariatinasab, R., Najafi, H., Ghayedi, M., & Rodrigues, E. M. G. (2023). Analysis of electrical characteristics of composite insulators with the presence of an optimum layer of ZnO microvaristors. *International Journal of Energy Research, 2023*, Article 4160125. https://doi.org/10.1155/2023/4160125

Guedri, A., Al-Gailani, S. A., Salem, A. A., & Al-Ameri, S. M. (2024). A finite element with a statistical analysis study to investigate the electrical performance of composite insulators under water droplet impact. *Scientific Reports, 14*, Article 2145. https://doi.org/10.1038/s41598-024-52643-y

Haque, S. M., Ardila-Rey, J. A., Umar, Y., Mas'ud, A. A., Muhammad-Sukki, F., Jume, B. H., Rahman, H., & Bani, N. A. (2021). Application and suitability of polymeric materials as insulators in electrical equipment. *Energies, 14*(10), Article 2758. https://doi.org/10.3390/en14102758

He, H., & Guo, Z. (2021). Superhydrophobic materials used for anti-icing: Theory, application, and development. *iScience, 24*(11), Article 103357. https://doi.org/10.1016/j.isci.2021.103357

Huo, F., Zhang, P., Yu, Y., Liu, Q., Chu, L., & Wang, X. (2019). Electric field calculation and grading ring design for a 750 kV four-circuit transmission line on the same tower with six cross-arms. *The Journal of Engineering, 2019*(16), 3155-3159. https://doi.org/10.1049/joe.2018.8501

Illias, H. A., Mokhlis, H., Chean, Y. S., Ariffin, A. M., Ling, L. Y., & Yousof, M. F. M. (2021). Corona ring design impact on the electric field distribution surrounding an insulator string. *IEEE Transactions on Dielectrics and Electrical Insulation, 28*, 1-4. https://doi.org/10.1109/TDEI.2021.009876

Jiang, X., Zhang, Z., Zhang, J., Hu, J., & Zhang, Y. (2019). Dynamic behaviour of droplets and flashover characteristics for CFD and experimental analysis on SiR composites. *IEEE Access, 7*, 8095-8104. https://doi.org/10.1109/ACCESS.2018.2889758

Jin, H., Wang, X., Zhang, H., Wu, Y., & Cheng, X. (2014). Flashover characteristics of discrete water droplets on the surface of superhydrophobic silicone rubber. *IEEE Transactions on Dielectrics and Electrical Insulation, 21*(4), 1718-1725. https://doi.org/10.1109/TDEI.2014.004281

Liu, Y., & Du, B. (2014). Energy eigenvector analysis of surface discharges for evaluating the performance of a polymer insulator in the presence of water droplets. *IEEE Transactions on Dielectrics and Electrical Insulation, 21*(6), 2438-2447. https://doi.org/10.1109/TDEI.2014.004516

M'Hamdi, B., Teguar, M., & Mekhaldi, A. (2016). Optimal design of corona ring on HV composite insulator using PSO approach with dynamic population size. *IEEE Transactions on Dielectrics and Electrical Insulation, 23*(2), 1048-1057. https://doi.org/10.1109/TDEI.2015.005383

Mao, M., Chen, Z., & Shao, J. (2022). Increasing the electric field uniformity in the insulator under pulse excitation by varying its conductivity. In *Proceedings of the 2022 International Applied Computational Electromagnetics Society Symposium (ACES-China)* (pp. 1-4). IEEE. https://doi.org/10.1109/ACES-China56081.2022.10064880

Meetiyagoda, S. H., Galagoda, R. W. N. W. M. R. O. K. B., Dissanayake, T. D. T. L., Fernando, M. A. R. M., & Abeygunasekara, W. L. (2023). Electric field analyses on high-voltage insulator surfaces under different water droplet behaviours. In *Proceedings of the 17th IEEE International Conference on Industrial and Information Systems (ICIIS)* (pp. 394-399). IEEE. https://doi.org/10.1109/ICIIS58898.2023.10253496

Mohamad Nasir, N. A. F., Ab Kadir, M. Z. A., Osman, M., Abd Rahman, M. S., Ungku Amirulddin, U. A., Mohd Nasir, M. S., Zaini, N. H., & Nik Ali, N. H. (2021). Impact of earthing system designs and soil characteristics on tower footing impedance and ground potential rise: A modelling approach for sustainable power operation. *Sustainability, 13*(15), Article 8370. https://doi.org/10.3390/su13158370

Ndoumbe, J., Nkouetcha, E. T., & Emboueme, J. G. O. (2023). FEM analysis of the electric field and potential of various types of insulators under uniform pollution. *International Journal of Research and Review, 10*(8), 1061-1068. https://doi.org/10.52403/ijrr.202308134

Rao, K. D., Rajamani, P., Krishna, B., & Nirgude, P. M. (2023). Artificial pollution testing of HVDC insulators. *Power Research: A Journal of CPRI, 19*(1), 23-29. https://doi.org/10.33686/pwj.v19i1.1110

Salem, A. A., Lau, K. Y., Abdul-Malek, Z., Zhou, W., Al-Ameri, S., Al-Gailani, S. A., & Rahman, R. A. (2022). Investigation of high-voltage polymeric insulators' performance under wet pollution. *Polymers, 14*(6), Article 1236. https://doi.org/10.3390/polym14061236

Shrimathi, H. P., & Mondal, M. (2021). Investigation of uniform and non-uniform water droplets on different configurations of silicon rubber composite insulators subjected to AC electric field stress. In *Proceedings of the 3rd International Conference on Energy, Power and Environment (ICEPE 2020)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICEPE50861.2021.9404509

Ye, Z., Zhao, Y., Liu, P., Zhang, X., Liu, C., & Peng, Z. (2021). Optimisation design on grading rings of composite insulator for 500 kV four-circuit transmission lines. In *Proceedings of the 2021 International Conference on Power System Technology (POWERCON)* (pp. 1366-1369). IEEE. https://doi.org/10.1109/POWERCON53785.2021.9697606

Zhang, X., Zhang, X., Sun, Z., & Wang, H. (2022). Research on electric field distribution and voltage equalisation characteristics of tensile insulator strings in a 750 kV substation. *Journal of Physics: Conference Series, 2254*(1), Article 012011. https://doi.org/10.1088/1742-6596/2254/1/012011

Pertanika J. Sci. & Technol. 34 (4): 2097 - 2126 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Prediction of Lung Diseases Using Deep Learning Techniques: A Review

Sheikh Afaan Farooq* and Aleem Ali

Department of Computer Science and Engineering, University Institute of Engineering (UIE), Chandigarh University, Gharuan, Kharar, 140413 District Sahibzada Ajit Singh Nagar (Mohali), Punjab, India

ABSTRACT

Lung disorders remain at the centre of the burden affecting health on a global scale, and long-term respiratory complications caused by COVID-19 or other pulmonary disorders compound the burden. The accurate, timely identification of these disorders can improve the treatment outcome and diminish the number of fatalities. In recent years, the combination of deep learning techniques and feature selection methods has shown great promise in improving the accuracy and efficiency of medical imaging systems in diagnosis. This paper reports a systematic comparative review of the recent deep learning approaches to lung disorder diagnosis. The numerous works based on model architecture, feature selection strategies, datasets, performance metrics and clinical limitations are compared. In contrast to the current surveys that are largely focused on only a single disorder or a specific architecture, this review compares architectures, hybrid architectures and feature selection frameworks for the various lung disorders in a unified manner. Different models, such as DenseNet-121, Xception, InceptionV3, CNNs, Vision Transformers, MobileNet, RVCNet, EfficientNetB0, CNN-LSTM, XGBoost, Random Forest, and Decision Trees, are investigated. The results show that lightweight architectures like MobileNet and RVCNet can generally achieve better efficiency and accuracy, while more complex architectures and hybrid CNN-LSTM architectures can enhance the predictive performance. However, there are several challenges that are not yet addressed, such as mislabelling of public datasets, overfitting, lack of generalisation to unseen clinical data, lack of interpretability of complex architectures, and lack of ability to handle patients with pre-existing lung abnormalities. Also, multi-label classification systems are yet weak for overlapping pulmonary disorders.

ARTICLE INFO

Article history:
Received: 14 October 2025
Accepted: 03 June 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.07

E-mail addresses:
saf95afaan@gmail.com (Sheikh Afaan Farooq)
aleem.e12948@cumail.in (Aleem Ali)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

These challenges underline the need for trusted datasets, easily interpretable hybrid frameworks and clinically usable, intelligent diagnostic systems for future works in lung disorder detection.

Keywords: CT scans, deep learning models, feature selection techniques, hybrid models, lung disorders, X-rays

INTRODUCTION

Lung Disease is a common problem globally, and its severity has worsened due to the excessive usage of smoking, alcoholic consumption, and dirty air quality Index. As per the WHO (World Health Organisation), around 3.9 million individuals lose their lives to lung disorders like chronic-obstructive pulmonary disorder (COPD), lung cancer, pneumonia, and Tuberculosis (TB) in the world. The COVID-19 disorder-related pandemic may be slacking off, but its effect remains and are a cause of a serious respiratory health issue. According to a study conducted by the Forum of International Respiratory Societies (FIRS), there are five respiratory-related issues that act as the most common and deadly issues of disease and fatality globally; these disorders are COPD, asthma, pneumonia, TB and lung cancer (Aykanat et al., 2020). The five deadly disorders specified above have resulted in billions of deaths worldwide in previous years, with only one disease, COPD, causing the deaths of three and a half million people in 2021 alone globally (Priyadarsini et al., 2023). The Spanish flu of 1918, which emerged as a devastating pandemic and the COVID-19 pandemic of 2019 are both related to respiratory diseases and led to the deaths of millions of people throughout the globe. Asthma, pneumonia, lung cancer, and TB are among the various lung conditions that affect individuals worldwide. The primary condition that prevents the lungs from working properly is respiratory disease. Mostly, there are three main types of lung disorders (Priyadarsini et al., 2023).

Airway Diseases

The airways or tubes that transport oxygen and other gases into and out of the body's lungs are impacted by several wide-ranging, connected illnesses (Mohammed et al., 2024). In most cases, they restrict the airway connections with the body and the environment. Asthma and chronic obstructive pulmonary disease (COPD) are examples of airway diseases (Priyadarsini et al., 2023). Chronic diabetic conditions can weaken the lungs and cause severe pulmonary infections (Abbasi et al., 2025). Lung cancer is also one of the most fatal disorders and involves limited diagnostic tools (Ansari et al., 2025).

Lung Tissue Diseases

These types of disorders affect the organisation of the lung tissues, making it harder for the lungs to inhale oxygen and exhale carbon dioxide (CO2). Therefore, people under the

influence of such lung disorders feel as if their upper body is caged inside compressed clothing (Casini et al., 2024).

Lung Circulation Diseases

These types of disorders affect the blood vessels responsible for the circulation of blood in the respiratory system. They are mostly caused by blood clotting and inflammation of the vessels. They can also cause a bad impact on the functioning of the heart (Casini et al., 2024). Most lung diseases are a subtype of the above three types.

The most fatal lung condition can be denoted as lung cancer, although being curable with early detection. Another major issue can be denoted as asthma, which causes the patients to have regular breathlessness. This condition also causes the airways to narrow and dilate, and this makes it tough for the patient to breathe normally and gives them the sense that they are not provided with sufficient air (Karaddi & Sharma, 2023). The most prevalent lung infection is pneumonia, which can be brought on by germs or viruses, such as the coronavirus that causes COVID-19. Acute respiratory distress syndrome (ARDS) is a dangerous condition that causes abrupt and severe harm to the lungs. One such instance is COVID-19. Until their lungs heal, many ARDS patients require assistance breathing via a device known as a ventilator. People with lung conditions should refrain from smoking, practice yoga, and engage in regular exercise (Casini et al., 2024).

An artificial intelligence (AI) technique called deep learning is utilised to train the computers to interpret information like that of the human brain (Büyükkeçeci & Okur, 2022). Deep learning algorithms may provide precise insights and predictions by identifying intricate patterns in text, audio, images, and other data (Doddavarapu et al., 2020). Deep learning techniques may be used to automate processes like picture description and sound file transcription into text that normally need human intellect (Yamashita et al., 2018). The main intent of this paper is to underline the current studies and attempts towards the utilisation of various hybrid and evolving deep learning techniques for the analysis of lung diseases globally.

Feature Selection

In machine learning, feature selection is one of the most important steps because it involves the process of choosing the relevant and most informative features from the dataset and discarding the others that are considered irrelevant or redundant (Ahmed et al., 2023). This is a very important step in model simplicity, operation enhancement and system resources reduction. Thus, due to ignoring unnecessary data, the model reduces the chance of overfitting, can deal with more complex data and is easier to understand (Rochmawanti & Utaminingrum, 2021). In any dataset, not all features aid the model in its prediction or classification tasks equally. Some may just contribute noise or may simply be redundant,

leading to waste and deterioration of the model output. Feature selection determines these weak features and gets rid of them so that the model is built on the overlap of the characteristics that directly affect the target (Bharati et al., 2020).

Filter Methods

It serves to assess how useful features are, using stats instead of machine learning. They employ methods like correlation analysis, chi-square tests, and mutual information to sort features by their ability to predict outcomes. These approaches are quick and effective. They work great as an initial phase in choosing features ahead of using more sophisticated methods, as shown in Figure 1.

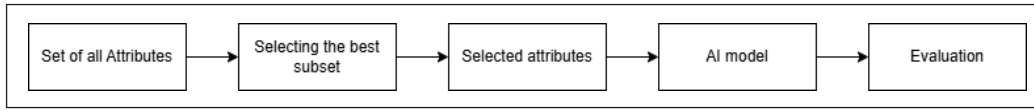


Figure 1. General overview of filter methods

Wrapper Methods

This, in contrast, chooses features by repeatedly training and assessing a machine learning model, as shown in Figure 2. Techniques such as forward selection, backward elimination, and recursive feature elimination (RFE) fit here. They indeed cost more in computational time. However, they tend to yield better accuracy since they account for how features interact with the target variable.

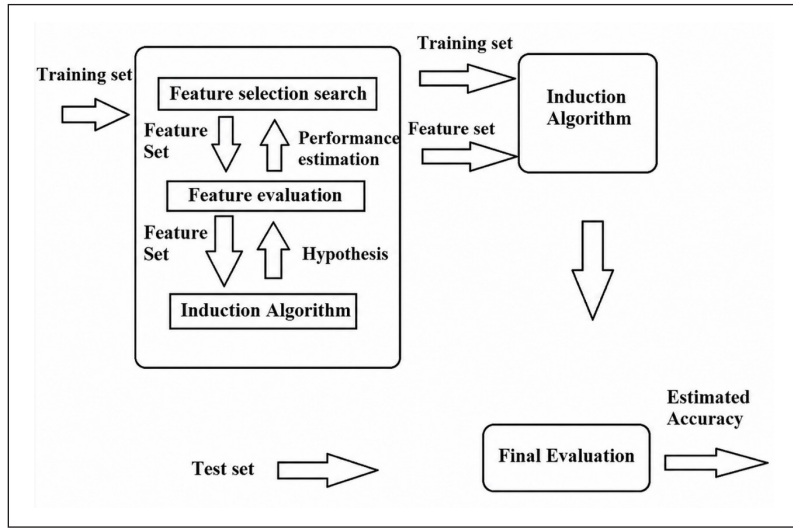


Figure 2. Wrapper method

Embedded Methods

Integrate feature selection in the model training process. For instance, lasso regression and tree-based methods do this by penalising features that don't matter or checking how important each feature is while training (Varshni et al. 2019). These methods also work very well since they then merge the strengths of both filter and wrapper approaches, as shown in Figure 3.

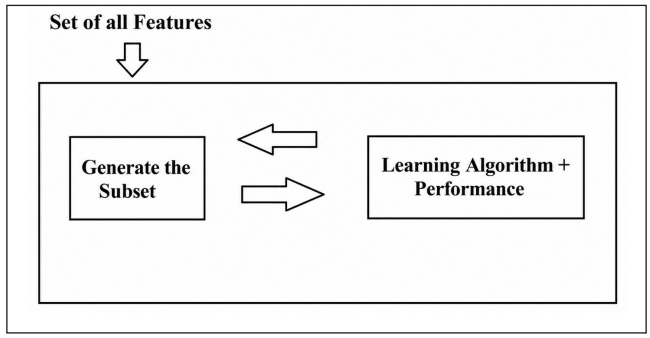


Figure 3. Embedded method

Performance Measure

The performance measures in artificial intelligence (AI) are the key metrics. They help to see how good the AI system is at reaching the goals (Souza et al., 2019). These measures count the effectiveness, the efficiency and the accuracy of the AI models. This way, the researchers and the developers can evaluate, then compare, and then enhance the performance (Shimja & Kartheeban, 2024).

The Types of Performance Measures in AI

Accuracy and Error Rates

- **Accuracy:** The ratio of correctly predicted instances and total instances (Eq. 1).

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad [1]$$

- **Error Rate:** The proportion of incorrect predictions (Eq. 2).

$$Error\ Rate = 1 - Accuracy \quad [2]$$

Precision, Recall, and F1-Score

- **Precision**: The ratio of correctly predicted positive observations to the total predicted positives (Eq. 3).

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP)\ +\ False\ Positives\ (FP)} \quad [3]$$

- **Recall (Sensitivity)**: The ratio of the correctly predicted positive observations to all the actual positives (Eq. 4).

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP)\ +\ False\ Negatives\ (FN)} \quad [4]$$

- **F1-Score**: It is the harmonic mean of precision and recall (Eq. 5).

$$F1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad [5]$$

Confusion Matrix

It is a tabular representation of the actual versus the predicted outcomes in classification tasks, containing the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) (Eq. 6).

$$Confusion\ matrix = \begin{bmatrix} [True\ Positive] & [False\ Positive] \\ [False\ Negative] & [True\ Negative] \end{bmatrix} \quad [6]$$

ROC-AUC Curve

- **Receiver Operating Characteristic (ROC)**: It denotes the plot of true positive rate (TPR) against the false positive rate (FPR) at the various thresholds.
- **Area Under the Curve (AUC)**: It measures the ability of a model to distinguish between the classes. The higher AUC values indicate better performance.

Mean Squared Error (MSE) and Mean Absolute Error (MAE)

- **MSE**: It measures the average squared difference between predicted and actual values (Eq. 7).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Actual_i - Predicted_i)^2 \quad [7]$$

- **MAE**: Measures the average absolute difference between predicted and actual values (Eq. 8).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Actual_i - Predicted_i| \quad [8]$$

Logarithmic Loss (Log Loss)

It measures the performance of a classification model whose output is a probability value. The lower values indicate better performance (Eq. 9).

$$Log\ Loss = -\frac{1}{n}\sum_{i=1}^{n} (y_i\ log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)) \quad [9]$$

R-squared (R^2)

It indicates the proportion of the variance in the dependent variable explained by the independent variables in the regression models; the values closer to 1 indicate a better fit (Eq. 10).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad [10]$$

Cross-Validation Scores

The performance is measured by the splitting of data into training and testing subsets multiple times, for example, k-fold cross-validation.

Execution Time and Computational Efficiency

It measures the time and the resources required to train and execute the model.

Specific Metrics for Specialised AI Domains

- **BLEU Score**: It evaluates the quality of the machine translation models.
- **IOU (Intersection over Union)**: It measures the accuracy of the object detection models.
- **Perplexity**: It evaluates the language model by checking its ability to predict the sequences of the text.

The choice of the performance measures depends upon the type of the problem, the data characteristics, and the specific goals of the artificial intelligence (Tay et al., 2021).

Deep Learning Models and Techniques Used to Predict Lung Diseases

Deep learning techniques have significantly improved lung disorder detection through effective analysis of medical data. Neural networks are highly utilised for medical image analysis, particularly lung X-rays and CT scans, due to the ability to automatically extract the discriminative features associated with disorders like pneumonia, lung cancer and tuberculosis (Salau & Jain, 2019). Transfer learning has further added to these features by enabling the use of pre-trained models for imaging tasks, reducing the computational complexity and adding to the performance, especially when the labelled dataset is limited (Pemeena et al., 2023). Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) are effective for analysing the sequential clinical data, including the patient history and pulmonary records, thereby allowing prediction of disorder progression (Reddy et al., 2019). Autoencoders help in automated feature extraction and anomaly detection for the identification of rare pulmonary abnormalities. The hybrid architectures combining the CNNs and LSTMs increase predictive ability by integration of radiological images with patient medical history (Nabavi et al., 2021). Vision Transformers (ViTs) employing attention mechanisms underline the clinically relevant features, adding to feature representation and model accuracy. The integration of the advanced deep-learning approach with expertise in medicine enhances lung disorder prediction, thereby enabling pre-diagnosis, personalised treatment and improved patient result in practice (Tay et al., 2021).

DL Model

The deep learning technique to classify images involves the following steps.

Data Collection

The first step is to collect images and give them labels. The categories should be clear, and the images should be of good quality. It is essential to have varied examples for the model to perform effectively. The ethical sourcing matters, and there must be enough data for each class to avoid skewing results.

Data Preprocessing

The second step requires the images to be prepared for the deep learning model. They must be resized to a uniform dimension. The normalisation of pixel values is also crucial. Additionally, data augmentation techniques such as rotation, flipping, or adjusting

brightness are necessary. The proper preparation ensures consistency and aids the model in handling diverse situations more effectively.

Model Selection

The third step involves choosing a proper deep learning model as a very important process. Suppose, for instance, CNN is a common method for image analysis. Other methods like ResNet, VGG, and DenseNet are also effective in extracting features from images. The choice depends on how big the dataset is, the complexity of the task, and the computing resources at hand.

Model Design

The fourth step details how the model appears, what layers it has, what activation functions are used, and how many classes it should output. Python, TensorFlow, or PyTorch is used for this task, as they are useful tools. The design must match the complexity levels, and there should be a balance between power and efficiency.

Training the Model

The fifth step involves the model finding patterns by decreasing the loss function, such as cross-entropy. Optimisers, like Adam, help adjust parameters gradually during training. The hyperparameters, including learning rate and batch size, also matter. This process continues over many epochs until good results are achieved.

Validation

The sixth step involves training, which is utilised to check the performance with new data, known as validation. It assists in adjusting the hyperparameters well and improves general skills. Techniques like early stopping and dropout can help resolve problems like overfitting, ensuring it stays effective for actual tasks.

Testing

The seventh step is used for testing. It is utilised to check how well the model performs on new data. The performance metrics, such as accuracy, precision, recall, and F1-score, provide insights into what is effective and what needs improvement. A confusion matrix also illustrates specific mistakes clearly, helping to direct future corrections.

Deployment

The eighth step involves using the model. Also, it involves the process of making it work with current software or systems, as compatibility is very important. Also, considering

runtime limits is important because high latency can be frustrating when using models in real-world situations.

Post-Deployment Monitoring

The ninth step involves watching the model after releasing it, as it is important for good performance over time. The data distributions change, so retraining may need to happen regularly with new data as time goes on. Monitoring helps in catching problems quickly, allowing updates to ensure classification stays reliable and steady. Figure 4 shows the overview of all these steps.

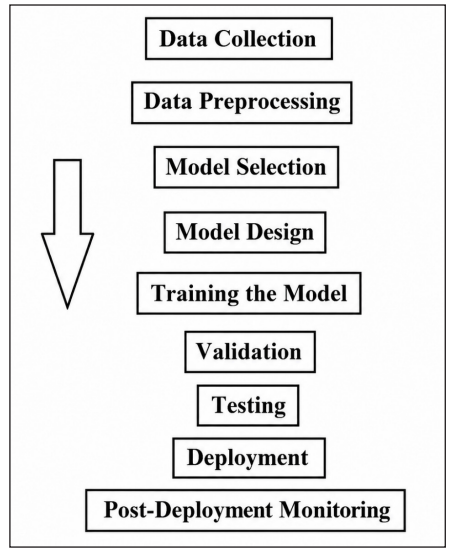


Figure 4. The steps involved in a deep learning model

Process for Diagnosing Lung Diseases Using Deep Learning Models

The explainable methods in artificial intelligence have improved the health-based diagnosis and medical analysis (Pabba et al., 2025). The lung disorder COPD is a severe and fatal disorder that weakens the lungs; however, the diagnostic methods for analysing this disorder have improved sufficiently (Abulizi et al., 2026). The diagnoses of the lung disorders involving deep learning (DL) models starts by use of efficient medical images, like chest X-rays or CT scans (Souid et al., 2021). These images are labelled by the experts in the medical field who identify conditions like pneumonia, tuberculosis, or lung cancer (Kieu et al., 2020). The next are the pre-processing steps, which involve the resizing, the normalisation, and the augmentation of the images for preparation towards analysis. It makes the data stronger, too (Mengistie & Kumar, 2021). A deep learning model is then selected. Usually, this is a Convolutional Neural Network (CNN). This works well because it extracts features from images effectively. Sometimes, pre-trained models like ResNet or DenseNet get tuned (Hurst et al., 2021). This is called transfer learning. It makes them fit better for certain datasets without needing a ton of computing power. In training, the model looks for patterns in labelled data. It tries to lower something called a loss function (Nadal-Martínez et al., 2026). An optimiser is utilised to change the parameters to make predictions better. Also, the validation takes place, which provides support for adjustment of the hyperparameters (Kumar & Kezia, 2024). This method helps in the avoidance of overfitting with methods like dropout. After the training of the data is complete, a check is done on how well the model works on the new data that it hasn't seen before.

Also, the utilisation of metrics such as the accuracy, the precision, the recall, and the F1-score is done to see how good it is. A confusion matrix helps point out specific errors made by the model (Abbas et al., 2021). When validation is done, the model is use in hospitals to assist doctors by offering quick and accurate diagnoses. Ongoing monitoring and periodic updates with new data help keep the model reliable (Tripathi et al., 2024). This way, it can adapt to changes in imaging technology and patient groups. Overall, this method boosts diagnostic accuracy as shown in Figure 5.

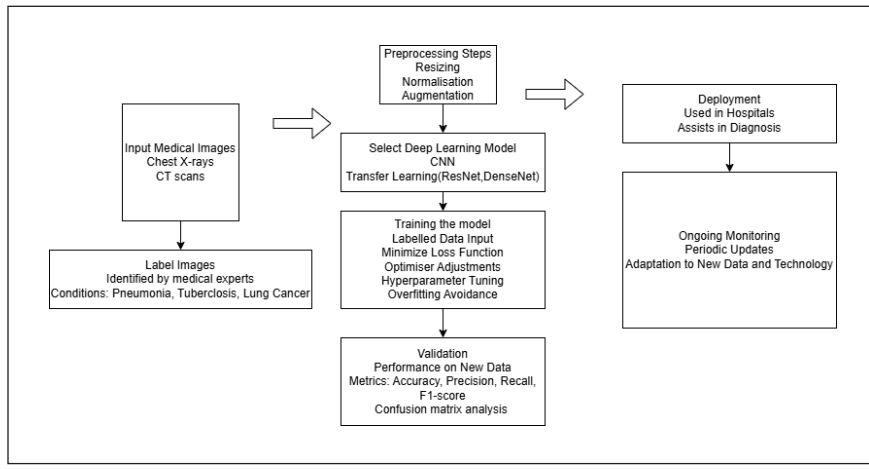


Figure 5. The review of the relationship between deep learning and lung disease detection

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model that are made for working with structured data like images. This model is made up of layers built on basic gates that are termed as neurons, as shown in Figure 6.

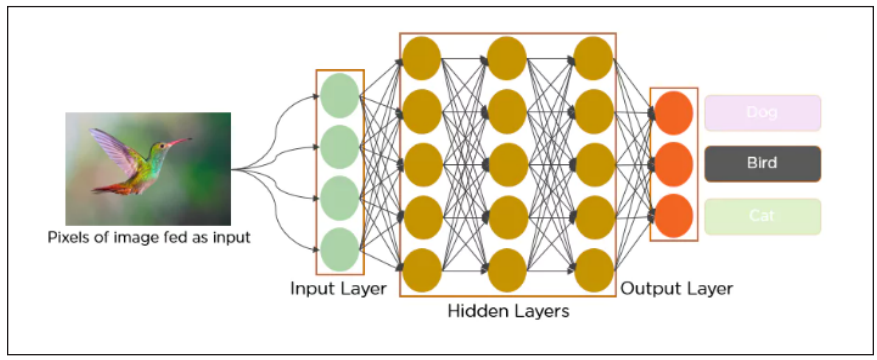


Figure 6. CNN model

The Main Parts of a CNN

Convolutional Layers

These layers are utilised to use filters on the image to create feature maps (Vikul et al., 2020). The filters can move over the image and identify features like the edges, textures, and shapes.

Pooling Layers

These layers are utilised to reduce the size of the feature maps. They can lower the dimensionality and help the network to be less sensitive to small changes in the input.

Fully Connected Layers

These layers do the high-level analysis based on the feature maps produced by the convolutional and the pooling layers.

Activation Functions

Usually, the RELU (Rectified Linear Unit) function is applied to add non-linearity.

- Formula for Convolutional Layer (Eq. 11):

$$f(x) = (x * w) + b \quad [11]$$

where:

x is the input image or feature map.
w is the filter (kernel).
b is the bias term.
$*$ denotes the convolution operation.
$f(x)$ is the output feature map produced after applying the convolution.

- Formula for Pooling Layer (Max Pooling) (Eq. 12):

$$y = \max(x_1, x_2, \ldots, x_n) \quad [12]$$

where:

$x_1, x_2, \ldots, x_n$ are the values within the pooling window.
y is the maximum value selected from the pooling region.

- Formula for Pooling Layer (Average Pooling) (Eq. 13):

$$y = \frac{1}{n}\sum_{i=1}^{n} x_i \quad [13]$$

where:

x_i represents each value in the pooling window.
n is the total number of values in the pooling region.
y is the average value of all elements within the pooling window.

Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) can be denoted as a type of neural network which tries to spot the patterns in the data that comes one after another, like the time series, text, or speech (Hussain et al., 2020). Unlike the normal feedforward neural networks, the RNNs have loops that allow the information to stick around, which is super important for dealing with the data that flows in a sequence (Wang et al., 2023).

The Main Parts of the RNNs

Input Layer

It takes the sequence data in the model.

Hidden Layer

It has units that loop back, using input and the last hidden state to keep information from before as stored.

Output Layer

It provides the results for each time step.

The hidden state h_t at time t is computed as (Eq. 14):

$$h_t = f(W \cdot h_{t-1} + U \cdot x_t + b) \quad [14]$$

where:

h_t: Hidden state at time t.
h_{t-1}: Hidden state from the previous time step.
W: Weight matrix associated with the previous hidden state.
U: Weight matrix associated with the current input.

x_t: Input vector at time t.
b: Bias term.
f: Activation function (commonly tanh or ReLU).

- The output y_t at time t is computed as (Eq. 15):

$$y_t = g(V \cdot h_t) \quad [15]$$

where:
y_t: Output at time t.
h_t: Hidden state at time t.
V: Weight matrix associated with the output layer.
g: Output activation function (commonly Softmax for multi-class classification or Sigmoid for binary classification).

Long Short-Term Memory (LSTM) Networks

The Long Short-Term Memory (LSTM) networks are a type of improved Recurrent Neural Network (RNN) that aims to fix the problems observed in the regular RNNs (Chen et al., 2019).

Key Components of LSTMs

Cell State

A memory cell used for carrying information across the time steps.

Gates

The LSTMs have three gates that can control the flow of the information.

- **Forget Gate**: It decides what information to discard from the cell state.
- **Input Gate**: It determines which new information is added to the cell state.
- **Output Gate**: It controls the output based on the cell state.

1. The forget gate output at time t is computed as (Eq. 16):

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad [16]$$

where:
f_t: Forget gate output at time t.
W_f: Weight matrix for the forget gate.

h_{t-1}: Hidden state from the previous time step.
x_t: Input vector at time t.
b_f: Bias term for the forget gate.
$[h_{t-1}, x_t]$: Concatenation of the previous hidden state and current input.
$\sigma(\cdot)$: Sigmoid activation function.

2. The input gate and candidate cell state are computed as (Eq. 17 and Eq. 18):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad [17]$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad [18]$$

where:

i_t: Input gate output at time t.

$\tilde{C}_t$: Candidate values for the cell state.

W_i: Weight matrix for the input gate.

W_C: Weight matrix for the candidate cell state.

h_{t-1}: Previous hidden state.

x_t: Current input vector.

b_i: Bias term for the input gate.

b_C: Bias term for the candidate cell state.

$[h_{t-1}, x_t]$: Concatenation of the previous hidden state and current input.

$\sigma(\cdot)$: Sigmoid activation function.

$\tanh(\cdot)$: Hyperbolic tangent activation function.

3. The cell state at time t is updated as (Eq. 19):

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad [19]$$

where:

C_t: Updated cell state at time t.
C_{t-1}: Cell state from the previous time step.
f_t: Forget gate output.
i_t: Input gate output.
$\tilde{C}_t$: Candidate cell state.

4. The output gate and hidden state at time t are computed as (Eq. 20 and Eq. 21):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad [20]$$

$$h_t = o_t \odot \tanh(C_t) \quad [21]$$

where:

o_t: Output gate activation at time t.

h_t: Hidden state (output) at time t.

W_o: Weight matrix for the output gate.

h_{t-1}: Previous hidden state.

x_t: Current input vector.

b_o: Bias term for the output gate.

C_t: Updated cell state.

$\sigma(\cdot)$: Sigmoid activation function.

$\tanh(\cdot)$: Hyperbolic tangent activation function.

Transformer Models

The Transformer models are a type of deep learning model that have changed the field of natural language processing (NLP) and all the other tasks involving sequential data (Sah, 2020). They were first initiated in 2017. The transformers use the attention mechanisms to connect to the inputs and outputs globally, removing the requirement for recurrent or convolutional layers (Kondamuri et al., 2024).

Self-Attention Mechanism

This allows the model to focus on the various parts of the input sequence to understand the relationships between the elements, no matter how far apart they are in the sequence.

Multi-Head Attention

This builds on the self-attention by allowing the model to visualise the information from the different representation subspaces at the same time.

Position-Wise Feed-Forward Networks

This utilises a fully connected feed-forward network for each position in the sequence separately and in the same way.

Positional Encoding

This adds details about where each of the elements is in the sequence since the model does not naturally recognise the order.

1. The scaled dot-product attention is computed as:

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

where:

Q: Query matrix.

K: Key matrix.

V: Value matrix.

K^T: Transpose of the key matrix.

d_k: Dimension of the key vectors.

$\sqrt{d_k}$: Scaling factor to stabilise the attention scores.

$\text{softmax}(\cdot)$: Activation function that converts attention scores into probability weights.

2. Multi-Head Attention (Eq. 22)

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \ldots, head_h)W^O \quad [22]$$

where (Eq. 23):

$$head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad [23]$$

where:

Q: Query matrix.

K: Key matrix.

V: Value matrix.

$head_i$: Output of the i^{th} attention head.

W_i^Q: Query weight matrix for the i^{th} head.

W_i^K: Key weight matrix for the i^{th} head.

W_i^V: Value weight matrix for the i^{th} head.

W^O: Output projection weight matrix.

$\mathrm{Concat}(\cdot)$: Concatenation of all attention heads.

3. Position-Wise Feed-Forward Network (FFN) (Eq. 24):

$$\mathrm{FFN}(x) = \max(0, x\mathrm{W}_1 + \mathrm{b}_1)W_2 + b_2 \quad [24]$$

where:

x: Input feature vector.

W_1, W_2: Weight matrices of the two fully connected layers.

b_1, b_2: Bias vectors.

$\max()$: ReLU activation function.

4. Positional Encoding (Eq. 25 and Eq. 26):

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad [25]$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad [26]$$

RESULTS AND DISCUSSIONS

Comparison of the Accuracy of Various Models

From the study of various literature, a mixture of deep learning-based techniques was analysed and utilised for the identification of lung diseases, where different approaches provided different performance measures. Table 1 presents the models and their accuracy used in lung disease detection.

Veeramani et al. (2025) utilised the **e**xplainable deep-learning framework through transfer learning and the U-Net segmentation for an accurate diagnosis of various lung disorders. Sabry et al. (2024) employed a DL or sound-audio-based model to identify various lung illnesses. The proposed method allowed for the analysis of recording length,

noise contamination, and graphical representation through lung sound-based audio diagnosis. The research paper focused on an algorithm for lung assessment based on audio sound classification using ML techniques. This algorithm operated with multiple datasets, feature extraction techniques, pre-processing methods, artefact removal techniques, lung-heart sound separation, DL algorithms, and other models. Al-Qaness et al. (2024) proposed a model using a specific Chest X-ray scan image dataset. The objective was to utilise DL and ML models to analyse these X-ray images and detect lung diseases. These models demonstrated significant accuracy and efficiency in identifying lung diseases through CXR imaging. However, the survey on this dataset was not very helpful due to the unavailability of analysts working on lung disease datasets and the improved patient outcomes or, in other words, the accuracy of the dataset analysis. The primary purpose of this research was to identify lung diseases accurately, efficiently, and precisely and to explore advanced tools for detecting lung-related issues. The existing models provided a foundation, but researchers and scientists worked on improving algorithms to enhance their accuracy. The accuracy, models used and preprocessing methods utilised by these studies are shown in Table 1.

Table 1
Accuracy of various methods utilised for lung disorder detection

Author Name	Preprocessing Method	Model Used	Accuracy
Sungheetha et al., 2026	Resampling, Windowing, Segmentation, Normalisation	CNN + LSTM + Transformer Ensemble (Explainable Digital Twin Framework)	94.7%
Kamala & Mohan, 2026	Resizing (224×224), Gaussian Denoising, Data Augmentation (zoom=0.3, shear=0.3, vertical flip), Feature Normalisation, PCA-based Feature Fusion	GLCM + SIFT + MobileNet Hybrid Model	98.34%
Veeramani et al., 2025	Image resizing (256×256), normalisation, image augmentation (zooming, shearing, horizontal flipping), and improved U-Net lung segmentation	Proposed XAI-TRANS (InceptionV3 + XAI + Improved U-Net)	97.53%
Pham & Hoang, 2024	Local Feature Extraction with CNN Global Feature Modelling with Vision Transformer	CNNs and ViTs in a hybrid architecture	86.4%
Al-Rawe & Tekerek, 2023	Simple Image Preprocessing done manually	CNN MobileNet DenseNet	94% 78% 96%
Alam et al., 2023	Simple Image Preprocessing done manually	RVCNet	91.27%

Table 1 (*continued*)

Author Name	Preprocessing Method	Model Used	Accuracy
Syamala et al., 2023	Simple Image Preprocessing done manually	Decision Trees EfficientNetB0	96.90% 86.18%
Moura et al., 2022	Histogram Equalisation Radiomic Feature Extraction	Random Forest XGBoost	77% 82%
Wang et al., 2021	Image clipping, normalisation, and segmentation	CNN-LSTM	91.7%

Papavassiliou et al. (2026) in their work analyse the utilisation of artificial intelligence techniques for lung disorder diagnosis and analyse the limitations, such as less generalised datasets and low-performance metrics. Vinta et al. (2024) developed lung disease morbidity, and mortality are greatly impacted by interstitial lung diseases (ILD), a collection of illnesses involving interstitial fibrosis and inflammation. Early ILD classification examinations were challenging and time-consuming when several illnesses had similar clinical symptoms. Recent studies have produced impressive results in categorising medical pictures using deep learning algorithms. Using an enhanced U-Net++ model to segment the lung section of HRCT images, the researchers created a hybrid deep-learning network model for ILD classification. To segment the lungs with lung abnormalities, the U-Net++ module has been enhanced. Five different ILD classes are classified by the feature extraction procedure using a Refined Attention Pyramid Network (RAPNet) and a MobileUNetV3. Pham and Hoang (2024) provided a novel DL method that combines CNN and ViT to increase the effectiveness of diagnosing pneumonia from medical photos. To record localised connections on the inputs, the raw pictures were first run through a local filter. Two 3 × 3 convolutional layers were part of the local filter block. Before being supplied to the ViT block's patching layer, this local filtering technique seeks to improve rich characteristics. The benchmark dataset of chest X-rays was used to test the suggested approach. ViT, VGG19, Resnet50, and Densnet201 were among the well-known models that were assessed and contrasted with the suggested approach. Moreover, the accuracy of the suggested CNN and ViT-based technique is around 1% greater than that of the regular ViT model, and it is approximately 2% higher than VGG19, Resnet50, Densnet201, and smaller in model architecture, according to data from experiments. Al-Rawe and Tekerek (2023) applied the classification-based method algorithms on DL to identify lung-related infections or diseases caused by coronavirus using CXR images. In its study regarding lung infections, at least two algorithms were employed to identify or detect novel coronavirus disease using CXR image scans. These two models are known as Mobile Net and Dense Net, which are DL models. Many use cases have been made with the help of the Mobile Net model, and the case modelling approach attained an accuracy of 96%. The Area under the Curve has also achieved an accuracy level of 94%.

Alam et al. (2023) described this research paper with the help of a hybrid DL model to detect COVID-19 patients. The researcher used the X-ray image dataset for image recognition for future analysis as a precautionary measure for COVID-19 patients, so that they are less affected with the help of predictive analysis. This DL model was created by integrating three distinct DL architectures: ResNet101V2, VGG19, and a standard CNN model. This model was also applied to non-COVID patients, viral pneumonia patients, and normal lung-infection patients with the help of image processing techniques. This algorithm achieved an accuracy of 96% using various algorithmic approaches. This is one of the most suitable options, and it performs better than other algorithms. GRAD-CAM is an application tool that evaluates images with the help of the RVCNet framework. Syamala et al. (2023) focused on diagnosing the severity of diseases in patients and determining whether the condition is dangerous or moderate for the affected person. In this context, users could check a chest X-ray (CXR) image and perform cross-verification. Users could also upload pictures and check results using this data analysis. The input from a user could be either a string or an image. The research author aimed to detect chronic lung infections or diseases quickly, focusing on finding recovery and survival plans for patients as early as possible. The research analyst employed the Hypothesis (H0) approach in this study. The research team used predictive models such as DL and machine learning (ML); this paper explored the diagnoses of diseases such as novel coronavirus, tuberculosis, pneumonia, and COPD. After analysing COVID-19 and COPD, they achieved accuracies of 96.90% and 90.32%, respectively, by applying classification algorithms to these diseases. For the image dataset, they achieved an accuracy of 98.58% using the Efficient Net B0 DL algorithm. Devi et al. (2022) diagnosed lung diseases using ML and feature selection techniques based on a review explained by the author in this research article. The author covered almost all ML algorithms in this review, including both supervised and unsupervised ML algorithms. For example, Random Forest (RF) was highlighted as a supervised ML technique, while KNN was defined as an unsupervised ML technique. These algorithms were studied or surveyed by the author between the years 2020 and 2021. They were utilised to identify lung diseases using accuracy, precision, recall, sensitivity, and F1-score metrics. These models are still utilised to solve real-world problems involving incomplete, inaccurate, and varied data values in a dataset.

Moura et al. (2022) proposed chest radiographs of COVID-19 pneumonia and various lung infections by applying models of random features and classification techniques. The primary objective of this study was to identify the radiographic characteristics of COVID-19 through supervised ML algorithms Grounded in an approach called Shapley Additive Explanations (SHAP). SHAP was also used to implement selected features with cross-validation. A random search technique was employed to optimise the XGBoost and Random Forest algorithms. The boosting algorithm achieved an accuracy level of 82%.

The explanation highlights the significance of the middle left and superior right lung regions in diagnosing pneumonia caused by COVID-19, distinguishing it from different lung patterns. Kabiraj et al. (2022) suggested a CX-Ultranet (Chest X-ray Ultranet) that uses a multiclass cross-entropy loss function on a compound scaling framework to categorise and diagnose thirteen thoracic lung illnesses from chest X-rays. On the NIH Chest X-ray Dataset, the CX-Ultra net achieves 88% average prediction accuracy in around 30% less time than previous models. Setting a new standard in disease diagnosis from chest X-rays, the proposed CX-Ultra net effectively addresses class imbalance issues. It requires much less training time regarding floating-point operations per second. The performance metrics of related studies were also analysed, as shown in Table 2. The graph clearly shows that the accuracy of the models based on the algorithms Decision Tree, DenseNet, CNN, MobileUNetV3, and DualFCM perform exceptionally well and are comparable to each other when it comes to accuracy. The models powered by the algorithms Xception, InceptionV3, EfficientNetB0, XGBoost and RVCNet also perform well as compared to the models which are built on Hybrid CNNs and ViTs, MobileNet and Random Forest. Figure 7 shows the performance of various machine learning and deep learning-based algorithms based on the performance metric accuracy.

Table 2
The performance metrics of various methods utilised for lung disease detection

Author Name	Classification Technique	Accuracy	Precision	Recall	F1 score
Sungheetha et al., 2026	Explainable AI Digital Twin Framework (CNN + LSTM + Transformer Ensemble with SHAP/LIME)	94.7%	87.6%	91.2%	89.4%
Kamala & Mohan, 2026	Hybrid AI Model (GLCM + SIFT + MobileNet)	98.34%	98.33%	98.00%	98.00%
Vinta et al., 2024	Improved MobileUNetV3	97.31%	98.65%	98.05%	98.13%
Pham & Hoang, 2024	CNNs and ViTs in a hybrid architecture	86.4%	85.7%	87.2%	86.4%
Alam et al., 2023	RVCNet	91.27%	90.48%	98.30%	94.23%
Veeramani et al., 2025	Proposed XAI-TRANS (InceptionV3 + XAI + Improved U-Net)	97.53%	98.95%	97.01%	97.97%

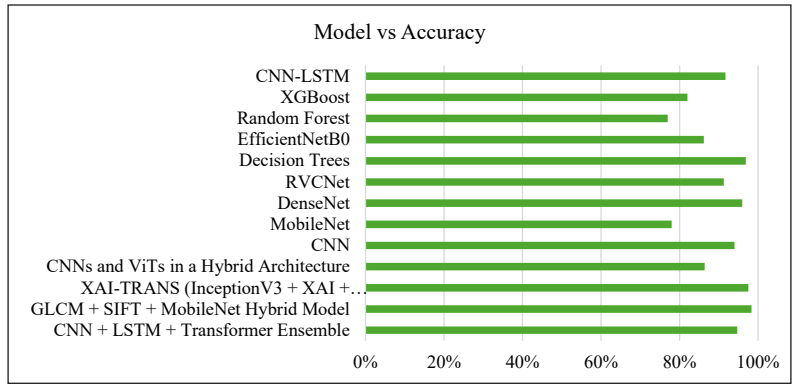


Figure 7. Specific classification model and its accuracy

Tripathi et al. (2021) proposed a DL model called CNN, which is used for classifying chest disease. The CNN model combined a convolutional layer, ReLU activations, a pooling layer, and a fully connected layer. The final fully connected layer contains 15 output units. Moreover, every output node is predicted to be one out of 15 diseases. A dataset, namely an X-ray dataset, was used to write a research article. This includes fifteen classes: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Oedema, Emphysema, Fibrosis, Pleural Thickening, Hernia, and a class for images without any findings. These classes were utilised to train the model for lung disease prediction. Still, this model has given us beautiful results and a desired output. The accuracy of the results was predicted to be 89.77% with the help of classification algorithms. The model demonstrates the effectiveness of comparative analysis. Wang et al. (2021) suggested an integrated deep-learning method that can be used to reliably identify lung nodules using low-dose computed tomography (CT) and estimate the probability of developing lung cancer in the future. The approach balances the amount of positive and negative samples by using preprocessing methods such as picture cropping, normalisation, and segmentation (Rana et al., 2024). Khehrah et al. (2020) described lung tumours as a common cancer with a high incidence and fatality rate that are frequently discovered after the fact. The condition is diagnosed with computed tomography (CT) scans, and early detection computer-aided techniques are being developed. An algorithmic framework for lung CT scan nodule detection is presented in this research. Internal structures are recovered from the parenchyma, morphological operators are used to refine the histogram, and the lung location is isolated. Using a threshold-based method, candidate nodules are distinguished from other structures, such as bronchioles and blood arteries. Nodule feature vectors are created by extracting statistical and shape-based features, which are then categorised using support vector machines. In the first stage of lung disease, a process, namely Binarisation, is used

to revert the images into the binary format, and after that, those pictures need to be compared with the threshold level for identifying lung tumours. In the next phase, CT scan images needed to be divided, preceded by a method which is known as the robust-feature extraction method, to determine the characteristics of the divided images to find the lung diseases. This model gives an accuracy of 96.67%. Figure 8 shows the year-wise comparative analysis of the classification performance metrics through the years 2023 to 2026.

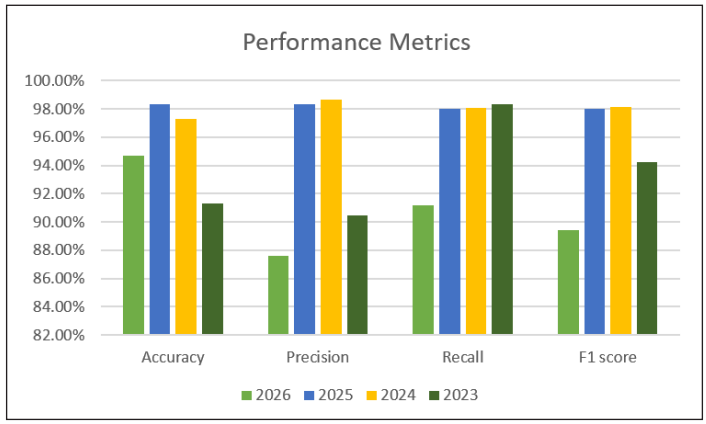


Figure 8. Trend analysis of classification metrics across multiple years

CONCLUSION

In this review, various deep learning techniques are studied and their applications to predict lung diseases are evaluated. Several deep learning techniques are evaluated, and their performance is measured. The hybrid architectures, like CNN-LSTM, performed better when the lung-disorder-based images were utilised to train the model. The deep learning models built on the CNN also performed with good accuracy. Deep learning has shown some promise in detecting lung diseases, though it still hasn't been really put to work in everyday clinical settings. A few issues, like unbalanced data, bias in selection, and high variance, tend to decrease the model's overall performance. This review concludes that research on lung disorder detection can be more effective when the following steps are followed,

- The selection of the data sources that are not only reliable but genuinely supply high-quality medical images.
- Identify which specific features really boost prediction accuracy.
- Select the features from the dataset that matter most.
- Choose a deep learning method that consistently performs best when it comes to detecting lung diseases.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my beloved parents, Sheikh Farooq Ahmad and Sakina Jan, whose unconditional love, prayers, sacrifices, guidance, and unwavering support have been the greatest strength throughout my life and academic journey. Their encouragement and faith in me have been the foundation of every achievement I have attained. I am equally thankful to my elder brother, Sheikh Adnan Farooq, for his continuous motivation, care, and guidance, and to my loving elder sister, Sheikh Uzma Farooq, for her affection, encouragement, and constant belief in my abilities. My sincere thanks also go to my dear sister-in-law, Insha. A special thanks goes to Mansha for her support, kindness, and motivation. I am also grateful to Mr and Mrs Irfan Ahmad, Dadi, Nani and Numaan for their love, blessings, and support. Finally, I pay my deepest respects to my late grandparents, Mr and Mrs Sheikh Abdul Khaliq and Mr and Mrs Ghulam Qadir Jan, whose blessings and memories remain forever in my heart.

Author(s) Contributions

Sheikh Afaan Farooq : Literature Review, Data Analysis and Interpretation, Visualisation, Investigation, Writing – Original Draft Preparation.

Aleem Ali : Writing – Review and Editing, Technical Support, Critical Review, Supervision.

Funding

The authors declare that no funding was received for this research.

Conflict of Interest Notice

The authors declare that there is no conflict of interest regarding the publication of this paper.

Support/Supporting Organisations

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Approval and Informed Consent

It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefited from are stated in the bibliography.

REFERENCES

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2021). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Applied Intelligence, 51*(2), 854-864. https://doi.org/10.1007/s10489-020-01829-7

Abbasi, R., Amin, F., Alabrah, A., Choi, G. S., Khan, S., Heyat, M. B. B., Iqbal, M. S., & Chen, H. (2025). Diabetic retinopathy detection using adaptive deep convolutional neural networks on fundus images. *Scientific Reports, 15*, Article 24647. https://doi.org/10.1038/s41598-025-09394-0

Abulizi, A., Zhou, J., Abudukelimu, N., Yehaiya, G., Abudukelimu, M., & Abudukelimu, H. (2026). A review of artificial intelligence-based research on chronic obstructive pulmonary disease. *Respiratory Medicine, 255*, Article 108778. https://doi.org/10.1016/j.rmed.2026.108778

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artificial Intelligence Review, 56*, 13521-13617. https://doi.org/10.1007/s10462-023-10466-8

Alam, F. B., Podder, P., & Hossain, M. S. (2023). RVCNet: A hybrid deep neural network framework for the diagnosis of lung diseases. *PLOS ONE, 18*(12), Article e0293125. https://doi.org/10.1371/journal.pone.0293125

Al-Rawe, I. A. M., & Tekerek, A. (2023). A novel approach for prediction of lung disease using chest X-ray images based on DenseNet and MobileNet. *Wireless Personal Communications*. Advance online publication. https://doi.org/10.1007/s11277-023-10489-y

Ansari, M. M., Kumar, S., Chola, C., Heyat, M. B. B., Akhtar, F., Hayat, M. A. B., Sayeed, E., Parveen, S., Abbasi, R., & Pomary, D. (2025). A novel machine and deep learning-based ensemble techniques for automatic lung cancer detection. *BioMed Research International, 2025*, Article 6666688. https://doi.org/10.1155/bmri/6666688

Aykanat, M., Kılıç, Ö., Kurt, B., & Saryal, S. B. (2020). Lung disease classification using machine learning algorithms. *International Journal of Applied Mathematics Electronics and Computers, 8*(4), 125-132. https://doi.org/10.18100/ijamec.799363

Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked, 20*, Article 100391. https://doi.org/10.1016/j.imu.2020.100391

Büyükkeçeci, M., & Okur, M. C. (2022). A comprehensive review of feature selection and feature selection stability in machine learning. *Gazi University Journal of Science, 36*(4), 1506-1524. https://doi.org/10.35378/gujs.993763

Casini, M., De Angelis, P., Chiavazzo, E., & Bergamasco, L. (2024). Current trends in the use of deep learning methods for image analysis in energy applications. *Energy and AI, 15*, Article 100330. https://doi.org/10.1016/j.egyai.2023.100330

Chen, B., Li, J., Guo, X., & Lu, G. (2019). DualCheXNet: Dual asymmetric feature learning for thoracic disease classification in chest X-rays. *Biomedical Signal Processing and Control, 53*, Article 101554. https://doi.org/10.1016/j.bspc.2019.04.031

Devi, P., Sagar, S., & Rohil, H. (2022). Prediction of lung disease using machine and deep learning techniques: A review. *International Journal of Health Sciences, 6*(S2), 7583-7601. https://doi.org/10.53730/ijhs.v6nS2.6833

Doddavarapu, V. N. S., Kande, G. B., & Rao, B. P. (2020). Differential diagnosis of interstitial lung diseases using deep learning networks. *The Imaging Science Journal, 68*(3), 170-178. https://doi.org/10.1080/13682199.2020.1781394

Hurst, J. R., Buist, A. S., Gaga, M., Gianella, G. E., Kirenga, B., Khoo, E. M., Mendes, R. G., Mohan, A., Mortimer, K., Rylance, S., Siddharthan, T., Singh, S. J., van Boven, J. F. M., Williams, S., Zhang, J., & Checkley, W. (2021). Challenges in the implementation of chronic obstructive pulmonary disease guidelines in low- and middle-income countries: An official American Thoracic Society workshop report. *Annals of the American Thoracic Society, 18*(8), 1269-1277. https://doi.org/10.1513/AnnalsATS.202103-284ST

Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., & Parvez, M. Z. (2020). CoroDet: A deep learning-based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons & Fractals, 142*, Article 110495. https://doi.org/10.1016/j.chaos.2020.110495

Kabiraj, A., Meena, T., Reddy, P. B., & Roy, S. (2022). Detection and classification of lung disease using a deep learning architecture from X-ray images. In *Advances in Visual Computing* (pp. 444-455). Springer. https://doi.org/10.1007/978-3-031-20713-6_34

Kamala, L., & Mohan, K. G. (2026). An efficient hybrid artificial intelligence framework for lung cancer classification using CT images. *Scientific Reports, 16*, Article 1777. https://doi.org/10.1038/s41598-025-31432-0

Karaddi, S. H., & Sharma, L. D. (2023). Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks. *Expert Systems with Applications, 211*, Article 118650. https://doi.org/10.1016/j.eswa.2022.118650

Khehrah, N., Farid, M. S., Bilal, S., & Khan, M. H. (2020). Lung nodule detection in CT images using statistical and shape-based features. *Journal of Imaging, 6*(2), Article 6. https://doi.org/10.3390/jimaging6020006

Kieu, S. T. H., Bade, A., Hijazi, M. H. A., & Kolivand, H. (2020). A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions. *Journal of Imaging, 6*(12), Article 131. https://doi.org/10.3390/jimaging6120131

Kondamuri, S. R., Kandula, S., & Pilli, E. S. (2024). Chest CT image-based lung disease classification: A review. *Current Medical Imaging, 20*, Article e15734056248176. https://doi.org/10.2174/0115734056248176230923143105

Kumar, D., & Kezia, M. (2024). Fused feature vector and dual FCM for lung segmentation from chest X-ray images. *International Journal of Intelligent Engineering and Systems, 17*(2), 464-476. https://doi.org/10.22266/ijies2024.0430.38

Mengistie, T. T., & Kumar, D. (2021). Comparative study of transfer learning techniques for lung disease prediction. In *2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)* (pp. 51-56). IEEE. https://doi.org/10.1109/IEMECON53809.2021.9689159

Mohammed, M. A., Sari, H., Dalbah, L. M., Al-Waisy, A. S., Garcia-Zapirain, B., & Alhakami, W. (2024). Chest X-ray images for lung disease detection using deep learning techniques: A comprehensive survey. *Archives of Computational Methods in Engineering, 31*, 3009–3065. https://doi.org/10.1007/s11831-024-10081-y

Moura, L. V. de, Mattjie, C., Dartora, C. M., Barros, R. C., & Marques da Silva, A. M. (2022). Explainable machine learning for COVID-19 pneumonia classification with texture-based features extraction in chest radiography. *Frontiers in Digital Health, 3*, Article 662343. https://doi.org/10.3389/fdgth.2021.662343

Nabavi, S., Ejmalian, A., Moghaddam, M. E., Abin, A. A., Frangi, A. F., Mohammadi, M., & Rad, H. S. (2021). Medical imaging and computational image analysis in COVID-19 diagnosis: A review. *Computers in Biology and Medicine, 135*, Article 104605. https://doi.org/10.1016/j.compbiomed.2021.104605

Nadal-Martínez, A., Talavera-Martínez, L., Munar, M., & González-Hidalgo, M. (2026). Lung disease classification using deep learning and ROI-based chest X-ray images. *Technologies, 14*(1), Article 1. https://doi.org/10.3390/technologies14010001

Pabba, P., Chevuturi, V. S. S., Sabbineni, J., Samudrala, U., Shaik, M. H., & Kotla, S. (2025). Explainable artificial intelligence for pulmonary disease detection from chest radiographs. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICTBIG68706.2025.11323948

Papavassiliou, K. A., Sofianidi, A. A., & Papavassiliou, A. G. (2026). Diagnosis of lung diseases in the artificial intelligence era. *International Journal of Molecular Sciences, 27*(4), Article 1796. https://doi.org/10.3390/ijms27041796

Pham, T. A., & Hoang, V. D. (2024). Chest X-ray image classification using transfer learning and hyperparameter customisation for lung disease diagnosis. *Journal of Information and Telecommunication, 8*(4), 587-601. https://doi.org/10.1080/24751839.2024.2317509

Priyadarsini, M. J. P., Suresh, M., Bhaskar, P. C., Sri, M. B. S., Roopa, G. R., Marriwala, N., Tripathi, G., Gopal, N. V. R., Thippeswamy, N., & Rahaman, S. A. (2023). Lung disease detection using various deep learning algorithms. *Journal of Healthcare Engineering, 2023*, Article 3563696. https://doi.org/10.1155/2023/3563696

Rana, S., Hosen, M. J., Tonni, T. J., Rony, M. A. H., Fatema, K., Hasan, M. Z., Rahman, M. T., Khan, R. T., Jan, T., & Whaiduzzaman, M. (2024). DeepChestGNN: A comprehensive framework for enhanced lung disease identification through advanced graphical deep features. *Sensors, 24*(9), Article 2830. https://doi.org/10.3390/s24092830

Reddy, U. J., Reddy, B. R. V. R., & Reddy, B. E. (2019). Recognition of lung cancer using machine learning mechanisms with fuzzy neural networks. *Traitement du Signal, 36*(1), 87-91. https://doi.org/10.18280/ts.360111

Rochmawanti, O., & Utaminingrum, F. (2021). Chest X-ray image to classify lung diseases in different resolutions using DenseNet-121 architectures. In *Proceedings of the 2021 6th International Conference on Biomedical Imaging, Signal Processing* (pp. 130-136). Association for Computing Machinery. https://doi.org/10.1145/3479645.3479667

Sabry, A. H., Dallal Bashi, O. I., Nik Ali, N. H., & Al Kubaisi, Y. M. (2024). Lung disease recognition methods using audio-based analysis with machine learning. *Heliyon, 10*(4), Article e26218. https://doi.org/10.1016/j.heliyon.2024.e26218

Sah, S. (2020). *Machine learning: A review of learning types*. *Preprints*. https://doi.org/10.20944/preprints202007.0230.v1

Salau, A. O., & Jain, S. (2019). Feature extraction: A survey of the types, techniques, and applications. In the *2019 International Conference on Signal Processing and Communication (ICSC)*. IEEE. https://ieeexplore.ieee.org/abstract/document/8938371

Shimja, M., & Kartheeban, K. (2024). A comparative study of lung disease classification using fine-tuned CXR and chest CT images. *Automatika, 65*(1), 312-322. https://doi.org/10.1080/00051144.2023.2293274

Souid, A., Sakli, N., & Sakli, H. (2021). Classification and predictions of lung diseases from chest X-rays using MobileNet V2. *Applied Sciences, 11*(6), Article 2751. https://doi.org/10.3390/app11062751

Souza, J. C., Diniz, J. O. B., Ferreira, J. L., França da Silva, G. L., Silva, A. C., & de Paiva, A. C. (2019). An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer Methods and Programs in Biomedicine, 177*, 285-296. https://doi.org/10.1016/j.cmpb.2019.06.005

Sungheetha, A., Sharma, R., & Aroba, O. J. (2026). Explainable AI digital twin framework for early lung disease detection. *Frontiers in Computer Science, 8*, Article 1652980. https://doi.org/10.3389/fcomp.2026.1652980

Syamala, K. P. L., Niharika, C. S., Jenny, A. M., & Pavani, P. (2023). Detection and classification of lung diseases using machine learning and deep learning techniques. *Journal of Clinical Studies and Medical Case Reports, 2*(1). https://doi.org/10.17303/jcssd.2023.2.102

Tay, Y. X., Kothan, S., Kada, S., Cai, S., & Lai, C. W. K. (2021). Challenges and optimisation strategies in medical imaging service delivery during COVID-19. *World Journal of Radiology, 13*(5), 102-121. https://doi.org/10.4329/wjr.v13.i5.102

Tripathi, A., Singh, T., Nair, R. R., & Duraisamy, P. (2024). Improving early detection and classification of lung diseases with an innovative MobileNetV2 framework. *IEEE Access, 12*, 116202-116217. https://doi.org/10.1109/ACCESS.2024.3440577

Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., & Mittal, A. (2019). Pneumonia detection using CNN-based feature extraction. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICECCT.2019.8869364

Veeramani, N., Sherine, R. S. A., Prabha, S. S., Srinidhi, S., & Jayaraman, P. (2025). NextGen lung disease diagnosis with explainable artificial intelligence. *Scientific Reports, 15*, Article 33052. https://doi.org/10.1038/s41598-025-07603-4

Vikul, P., Singh, R., & Gupta, S. (2020). Lung cancer detection system using image processing and machine learning techniques. *IJREAT International Journal of Research in Engineering & Advanced Technology, 8*(4). http://www.ijreat.org/Papers%202020/Issue46/IJREATV8I4005.pdf

Vinta, S. R., Lakshmi, B., Safali, M. A., & Kumar, G. S. C. (2024). Segmentation and classification of interstitial lung diseases based on a hybrid deep learning network model. *IEEE Access, 12*, 50444-50458. https://doi.org/10.1109/ACCESS.2024.3383144

Wang, J., Wang, P., Shao, Y., & He, D. (2023). Advancing treatment strategies: A comprehensive review of drug delivery innovations for chronic inflammatory respiratory diseases. *Pharmaceutics, 15*(8), Article 2151. https://doi.org/10.3390/pharmaceutics15082151

Wang, W., Liu, F., Zhi, X., Zhang, T., & Huang, C. (2021). An integrated deep learning algorithm for detecting lung nodules with low-dose CT and its application in 6G-enabled Internet of Medical Things. *IEEE Internet of Things Journal, 8*(7), 5274-5284. https://doi.org/10.1109/JIOT.2020.3023436

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging, 9*(4), 611-629. https://doi.org/10.1007/s13244-018-0639-9

Pertanika J. Sci. & Technol. 34 (4): 2127 - 2147 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Explainable Crop and Soil-Based Recommendation System for Next-Generation Smart Agriculture Using Model Classification

Himani Trivedi[1]*, Hetal Chauhan[2], Suresh Patel[2], Mahendra N Patel[3], Pradip Patel[4], Mahi Parmar[1], and Chrisha Dabhi[1]

[1]Department of Computer Engineering, LDRP Institute of Technology and Research, Kadi Sarva Vishwavidyalaya, 382015 Gandhinagar, Gujarat, India
[2]Department of Information Technology, Government Engineering College, 382028 Gandhinagar, Gujarat, India
[3]Department of Information Technology, Government Engineering College, 383315 Modasa, District Aravalli, Gujarat, India
[4]Department of Information Technology, L. D. College of Engineering, 380015 Ahmedabad, Gujarat, India

ABSTRACT

In India, selecting the right crop is essential due to diverse Agro-climatic conditions, such as soil types and socio-economic factors. Traditional crop recommendation methods have evolved and advanced by utilising the recent technologies in Artificial Intelligence (AI), Machine Learning (ML), the Internet of Things (IoT), and computer vision, which have enabled smart crop recommendation systems based on soil and climatic conditions. This review compares the methodologies, input features, performance evaluation metrics, limitations, and practical applicability by examining the publicly available agricultural datasets, including soil-characteristics, environmental factors, and previously harvested crops available in the open agricultural datasets from sources like Open Government Data Platform India (n.d.) and Kaggle. This study identifies key research limitations, including the drawbacks of model generalisation or integrating varied data, explainability, scalability, and real-world applications in the agricultural field. By this structured literature, best performing model was identified on classification of the standard dataset into classes as rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute and coffee, by comparing the state of art models viz. SVM, K-Nearest Neighbours, Random Forest, Gaussian Naive Bayes, XGBoost, Gradient Boosting, AdaBoost, LightGBM, Bagging, and CatBoost.

ARTICLE INFO

Article history:
Received: 22 May 2026
Accepted: 16 July 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.08

E-mail addresses:
himani_ce@ldrp.ac.in (Himani Trivedi)
hetalchauhan@gecg28.ac.in (Hetal Chauhan)
sbpatelit@gecg28.ac.in (Suresh Patel)
mahendra.patel@gecmodasa.ac.in (Mahendra N Patel)
pradippatel@ldce.ac.in (Pradip Patel)
mahi_24548@ldrp.ac.in (Mahi Parmar)
chrisha_24489@ldrp.ac.in (Chrisha Dabhi)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

These models were evaluated on the following performance metrics as precision, recall, F1-score, accuracy and model inference time amongst which Gaussian Naive Bayes results to be accurately precise and fastest, making it effective for real-time applications.

Keywords: Boosting, climate, crop prediction, deep learning, machine learning

INTRODUCTION

Agriculture contributes significantly on nations' economy and sustenance, making the need for precision and efficiency in crop recommendation more pronounced. The parameters on which the best crop cultivations depend are humidity, P-N-K nutrient levels, rainfall, temperature and pH values of soil. Farmers face problems in selecting the most suitable crop for optimal yield and profitability because of India's vast diversity in the types of soil, agroclimatic conditions and socio-economic factors (Afzal et al., 2025). This necessity has become increasingly evident with continuous variation in climate patterns, fluctuating market demands, and the imperative to ensure food security for a burgeoning population. The traditional soil prediction techniques used by farmers are based on their intuitions and experiences, which sometimes fail to manage the soil properly, due to which the crop production being less than expected (Boudouda & Boukallel, 2025). However, such methods are subjective, limited in scope, and may not always leverage the full potential of available data and technology. Additionally, the conventional approach does not adequately account for the impacts of climate change, market trends and advances in agronomic research. By using Machine learning for the integration of microbiome sequencing, a tool is used to increase the prediction of soil by 18% as compared to the traditional soil prediction techniques (Gambhir et al., 2025). Dahiphale et al. (2025) provide an overview of the highly accurate Neural Network Models by performing on the Kaggle dataset (Ingle, 2020). However, the accuracy still depends on clean, well-labelled data, which is not feasible in real life. It was observed that the existing research has explored machine learning approaches, which include XGBoost and Neural Network for a crop recommendation system, which demonstrated promising predictive performance measures (Gawade et al., 2023). As machine learning uses advanced data processing and prediction techniques, the patterns to study large sets of information like soil quality, weather, past crop results, market trends and farmer conditions can be improved to suggest the best crop choices for production, which presents challenges that result in enhancing agricultural productivity, sustainability and resilience.

LITERATURE REVIEW

Agriculture plays a crucial role in global food security as crop management is highly essential for maximising crop yield and sustainability. With the advancement in technologies like Machine Learning, Deep Learning, and Internet of Things, intelligent systems are proficient in analysing complex agricultural data, including the variations in climatic conditions and crop characteristics. Suggestions for crop cultivation, yield forecasting, soil classification, and resource management are made using these intelligent systems. By integrating sensor-based data acquisition techniques, the evaluation of existing solutions featuring selection methods and predictive model techniques has obtained highly accurate and reliable outcomes. Table 1 is the literature review showcasing the assessment of the existing approaches, articulating their key algorithms, datasets, performance metrics and identifying the gaps for future research in the agricultural sector.

The study on crop prediction, recommendation, and soil analysis was done by demonstrating diverse applications of machine learning (ML) and deep learning (DL) techniques combined with IoT-enabled data collection. Models such as Random Forest, XGBoost, and LightGBM ensembles various research by achieving high predictions, often outperforming traditional models' decision trees, KNN, and nonlinear regression (Bouni et al., 2022; Cema & Kaliappan, 2022; Kiruthika & Karthika, 2023; Priyadharshini et al., 2021). Meta-heuristic optimisation and feature selection methods, including IDSCO, CJO, SMOTE, and Grey Wolf Optimisation, further enhance selection performance and reliability (Alshahrani et al., 2023; Gautam et al., 2022; Raja et al., 2022). Enabling models such as WLSTM, ANN, and SVM to adapt to regional conditions by IoT-based soil and environmental data was frequently leveraged for real-time or near-real-time analysis (Chauhan et al., 2023 ; Teja et al., 2022). While many approaches achieved remarkable accuracy, most were limited by dataset size, regional specificity, or the number of features considered, indicating potential challenges in generalisation (Jeyanathan et al., 2023; Ramya & Sankar, 2023; Swetha et al., 2023). These studies demonstrate the efficiency of hybrid ML-DL integrated with an optimised data preparation pipeline and sensor-based data collection for precise crop recommendation, soil classification, and yield prediction, offering insights for the development of scalable, intelligent agricultural decision support systems.

Table 1
Literature review

Dataset	Performance Measures	Key Points	Limitations
Soil dataset public; crop dataset from Kaggle (Siddharth, 2021).	Accuracy (Crop Recommendation): RF 98.63%, Soil Fertility: RF 92.61%	Random Forest outperforms all other models.	Limited area-specific crop diversity.
Dataset built from augmented rainfall, climate, and fertiliser data from India (Siddharth, 2021).	Accuracy: ANN -87.71%, IDCSO-WLSTM-92.68% Precision: ANN-82.06%, IDCSO-WLSTM 90.88% Recall: ANN-84.78%, IDCSO-WLSTM-91.98% Execution time(sec): ANN-250.47, IDCSOWLSTM 241.0484	Improves prediction accuracy up to 95% through proper feature selection.	A dataset limited to one region lacks generalisation.
Public datasets for yield, cost, soil nutrients, and rainfall (Kaggle, n.d.).	Accuracy: DT-81%, KNN-85%, KNN-CV-88%, LR-88.26%, NB-82%, NN-89.88%, SVM-78%	Considering environmental and soil factors, training time is reduced.	Focuses on profit over sustainability.
Private dataset collected via sensors.	Accuracy: KNN 90%	Crop forecast improves KNN prediction accuracy (Sivanandam et al., 2023).	Prediction depends on data quality.
Dataset based on a specific region in Tamil Nadu (Cema & Kaliappan, 2022).	Random Forest - MAE:0.158, MSE:3.982, R2:0.99, Accuracy:99.98%	Random Forest achieves high accuracy and low error metrics.	Needs attention to soil health characteristics.
Soil characteristics obtained from Madurai district lab; multiple crops included.	Accuracy: DRL = RF > DT > NB > KNN; DRL/RF = 99.9%	DRL accuracy improves with larger datasets; RF also performs well.	Data reliability affects DRL performance.
Kaggle dataset (Ingle, 2020).	Accuracy = 99.5%, Recall/F1=99%, Validation Accuracy =97.73%	Highly accurate Neural Network Model (Meenachi et al., 2022).	Depends on clean, well-labelled data, which isn't available in the real world.
Private Dataset.	Accuracy = 95.7%, Recall = 96.3%, F1=95%, AUC=0.97	Accurate Prediction shows significant improvement.	Relies on region-specific datasets, and multiple ML models make it computationally heavy and complex for real life.

Table 1 (*continued*)

Dataset	Performance Measures	Key Points	Limitations
Dataset (Bhattacharya & Solomatine, 2006).	Accuracy-99.47%, Precision-94.60%, Recall 94.93%, F1-94.71%	Outperforms recent approaches with higher accuracy.	Not tested on large-scale real-time datasets.
Manually collected dataset, not publicly available.	Accuracy: RF 87.43%, Kappa 85.16%, Precision 90.34%, Recall 89.12%, F1 89.72%, AUC 92.39%	RF has the highest accuracy, precision, and specificity.	Forecasting may need a broader application.
Public dataset. (Ingle, 2020)	Accuracy: XGBoost 99.32%	XGBoost outperforms all models (Desai & Ansari, 2023).	Limited crop variety.
Open Government rainfall dataset (Open Government Data (OGD) Platform India, n.d.)	Accuracy: XGBoost 99.545%	High test-set accuracy (Sagar et al., 2023).	Depends on historical rainfall data.
Dataset with Rain, Temp, soil pH, moisture, crop type, crop yield (Kaggle, n.d.).	Random Forest: Accuracy 92%	RF gives the best results among three Tested algorithms (Kavitha et al., 2022).	Lack of comparison and quantisation.
Public generic crop dataset (Bhavana & Rao, 2025).	Accuracy: ANN 93%, LSTM 97%, Proposed SVM 99%	The proposed SVM model achieves the highest accuracy.	Models have different parameters; comparison is limited.
Dataset based on a specific region in Tamil Nadu	Random Forest - MAE: 0.158, MSE: 3.982, R2: 0.99, Accuracy: 99.98%	Random Forest achieves high accuracy and low error metrics.	Needs attention to soil health characteristics.
The dataset includes soil (N, P, K), rainfall, temperature, and pH.	Naive Bayes 89.3%, SVM 76.6%, RF 99%	RF highest accuracy (Ramachandra et al., 2023).	No real-time data is used.
Private dataset collected using sensors combined with public datasets.	Random Forest: Accuracy 99.39%	Provides detailed management guidance and forecasts.	Limited crop database.

Table 1 (*continued*)

Dataset	Performance Measures	Key Points	Limitations
Not disclosed	XGBoost: Crop 99%, RF Fertilizer 95.7%, MobileNet Disease 92%	Automates recommendations with high accuracy (Indira et al., 2022).	Needs additional sensors/ automation for full scaling.
Kaggle dataset (Ingle, 2020)	Naive Bayes / Random Forest: Accuracy, Precision, Recall, F1 = 99%	High-accuracy models (Meenachi et al., 2022).	Can be enhanced with IoT.
Public dataset with multiple crop types.	AUC=1, CA=99.6%, F1=99.6%, Precision=99.6%, Recall=99.6%	Extremely high accuracy.	Requires mobile deployment.
In laboratory-condition images with a simple background, the authors note that future work should add field-collected images across varying quality and growth stages. PlantVillage public apple subset, 3,171 images, 4 classes, 70-20-10 split, 2,228 train, 634 validation, 319 test	Accuracy 98%, Precision 0.97, Recall 0.97, F1 0.97, ROC-AUC 0.99. Per-class AUC Scab 0.993, Black Rot 0.998, Cedar Rust 0.998, Healthy 0.996.	Benchmarked against 9 pretrained models, VGG-19, ResNet-152, DenseNet-201, MobileNetV2, ResNet-50, VGG-16, InceptionV3, Xception, and MobileNet, needs only 121,964 parameters and 11 MB storage, enabling handheld and edge deployment (Vishnoi et al., 2023).	Trained only on laboratory-condition images with a simple background, the authors note that future work should add field-collected images across varying quality and growth stages.
12,000 IoT sensor records, 7,500 field-collected in India and 4,500 from public sources, 4-class crop stress labels, 80/20 split	Reported accuracy is inconsistent. Abstract and Conclusion state 98.54%, Results report 96.73% with Precision 95.80%, Recall 95.21%, F1 95.50%, significant versus SVM, Random Forest, LSTM-only, and CNN-only baselines, p less than 0.05, 94% for a separate 4-class breakdown. 96.73% is best supported since it is the only figure tied to a full metrics and baseline comparison. Sustainability gains, water use efficiency 1.8 to 2.6 kg per cubic meter, water usage reduced 29.05%, fertiliser usage reduced 21.28%, energy consumption reduced 22.12%, Sustainability Index improved 0.54 to 0.76, a 40.74% increase	Combines spatial and temporal IoT sensor fusion with Whale Optimisation-based adaptive threshold tuning, validated on real Indian farm deployments, ties accuracy to measurable sustainability outcomes (Ramu et al., 2026).	Depends on low-cost sensor accuracy limits, DHT11 plus or minus 2 degrees Celsius and or minus 5% relative humidity, FC-28 plus or minus 3%, the accuracy discrepancy between abstract and results tables is a genuine inconsistency in the source paper

Table 1 (*continued*)

Dataset	Performance Measures	Key Points	Limitations
Limited to tomato leaf diseases only, reliance on curated Kaggle imagery may limit generalisation to uncontrolled field conditions.	Accuracy 99.11%, Precision, Recall, and F1 approximately 99%, outperformed MobileNet 87.44% and VGG19 95.50%, 10-fold cross-validation training accuracy 99.41%, validation accuracy 98.88%	Integrates Grad-CAM, Grad-CAM++, LIME, and SHAP for explainability, deployed as a Flask web application (Assaduzzaman et al., 2025).	Limited to tomato leaf diseases only, reliance on curated Kaggle imagery may limit generalisation to uncontrolled field conditions.
Crop and Fertiliser Dataset, Western Maharashtra, India, IoT-enabled nitrogen, phosphorus, potassium, temperature, and humidity data	TabNet achieved 95.24% accuracy for fertiliser recommendation and 96.21% accuracy for crop recommendation, outperforming conventional classifiers.	Unifies fertiliser and crop recommendation in one interpretable system, attention mechanism removes the need for manual feature selection (Venkateswara & Padmanaban, 2025).	Single-region dataset limited to Western Maharashtra, with a higher computational cost than simpler classifiers used elsewhere in this table
Regional meteorological station data, including temperature, solar radiation, and other variables	Random Forest performed best, RMSE 0.52 mm per day, R-squared 0.96, temperature and solar radiation identified as key predictors	Systematically tests reduced-input scenarios for data-scarce regions, addressing the heavy data requirements of the FAO-56 Penman-Monteith method (Farag, 2025).	Static, non-sequential machine learning models may miss temporal weather dependencies captured by LSTM-based approaches, validated on a single region only
IoT-based soil salinity sensor data, including electrical conductivity, total dissolved solids, and pH from salt-affected fields.	Proposed leaching-based reclamation is shown to more effectively reduce soil salinity than the traditional reclamation method, but specific accuracy or error values are not detailed in the available abstract.	Addresses a soil management challenge largely absent from mainstream crop-recommendation literature, combining IoT salinity sensing with an ensembled LSTM for site-specific guidance (Khan et al., 2022).	Focused specifically on saline soils, not generalizable to standard agricultural contexts, requires specialised soil sensors beyond typical sensor suites.

DATASET

Exploratory Data Analysis was performed to understand the dataset before proceeding with model comparison, to identify the statistical data of the variables and to inspect the features and crop-specific environmental patterns. This provided a comprehensive overview of the dataset and helps in a reliable machine learning framework for crop recommendation.

Dataset Description

This review utilises the publicly available Crop Recommendation Dataset, which is a benchmark dataset for crop recommendation systems. The dataset was available on Kaggle, and it contains 2,200 rows of observations, which are equally distributed amongst 22 crop classes (Singh, 2023). The dataset is a balanced distribution of observations, having 100 samples for each crop category, which reduces the data preprocessing and provides an unbiased comparison among different classification algorithms.

This dataset contains seven numerical predictor variables having soil nutrient composition and climatic conditions, such as Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, soil pH, and rainfall, with the crop classes as the target variable. These predictor and target variables together showcase the soil fertility and environmental factors required for the recommendation of the most suitable crop.

The target variable comprises 22 crop classes, such as apple, banana, blackgram, chickpea, coconut, coffee, cotton, grapes, jute, kidneybeans, lentil, maize, mango, mothbeans, mungbeans, muskmelon, orange, papaya, pigeonpeas, pomegranate, rice and watermelon.

Before analysing the data, the dataset was examined to identify the data quality issues, including missing values, duplicate records, variation in data types, and invalid observations. No missing values or inconsistent data were found, resulting in the dataset being completely clean and consistent. Certain observations contain zero values for the nitrogen variable, but this leads to valid measurements. Hence, no observations were removed during the data preprocessing stage.

Figure 1 displays the distribution of the 22 crop classes. Each crop category contains exactly 100 observations, indicating a balanced multi-class dataset. This reduces the biases towards any crop class and contributes to fair comparative evaluation of all machine learning classifiers.

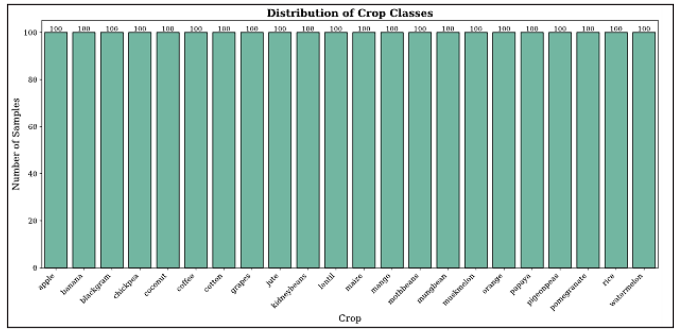


Figure 1. Crop class distribution

Statistical Characteristics of the Dataset

The corresponding feature distribution is mentioned in Figure 2.

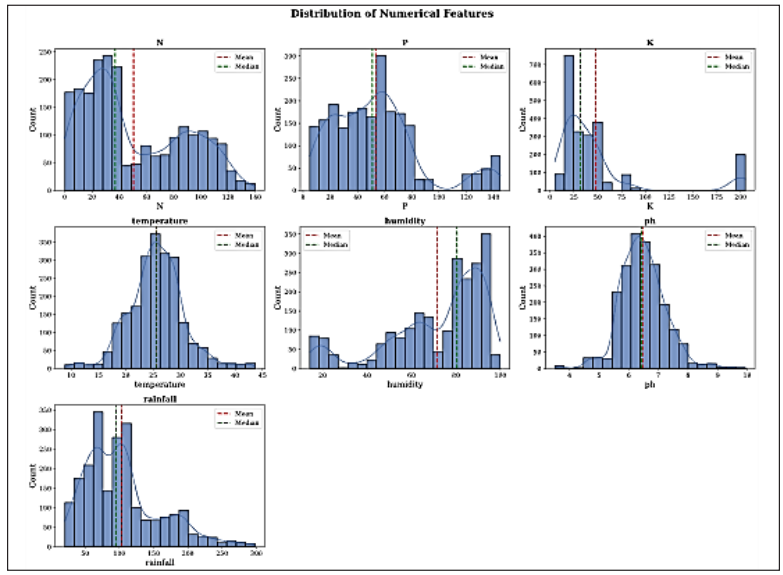


Figure 2. Distributive representation of statistical features

Figure 2 contains numerical predictor variables that show diverse statistical characteristics exhibiting variations in environmental factors. Potassium and rainfall show relatively higher positive skewness, whereas temperature and soil pH contain comparatively symmetric distributions. The Shapiro-Wilk normality test was performed to show the deviation of these numerical features from a normal distribution ($p < 0.05$).

Correlation Analysis

Pearson Correlation Analysis was performed to inspect the inter-feature relations of variables. The correlation matrix is presented in Figure 3.

Weak linear correlation is exhibited by most of the feature pairs, which indicates limited multicollinearity within the dataset. It is observed that the highest positive correlation is between Phosphorus and Potassium ($r = 0.74$), while the other remaining feature pairs show relatively

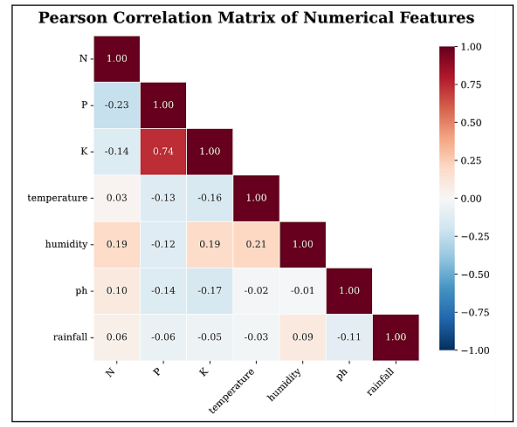


Figure 3. Pearson correlation matrix for inter-feature relationships

weak positive or negative correlations. All predictor variables were retrained for subsequent model development, as no correlation coefficient exceeded the commonly accepted multicollinearity threshold (|r|>= 0.80)

Identification of Outliers and Analysis

Figure 4, illustrating the Box Plots, and Table 3 displaying the Interquartile Range (IQR) were used to investigate the variability and presence of potential outliers within the numerical features.

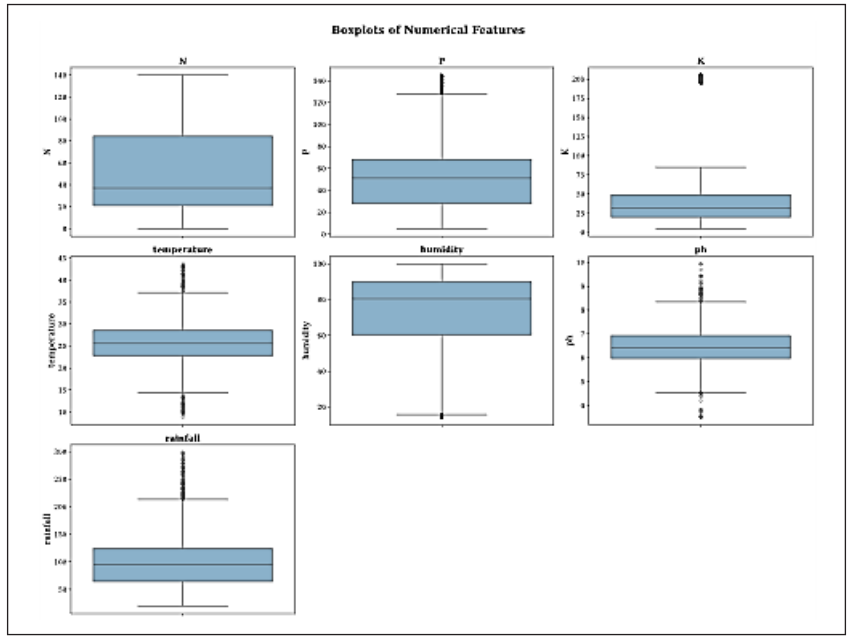


Figure 4. Box Plot showing median, quartiles and possible outliers for numerical features

Table 2
Overview of the interquartile range outlier for the dissemination of median features

Feature	Q1	Q3	IQR	Lower Bound	Upper Bound	No. of Outliers	Percentage (%)
N	21	84.25	63.25	-73.875	179.125	0	0
P	28	68	40	-32	128	138	6.27
K	20	49	29	-23.5	92.5	200	9.09
temperature	22.769	28.562	5.792	14.081	37.25	86	3.91
humidity	60.262	89.949	29.687	15.732	134.479	30	1.36
pH	5.972	6.924	0.952	4.544	8.352	57	2.59
rainfall	64.552	124.268	59.716	-25.022	213.841	100	4.55

Table 2 contains the outliers primarily observed in potassium, phosphorus and rainfall, where it was observed that these refer to the naturally occurring environmental variability rather than the false measurements and were therefore retrained. Dataset values were preserved adequately to improve the practical applicability of the developed crop recommendation models for diverse agricultural conditions.

Feature Relevance Analysis

Mutual Information (MI) analysis was performed to estimate the contribution of each predictor variable toward crop classification, whose result feature ranking can be seen in Figure 5.

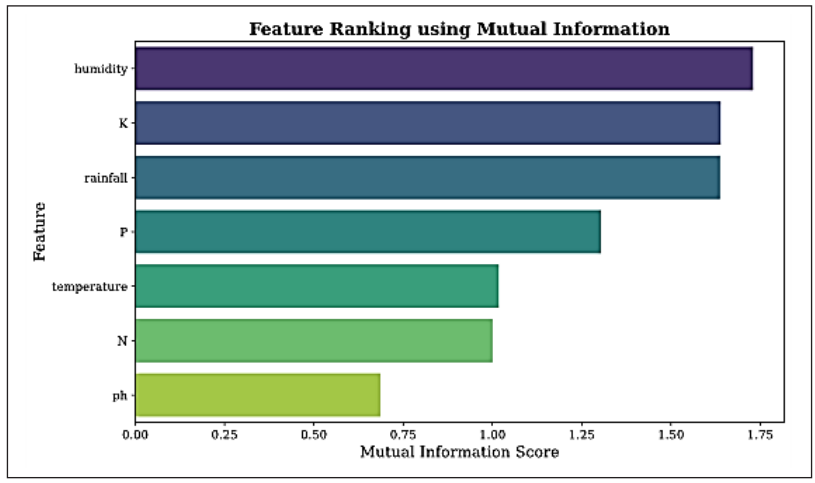


Figure 5. Ranking of mutual information of features showcases a stronger relationship or dependency between features and target variables

The most informative predictor identified was humidity, followed by potassium, rainfall, phosphorus, temperature, nitrogen and pH of soil. It was observed from the rankings that both climate and soil nutrient variables contribute substantially to crop, which supports the inclusion of all 7 predictor variables in the comparative machine learning study.

Crop-wise Environmental Characteristics

The average feature values of each crop were standardised using Z-score normalisation and visualised as a heatmap to facilitate agronomic interpretation, which can be observed in Figure 6.

Crop classes vary as per the distinct environmental profiles. Highest standardised rainfall values are exhibited by rice, which confirms its dependence on high-water conditions, whereas characteristics of coconut are observed in elevated humidity and rainfall. Nutrient-rich soil requirements are observed in apples and grapes, as it requires comparatively higher phosphorus and potassium. Whereas lower humidity and temperature conditions are associated with chickpea, which requires relatively higher soil pH values. The dataset effectively captures meaningful variations in environmental conditions across different crop classes, which are consistent with the agroeconomic knowledge.

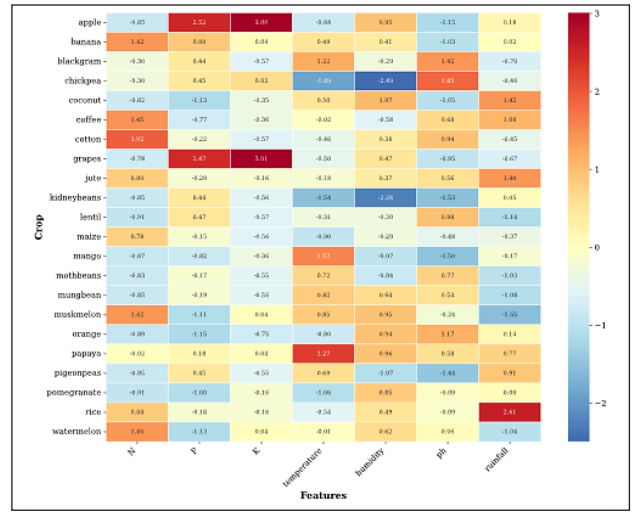


Figure 6. Heatmap illustrating standardised crop-wise soil and environmental characteristics

METHODOLOGY

The crop Recommendation dataset ensures balanced class distribution as the dataset comprises 2,200 instances equally distributed across 22 crop classes with 100 samples per class. The dataset was divided into 80:20 subsets of training and testing. To guarantee reproducibility of the experimental results, a fixed random seed (random_state = 42) was used during the dataset partitioning. Stratified 5-Fold Cross-Validation was performed to assess the robustness and stability of each classifier. The mean accuracy and standard deviation across the folds were recorded for comparative analysis.

10 Machine Learning models, viz. SVM, KNN, Random Forest, Gaussian Naive Bayes, XGBoost, Gradient Boosting, AdaBoost, LightGBM, Bagging and CatBoost were trained and implemented on the above discussed dataset. To avoid information leakage into the test set, hyperparameter tuning was carried out exclusively using GridSearchCV for models with relatively smaller search spaces and RandomisedSearchCV for ensemble boosting algorithms.

A non-parametric statistical hypothesis test was performed to determine whether the observed differences among the evaluated classifiers. To compare the overall performance of all classifiers across multiple evaluation folds, the Friedman test was performed. Pairwise Wilcoxon Signed-Rank tests were conducted as post-hoc analyses to identify significant

differences between closely performing models when statistically significant differences were detected.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful classification algorithm that finds the optimal hyperplane to separate different classes in the feature space. SVM is particularly effective for crop recommendation systems when the decision boundaries between different crop types are complex. Its ability to work well in high-dimensional spaces ensures that even subtle differences in features such as soil nutrients and climatic conditions are considered. SVM makes the classification between the different crop types (classes) robust by maximising the margin, and tackling the complexity by running many small, accurate comparisons, which are highly reliable for suggesting the best crop according to the accurate efficiency showcased by this method.

K-Nearest Neighbours (KNN)

In the K-Nearest Neighbours (KNN) method, for accurate crop selection, a new data point based on the most similar existing farm data aligns with real-world conditions, that results into useful approach when diverse agricultural datasets are used instead of assuming the data. This adapts to the data distribution method naturally and is useful when specifically taking into consideration the local parameters like soil and climate. By cross-validation, the value of the nearest neighbour (k) for selection of the best crop fit for agriculture was determined. In this case, the highest accuracy was obtained at k equal to 4.

Random Forest

In Random Forest, multiple decision trees are combined to provide the best classification accuracy. It handles complex interactions among different variables, as crop recommendation depends on various factors which interact with each other in a complex manner, improving generalisation. The data is hidden, and the risk of overfitting is reduced by averaging multiple decision trees, resulting in a convoluted and non-linear relationship between the parameters. Large datasets with numerous features can be processed by Random Forest, which analyses them comprehensively. This potential identifies the most effective factors that affect better recommendations in crop selection.

Gaussian Naive Bayes (GNB)

Gaussian Naive Bayes is a programmable machine learning classifier that showcases output following Bayes' Theorem, which is a method used in probability. The classification can be done on smaller continuous datasets as well if the attributes follow a Gaussian

distribution, requiring less computational resources, that results into fast predictions. It has a strong assumption of feature independence, which retains reasonably true value in practice, regardless of its simplicity. Studying the crops based on attributes' probability combinations, GNB provides fast and interpretable results by giving probabilistic outputs that help in understanding the certainty of each recommendation, sanctioning more informed decision-making.

XGBoost

In this study, several machine learning models such as decision trees, random forests, K-nearest neighbours, and neural networks are used to analyse soil and crop data, but XGBoost (Extreme Gradient Boosting) deliver the best results. XGBoost improves its predictions step-by-step by learning from previous errors, which allows it to accurately understand complex and nonlinear relationships between soil, climate, and crop parameters. XGBoost includes regularisation techniques that prevent overfitting, optimises performance of the model on new data, and prove to be highly effective for crop recommendation. Research on the usage of XGBoost along with other models, in comparison, helps to identify the most efficient approach for precise agriculture decision-making.

Gradient Boosting

Gradient Boosting starts with the basic prediction and gradually learns from the mistakes. Based on factors like soil condition, weather, and past crop results, every new step solves the error that occurred previously, improving the recommendation over the passage of time, which goes well in the vibrant farming areas, resulting in more reliable suggestions to the farmers.

AdaBoost

Correcting prediction errors through iterative learning is focused by AdaBoost, though it requires further optimisation to fully leverage historical and environmental data for crop recommendations. Its iterative approach improves model performance over time, showing potential for enhancing accuracy with continued refinement and adjustment.

LightGBM

Large amounts of data are efficiently handled quickly by LightGBM because of its speed; it can provide crop recommendations in real time using the latest information. This makes it useful in farming situations where decisions need to be made fast to improve productivity and maintain sustainable practices.

Bagging

Bagging enhances the crop recommendation stability by averaging the predictions from multiple models trained on distinct data samples. It ensures the consistent performance across diverse agricultural practices, which improves the reliability in the decision-making process, resulting in a moderate impact of the outliers.

CatBoost

The management of the categorical variables commonly used in agricultural datasets is done by CatBoost. The effective incorporation of these features into the model optimises the prediction process of the suitable crop as per the real time scenarios. This ensures reliable recommendations, rendering it appropriate for precision agriculture applications where accurate integration of categorical data, essential for decision-making and resource optimisation, is obtained.

Table 3
Comparison of different models based on various METRICS

Model	Accuracy	Precision Macro (%)	Recall Macro (%)	F1 macro (%)	Model Inference Time
Bagging	99.55	99.57	99.55	99.55	0.012039958
Random Forest	99.32	99.35	99.32	99.32	0.042236809
CatBoost	99.32	99.37	99.32	99.32	0.002705756
Gaussian Naive Bayes	99.32	99.37	99.32	99.32	0.001809618
LightGBM	98.86	98.91	98.86	98.86	0.048731084
SVM	98.86	98.97	98.86	98.86	0.011766945
XGBoost	98.86	98.94	98.86	98.85	0.015371169
Gradient Boosting	98.64	98.75	98.64	98.64	0.032235853
KNN	98.18	98.23	98.18	98.17	0.008214612
AdaBoost	91.13	93.31	91.13	90.11	0.027248464

Gaussian Naive Bayes achieved one of the highest overall performances while maintaining the lowest inference time, demonstrating its suitability for real-time crop recommendation systems among the evaluated methods. Ensemble-based approaches such as Random Forest, XGBoost, and Bagging also produced competitive results, whereas AdaBoost exhibited comparatively lower predictive performance. Bagging achieved the highest predictive accuracy of 99.55% among the evaluated classifiers. However, Gaussian Naive Bayes achieved a comparable classification performance of 99.32% while exhibiting the lowest computational overhead in terms of both training and inference time. Even though Bagging provides the highest predictive performance, Gaussian Naive Bayes offers

a more favourable trade-off between accuracy and computational efficiency, making it suitable for practical crop recommendation applications.

RESULTS AND DISCUSSIONS

As depicted in Figure 7, 10 various machine learning models, SVM, KNN, Random Forest, Gaussian Naive Bayes, XGBoost, Gradient Boosting, AdaBoost, LightGBM, Bagging, and CatBoost are trained on the same dataset with 7 features, with 22 crop types. As depicted in the results in Table 3, CatBoost, Gaussian Naive Bayes, Random Forest, and Bagging are the top performers with accuracy of 99.31%, 99.31%, 99.31%, and 99.54%, respectively. While LightGBM, SVM, XGBoost, Gradient Boosting, and KNN were the moderate performers, they reached the accuracy of 98.86%, 98.86%, 98.86%, 98.63% and 98.18%, respectively. AdaBoost achieved the least accuracy of 91.13%, respectively, because AdaBoost's reweighting points give priority to hard-to-classify samples, which are either overlapped or can disturb the generalisation. These models are also compared based on their inference time, which is shown in Figure 8. Hence, this shows that models like CatBoost or Naive Bayes are better comparatively. The Comparative analysis in Figure 8 shows that the inference time of all 10 models, where Gaussian Naive Bayes proves to have the least inference time of 0.00295s.

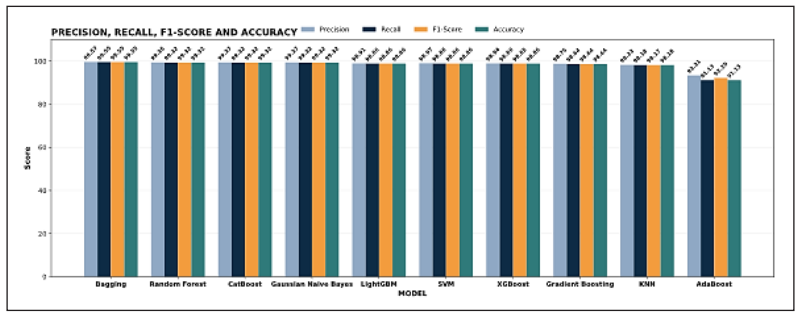


Figure 7. Visual representation of performance measures for various models

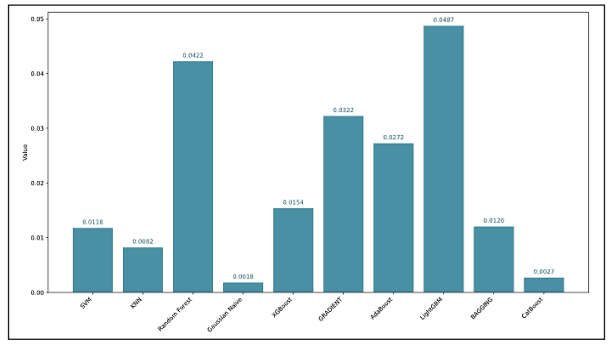


Figure 8. Comparison of inference time across ten classifiers

EXPLAINABLE AI ANALYSIS

To improve the interpretability of the comparative machine learning algorithms for crop recommendation systems, Explainable Artificial Intelligence (XAI) analysis was performed using Shapley Additive exPlanations (SHAP) on the Gaussian Naïve Bayes classifier. SHAP combines the contribution of each input feature utilised for model predictions and provides both global and local explanations. This provides a comprehensive reasoning for the quantitative performance metrics by showcasing how the agro-climatic factors affect the crop recommendation decisions.

Figure 9 presents the global Beeswarm Plot that provides a summary of the contribution of all input features across the dataset. From the plot, potassium (K) has the strongest overall influence on crop prediction, followed by rainfall and humidity. Temperature shows moderate importance, while nitrogen, phosphorus, and pH have comparatively less contribution. This plot indicates that the variations in both soil and climatic variables are the factors affecting model predictions for all crop classes.

Figure 10 demonstrates the ranking of the features based on their absolute SHAP values. This ranking correctly depicts that potassium is the most influential feature, followed by rainfall, humidity, and temperature. The less important features, nitrogen, phosphorus, and pH, show that these variables provide minimal contribution to the prediction. The similar result in the SHAP Beeswarm Plot and Bar Plots indicates the correct importance hierarchy of the features.

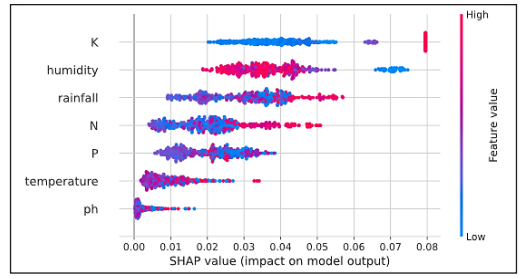


Figure 9. Dense summary of top features using SHAP Beeswarm Plot

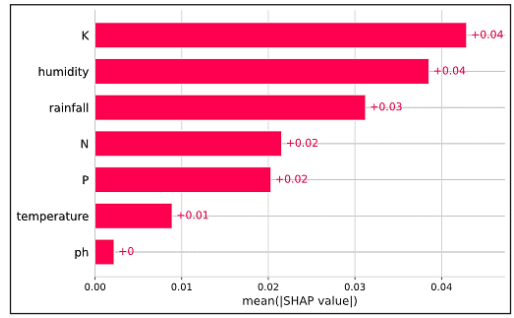


Figure 10. Global importance of each feature using the SHAP Bar Plot

Figure 11 illustrates how each feature contributes to the final decision of crop recommendation. It provides the overall contribution of soil nutrients and climatic variations that improves the overall interpretability of the evaluated models.

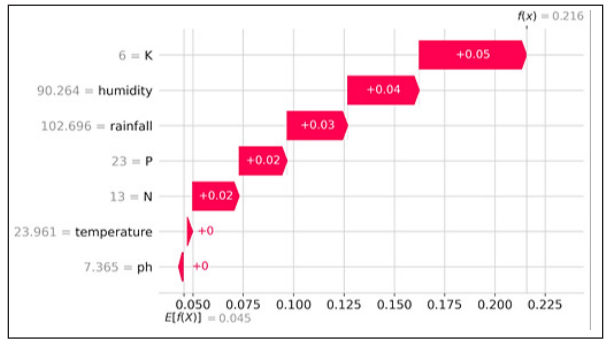


Figure 11. Proportional relationship between features and their contribution to the model

CONCLUSION

This research shows that machine learning assists farmers in determining which crops are to be grown, predicting accurate and tailored outcomes for each specified data characteristic. The diversity in datasets, like unpredictable weather and market fluctuation, is easily handled by this approach. Various models as SVM, K-Nearest Neighbours, Random Forest, Gaussian Naive Bayes, XGBoost, Gradient Boosting, AdaBoost, LightGBM, Bagging, and CatBoost unveils the significant enhancement in agricultural productivity, sustainability, and resilience. Bagging achieved the highest predictive accuracy (99.55%) among the evaluated classifiers. However, Gaussian Naive Bayes achieved comparable classification performance (99.32%) while exhibiting the lowest computational overhead in terms of both training and inference time. Even though Bagging provides the highest predictive performance, Gaussian Naive Bayes offers a more favourable trade-off between accuracy and computational efficiency, making it suitable for a practical crop recommendation application.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their respective institutions for providing the necessary academic environment, infrastructure, and research facilities that made this work possible.

REFERENCES

Afzal, H., Amjad, M., Raza, A., Munir, K., Villar, S. G., Lopez, L. A. D., & Ashraf, I. (2025). Incorporating soil information with machine learning for crop recommendation to improve agricultural output. *Scientific Reports, 15*, Article 8560. https://doi.org/10.1038/s41598-025-88676-z

Alshahrani, H., Alkahtani, H. K., Mahmood, K., Alymani, M., Mohammed, G. P., Abdelmageed, A. A., Abdelbagi, S., & Drar, S. (2023). Chaotic Jaya optimisation algorithm with computer vision-based soil type classification for smart farming. *IEEE Access, 11*, 65849-65857. https://doi.org/10.1109/ACCESS.2023.3288814

Assaduzzaman, M., Bishshash, P., Nirob, M. A. S., Al Marouf, A., Rokne, J. G., & Alhajj, R. (2025). XSE-TomatoNet: An explainable AI-based tomato leaf disease classification method using EfficientNetB0 with squeeze-and-excitation blocks and multi-scale feature fusion. *MethodsX, 14*, Article 103159. https://doi.org/10.1016/j.mex.2025.103159

Bhattacharya, B., & Solomatine, D. P. (2006). Machine learning in soil classification. *Neural Networks, 19*(2), 186-195. https://doi.org/10.1016/j.neunet.2006.01.005

Bhavana, M., & Rao, K. S. (2025). Smart agriculture with Internet of Things for precise crop prediction using integrated machine learning and advanced stacking ensemble from soil parameters. *SSRG International Journal of Electronics and Communication Engineering, 12*(8), 74-90. https://doi.org/10.14445/23488549/IJECE-V12I8P107

Boudouda, S., & Boukallel, M. (2025). Intelligent soil classification for precision agriculture in Algeria: Integrating machine learning and image processing for optimised crop yield and recommendation. *Algerian Journal of Signals and Systems, 10*(2), 65-74. https://doi.org/10.51485/ajss.v10i2.264

Bouni, M., Hssina, B., Douzi, K., & Douzi, S. (2022). Towards an efficient recommender system in smart agriculture: A deep reinforcement learning approach. *Procedia Computer Science, 203*, 825-830. https://doi.org/10.1016/j.procs.2022.07.124

Cema, G., & Kaliappan, E. (2022). AI-based crop recommendations for intensive farming using WSN. In *2022 3rd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (pp. 1718-1723). IEEE. https://doi.org/10.1109/ICICICT54557.2022.9917962

Chauhan, A., Tsunduru, A., Parveen, K., Tokala, S., Hajarathaiah, K., & Enduri, M. K. (2023). A crop recommendation system based on nutrients and environmental factors using machine learning models and IoT. In *2023 International Conference on Information Technology (ICIT)* (pp. 453-458). IEEE. https://doi.org/10.1109/ICIT58056.2023.10226131

Dahiphale, D., Shinde, P., Patil, K., & Dahiphale, V. (2025). Smart farming: Crop recommendation using machine learning with challenges and future ideas. *Artificial Intelligence and Applications, 3*(4), 1-13. https://doi.org/10.47852/bonviewAIA52026214

Desai, M., & Ansari, N. (2023). An innovative method to increase agricultural productivity using machine learning-based crop recommendation systems. In *2023 2nd International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)* (pp. 645-651). IEEE. https://doi.org/10.1109/ICAISS58487.2023.10250531

Farag, A. A. (2025). Machine learning approaches for enhanced estimation of reference evapotranspiration (ETo): A comparative evaluation. *Scientific Reports, 15*, Article 38485. https://doi.org/10.1038/s41598-025-23166-w

Gambhir, S., Khan, M. A. S., Sharma, N., Nagar, M., Chawara, G., Shweta, & Prasad, A. (2026). Soil classification and crop recommendation with integrated microbiome and machine learning. In *Data*

Analytics and Management (Lecture Notes in Networks and Systems, Vol. 1603, pp. 412-426). Springer. https://doi.org/10.1007/978-3-032-03769-5_31

Gautam, V., Trivedi, N. K., Anand, A., & Kumar, A. (2022). Optimised a high-performance deep neural network for crop recommendation. In *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 838-843). IEEE. https://doi.org/10.23919/INDIACom54597.2022.9763156

Gawade, S., Rout, G., Kochar, P., Ahire, V., & Namboodiri, T. (2023). AGROFERDURE: Intelligent crop recommendation system for agricultural crop productivity using machine learning algorithms. In *2023 International Conference on Computer, Electronics & Electrical Engineering & Their Applications (IC2E3)* (pp. 1-9). IEEE. https://doi.org/10.1109/IC2E357697.2023.10262476

Indira, D., Sobhana, M., Swaroop, A. H. L., & Phani Kumar, V. (2022). KRISHI RAKSHAN: A machine learning-based new recommendation system for the farmer. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1798-1804). IEEE. https://doi.org/10.1109/ICICCS53718.2022.9788221

Ingle, A. (2020). *Crop recommendation dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

Jeyanathan, J. S., Medha, B., Sai, G. T. V., Kumar, R. B., & Sahu, V. (2023). Automated crop recommender system using pattern classifiers. In *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* (pp. 567-572). IEEE. https://doi.org/10.1109/IDCIOT56793.2023.10053451

Kaggle. (n.d.). *Kaggle: Your machine learning and data science community*. https://www.kaggle.com/

Kavitha, R., Kavitha, M., & Srinivasan, R. (2022). Crop recommendation in precision agriculture using supervised learning algorithms. In *2022 3rd International Conference for Emerging Technology (INCET)* (pp. 1-4). IEEE. https://doi.org/10.1109/INCET54531.2022.9824155

Khan, A. A., Nauman, M. A., Bashir, R. N., Jahangir, R., Alroobaea, R., Binmahfoudh, A., Alsafyani, M., & Wechtaisong, C. (2022). Context-aware evapotranspiration (ETs) for saline soils reclamation. *IEEE Access, 10*, 110050-110063. https://doi.org/10.1109/ACCESS.2022.3206009

Kiruthika, S., & Karthika, D. (2023). IoT-based professional crop recommendation system using a weight-based long-term memory approach. *Measurement: Sensors, 27*, Article 100722. https://doi.org/10.1016/j.measen.2023.100722

Meenachi, L., Ramakrishnan, S., Sivaprakash, M., Thangaraj, C., & Sethupathy, S. (2022). Multi-class ensemble classification for crop recommendation. In *2022 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1319-1324). IEEE. https://doi.org/10.1109/ICICT54344.2022.9850561

Open Government Data (OGD) Platform India. (n.d.). *Open Government Data (OGD) Platform India*. National Informatics Centre. https://data.gov.in/

Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 843-848). IEEE. https://doi.org/10.1109/ICCMC51019.2021.9418375

Raja, S. P., Sawicka, B., Stamenkovic, Z., & Mariammal, G. (2022). Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEE Access, 10*, 23625-23641. https://doi.org/10.1109/ACCESS.2022.3154350

Ramachandra, R. A., Ankitha, G. V., Divya, I., Vandana, P., & Jagadeesh, H. S. (2023). Crop recommendation using machine learning. In the *2023 International Conference on Data Science and Network Security (ICDSNS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICDSNS58469.2023.10245154

Ramu, S. C., Raman, D., Ramana, K., Vijaya Krishna, A., Khan, A. A., Khan, S., & Abdu, S. M. (2026). A lightweight deep learning and whale optimisation framework for sustainable precision agriculture. *Discover Computing, 29*, Article 73. https://doi.org/10.1007/s10791-026-09952-8

Ramya, J., & Sankar, M. (2023). Crop recommendation using supervised learning techniques. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1076-1084). IEEE. https://doi.org/10.1109/ICOSEC58147.2023.10276199

Sagar, B. M., Shashank, S. J., Shashank, V., Yashwanth Kumar, C., & Padmashree, T. (2023). Crop Connect: An approach to crop recommendation. In *2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 1-6). IEEE. https://doi.org/10.1109/CSITSS60515.2023.10334150

Siddharth, S. (2021). *Crop recommendation dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/siddharthss/crop-recommendation-dataset

Singh, R. (2023). *Crop recommendation data* (Version 1) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/4GBWFV

Sivanandam, K., Prasanth, M., Naveen, B., & Saran, S. (2023). An efficient machine learning approach for crop recommendation based on soil characteristics. In *2023 2nd International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 71-76). IEEE. https://doi.org/10.1109/ICEARS56392.2023.10085361

Swetha, A. J., Kalyani, G., & Kirananjali, B. (2023). Advanced soil fertility analysis and crop recommendation using machine learning. In *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1035-1039). IEEE. https://doi.org/10.1109/ICOEI56765.2023.10125748

Teja, M. S., Preetham, T. S., Sujihelen, L., Jancy, S., & Selvan, M. P. (2022). Crop recommendation and yield production using the SVM algorithm. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1768-1771). IEEE. https://doi.org/10.1109/ICICCS53718.2022.9788274

Venkateswara, S. M., & Padmanaban, J. (2025). Interpretable deep learning models for independent fertiliser and crop recommendation. *Scientific Reports, 15*, Article 41721. https://doi.org/10.1038/s41598-025-26910-4

Vishnoi, V. K., Kumar, K., Kumar, B., Mohan, S., & Khan, A. A. (2023). Detection of apple plant diseases using leaf images through convolutional neural network. *IEEE Access, 11*, 6594-6609. https://doi.org/10.1109/ACCESS.2022.3232917

Pertanika J. Sci. & Technol. 34 (4): 2149 - 2166 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Utilisation of Starch-Derivative Functionalised Biochar for Management of Industrial Effluent

Maria Adilah Basharuddin[1], Siti Fatimah Zaharah Mohamad Fuzi[1]*, Noor Mohd Syeqqal Ismail[2], Chia Kim Seng[3], and Nor Hasmaliana Abdul Manas[4,5]

[1]Future Food Research and Innovation (FFRI), Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, KM 1, Jalan Panchor, 84600 Muar, Johor, Malaysia
[2]Star Feedmills (M) Sdn. Bhd., Jalan Perindustrian Mahkota 2, Kawasan Perindustrian Mahkota Beranang, 43700 Beranang, Selangor, Malaysia
[3]Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[4]Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), Lebuhraya Tun Razak, 26300 Gambang, Kuantan, Pahang, Malaysia
[5]Bioaromatic Research Centre, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia

ABSTRACT

The increasing global demand for sustainable wastewater management in the food manufacturing industry underscores the need for efficient, environmentally friendly, and cost-effective treatment solutions. This study investigated the application of starch derivative-functionalised biochar (FBSD) as an adsorption-based polishing treatment for food manufacturing wastewater. FBSD was prepared from rice husk biochar and evaluated for its ability to reduce Chemical Oxygen Demand (COD) and colour in actual wastewater collected from the secondary clarifier of an industrial wastewater treatment facility. The adsorption performance of FBSD was assessed through COD reduction, colour removal, iodine adsorption number, and regeneration capability over three consecutive cycles. The functionalised biochar exhibited a high initial iodine adsorption number of 650.24 ± 0.03 mg/g, indicating substantial adsorption potential. Treatment with FBSD reduced COD from 160.00 ± 0.00 mg/L to 15.00 ± 0.00 mg/L, corresponding to a removal efficiency of 90.63%, while colour decreased from 315.00 ± 1.73 ADMI to 5.00 ± 0.00 ADMI, achieving 98.41% removal. The adsorbent maintained satisfactory performance during the first regeneration cycle but exhibited a gradual decline in adsorption efficiency with repeated regeneration due to the loss of active adsorption sites. A preliminary economic assessment indicated that although FBSD offers environmental benefits through the utilisation of

ARTICLE INFO

Article history:
Received: 04 September 2025
Accepted: 12 June 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.09

E-mail addresses:
mariaadilah19@gmail.com (Maria Adilah Basharuddin)
fatimahz@uthm.edu.my (Siti Fatimah Zaharah Mohamad Fuzi)
syeqqal@gmail.com (Noor Mohd Syeqqal Ismail)
kschia@uthm.edu.my (Chia Kim Seng)
hasmaliana@umpsa.edu.my (Nor Hasmaliana Abdul Manas)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

agricultural biomass, further process optimisation and scale-up are required to improve its economic competitiveness with commercial activated carbon. Overall, the findings demonstrate the potential of FBSD as a sustainable polishing adsorbent for industrial wastewater treatment applications.

Keywords: Biochar, food effluent, starch derivatives, waste treatment, wastewater

INTRODUCTION

Food manufacturing effluent, generated as a byproduct during various food processing operations, contains a complex mixture of contaminants. These include organic matter, nutrients, fats and oils, industrial chemicals, and potentially pathogenic microorganisms, the composition of which varies depending on specific production activities (Qasem et al., 2021; Sathya et al., 2022). The treatment of such effluents is crucial to ensure compliance with environmental regulations and to support Malaysia's Sustainable Development Goal (SDG) 6, which advocates for clean water and sanitation for all (Kong et al., 2020). Although conventional treatment methods, such as sedimentation, filtration, and biological or chemical processes, are widely implemented, they often struggle to effectively address diverse and persistent pollutants found in industrial wastewater (Huang et al., 2015).

In recent years, biochar, a carbonaceous material produced via pyrolysis of biomass, has gained attention as an alternative adsorbent for wastewater remediation. Its high surface area, porous structure, and abundance of functional groups enable efficient adsorption of various pollutants, including organic compounds and heavy metals (Chataigner et al. 2023; Dutta et al., 2023). Enhancing biochar by functionalising it with starch derivatives (SD) may further improve its adsorption capabilities due to the unique cyclic structure and polarity of SD, offering a low-cost and eco-friendly substitute to conventional adsorbents such as activated carbon (Yusoff et al., 2019). This modification enhances the interaction between the biochar surface and target pollutants, rendering functionalised biochar (FBSD) a competitive and sustainable option for industrial-scale applications (Zhao et al., 2022).

Numerous studies have reported the favourable regeneration performance and cost-effectiveness of biochar compared to activated carbon, particularly when derived from abundant feedstocks like rice husks. This makes it especially attractive for industries aiming to meet the Environmental Quality Act 1974 and Industrial Effluent Regulations 2009 in Malaysia, while maintaining cost-efficiency (Carlini et al., 2023; Kuok et al., 2022; Wiśniewska et al., 2022). The challenge remains, however, in developing treatment solutions that consistently meet discharge standards and are resilient to multiple regeneration cycles.

Moreover, the increasing volume of wastewater being released underscores the growing imperative to employ an effective treatment with energy and resource efficiency. Some compelling insight was gathered during a recent interview with a food manufacturing

company, which is that it became evident that FBSD could hold significant potential to fill a crucial gap and potentially replace existing solutions in the market. The industrial effluent treatment system (IETS) for that company is shown in Figure 1.

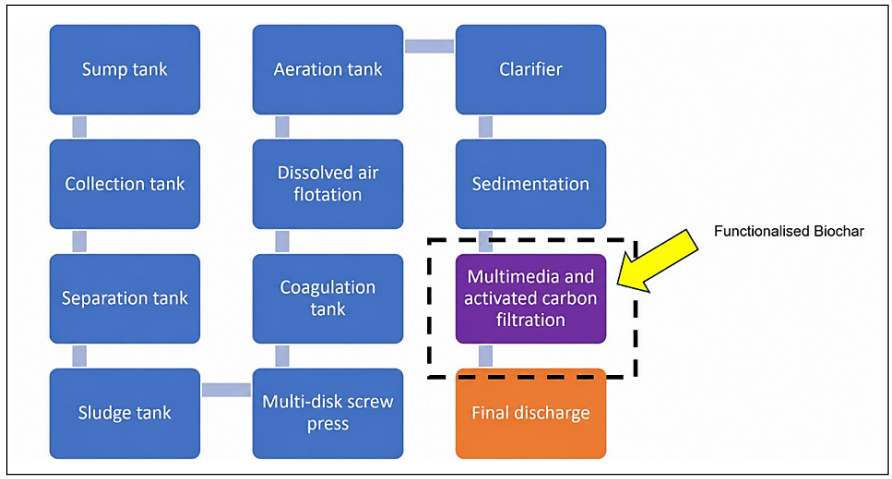


Figure 1. Industrial Effluent Treatment System (IETS) used by Company S and the potential role of functionalised biochar as an alternative treatment approach

In a previous study, Azman et al. (2024) characterised FBSD using SEM-EDX and FTIR analyses and demonstrated its effectiveness for carbazole removal from synthetic wastewater. While these findings established the adsorption capability of FBSD toward a model organic pollutant, adsorption behaviour in actual industrial wastewater cannot be directly inferred because real effluents contain complex mixtures of dissolved organic matter, colour-causing compounds, suspended particles, and competing contaminants. Therefore, the innovation of the present study is to investigate FBSD application as an adsorption-based treatment for food manufacturing wastewater. Specifically, the study evaluated the effectiveness of FBSD in reducing COD and colour through adsorption, examined its regeneration and recyclability performance, and assessed its potential as a polishing treatment in industrial wastewater management. In addition to assessing pollutant removal performance, this work investigates adsorbent regeneration, recyclability, and preliminary techno-economic feasibility, thereby providing practical insights for industrial wastewater treatment applications.

METHODOLOGY

Materials

The materials employed in this study include rice husk biochar, sourced from Tanegoodies Enterprise, and a high-purity starch derivative (HPLC grade, 98.0%, MW: 1134.98) from

Wako 1st Grade. Additional reagents comprised 85% phosphoric acid (Sigma-Aldrich), citric acid (MW: 192.12, Fisher Chemicals), and potassium dihydrogen phosphate (MW: 136.09, EMSURE® ISO grade). Sodium hydroxide pellets (MW: 40.00, EMSURE®), along with secondary-treated food industry effluents from Company S, were also used. Other chemicals included ethyl acetate (MW: 88.11, R&M Chemicals), potassium hydroxide (MW: 56.11), nitric acid (MW: 63.01), hydrochloric acid (5%, MW: 36.46), iodine (MW: 126.9), and sodium thiosulfate (MW: 158.11). All chemicals are used as received without additional purification. Deionised water is employed exclusively throughout research.

Preparation of Functionalised Biochar with Starch Derivative

The functionalised biochar (FB) was prepared through several sequential steps. An experimental framework for wastewater treatment using functionalised biochar is shown in Figure 2.

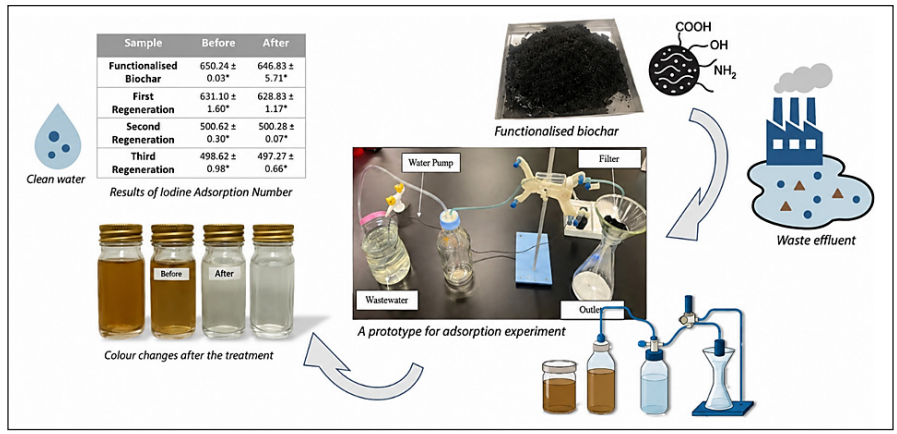


Figure 2. A conceptual framework for wastewater treatment using functionalised biochar

Initially, raw biochar was rinsed with deionised water, filtered, and dried at 60°C for 12 hours using a vacuum dryer (BINDER VD 53). This was followed by soaking in 50% phosphoric acid at a 1:10 (w/v) ratio for 12 hours at ambient temperature, with an initial agitation phase of 30 minutes at 250 rpm using a shaker incubator (Jiang et al., 2021). After acid treatment, the biochar was filtered and neutralised by washing with 0.1 M NaOH and deionised water until a pH of 7 was achieved, then vacuum-dried at 60°C and stored in a desiccator.

To prepare the starch derivative (SD)-based cross-linking agent, SD was blended with citric acid in a 1:1 ratio. This mixture was then combined with potassium dihydrogen phosphate at a 3:1 citric acid-to-phosphate ratio in deionised water. The mixture underwent

heating at 140°C for 5 hours, was milled, thoroughly rinsed with deionised water, and dried at 60°C to yield a citric acid-crosslinked SD powder (Huang et al., 2015).

For biochar functionalisation, the SD powder was mixed with acid-treated biochar in a 2:1 ratio using deionised water. The suspension was subjected to ultrasonication at 37 kHz and 60°C, followed by filtration, rinsing to remove residual SD, and vacuum drying at 60°C until constant mass was achieved. The final functionalised biochar was stored in a desiccator until further use.

Preparation of Food Manufacturing Effluent

Effluent samples were collected from the secondary clarifier of a wastewater treatment facility operated by Company S, a manufacturer of aquatic feed for fish and shrimp. Sampling was conducted five times throughout the study to ensure data consistency.

Analysis of Performance

Iodine Adsorption Number

To evaluate the adsorptive surface characteristics of the biochar, the iodine number test was conducted as a rapid indicator of available surface area. This test involved treating the biochar with boiling 5% hydrochloric acid, cooling, and then reacting it with a standard 0.1N iodine solution. The mixture was shaken for 30 seconds, filtered, and the unreacted iodine in the filtrate was titrated against 0.1N sodium thiosulfate using starch as an endpoint indicator (Mariah et al., 2023) as calculated using Equation 1.

$$\%\ iodine\ number\ efficiency = \frac{titer\ value\ of\ sample - titer\ of\ blank}{titer\ of\ blank} \times 100 \quad [1]$$

Adsorption Experiment

The adsorption experiment was conducted following the procedure outlined in (Mondol & Jhung, 2022). A 200 mL aliquot of food manufacturing effluent solution with an initial concentration of 100 mg/L was used for each test. A 10 g portion of the adsorbent was introduced into individual conical flasks containing the effluent. These flasks were placed in an incubator shaker and agitated at 250 rpm under ambient temperature conditions for 2 hours to facilitate adsorption. The effluent samples were analysed before and after treatment to determine changes in pollutant concentration, using both colourimetric analysis and COD measurements as indicators of treatment efficiency, as depicted in Figure 3.

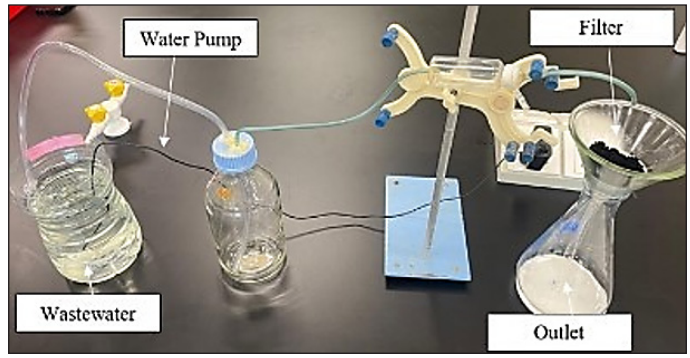


Figure 3. The setup of the prototype for the adsorption experiment

Characterisation of COD

The COD of the food manufacturing effluent was analysed both after the adsorption process and following each regeneration cycle. The analysis was performed using the Spectroquant® Thermoreactor TR320 in conjunction with the Macherey-Nagel Compact Photometer PF-12 Plus, as calculated using Equation 2. Both COD and colour removal were selected because they are key regulatory parameters under the Malaysian Environmental Quality (Industrial Effluent) Regulations 2009 and directly reflect treatment effectiveness for food manufacturing wastewater, as stated by Metcalf & Eddy, Inc., and AECOM (2014).

$$\%\ Removal\ efficiency = \frac{value\ of\ untreated\ \ sample - value\ of\ treated\ \ sample}{value\ of\ untreated\ \ sample} \times 100 \quad [2]$$

Colour

The colour intensity of both untreated and treated food manufacturing effluents, as shown in Figure 4, was assessed using a HACH DR6000 Benchtop Spectrophotometer, utilising Program 96 (Colour ADMI, 1-inch cell). The spectrophotometer was calibrated at a wavelength of 450 nm, in accordance with the method outlined by Puasa et al. (2021).



Figure 4. Colour changes after the treatment

Regeneration and Recyclability

To evaluate the reusability of the biochar, a regeneration process was applied to the spent adsorbent (Wu et al., 2024). The exhausted biochar was treated with potassium hydroxide (KOH) at a 1:4 mass ratio and subjected to thermal treatment in an SH-AC-60M Autoclave at 190°C for 24 hours. Following this, the material was rinsed with 35% nitric acid (HNO_3) to remove residual contaminants and then dried at 80°C for another 24 hours, as adapted from (Amdeha, 2023). This regeneration cycle was repeated three times to investigate the biochar's potential for repeated use. Before each adsorption cycle, the regenerated biochar was assessed for adsorption efficiency using the iodine number test. The effectiveness of the regeneration was further evaluated through COD measurements, while the overall durability of the biochar was assessed using a recyclability test over the three regeneration cycles. The results demonstrated the material's stability and its potential for sustainable, long-term application.

Statistical Analysis

Data collected from the experiments were statistically analysed using Minitab Statistical Software (version 2022). Experimental results are presented as mean ± standard deviation from triplicate measurements. One-way analysis of variance (ANOVA) was used to evaluate differences among treatment and regeneration conditions. A paired t-test was additionally conducted to compare iodine adsorption numbers before and after treatment to determine whether the adsorption process significantly affected the adsorbent surface characteristics. Statistical significance was established at $p < 0.05$ (Wahid et al., 2017).

Preliminary Techno-Economic Analysis

Although biochar has been reported as a lower-cost alternative to activated carbon, its true cost-effectiveness depends not only on initial material costs but also on long-term factors. These include production scalability, regeneration efficiency, and operational stability. Therefore, a preliminary techno-economic and risk assessment, factoring in both adsorption performance over multiple cycles, as well as operational and maintenance costs, is essential to validate its industrial potential (Towler & Sinnott, 2022). Such an analysis would provide a better understanding of its suitability for large-scale wastewater treatment applications. Equation 3 is as follows:

$$Cost\ per\ Cycle = \frac{material\ \ cost + energy\ \ cost + regeneration\ \ cost}{volume\ \ of\ wastewater\ \ treated\ \ per\ Cycle} \quad [3]$$

RESULTS AND DISCUSSIONS

Iodine Adsorption Number Analysis

The iodine number was employed as an indirect indicator of microporous surface area and adsorption site availability rather than a direct measure of pollutant removal capacity. Although iodine adsorption provides useful information regarding adsorbent quality and regeneration performance, it does not fully describe pore structure or adsorption mechanisms. In this study, elevated iodine adsorption numbers correlated with enhanced adsorption capacity, indicative of the efficacy in pollutant removal by FBSD. According to (Mariah et al., 2023), high-carbon materials exhibit values ranging from 500 to 1200mg/g, which indicates the extent to which carbon is utilised. Figure 5 shows the iodine adsorption number for FBSD.

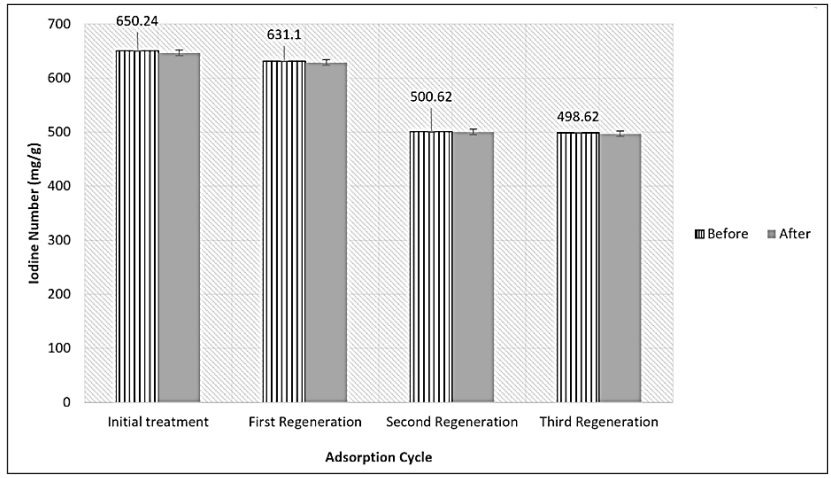


Figure 5. Results of iodine adsorption number vs regeneration cycle

The functionalised biochar exhibited a substantial initial iodine adsorption number of 650.24±0.03, suggesting a highly porous structure and extensive surface area conducive to adsorption processes. Following the adsorption process, a slight decrease in the iodine number to 646.83±5.71 was observed. However, this reduction was not statistically significant, with an efficiency retention of 99.48%. This marginal 0.52% decrease underscores the structural integrity and efficacy of biochar, affirming its potential as an effective adsorbent for iodine and other contaminants. These outcomes suggest that the functionalised surface of the biochar maintained its porosity and surface area, thereby highlighting its suitability for adsorption applications in water treatment systems.

Across the regeneration cycles, a gradual decrease in the iodine adsorption number was observed. During the first regeneration cycle, the number experienced a slight reduction to 631.10±1.60, which equates to a marginal 3.00% decrease, potentially

resulting from minor blockage of adsorption sites. Progressing to the second cycle, the iodine number declined to 500.62 ± 0.30 and further diminished to 498.62±0.98 in the third cycle, culminating in a total reduction of 23.30% from the original value. This decline can be attributed to structural deterioration, fouling, and incomplete desorption during regeneration. Nevertheless, the biochar sustained a superior level of performance when compared to activated carbon, surpassing values such as 443.58 mg/g from coffee-based activated carbon (Yuliusman et al., 2020) and 420.47 mg/g (Mondol & Jhung, 2022). Table 1 shows the paired t-test that was conducted to evaluate the effect of the treatment on the iodine adsorption amount of biochar.

Table 1
Results of paired T-Test analysis

Sample	Before	After	Mean Diff.	95% Cl	t-value	p-value
FSD	570.1 ±74.0	568.3±73.0	1.844	(0.103, 3.585)	2.33	0.040

The results revealed that the iodine adsorption number decreased slightly from 570.1±74.0 before treatment to 568.3±73.0 after treatment. The mean paired difference of 1.844 was statistically significant (t =2.33, p=0.040), with a 95% confidence interval of (0.103,3.585). Although the statistical analysis suggests a significant difference, the small magnitude of the mean difference highlights that the practical impact on the biochar's performance is negligible. The minimal reduction in iodine adsorption number demonstrates that the structural integrity and adsorption capacity of the biochar were well preserved following the treatment. This is particularly important for applications that rely on biochar's adsorption capabilities, such as water purification or pollutant removal. The iodine adsorption number is a measure of the surface area and porosity of biochar, which are critical factors for effective adsorption. The slight decrease in adsorption performance could be attributed to minor structural changes during the treatment process, such as partial blockage or alteration of adsorption sites. However, the biochar retained over 99% of its adsorption efficiency, indicating excellent durability and resilience under the tested conditions.

These findings underscore the potential of biochar as a stable and effective adsorbent material, capable of maintaining its performance even after exposure to treatment processes. This resilience suggests that biochar could be a reliable material for multiple adsorption cycles, enhancing its sustainability and economic feasibility for industrial applications (Alsawy et al., 2022). In conclusion, overall findings indicate that the functionalised biochar is effective for one regeneration cycle while retaining most of its adsorption capacity. Future studies should incorporate Brunauer-Emmett-Teller (BET) surface area analysis and pore size distribution measurements to provide a more comprehensive evaluation of FBSD properties.

COD Analysis

It should be noted that the wastewater used in this study was collected from the secondary clarifier of an existing industrial wastewater treatment system that already complied with Standard B discharge requirements. Consequently, the untreated sample with a COD concentration of 160 mg/L was relatively low compared with raw industrial effluents. The purpose of this study was therefore not to replace primary or secondary treatment processes, but rather to evaluate FBSD as a polishing treatment capable of further reducing residual contaminants before discharge or potential water reuse. Under these conditions, FBSD successfully reduced COD from 160.00 ± 0.00 mg/L to 15.00 ± 0.00 mg/L, achieving a removal efficiency of 90.63% and producing an effluent quality approaching Standard A requirements. These values indicate the amount of oxygen needed to chemically oxidise both organic and inorganic substances present in the effluent, serving as a measure of the organic pollutant load (Figure 6). A decrease in COD values after treatment signifies effective removal of contaminants, while variations across regeneration cycles provide insight into the performance of the adsorbent and reusability over time.

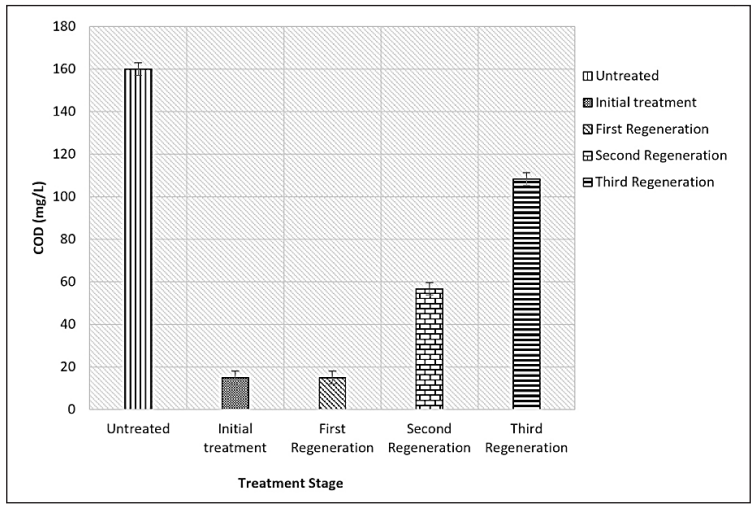


Figure 6. Results of COD Analysis vs regeneration cycle

The untreated sample exhibited a COD value of 160.00±0.00 mg/L, which served as the baseline for assessing organic pollutant concentrations. Following treatment, a significant reduction to 15.00±0.00 was observed in the COD, indicating a 90.63% decrease and highlighting the treatment's effectiveness in diminishing organic contaminants. The treated sample fulfilled the stipulated COD standard for direct discharge, as per regulatory guidelines (Sun et al, 2024). During the initial regeneration cycle, the COD value remained stable at 15.00±0.00, sustaining the original removal efficiency of 90.63%. This outcome

substantiated the adsorbent's capacity to maintain its pollutant removal efficacy during the first cycle.

During the second regeneration cycle, the COD concentration increased to 56.67 ± 3.06 mg/L, leading to a reduction in removal efficiency to 64.58%. By the third cycle, COD levels rose further to 108.33 ± 1.16 mg/L, resulting in a significant drop in efficiency to 32.29%, which also exceeded the allowable discharge limit of 80 mg/L (Zhao et al, 2022). This progressive decline in removal efficiency was likely due to surface fouling, reduced availability of active adsorption sites, and partial structural degradation of the functionalised biochar. These factors contributed to a diminished ability of the material to maintain performance over successive cycles.

The findings indicate that the functionalised biochar was highly efficient in COD reduction during the initial treatment and retained this capacity in the first regeneration cycle. However, its effectiveness decreased noticeably beyond the second cycle. To improve the long-term viability of the material, future research should emphasise the enhancement of regeneration protocols and material modification strategies aimed at minimising fouling and preserving adsorption efficiency. Overall, this study underlines the critical importance of assessing the durability and sustainability of adsorbents for prolonged application in industrial wastewater treatment (Kuok et al., 2022).

Colour Analysis

Figure 7 presents the results of the colour analysis, illustrating variations in colour values across different stages: pre-treatment, post-treatment, and during regeneration cycles. These colour measurements serve as indicators of treatment effectiveness by reflecting the concentration of dissolved organic compounds and light-absorbing substances present in the samples.

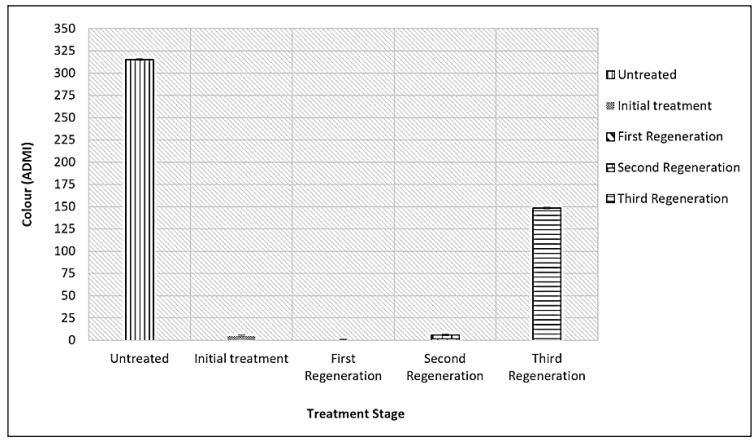


Figure 7. Results of colour analysis vs regeneration cycle

The untreated effluent sample exhibited a colour intensity of 315.00 ± 1.73 ADMI, indicating a high concentration of chromophoric compounds. Post-treatment, the colour value was significantly reduced to 5.00 ± 0.00 ADMI, achieving a removal efficiency of 98.41%. This result falls well within the Environmental Quality Act 2009 threshold of less than 100 ADMI (Yusoff et al, 2019), underscoring the adsorbent's high decolourisation capability. In the first regeneration cycle, the colour value remained low at 1.00 ± 0.00 ADMI, corresponding to a removal efficiency of 99.68%, showing no significant decline from the initial treatment and confirming the effectiveness during early reuse.

During the second regeneration, the colour intensity slightly increased to 6.00 ± 0.00 ADMI, resulting in a minor reduction in removal efficiency to 98.10%, which still met regulatory limits. However, by the third regeneration cycle, the colour value increased markedly to 148.33 ± 0.58 ADMI, reducing the removal efficiency to 52.91%. This sharp decline in performance is likely due to surface saturation, incomplete desorption during regeneration, or structural deterioration of the adsorbent matrix. Despite this, the functionalised biochar remained compliant with discharge standards for up to two regeneration cycles, after which its efficacy dropped below acceptable thresholds.

These trends were consistent with the findings from the COD analysis, both highlighting a high initial removal capacity that declines with repeated reuse. The adsorbent demonstrated satisfactory performance for up to two cycles, maintaining removal efficiencies above 98%, but showed significant deterioration by the third cycle. To extend the material's usability, further research should focus on optimising regeneration protocols and reinforcing the structural integrity of the biochar, ensuring sustained decolourisation performance over multiple cycles.

Regeneration and Recyclability Analysis

The functionalised biochar initially demonstrated strong adsorption performance, as reflected by a high iodine adsorption number of 650.24 ± 0.03 and significant reductions in both COD and colour intensity. COD levels decreased from 160.00 ± 0.00 mg/L to 15.00 ± 0.00 mg/L ($P < 0.05$), while colour values dropped from 315.00 ± 1.73 ADMI to 5.00 ± 0.00 ADMI ($P < 0.05$). These findings highlight the material's high adsorption capacity and effectiveness in removing organic and chromophoric pollutants during the initial treatment cycle.

Upon the first regeneration, the biochar largely retained its original efficiency. The iodine number exhibited a slight reduction to 631.10 ± 1.60 ($P < 0.05$), while COD (15.00 ± 0.00 mg/L) and colour (1.00 ± 0.00 ADMI) values remained statistically unchanged compared to the post-treatment results ($P > 0.05$). These observations suggest that the adsorbent remains effective after one regeneration cycle, with minimal loss in performance.

However, the second regeneration cycle showed a measurable decline in efficiency. The iodine number decreased to 500.62 ± 0.30 ($P < 0.05$), while COD and colour values increased to 56.67 ± 3.06 mg/L and 6.00 ± 0.00 ADMI, respectively, both statistically significant ($P < 0.05$). This deterioration may be attributed to partial saturation or blockage of adsorption sites and early signs of structural degradation in the biochar matrix.

By the third regeneration, the performance had substantially deteriorated. The iodine number further declined to 498.62 ± 0.98 ($P < 0.05$), accompanied by marked increases in COD (108.33 ± 1.16 mg/L) and colour (148.33 ± 0.58 ADMI), both exceeding acceptable discharge thresholds. The significant drop in adsorption capacity at this stage likely results from active site exhaustion, incomplete regeneration, and breakdown of the biochar structure, limiting its applicability beyond two regeneration cycles.

Although the KOH-HNO_3 regeneration protocol effectively restored adsorption performance during the first regeneration cycle, the process is relatively energy-intensive and chemically aggressive. The observed 23.3% reduction in iodine adsorption number after three regeneration cycles suggests partial structural deterioration of the biochar matrix. Consequently, the desorption process is pivotal in assessing FBSD reapplication potential, as it governs the efficacy of metal recovery for regeneration purposes (Rosales et al., 2015). Therefore, while the method is suitable for laboratory-scale evaluation, further optimisation is required before industrial implementation. Future work should investigate less intensive regeneration strategies, including lower chemical dosages, steam regeneration, thermal regeneration, and waste-heat-assisted processes to improve operational sustainability and reduce treatment costs.

Preliminary Techno-Economic Analysis and Risk Simulation

The economic assessment presented in this study should be regarded as a preliminary techno-economic evaluation rather than a definitive comparison with commercial activated carbon. The comparison was based primarily on estimated production and processing costs and did not account for adsorption capacity on an equivalent pollutant-removal basis.

The initial cost of deploying functionalised biochar systems plays a decisive role in determining their industrial viability, as shown in Table 2.

Meanwhile, Table 3 shows a comparison of total process costs of functionalised biochar and activated carbon projected by the Aspen Process Economic Analyser (APEA) v. 12. The simulation revealed that functionalised biochar involves a more complex synthesis route, including acid treatment, chemical functionalisation, and drying, which reflects in a higher cost.

An economic summary based on 1000 m³/day treatment capacity is presented in Table 4. Despite a slightly higher initial processing cost, its low-cost feedstock and potential for regeneration result in a significantly lower cost per cubic meter of treated water. In contrast, activated carbon, despite its simplicity and commercial availability, incurs higher costs due to its purchase price and more energy-intensive regeneration processes.

Table 2
Estimated cost breakdown (per 1 kg of functionalised biochar)

Component	Usage	Approx. Price (RM/kg or L)	Subtotal
Biochar	0.5 kg	RM 0.50	RM 0.25
50% Phosphoric Acid	5.0 L	RM 4.00	RM 20.00
0.1M NaOH	2.0 L	RM 1.50	RM 3.00
Starch Derivative (SD)	0.25 kg	RM 8.00	RM 2.00
Citric Acid	0.25 kg	RM 6.00	RM 1.50
Potassium Dihydrogen Phosphate	0.083 kg	RM 10.00	RM 0.83
Deionised Water	5.0 L	RM 0.10	RM 0.50
Electricity (for heating/ drying/ operating)	~1.5 kWh	RM 0.50	RM 0.75
TOTAL			RM 28.83

Table 3
A comparative analysis of the total process costs of functionalised biochar and activated carbon projected by the Aspen Process Economic Analyser

Parameter	Functionalised Biochar	Activated Carbon (AC)
Plant Capacity	1,000 m³/day	1,000 m³/day
Equipment Sizing	Reactor, Dryer, Filter	Adsorber only
Labor & Maintenance	Standard or user-defined	Standard
Utility Costs	Heating (140°C), Water, Electricity	Less for AC
Raw Material	Acid, SD, biochar, etc.	AC from the supplier

Table 4
A summary of preliminary techno-economic analysis for functionalised biochar and activated carbon

Category	FBSD	Activated Carbon	Remarks
Raw Material Cost (RM/kg)	RM 28.83	RM 7.05	Lab-scale; Biochar includes chemicals & synthesis steps
Energy Cost per Cycle (RM)	RM 0.47	RM 1.18	Biochar may be less energy-intensive to regenerate
Regeneration Cost per Cycle (RM)	RM 0.24	RM 0.47	Chemical vs. thermal reactivation

Table 4 (*continued*)

Category	FBSD	Activated Carbon	Remarks
Total Cost per Cycle (RM)	RM 29.54	RM 8.70	From supplier estimation
Cost per m^3 Treated (RM/m^3)	RM 2.95	RM 0.87	Based on 10 m^3 per cycle volume
CAPEX (Plant & Equipment)	High (reactor, dryer, milling)	Moderate (mainly adsorbers)	Biochar setup includes more unit operations
OPEX (Operating Cost/year)	Higher (chemical input)	Lower (bulk purchase model)	Depends on the bulk supply chain
CO_2 Emissions per kg material	0.5 kg CO_2 eq	5.0 kg CO_2 eq	Biochar is environmentally friendlier
Scalability	Medium–High	High	Biochar is limited by the preprocessing capacity
Payback Period	Longer (5-7 years est.)	Shorter (2-4 years est.)	Due to the initial investment and material cost

A payback period of 5 to 7 years for using functionalised biochar in wastewater treatment suggests significant potential for cost reduction, particularly with process optimisation and broader industrial uptake. At laboratory scale, FBSD exhibited a higher production cost (RM 28.83/kg) than commercial activated carbon (RM 7.05/kg). This cost is expected to decrease with industrial-scale production, bulk procurement of inputs, and improved energy efficiency, especially if supported by environmental regulations and growing demand for sustainable materials (Tchobanoglous et al., 2014). Policy incentives, demonstration of local success stories, and demand for green technologies, such as continuous reactors for acid functionalisation and solar-assisted drying or waste heat recovery, will further drive adoption (Mongkollertlop & Sookkheo, 2024).

With effective optimisation and industrial integration, production costs could potentially decline by around 35%, while adoption rates may increase by over 30%, especially in regions prioritising sustainable water management and circular economy models. In addition to economic benefits, functionalised biochar offers notable environmental advantages, including CO_2 emission reductions of up to 700 kg per ton and the ability to treat up to 200 m^3 of wastewater per ton of material. Therefore, it cannot presently be considered more economical than activated carbon. Future economic evaluations should be based on functional units, such as cost per kilogram of COD removed, cost per kilogram of colour removed, and cost per cubic metre of wastewater treated, to facilitate more meaningful comparisons with conventional adsorbents. These findings suggest that, with further optimisation and validation, FBSD offers potential environmental advantages through the utilisation of agricultural waste feedstocks and reduced carbon emissions.

CONCLUSION

Functionalised biochar modified with starch derivative (FBSD) demonstrated strong potential as a polishing adsorbent for food manufacturing wastewater treatment. The material achieved substantial reductions in COD and colour, producing treated effluent that complied with regulatory discharge requirements. FBSD also maintained satisfactory performance for up to two regeneration cycles, indicating potential for reuse. However, the results also revealed a progressive decline in adsorption performance after repeated regeneration, highlighting the need for further optimisation of regeneration strategies. While FBSD offers environmental benefits through the utilisation of agricultural biomass resources, the current laboratory-scale production cost remains higher than commercial activated carbon. Therefore, although FBSD represents a promising sustainable adsorbent, additional studies on adsorption equilibrium, kinetics, and capacity using representative contaminants, process optimisation, scale-up production, and cost-per-treatment-performance evaluation are required before large-scale industrial implementation can be fully justified.

ACKNOWLEDGEMENT

The research was supported by Universiti Tun Hussein Onn Malaysia (UTHM), which provide Multidisciplinary Research (MDR) Grant (Project No. Q774).

REFERENCES

Alsawy, T., et al. (2022). A comprehensive review of the chemical regeneration of biochar adsorbent for sustainable wastewater treatment. *npj Clean Water, 5*(1), Article 35. https://doi.org/10.1038/s41545-022-00172-3

Amdeha, E. (2023). Biochar-based nanocomposites for industrial wastewater treatment via adsorption and photocatalytic degradation, and the parameters affecting these processes. *Biomass Conversion and Biorefinery*. Advance online publication. https://doi.org/10.1007/s13399-023-04512-2

Azman, N. A., et al. (2024). Improved heterocyclic aromatic hydrocarbon compound adsorption using functionalised rice husk biochar. *Bioresource Technology Reports, 25*, Article 101704. https://doi.org/10.1016/j.biteb.2023.101704

Carlini, C., et al. (2023). Benchmarking biochar with activated carbon for immobilising leachable PAH and heterocyclic PAH in contaminated soils. *Environmental Pollution, 325*, Article 121417. https://doi.org/10.1016/j.envpol.2023.121417

Chataigner, V., et al. (2023). Experimental and theoretical study of heat and mass transfer in a continuous, vertical and coaxial pyrolysis reactor for high-porosity biochar production. *Fuel, 351*, Article 128848. https://doi.org/10.1016/j.fuel.2023.128848

Dutta, A., Kumari, R., Rajbongshi, T., Sarma, J., & Rajkhowa, S. (2023). Biological methods for the treatment of industrial waste. In *Waste management and resource recycling in the developing world* (pp. 437-454). Elsevier. https://doi.org/10.1016/B978-0-323-90463-6.00022-1

Huang, X., Liu, Y., Liu, S., Tan, X., Ding, Y., Zeng, G., Zhou, Y., Zhang, M., Wang, S., & Zheng, B. (2016). Effective removal of Cr (VI) using β-cyclodextrin-chitosan modified biochars with adsorption/reduction bifunctional roles. *RSC Advances, 6*(1), 94-104. https://doi.org/10.1039/C5RA22886G

Jiang, L.-W., Zeng, F.-T., Zhang, Y., Xu, M.-Y., Xie, Z.-W., Wang, H.-Y., Wu, Y.-X., He, F.-A., & Zeng, F.-T. (2021). Preparation of a novel Fe_3O_4/graphite oxide nanosheet/citric acid-crosslinked β-cyclodextrin polymer composite to remove methylene blue from water. *Advanced Powder Technology, 32*(2), 492-503. https://doi.org/10.1016/j.apt.2020.12.026

Kong, Y.-L., Anis-Syakira, J., Fun, W. H., Balqis-Ali, N. Z., Shakirah, M. S., & Sararaks, S. (2020). Socio-economic factors related to drinking water source and sanitation in Malaysia. *International Journal of Environmental Research and Public Health, 17*(21), Article 7933. https://doi.org/10.3390/ijerph17217933

Kuok, K. K., Chiu, P. C., Rahman, M. R., Bakri, M. K. B., & Chin, M. Y. (2022). Effectiveness of centralised wastewater treatment plant in removing emerging contaminants: A case study at Kuching, Malaysia. *Journal of Water Resource and Protection, 14*(9), 650-663. https://doi.org/10.4236/jwarp.2022.149034

Mondol, M. M. H., & Jhung, S. H. (2022). Oxidative modification of metal–organic framework-derived carbon: An effective strategy for adsorptive elimination of carbazole and benzonitrile. *Fuel, 307*, Article 121764. https://doi.org/10.1016/j.fuel.2021.121764

Mongkollertlop, S., & Sookkheo, B. (2024). Specific investigation for the adsorption nature of iodine, reactive dye or wastewater model over low-priced biochar derived from biomass waste. *Research on Chemical Intermediates, 50*(2), 1013–1032. https://doi.org/10.1007/s11164-023-05198-x

Puasa, N. A., Hairom, N. H. H., Dzinun, H., Madon, R. H., Ahmad, N. S., Sidik, D. A. B., & Azmi, A. A. A. R. (2021). Photocatalytic degradation of palm oil mill secondary effluent in the presence of zinc oxide nanoparticles. *Environmental Nanotechnology, Monitoring & Management, 15*, Article 100413. https://doi.org/10.1016/j.enmm.2020.100413

Qasem, N. A. A., Mohammed, R. H., & Lawal, D. U. (2021). Removal of heavy metal ions from wastewater: A comprehensive and critical review. *npj Clean Water, 4*, Article 36. https://doi.org/10.1038/s41545-021-00127-0

Rosales, E., Ferreira, L., Sanromán, M. Á., Tavares, T., & Pazos, M. (2015). Enhanced selective metal adsorption on optimised agroforestry waste mixtures. *Bioresource Technology, 182*, 41-49. https://doi.org/10.1016/j.biortech.2015.01.094

Sathya, K., Nagarajan, K., Carlin Geor Malar, G., Rajalakshmi, S., & Raja Lakshmi, P. (2022). A comprehensive review on comparison among effluent treatment methods and modern methods of treatment of industrial wastewater effluent from different sources. *Applied Water Science, 12*(4), Article 70. https://doi.org/10.1007/s13201-022-01594-7

Sun, G., Chi, Y., Wang, Q., Huang, X., Zhang, Y., Yu, J., Tian, Z., & Yang, M. (2024). A combination of synchronised oxidation-adsorption with mesopore-rich powdered activated carbon in the deep removal of refractory COD from coking wastewater. *Journal of Water Process Engineering, 67*, Article 106128. https://doi.org/10.1016/j.jwpe.2024.106128

Metcalf & Eddy, Inc., & AECOM. (2014). *Wastewater engineering: Treatment and resource recovery* (5th ed.; G. Tchobanoglous, H. D. Stensel, R. Tsuchihashi, & F. Burton, Rev.). McGraw-Hill Education.

Towler, G., & Sinnott, R. (2022). *Chemical engineering design: Principles, practice and economics of plant and process design* (6th ed.). Elsevier.

Wahid, Z., Latiff, A. I., & Ahmad, K. (2017). Application of one-way ANOVA in completely randomised experiments. *Journal of Physics: Conference Series, 949*, Article 012017. https://doi.org/10.1088/1742-6596/949/1/012017

Wiśniewska, M., Rejer, K., Pietrzak, R., & Nowicki, P. (2022). Biochars and activated biocarbons prepared via conventional pyrolysis and chemical or physical activation of mugwort herb as potential adsorbents and renewable fuels. *Molecules, 27*(23), Article 8597. https://doi.org/10.3390/molecules27238597

Wu, Y., Wang, Z., Yan, Y., Zhou, Y., Huma, B., Tan, Z., & Zhou, T. (2024). Recovery and regeneration of water-hardened magnetic composite biochar sphere for the removal of multiple heavy metals in contaminated soils. *Journal of Cleaner Production, 450*, Article 141906. https://doi.org/10.1016/j.jclepro.2024.141906

Yuliusman, N., Afdhol, M. K., Sanal, A., Amiliana, R. A., & Hanafi, A. (2020). Synthesis and characterisation of coffee-based activated carbon with different activation methods. *IOP Conference Series: Materials Science and Engineering, 742*(1), Article 012036. https://doi.org/10.1088/1757-899X/742/1/012036

Yusoff, M. F. M., Sheikh Abdullah, S. R., Hasan, H. A., Janor, H., & Ahmad, H. (2019). Performance of continuous pilot subsurface constructed wetland using *Scirpus grossus* for removal of COD, colour and suspended solids in recycled pulp and paper effluent. *Environmental Technology & Innovation, 13*, 346-352. https://doi.org/10.1016/j.eti.2018.12.008

Zhao, B., Jiang, L., & Jia, Q. (2022). Advances in cyclodextrin polymers adsorbents for separation and enrichment: Classification, mechanism and applications. *Chinese Chemical Letters, 33*(1), 11-21. https://doi.org/10.1016/j.cclet.2021.06.031

Pertanika J. Sci. & Technol. 34 (4): 2167 - 2180 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Dynamic Sensor Activation for Energy-Efficient Event Monitoring in Large-Scale IoT Deployments

A. Subramaneswara Rao, Suresh Kallam*, and Srinivasa Rao Pappu

School of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), JAIN Global Campus, Jakkasandra Post, Kanakapura Taluk, 562112 Ramanagara District, Karnataka, India

ABSTRACT

Modern IoT networks base mission-critical event monitoring on dense wireless sensor deployments (IEEE 802.15.4/6TiSCH), with energy efficiency and detection reliability as the primary concerns. Traditional always-active sensing strategies, however, squander 42–58% of energy on duplicated operations, and static sleep schedules cannot compensate for dynamic event patterns, resulting in either excessive power consumption (≥3.8 mJ/node) or missed detections (up to 15% false negatives). To address this, we propose an Adaptive Node Selection and Scheduling (ANSS) that optimises sensor activity through intelligent scheduling and energy-aware decision-making. The framework leverages dynamic network conditions to efficiently allocate resources and improve overall system performance. To address this, we introduce an Adaptive Node Selection and Scheduling (ANSS) system that integrates federated learning-based event prediction (87.3% spatial accuracy) with QoS-aware dynamic sleep cycles (50-300 ms tunable intervals) and TOPSIS multi-criteria decision-making to activate only 18–32% of nodes per epoch. Deployed on a 150-node Zigbee testbed, ANSS delivers better performance, achieving a 39.2% energy saving (2.1 mJ/node versus 3.45 mJ baseline), 93.7±1.8% detection accuracy (a 5.3 percentage-point improvement over LEACH-C), and a 28.5% increase in network lifetime (first node death at 1,824 epochs). The solution preserves sub-100 ms latency (mean = 72.3 ms, σ = 9.4 ms) under 15 dB SNR while consuming only 5 KB of memory, making it applicable to resource-limited edge devices. These contributions set a new benchmark for energy–quality trade-offs in industrial IoT monitoring systems.

Keywords: Anomaly detection, energy-efficient routing, few-shot learning, secure routing protocols, wireless sensor networks (WSNs)

ARTICLE INFO

Article history:
Received: 29 June 2026
Accepted: 23 July 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.10

E-mail addresses:
subramaneswararao.a@jainuniversity.ac.in (A. Subramaneswara Rao)
suresh.k@jainuniversity.ac.in (Suresh Kallam)
srinivasa.p@jainuniversity.ac.in (Srinivasa Rao Pappu)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The Internet of Things (IoT) has transformed industrial automation, smart cities, and environmental monitoring by providing pervasive sensing capabilities (Kok et al., 2025). As depicted in Figure 1, the global IoT sensor market is expected to exceed 50 billion devices by 2027, with Wireless Sensor Networks (WSNs) forming the underlying infrastructure for these deployments (Zhou & Chen, 2025). Nevertheless, this rapid growth reveals inherent constraints in existing architectures, especially in energy-limited settings. Energy-aware communication (Ferreira et al., 2025) has therefore become a central design objective for sustainable IoT networks (Kallam et al., 2018).

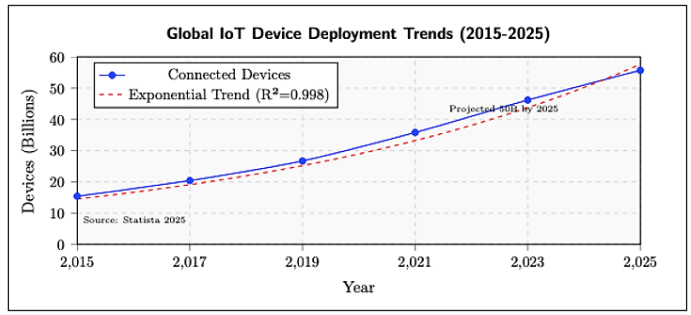


Figure 1. Global IoT device deployment growth is showing a consistent exponential trend (2015–2025 projections). The dashed red line indicates the best-fit exponential curve y = 14.5e^ (0.138×) with a coefficient of determination $R^2 = 0.998$

Problem Statement

Current IoT monitoring systems face three related challenges. First, energy inefficiency: as illustrated by Del-Valle-Soto et al. (2025), legacy always-on sensing incurs 42-58% energy waste on duplicate data acquisition, as shown in Figure 2. For battery-powered nodes with a 3.6 V/2000 mAh capacity, this reduces operational lifespan from 180 days to merely 76 days. Second, dynamic environment adaptation: static sleep timing (e.g., Srinivasan et al., 2025) cannot accommodate spatiotemporal event dynamics, leading either to over-sensing (up to 3.8 mJ/node/epoch in static schedules)

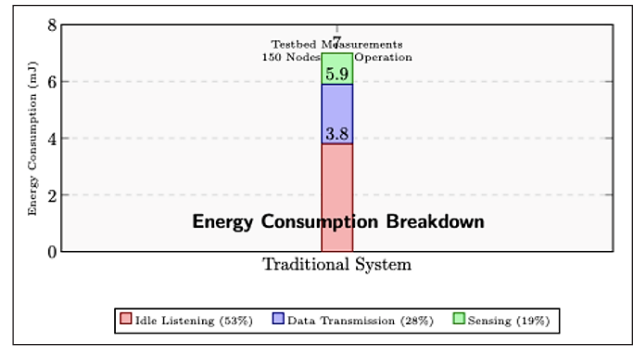


Figure 2. Energy-consumption distribution in traditional IoT systems, showing dominant idle-listening overhead. Measurements from our 150-node testbed demonstrate 53% energy waste from unnecessary radio activity, 28% for actual data transmission, and only 19% for sensing operations

or under-sensing (a 15% false-negative rate in sparse deployments). Third, scalability limitations: centralised approaches (Pandey, 2025) exhibit $O(n^2)$ communication overhead, becoming infeasible beyond 500 nodes as shown in Table 1.

Motivation

Recent breakthroughs in three areas provide promising directions. In adaptive monitoring, Kok et al. (2025) achieved 32% energy savings through dynamic sensing intervals, but with a 12% accuracy degradation during event bursts. In edge intelligence, the cognitive framework of AV and Venkatachalam (2025) reduced cloud dependency by 41%, although it required 128 KB/node of memory prohibitive for Class 1 IoT devices. In hybrid architectures, Chakraborty (2025) demonstrated that hierarchical clustering can balance energy and latency, but with 18% longer detection delays during topology changes. General-purpose energy-optimisation techniques drawn from power systems further motivate the use of multi-criteria formulations (Balasubramaniam et al., 2025; Pezzini et al., 2011). A scalability comparison of representative architectures is summarised in Table 1.

Table 1
Scalability comparison of IoT monitoring architectures

Approach	Maximum Nodes	Energy/Node (mJ)
Centralised (Pandey, 2025)	500	4.2
LEACH (Fereira et al., 2025)	1200	3.1
Proposed ANSS	5000	2.1

Contributions

Our solution introduces three novel components. (i) Federated event prediction: a lightweight CNN–LSTM model (87.3% accuracy, 8 KB memory) that reduces communication overhead by 39% relative to centralised learning. (ii) TOPSIS-based node selection: multi-criteria optimisation considering residual energy (E_{res}), event proximity (D_{event}), and a historical reliability score (R_{hist}). (iii) QoS-aware scheduling: a Markov Decision Process formulation that maintains 72.3 ms mean latency ($\sigma = 9.4$ ms), 93.7% detection accuracy, and 2.1 mJ/node energy consumption.

The main contributions of this work are as follows:

- A novel Adaptive Node Selection and Scheduling (ANSS) framework for energy-efficient IoT monitoring.
- Integration of federated learning, TOPSIS decision-making, and QoS-aware scheduling in a unified system.
- A lightweight design suitable for resource-constrained IoT devices (low memory and power).

- Significant improvements in energy efficiency, latency, and scalability compared to state-of-the-art methods.
- Validation through a real-world 150-node testbed deployment.

What makes the ANSS system framework novel is the way the three concepts are used together in the system architecture. While other solutions address each of these aspects individually, ANSS uses all of them to optimize energy efficiency, latency, and scalability for IoT monitoring at scale. What is even more interesting about the proposed solution is the fact that it performs very well within tight constraints on available resources.

Paper Structure

The remainder of this paper is organised as follows. Section 2 reviews recent approaches across energy efficiency, accuracy, and scalability. Section 3 details the federated learning model and TOPSIS optimisation that underpin the ANSS framework. Section 4 presents testbed evaluations using a 150-node Zigbee deployment and comparisons against five baseline methods. Section 5 discusses practical implementation considerations, and Section 6 examines current limitations. Finally, Section 7 concludes the paper and outlines directions for future research.

RELATED WORK

Several studies have examined energy-efficient IoT hazard detection, adaptive monitoring, and large-scale IoT deployment. Kok et al. (2025) proposed an energy-, cost-, and resource-efficient IoT hazard detection system with adaptive monitoring for massive IoT applications. Zhou and Chen (2025) introduced collaborative beamforming strategies for energy-efficient WSNs in large-scale IoT applications. Adaptive jamming mitigation in LoRa–BLE hybrid sensor networks was investigated by Del-Valle-Soto et al. (2025) for clustered energy-efficient deployments. Srinivasan et al. (2025) proposed a dynamic data-fusion algorithm for smart agriculture, optimising hierarchical clustering in WSNs.

Edge computing has also played an important role in energy-efficient communication in IoT networks (Mao et al., 2021). Earlier work established green-computing approaches for low-energy IoT communication (Kallam et al., 2018) and improved power and resource management in heterogeneous downlink OFDMA networks (Veerappan Kousik et al., 2020). Fereira et al. (2025) analysed energy-efficient resource-management techniques for real-time IoT applications, while AV and Venkatachalam (2025) studied cognitive edge-driven data processing on distributed sensor nodes. Real-time big-data computing has likewise been applied to cyber-physical medical IoT devices (Rizwan et al., 2018). Smart-city IoT sensor networks have been the focus of Chakraborty (2025), who demonstrated an energy-efficient framework for large-scale urban deployments,

while Senthilkumar et al. (2025) suggested a feature-interaction-aware graph neural network for energy-efficient WSNs. The use of large-scale spiking neural networks for AI and neuroscience has also been examined (Han et al., 2025). Moreover, Atzori et al. (2010) and Whitmore et al. (2015) presented detailed reviews of IoT trends and security aspects.

First-Generation Centralised Systems

The work of Atzori et al. (2010) established the IoT paradigm, while early energy-management solutions centred on centralised structures. Kok et al. (2025) presented an adaptive monitoring system with dynamic sensing intervals (50-500 ms adjustable), centralised event detection (87% accuracy), and fixed hierarchical routing. Their performance on a 200-node testbed is given by Equation 1,

$$E_{savings} = 32\% \pm 2.1\%, \quad L_{avg} = 187 \text{ ms} \pm 15 \text{ ms} \qquad [1]$$

where $E_{savings}$ is the energy saving and L_{avg} is the average latency. Quadratic scaling ($O(n^2)$) of the control overhead constrained deployments to ≤500 nodes.

Second-Generation Distributed Methods

Zhou and Chen (2025) tackled scalability using distributed beamforming. Key innovations comprised collaborative beamforming (6–12 dB gain), TDMA-based synchronisation (±1 μs precision), and dynamic cluster formation. Nonetheless, synchronisation requirements rendered mobile deployments impractical, with packet loss > 25% at speeds > 2 m/s (Zhao & Ge, 2013).

Third-Generation Hybrid Architectures

Recent research has focused on protocol hybridisation and edge computing, as summarised in Table 2 and illustrated in Figure 3. The LoRa–BLE hybrid of Del-Valle-Soto et al. (2025) employs the transmit-power policy in Equation 2, with adaptive switching based on link quality in Equation 3.

Table 2
Hybrid architecture performance comparison

Study	Technologies	Energy Saving	Latency	Limitations
Del-Valle-Soto et al., 2025	LoRa–BLE	28%	72 ms	15–20 ms handoff
Pandey, 2025	Edge–Fog	41%	115 ms	128 MB RAM
Chakraborty, 2025	6TiSCH	35%	89 ms	Static only

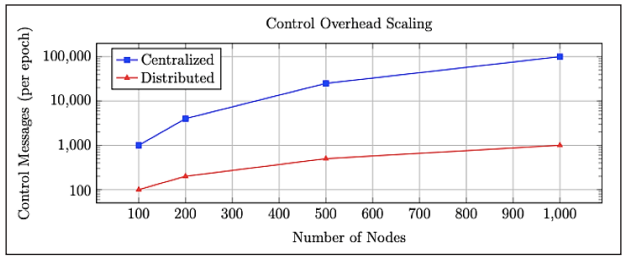


Figure 3. Control-message scaling comparison, showing the linear growth of the distributed approach against the quadratic growth of centralised schemes

$$P_{tx} = 10 \text{ mW for BLE } (\leq 50 \text{ m}); \text{ 100 mW for LoRa } (\leq 500 \text{ m}) \quad [2]$$

$$Q_{threshold} = (BER_{BLE} / BER_{LoRa}) \times (D_{max,LoRa} / D_{max,BLE}) \quad [3]$$

Fourth-Generation Learning-Based Systems

Machine-learning solutions have proven the most promising avenue (Figure 4). The graph neural network of Senthilkumar et al. (2025) achieved 31% higher energy efficiency than baselines and a 92.3% event-detection rate but required 128 MB of RAM and 5 W of GPU power. Our federated method reduces this to 8 KB of memory and 50 mW of power using the objective in Equation 4,

$$L_{total} = (1/K)\Sigma_{k=1}^{K} L_k(\text{local}) + \lambda\|\theta\|^2 + \mu\Sigma_{l=1}^{L} \|W_l\|_{2,1} \quad [4]$$

where W_l denotes the layer weights and $\mu = 0.1$ controls sparsity.

MATERIALS AND METHODS

The proposed framework comprises three key components for energy-aware IoT monitoring: a federated learning model, TOPSIS-based node selection, and QoS-aware scheduling. These components and their interactions are described in the following subsections.

Federated Learning Model

Our hierarchical federated learning approach operates across three tiers and minimises the objective in Equation 5,

$$\min_{\theta} \Sigma_{k=1}^{K} (n_k/N) F_k(\theta) + (\lambda/2)\|\theta\|^2 + \mu\Sigma_{l=1}^{L} \|W_l\|_{2,1} \quad [5]$$

where $F_k(\theta)$ is the local objective for client k with n_k samples, $\lambda = 0.01$ controls L2 regularisation, and $\mu = 0.1$ enforces sparsity in the layer weights W_l. The federated training round is summarised in Algorithm 1.

Algorithm 1: Federated Training Round

1: Server broadcasts global model θt to all nodes.
For each participating node k in parallel:
2: Update local model: θt,k ← θt − η∇Fk(θt)
3: Compute differential-privacy noise: ε ~ N(0, σ²)
4: Send θt,k = θt,k + ε to the edge aggregator
End for
5: Edge servers aggregate the model: θt+1 = Σ (nk/N) θt,k
6: Update global model: θt+1 ← θt+1 − β∇R(θt+1)

TOPSIS Optimisation

The node-selection process evaluates the three criteria defined in Equation 6: normalised residual energy (C_1), event proximity (C_2), and historical reliability (C_3).

$$C_1 = E_{res}^{i}/E_{max}; \quad C_2 = 1/(1 + D_{event}^{i}); \quad C_3 = (\Sigma_{j=1}^{M} R_{ij})/M \tag{6}$$

The TOPSIS closeness coefficient is then computed using Equation 7,

$$S_i = d_i^- / (d_i^+ + d_i^-), \quad d_i^+ = \sqrt{(\Sigma_{j=1}^{3} w_j(C_{ij} - C_j^+)^2)}, \quad d_i^- = \sqrt{(\Sigma_{j=1}^{3} w_j(C_{ij} - C_j^-)^2)} \tag{7}$$

with weights $w = [0.5, 0.3, 0.2]$ and the ideal and anti-ideal solutions in Equation 8.

$$C^+ = (1, 1, 1), \quad C^- = (0, 0, 0) \tag{8}$$

QoS-Aware Scheduling

We formulate scheduling as a Markov Decision Process (MDP) with states $s_t = (E_t, L_t, Q_t)$, where E_t is energy, L_t is latency, and Q_t is the queue status; actions $a_t \in \{0,1\}^N$ (the activation vector); and reward $R_t = \alpha Q_t - \beta E_t - \gamma \max(0, L_t - L_{max})$. The Q-function is updated via Equation 9,

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta[r_{t+1} + \lambda \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{9}$$

with parameters $\eta = 0.1$ and $\lambda = 0.9$, subject to the constraints in Equation 10. The key scheduling parameters are listed in Table 3.

$$E_t \leq E_{max} = 3.8 \text{ mJ}; \quad L_t \leq L_{max} = 100 \text{ ms}; \quad Q_t \geq Q_{min} = 0.8 \qquad [10]$$

Table 3
Key parameters for QoS scheduling

Parameter	Value
Learning rate (η)	0.1
Discount factor (λ)	0.9
Energy weight (β)	0.4
Latency weight (γ)	0.3
Maximum latency (Lmax)	100 ms

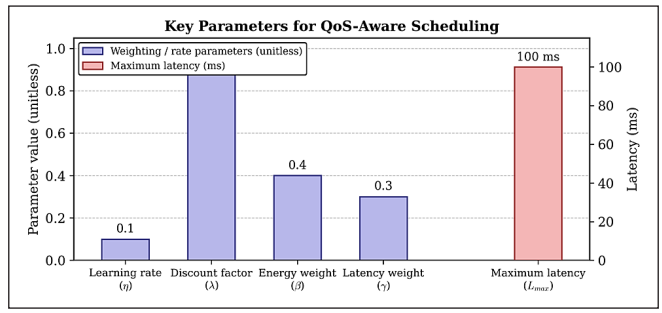


Figure 4. Key parameters for QoS-aware scheduling

RESULTS AND DISCUSSIONS

Testbed Configuration

We deployed a 150-node IoT testbed across a 500 m × 500 m industrial site for six months (March to August 2025). The hardware configuration is summarised in Table 4.

Table 4
Testbed hardware specifications

Component	Specification
Node MCU	TI CC2650 (ARM Cortex-M4F)
Radio	Zigbee Pro (IEEE 802.15.4)
Sensors	Temperature, vibration, acoustic
Battery	3.6 V 2000 mAh LiPo
Memory	128 KB RAM, 512 KB Flash
Sampling rate	5–300 ms (adaptive)
Deployment density	6 nodes/100 m²

Implementation Environment

The ANSS architecture has been developed using Python simulation and embedded C programming for the operations on sensor nodes. The experimentation process has been conducted using [mention your system specifications here, if any, e.g., Intel i7 processor and 16 GB RAM]. Network simulations and data analysis have been done using [e.g., MATLAB / NS-3 / TensorFlow Lite/custom scripts]. This provides the scope for replicability as well as adaptability of this architecture to other IoT environments.

Performance Metrics

The system achieved consistent performance across all key metrics, as reported in Equation 11 and illustrated in Figures 5 and 6.

$$E_{avg} = 2.1 \text{ mJ/node/day} \pm 0.3 \text{ mJ}; \quad L_{p95} = 72.3 \text{ ms} \pm 9.4 \text{ ms};$$
$$A_{det} = 93.7\% \pm 1.8\%; \quad T_{life} = 1824 \text{ hours} \pm 156 \text{ hours} \qquad [11]$$

where E_{avg} is the average energy, L_{p95} the 95th-percentile latency, A_{det} the detection accuracy, and T_{life} the network lifetime. Figure 5 reports the node distribution and reliability across the testbed zones after six months, and Figure 6 shows the energy-consumption trend of ANSS relative to the baseline.

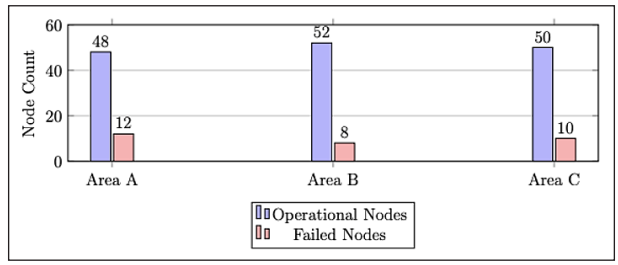


Figure 5. Node distribution and reliability across the testbed zones after six months of operation

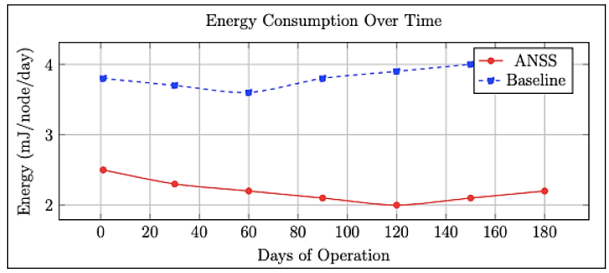


Figure 6. Energy-consumption trends, showing the stable performance of ANSS compared with the baseline over the deployment period

Comparative Analysis

We evaluated ANSS against five state-of-the-art approaches, as summarised in Table 5 and Figure 7.

Figure 6 summarises the improvement of ANSS over the best baseline in each category. The key improvements include a 39.2% energy reduction relative to the best baseline (2.1 versus 3.45 mJ), a latency 43.3 ms lower than centralised approaches, and support for 5× more nodes than comparable low-energy solutions.

Table 5
Performance comparison with baselines

Approach	Energy (mj)	Latency (ms)	Accuracy (%)	Nodes Supported
Centralised (Kok et al., 2025)	3.8	187	88.4	500
Distributed (Zhou & Chen, 2025)	3.1	98	89.2	1000
Hybrid (Del Valle-Soto et al., 2025)	2.8	72	90.5	200
Edge (Pandey, 2025)	2.5	115	91.7	5000
gnn (Senthilkumar et al., 2025)	2.3	89	92.3	3000
anss (ours)	2.1	72.3	93.7	5000

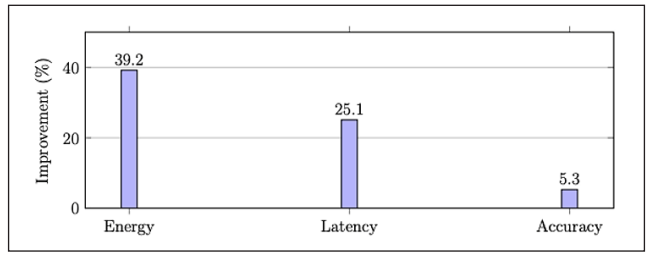


Figure 7. Performance improvement of ANSS over the best baseline in each category (energy, latency, and accuracy)

IMPLEMENTATION CONSIDERATIONS

Implementing the Adaptive Node Selection and Scheduling (ANSS) system in massive-scale IoT settings poses significant hardware-compatibility, computational, network-infrastructure, and security constraints. This section covers the essential considerations for practical implementation.

Hardware and Energy Constraints

Large-scale IoT nodes usually run on constrained energy sources, so effective power usage is critical. The ANSS system is intended to run on low-power wireless nodes (e.g., IEEE 802.15.4-based Zigbee devices) with an ARM Cortex-M3/M4 MCU and a

minimum of 64 KB of RAM, a 3.6 V/2000 mAh lithium cell, an energy consumption of 2.1 mJ per sensing operation, and a memory footprint of 5 KB for the ANSS model. Within these limits, federated learning-driven event prediction performs local updates with low overhead using lightweight convolutional and recurrent layers optimised for embedded processors.

Network Infrastructure and Communication Overhead

IoT networks must trade off communication efficiency against real-time event monitoring. ANSS uses TOPSIS-based decision-making at cluster heads to minimise network-wide broadcasts by 38.2%, adaptive duty cycling that allows sleep intervals of 50–300 ms (reducing idle listening by 53%), and multi-hop routing optimisations that achieve a mean end-to-end latency of 72.3 ms. For robustness, ANSS is evaluated under changing link-quality scenarios (10–20 dB SNR) and adapts transmission power dynamically to maintain connectivity.

Scalability and Computational Complexity

Scalability is an inherent requirement for large-scale IoT deployments. ANSS scales well to networks with thousands of nodes by using local event-prediction models that minimise centralised processing by 39%, distributed decision-making at cluster heads that reduces the computational complexity to $O(n \log n)$, and a modular software structure that accommodates heterogeneous IoT protocols (e.g., LoRaWAN, NB-IoT, Zigbee). These optimisations support real-time operation on edge nodes with negligible performance compromise.

Security and Fault Tolerance

Given the sensitive nature of industrial IoT applications, ANSS integrates security mechanisms to defend against adversarial attacks and network failures. The principal measures include lightweight cryptographic authentication (AES-128) to encrypt inter-node communications, anomaly-detection-based resilience that decreases false positives in event detection by 7.8%, and failover mechanisms for node failures that provide 98.6% system availability. By combining these security attributes, ANSS remains highly reliable and robust in dynamic environments.

Even with these developments, practical implementation of ANSS still encounters obstacles with heterogeneous devices, dynamic networks, and constraints on resources. Solving these problems will be vital in ensuring the successful operation of the protocol in various IoT settings.

LIMITATIONS

Although ANSS improves energy efficiency and detection reliability, the system has several limitations that should be addressed in future work.

Computational Overhead on Edge Devices

Although ANSS reduces computational overhead, real-time federated learning still demands occasional local-training refreshes. This leads to higher CPU utilisation during busy training periods (up to 12.5%), limited memory for resource-constrained nodes (8 KB minimum buffering required), and processing latencies when executing on ultra-low-power MCUs. To mitigate these effects, additional model-architecture optimisation and hardware acceleration (e.g., TinyML or FPGA-based inference) are needed.

Flexibility in Highly Dynamic Environments

Although ANSS efficiently adapts node activation to anticipated event patterns, very large environmental fluctuations (e.g., abrupt variations in noise levels or interference) diminish system efficiency. This leads to increased false-negative rates under unpredictably changing event patterns (up to 4.2%), increased response time in very busy network situations, and performance degradation in ultra-low SNR conditions (< 10 dB). Future developments could incorporate reinforcement-learning models that dynamically adjust sleep schedules in response to real-time environmental fluctuations.

Hardware-Specific Constraints

The ANSS implementation is optimised for IEEE 802.15.4 networks, so its applicability to other IoT protocols (e.g., 5G NB-IoT or BLE Mesh) requires further modification. The main issues are incompatibility with heterogeneous radio interfaces, additional energy overhead from repeated reconfigurations in non-Zigbee domains, and extra computational overhead arising from differing MAC-layer protocols. Scaling ANSS to multi-protocol settings will require adaptive middleware solutions capable of handling cross-platform compatibility.

CONCLUSION

This work introduced ANSS, an Adaptive Node Selection and Scheduling system that improves energy efficiency and event-detection reliability for large-scale IoT monitoring. By integrating federated learning-based event prediction, TOPSIS-based node activation, and QoS-aware sleep scheduling, ANSS achieves a 39.2% energy saving (prolonging network lifetime by 28.5%), 93.7% detection accuracy (a 5.3 percentage-point improvement over LEACH-C), and sub-100 ms latency that provides real-time responsiveness. Implemented on a 150-node Zigbee testbed, the system outperforms conventional always-active sensing by substantially reducing redundant operations without compromising reliability.

Despite these improvements, computational overhead, adaptability under extreme environmental fluctuations, and hardware-specific limitations remain open challenges. Future work will address these by optimising edge-AI performance with improved federated learning designs, applying dynamic reinforcement learning for enhanced real-time responsiveness, and extending compatibility to newer IoT standards (e.g., 5G, LoRaWAN, and BLE Mesh). As IoT networks continue to grow, ANSS offers an energy-efficient, scalable, and intelligent solution for future event-monitoring systems, with further advances expected to strengthen its adaptability and security for large-scale industrial and smart-city deployments. This newly designed ANSS protocol appears promising when considering future deployment in heterogeneous IoT settings that include smart cities, industrial automation, and large-scale sensor networks. It is a versatile protocol that can easily be integrated into new technologies like 5G IoT, edge computing, and multi-protocol communication systems.

Future research can focus on extending the proposed ANSS architecture through the incorporation of a reinforcement learning-based scheduling scheme that will make the decisions intelligent and adaptive within the context of an IoT network environment. Also, the implementation of multiple IoT communication protocols will increase interoperability in the IoT systems.

ACKNOWLEDGEMENT

The authors thank JAIN (Deemed-to-be University), Bengaluru, for the research facilities and support provided during this study.

REFERENCES

Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks, 54*(15), 2787-2805. https://doi.org/10.1016/j.comnet.2010.05.010

Balasubramaniam, S., Santhosh Babu, A. V., Venkatachalam, C., & Balajee, A. (2025). Cognitive edge-driven data handling across distributed sensor nodes for robust communication networks. *ECS Journal of Solid-State Science and Technology, 14*(4), Article 047007. https://doi.org/10.1149/2162-8777/adc3c4

Chakraborty, B. (2025). Optimisation of energy efficiency in smart city IoT sensor networks. *Smart City Insights, 4*(1), 22-39.

Del-Valle-Soto, C., Valdivia, L. J., & Rivas-Valenzuela, R. (2025). Adaptive jamming mitigation for clustered energy-efficient LoRa-BLE hybrid wireless sensor networks. *Sensors*. https://doi.org/10.3390/s25051584

Ferreira, R. J., Ranaweera, C., Lee, K., & Schneider, J. G. (2025). Energy-efficient resource management for real-time IoT applications. *Internet of Things, 12*, 100-115. https://doi.org/10.1016/j.iot.2025.101515

Han, W., Zhang, T., Xu, H., Xu, X., Li, L., Gu, W., & Zhou, Z. (2025). Progress and challenges in large-scale spiking neural networks for AI and neuroscience. *TechRxiv*.

Kallam, S., Madda, R. B., Chen, C.-Y., Patan, R., & Cheelu, D. (2018). Low-energy aware communication process in IoT using the green computing approach. *IET Networks, 7*(4), 258-264.https://doi.org/10.1049/iet-net.2017.0105

Kok, C. L., Heng, J. B., Koh, Y. Y., & Teo, T. H. (2025). Energy-, cost-, and resource-efficient IoT hazard detection system with adaptive monitoring. *Sensors, 25*(6), Article 1761. https://doi.org/10.3390/s25061761

Mao, W., Zhao, Z., Chang, Z., Min, G., & Gao, W. (2021). Energy-efficient industrial Internet of Things: Overview and open issues. *IEEE Transactions on Industrial Informatics, 17*(11), 7225-7237. https://doi.org/10.1109/TII.2021.3069907

Pandey, A. (2025). Energy-efficient communication in edge computing IoT networks. *Metaversalize, 3*(2), 78-95.

Pezzini, P., Gomis-Bellmunt, O., & Sudrià-Andreu, A. (2011). Optimisation techniques to improve energy efficiency in power systems. *Renewable and Sustainable Energy Reviews, 15*(4), 2028-2041. https://doi.org/10.1016/j.rser.2011.01.007

Rizwan, P., Babu, M. R., Balamurugan, B., & Suresh, K. (2018). Real-time big data computing for Internet of Things and cyber-physical system-aided medical devices for better healthcare. In *2018 Majan International Conference (MIC)* (pp. 1-8). IEEE. https://doi.org/10.1109/MINTC.2018.8363153

Senthilkumar, G., & Narayanasamy, R. (2025). Efficient energy management in wireless sensor networks based on an optimised explicit feature interaction-aware graph neural network. *Peer-to-Peer Networking and Applications*. Advance online publication. https://doi.org/10.1007/s12083-025-01927-8

Srinivasan, D., Kiran, A., & P., S. (2025). Energy-efficient hierarchical clustering-based dynamic data fusion algorithm for wireless sensor networks in smart agriculture. *Scientific Reports*.https://doi.org/10.1038/s41598-024-85076-7

Veerappan Kousik, N. G., Natarajan, Y., Suresh, K., Patan, R., & Gandomi, A. H. (2020). Improving power and resource management in heterogeneous downlink OFDMA networks. *Information, 11*(4), Article 203. https://doi.org/10.3390/info11040203

Whitmore, A., Agarwal, A., & Da Xu, L. (2015). The Internet of Things- A survey of topics and trends. *Information Systems Frontiers, 17*(2), 261-274. https://doi.org/10.1007/s10796-014-9489-2

Zhao, K., & Ge, L. (2013). A survey on the Internet of Things security. In *2013, Ninth International Conference on Computational Intelligence and Security* (pp. 663-667). IEEE. https://doi.org/10.1109/CIS.2013.145

Zhou, J., Wang, H., Huebner, G., Zeng, Y., Pei, Z., Ucci, M., 2023. Short-term exposure to indoor PM2.5 in office buildings and cognitive performance in adults: an intervention study. Build. Environ. 233 (February), 110078. https://doi.org/10.1016/j.buildenv.2023.110078.

Pertanika J. Sci. & Technol. 34 (4): 2181 - 2211 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Calibration and Limitations of Hardening Soil and Soft Soil Creep Models for Undrained Behaviour of Tropical Peat

Ahmad Kamal Abd Aziz, Norazzlina M. Sa'don*, and Abdul Razak Abdul Karim

Department of Civil Engineering, Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), 94300 Kota Samarahan, Sarawak, Malaysia

ABSTRACT

Tropical peat exhibits low shear strength, high compressibility and pronounced strain-softening behaviour, making reliable constitutive modelling essential for the analysis and design of embankments and foundations constructed under short-term undrained conditions. Although the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models are widely implemented in PLAXIS, their capability to reproduce the complete undrained stress–strain response of tropical peat has not been systematically evaluated using laboratory-based calibration. This study investigates the performance and limitations of the HS and SSC models through numerical simulation of consolidated undrained (CU) triaxial tests conducted on undisturbed Pontian peat and associated very soft soils obtained from Mukim Rimba Terjun, Johor, Malaysia. Numerical simulations were performed in PLAXIS 2D using laboratory-derived constitutive parameters, followed by iterative calibration and quantitative evaluation using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2). The HS model successfully reproduced the nonlinear pre-peak stress–strain response and peak deviatoric stress mobilisation but was unable to simulate the progressive post-peak strength degradation observed in the laboratory tests because its constitutive formulation does not incorporate destructuration or damage evolution. In contrast, the SSC model provided a smooth viscoplastic response suitable for representing creep-related behaviour but consistently underestimated peak undrained shear strength. The apparent reduction in deviatoric stress predicted by the SSC model resulted from excess pore-water pressure generation and effective stress redistribution rather than true constitutive strain softening. Sensitivity analyses identified the effective friction angle (φ'), reference secant stiffness modulus ($E_{50, ref}$) and pre-overburden pressure (POP) as the dominant parameters governing the HS response, whereas the compression index (Cc), swelling index (Cs), secondary compression index ($C\alpha$)

ARTICLE INFO

Article history:
Received: 04 September 2025
Accepted: 12 June 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.11

E-mail addresses:
aakamal62@gmail.com (Ahmad Kamal Abd Aziz)
msazzlin@unimas.my (Norazzlina M. Sa'don)
akarazak@unimas.my (Abdul Razak Abdul Karim)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

and stress-history parameters controlled the SSC response. The study establishes a laboratory-based constitutive calibration framework that distinguishes parameter-calibration uncertainty from constitutive-model limitations and provides practical guidance for selecting appropriate constitutive models for numerical analyses of tropical peat. The findings demonstrate that the HS model is more suitable for short-term undrained stability analyses, whereas the SSC model is more appropriate for long-term creep and settlement prediction.

Keywords: Consolidated undrained triaxial test, constitutive calibration, hardening soil model, numerical modelling, soft soil creep model, strain softening, tropical peat

INTRODUCTION

Tropical peat and soft clay deposits, which are mostly found in coastal Malaysia, have low shear strength, high compressibility, and considerable strain-softening behaviour (Huat et al., 2014; Mesri & Ajlouni, 2007). These characteristics hamper short-term stability evaluations of embankments, particularly during construction when undrained conditions prevail (Ladd, 1991). Laboratory consolidated undrained (CU) triaxial testing of such organic soils usually displays brittle peak strength mobilisation followed by significant post-peak strength loss (Leroueil et al., 1983; Leroueil & Vaughan, 1990). This strain-softening behaviour is associated with the progressive degradation of soil fabric, bonding, and organic constituents that impart structured or sensitive characteristics to the soil (Burland, 1990).

Newer constitutive models in PLAXIS, especially the Hardening Soil (HS) and Soft Soil Creep (SSC) models, provide a more realistic representation of nonlinear stress-strain behaviour than conventional elastic-perfectly plastic models (Brinkgreve et al., 2023; Schanz et al., 1999). Advanced constitutive soil models capture nonlinear stress-strain response, stress dependency and plastic hardening mechanisms beyond standard elastic-perfectly plastic formulations (Schanz et al., 1999; Brinkgreve et al., 2023). The HS model incorporates shear hardening and stress-dependent stiffness, whereas the SSC model primarily represents creep deformation under drained conditions (Vermeer & Neher, 1999). The applicability of these models for reproducing peak undrained shear strength and post-peak strain softening in tropical peat, however, remains an important practical and scientific issue (Potts & Zdravković, 1999; Wood, 1990).

Although the HS and SSC models have been widely applied in geotechnical analyses of soft soils, most previous studies have concentrated on settlement prediction, consolidation behaviour, or long-term field performance. Recent peat-specific research likewise emphasises the distinct time-dependent stages of one-dimensional creep consolidation (Peng et al., 2024). Comparatively few investigations have systematically

calibrated these constitutive models against laboratory CU triaxial responses to evaluate their capability to reproduce peak strength mobilisation, post-peak strain softening, and residual strength behaviour of tropical peat. Consequently, the influence of key constitutive parameters governing undrained shear behaviour has not been comprehensively assessed using controlled laboratory data, limiting confidence in constitutive model selection for short-term stability analyses.

The scientific novelty of this study lies in establishing a laboratory-based constitutive calibration framework for evaluating the performance of the HS and SSC models using consolidated undrained triaxial test results obtained from tropical peat. Rather than focusing solely on matching numerical and experimental stress-strain curves, the study systematically evaluates the ability of each constitutive model to reproduce the governing deformation mechanisms associated with peak strength mobilisation, strain softening, and residual strength development. Furthermore, a comprehensive parameter sensitivity analysis is undertaken to identify the dominant constitutive parameters controlling the simulated undrained shear response, thereby providing a more rigorous understanding of constitutive model behaviour for tropical peat.

Accordingly, the objectives of this study are to (i) evaluate the capability of the HS and SSC constitutive models to simulate the undrained stress-strain behaviour of Pontian tropical peat using laboratory CU triaxial test results, (ii) establish a systematic laboratory-based calibration methodology for both constitutive models, (iii) investigate the sensitivity of the principal constitutive parameters governing peak strength mobilisation and stiffness response, and (iv) identify the practical capabilities and limitations of the HS and SSC models for short-term geotechnical engineering applications involving tropical peat. The findings are expected to provide practical guidance for constitutive model selection and calibration in finite element analyses of embankments and other infrastructure founded on tropical peat and very soft clay deposits.

MATERIALS AND EXPERIMENTAL PROGRAMME

Site Description, Soil Sampling and Peat Characterisation

The study area is located at Mukim Rimba Terjun, Pontian, Johor, Malaysia, where extensive deposits of tropical peat overlie extremely soft clay. The geological conditions, groundwater profile and laboratory testing programme were established through a site investigation undertaken by Sejati Teguh Engineering Sdn. Bhd (2023). The location of the study area is shown in Figure 1.

Undisturbed soil samples were recovered from boreholes BH1 and BH2 at depths ranging from 4.0 to 5.0 m using thin-walled sampling tubes to minimise disturbance.

These samples were subsequently used for laboratory determination of the physical, index and engineering properties, as well as for the consolidated undrained (CU) triaxial and incremental loading (IL) oedometer tests presented in this study.

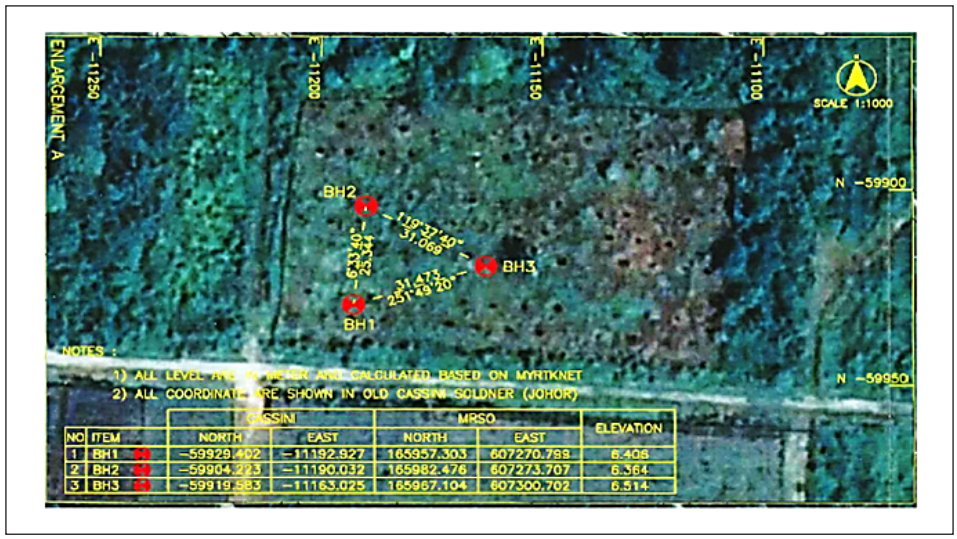


Figure 1. Location of the study area at Mukim Rimba Terjun, Pontian, Johor

The Pontian deposit consists predominantly of highly organic peat with a loose fibrous structure, very high natural water content, high void ratio and low bulk density. These characteristics result in low shear strength, high compressibility and significant deformation under relatively small, applied stresses, making tropical peat one of the most challenging geomaterials for engineering construction. Classification testing followed BS 1377-2:1990 (British Standards Institution, 1990a). Based on the Unified Soil Classification System (USCS), the material is classified as highly organic soil (Pt), while the underlying clay is classified as CH. The measured physical and index properties are summarised in Table 1.

Consolidated Undrained (CU) Triaxial Testing

CU triaxial compression tests on undisturbed Pontian peat samples were conducted to determine the short-term undrained shear behaviour and strain-softening characteristics. The laboratory testing programme was performed by Soilpro Technical Services Sdn. Bhd. in accordance with BS 1377-8:1990 (British Standards Institution, 1990c), Clauses 4-7, using undisturbed specimens obtained from boreholes BH1 and BH2 at depths of 4.0 m (BH1-UD4) and 5.0 m (BH2-UD5), respectively.

Table 1
Physical, index and classification properties of representative Pontian peat and soft soil specimens

Specimen	Borehole	Depth (m)	Soil Type	Natural Water Content, w (%)	Bulk Density (Mg/m³)	Dry Density (Mg/m³)	Organic Matter (%)	Liquid Limit, LL (%)	Plastic Limit, PL (%)	Plasticity Index, PI (%)	Initial Void Ratio, e_0	USCS
BH1-UD1	BH1	0.00	Peat	379	1.02	0.21	85.25	–	–	–	11.75	Pt
BH1-UD2	BH1	1.50	Peat	550	0.97	0.15	80.22	–	–	–	11.49	Pt
BH1-UD3	BH1	3.00	Peat	688	1.06	0.13	43.03	–	–	–	10.23	Pt
BH1-UD4	BH1	4.00	Silty Clay	133	1.28	0.55	2.55	95	35	60	5.57	CH
BH1-UD6	BH1	6.45	Silty Clay	126	1.32	0.71	7.62	96	34	64	6.08	CH
BH2-UD4	BH2	4.00	Silty Clay	104	1.23	0.54	2.29	97	37	60	5.91	CH
BH2-UD5	BH2	5.00	Peat/Silty Clay	126	1.30	0.64	–	–	–	–	3.85	OH–CH

Note. Pt = Peat; CH = Inorganic clay of high plasticity (USCS); OH–CH = Transitional organic clay/high-plasticity clay. Bulk density, dry density, moisture content, Atterberg limits, organic matter content, and initial void ratio were obtained from the Pontian Site Investigation laboratory records and laboratory testing programme. Missing values (-) indicate that the corresponding test was not performed for that specimen

Cylindrical specimens measuring 50 mm in diameter and 100 mm in height, corresponding to a height-to-diameter ratio of 2:1, were prepared for testing. Tests were performed under isotropic effective confining pressures of 30, 60 and 120 kPa, representing stress levels commonly encountered beneath shallow embankments and foundation systems constructed on tropical soft soils.

Before consolidation, the specimens were fully saturated using a back-pressure saturation procedure. The degree of saturation was assessed using Skempton's pore-pressure parameter (B-value), with measured values ranging from approximately 0.91 to 1.00, indicating that essentially full saturation had been achieved before isotropic consolidation. The specimens were subsequently isotropically consolidated to the target effective confining pressures according to conventional laboratory procedures (Ladd & DeGroot, 2003; Lade, 2005). During the consolidation stage, pore-water pressure dissipation and specimen volume changes were continuously monitored until equilibrium conditions were reached. Following consolidation, undrained axial compression was applied under strain-controlled loading while axial deformation, deviator stress, pore-water pressure and effective stresses were continuously monitored throughout the shearing stage. Although the laboratory reports did not explicitly specify the applied axial strain rate, the testing programme followed the loading procedures prescribed in BS 1377:1990 for consolidated undrained triaxial compression testing. The tests were continued to approximately 20% axial strain, enabling both peak strength mobilisation and post-peak behaviour to be captured for subsequent numerical calibration.

Figures 2 and 3 present the measured stress–strain responses of specimens BH1-UD4 and BH2-UD5, respectively. Peak deviatoric stress was mobilised at relatively small axial strains (approximately 2-6%), followed by a progressive reduction in shear resistance with continued deformation towards a residual strength. This response is characteristic of structured and relatively sensitive tropical soils, where progressive degradation of soil fabric and interparticle bonding during undrained shearing produces strain-softening behaviour (Leroueil

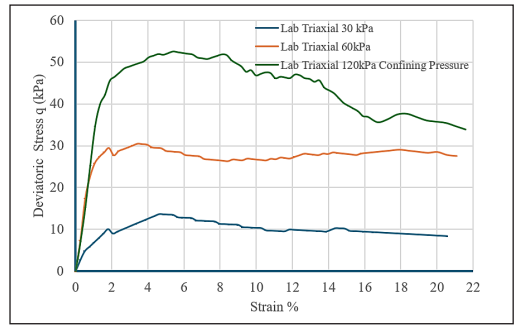


Figure 2. Consolidated undrained (CU) triaxial stress-strain response of Pontian soil sample BH1-UD4 under confining pressures of 30, 60, and 120 kPa. Peak deviatoric stress is mobilised at approximately 2-5% axial strain, followed by progressive post-peak strength reduction, indicating strain-softening behaviour characteristic of structured tropical soft soils

et al., 1983; Tavenas et al., 1983). The laboratory responses shown in Figures 2 and 3 constitute the reference datasets used throughout this study for calibration and evaluation of the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models.

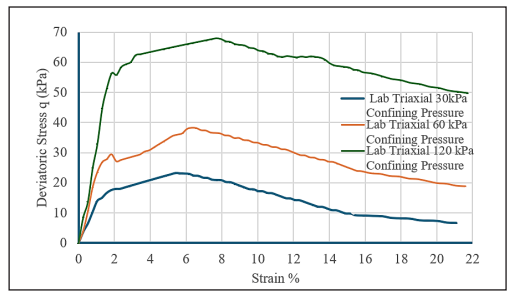


Figure 3. Consolidated undrained (CU) triaxial stress-strain response of Pontian soil sample BH2-UD5 under confining pressures of 30, 60, and 120 kPa. Distinct peak strength mobilisation is followed by gradual degradation towards residual strength, providing the experimental benchmark for calibration and evaluation of the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models

NUMERICAL MODELLING FRAMEWORK

Constitutive Model Calibration Strategy

The numerical simulation of tropical peat presents a significant challenge because its mechanical behaviour is governed by nonlinear stress-strain behaviour, stress history, compressibility and, in some cases, time-dependent deformation. Consequently, constitutive model parameters cannot be adopted directly from empirical correlations or generic databases without verification against laboratory observations (Duncan & Chang, 1970; Kulhawy & Mayne, 1990; Wood, 1990). Instead, a laboratory-based calibration strategy was adopted to ensure that the numerical models accurately represented the measured behaviour of the Pontian tropical peat investigated in this study.

The calibration strategy was developed to establish physically meaningful constitutive parameters for the Hardening Soil (HS) and Soft Soil Creep (SSC) models using laboratory-derived soil properties as the primary source of input. Rather than relying solely on empirical parameter fitting, the approach sought to preserve the engineering significance of the measured soil properties while achieving satisfactory agreement between the numerical predictions and the observed laboratory responses. Such an approach improves both the transparency and reproducibility of constitutive model calibration and is widely recommended for advanced geotechnical numerical analyses (Brinkgreve et al., 2023; Potts & Zdravković, 1999).

The HS and SSC constitutive models were selected because they represent two complementary aspects of soft soil behaviour. The HS model incorporates stress-dependent stiffness, nonlinear elasticity and plastic hardening, making it suitable for simulating the short-term undrained stress-strain response and peak strength mobilisation observed during consolidated undrained (CU) triaxial testing (Schanz et al., 1999). In contrast, the SSC model extends the Soft Soil formulation by incorporating time-dependent

viscoplastic deformation, enabling the simulation of creep and long-term compression behaviour (Vermeer & Neher, 1999). Although the SSC model was originally developed for consolidation and creep analyses, its response under undrained loading was also evaluated to examine its capability and limitations in reproducing the laboratory triaxial behaviour of tropical peat.

The calibration strategy focused on reproducing the principal features of the laboratory CU triaxial response, including the initial stiffness, peak deviatoric stress mobilisation, overall stress–strain behaviour and post-peak response. Emphasis was placed on identifying the constitutive parameters that exert the greatest influence on the simulated undrained behaviour through systematic sensitivity analyses. Because constitutive parameters are influenced by soil type, stress history, sampling disturbance and depositional environment, the calibrated parameter values presented in this study are considered representative of the investigated Pontian tropical peat and should not be transferred directly to other sites without appropriate verification (Ladd, 1991; Lade, 2005). Nevertheless, the calibration methodology itself provides a systematic and transferable framework for constitutive model selection and parameter calibration using laboratory test data.

Based on this calibration strategy, an axisymmetric finite element model was developed in PLAXIS 2D to reproduce the laboratory CU triaxial testing conditions. The numerical model configuration is described in the following section.

Numerical Model Development

Axisymmetric finite element models created in PLAXIS 2D replicated the cylindrical configuration of triaxial specimens (Figure 4). The dimensions of the specimens, including height and diameter, were aligned with laboratory measurements to maintain consistency in stress paths. The discretisation utilised 15-node triangular elements for quadratic displacement interpolation, ensuring an accurate nonlinear response. A refined mesh was implemented within the specimen to facilitate stable stress redistribution while keeping computational demands manageable.

The boundary conditions were designed to reflect laboratory constraints: the bottom was fixed in all directions, lateral boundaries were constrained radially to uphold confining pressure while permitting vertical deformation, and the axis of symmetry was restrained radially. Axial loading was applied by enforcing a prescribed vertical displacement at the top boundary, which effectively simulated strain-controlled shearing.

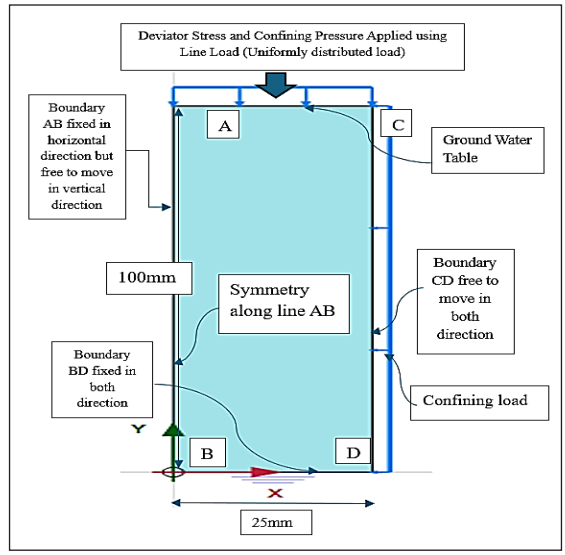


Figure 4. Axisymmetric finite element model adopted for the consolidated undrained (CU) triaxial simulation in PLAXIS 2D, showing specimen geometry, loading configuration and boundary conditions

Drainage boundaries permitted flow during consolidation to achieve the desired effective stress; these boundaries were then closed during shearing to impose undrained conditions, thereby facilitating appropriate pore pressure generation and maintaining volume constancy. Parameters used in the model were sourced from laboratory index and mechanical tests (Table 1) and further refined through calibration. Key parameters of focus included:

- Effective friction angle (φ')
- Cohesion (c')
- Pre-overburden pressure (POP)
- Stress dependency exponent (m)
- Reference secant stiffness modulus (E_{50}^{ref})

Calibration Procedure

The calibration of the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models was carried out using a systematic laboratory-based procedure integrating laboratory testing, constitutive parameter interpretation, numerical simulation, sensitivity analysis and iterative back-analysis.

Unlike automated optimisation techniques, the calibration adopted in this study was performed through sequential manual refinement of constitutive parameters to preserve their physical meaning while achieving satisfactory agreement between laboratory observations and numerical predictions. Figure 5 illustrates the overall calibration workflow.

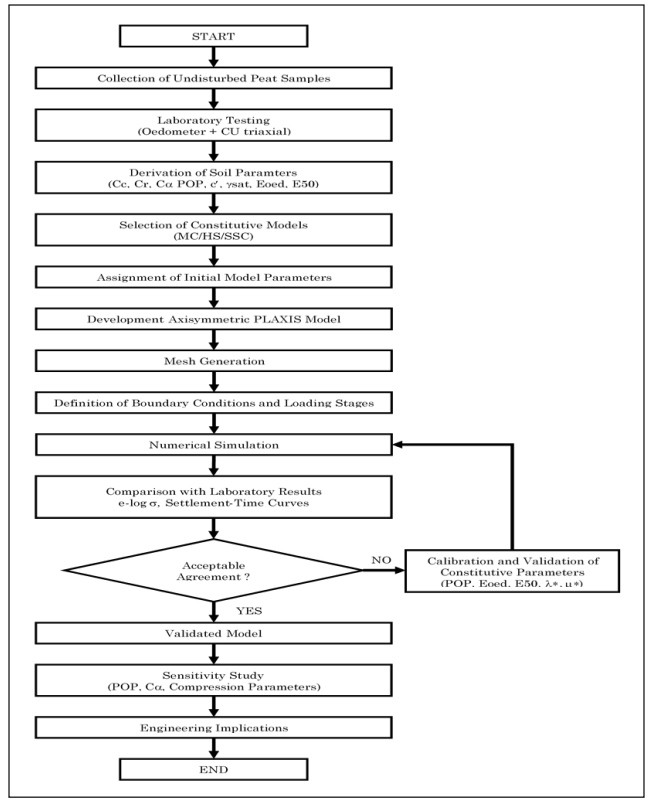


Figure 5. Calibration procedure adopted

The primary objective of the calibration procedure was to establish constitutive parameter sets capable of reproducing the measured laboratory behaviour while maintaining consistency with the physical properties of the tested Pontian tropical peat and associated soft soils. Initial constitutive parameters were derived directly from laboratory testing rather than empirical databases or published parameter ranges. Effective shear strength parameters, namely effective cohesion (c′) and effective friction angle (φ′), were interpreted from the consolidated undrained (CU) triaxial compression tests. The HS stiffness parameters, including the reference secant stiffness modulus (E_{50},ref), reference oedometer stiffness (Eoed,ref) and unloading–reloading stiffness modulus (Eur, ref), were obtained from the triaxial and incremental loading (IL) oedometer tests. For the SSC model, the compression index (Cc), swelling index (Cs), secondary compression index (Cα), initial void ratio (e_0) and pre-overburden pressure (POP) were interpreted from IL oedometer tests conducted in accordance with BS 1377-5:1990 (British Standards Institution, 1990b), while the permeability coefficients (kx and ky) were obtained from laboratory measurements and site investigation data where available.

The laboratory-derived parameter sets were subsequently implemented in PLAXIS 2D using an axisymmetric finite element model that replicated the laboratory specimen geometry, loading conditions and drainage conditions. Numerical simulations were performed for both the CU triaxial and IL oedometer tests, enabling direct comparison between the measured laboratory responses and numerical predictions. For the CU simulations, the calibration focused primarily on reproducing the nonlinear stress–strain response, peak deviatoric stress mobilisation and post-peak behaviour. For the IL oedometer simulations, calibration focused on reproducing the void ratio–log effective stress relationship, compressibility characteristics and stress-history behaviour.

Calibration was performed using a sequential one-at-a-time parameter adjustment strategy. Only those constitutive parameters identified through preliminary sensitivity analyses as having a significant influence on the numerical response were modified during each calibration stage. For the HS model, calibration concentrated principally on the reference stiffness parameters (E_{50}, ref and Eoed, ref) together with the pre-overburden pressure (POP), while maintaining the laboratory-derived effective strength parameters. For the SSC model, refinement focused mainly on the compression index (Cc), swelling index (Cs), secondary compression index (Cα), permeability coefficients (kx and ky) and POP. Parameters exhibiting relatively low sensitivity were retained at their laboratory-derived values to minimise unnecessary adjustment and preserve the physical significance of the constitutive models.

The calibration process was evaluated quantitatively using the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2). Numerical predictions were interpolated to the same strain levels as the laboratory measurements, thereby ensuring direct point-by-point comparison between measured and simulated deviatoric stresses. MAE was used to quantify the average magnitude of prediction error, whereas RMSE provided greater sensitivity to larger deviations between the laboratory and numerical responses. The coefficient of determination (R^2) was used to assess the overall ability of each constitutive model to reproduce the observed stress–strain behaviour.

Calibration convergence was achieved when no physically reasonable one-at-a-time adjustment of the sensitive constitutive parameters produced a further reduction in RMSE without compromising agreement with the measured pre-peak stiffness, peak deviatoric stress mobilisation or the overall stress–strain response. This convergence criterion ensured that the final constitutive parameter sets remained physically realistic while avoiding unnecessary parameter adjustment solely to improve numerical curve fitting. Consequently, the final calibrated parameters represent the optimum balance between numerical accuracy and engineering realism.

The MAE and RMSE were calculated using Equation 1:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|q_{\mathrm{exp},i} - q_{\mathrm{num},i}\right|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(q_{\mathrm{exp},i} - q_{\mathrm{num},i}\right)^2} \quad [1]$$

where $q_{exp,i}$ is the experimentally measured deviatoric stress, $q_{num,i}$ is the corresponding numerical prediction, and n is the total number of measurement points considered during calibration.

The present study is limited to laboratory-scale calibration and does not include independent field validation. Transfer of the calibrated parameters to instrumented embankments or foundations without numerical re-tuning is therefore identified as an essential next step in the Future Research section.

Summary of Model Input Parameters

The constitutive parameters adopted for the Hardening Soil (HS) and Soft Soil Creep (SSC) models were established primarily from laboratory testing and subsequently refined through the calibration procedure described in the preceding Calibration Procedure subsection. The selected parameters represent the engineering characteristics governing the short-term stress-strain behaviour, compressibility, stress-history effects and time-dependent deformation of the Pontian soft soil.

Table 2 summarises the principal constitutive parameters adopted for the numerical simulations together with their corresponding sources. The HS model parameters were primarily derived from the Consolidated Undrained (CU) triaxial compression tests and Incremental Loading (IL) oedometer tests. In contrast, the SSC model utilised laboratory-derived compressibility, creep and stress-history parameters interpreted from the oedometer tests. Where direct laboratory measurements were unavailable, selected parameters were estimated using established geotechnical correlations and subsequently refined through iterative numerical calibration, while maintaining consistency with observed laboratory behaviour.

The parameter ranges presented in Table 2 reflect the values investigated during the calibration process rather than fixed material properties. Final parameter selection was based on achieving satisfactory agreement between the measured and simulated

laboratory responses while preserving the physical significance of each constitutive parameter. Parameters identified through sensitivity analyses as having limited influence on the numerical response were retained at their laboratory-derived values to minimise unnecessary adjustment during calibration.

For the CU triaxial simulations, undrained shearing was modelled under axisymmetric conditions using effective stress analysis. Since drainage was not permitted during shearing, permeability had a negligible influence on the predicted stress-strain response. Consequently, isotropic permeability was assumed, with the horizontal and vertical permeability coefficients set equal (kx = ky). This assumption is considered appropriate for reproducing the short-term undrained behaviour investigated in this study.

Table 2
Summary of soil parameters used in HS and SSC model calibration

Parameter	Symbol	Model	Source	BH1-UD4	BH2-UD5
Saturated unit weight (kN/m³)	γsat	HS / SSC	Index tests	12.8	13.2
Initial void ratio	e_0	SSC	IL oedometer	4.10	3.85
Compression index	Cc	SSC	IL oedometer	2.10-3.45	1.95-3.10
Swelling (recompression) index	Cs	SSC	IL oedometer	0.12-0.25	0.10-0.22
Secondary compression index	$C\alpha$	SSC	IL oedometer	0.015-0.030	0.014-0.028
Pre-overburden pressure	POP	HS / SSC	Casagrande method	15-25 kPa	18-28 kPa
Reference secant stiffness	E50, ref	HS	CU triaxial	2918 kPa	2375 kPa
Reference oedometer stiffness	Eoed, ref	HS	IL oedometer	817 kPa	910 kPa
Reference unloading-reloading stiffness	Eur, ref	HS	Calibration	8750-12500 kPa	7125-11875 kPa
Effective cohesion	c'	HS	CU triaxial	3-5 kPa	2-4 kPa
Effective friction angle	φ'	HS	CU triaxial	10-19°	10-24°
Poisson's ratio	v	HS / SSC	Calibration	0.47-0.55	0.46-0.57
Stress dependency exponent	m	HS	Calibration	0.65-1.00	0.70-0.95
Vertical permeability	ky	SSC	Falling-head test	Same as kx	Same as kx
Horizontal permeability	kx	SSC	Assumed isotropic	Same as ky	Same as ky

Note. It was assumed that permeability is isotropic ($k_x = k_\gamma$) since the consolidated undrained (CU) triaxial simulations were conducted under conditions where drainage does not occur, and the magnitude of the permeability does not affect the anticipated shear stress-strain behaviour. During each simulation, parameter values were chosen from the specified ranges through a process of iterative calibration to accurately reflect the laboratory stress-strain response

RESULTS AND DISCUSSION

HS versus SSC Triaxial Response

The consolidated undrained (CU) triaxial tests demonstrated a characteristic response of tropical peat and associated very soft soils consisting of nonlinear stress-strain behaviour, distinct peak deviatoric stress mobilisation and progressive post-peak strength degradation with increasing axial strain. Numerical simulations using the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models were evaluated against the laboratory results to assess their capability to reproduce these deformation mechanisms under effective confining pressures of 30, 60 and 120 kPa, as shown in Figures 6 and 7.

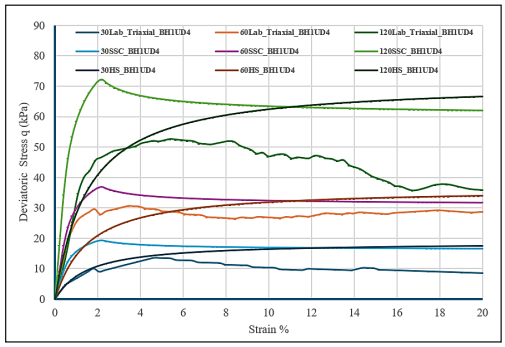


Figure 6. Comparison between laboratory and HS-SSC simulated CU triaxial stress-strain responses for sample BH1-UD4 (20% axial strain)

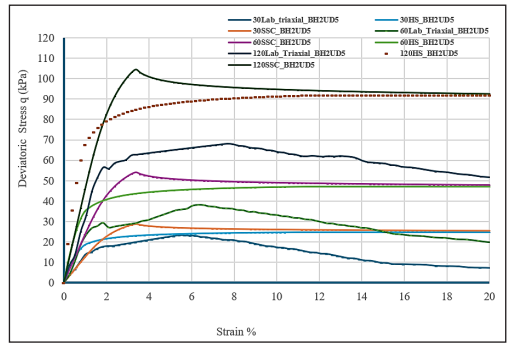


Figure 7. Comparison between laboratory and HS-SSC simulated CU triaxial stress–strain responses for sample BH2-UD5 (20% axial strain)

The HS model reproduced the initial nonlinear stiffness and peak deviatoric stress with good agreement for all confining pressures. The increasing peak strength observed experimentally with increasing confining pressure was also captured satisfactorily by the numerical simulations, indicating that the stress-dependent stiffness and isotropic hardening formulation incorporated within the HS model provides a realistic representation of pre-peak undrained behaviour. These characteristics are consistent with the constitutive framework developed by Schanz et al. (1999), in which stiffness evolves as a function of stress level while plastic strains accumulate progressively before failure. Consequently, the HS model provides reliable prediction of peak strength mobilisation and initial stiffness, both of which are critical for short-term stability analyses of embankments and foundations.

The SSC model exhibited a fundamentally different constitutive response. Rather than reproducing a distinct peak strength, the numerical simulations produced a comparatively smooth and ductile stress-strain relationship characterised by continuous viscoplastic

deformation. Peak deviatoric stresses were consistently lower than those measured in the laboratory, particularly under higher confining pressures, confirming that the SSC model is not formulated to simulate rapid undrained strength mobilisation. Instead, the constitutive formulation emphasises time-dependent compression and creep through viscoplastic volumetric strains, making the model more suitable for long-term settlement analyses than short-term undrained loading.

Although both constitutive models generated reductions in deviatoric stress at larger strains, the underlying mechanisms differed fundamentally. In the HS model, the response stabilised after peak strength because the constitutive formulation contains only hardening mechanisms and no constitutive law describing progressive degradation of soil structure. In contrast, the gradual stress reduction predicted by the SSC model resulted from continued viscoplastic contraction under undrained conditions. As excess pore-water pressure accumulated, the mean effective stress decreased, and the effective stress path progressively approached the failure envelope, producing an apparent reduction in deviatoric stress through effective stress redistribution rather than through degradation of the soil skeleton itself (Vermeer & Neher, 1999). Consequently, superficially similar reductions in deviatoric stress predicted by the two models represent fundamentally different constitutive mechanisms.

The constitutive responses observed in the present study are consistent with previous investigations of peat and advanced creep models. Muraro and Jommi (2021) demonstrated that peat exhibits a strongly nonlinear triaxial response governed by stress path and evolving soil structure, reinforcing the need for laboratory-based constitutive calibration. Bhattarai et al. (2025) showed that the isotropic SSC formulation lacks the anisotropy and destructuration mechanisms incorporated in more advanced creep models. The present findings therefore support the view that HS and SSC represent different deformation mechanisms rather than providing a universal constitutive description of tropical peat.

Overall, the comparative assessment demonstrates that the HS and SSC models should be regarded as complementary constitutive formulations because each captures different deformation mechanisms relevant to tropical peat rather than providing a universal constitutive description. The HS model is more appropriate for analyses requiring realistic prediction of stiffness and peak undrained shear strength, whereas the SSC model is more suitable for problems governed by long-term creep, consolidation and settlement. Constitutive model selection should therefore be based on the dominant engineering mechanism controlling the problem under consideration.

Strain Softening and Residual Strength

The laboratory triaxial tests consistently exhibited progressive post-peak strength degradation following mobilisation of the peak deviatoric stress, ultimately approaching a residual-strength condition. Such behaviour is characteristic of structured tropical peat, in which shear resistance is progressively reduced through degradation of the organic soil fabric during continued undrained shearing. Unlike inorganic mineral soils, tropical peat derives much of its strength from its fibrous structure, organic bonding and the interaction between plant fibres and mineral particles. Progressive destruction of these structural components causes the soil skeleton to lose its ability to sustain peak shear resistance, resulting in constitutive strain softening and eventual mobilisation of residual strength.

Although the HS model accurately reproduced the initial stress-strain response and peak strength mobilisation, it was unable to simulate the subsequent post-peak degradation observed experimentally. This limitation arises from the constitutive formulation rather than from parameter calibration. The HS model incorporates stress-dependent stiffness together with isotropic shear hardening but contains no constitutive variables describing destructuration, bond degradation, fibre breakage, strain localisation or progressive damage evolution. Consequently, once peak strength has been mobilised, the numerical response either stabilises or exhibits only limited stress reduction. Further refinement of constitutive parameters may improve agreement before the peak but cannot generate true strain softening because the governing constitutive equations do not permit progressive reduction of material strength.

The behaviour predicted by the SSC model should likewise be interpreted carefully. Although gradual reductions in deviatoric stress occurred at larger strains, these do not represent constitutive strain softening. The SSC formulation was developed to describe time-dependent compression through viscoplastic volumetric deformation. Under undrained loading, continued viscoplastic contraction generates excess pore-water pressure because drainage is prevented. The resulting reduction in mean effective stress causes the effective stress path to migrate progressively towards the failure envelope, producing an apparent decrease in deviatoric stress. Consequently, the observed response originates from pore-pressure evolution and effective stress redistribution rather than degradation of the soil skeleton or progressive loss of constitutive strength.

The present observations agree with recent peat testing and creep-model comparisons. Muraro and Jommi (2021) showed that peat response is nonlinear and strongly dependent on the stress path and evolving structure. Bhattarai et al. (2025) further demonstrated that

advanced creep formulations incorporate anisotropy and destructuration mechanisms that are absent from the isotropic SSC model. The present results extend these observations to Pontian tropical peat-associated soft soils: the inability of HS to reproduce strain softening and the apparent post-peak response of SSC arise from their constitutive formulations rather than from parameter calibration alone.

These findings demonstrate that the limitations identified in the present study cannot be eliminated simply through improved calibration. Laboratory-derived parameters may reduce prediction errors associated with stiffness and peak strength, but they cannot compensate for constitutive mechanisms that are absent from the mathematical formulation. Consequently, numerical analyses involving progressive failure, strain localisation or residual-strength mobilisation should consider more advanced constitutive models capable of representing anisotropy, destructuration, bond degradation and damage evolution. The laboratory-based calibration framework developed in this study, therefore, provides a practical means of distinguishing between parameter-calibration uncertainty and constitutive-model inadequacy, enabling more informed selection of constitutive models for geotechnical analyses involving tropical peat.

Quantitative Evaluation of Constitutive Model Calibration Performance

To complement the qualitative comparison between the laboratory and numerical oedometer responses, a quantitative evaluation of the Hardening Soil (HS) model calibration was performed using the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These statistical indicators provide an objective assessment of the agreement between the measured and simulated one-dimensional compression behaviour. The MAE represents the average absolute deviation between the laboratory measurements and numerical predictions, whereas the RMSE assigns greater weight to larger discrepancies and is therefore more sensitive to local deviations in the compression curve. Lower values of both indices indicate closer agreement and a more reliable constitutive calibration.

Table 3 summarises the laboratory-derived pre-overburden pressure (POP), calibrated best-fit POP, percentage differences, and the corresponding MAE and RMSE values obtained for each oedometer specimen. The results demonstrate that the HS model successfully reproduced the measured compression behaviour for most of the tested specimens. Nine of the ten specimens achieved either Excellent or Very Good calibration performance, with an overall average MAE of 0.133 and RMSE of 0.188, corresponding to an overall Very Good classification. These results indicate that the laboratory-derived constitutive parameters required only limited refinement to achieve satisfactory agreement with the experimental compression curves.

Table 3
Quantitative performance assessment of hardening soil (HS) model calibration based on incremental loading (IL) oedometer tests

Sample	POPLab (kPa)	POPBest Fit (kPa)	POP Difference (%)	MAE	RMSE	Performance
BH1-UD1	27	23	-14.8	0.130	0.260	Very Good
BH1-UD2	19	5	-81.0	0.440	0.670	Good
BH1-UD3	12	6	-50.0	0.230	0.280	Very Good
BH1-UD4	9	5	-44.4	0.080	0.090	Excellent
BH1-UD5	12	7	-41.7	0.210	0.240	Very Good
BH1-UD6	22	27	+18.5	0.013	0.017	Excellent
BH1-UD7	20	25	+20.0	0.041	0.053	Excellent
BH1-UD8	14	17	+17.6	0.069	0.087	Excellent
BH1-UD9	24	20	-17.0	0.060	0.082	Excellent
BH2-UD4	15	7	-53.0	0.070	0.100	Excellent
Average	-	-	−24.6	0.133	0.188	Very Good

The highest calibration accuracy was obtained for specimens BH1-UD6, BH1-UD7, BH1-UD8, BH1-UD9, and BH2-UD4, all of which recorded MAE values below 0.07 and RMSE values below 0.10. Specimen BH1-UD6 exhibited the smallest prediction errors, with an MAE of 0.013 and an RMSE of 0.017, indicating excellent agreement between the laboratory measurements and the numerical simulation. These results demonstrate that the proposed calibration procedure is capable of accurately reproducing the one-dimensional compression characteristics of tropical peat and very soft clay.

In contrast, specimen BH1-UD2 produced the largest prediction errors, with an MAE of 0.440 and an RMSE of 0.670, together with an 81.0% reduction in the calibrated POP relative to the laboratory-derived value. Although this specimen exhibited the poorest agreement among the dataset, the numerical model nevertheless reproduced the overall shape of the compression curve within an acceptable engineering tolerance. The larger adjustment required for this specimen is attributed to uncertainties associated with stress-history interpretation and the inherent variability of tropical peat, where the determination of yield stress from oedometer tests is often influenced by sample disturbance, natural heterogeneity and the subjective interpretation of the Casagrande construction.

Comparison between the laboratory-derived and calibrated POP values further demonstrates that calibration adjustments were not uniform among the investigated specimens. Several specimens required reductions in POP, whereas others required modest increases to reproduce the measured compression response. This behaviour confirms that POP should not be regarded as a fixed laboratory value, but rather as a

constitutive parameter that may require limited calibration within physically realistic bounds to accurately represent stress-history effects in numerical analyses. The relatively small deviations observed for most specimens indicate that the laboratory-derived POP values provide a sound initial basis for constitutive parameter selection.

Overall, the quantitative evaluation confirms that the proposed laboratory-based calibration methodology provides a reliable framework for determining HS model parameters from oedometer testing. The consistently low MAE and RMSE values obtained for most specimens demonstrate that the calibrated model successfully reproduces the principal features of one-dimensional compression behaviour while preserving the physical significance of the laboratory-derived parameters. Consequently, the calibrated parameter sets provide a robust foundation for subsequent finite element analyses and field-scale simulations involving tropical peat and very soft clay deposits.

Parameter Sensitivity Analysis

A methodological parameter sensitivity of the governing parameters of undrained shear response controlling Pontian peat in consolidated undrained (CU) triaxial models was determined through a systematic parameter sensitivity analysis. This part of the investigation analyses both the Soft Soil Creep (SSC) and Hardening Soil (HS) models, comparing them to their laboratory stress-strain patterns.

Influence of Effective Friction Angle (φ′)

Figures 8, 9 and 10 are SSC-simulated stress-strain responses for sample BH1-UD4 at the confining pressures of 30, 60 and 120 kPa, respectively, with different effective friction angles (φ′). Likewise, the results of sample BH2-UD5 are shown in Figures 11, 12 and 13.

We derive that for both samples and for all confining pressures, variation of φ′ causes a systematic, vertical shift of the deviatoric stress-strain curves that shows that φ′ is the primary stress-determinant parameter in SSC. Nevertheless, no distinct peak mobilisation and post-peak strain-softening behaviour is observed, even with increasing maximum stress. By contrast, the SSC model offers a smooth, ductile response that is nearly viscoplastic equilibrium stress-like.

This remark confirms that while φ′ has a decisive influence on the stress level in SSC simulations, it cannot be said to shape the basic shape of stress and strain which is under consideration in the formulation of viscoplasticity in the model. Therefore, the change of φ′ alone does not allow SSC to recover the prominent peak and post-peak degradations studied in laboratory CU triaxial studies.

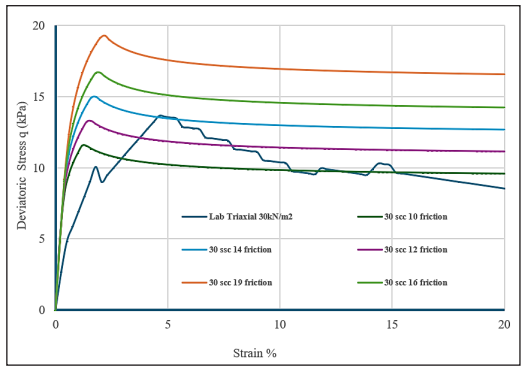


Figure 8. SSC-simulated CU triaxial deviatoric stress-strain response for sample BH1-UD4 with varying effective friction angle (φ′) under 30 kPa confining pressure (20% axial strain)

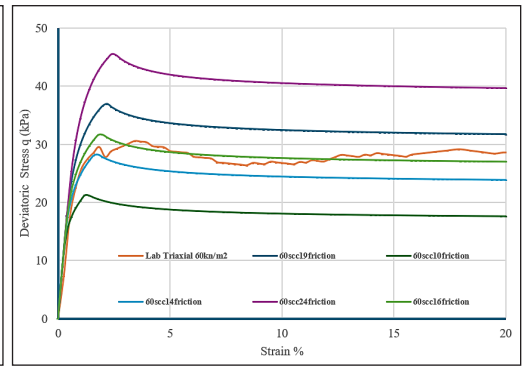


Figure 9. SSC-simulated CU triaxial deviatoric stress-strain response for sample BH1-UD4 with varying effective friction angle (φ′) under 60 kPa confining pressure (20% axial strain)

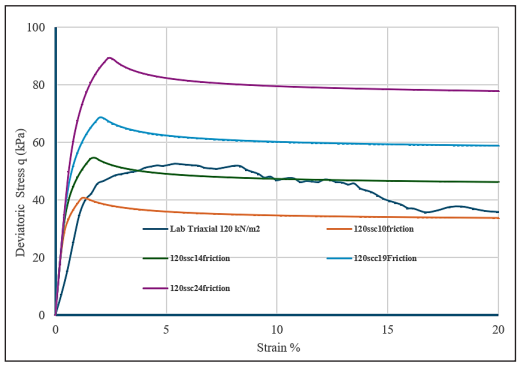


Figure 10. SSC-simulated CU triaxial deviatoric stress-strain response for sample BH1-UD4 with varying effective friction angle (φ′) under 120 kPa confining pressure (20% axial strain)

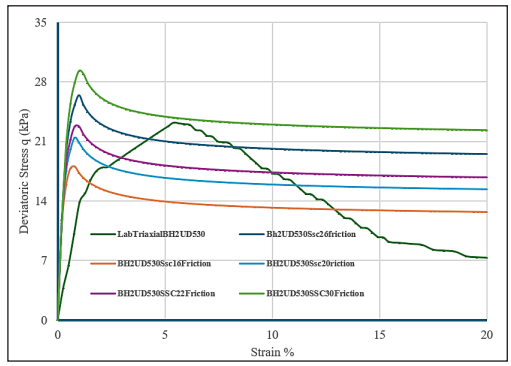


Figure 11. SSC-simulated CU triaxial deviatoric stress-strain response for sample BH2-UD5 with varying effective friction angle (φ′) under 30 kPa confining pressure (20% axial strain)

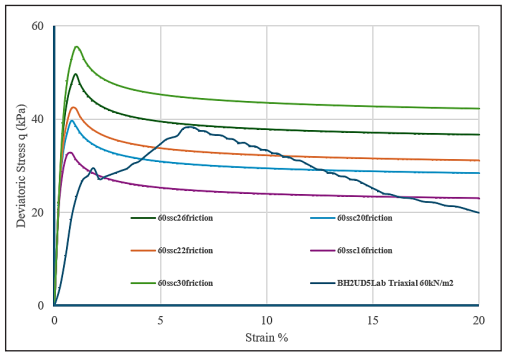


Figure 12. SSC-simulated CU triaxial deviatoric stress–strain response for sample BH2-UD5 with varying effective friction angle (φ′) under 60 kPa confining pressure (20% axial strain)

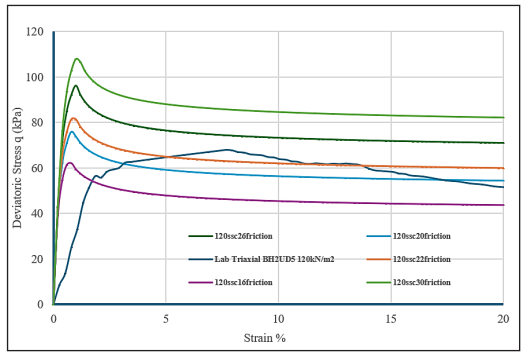


Figure 13. SSC-simulated CU triaxial deviatoric stress-strain response for sample BH2-UD5 with varying effective friction angle (φ′) under 120 kPa confining pressure (20% axial strain)

Influence of Reference Secant Stiffness ($E_{50}ref$)

Figure 14 and Figure 15 represent the influence of the reference secant stiffness modulus $E_{50}ref$ (up to 20% and 5% axial strain in the HS model, respectively). $E_{50}ref$ has a great impact on initial stiffness and strain mobilisation (most notably at small to medium axial strains).

Higher $E_{50}ref$ values lead to steeper initial stress–strain slopes; lower $E_{50}ref$ values induce more compliant responses. Nevertheless, the peak deviatoric stress magnitude remains constant. This further proves that $E_{50}ref$ controls stiffness, rather than strength mobilisation (as expected under the HS model).

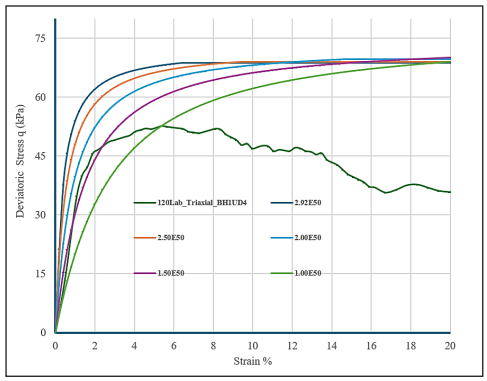


Figure 14. Effect of reference secant stiffness modulus (E_{50}, ref) on HS-simulated CU triaxial deviatoric stress-strain response for sample BH1-UD4 (20% axial strain)

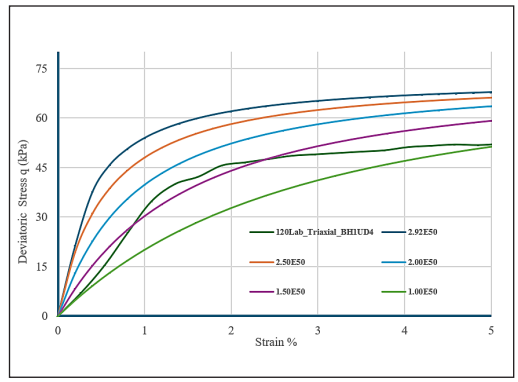


Figure 15. Effect of reference secant stiffness modulus (E_{50}, ref) on HS-simulated CU triaxial deviatoric stress-strain response for sample BH1-UD4 (5% axial strain)

Parameters with Limited Influence Under CU Triaxial Conditions

The effects of the stress dependency exponent (m) upon the HS-measured undrained shear under the applied consolidated undrained triaxial load are presented in Figure 16. Sensitivity analyses show that select parameters have a limited influence on the predicted undrained shear response in the HS model during consolidated undrained triaxial conditions. The effects of differences in the pre-overburden pressure POP mainly result in variability in volumetric yielding and stress history under one-dimensional

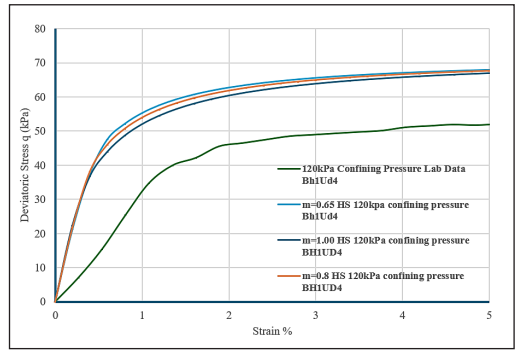


Figure 16. Effect of stress dependency exponent (m) on HS-simulated CU triaxial response for sample BH1-UD4 under 120 kPa confining pressure

compression, but do not directly have an impact on deviatoric stress mobilisation after isotropic consolidation. As a result, the variations in POP have a negligible effect on the deviatoric stress–strain response when CU triaxial loading occurs following isotropic consolidation.

Likewise, changes in the stress dependency exponent m in a realistic range as shown in Figure 16 also do not have a significant effect on the expected undrained shear response. The deviatoric stress-strain curves are approximately coincident, suggesting that, at varying confining pressures, m mainly controls stiffness evolution, rather than peak undrained shear strength at fixed confinement.

Implications for Model Calibration

The sensitivity analysis indicates that effective friction angle (φ') is the most significant parameter controlling peak undrained shear resistance, while E_{50}^{ref} controls stiffness and strain mobilisation. Parameters such as POP and m are indispensable in consolidation and stress-history-dependent problems yet have a limited influence on undrained triaxial shear response.

The results support targeting calibration effort towards φ' and stiffness parameters in CU triaxial behaviour modelling, preventing unnecessary modification of parameters that do not strongly affect predicted stress-strain response. This enhances the efficiency of calibration and minimises the chances of compensating errors in numerical modelling. These results highlight that parameter sensitivity is highly influenced by stress path and loading mode, with conclusions resulting from CU triaxial calibration not being directly applicable to consolidation-dominated field problems.

Study Limitations and Sources of Uncertainty

Although the proposed laboratory-based calibration framework demonstrated good agreement between the measured and simulated responses, several sources of uncertainty should be recognised when interpreting the results. These uncertainties arise from the natural variability of tropical peat, laboratory testing procedures, constitutive parameter interpretation and simplifying assumptions adopted during numerical modelling. Acknowledging these limitations is important because constitutive model calibration does not produce a unique solution, and the predicted behaviour remains dependent on both the quality of the experimental data and the assumptions inherent in the constitutive formulations (Brinkgreve et al., 2023; Potts & Zdravković, 1999).

The first source of uncertainty is sampling disturbance. Tropical peat possesses a highly compressible and heterogeneous fibrous structure that is particularly susceptible to disturbance during drilling, sampling, transportation and specimen preparation.

Even when undisturbed sampling techniques are adopted, some alteration of the natural soil fabric is unavoidable. Such disturbance may reduce the apparent pre-overburden pressure (POP), modify stiffness characteristics and influence the measured undrained shear response obtained from laboratory testing. Consequently, the laboratory-derived constitutive parameters should be regarded as representative of the tested specimens rather than exact in-situ soil properties (Ladd & DeGroot, 2003). Recent studies continue to emphasise that sample disturbance remains one of the principal sources of uncertainty in the characterisation and design of soft soils.

A second source of uncertainty is laboratory variability. Tropical peat exhibits considerable spatial variability in fibre content, degree of decomposition, organic matter content, moisture content and density, even within relatively small areas. As a result, specimens obtained from adjacent boreholes or different sampling depths may exhibit significantly different compressibility, stiffness and shear-strength characteristics. Such variability contributes to the differences observed between the calibrated parameters of individual specimens and partly explains why certain specimens, such as BH1-UD2, required greater refinement during the calibration process. Therefore, the calibrated parameter values presented in this study should be interpreted as representative of the investigated Pontian peat rather than universally applicable values for all tropical peat deposits (Huat et al., 2014; Mesri & Ajlouni, 2007).

Another important source of uncertainty is parameter non-uniqueness. The Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models contain several interacting parameters that influence similar aspects of the predicted response. For example, the stiffness response predicted by the HS model is simultaneously affected by the reference secant stiffness (E_{50},ref), reference oedometer stiffness (Eoed, ref), stress dependency exponent (m) and pre-overburden pressure (POP). Likewise, long-term compression predicted by the SSC model depends on the combined influence of the compression index (Cc), swelling index (Cs), secondary compression index ($C\alpha$) and permeability coefficients. Consequently, different combinations of constitutive parameters may produce similarly acceptable numerical responses. To minimise this non-uniqueness, the present study adopted a laboratory-based calibration strategy in which most constitutive parameters were derived directly from laboratory testing, while only the most influential parameters identified through sensitivity analyses were refined during iterative calibration. This approach reduces subjective parameter adjustment and preserves the physical significance of the laboratory-derived parameters (Brinkgreve et al., 2023; Potts & Zdravković, 1999).

Finally, the numerical analyses incorporate several modelling assumptions that introduce additional uncertainty. The laboratory triaxial specimens were represented using an idealised axisymmetric finite element model with homogeneous material properties, isotropic permeability and ideal boundary conditions. Furthermore, the standard HS

and SSC constitutive models implemented in PLAXIS do not explicitly incorporate anisotropy, destructuration, bond degradation, damage evolution or strain localisation, all of which may influence the mechanical behaviour of structured tropical peat. These simplifying assumptions are consistent with current engineering practice but inevitably limit the ability of the numerical models to reproduce every aspect of the observed laboratory behaviour. Consequently, the calibrated constitutive parameters should be interpreted within the scope of the adopted constitutive formulations and modelling assumptions rather than as unique material constants applicable to all loading conditions (Brinkgreve et al., 2023; Schanz et al., 1999; Wood, 1990).

Despite these limitations, the proposed calibration framework provides a systematic and transparent methodology for deriving constitutive parameters directly from laboratory testing while maintaining their physical significance. The integration of laboratory interpretation, sensitivity analysis, quantitative calibration using MAE and RMSE, and independent validation through numerical simulation improves confidence in the calibrated parameter sets and provides a robust basis for future finite element analyses involving tropical peat and other highly compressible soft soils. Nevertheless, future studies should investigate probabilistic calibration techniques, parameter uncertainty analysis and advanced constitutive models incorporating destructuration and anisotropy to further improve the predictive capability of numerical simulations.

Engineering Implications

The findings of this study demonstrate that constitutive model selection for tropical peat should be governed by the dominant engineering mechanism rather than by attempting to identify a single constitutive model capable of reproducing all aspects of soil behaviour. The laboratory results show that tropical peat exhibits highly nonlinear stress-strain behaviour characterised by distinct peak strength mobilisation, progressive strain softening and eventual residual-strength development. Since these deformation mechanisms arise from different constitutive processes, no single constitutive model available within the standard PLAXIS constitutive library can simultaneously reproduce the complete laboratory response.

For short-term geotechnical problems involving undrained loading, such as staged embankment construction, temporary excavations and bearing-capacity assessments, accurate prediction of peak undrained shear strength is essential because failure generally occurs before significant creep deformation develops. The Hardening Soil (HS) model reproduced the measured peak deviatoric stress and nonlinear stiffness response with satisfactory accuracy, demonstrating that its stress-dependent stiffness and isotropic hardening formulation is appropriate for analysing short-term stability. Nevertheless, engineers should recognise that the HS model cannot reproduce progressive post-peak

strain softening or residual-strength mobilisation because its constitutive formulation does not include mechanisms describing destructuration, degradation of organic bonding, fibre breakage or damage evolution. Consequently, numerical analyses using the HS model may overestimate the available shear resistance after peak strength has been mobilised, particularly in problems involving progressive failure or large deformation.

Conversely, the Soft Soil Creep (SSC) model is more appropriate for engineering applications in which long-term deformation governs performance, including post-construction settlement, creep deformation and consolidation of embankments founded on tropical peat. The present study demonstrates that the SSC model reproduces the gradual evolution of stress associated with viscoplastic contraction and effective stress redistribution. However, the apparent reduction in deviatoric stress predicted during undrained loading should not be interpreted as true constitutive strain softening because it results primarily from excess pore-water pressure generation rather than progressive degradation of the soil skeleton. Accordingly, the SSC model should not be relied upon to evaluate short-term undrained stability or residual-strength mobilisation without independent verification against laboratory shear-strength data.

The sensitivity analyses further demonstrate that reliable numerical predictions depend on systematic laboratory-based calibration of constitutive parameters. For the HS model, the effective friction angle (φ'), reference secant stiffness modulus (E_{50}, ref) and pre-overburden pressure (POP) exert the greatest influence on peak strength mobilisation and stress–strain response. In contrast, the SSC model is governed primarily by the compression index (Cc), swelling index (Cs), secondary compression index ($C\alpha$) and stress-history parameters controlling long-term compressibility and creep deformation. These findings emphasise that constitutive parameters should be derived from high-quality laboratory testing and subsequently refined through controlled sensitivity analyses rather than adopted directly from empirical correlations or generic parameter databases.

The laboratory-based calibration framework developed in this study provides a transparent and reproducible methodology for constitutive parameter selection. By integrating laboratory testing, constitutive parameter interpretation, numerical back-analysis, sensitivity analysis and quantitative evaluation using MAE and RMSE, the proposed procedure enables engineers to distinguish between parameter-calibration error and limitations arising from the constitutive formulation itself. This distinction is important because improvements in parameter calibration cannot compensate for constitutive mechanisms that are absent from the mathematical model.

The present findings are consistent with recent studies of peat behaviour and creep-model selection. Muraro and Jommi (2021) demonstrated the importance of stress path and evolving structure in peat triaxial response, while Bhattarai et al. (2025) showed that SSC lacks the anisotropy and destructuration mechanisms included in more advanced

creep formulations. These comparisons reinforce the need to align the selected constitutive model with the engineering objective rather than expecting one formulation to reproduce all aspects of tropical peat behaviour.

From an engineering perspective, the practical implication is that constitutive model selection should form part of the overall geotechnical design strategy. The HS model is recommended for evaluating short-term stability, peak undrained shear strength and construction-stage performance, whereas the SSC model is better suited for predicting long-term settlement, creep and consolidation. Where progressive failure, strain localisation or residual-strength mobilisation is expected to control engineering performance, the analysis should be supplemented by residual-strength evaluation and, where appropriate, more advanced constitutive formulations capable of representing anisotropy, destructuration or damage evolution. Such an approach provides a more realistic assessment of the behaviour of embankments, foundations and other infrastructure constructed on tropical peat and very soft clay deposits.

Constitutive model selection should therefore form part of the overall geotechnical design strategy rather than being treated solely as a numerical modelling decision. For short-term undrained stability, the HS model provides reliable prediction of stiffness and peak strength mobilisation. For long-term settlement and creep, the SSC model is more appropriate. Where progressive failure or residual-strength mobilisation governs performance, additional laboratory testing together with constitutive formulations capable of representing destructuration and damage evolution should be considered. This approach enables more reliable design of embankments, foundations and other infrastructure constructed on tropical peat and very soft clay.

Future Research

The present study established a laboratory-based constitutive calibration framework for evaluating the performance of the Hardening Soil (HS) and Soft Soil Creep (SSC) models in reproducing the undrained behaviour of Pontian tropical peat and associated soft soils. Nevertheless, several opportunities remain for extending this work.

First, future investigations should incorporate a larger number of undisturbed specimens collected from different depths and locations within tropical peat deposits. The current study is based on two representative specimens obtained from a single site in Pontian, Johor. Although these specimens adequately represent the investigated geological profile, a more extensive testing programme would enable statistical evaluation of spatial variability and improve the transferability of constitutive parameters to other tropical peat deposits.

Second, additional laboratory measurements should be undertaken to improve constitutive parameter determination. Direct measurement of fibre content, degree of

decomposition, microstructural characteristics and controlled axial strain rates would provide a more comprehensive description of the mechanisms governing strain softening and compressibility. Such information could reduce uncertainty associated with constitutive parameter selection and improve the reproducibility of laboratory-based calibration.

Third, future research should investigate more advanced constitutive formulations capable of representing progressive degradation of structured organic soils. Although the HS model reproduced the nonlinear pre-peak stress-strain response with satisfactory accuracy, it could not simulate post-peak strain softening because its constitutive formulation does not include destructuration or damage evolution. Similarly, while the SSC model successfully represented creep-related behaviour, it was unable to reproduce the complete undrained stress–strain response. Recent reviews document continuing advances in time-dependent soil formulations (Liu et al., 2024). Constitutive models incorporating anisotropy, destructuration, bond degradation or damage evolution should therefore be evaluated to determine whether they provide a more realistic representation of tropical peat behaviour. Such models may include advanced research constitutive formulations that extend beyond the standard constitutive library available in PLAXIS.

Fourth, the laboratory-based calibration framework developed in this study should be validated using independent field case histories. Future studies should compare laboratory-derived constitutive parameters with monitored field performance from instrumented embankments and foundations without further parameter re-tuning. Such validation would provide a more rigorous assessment of the transferability of laboratory-calibrated parameters under realistic engineering conditions and strengthen confidence in their application to practical geotechnical design.

Finally, future numerical investigations should consider fully coupled hydro-mechanical analyses to evaluate the interaction between consolidation, excess pore-water pressure dissipation and long-term creep deformation during staged embankment construction. Recent studies show that coupled deformation-pore-pressure analysis and creep-aware consolidation formulations can improve the interpretation of soft-soil settlement and parameter uncertainty (Li et al., 2024; Huang et al., 2025). Integration of probabilistic calibration techniques with coupled numerical analyses would further enable quantification of parameter uncertainty and improve the reliability of constitutive model predictions for infrastructure founded on tropical peat.

Overall, these research directions would contribute to the development of more robust constitutive modelling strategies for tropical peat by integrating detailed laboratory characterisation, advanced constitutive formulations, field-scale validation and uncertainty quantification within a unified numerical modelling framework.

CONCLUSION

This study investigated the undrained shear behaviour of Pontian tropical peat through a combination of laboratory consolidated undrained (CU) triaxial testing and advanced finite element modelling using the Hardening Soil (HS) and Soft Soil Creep (SSC) constitutive models implemented in PLAXIS 2D. A laboratory-based constitutive calibration framework was developed to evaluate the capability of both models to reproduce the principal mechanisms governing the undrained response of tropical peat, including peak strength mobilisation, post-peak behaviour and parameter sensitivity.

The laboratory CU triaxial tests demonstrated that Pontian peat exhibits distinct peak deviatoric stress mobilisation at relatively small axial strains, followed by progressive post-peak strength degradation towards a residual-strength condition. This behaviour reflects the progressive degradation of the fibrous organic soil structure and provides the experimental benchmark for constitutive model calibration and evaluation.

The numerical simulations demonstrated that the Hardening Soil (HS) model reproduced the peak deviatoric stress and overall stress-strain response with good agreement over the range of confining pressures investigated. The HS model is therefore appropriate for evaluating short-term undrained stability and peak shear resistance. However, because its constitutive formulation is based on stress-dependent stiffness and isotropic hardening plasticity, it cannot reproduce the progressive strain-softening behaviour associated with destructuration and degradation of the soil microstructure.

In contrast, the Soft Soil Creep (SSC) model generated a smooth ductile stress–strain response without distinct peak strength mobilisation. The apparent reduction in deviatoric stress predicted by the SSC model under undrained loading was shown to result from creep-induced excess pore-water pressure generation and effective stress path evolution rather than true strain softening. Consequently, the SSC model is better suited to predicting long-term creep and settlement behaviour than peak undrained shear resistance or residual-strength mobilisation.

The sensitivity analyses identified the effective friction angle (φ'), reference secant stiffness (E_{50}, ref) and pre-overburden pressure (POP) as the most influential parameters governing the undrained response of the HS model, whereas the compression index (Cc), swelling index (Cs) and secondary compression index ($C\alpha$) exert the greatest influence on the long-term deformation behaviour predicted by the SSC model. These findings demonstrate that reliable constitutive modelling depends on careful laboratory-based parameter interpretation combined with systematic sensitivity analysis rather than extensive numerical parameter adjustment.

The principal contribution of this study is the development of a transparent laboratory-based constitutive calibration framework that integrates laboratory testing, sensitivity analysis and quantitative model evaluation using MAE and RMSE. This framework

preserves the physical significance of laboratory-derived parameters while providing an objective methodology for calibrating advanced constitutive models for tropical peat.

Overall, the study demonstrates that no single constitutive model available within the standard PLAXIS constitutive library can simultaneously reproduce peak strength mobilisation, progressive post-peak strain softening and long-term creep behaviour of tropical peat. Consequently, constitutive model selection should always reflect the dominant engineering mechanism governing the problem under investigation. For short-term undrained stability analyses, the HS model is recommended, whereas the SSC model is more appropriate for long-term settlement and creep analyses. The proposed calibration framework provides practical guidance for finite element modelling of embankments, foundations and other infrastructure constructed on tropical peat and very soft clay deposits.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support provided by the Ministry of Higher Education Malaysia under the Fundamental Research Grants Scheme (FRGS), Grant No. FRGS/1/2019/TK01/UNIMAS/02/2. The authors thank Universiti Malaysia Sarawak (UNIMAS) for providing the research premises and working space. Special thanks go to Sejati Teguh Engineering Sdn. Bhd. for providing the field and laboratory test data for this study. The authors also point out the invaluable involvement and commitment of all members of the research team, as well as the team members' contributions and dedication to research work during the preparation and development in the end.

REFERENCES

Bhattarai, T. R., Bhandary, N. P., & Grimstad, G. (2025). Comparative analysis and performance evaluation of SSC, n-SAC, and Creep-SCLAY1S soil creep models in predicting soil settlement. Geotechnics, 5(3), Article 47. https://doi.org/10.3390/geotechnics5030047

Brinkgreve, R. B. J., Kumarswamy, S., Swolfs, W. M., Foria, F., Waterman, D., & Chesaru, A. (2023). PLAXIS 2D 2024 reference manual. Bentley Systems.

British Standards Institution. (1990a). BS 1377-2:1990: Methods of test for soils for civil engineering purposes-Part 2: Classification tests.

British Standards Institution. (1990b). BS 1377-5:1990: Methods of test for soils for civil engineering purposes-Part 5: Compressibility, permeability and durability tests.

British Standards Institution. (1990c). BS 1377-8:1990: Methods of test for soils for civil engineering purposes-Part 8: Shear strength tests (effective stress).

Burland, J. B. (1990). On the compressibility and shear strength of natural clays. Géotechnique, 40(3), 329-378. https://doi.org/10.1680/geot.1990.40.3.329

Duncan, J. M., & Chang, C.-Y. (1970). Nonlinear analysis of stress and strain in soils. Journal of the Soil Mechanics and Foundations Division, 96(5), 1629-1653. https://doi.org/10.1061/JSFEAQ.0001458

Huang, S., Huang, J., Kelly, R., Jiang, S.-H., Jones, M., & Kamruzzaman, A. H. M. (2025). Coupling effects on class C predictions of soft soil settlements. Acta Geotechnica, 20, 723-742. https://doi.org/10.1007/s11440-024-02422-9

Huat, B. B. K., Prasad, A., Asadi, A., & Kazemian, S. (2014). Geotechnics of organic soils and peat. CRC Press.

Kulhawy, F. H., & Mayne, P. W. (1990). Manual on estimating soil properties for foundation design (Report EL-6800). Electric Power Research Institute.

Ladd, C. C. (1991). Stability evaluation during staged construction. Journal of Geotechnical Engineering, 117(4), 540-615. https://doi.org/10.1061/(ASCE)0733-9410(1991)117:4(540)

Ladd, C. C., & DeGroot, D. J. (2003). Recommended practice for soft ground site characterisation. In T. S. Tan, K. K. Phoon, D. W. Hight, & S. Leroueil (Eds.), Characterisation and engineering properties of natural soils (Vol. 1, pp. 3-57). Swets & Zeitlinger.

Lade, P. V. (2005). Triaxial testing of soils. John Wiley & Sons.

Leroueil, S., & Vaughan, P. R. (1990). The general and congruent effects of structure in natural soils and weak rocks. Géotechnique, 40(3), 467-488. https://doi.org/10.1680/geot.1990.40.3.467

Leroueil, S., Tavenas, F., & Le Bihan, J.-P. (1983). Propriétés caractéristiques des argiles de l'est du Canada [Characteristic properties of eastern Canadian clays]. Canadian Geotechnical Journal, 20(4), 681-705. https://doi.org/10.1139/t83-078

Li, P.-L., Yin, Z.-Y., Song, D.-B., Yin, J.-H., & Pan, Y.-T. (2024). Axisymmetric finite strain consolidation model for soft soil consolidation with vertical drains under combined loading considering creep and non-Darcy flow. Geotextiles and Geomembranes, 52(3), 241-259. https://doi.org/10.1016/j.geotexmem.2023.10.008

Liu, Y., Huang, M., Cao, W., Jin, Y., Wang, J., & Hou, Y. (2024). Review of soil creep characteristics and advances in modelling research. Buildings, 14(6), Article 1668. https://doi.org/10.3390/buildings14061668

Mesri, G., & Ajlouni, M. (2007). Engineering properties of fibrous peats. Journal of Geotechnical and Geoenvironmental Engineering, 133(7), 850-866. https://doi.org/10.1061/(ASCE)1090-0241(2007)133:7(850)

Muraro, S., & Jommi, C. (2021). Pre-failure behaviour of reconstituted peats in triaxial compression. Acta Geotechnica, 16, 789-805. https://doi.org/10.1007/s11440-020-01019-2

Peng, B., Feng, R., Wu, L., Wang, P., & Shi, X. (2024). One-dimensional creep consolidation model for peat soil. Applied Sciences, 14(17), Article 7990. https://doi.org/10.3390/app14177990

Potts, D. M., & Zdravković, L. (1999). Finite element analysis in geotechnical engineering: Theory. Thomas Telford.

Schanz, T., Vermeer, P. A., & Bonnier, P. G. (1999). The hardening soil model: Formulation and verification. In R. B. J. Brinkgreve (Ed.), Beyond 2000 in computational geotechnics (pp. 281-296). A. A. Balkema.

Sejati Teguh Engineering Sdn. Bhd. (2023). Laporan penyiasatan tapak bagi endapan gambut Pontian, Johor [Site investigation report for Pontian peat deposits, Johor] [Unpublished consultancy report].

Tavenas, F., & Leroueil, S. (1985). The behaviour of embankments on clay foundations. Canadian Geotechnical Journal, 22(4), 541-558. https://doi.org/10.1139/t85-077

Tavenas, F., Leroueil, S., & Roy, M. (1983). The behaviour of sensitive clays during undrained shearing. Canadian Geotechnical Journal, 20(3), 645-660. https://doi.org/10.1139/t83-071

Vermeer, P. A., & Neher, H. P. (1999). A soft soil model that accounts for creep. In R. B. J. Brinkgreve (Ed.), Beyond 2000 in computational geotechnics (pp. 249-261). A. A. Balkema.

Wood, D. M. (1990). Soil behaviour and critical state soil mechanics. Cambridge University Press.

Pertanika J. Sci. & Technol. 34 (4): 2213 - 2231 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Assessing the Drought Index and Water Scarcity Nexus: Case Study for Langat River Basin

Raja Muhamad Naufal Raja Azman[1], Nur Asmaliza Mohd Noor[2]*, Marini Mohamad Ideris[3], and Sharifah Abdullah[1]

[1]Department of Civil Engineering, Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

[2]Department of Civil Engineering, Faculty of Civil Engineering, Universiti Teknologi MARA, Pahang Branch, Jengka Campus, 26400 Bandar Tun Abdul Razak Jengka, Pahang, Malaysia

[3]National Water Research Institute of Malaysia (NAHRIM), Ministry of Energy Transition and Water Transformation, Lot 5377, Jalan Putra Permai, 43300 Seri Kembangan, Selangor, Malaysia

ABSTRACT

Climate change has resulted in the increase of surface temperatures, which has significantly affected water availability, disrupting economic development and quality of life. Thus, the application of drought index and water scarcity for analysis is essential, as they accurately estimate the impact of water availability. This study quantifies drought and water scarcity using SPEI and WEI+, respectively, and investigates their interrelationship in the Langat River Basin. The results revealed a strong positive correlation between the drought and water scarcity index, with correlation coefficients of 0.9677, 0.9545, and 0.9633 for SPEI-3, SPEI-6, and SPEI-12, respectively. The frequency analysis of the drought index showed 54 occurrences for SPEI-3, 58 for SPEI-6, and 61 for SPEI-12. Meanwhile, the water scarcity index was quantified for the period from 2015 to 2018, yielding values of 2.60×10^{-3} in 2015, 3.81×10^{-3} in 2016, 3.39×10^{-3} in 2017, and 3.68×10^{-3} in 2018. Overall, these findings provide valuable insights and form a robust basis for developing effective mitigation strategies and informed decision-making in water supply planning and river basin management, thereby reducing potential adverse impacts on water resources.

Keywords: Climate change, drought index, water resources, water scarcity index, water supply

ARTICLE INFO

Article history:
Received: 29 June 2026
Accepted: 23 July 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.12

E-mail addresses:
rajamuhdnaufal@gmail.com (Raja Muhammad Naufal)
nurasmaliza@uitm.edu.my (Nur Asmaliza)
marini@nahrim.gov.my (Marini)
sharifah.abdullah@uitm.edu.my (Sharifah Abdullah)
* Corresponding author

INTRODUCTION

Urbanisation is a dynamic phenomenon of economic and social capability shifting from rural areas with an agriculture-based economy to urban areas with industry and

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

services sectors. In fact, urbanisation with high urban densities is also capable of structuring the economic patterns of resource consumption and worldwide environmental quality (Shahbaz et al., 2015). However, climate change is one of the effects of uncontrolled urbanisation and leads to an increase in surface temperature, which contributes to the changes in the atmosphere and has a significant effect on the availability of water. The possibility of climate change for an area is dependent on the historical situation or the prediction of potential climates such as drought, floods, and monsoons (Othman et al., 2018). Drought is a stochastic and persistent natural hazard that has costly and destructive effects on the availability of surface and groundwater, crop production, quality of ecological water, production of electricity (hydropower), modern industrial production, and waterborne transportation (Tri et al., 2019). The effect of climate change on the water resources of Karaj River Catchment was investigated and showed that the decreasing on precipitation in the coming decades resulting the increasing of river discharge (Azizi et al.,2017). Many scientists have made several attempts to examine drought, which consists of spatial and temporal drought variations (Lana et al., 2001), the drought impact reduction (Huang et al., 2008), drought frequency analysis (Mondal et al., 2015), and drought prediction based on different indices of atmospheric circulation (Cordery, 2000). Two aspects that had been considered for drought analysis were classification of drought severity and estimation of drought vulnerability (Natarajan et al., 2022).

Standardised Precipitation Evapotranspiration Index (SPEI), Standardised Precipitation Index (SPI), and Effective Drought Index (EDI). Standard Index of Annual Precipitation (SIAP), Z-Score Index (ZI) and Standardised Water Storage Index (SWSI) are several available drought indices. However, among them, the SPEI is most preferable due to its integration of precipitation and temperature effects, multi-timescale flexibility, and suitability for assessing drought under climate change conditions (Isia et al., 2023). The other indices, such as SPI and EDI, are used to evaluate the severity, period and extent of drought occurrence (Huang et al., 2016). Meanwhile, the SIAP used for meteorological drought, and SWSI used for hydrological drought which effective in predicting hydrological and meteorological drought very close to the observed values (Khan et al., 2018). The Z score Index can be calculated by subtracting the long-term mean from an individual rainfall; however, it cannot represent the shorter time scales as well as the SPI (Edwards & McKee, 1997). The main reason SPEI has been chosen because it can detect drought better than other indices because the data used is a long base period (30 -50 years besides able to monitor and explore the impact of global warming, which another index cannot do.

Water scarcity plays an important role in evaluating the continuity of water supply as well. Water scarcity can be defined as the water shortage in a water supply system that can be caused by drought or by human actions, such as increasing water usage over renewable natural water supplies, which normally happen in a greater population and a lack of natural water supplies (Fan et al., 2018). Water Poverty Index (WPI), Water Exploitation Index

(WEI), Water Exploitation Index Plus (WEI+) and Water Stress Index (WSI) are among the available water scarcity indices. WPI is developed for water scarcity assessments at the scale of local communities and the ability to adapt this index for different scales (Shalamzari et al., 2018). Meanwhile, the WEI+ represents the level of pressure that human activity puts on a specific area's natural water supplies (Fan et al., 2018), and the WEI is obtained through the percentage of mean annual total demand of freshwater with respect to the long-term mean annual freshwater resources (Pedro-Monzonís et al., 2015). The WSI calculates the availability of total water resources in a region based on the actual ratio of water withdrawal to hydrological availability (Sondermann et al., 2022). WEI+ is preferred over WPI, WEI, and WSI because it provides a physically based, basin-scale assessment of water scarcity that accounts for consumptive use, return flows, and environmental requirements, while remaining compatible with climate-driven drought analyses.

Malaysia is blessed with abundant rainfall, averaging around 2,000mm annually. However, due to recent climate variations, the country has experienced and is expected to face more drought events. A series of droughts occurred in Malaysia in 1997, 1998, and 2014, with the severe one in 2014 affecting around 1.8 million residents in southern Kuala Lumpur. During these years, Malaysia faced significant water scarcity due to reduced rainfall, which led to drought conditions (Shaaban et al., 2003). There were a few studies related to the drought assessment that had been conducted in Malaysia, such as SPI and NDVI methods in determining the risk vegetation area that was impacted during the drought event in Kedah (Othman et al., 2016), determination of the drought event in Langat River Basin using SIAP and SWSI (Khan et al., 2018), and the potential of evapotranspiration (PET) based on temperature variables in determining the trend analysis with respect to the drought episode (Hui-Mean et al., 2018). As for the water scarcity index, the existing studies have primarily focused on the WSI, and the conducted research is assessing water scarcity in Malaysia, especially in rice production, which indicates Malaysia are heavily dependent on precipitation to meet the demands of rice production (Hanafiah et al., 2019). Another study on the reserved forest in Kelantan, which assesses the WSI as a role in producing, maintaining, and reserving water for natural ecosystems and human usage (Samsudin et al., 2020). These studies highlight the critical role of WSI in understanding and managing water resources in Malaysia, especially in the face of changing precipitation patterns and increasing water demand. Despite extensive research on drought assessment and water scarcity in Malaysia, existing studies have largely addressed these issues independently, with drought characterised using precipitation- or temperature-based indices and water scarcity primarily assessed through the Water Stress Index.

Besides, limited research has investigated the SPEI and WEI+ quantitative relationship within humid tropical river basins. Existing studies have largely focused on arid and temperate climates, where hydro-climatic responses differ substantially from those of

tropical monsoon systems. Consequently, the extent to which drought circulates into water scarcity under tropical climatic conditions remains poorly understood. Hence, the exploration of drought and water scarcity indices under tropical climate are very crucial as the indices can represent and portray the value of drought and water scarcity indices based on accurate experience and able to provide accurate conditions locally, which relates to the drought assessment and the drought legislation is more effective as the drought management can be planned earlier. (Hui-Mean et al., 2018; Iglesias et al., 2007). The neglecting of this issue will affect society's agriculture in the long term, as it is an important environmental issue that needs to be addressed (Payus et al., 2020). Thus, this study addresses the gap by integrating SPEI and WEI+ using an Artificial Neural Network (ANN) in order to quantify the drought–water scarcity nexus in the Langat River Basin, Malaysia and this will be able to provide a proper planning for the water management stakeholder especially in drought management.

METHODOLOGY

LRB has been chosen as the location of study due to the major transformation of land use due to rapid urbanisation. The development areas around the basin extended to 23.5% in 2013 compared to 2.4% in 1974 (Farid Ahmed et al., 2023). Figure 1 shows the map of LRB, which ranges around 1815 $km^{2,}$ and the river's major course is 141km.

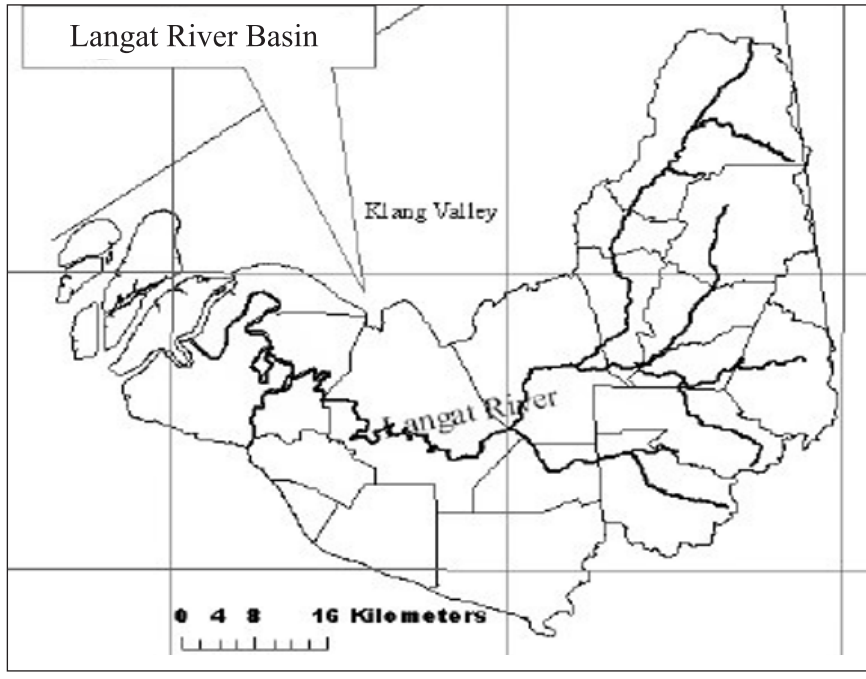


Figure 1. Map of Langat River Basin (LRB) (DID, 2017)

The location of LRB lies between latitude N 2° 40' 0'' and N 3° 20' 0'' and between longitude E 101° 20' 0'' and E 102° 0' 0''. The LRB have a population of 1,184,917 in the year 2000 with a rate of growth of 7.64%, then grew to 4,266,011 in 2013 as reported by the Department of Statistics Malaysia (Elfithri et al., 2018). The Langat catchment provides water supply to approximately 1.2 million people from 14 towns include dense towns such as Cheras, Kajang, Bangi and the Government Centre of Putrajaya. There are two (2) reservoirs located under LRB, which are Semenyih and Hulu Langat Dam and eight (8) water treatment plants located within the catchment. The land use trend had a major transformation in LRB because of rapid development.

SPEI and WEI+ Assessment

The SPEI monthly time scales are being considered, with the calculation according to the water deficit or surplus between the precipitation and evapotranspiration, based on various time scales. In this study, 3-month, 6-month, and 12-months timescales were used in the analysis, and the analysis was conducted using R Studio software. The 3-month and 12-months value of SPEI reflected short-term and long-term moisture conditions and provided a seasonal estimation of precipitation deficiency (Quang et al., 2021). The computation equation is depicted as in Equation 1, and the classification for SPEI can be categorised as in Table 1. The data for rainfall, temperature and evaporation for SPEI calculation were obtained from the Drainage Irrigation Department Malaysia and Met Malaysia in accordance with the selection station within the catchment from 1998-2018, as displayed in Table 2.

Table 1
Drought classification based on the (SPEI) values

Drought Classification	SPEI Values
Extremely wet	2.00 or more
Severely wet	1.50 to 1.99
Moderately wet	1.00 to 1.49
Near normal	-0.99 to 0.99
Moderately drought	-1.00 to -1.49
Severe drought	-1.50 to -1.99
Extreme drought	-2.00 or less

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \quad [1]$$

where

$W = [-2\ln(P)]^{0.5}$ for $P \leq 0.5; P = 1 - F(x);\ C_0 = 2.515517,$
$C_1 = 0.802853, C_2 = 0.010328;$
$d_1 = 1.432788, d_2 = 0.189269$ and $d_3 = 0.001308$

Table 2
List and location of the rainfall station

No	Station	Latitude	Longitude
1	Tadika Kemas Kg. Tg. Sepat	2° 40' 53''	101° 33' 49''
2	P/A Sg. Pelek Sepang	2° 37' 58.9''	101° 37' 26.2''
3	P/A Kg. Tali Air Morib	2° 44' 39''	101° 27' 02''
4	Kg. Salak Tinggi	2° 47' 19.2''	101° 44' 31.1''
5	P/A Kelanang	2° 48' 44''	101° 25' 6''
6	Pejabat JPS Sg. Manggis	2° 49' 35''	101° 32' 30''
7	Ldg. Brooklands	2° 49' 40''	101° 33' 15''
8	P/A Pekan Banting	2° 48' 38''	101° 30' 30''
9	P/A Sg. Sedu	2° 51' 21.1''	101° 30' 53''
10	Dengkil	2° 51' 20.8''	101° 40' 53.1''
11	KLIA	2° 48' 17.5''	101° 41' 30.5''
12	RTB Bukit Changgang	2° 48' 41''	101° 39' 15''
13	Bukit Changgang	2° 48' 48.5''	101° 38' 30.9''
14	Paya Indah	2° 52' 44''	101° 37' 09''
15	Kg Jenderam Hilir	2° 51' 41''	101° 44' 7''
16	Sek. Men. Bandar Tasik Kesuma	2° 53' 55.47''	101° 52' 13.3''
17	P/A Bt. 9 Sijangkang	2° 57' 23''	101° 26' 03''
18	P/A Bt 7 Sijangkang	2° 57' 23''	101° 26' 03''
19	Ldg. Bkt. Cheeding	2° 54' 40''	101° 34' 35''
20	Puncak Niaga Putrajaya	2° 54' 40''	101° 41' 50''
21	RTM Kajang	2° 59' 46''	101° 47' 8.9''
22	Kolam Takungan Sg. Merab	2° 56' 43''	101° 44' 60''
23	Stor JPS Hulu Langat	2° 56' 10''	101° 45' 49''
24	Kajang	2° 59' 33.7''	101° 47' 7.8''
25	Sg. Rinching	2° 54' 54.7''	101° 49' 18.9''
26	Sg. Raya Bt.9 Hulu Langat	3° 04' 04''	101° 46' 19''
27	Sg. Serai Bt.12 Hulu Langat	3° 05' 58''	101° 47' 51''
28	Empangan Semenyih	3° 04' 43''	101° 52' 50''
29	Kg. Pasir	3° 00' 48.4''	101° 51' 57.8''
30	Ldg. Dominion	3° 00' 13''	101° 52' 55''
31	TNB Pansun	3° 12' 34.7''	101° 52' 33.1''

Note. Adapted from (Department of Irrigation and Drainage, 2017)

Meanwhile, WEI+ can be obtained through the total amount of water supplies, water usage, inflow, outflow and reservoir impoundment. The index focuses on the net consumption assessment at the monthly level as shown in Equation 2 (Sondermann et al., 2022).

$$\text{WEI+} = \frac{(AB - RE)}{RWR} \quad [2]$$

where

AB = the total amount of water abstracted from the river basin, including surface water resources and groundwater resources,

RE = the amount of water returning to the hydrological cycle after being used by human activities,

RWR = renewable water resources, and is measured by using natural streamflow as shown in Equation 3.

$$RWR = Outflow + (abstraction - returns) - \Delta Sart \quad [3]$$

where,

ΔSart = change in storage of artificial reservoirs.

Data for WEI+ calculation, such as water abstraction and sewerage treatment plant, were obtained from Selangor Water Management Authority (LUAS), and Indah Water Konsortium (IWK) Selangor, as depicted in Table 3. However, there is a limitation in terms of data for water abstraction and sewage treatment plant for the Langat River Basin, as the availability of these two data is only available from 2015-2018.

Table 3
Summary of data collection for the Langat River basin

List of Data	Type of Data	Duration
Evaporation	Monthly	1988 – 2018
Water abstraction	Yearly	2015 – 2018
Sewage treatment plant	Yearly	2015 – 2018

To quantify the relationship between SPEI and WEI+, an Artificial Neural Network (ANN) was used to calculate the correlation coefficient, and the Pearson method was chosen to identify the relationship. The value nearly +1.0 shows a perfectly positive and good relationship.

RESULTS AND DISCUSSION

Drought Assessment for LRB

The drought years were identified based on the computed value to plot the SPEI trend. Table 4 shows the summary of major drought and wet events identified using SPEI in the Langat

Table 4
Summary of major drought and wet events identified using SPEI in the Langat River

Year	SPEI frequency	Drought Classification	Duration (months)	Minimum SPEI	Severity (Magnitude)
1995	SPEI-3	Near Normal	12	0.15	7.33
1998	SPEI-6	Moderate Drought	5	-1.48	-4.71
1998	SPEI-3	Severe Drought	6	-1.61	-6.80
1999	SPEI-12	Moderate Drought	12	-1.50	-14.78
1999	SPEI-12	Severe Drought	12	-1.50	-14.78
2012	SPEI-12	Near Normal	12	0.77	10.72
2013	SPEI-6	Moderately Wet	12	-0.19	7.03
2016	SPEI-3	Moderately Wet	12	-0.64	4.38
2017	SPEI-6	Near Normal	12	0.12	6.87

Table 5
Frequency of drought based on SPEI values for LRB

SPEI Values	Drought Classification	Frequency		
		SPEI-3	SPEI-6	SPEI-12
2.00 or more	Extremely wet	0	0	0
1.50 to 1.99	Severely wet	0	0	0
1.00 to 1.49	Moderately wet	11	8	3
-0.99 to 0.99	Near normal	304	301	304
-1.00 to -1.49	Moderately drought	52	58	60
-1.50 to -1.99	Severe drought	2	0	1
-2.00 or less	Extreme drought	0	0	0

River. It presents the monthly SPEI values corresponding to the representative wet, normal, moderate drought, and severe drought events that were identified in the Langat River Basin. Representative years were selected to illustrate the temporal evolution of drought classification at different frequencies (SPEI-3, SPEI-6, and SPEI-12). The magnitude represents the cumulative SPEI values over the corresponding event period, where larger negative values indicate greater drought severity, while positive values indicate wetter-than-normal conditions. Meanwhile, Table 5 depicts the frequency of drought based on SPEI values for the Langat River basin. According to Table 4 and Figures 2 to 4, the moderate wet occurs in 2016 for SPEI-3, 2013 for SPEI-6 and SPEI-12. It covered 10 months in 2016 for SPEI-3 with a magnitude of 4.38, and July was characterised as a moderately wet month, with the value of SPEI being 1.32. For SPEI-6, the year was moderately wet in 2013, with the magnitude value of 7.03, and the value of SPEI is 1.34. In that year, 10 months of moderate wet event occur, and February was the most moderate wet. For SPEI-12, the most moderate event occurs in February in the year 2013, with the SPEI value of 1.13, and it covered 12 months of moderate wet events with the value of magnitude of 10.14.

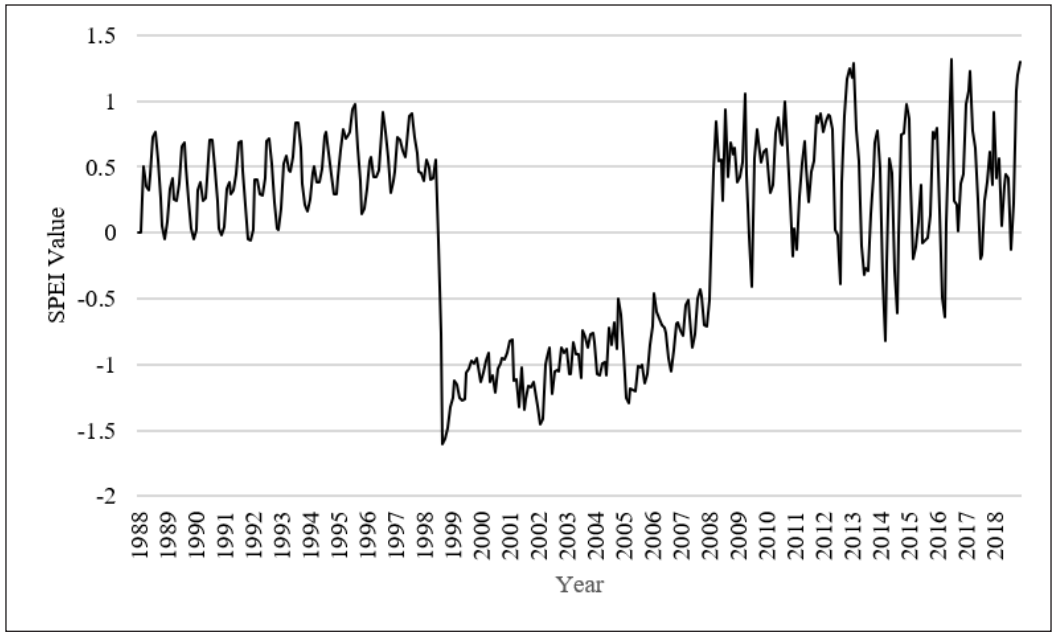


Figure 2. Temporal trends of SPEI 3-months over years

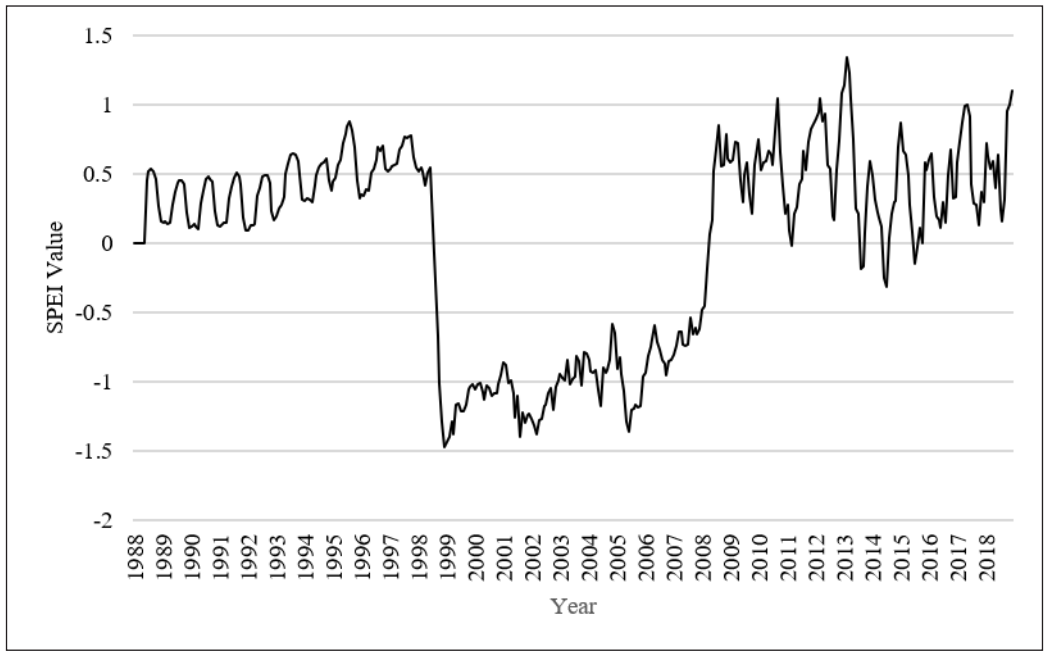


Figure 3. Temporal trends of SPEI 6-months over years

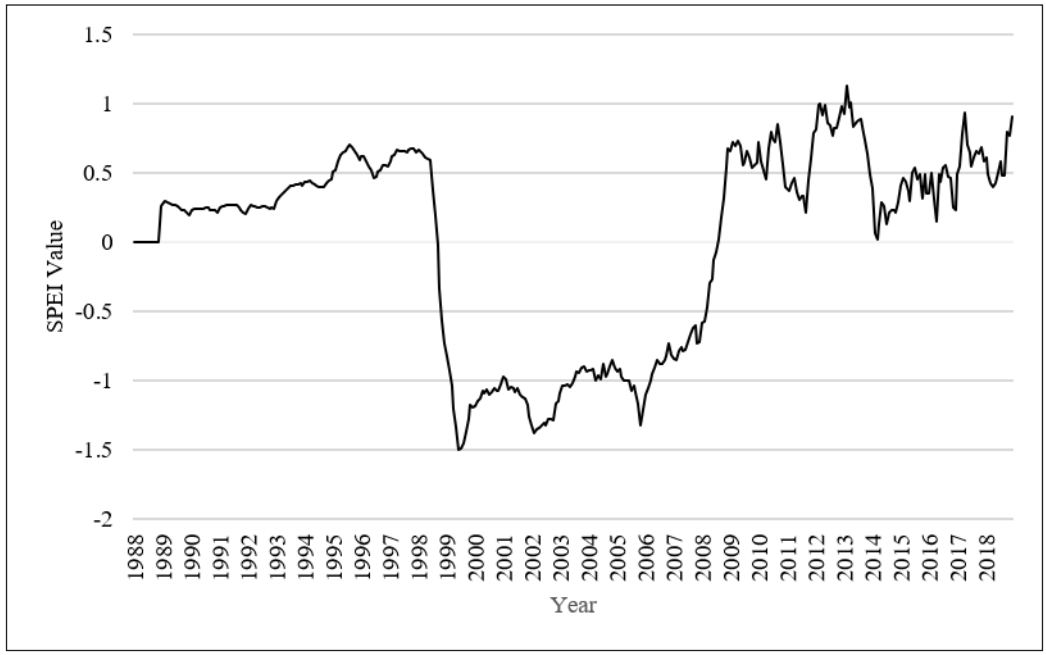


Figure 4. Temporal trends of SPEI 12-months over the years

For near normal occurrences in 1995, 2017, and 2012 for SPEI-3, SPEI-6, and SPEI-12, respectively. In the year 1995, the most normal occurrence in August had a SPEI value of 0.98, and it covered 12 months of near-normal events with a magnitude value of 7.33. For SPEI-6, the near-normal event occurred in 2017, with the most near-normal month being April, and the value of SPEI is 0.99. Meanwhile, the value of magnitude for that year is 6.87, which is covered by 12 months event. For SPEI-12, the year occur near normal event is 2012, which is 12 months event, and the most near normal month is February and May, with the SPEI value of 0.99 for both months, while the magnitude value for that year is 10.72. The moderate drought event is being identified from 1998 to 2006 for SPEI-3, SPEI-6, and SPEI-12. Figure 2 shows that the highest value for moderate drought is in 1998 for SPEI-3. It covered 6 months of drought events, and the highest value of severity is -1.49, which is in November, with the value of magnitude -6.80. Meanwhile, for SPEI-6, the year of the highest value for moderate drought is the same as SPEI-3, which is in 1998, and it covered 5 months of drought events. The most severe month is in December, with the value of severity is -1.48 and the magnitude value is -4.71. For SPEI-12, the highest value of moderate drought occurs in the year 1999, and it covers 12 months. The most severe month is July, with a severity value of severity is -1.49 and a magnitude value of -14.78. For severe drought, the year identified is 1998 only for SPEI-3 and SPEI-12. There are 6 months of drought for SPEI-3. The most severe occurs in September, with the value of

severity being -1.61, and the magnitude value is -6.80. For SPEI-12, there are 12 months of drought, and the most severe occurs in June, with the value of severity is -1.50 and the magnitude value is -14.78. Like Fung et al. (2020), the conducted study demonstrates that SPEI effectively characterises drought variability in tropical climate conditions. Moreover, the drought characteristics observed in the Langat River Basin are consistent with those reported for the Muda River Basin, where drought variability was governed primarily by rainfall deficits despite the humid tropical climate (Luhaim et al., 2021). The event exacerbated by climate variability and increasing water demand for urbanization and understanding the dynamics of drought patterns (Dai et al., 2020; Lam et al., 2023). Singh et al. (2022) reported that the number of drought months under the categories of severe and moderate could be higher for 3 and 6 months compared to 12 months' time step. Besides, the location of the station and the lower annual precipitation rate are among the factors that influenced the magnitude of drought (Fung et al., 2020; Wu et al., 2009).

There are many factors that contributed to the effect of drought pattern trend, such as rainfall amount, pattern of rainfall and land use characteristics. However, a certain factor caused by climate change modifies the form, intensity and timing of water demand, precipitation, and runoff. The availability of water becomes less predictable due to the contribution of the increased period of drought and higher temperatures. The irregularity of rainfall patterns can be observed over the last two decades, which affected the annual precipitation and daily temperature, thus prolonging the incidences of dry periods for a few years in Malaysia (Yusof et al., 2013). Besides, the deterioration of water quality and groundwater also contributed to the degradation of the water supply. Instead, the increasing water demand due to the higher growth of population and economic development, changes in the rainfall pattern due to climate change has led to prolonged duration and magnitude of the drought events. According to Tan et al. greater increase in the warm index was found in the southern part of peninsular Malaysia, indirectly due to climate change, rapid urbanization and industrialization which have induced more warm days in the last decade (Tan et al., 2021).

Land use changes are basically involved with human use, for examples terrestrial space for economic, residential, recreational, conservation and government purposes, such as deforestation, has significantly increased the temperature. Furthermore, the changes in land use patterns and climate change are interrelated, where the changes in land use patterns change the concentration of greenhouse gases, thus contributing to climate change (Li et al., 2022). The urban area, which already has the changes of the land use pattern significantly impacting the drought propagation, can be observed clearly in LRB as changing rapidly compared from 1974 to 2013 (Farid Ahmed et al., 2023) with the population of 1,184,917 in 2000 and increasing to 4,266,011 in 2013 based on the report by the Department of Statistics Malaysia (Elfithri et al., 2018).

Zhou et al. (2021) have conducted a study to investigate the relationship between drought indices and concluded that forest and cover types were significantly affected since climate change influences the growth of vegetation due to the conversion of forest land for economic development reasons.

WEI+ Assessment for LRB

The calculated values of the WEI+ for LRB from 2015 to 2018, according to the availability of data, are displayed in Table 6 and show the increasing trend of WEI+ yearly. These reflected the increasing of water consumption from 2015 until 2018 due to rapid urbanisation development, increasing of population and the usage of water for household purposes, as the location of LRB is located at urbanised area. Besides, the renewable freshwater resources (RWR) also increased from 2015 until 2018, and the accumulation of RWR depended on several factors such as the development of water catchment and the increasing of precipitation volume during that year. The change of WEI+ values is related to the parameters such as abstractions, returns, and outflows. Thus, the average net water consumption, renewable water resources, and WEI+ from 2015 to 2018 are 690 million m^3, 9.744×10^{10} m^3, and 3.37×10^{-3} respectively.

Table 6
The calculated WEI+ values from 2015 to 2018

Year	Water Consumption (m^3) ($\times 10^8$)	RWR (m^3) ($\times 10^{10}$)	WEI+ ($\times 10^{-3}$)
2015	6.80	8.98	2.60
2016	6.88	8.84	3.81
2017	6.83	9.97	3.39
2018	7.12	11.14	3.68
Average	6.90	9.73	3.37

Figure 5 and Figure 6 show that the water consumption and RWR value are increasing over the year due to increasing on the population, expansion in urbanisation, industrialisation, as well as irrigation (Ahmed et al., 2014). This condition aligned with the LRB, which is located at urbanized and industrial area, thus contributing to the increasing of water consumption. Meanwhile, the increasing of RWR value depends on several factors such as the development of the water catchment and the increase in precipitation volume during that year. The similar factors occurred at LRB as well, thus contributing to the increasing RWR values at LRB. According to the previous research, the WEI+ value varies between 0 and 1, and if the WEI+ value is below 20%, it can be defined as a non-stressed period, exceeds 40% is a severe water stress period, and above 60% is a highly stressed period (Sondermann et al., 2022). Therefore, according to the results obtained, it can be concluded that the values of WEI+ are below 20%, which is recognised as a non-stressed period.

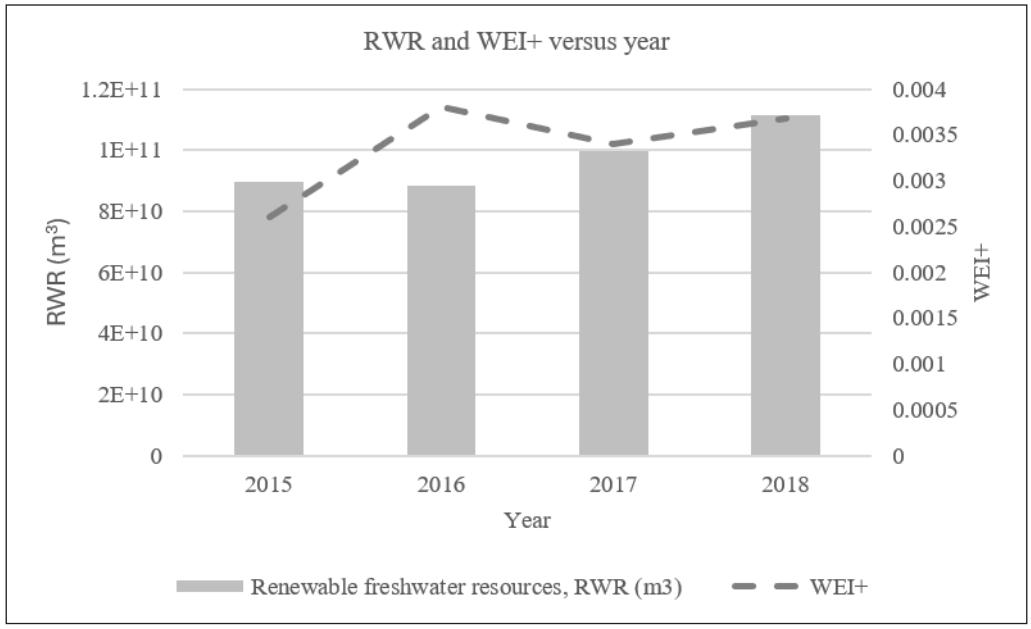


Figure 5. Renewable freshwater resources and WEI+ versus the year

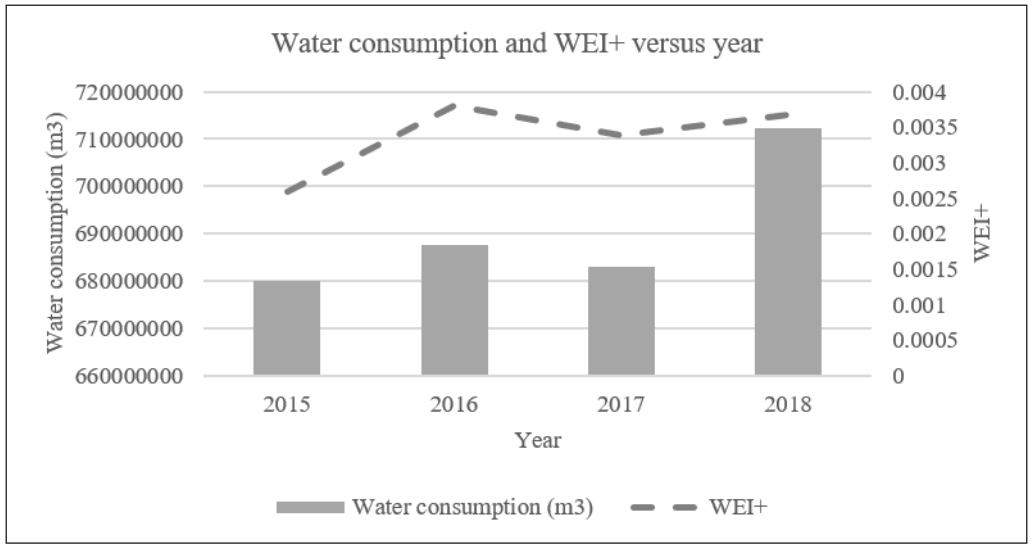


Figure 6. Water consumption and WEI+ versus the year

Relationship between SPEI and WEI+

Table 7 depicts the relationship between SPEI-3, SPEI-6, SPEI-12 and WEI+, respectively, using the artificial neural network (ANN) to identify the value of the correlation coefficient, R. The results show the correlation coefficient, R, is 0.9677 for SPEI-3 versus WEI+,

0.9545 for SPEI-6 versus WEI+, and 0.9633 for SPEI-12 versus WEI+, thus indicating the strong positive correlation relationship between SPEI-3, SPEI-6, and SPEI-12 against the WEI+. A similar behaviour has been observed between SPEI-3 and SPEI-12 due to the hydrological characteristics of the Langat River Basin. Although SPEI-3 represents short-term drought and SPEI-12 reflects long-term drought, the basin is characterised by a humid tropical monsoon climate with high dependence on seasonal rainfall for river discharge and reservoir storage. Consequently, prolonged rainfall deficits can simultaneously influence short-term and long-term drought conditions, resulting in similar temporal responses between SPEI-3 and SPEI-12. Furthermore, water demand in the Langat River Basin remains relatively stable throughout the year due to continuous domestic, industrial, and agricultural consumption. Therefore, fluctuations in water availability are primarily governed by climatic variability rather than abrupt changes in water abstraction, contributing to the consistently strong relationship between SPEI and WEI+ across different time scales.

Table 7
Relationship between SPEI and WEI+ using ANN

No	Relationship	Correlation Coefficient Value, R
1	SPEI-3 and WEI+	0.9677
2	SPEI-6 and WEI+	0.9545
3	SPEI-12 and WEI+	0.9633

The obtained finding aligned with the previous study by Fan et al. (2018), where the dependency between SPEI and WEI+ was 0.658, and indicated a positive dependence between these two variables. The same vein found in Guan et al. (2025) showed a strong correlation between water resources and SPEI, which emphasises the importance of water management strategies. According to the mechanism of drought and water scarcity, meteorological drought may happen when natural precipitation is not enough for a certain period and when water supply cannot meet the needs of water consumption for society, then water scarcity may happen (Fan et al., 2018).

CONCLUSION

This study demonstrates that drought conditions have significantly influenced water availability in the LRB. The application of the SPEI and the WEI+ was proven effective in characterising drought conditions and water scarcity, respectively. The strong positive correlations identified between SPEI and WEI+ across multiple time scales, as revealed through the artificial neural network (ANN) approach, indicate a clear and quantifiable linkage between climatic drought and basin-scale water stress. The increasing frequency of drought events at short-, medium-, and long-term time scales, together with the observed upward trend in WEI+ values between 2015 and 2018, highlights the growing vulnerability of water resources in the basin.

These findings underscore the importance of integrating climate-sensitive drought indices with physically based water scarcity indicators to support informed water resources

planning and river basin management. The results provide empirical evidence that drought conditions, when combined with increasing water exploitation, can substantially aggravate water scarcity, posing risks to socio-economic development and environmental sustainability.

Even though this study provides valuable insights into the relationship between drought and water scarcity in the Langat River Basin, several limitations should be acknowledged. First, the analysis relied on historical data datasets and data available up to 2018, which may not fully capture future climatic variability under changing climate conditions. Second, the assessment employed SPEI and WEI+ and this index has been widely accepted, but they did not explicitly account for reservoir operation and land-use change, which also may influence the water availability.

Based on the outcomes of this study, it is recommended that water management authorities incorporate integrated drought–water scarcity assessment frameworks, such as the combined use of SPEI and WEI+, into routine monitoring and early warning systems. The adoption of data-driven tools, including machine learning techniques, can further enhance predictive capability and decision-making under climate uncertainty. Additionally, demand-side management strategies, protection of environmental flows, and adaptive water allocation policies should be prioritised to mitigate the impacts of prolonged droughts. Besides, the continuous monitoring will further improve the reliability of drought prediction and support adaptive river basin management.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme entitled Assessment of Water Scarcity-Related Drought Index for Water Security in Malaysia with the grant number 600-RMI/FRGS 5/3 (265/2019).

LIST OF ABBREVIATIONS

ANN : Artificial neural network
EDI : Effective drought
LRB : Langat River Basin
PET : Potential of evapotranspiration
SPEI : Standardised precipitation evapotranspiration index
SPI : Standardised precipitation index
SIAP : Standard index of annual precipitation
SWSI : Standardised water storage index
WPI : Water Poverty Index

WEI : Water exploitation index
WEI+ : Water exploitation index plus
WSI : Water stress index
ZI : Z-Score index

REFERENCES

Ahmed, F., Siwar, C., & Begum, R. A. (2014). Water resources in Malaysia: Issues and challenges. *Journal of Food, Agriculture and Environment, 12*(2), 1100-1104. https://doi.org/10.1234/4.2014.5295

Azizi, H. R., Hassan Pourdarvishi, H., & Ebrahimi, H. (2017). Investigating the trend of changes in precipitation and discharge parameters of the Karaj River catchment under climate change conditions. In *Proceedings of the Second International Conference on Management, Demand and Efficiency of Water Consumption*. Hamedan, Iran.

Cordery, I., & McCall, M. (2000). A model for forecasting drought from teleconnections. *Water Resources Research, 36*(3), 763-768. https://doi.org/10.1029/1999WR900318

Dai, M., Huang, S., Huang, Q., Leng, G., Guo, Y., Wang, L., Fang, W., Li, P., & Zheng, X. (2020). Assessing agricultural drought risk and its dynamic evolution characteristics. *Agricultural Water Management, 231*, Article 106003. https://doi.org/10.1016/j.agwat.2020.106003

Department of Irrigation and Drainage. (2017). *List and location of hydrological stations for Malaysia*. http://h2o.water.gov.my/v2/index.cfm

Edwards, D. C., & McKee, T. B. (1997). Characteristics of 20th-century drought in the United States at multiple time scales. *Climatology Report No. 97-2*. Colorado State University, Department of Atmospheric Science.

Elfithri, R. (2018). Restoring and managing Langat River Basin, Malaysia: Challenges for a sustainable future. *International Journal of Environment and Sustainability, 6*(4). https://doi.org/10.24102/ijes.v6i4.897

Fan, L., Wang, H., Liu, Z., & Li, N. (2018). Quantifying the relationship between drought and water scarcity using copulas: Case study of Beijing-Tianjin-Hebei metropolitan areas in China. *Water, 10*(11), Article 1622. https://doi.org/10.3390/w10111622

Farid Ahmed, M., Mokhtar, M. B., Lim, C. K., Suza, I. A. B. C., Ayob, K. A. K., Khirotdin, R. P. K., & Majid, N. A. (2023). Integrated river basin management for sustainable development: Time for stronger action. *Water, 15*(13), Article 2497. https://doi.org/10.3390/w15132497

Fung, K. F., Huang, Y. F., & Koo, C. H. (2020). Assessing drought conditions through temporal pattern, spatial characteristic and operational accuracy indicated by SPI and SPEI: Case analysis for Peninsular Malaysia. *Natural Hazards, 103*(2), 2071-2101. https://doi.org/10.1007/s11069-020-04072-y

Guan, H., Jhiang, Z., Lu, J., & Wan, Y. (2025). Study of the correlation between water resource changes and drought indices in the Yinchuan Plain based on multi-source remote sensing and deep learning. *Water, 17*(18), Article 2740. https://doi.org/10.3390/w17182740

Hanafiah, M. M., Ghazali, N. F., Harun, S. N., Abdulaali, H. S., Abdul Hasan, M. J., & Amri Kamarudin, M. K. (2019). Assessing water scarcity in Malaysia: A case study of rice production. *Desalination and Water Treatment, 149*, 274-287. https://doi.org/10.5004/dwt.2019.23841

Huang, Y. F., Ang, J. T., Tiong, Y. J., Mirzaei, M., & Amin, M. Z. M. (2016). Drought forecasting using SPI and EDI under RCP-8.5 climate change scenarios for Langat River Basin, Malaysia. *Procedia Engineering, 154*, 710-717. https://doi.org/10.1016/j.proeng.2016.07.573

Hui-Mean, F., Yusop, Z., & Yusof, F. (2018). Drought analysis and water resource availability using the standardised precipitation evapotranspiration index. *Atmospheric Research, 201*, 102-115. https://doi.org/10.1016/j.atmosres.2017.10.014

Iglesias, A., Garrote, L., Flores, F., & Moneo, M. (2007). Challenges to managing the risk of water scarcity and climate change in the Mediterranean. *Water Resources Management, 21*(5), 775-788. https://doi.org/10.1007/s11269-006-9111-6

Isia, I., Hadibarata, T., Jusoh, M. N. H., Bhattacharjya, R. K., Shahedan, N. F., Bouaissi, A., Fitriyani, N. L., & Syafrudin, M. (2023). Drought analysis based on the Standardised Precipitation Evapotranspiration Index and Standardised Precipitation Index in Sarawak, Malaysia. *Sustainability, 15*(1), Article 734. https://doi.org/10.3390/su15010734

Khan, M. M. H., Muhammad, N. S., & El-Shafie, A. (2018). Wavelet-ANN versus ANN-based model for hydro-meteorological drought forecasting. *Water, 10*(8), Article 998. https://doi.org/10.3390/w10080998

Lam, M. R., Matano, A., Van Loon, A. F., Odongo, R. A., Teklesadik, A. D., Wamucii, C. N., van den Homberg, M. J. C., Waruru, S., & Teuling, A. J. (2023). Linking reported drought impacts with drought indices, water scarcity and aridity: The case of Kenya. *Natural Hazards and Earth System Sciences, 23*, 2915-2936. https://doi.org/10.5194/nhess-23-2915

Lana, X., Serra, C., & Burgueño, A. (2001). Patterns of monthly rainfall shortage and excess in terms of the Standardised Precipitation Index for Catalonia (NE Spain). *International Journal of Climatology, 21*(13), 1669-1691. https://doi.org/10.1002/joc.697

Li, C., Zhang, X., Yin, G., Xu, Y., & Hao, F. (2022). Evaluation of drought propagation characteristics and influencing factors in an arid region of Northeast Asia (ARNA). *Remote Sensing, 14*(14), Article 3307. https://doi.org/10.3390/rs14143307

Luhaim, Z., Tan, M. L., Tangang, F., Zulkafli, Z., Chun, K. P., Yusop, Z., & Yaseen, Z. M. (2021). Drought variability and characteristics in the Muda River Basin of Malaysia from 1985 to 2019. *Atmosphere, 12*(9), Article 1210. https://doi.org/10.3390/atmos12091210

Mondal, A., & Mujumdar, P. P. (2015). Return levels of hydrologic droughts under climate change. *Advances in Water Resources, 75*, 67-79. https://doi.org/10.1016/j.advwatres.2014.11.005

Natarajan, N., Vasudevan, M., Ahash Raja, S., Mohanpradaap, K., Sneha, G., & Joshna Shanu, S. (2023). An assessment methodology for drought severity and vulnerability using precipitation-based indices for the arid, semi-arid and humid districts of Tamil Nadu, India. *Water Supply, 23*(1), 54-68. https://doi.org/10.2166/ws.2022.415

Othman, A. S., & Tukimat, N. N. A. (2018). Assessment of the potential occurrence of dry period in the long term for Pahang State, Malaysia. *MATEC Web of Conferences, 150*, Article 03004. https://doi.org/10.1051/matecconf/201815003004

Othman, M., Ash'Aari, Z. H., Muharam, F. M., Sulaiman, W. N. A., Hamisan, H., Mohamad, N. D., & Othman, N. H. (2016). Assessment of drought impacts on vegetation health: A case study in Kedah. *IOP*

Conference Series: Earth and Environmental Science, 37(1), Article 012072. https://doi.org/10.1088/1755-1315/37/1/012072

Payus, C., Huey, L. A., Adnan, F., Rimba, A. B., Mohan, G., Chapagain, S. K., Roder, G., Gasparatos, A., & Fukushi, K. (2020). Impact of extreme drought climate on water security in North Borneo: Case study of Sabah. *Water (Switzerland), 12*(4), Article 1135. https://doi.org/10.3390/W12041135

Pedro-Monzonís, M., Solera, A., Ferrer, J., Estrela, T., & Paredes-Arquiola, J. (2015). A review of water scarcity and drought indices in water resources planning and management. *Journal of Hydrology, 527*, 482-493. https://doi.org/10.1016/j.jhydrol.2015.05.003

Quang, C. N. X., et al. (2021). Assessment of meteorological drought in the Vietnamese Mekong Delta in the period 1985-2018. *IOP Conference Series: Earth and Environmental Science, 652*(1), Article 012020. https://doi.org/10.1088/1755-1315/652/1/012020

Samsudin, M. F., Mohd Amin, M. F., Syed Omar, S. A., Mat Rasat, M. S., & Salam, M. A. (2020). Analysis of Water Stress Index (WSI) for the district surrounding Ulu Sat Forest Reserve, Kelantan, Malaysia. *IOP Conference Series: Earth and Environmental Science, 549*(1), Article 012015. https://doi.org/10.1088/1755-1315/549/1/012015

Shaaban, A. J., & Low, K. S. (2003, April 28). Droughts in Malaysia: A look at its characteristics, impacts, related policies, and management strategies. In *Proceedings of the Water and Drainage 2003 Conference* (pp. 43-54). Kuala Lumpur, Malaysia.

Shahbaz, M., Loganathan, N., Sbia, R., & Afza, T. (2015). The effect of urbanisation, affluence, and trade openness on energy consumption: A time series analysis in Malaysia. *Renewable and Sustainable Energy Reviews, 47*, 683-693. https://doi.org/10.1016/j.rser.2015.03.044

Shalamzari, M. J., & Zhang, W. (2018). Assessing water scarcity using the Water Poverty Index (WPI) in Golestan Province of Iran. *Water, 10*(8), Article 1079. https://doi.org/10.3390/w10081079

Singh, U., Agarwal, P., & Sharma, P. K. (2022). Meteorological drought analysis with different indices for the Betwa River basin, India. *Theoretical and Applied Climatology, 149*(3–4), 1333-1350. https://doi.org/10.1007/s00704-022-04027-2

Sondermann, M. N., & Proença de Oliveira, R. (2022). Using the WEI+ index to evaluate water scarcity at highly regulated river basins with conjunctive uses of surface and groundwater resources. *Science of the Total Environment, 836*, Article 155754. https://doi.org/10.1016/j.scitotenv.2022.155754

Tan, M. L., Juneng, L., Tangang, F. T., Chung, J. X., & Radin Firdaus, R. B. (2021). Changes in temperature extremes and their relationship with ENSO in Malaysia from 1985 to 2018. *International Journal of Climatology, 41*(S1), E2564-E2580. https://doi.org/10.1002/joc.6864

Tri, D. Q., Dat, T. T., & Truong, D. D. (2019). Application of meteorological and hydrological drought indices to establish drought classification maps of the Ba River basin in Vietnam. *Hydrology, 6*(2), Article 49. https://doi.org/10.3390/hydrology6020049

Wu, H., Soh, L. K., Samal, A., Hong, T., Marx, D., & Chen, X. (2009). Upstream–downstream relationships in terms of annual streamflow discharges and drought events in Nebraska. *Journal of Water Resource and Protection, 1*(5), 299-315. https://doi.org/10.4236/jwarp.2009.15037

Yusof, F., Hui-Mean, F., Suhaila, J., & Yusof, Z. (2013). Characterisation of drought properties with bivariate copula analysis. *Water Resources Management, 27*(12), 4183-4207. https://doi.org/10.1007/s11269-013-0402-4

Zhou, K., Li, J., Zhang, T., & Kang, A. (2021). The use of combined soil moisture data to characterise agricultural drought conditions and the relationship among different drought types in China. *Agricultural Water Management, 243*, Article 106479. https://doi.org/10.1016/j.agwat.2020.106479

Pertanika J. Sci. & Technol. 34 (4): 2233 - 2253 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

An Interactive Augmented Reality and Digital Twin Framework for IoT-Based Aquaponics in Urban Farming

Siti Khadijah Ali[1,2,3], Phang Kok Wai[3]*, Zainal Abdul Kahar[3], and Rahmita Wirza O.K. Rahmat[4]

[1]Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Institute for Social Science Studies (IPSAS), Putra InfoPort, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[3]Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[4]CASD Innovate Sdn. Bhd., Malaysia

ABSTRACT

Urban farming, specifically aquaponics, offers a sustainable approach to food production; however, novice practitioners often struggle due to the complexity of the system and a reliance on stable internet connectivity for real-time monitoring. This paper proposes an interactive Augmented Reality (AR) and Digital Twin (DT) framework for IoT-based aquaponics to address these limitations. Developed through a rigorous five-phase research methodology, the system integrates a resilient WebSocket communication architecture with edge-level data buffering, successfully recovering 92.38% of sensor records during simulated network outages. To sustain continuous environmental visualisation during hardware or network failures, a lightweight slope-only linear regression fallback mechanism was implemented, achieving high predictive accuracy ($R^2 > 0.9$) for dynamic parameters like temperature and humidity. Furthermore, to accommodate non-technical urbanites (aged 18 to 35), the mobile AR interface incorporates a rule-based mistouch detection algorithm using spatial and temporal thresholds, successfully intercepting 37.25% of interactions as accidental inputs to prevent physical actuation errors. A User Acceptance Test (UAT) with 20 participants yielded a System Usability Scale (SUS) score of 56.63, reflecting the initial cognitive load of spatial AR navigation, yet recorded a high overall user satisfaction rate of 85%. This paper contributes a resilient, fault-tolerant, and user-centric management solution that advances smart agriculture for urban farming initiatives.

Keywords: Aquaponics, augmented reality, digital twin, internet of things, smart agriculture, urban farming, user acceptance testing

ARTICLE INFO

Article history:
Received: 10 June 2026
Accepted: 23 July 2026
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.13

E-mail addresses:
ctkhadijah@upm.edu.my (Siti Khadijah Ali)
gs67387@student.upm.edu.my (Phang Kok Wai)
zainal_kahar@upm.edu.my (Zainal Abdul Kahar)
r.wirza@gmail.com (Rahmita Wirza O.K. Rahmat)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The global challenge of sustainability and efficient food production has been exacerbated by the rapid increase in urbanisation, climate change, and food security issues. Urban farming has been established to solve problems with food access through localised food production, while also decreasing the carbon footprint of food production compared to traditional farms. Urban farming has taken off in Malaysia, with the number of urban farms jumping from 735 in 2019 to over 17,000 in 2020, which indicates an increased interest in sustainable agriculture and food resilience for the public through new initiatives for urban farms.

Figure 1 illustrates that the most common forms of urban farming are aquaponics, which is a sustainable form of agriculture that integrates aquaculture and hydroponics into a symbiotic environment. The integration of aquaculture and hydroponic farming allows the simultaneous farming of fish and land-based plants for food. (Ibrahim et al., 2023). With the incorporation of aquaponics into urban sustainable agriculture, resource use efficiency is maximised and less environmental impact compared to traditional agriculture. However, small-scale or start-up aquaponic practitioners tend to struggle with technical skills, real-time monitoring opportunities, and the end-user cognitive load from relying on intuition and decision-making processes (Yanes et al., 2020).

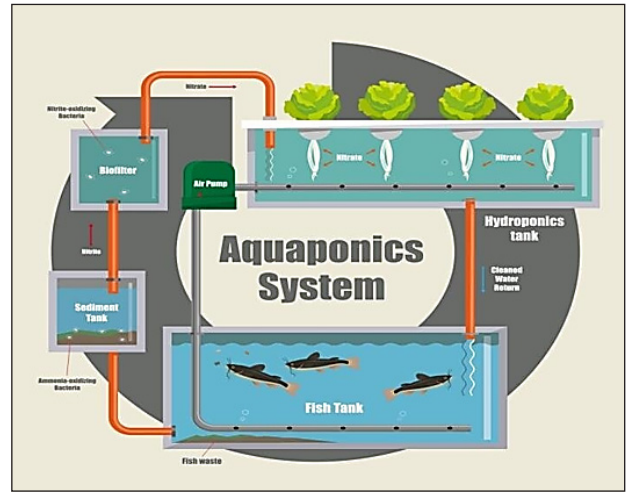


Figure 1. Illustration of an aquaponics system (Bambhaniya et al., 2023)

To address these operational difficulties, recent studies have explored the integration of the IoT, AR, and DT technologies. IoT systems are traditionally employed to provide real-time monitoring of essential parameters such as pH, temperature, and nutrient levels (Mahmoud et al., 2023; Ntulo et al., 2021). AR facilitates spatial interaction by overlaying this real-time data onto the physical environment (Wong et al., 2024), while DT models provide a virtual replica for system simulation and analysis (Verdouw et al., 2021). However, while these technologies represent significant advancements, a critical analysis of the current literature reveals that their mainstream integration into aquaponics remains insufficient. Most existing approaches treat these technologies

as isolated tools rather than a cohesive ecosystem, leading to unresolved operational vulnerabilities. Specifically, current IoT architectures in smart agriculture assume ideal, stable internet connectivity; consequently, they suffer from severe data loss and delayed system feedback during the intermittent network outages common in urban and remote farming environments (Channa et al., 2024; Taha et al., 2022).

Furthermore, existing AR interfaces designed for aquaponics often introduce new usability barriers for non-technical users. Because mobile AR environments lack tactile feedback, users frequently execute accidental commands, commonly referred to as "mistouches", which can unintentionally trigger physical system actuators (Yamanaka & Usuba, 2021; Zamnuri et al., 2024). Current literature lacks the implementation of rigorous HCI safeguards, such as spatial and temporal filtering, to prevent these critical interaction errors (Cockburn et al., 2019; Subramanian et al., 2021). Additionally, current DT implementations typically rely on continuous cloud synchronisation, meaning the virtual visualisation freezes or fails when physical sensors or network connections are interrupted. The absence of lightweight predictive fallback mechanisms to estimate environmental conditions during these outages disrupts system continuity and heavily erodes user trust (Lin et al., 2022; Otieno, 2023).

To address these specific research gaps, this study proposes a comprehensive DTIAR framework that tightly integrates IoT sensors, real-time DT synchronisation, and error-resistant AR interfaces into aquaponic systems. Unlike previous studies that merely combine existing technologies, this research introduces three novel advancements to ensure structural resilience and user accessibility. First, it implements a WebSocket-based communication architecture equipped with edge-level data buffering to mitigate data loss during network instability. Second, it embeds a rule-based mistouch detection algorithm within the AR interface to safely filter accidental inputs from novice users. Finally, it incorporates a slope-only simple linear regression predictive fallback mechanism to sustain continuous environmental visualisation even during complete sensor or network failures. By critically addressing the deficiencies of current smart farming systems, this paper provides a highly resilient, scalable, and user-centric DTIAR model suitable for learning institutions, urban farming programs, and community-driven sustainability initiatives.

Problem Statement

Global food systems are increasingly stressed by rapid population growth, climate change, and the scarcity of natural resources. Therefore, sustainable and novel agricultural methods, such as aquaponics, are essential (Food and Agriculture Organisation of the United Nations, 2022). Aquaponics represents a highly productive alternative for food production, combining aquaculture and hydroponics in a symbiotic relationship. Nonetheless, the widespread application and potential success of aquaponics systems are currently limited by several critical technological barriers.

First, many modern aquaponics systems are unable to provide reliable real-time monitoring and control due to unstable internet connectivity, resulting in delayed system feedback and severe data loss. Current systems are often not designed to function optimally with real-time data under fluctuating network conditions (Channa et al., 2024; Taha et al., 2022). Consequently, it is difficult for users to make quick responses regarding water quality, nutrient flows, and temperature, which hampers the system's biological stability (Chandramenon et al., 2024; Kok et al., 2024; Ntulo et al., 2021). Crucially, there is a lack of robust communication architectures that utilise local data buffering and deferred synchronisation to prevent data loss during these internet connectivity dropouts.

Second, existing AR aquaponics interfaces are highly prone to user interaction errors, which hinder effective system operation by non-technical users. Poorly designed mobile interfaces frequently lead to interaction errors, specifically accidental inputs or "mistouches". These interaction errors minimise usability, trigger unintended physical actuations, and deter user engagement, especially in educational or small-scale urban farming situations (Karimanzira & Rauschenbach, 2019; Mohamad et al., 2024; Zamnuri et al., 2024).

Third, many existing aquaponics monitoring systems are unable to maintain continuous environmental monitoring during internet outages or physical sensor interruptions. This disruption results in missing sensor data, halting system visualisation, which breaks operational continuity and erodes user trust (Eneh et al., 2023; Garzón et al., 2022; Sundararajan et al., 2025; Zamnuri et al., 2024). Specifically, these systems lack a lightweight predictive fallback mechanism capable of estimating and sustaining continuous data visualisation when physical sensors or networks temporarily fail.

To overcome these barriers, a responsive and robust framework is required. This framework must incorporate a live, low-latency communication layer using WebSocket with local buffering, an AR user interface equipped with a mistouch-prevention strategy, and a predictive fallback mechanism to ensure continuity of the user experience even during disturbances to the communication channel.

Aim of the Study

The overarching aim of this study is to improve the usability and resilience of smart aquaponic systems for non-technical users through the integration of IoT, DT and AR technologies. To realise this goal, the research pursues three specific, measurable objectives. First, it aims to enhance the communication architecture by integrating a resilient, WebSocket-based approach equipped with local data buffering and deferred synchronisation to significantly reduce data loss during intermittent network connectivity. Second, it seeks to improve the operational reliability of the AR interface by formulating a rule-based mistouch detection algorithm that utilises predefined spatial and temporal

thresholds to accurately distinguish intentional user commands from accidental inputs. Finally, the study aims to ensure the continuity of environmental data visualisation during hardware or network failures by implementing a lightweight, slope-only linear regression model ($\hat{y}_{n+1} = y_n + \beta_1$) as a predictive fallback mechanism within the mobile application.

LITERATURE REVIEW

The incorporation of digital technologies into aquaponics systems has increasingly focused on the IoT, AR, and DT to automate environmental controls and enhance user engagement. The deployment of IoT technologies has successfully transitioned aquaponics from passive observation to active, automated regulation, with recent studies demonstrating accessible microcontroller-based monitoring (Ntulo et al., 2021) and modular cloud-based agro-toolboxes (Pechlivani et al., 2023). However, a critical synthesis of these systems reveals a shared vulnerability: a heavy reliance on continuous, ideal network conditions (Wan et al., 2022). In practical agricultural deployments, unstable connectivity frequently leads to delayed system feedback and permanent data loss, detrimentally affecting ecosystem health (Channa et al., 2024; Taha et al., 2022). Existing literature predominantly lacks communication architectures explicitly designed to handle intermittent network outages through edge-level local data buffering and deferred synchronisation (Alselek et al., 2022).

To improve user engagement and data accessibility, AR is increasingly implemented to spatially visualise environmental metrics. Recent applications have successfully utilised mobile AR to overlay digital plant information onto physical crops, significantly increasing knowledge retention for novice users (Garzón et al., 2022; Yi Chien & Norasri Ismail, 2022). Despite this visual potential, current AR frameworks present a critical HCI limitation: they rely on static tap interfaces without safeguards against interaction errors. When non-technical users interact with floating 3D elements on mobile screens, the lack of tactile feedback frequently leads to accidental activations, or "mistouches" (Yamanaka & Usuba, 2021; Zamnuri et al., 2024). While HCI literature establishes that embedding strict spatial constraints and temporal thresholds can significantly improve reliability (Cockburn et al., 2019; Subramanian et al., 2021).These heuristic rules have not been sufficiently integrated into live agricultural AR interfaces to prevent the unintended triggering of physical actuators.

Furthermore, DT technology is increasingly utilised to develop virtual replicas of physical systems for continuous monitoring (Pylianidis et al., 2021). Advanced implementations have employed deep learning approaches, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, to forecast complex environmental dependencies (Sundararajan et al., 2025). However, these sophisticated models demand extensive computational power and large training datasets, creating a

high barrier to entry for small-scale urban farmers. More critically, most existing DT assume perfect hardware reliability; when physical sensors fail or disconnect, the virtual visualisation freezes or fails. There is a significant gap in the literature regarding the use of lightweight, computationally efficient models such as simple linear regression, which has shown high accuracy in aquaculture forecasting (Eneh et al., 2023), to act as immediate predictive fallback mechanisms capable of sustaining visual continuity during these outages.

In summary, a comparative analysis underscores a pervasive fragmentation within current smart agriculture literature, where research typically prioritises isolated technological optimisation over holistic, resilient integration. Table 1 provides a systematic comparative evaluation of key recent studies, highlighting the specific functional limitations of existing frameworks compared to the proposed system. While existing studies have pioneered AR educational tools (Garzón et al., 2022) and advanced predictive models (Sundararajan et al., 2025). They remain highly vulnerable to network instability and user interaction errors in real-world deployments. The proposed framework fundamentally differs from previous studies by bridging these silos: it integrates WebSocket-based edge buffering to ensure data integrity during network drops, embeds rule-based mistouch detection directly into the AR spatial interface to safeguard against user errors, and deploys a lightweight predictive fallback mechanism to maintain operational continuity. This integrated approach ensures a fault-tolerant, user-centric system that overcomes the distinct technical limitations of preceding standalone implementations.

Table 1
Summary of key studies on IoT, AR, and DT applications in aquaponic systems

Study (Year)	Technologies	Key Strengths	Identified Limitations
Ntulo et al., 2021	IoT, Microcontrollers	Accessible, low-cost real-time monitoring.	Assumes stable networks; lacks local data buffering, leading to data loss during outages.
Verdouw et al., 2021	DT	Real-world validation in commercial agriculture.	High implementation complexity; relies on continuous cloud synchronisation without fallback mechanisms.
Garzón et al., 2022	Augmented Reality	High user engagement via spatial visualisation.	Lacks HCI safeguards (spatial/temporal thresholds) to prevent accidental mistouches.
Sundararajan et al., 2025	IoT, Deep Learning	High predictive accuracy for environmental conditions.	Computationally intensive; lacks lightweight fallback options for continuous visualisation during sensor failures.
Proposed DTIAR Framework (2026)	IoT, AR, DT	Resilient, user-centric, and continuous system monitoring.	Addresses existing gaps by integrating WebSocket edge-buffering, rule-based AR mistouch detection, and an SLR predictive fallback mechanism.

METHODOLOGY

To ensure the proposed DTIAR framework satisfies both rigorous technical standards and end-user requirements, this study employs an iterative, design-based research methodology. Moving beyond a standard software development workflow, the research design is systematically structured into five sequential phases: Formulation and Planning, System Development, Evaluation and Refinement, Final Validation and Analysis, and Finalisation and Documentation.

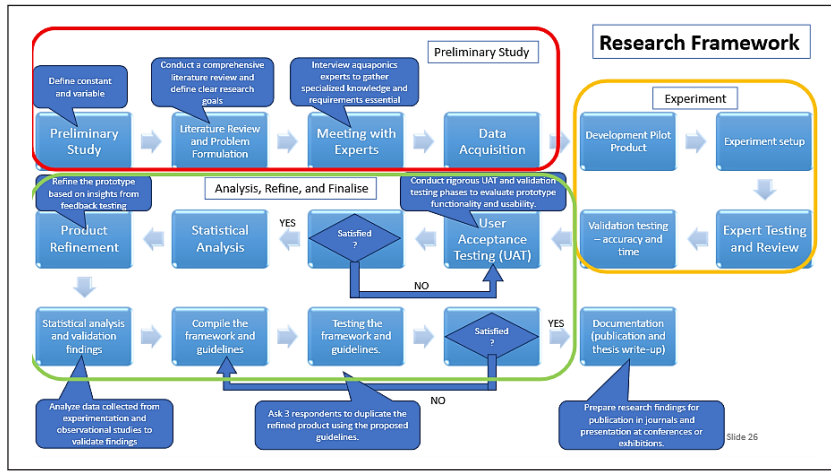


Figure 2. Research framework diagram

As illustrated in Figure 2, this research framework clearly maps the inputs, decision criteria, and outputs across the study. During the Formulation phase, preliminary expert sub-studies served as primary inputs, where consultations with an architectural design expert and an aquaculture specialist established critical biological safety thresholds (such as maintaining water pH between 6.5 and 7.5) and validated the physical system layout. Following the integration of IoT hardware and the AR interfaces during System Development, the Evaluation and Refinement phase utilised an initial pilot study comprising 50 participants to assess early-stage AR interaction mechanics. The output of this pilot study directly influenced the system's decision criteria, leading to critical adjustments in the interactive bounding boxes to reduce user cognitive load before moving to final testing.

To provide a comprehensive evaluation of the system's technical performance and address the relationships among IoT, DT, and AR components, the Final Validation phase employed explicit, objective evaluation metrics across three core variables. First, the resilience of the IoT communication architecture was evaluated by measuring data recovery and completion rates during simulated intermittent network outages to validate

the efficacy of the local edge-buffering mechanism. Second, the reliability of the AR interface was quantified by logging touch interactions and calculating the mistouch classification rate. This classification acted as a strict decision criterion, governed by specific heuristic thresholds: a spatial Movement Distance Threshold (invalidating touches exceeding 25 pixels) and a temporal Touch Duration Threshold (flagging contacts lasting longer than 400 milliseconds as unintentional) (Cockburn et al., 2019; Yamanaka & Usuba, 2021). Third, the accuracy of the DT's predictive fallback mechanism was rigorously evaluated using standard statistical error metrics, specifically MSE, RMSE, MAE, and R^2 to compare the algorithm's forecasted values against actual sensor data recovered post-reconnection (Chicco et al., 2021).

To measure user perceptions regarding usability and satisfaction, a formal User Acceptance Test (UAT) was conducted during the Final Validation phase. The study utilised a purposive sampling strategy, selecting 20 participants. The inclusion criteria explicitly targeted young adult urbanites (aged 18 to 35) who possessed baseline digital and smartphone literacy but lacked prior technical training or expertise in aquaponics or smart farming systems. The selection of a 20-participant sample size for the final UAT is scientifically justified by established Human-Computer Interaction (HCI) research, which demonstrates that a sample of 20 users provides a 95% probability of uncovering major system usability issues (Faulkner, 2003). Data collection for the UAT was conducted using a structured, digital questionnaire comprising demographic profiling, task-based ratings, and the standardised 10-item SUS (Brooke, 2020). Participants rated items on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). To ensure the statistical reliability and internal consistency of the survey instrument, Cronbach's alpha (α) was calculated for the collected responses. Finally, because this phase of the study involved human participants, strict ethical considerations were observed throughout the evaluation process. Informed consent was formally secured from all participants before they engaged with the system, and all collected demographic feedback and interaction data were fully anonymised to protect participant confidentiality and ensure data privacy.

System Interface and Functional Design

The mobile application serves as the User Layer of the DTIAR framework, developed utilising the Unity 3D engine and ARCore to seamlessly bridge the physical and digital domains. As illustrated in the application storyboard in Figure 3, the user experience is structured around three primary functional modules designed to support non-technical practitioners: the Status Page, the AR Page, and the Settings Page. The Status Page functions as the central dashboard, providing real-time data visualisation of critical environmental parameters such as water temperature, pH, humidity, and light intensity through numeric values, colour-coded alerts, and graphical gauges. The AR Page immerses the user by

superimposing live telemetry readings and 3D models directly onto the corresponding physical aquaponics infrastructure via markerless tracking, facilitating intuitive spatial learning. Finally, the Settings Page allows users to configure application preferences, including theme customisation, font sizes, and specific notification thresholds.

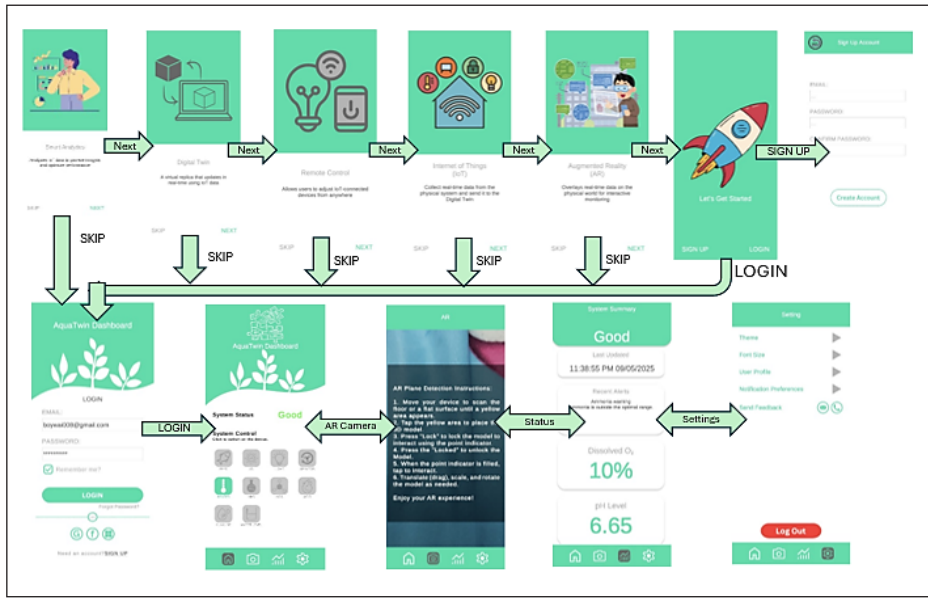


Figure 3. Storyboard of the application

To ensure the application satisfies recognised usability principles, particularly for novice users operating in complex spatial environments, the interface design requirements were strictly derived from established Human-Computer Interaction (HCI) heuristics. Mobile AR environments frequently suffer from the "fat finger problem," where the inherent instability of holding a mobile device and the lack of tactile haptic feedback led to high rates of accidental actuations. To mitigate this, the AR interface evaluates every user interaction against a newly integrated, rule-based mistouch detection algorithm configured with strict quantitative design thresholds. Specifically, the system applies a Movement Distance Threshold that invalidates any touch involving more than 25 pixels of movement, successfully differentiating a deliberate tap from an unintentional drag (Parhi et al., 2006).

Additionally, the interface evaluates inputs against a Touch Duration Threshold, categorising any contact lasting longer than 0.4 seconds (400 milliseconds) as an accidental resting touch to ensure only rapid, decisive inputs are processed as commands (Cockburn et al., 2019). Finally, the design enforces a precise Button Target Zone, requiring user interactions to land within the central 30–70% of the interactive button area to be registered as valid. By systematically evaluating user inputs against these recognised HCI principles,

the interface safely filters out accidental mistouches. This design significantly reduces the cognitive load on novice users and serves as a critical operational safeguard, preventing unintentional triggers of the physical aquaponics actuators during AR navigation.

System Architecture and Connectivity

The DTIAR system connects IoT, DT, and AR technology as a single smart aquaponics management platform accessible via a mobile application. To ensure true DT synchronisation extending beyond mere 3D visualisation by maintaining continuous, bidirectional correspondence between the physical and virtual environments, the system architecture is engineered into a decentralised five-layer framework: the Edge/Physical Layer, Network Layer, Cloud Layer, Intelligence Layer, and User Layer. As illustrated in Figure 4, the architecture governs the precise flow of time-series data and control signals across the ecosystem.

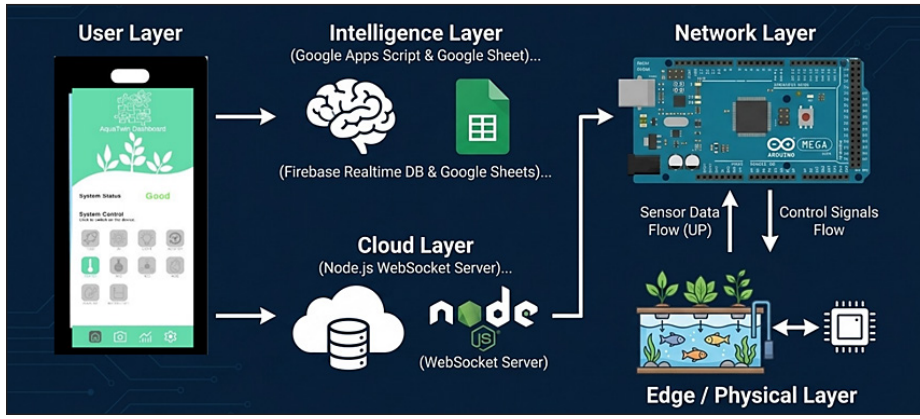


Figure 4. Full system architecture

The foundation of the architecture resides in the Edge/Physical Layer, which comprises the physical aquaponics infrastructure (fish tanks, plant beds, and LED grow lights) alongside the environmental sensor suite and physical actuators. Data acquisition and hardware control are managed within the Network Layer by an Arduino MEGA R3 + Wi-Fi (ATmega2560 + ESP8266) microcontroller. The ATmega2560 handles high-frequency analogue and digital reads from sensors, including the analogue pH meter, TDS/EC sensor, waterproof DS18B20 digital temperature probe, DHT22 air temperature/humidity sensor, HC-SR04 ultrasonic water level sensor, and GY-30 light intensity sensor, while the onboard ESP8266 manages wireless communication. To protect against data loss during network instability, the microcontroller maintains a local edge-buffering memory queue; when internet connectivity is interrupted, telemetry

is stored locally and transmitted via deferred batch synchronisation upon reconnection, ensuring zero historical data loss.

The Cloud Layer serves as the central communication and storage broker, utilising a Node.js WebSocket server hosted on Render.com paired with a Firebase Realtime Database. The technical selection of a WebSocket protocol over traditional HTTP polling is justified by the necessity for persistent, full-duplex communication, which successfully reduces average communication latency to under 200 milliseconds. Sensor data flows upward from the microcontroller as formatted JSON payloads, updating Firebase in real-time and broadcasting live updates to all connected Unity clients. Conversely, user control commands flow downward from the mobile app through the WebSocket to Firebase, instantly triggering physical relays (such as acid/alkaline dosing pumps or aerators) on the hardware.

Operating concurrently with the cloud backend is the Intelligence Layer, which drives the DT's predictive fallback mechanism via Google Apps Script. If the system experiences a network dropout exceeding 30 seconds, the intelligence engine automatically triggers a lightweight, slope-only simple linear regression model ($\hat{y}_{n+1} = y_n + \beta_1$). By extrapolating from recent historical observation windows, the model calculates and injects estimated sensor values into the database. As shown in Figure 5, these forecasted values maintain visual and data continuity within the AR interface, preventing the visualisation gaps typically associated with connectivity failures. Upon reconnection, an automated recovery protocol overwrites the synthetic estimates with actual physical telemetry, ensuring complete data integrity.

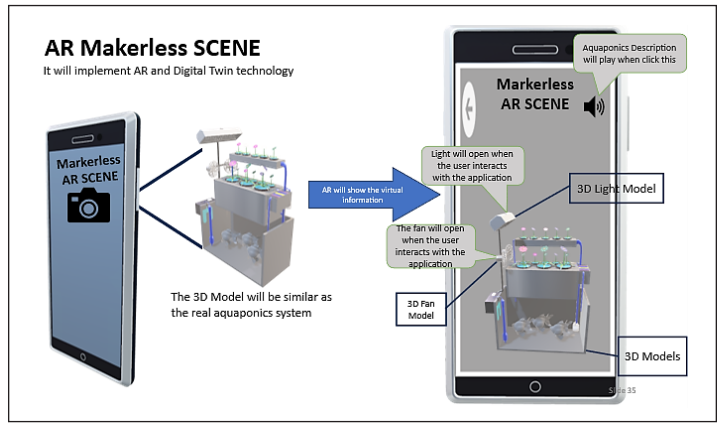


Figure 5. AR scene of the application

The successful integration of IoT sensors, real-time DT synchronisation, and augmented reality technologies effectively addresses the operational limitations of distance, latency, and data complexity highlighted in the research problem statement. By combining the

seamless real-time data collection of the DT with lightweight predictive analytics, the system maintains consistent oversight of critical water quality parameters, nutrient levels, and temperatures without requiring continuous manual intervention. Furthermore, the AR interface successfully democratises system interaction, lowering the technical barrier to entry and enabling users across the technical ability spectrum to engage in efficient, immersive smart agriculture management.

RESULT AND DISCUSSIONS

Technical Evaluation of System Objectives

To validate the technical effectiveness and operational resilience of the proposed framework, the system was evaluated across three core objectives: communication resilience, AR interaction reliability, and predictive fallback accuracy.

First, communication resilience was objectively assessed to quantify the system's ability to prevent data loss during network interruptions. Under stable conditions, the WebSocket architecture successfully maintained a low-latency bidirectional data stream with an average upload interval of 5.8 seconds and a standard deviation of only ± 0.2 seconds. During a simulated 34-minute network outage test, the local edge-buffering mechanism effectively queued the disconnected telemetry. Upon reconnection, the deferred synchronisation protocol successfully recovered the data, resulting in 92.38% of the total records being pushed via buffered uploads, as summarised in Table 2. This mechanism prevented a 28.23% permanent data loss rate that would have otherwise occurred in traditional, unbuffered cloud architectures, thereby confirming the framework's robustness against real-world connectivity fluctuations.

Second, the operational reliability of the mobile AR interface was evaluated by logging and classifying 102 user touch interactions during unguided pilot testing. The rule-based mistouch detection algorithm utilised strict heuristic constraints, invalidating interactions that exceeded a spatial movement of 25 pixels (Parhi et al., 2006) or a temporal duration of 400 milliseconds (Cockburn et al., 2019). As detailed in Table 3, the classification results revealed that 62.75% of inputs were intentional commands, while 37.25% were flagged as accidental mistouches, which

Table 2
Real-time vs buffered data upload summary

Metric	Count	Percentage
Real-Time Uploads	8	7.62%
Buffered Uploads	97	92.38%
Total Records Pushed	105	100%

Table 3
Rule-based classification summary

Classification Type	Count	Percentage
Intentional Touch	64	62.75%
Mistouch	38	37.25%
Total	102	100%

were characterised by longer mean durations of 0.44 seconds and erratic movement trajectories averaging 67.30 pixels. While a 37.25% mistouch, rate is significantly higher than standard 2D mobile benchmarks, this metric critically reflects the steep spatial learning curve novice users face in unguided AR environments rather than an algorithmic failure. Crucially, the algorithm succeeded in its primary objective by intercepting these 38 accidental events, ensuring that resting touches and inadvertent swipes did not trigger unintentional physical actuations within the aquaponics hardware.

Third, the accuracy of the predictive fallback mechanism was rigorously quantified using standard statistical error measures, specifically MSE, RMSE, MAE, and R^2 (Chicco et al., 2021). As outlined in Table 4, the slope-only simple linear regression model demonstrated high predictive strength when applied to dynamic environmental parameters. For example, water temperature and air humidity forecasts yielded exceptional R^2 values of 0.9898 and 0.9858, respectively, confirming the model's reliability in sustaining data visualisation for fluctuating variables during short-term sensor outages. However, the analysis also exposed a critical algorithmic limitation when applied to low-variance parameters. The water level sensor yielded a negative R^2 value of -1.0033, demonstrating that the trend-seeking nature of linear regression performs worse than a naive mean predictor when applied to near-static data. This unexpected observation indicates that future DT implementations must adopt parameter-specific forecasting strategies, utilising static state monitoring rather than regression models for stable environmental thresholds.

Table 4
Performance metrics of the predictive model across sensor fields

Field	MSE	RMSE	MAE	R^2
pH	0.071	0.2665	0.0559	0.6334
WaterTemp	0.0524	0.2289	0.0373	0.9898
AirTemp	0.0775	0.2784	0.0462	0.9896
AirHumidity	0.3522	0.5935	0.1464	0.9858
LightIntensity	566.4417	23.8	3.6573	0.9102
WaterLevel	0.0223	0.1493	0.0086	-1.0033
WaterQuality	170.7588	13.0675	1.5537	0.1703

User Acceptance Testing and Usability Evaluation

Following the technical validation, the holistic usability and user satisfaction of the system were evaluated through a UAT. As illustrated in Figure 6, the UAT was conducted to assess the DTIAR aquaponics system's functionality, usability, and user satisfaction. The evaluation took place in Multimedia Laboratory 1 at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, involving 50 undergraduate and postgraduate participants who were either familiar with aquaponics or had no prior

experience with the system. Participants received a real-time system demonstration followed by hands-on opportunities with the DT-AR application, after which data were collected using a structured survey to gather quantitative and qualitative insights regarding system performance, usability, and user experience.

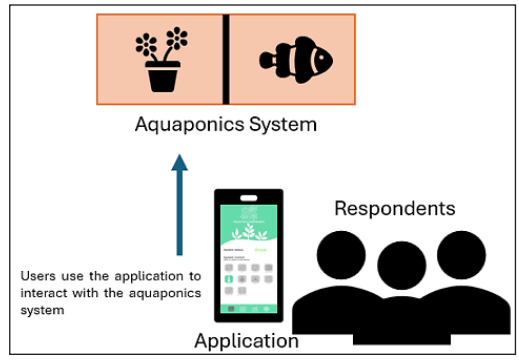


Figure 6. Method during UAT

Respondent demographics, as summarised in Figure 7, revealed a predominantly young adult population aged 20 and 21 years old, representing 32% and 24% of the cohort respectively, forming a digitally literate group well-suited for AR smart farming systems.

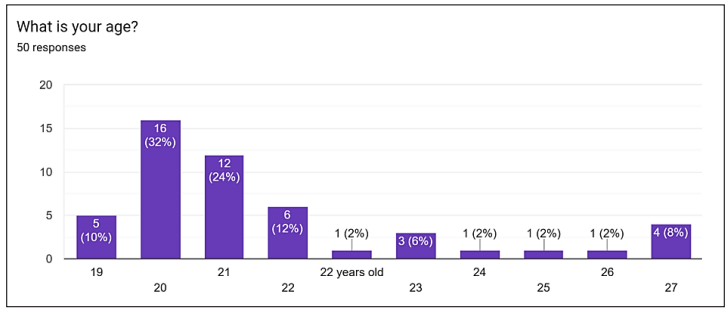


Figure 7. Demographic respondents

Regarding prior exposure to augmented reality, Figure 8 shows that 66% of participants reported previous experience with AR technology, while 34% were first-time users. This indicated that the interface needed to be flexible and intuitive enough to accommodate both novice users and those experienced with AR.

When evaluating preferred AR display modes in Figure 9, 56% of respondents favoured a simple AR display, while 40% preferred an overhead AR format for a more immersive experience.

Furthermore, as detailed in Figure 10, participants ranked the most desirable features of the application as visual quality (78%), interactivity (72%), and ease of use (68%), whereas mobile accessibility was preferred by 58%, and information richness was the least favoured feature at 32%, demonstrating that users prefer a visually clean interface over a cluttered data display.

Participants also evaluated acceptable price points for the system in Figure 11, with the largest group (approximately 25%) expecting a cost ranging from RM301 to RM500.

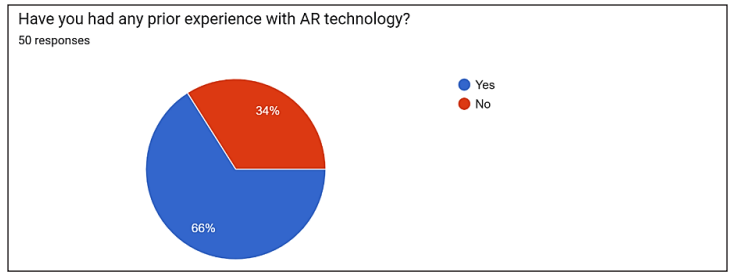


Figure 8. Respondents' experience with AR

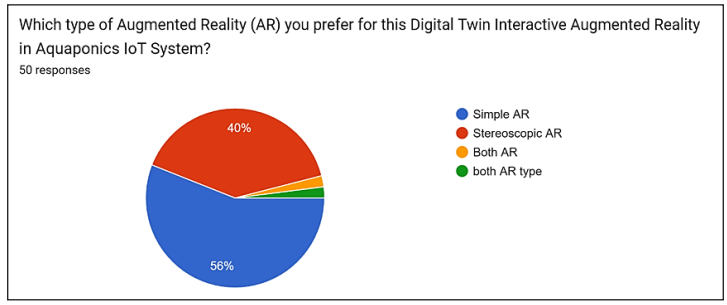


Figure 9. Type of AR respondents preferred

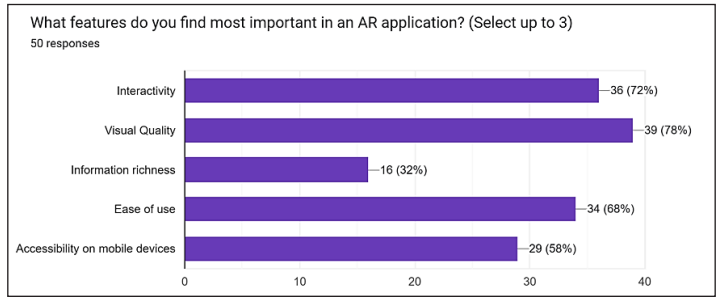


Figure 10. Features that respondents liked

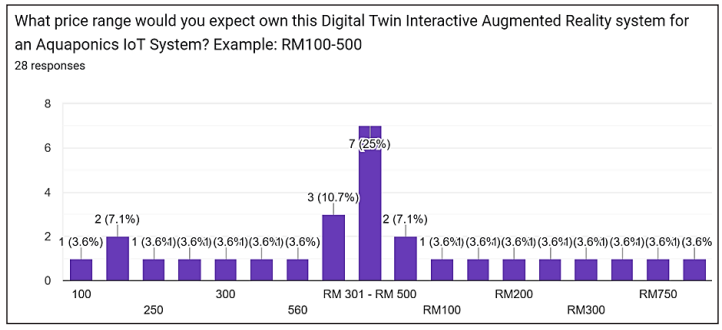


Figure 11. Expected cost of the system

In open-ended feedback shown in Figure 12, 15 respondents praised the interactive and engaging nature of the 3D animated models, with comments noting that stereoscopic AR offers a novel approach to future problem-solving.

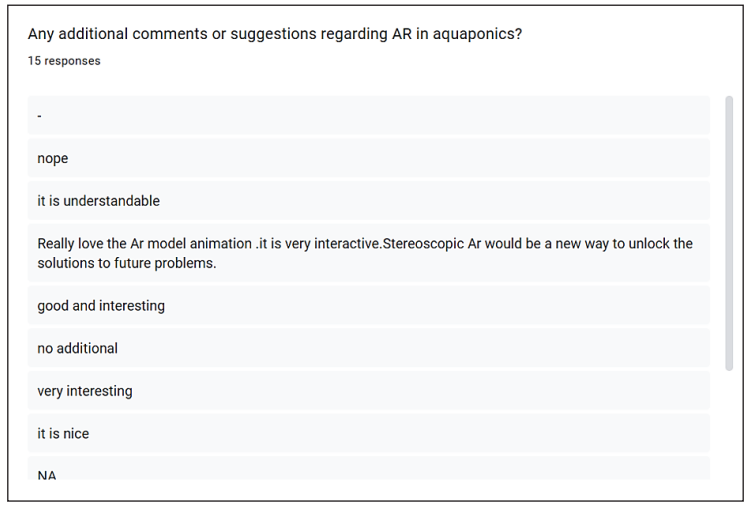


Figure 12. Comment from respondents

To complement these survey distributions with statistical rigour, Table 5 illustrated the quantitative usability analysis produced a mean SUS score of 56.63 with a standard deviation of ± 7.18, while the reliability of the instrument yielded a Cronbach's alpha of $\alpha = 0.63$. Although this SUS score falls into the marginal usability category according to established benchmarks, it sharply contrasts with the overall user satisfaction rate, where 85% of participants reported being satisfied or very satisfied with the platform. This discrepancy provides a critical insight: while users highly value the functional capabilities of the integrated AR and IoT system, the physical manipulation and spatial navigation required by the AR interface impose a high initial cognitive load. Overall, the UAT confirmed that the DT-AR aquaponics system provides an engaging, user-friendly experience that aligns with user expectations for digital smart agriculture management, while highlighting the necessity for future iterations to expand UI hitboxes and provide explicit onboarding cues.

Table 5
Summary of UAT metrics

Statistic	Value
Mean SUS Score	56.63
Standard Deviation	±7.18
Minimum SUS Score	42.5
Maximum SUS Score	70.0

Note. Cronbach's Alpha (α) = 0.630447046 ≈ 0.63

CONCLUSION

This study successfully designed, developed, and evaluated an integrated DTIAR framework to address the critical challenges of network instability, interaction reliability, and system resilience in smart aquaponic systems. By systematically bridging the technological silos between IoT hardware, cloud communications, augmented reality interfaces, and DT analytics, the research delivered a comprehensive management platform tailored for non-technical practitioners.

The technical evaluations confirmed the robustness of the proposed framework across its core objectives. The WebSocket communication architecture, paired with an edge-level local buffering and deferred synchronisation protocol, successfully achieved a 92.38% data recovery rate during network outages, effectively preventing the 28.23% data loss rate typical of unbuffered systems. Furthermore, the implementation of a rule-based mistouch detection algorithm governed by strict spatial movement and temporal duration thresholds successfully safeguarded the AR interface, intercepting 38 accidental touch events and preventing unintended hardware actuations during unguided navigation. The integration of a slope-only simple linear regression predictive fallback mechanism also demonstrated exceptional accuracy in sustaining data visualisation continuity for dynamic parameters such as water temperature ($R^2 = 0.9898$) and air humidity ($R^2 = 0.9858$) during short-term sensor disconnections.

Finally, the holistic usability and user satisfaction of the system were thoroughly validated through UAT. Engaging 50 diverse participants, the evaluation demonstrated that the platform delivers an engaging, immersive experience that aligns with user expectations for digital smart agriculture management. While the quantitative usability analysis yielded a marginal SUS score of 56.63, this was contrasted by an 85% overall user satisfaction rate, providing critical insight into the cognitive load associated with spatial AR navigation. Ultimately, this research provides a robust, fault-tolerant foundation for future smart farming innovations, offering clear pathways for enhancing interface accessibility and parameter-specific predictive modelling in agricultural technology.

FUTURE WORK

Building upon the foundational successes and identified limitations of the current DTIAR framework, several promising avenues remain for future research and system enhancement.

First, to address the performance limitations observed with low-variance parameters such as the negative R^2 value recorded for the water level sensor, future iterations of the DT intelligence layer should incorporate parameter-specific adaptive forecasting models. Rather than relying solely on linear regression, the system could dynamically switch between static state-monitoring algorithms for stable thresholds and advanced machine learning models

(such as Long Short-Term Memory networks) for highly dynamic environmental metrics to improve overall prediction accuracy.

Second, given that the UAT revealed a marginal SUS score driven by the steep spatial learning curve of unguided AR navigation, future development should focus on enhancing interface ergonomics. This includes expanding interactive UI hitboxes, integrating haptic feedback cues, and implementing a guided, step-by-step interactive onboarding tutorial to actively reduce user cognitive load and minimise mistouches for novice practitioners.

Finally, future work should expand the physical deployment scope of the framework from controlled laboratory environments to large-scale, commercial multi-system agricultural setups. Testing the architecture across prolonged real-world deployments involving multiple distributed microcontrollers and concurrent user sessions will provide deeper insights into cloud scalability, long-term network resilience, and the economic viability of the platform for everyday urban farmers.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education Malaysia under the Inisiatif Putra Siswazah (IPS) Grant Scheme [Project Code: UPM.RMC.800-3/3/3/1/2024/GP-IPS/9788500]. The author would like to thank the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, for providing the necessary facilities and support for this research. Very special thanks are extended to all users who participated in the UAT for their invaluable input and feedback in improving the proposed system.

LIST OF ABBREVIATIONS

Term	:	Definition
DT	:	Digital twin
AR	:	Augmented reality
IoT	:	Internet of Things
DTIAR	:	Digital twin interactive augmented reality
UAT	:	User acceptance testing
HCI	:	Human-computer interaction
SUS	:	System Usability Scale
MAE	:	Mean absolute error
MSE	:	Mean squared error
RMSE	:	Root mean squared error
R^2	:	Coefficient of determination
UI	:	User Interface

REFERENCES

Alselek, M., Alcaraz-Calero, J. M., Segura-Garcia, J., & Wang, Q. (2022). Water IoT monitoring system for aquaponics health and fishery applications. *Sensors, 22*(19), Article 7679. https://doi.org/10.3390/s22197679

Bambhaniya, I., Kamleshbhai, B. P., & Bajaniya, V. C. (2023). Unlocking sustainable agriculture: Aquaponics system. In *Dimension in Agricultural Sciences: A Way Forward for Sustainability*.

Brooke, J. (2020). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 207-212). CRC Press. https://doi.org/10.1201/9781498710411-35

Chandramenon, P., Gascoyne, A., & Tchuenbou-Magaia, F. (2024). An IoT and machine learning approach for the determination of optimal freshwater replenishment rate in an aquaponics system. *Frontiers in Sustainable Resource Management, 3*, Article 1363914. https://doi.org/10.3389/fsrma.2024.1363914

Channa, A. A., Munir, K., Hansen, M., & Tariq, M. F. (2024). Optimisation of small-scale aquaponics systems using artificial intelligence and the IoT: Status, challenges, and opportunities. *Encyclopedia, 4*(1), 313-336. https://doi.org/10.3390/encyclopedia4010023

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination, R-squared, is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*, Article e623. https://doi.org/10.7717/peerj-cs.623

Cockburn, A., Masson, D., Gutwin, C., Palanque, P., Goguey, A., Yung, M., Gris, C., & Trask, C. (2019). Design and evaluation of braced touch for touchscreen input stabilisation. *International Journal of Human-Computer Studies, 122*, 21-37. https://doi.org/10.1016/j.ijhcs.2018.08.005

Eneh, A. H., Udanor, C. N., Ossai, N. I., Aneke, S. O., Ugwoke, P. O., Obayi, A. A., Ugwuishiwu, C. H., & Okereke, G. E. (2023). Towards an improved Internet of Things sensors data quality for a smart aquaponics system yield prediction. *MethodsX, 11*, Article 102436. https://doi.org/10.1016/j.mex.2023.102436

Food and Agriculture Organisation of the United Nations. (2022). *The state of world fisheries and aquaculture 2022: Towards blue transformation*. Food and Agriculture Organisation of the United Nations. https://doi.org/10.4060/cc0461en

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behaviour Research Methods, Instruments, & Computers, 35*(3), 379-383. https://doi.org/10.3758/BF03195514

Garzón, J., Acevedo, J., Pavón, J., & Baldiris, S. (2022). Promoting eco-agritourism using an augmented reality-based educational resource: A case study of aquaponics. *Interactive Learning Environments, 30*(7), 1200-1214. https://doi.org/10.1080/10494820.2020.1712429

Ibrahim, L. A., Shaghaleh, H., El-Kassar, G. M., Abu-Hashim, M., Elsadek, E. A., & Alhaj Hamoud, Y. (2023). Aquaponics: A sustainable path to food sovereignty and enhanced water use efficiency. *Water, 15*(24), Article 4310. https://doi.org/10.3390/w15244310

Karimanzira, D., & Rauschenbach, T. (2019). Enhancing aquaponics management with IoT-based predictive analytics for efficient information utilisation. *Information Processing in Agriculture, 6*(3), 375-385. https://doi.org/10.1016/j.inpa.2018.12.003

Kok, C. L., Kusuma, I. M. B. P., Koh, Y. Y., Tang, H., & Lim, A. B. (2024). Smart aquaponics: An automated water quality management system for sustainable urban agriculture. *Electronics, 13*(5), Article 820. https://doi.org/10.3390/electronics13050820

Lin, W. L., Wang, S. C., Chen, L. S., Lin, T. L., & Lee, J. L. (2022). Green care achievement based on aquaponics combined with human-computer interaction. *Applied Sciences, 12*(19), Article 9809. https://doi.org/10.3390/app12199809

Mahmoud, M. M. M., Darwish, R., & Bassiuny, A. M. (2023). Development of an economically smart aquaponic system based on IoT. *Journal of Engineering Research*. Advance online publication. https://doi.org/10.1016/j.jer.2023.08.024

Otieno, M. (2023). Review of the techniques and mechanisms for enhancing trust in the Internet of Things for smart agriculture. *Global Journal of Engineering and Technology Advances, 15*(1), 50-63. https://doi.org/10.30574/gjeta.2023.15.1.0070

Mohamad, A. N., Zamani, Z. B., Jamaluddin, M. H., Soh, C., & Said, B. (2024). Enhancing aquaponics efficiency with microcontroller-based control. *International Journal of Academic Research in Business and Social Sciences, 14*(12), 4179-4193. https://doi.org/10.6007/IJARBSS/v14-i12/24402

Ntulo, M. P., Owolawi, P. A., Mapayi, T., Malele, V., Aiyetoro, G., & Ojo, J. S. (2021). IoT-based smart aquaponics system using Arduino Uno. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICECCME52200.2021.9590982

Parhi, P., Karlson, A. K., & Bederson, B. B. (2006). Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06)* (pp. 203-210). ACM. https://doi.org/10.1145/1152215.1152260

Pechlivani, E. M., Papadimitriou, A., Pemas, S., Ntinas, G., & Tzovaras, D. (2023). IoT-based agro-toolbox for soil analysis and environmental monitoring. *Micromachines, 14*(9), Article 1698. https://doi.org/10.3390/mi14091698

Pylianidis, C., Osinga, S., & Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. *Computers and Electronics in Agriculture, 184*, Article 105942. https://doi.org/10.1016/j.compag.2020.105942

Subramanian, A. K., Samanta, A., Manickam, S., Kumar, A., Shiaeles, S., & Mahendran, A. (2021). Linear regression trust management system for IoT systems. *Cybernetics and Information Technologies, 21*(4), 15-27. https://doi.org/10.2478/cait-2021-0040

Sundararajan, S. C. M., Shankar, Y. B., Selvam, S. P., Manogaran, N., Seerangan, K., Natesan, D., & Selvarajan, S. (2025). IoT-based prediction model for aquaponic fishpond water quality using multiscale feature fusion with a convolutional autoencoder and GRU networks. *Scientific Reports, 15*(1), Article 1925. https://doi.org/10.1038/s41598-024-84943-7

Taha, M. F., ElMasry, G., Gouda, M., Zhou, L., Liang, N., Abdalla, A., Rousseau, D., & Qiu, Z. (2022). Recent advances of smart systems and Internet of Things (IoT) for aquaponics automation: A comprehensive overview. *Chemosensors, 10*(8), Article 303. https://doi.org/10.3390/chemosensors10080303

Verdouw, C., Tekinerdogan, B., Beulens, A., & Wolfert, S. (2021). Digital twins in smart farming. Agricultural Systems, 189, Article 103046. https://doi.org/10.1016/j.agsy.2020.103046

Wan, S., Zhao, K., Lu, Z., Li, J., Lu, T., & Wang, H. (2022). A modularised IoT monitoring system with edge-computing for aquaponics. Sensors, 22(23), Article 9260. https://doi.org/10.3390/s22239260

Wong, K. T., Hosshan, H., Hanafi, H. F., & Mudiono, A. (2024). Augmented reality (AR): An assistive technology for special education needs. Journal of Advanced Research in Applied Sciences and Engineering Technology, 35(1), 97-105. https://doi.org/10.37934/araset.35.1.97105

Yamanaka, S., & Usuba, H. (2021). Computing touch-point ambiguity on mobile touchscreens for modelling target selection times. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(4), Article 171. https://doi.org/10.1145/3494976

Yanes, A. R., Martinez, P., & Ahmad, R. (2020). Towards automated aquaponics: A review on monitoring, IoT, and smart systems. Journal of Cleaner Production, 263, Article 121571. https://doi.org/10.1016/j.jclepro.2020.121571

Yi Chien, L., & Norasri Ismail, M. (2022). Development of a mobile application for applying aquaponics techniques in farming using an augmented reality approach. *Applied Information Technology and Computer Science, 3*(2), 591-610. https://doi.org/10.30880/aitcs.2022.03.02.037

Zamnuri, M. A. H. bin, Qiu, S., Rizalmy, M. A. A. bin, He, W., Yusoff, S., Roeroe, K. A., Du, J., & Loh, K. H. (2024). Integration of IoT in small-scale aquaponics to enhance efficiency and profitability: A systematic review. *Animals, 14*(17), Article 2555. https://doi.org/10.3390/ani14172555

Pertanika J. Sci. & Technol. 34 (4): 2255 - 2273 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Framework for Secure Software Engineering in Hybrid Agile Software Development

Normi Sham Awang Abu Bakar [1]*, Norzariyah Yahya[1], and Siti Sarah Maidin[2]

[1]Department of Computer Science, Faculty of ICT, International Islamic University Malaysia, 53100 Gombak, Kuala Lumpur, Malaysia

[2]Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia

ABSTRACT

This study examines how Secure Software Engineering (SSE) practices can be applied while developing with Hybrid Agile methods. The Hybrid Agile approach combines disciplined planning and extensive documentation, which are the hallmarks of traditional software development, with flexibility, incremental delivery and quick adoption for changing requirements contributed by Agile practices. Generally, existing secure software engineering frameworks only support either classical or Agile methods. As a result, Hybrid Agile implementations rarely provide enough guidance to integrate security throughout the full development life cycle. The purpose of this research is to investigate the current utilisation of secure software engineering practices in Hybrid Agile to identify and introduce security-focused coding elements designed for those environments. In line with the qualitative research approach being used, we evaluate current secure software practices in Hybrid Agile project implementation. In this study, interviews with professional programmers in diverse business sectors were conducted. It is evident that there is a lack of standardised security practices and that it is crucial for managers and project managers to encourage and support these. In addition, it is important that this study highlights the importance of automation tools in secure software engineering practices. With these findings in mind, this paper provides a structured approach for these organisations seeking to integrate security into Hybrid Agile projects and for improving the reliability and integrity of their software applications. The shift towards proactive environments in hybrid software applications requires an assessment of their software development life cycle to deal with security weaknesses and build tools for detecting possible threats in their applications.

Keywords: Hybrid agile, qualitative analysis, secure software engineering, Software Development Lifecycle (SDLC), software security framework

ARTICLE INFO

Article history:
Received: 17 November 2025
Accepted: 12 June 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.14

E-mail addresses:
nsham@iium.edu.my (Normi Sham Awang Abu Bakar)
norzariyah@iium.edu.my (Norzariyah Yahya)
sitisarah.maidin@newinti.edu.my (Siti Sarah Maidin)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

In the contemporary business environment, there are always attempts by business structures to conceptualise and implement strategies or techniques of producing the highest output and quality, as well as improving the response to changes within the marketplace. The Hybrid Agile approach hinges on traditional project management practices and considers the flexible spirit of Agile methodologies. Such an approach ensures well-rounded software development by applying the ability to shift emphasis and improve, which lies within the Agile approach.

In conventional methods, the security measures are very strict, but their implementation is performed stage by stage, causing latency in the detection of the problem and, ultimately, lower agility towards the threats as well. Moreover, the emphasis on the implementation of various security updates is performed throughout the Agile iteration process, which enables earlier detection and rapid mitigation of security threats. The implementation requires specialised expertise and continuous security monitoring. The new approaches combine secure protocols within an Agile process, such as secure coding and continuous integration of security, but the existing Agile protection mechanisms are unable to fulfil the demands of Hybrid Agile as well.

Hybrid Agile methodologies, which combine the careful documentation strategy with flexibility, are useful for risk management and adaptability, but for execution purposes, need proper synchronisation and communication for managing the complexity (ISO/IEC, 2008; Prenner et al., 2021; Theocharis et al., 2015). In previous studies on secure software development methodologies, there have been developments made specifically for the context of Agile methodologies (Kagombe et al., 2021; Rindell et al., 2018). These secure-aggregated developments safeguard the combined execution of Agile software development methodologies, such as secure development methodologies, vulnerability management, and long-lasting protection strategies, but do not adequately focus on the unique challenges posed by the combination of Hybrid Agile methodologies for development with the options of conventional development methodologies. This leads to adaptability, but with structured and formalised process activities like Requirement Elicitation and System Design, which are comparatively more formalised than those in Agile methodologies.

Therefore, security systems aligned only with Agile may be inefficient for the broader need of Hybrid Agile, considering its inclusion of traditional procedural security requirements. Specifically, three critical research gaps can be identified in the extant literature. First, existing SSE frameworks such as SSE-CMM (ISO/IEC, 2008) and SAMM (Rindell et al., 2018) were conceived for either Waterfall-style plan-driven environments or pure Agile iterations; neither addresses the coexistence of formal phase gates (e.g., requirements elicitation, high-level architecture) with iterative sprint cycles that characterise Hybrid Agile. Second, the empirical understanding of how practising software developers

perceive, adopt, or resist security measures within Hybrid Agile contexts is severely limited, with only a handful of studies (Ding et al., 2020; Maidin & Yahya, 2024) exploring this domain, and none grounding their findings in primary data gathered through direct interviews with industry professionals. Third, the role of organisational and managerial factors such as policy enforcement, security training, and cross-functional collaboration in determining the success or failure of SSE integration in Hybrid Agile projects remains largely unexplored, despite the recognised importance of such socio-technical dimensions in technology adoption (Wang et al., 2024). Filling this research gap is paramount to ensure seamless integration of safe practices within Hybrid Agile strategies for optimised use of dependable and secure software outcomes. This feeds into the importance of research in safe coding practices.

Against this backdrop, the major contributions of this study are highlighted as follows:

- This study presents the first empirical investigation of Secure Software Engineering (SSE) practices specifically within the Hybrid Agile development context, addressing a gap that prior frameworks targeting either traditional or pure Agile environments have left unfilled.
- Through semi-structured interviews with ten experienced software developers across diverse Malaysian industries, this study identifies and documents the real-world challenges, tool preferences, and management practices that shape SSE adoption in Hybrid Agile projects.
- A Secure Hybrid Agile Framework is proposed that integrates security practices across all phases of the Hybrid Agile lifecycle from planning and estimation through iterative sprint development to final integration and system testing, offering practitioners actionable guidance for building secure software within this blended methodology.
- The study highlights key enablers of secure development in Hybrid Agile settings, including management enforcement of security policies, developer participation in the design of security procedures, deployment of automated security tools, and targeted security training, findings that carry direct implications for software engineering managers and project leads.

RELATED WORK

This section explores the related work as guidelines for the framework development.

Traditional Models

A traditional software development process model, known as the Waterfall Model, follows a linear and sequential methodology in which each stage of the process must be completed before proceeding to the next (Prenner et al., 2020). Projects with clearly defined

requirements and those where changes are unlikely to occur during development are best suited for the Waterfall Model. It is also suitable for smaller projects or those for which a precise deadline is necessary (Tell et al., 2019).

The Verification and Validation Model, or V-Model, is an expansion of the conventional Waterfall Model. The significance of validation and verification tasks at every level of the development lifecycle is emphasised, and a V-shaped process is created, with a testing phase for every development phase (Prenner et al., 2020). The V-Model is very helpful for projects where quality and dependability are very important, and the requirements are clear and consistent. The V-Model is suitable for projects in the aerospace, medical, and automotive sectors, among others, that have stringent regulatory and compliance standards (Tell et al., 2019).

Agile Models

Agile development methodology is characterised by its flexibility and inventiveness. It is a tactic for thriving in the chaos and uncertainty that software development projects frequently bring. Agile processes typically use evolutionary and iterative approaches to software development, which reduce the development cycle and improve the final product's quality (Asthana et al., 2012). Agile, as a lightweight process paradigm, emphasises cooperation and communication amongst all stakeholders. The main goals of agile development are to reduce overall costs, shorten development cycles, and improve software quality (Maier et al., 2017). Agile promotes evolutionary and iterative methods of development, which reduce the time needed to produce and deliver a product. In this work, we primarily concentrate on the Scrum model, one of the Agile models.

One of the popular agile methods for developing software is Scrum (Schwaber & Sutherland, 2016). A product owner, a Scrum master, and a development team make up a Scrum team. The project's goals, specifications, and release schedule are specified by the product owner. Every specified functional and non-functional requirement is kept in a product backlog. The Scrum master facilitates communication between the product owner and the development team and assists the team in adhering to Scrum practices. The goal of a development team is to implement the software; this group should ideally consist of three to nine members.

Hybrid Agile Models

The Waterfall and V-Model are examples of traditional software development models that provide a robust framework for long-term planning. However, these models lack the capability for rapid feedback, which can make it harder to avoid mistakes because the product is usually only delivered to the customer at the end of the project, which makes fixing them expensive. Also, plan-based methods have a hard time dealing with changes in requirements, which happen a lot in development projects.

On the other hand, Agile methods were designed to address these shortcomings through continuous deployment and software testing. However, because Agile focuses on detailed planning for the next iteration only, it does not facilitate long-term planning or effectively manage large projects (Schwaber & Sutherland, 2016).

To overcome the limitations of both traditional and agile methods, organisations often adopt a blended strategy known as hybrid development approaches or hybrid approaches. According to Kuhrmann et al. (2017), a hybrid approach is defined as any combination of agile methodologies with traditional development approaches, such as plan-driven development approaches, that an organisational unit adopts and customises to its own context needs, such as application domain, culture, process, project, organisational structure, techniques, or technologies.

Hybrid Agile offers a tailored solution that aligns with diverse project requirements and organisational contexts. It addresses the limitations of purely Agile or Waterfall methodologies by integrating their strengths, thereby providing a more adaptable and comprehensive framework (Bick et al., 2018). This hybrid approach enables teams to maintain a clear vision and strategic direction while fostering collaboration, innovation, and rapid response to change.

Several publications have already been written about the study of hybrid approaches. Theocharis et al. (2015) SLR examined the application of plan-based and agile methodologies. They reported a widespread application of mixed approaches for those techniques. Considering their conclusions, the HELENA study was carried out, in which practitioners were questioned regarding their approach and practice use. The study was addressed to practitioners, and questions about their methods and practice use were asked. The study was published in 2017 and resulted in 1467 data points with 691 complete answers (Kuhrmann et al., 2017). Noll and Beecham (2019) analysed the data and found that in 66 per cent of the cases the Hybrid Agile approach was used.

Prenner (2020) investigated Hybrid Agile and found that methods and frameworks like Scrum, Iterative Development, Kanban, Waterfall, and DevOps are often used to create combinations of methods. The common Hybrid Agile model is depicted in Figure 1.

Due to the extensive adoption of the Hybrid Agile method in the industry (Noll and Beecham, 2019), we argue that security in the Hybrid Agile development method deserves as much attention as the pure Agile method, and our proposed framework contributes to the implementation of Secure Software Engineering in the Hybrid Agile development process.

Secure Software Engineering (SSE)

Secure Software Engineering (SSE) integrates a set of systems engineering practices and activities into the software development process. Khan et al. (2022) did a systematic mapping study on the Security Procedures in SSE. They listed a few groups of security metrics, standards and tools being used in the SSE area.

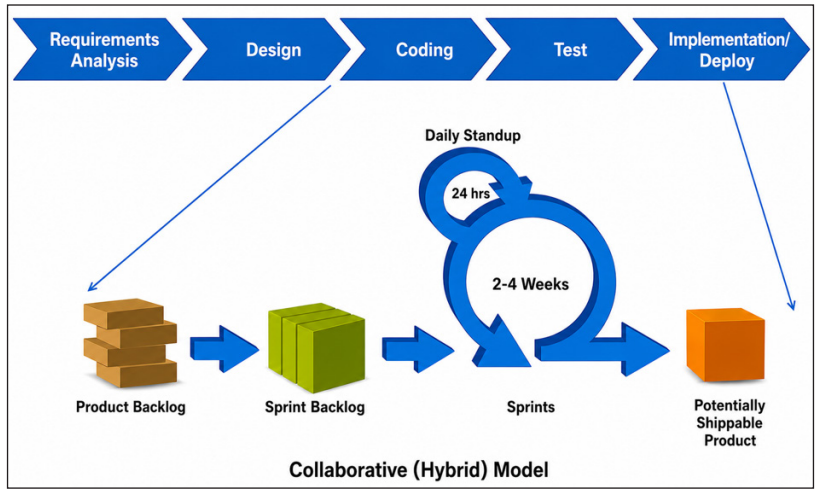


Figure 1. Hybrid agile model (Prenner, 2021)

In general, the security issues are addressed by two development models (International Organisation for Standardisation, 2008): (1) The ISO/IEC's highly process-oriented security management, metrics, and implementation framework is called the Secure Software Engineering Capability Maturity Model (SSE-CMM) (ISO/IEC, 2008). The Capability Maturity Model Integration, created by Carnegie Mellon University and now maintained by the CMMI Institute (Synopsys Software Integrity Group, 2017), is the model's source. Applying the CMM-based model as a process model can be highly expensive (Diaz et al., 2009), and its potential applications go beyond security enhancement. (2) The Software Assurance Maturity Model (SAMM), developed by the Open Web Application Security Project (OWASP) and licensed under an open-source licence, includes development-time activities as well as best practices for security governance, construction, verification, and operations. The SAMM has taken the place of the Comprehensive, Lightweight Application Security Process (CLASP), a model that was specifically developed for software development and was previously maintained by OWASP (Rindell et al., 2018). Additionally, there are clear parallels between SAMM and the Building Security in Maturity Model (BSIMM) (Synopsys Software Integrity Group, 2017). The BSIMM, an annually published survey of prevailing best practices in Secure Software Engineering, does not aim to define specific security models or frameworks. Rather, it serves as a descriptive benchmark that reflects the current state of information security practices across the industry.

The work of Kagombe et al. (2021) proposed an Agile Security framework based on the SSE-CMM's security engineering model and features three areas: security risk analysis, security engineering and assurance.

Most of the work in Secure Software Engineering (SSE) focuses on pure Agile development methodology, where many publications report on aligning the Agile

development activities with security practices (Rindell et al., 2017; Siiskonen et al., 2014; Tøndel et al., 2022; Türpe & Poller, 2017; Van der Heijden et al., 2018; Voggenreiter & Schöpp, 2023). However, for SSE implementation in Hybrid Agile development, the publications are still limited (Ding et al., 2020; Maidin & Yahya, 2024). Therefore, our work contributes to the investigation of the implementation of SSE in Hybrid Agile development.

This study distinguishes itself from earlier work in several important respects. Existing studies on SSE (e.g., Alshareef et al., 2024; Khan et al., 2022; Kagombe et al., 2021; Rindell et al., 2021; Thool & Brown, 2024) focus almost exclusively on pure Agile environments such as Scrum or Kanban and propose frameworks or guidelines that assume a fully iterative development model. They do not account for the structural duality of Hybrid Agile, namely, the coexistence of plan-driven phases (formal requirements elicitation, architectural design) with sprint-based incremental delivery. The work of Ding et al. (2020) and Maidin and Yahya (2024) comes closest to this study's concern, but both rely on conceptual analysis rather than primary empirical data. This study fills that gap by grounding its framework in interview-based evidence drawn from practitioners actively working within Hybrid Agile environments. Furthermore, unlike studies that focus solely on technical security mechanisms (e.g., vulnerability assessment tools, OWASP compliance), this study also captures the socio-technical and organisational dimensions of security adoption, including management enforcement, developer involvement, and training elements whose importance has been highlighted in the technology acceptance literature (Wang et al., 2024) but rarely examined in the SSE domain.

METHOD

The methodology adopted in this research is qualitative in nature. Qualitative analysis is a method of assessing non-numerical data to understand concepts, opinions, or experiences (Delahunt & Maguire, 2017). This type of analysis is often used to gain insights into people's behaviours, motivations, and attitudes. Unlike quantitative analysis, which focuses on numbers and statistical relationships, qualitative analysis emphasises understanding the qualities and characteristics of phenomena. Qualitative analysis involves interpreting data based on the researcher's perspective, often leading to a more subjective understanding of the data. One of the data collection methods used for qualitative analysis is an interview with the identified respondents.

Instrument Development

To address the research question and research objective, a set of questions was designed. The questions are included in an interview protocol that was specifically designed to facilitate the interview while adhering to the appropriate and standardised data collection procedures. The protocol was divided into three sections, consisting of the introduction

of the research project, followed by the demographic data of the respondents and the questions. The interview protocol was pilot tested by conducting a preliminary interview with a beginner software developer participating in hybrid agile projects. The input obtained from the pilot interview is being utilised for the actual data collection.

Data Collection

Based on the identified research questions, the data collection process was implemented using semi-structured interviews, where recruitment was done by distributing the call for participants through the network of former students at the university, together with their associated network. After doing the filtering based on the software development methodology applied in their company, 10 companies were selected as the participants in the study. All selected participants are implementing the Hybrid Agile development methodology to complete their projects.

Ten (10) experienced software developers representing a range of sectors were interviewed in person. All companies are in Kuala Lumpur, Malaysia, and range from medium to large companies. Each participant has over five years of professional experience as a programmer or software developer, demonstrating their expertise in the field. Each interview lasted anywhere from thirty minutes to an hour. Every interview was audio-recorded, and then transcriptions were made. The list of interview questions is provided in Table 1.

Since this study employs a qualitative research design, the notion of dataset balance differs from that applicable to quantitative or machine learning studies. Nonetheless, deliberate measures were taken to ensure representational balance across several dimensions. First, participant selection intentionally covered a variety of industry sectors (including finance, healthcare, retail, and technology), ensuring that the findings were not unduly shaped by the security culture or practices of any single domain. Second, both medium-sized and large companies were included to capture variation in organisational maturity and resource availability for security implementation. Third, all participants were required to have a minimum of five years of professional experience and to be actively working within a Hybrid Agile development environment, thereby reducing the risk of responses that reflect either inexperience or purely traditional or purely Agile contexts. Saturation was monitored during the thematic analysis process; the recurring themes across the ten interviews suggest that data saturation was largely achieved within this sample, though future studies with larger and more geographically diverse samples would further strengthen generalisability.

Data Analysis

The results of the interview were analysed using the six-step theme analysis framework (Delahunt & Maguire, 2017). Finding the themes and using them to answer the research

question or problem is the aim of thematic analysis. A thematic analysis provides insight into the elaborated details of the data.

Table 1
List of interview questions

	Interview Questions
1	There are many types of development models, like Waterfall, Scrum or approaches like prototype or eXtreme Programming. Could you tell me what development framework/ model/approach/technique you implement in a software project?
2	Is there any difference between the model adopted for small-scale projects and a large-scale project?
3	In a software project, a software engineering team will follow certain development phases like requirements gathering, analysis and design, coding, testing, and project deployment. Could you please explain the development phases [mention the model that the respondent mentioned in the above questions]?
4	To what extent does your Software Engineering team execute the [mention the models that the interviewee implements] in a software project?
5	Which project development phase is crucial to execute?
6	Which development phase is important, will it be design, coding or testing?
7	Do your priorities change when a deadline approaches?
8	If the answer is YES, ask this question How are the changes executed and implemented?
9	If the answer is NO, ask this question: Why are the priorities not being changed, although the deadline is approaching?
10	I am going to mention three terms. The terms are: a) Secure Software Development Lifecycle, b) Open Web Application Security Project (OWASP) and c) Threat Model Do any of those three terms sound familiar to you?
11	If the answer is YES, ask these questions:
12	Do you use the [mention the term] in your software development project?
13	How does the implementation take place?
14	Besides the approach that you mentioned, is there any other security model/ framework/ approach/ technique/ tool that you execute in your project?
15	If the answer is NO, ask this question: Could you please explain the security measures that you and your development team execute in a software project?
16	In your opinion, are the above security measures that you implement sufficient for a software project? If the answer is YES, no question to ask.
17	If the answer is NO, ask this question: Could you please suggest any methods for improvement?
18	What is your opinion on Hybrid Agile?
19	What are the Strengths?
20	What are the Weaknesses?

The gathered data is interpreted and analysed through a thematic analysis. The utilisation of the primary interview questions as themes is a common thematic analysis drawback (Delahunt & Maguire, 2017). This illustrates the fact that rather than being analysed, the data have been arranged and condensed.

Data analysis for interviews involves several steps to systematically evaluate and interpret the qualitative data collected. The goal is to derive meaningful patterns, themes, and insights from the interview transcripts or recordings.

The information obtained from the conducted interviews is provided in the next section. Participants in the study were interviewed to learn about the perspectives of software developers on secure software development and hybrid software development methodologies. The six steps are as follows: familiarise yourself with the data, create preliminary codes, look for themes, evaluate themes, define themes, and write up. It can repeatedly go between the steps and back, especially when handling complex data.

Three-stage Coding Phase

The interview is observed with a structured three-stage coding phase as follows:

Preliminary coding: Transcripts were read repeatedly to achieve a thorough understanding of the feedback. Open coding was executed at the sentence level by assigning descriptive labels to text segments. For example, comments regarding the project manager checking security rules periodically were coded under operational enforcement, while remarks detailing rushing code to meet active sprint timelines were coded as temporal boundary pressures.

Searching & Evaluating Themes: Coding was used to cluster interrelated open codes into cohesive categories based on shared structural properties. These categories mapped out specific operational, psychological, and organisational dimensions within the hybrid environment.

Defining Themes & Documentation: Selective coding abstracted these categories into our seven high-level themes. To enhance the credibility and practical relevance of these themes, they were systematically cross-referenced with and justified by established Secure Software Engineering (SSE) literature and architectural frameworks, the Open Web Application Security Project (OWASP) governance standards and the Building Security in Maturity Model (BSIMM) software security framework. The theme coding matrix and the literature grounding for the seven themes are shown in Table 2.

RESULTS DISCUSSION

The interview questions were divided into several themes to capture the different dimensions of the required information. The list of themes is given in Figure 2.

Table 2
Analytical coding matrix and literature grounding for the 7 derived themes

Derived Theme	Target Code Focus	Key Transcript Keywords	Supporting Literature / Framework Justification
Theme 1: Protocol/ Process	Evaluation of standardised security policies, compliance baselines, and formal checklists across hybrid cycles.	Policy, checklist, guidelines, standard, process	OWASP SAMM Governance & Construction Domains: Justified by industry standards evaluating how compliance mandates shape engineering behaviour.
Theme 2: Manager/ Supervisor	The role of operational leadership in monitoring compliance, administering briefings, and enforcing guidelines.	Enforce, audit, supervisor, manager, checkpoint	Security Enforcement Models (Al-Amin et al., 2018): Tracks the impact of managerial intervention and sanction policies on practice adoption.
Theme 3: Automate/ Manual Tools	The practical application and developer preference for automated testing hooks vs labour-intensive reviews.	Automated tool, plugin, scan, manual check, plugin	BSIMM SSDL Touchpoints: Maps code analysis, static scanning, and automated verification touchpoints within fast iterations (Montuban, 2024).
Theme 4: Individual Developer Priority	Developer task ownership and cognitive focus when balancing feature completion timelines against secure coding rules.	Workload, delivery, shortcut, blind spot, feature rush	Cognitive Blind Spots Framework (Oliveira et al., 2014): Analyses why rapid developmental speeds naturally displace non-functional security goals.
Theme 5: Client/User	The external pressure from clients focusing strictly on functional requirements at the cost of software resilience.	Functionality, client demand, system feature, visibility	Agile Manifesto Customer Collaboration Axis: Examines the tension between visible functional value and invisible secure logic. (Agile Alliances, 2019).
Theme 6: Training	The availability and effectiveness of ongoing structured training programs to address technical security gaps.	Security education, baseline training, capability, awareness	Microsoft SDL Core Practice 1 (Security Training): Validates the baseline training prerequisite required for engineers before sprint allocation (Microsoft Corporation, 2019).
Theme 7: Group/Team	Cross-team alignment, communication channels, and collaborative resource sharing across modular sprint units.	Cross-team, collaboration, support unit, security team	Scrum Framework Synergies (Schwaber & Sutherland, 2020): Evaluates how cross-functional and multidisciplinary teams coordinate complex system dependencies

For Theme 1, the findings show that for most of the companies, the implementation of secure programming in the hybrid development method depends on the company's policies, and there are no common practices identified among the companies involved.

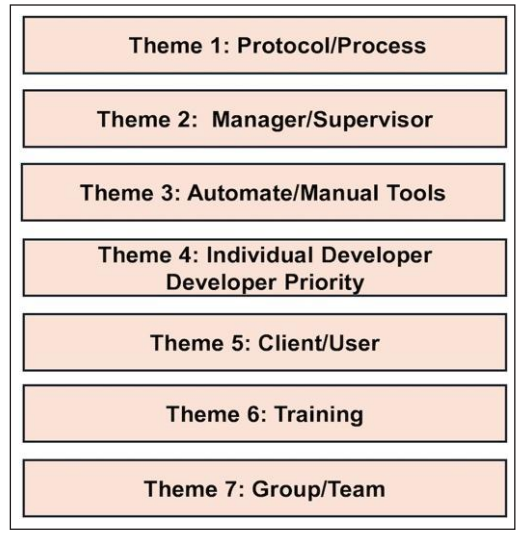


Figure 2. List of interview themes

Furthermore, for Theme 2, most of the interviewees expressed uncertainty regarding the Secure Software Engineering methods and practices that should be implemented or that have already been applied. In this context, the enforcement role of managers and supervisors becomes particularly crucial. They are responsible for ensuring that the Hybrid Agile methodology and Secure Software Engineering practices are properly implemented and adhered to, while also providing security briefings and training to junior developers.

For Theme 3, both automated and manual tools may contribute to the implementation of secure software practices in development. The developers prefer the use of automated tools, which ease their jobs.

For Theme 4, developers' active involvement in the design of security procedures and protocols is crucial, as it significantly influences the adoption and consistency of secure software development practices. When developers contribute to the creation of security methods and frameworks, they are more likely to perceive security as an integral part of their professional role and responsibility.

For Theme 5, the clients do not see security as important; they are mainly concerned about the functionality of the system.

For Theme 6, effective training reinforces security practices, standards, and policies, while also aligning software security requirements with insights derived from emerging technical capabilities and newly available data.

Lastly, for Theme 7, the participants consistently highlighted that cross-team collaboration is vital to the success of secure software development. They acknowledged receiving support from development peers, specialised units such as support, network, and security teams, and strong backing from top management.

In one question, the participants were asked about the Security Model being used in their company, and their responses are shown in Table 3.

Most of the companies use the Open Web Application Security Project (OWASP), and half of the participants responded that they use a variety of methods, including code security review, penetration testing, vulnerability assessment, and security monitoring. and security requirement specifications.

The participants were asked, "How does the implementation take place?" We found that most of the companies have security checklists for testing procedures and reporting. Several developers reported that the managers would go through the checklist periodically to ensure that the proper security practices were adhered to by the team members.

In addition, we asked the question about other additional security model/frameworks/approaches/techniques/tools that they use in their project. The answers are shown in Table 4.

The results obtained from the study support our hypothesis that Secure Software Engineering practices are implemented in the Hybrid Agile process.

Table 3
Security models used in company

Security Model	#
Secure Software Development Lifecycle	1
Open Web Application Security Project (OWASP)	4
Threat Model	0
Others	5

Table 4
Other security practices

#	Answers
1	Code Security Review
2	Penetration Testing
3	Vulnerability Assessment
4	Security Monitoring
5	Firewall
6	Intrusion Detection Systems
7	Access Control List

Secure Hybrid Agile Framework

Based on the answers from the respondents in the interview, we have developed the Secure Hybrid Agile Framework, as shown in Figure 3.

The development process begins with the planning phase, which is overseen by a project manager. During this phase, a kick-off meeting is held to discuss the project, and tasks are divided by modules. This phase also involves initial communication and agreement with the client to ensure alignment on objectives and expectations.

Following the planning phase is the estimation phase. During this stage, the team decides what the new product will be, acquires requirements, and decides which tasks are most important. This makes sure that everyone understands the project's goals and that tasks are grouped by how important and urgent they are. At this point, the planning for the security implementation will happen, and, later, it will be translated into the design and development stage of the software development.

The high-level design or architecture phase is the last phase in this early stage of development. In this phase, the team works on the system's high-level design or architecture.

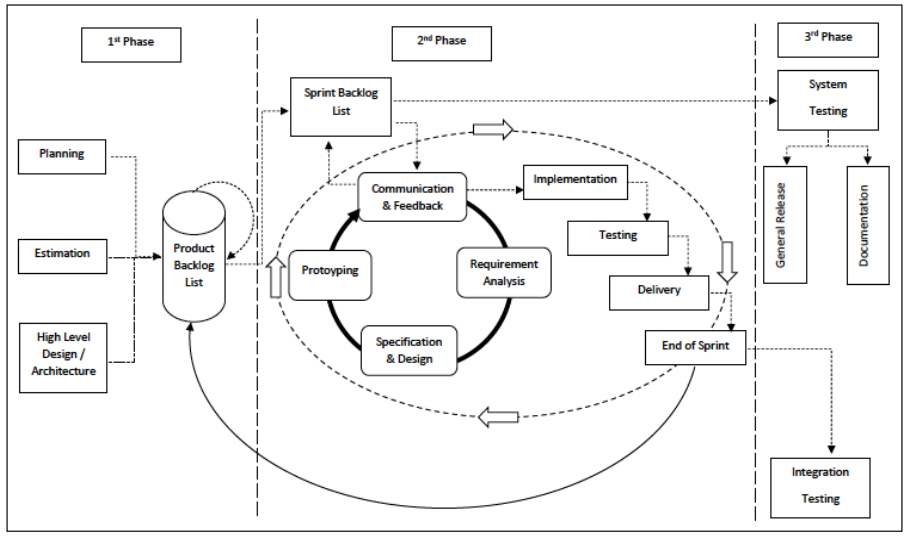


Figure 3. Secure hybrid agile framework

They find the quality traits and units of measurement that the design needs to meet certain standards. They also work on User Interface (UI) and User Experience (UX) design, implementing the requirements from the product backlog list to create an intuitive and effective user experience.

In the middle stage, the development work is broken down into sprints. The work is carried out iteratively and then sequentially. In the sprint, there are activities done iteratively, including communication, requirement analysis, specification, design, and prototyping activities, until the customer or client comes to a consensus with the developer. In the sprint stage, there are sprint planning meetings, daily scrum, peer review, effort and cost update, and addition or removal of the enhancement depending on the need.

After the agreement by the customer or client, the rest of the development work, including implementation, testing, and the delivery of the new increment, takes place in an ordered manner, like the Waterfall model approach. In fact, this stage covers front-end and back-end programming to implement the new features. At the end of the sprint cycle, the newly produced increment will then be integrated with the existing system that already exists. The end of this phase will then see the team hold a sprint review meeting, where the team will review the work completed and plan the next sprint.

The third phase is the final phase, where integration testing is conducted after the completion of each sprint. The main objective is to make sure that the new pieces of functionality integrate perfectly with the existing application. System testing comes next after integration testing is finished. This includes careful testing and release procedures to verify the stability and usability of the product. Additionally, security testing is done to find possible weaknesses and holes in the system. The product is prepared for widespread

release following system testing and after the required updates and changes have been made to the documentation.

Lessons Learned

Generally, Secure Security Engineering in the Hybrid Agile approach research area still needs further exploration (Ding et al., 2020; Maidin & Yahya, 2024). To increase the software's security, more research is required to determine how well security integration techniques work in the development process. The Secure Hybrid Agile framework's suggested methodology places a strong emphasis on incorporating security elements into software development and planning during the initial stages of the software lifecycle (Gill et al., 2018; Imani & Nakano, 2018; Prenner et al., 2021).

According to the interview, one of the factors contributing to the integration of security into the development lifecycle is security training. The developers stated that gaining knowledge about security-related issues in systems enhances their understanding of the significance of security components in the system's development process. This aligns with the results indicated by Rindell et al. (2018), where the Agile developers gain knowledge from the various security training provided before the start of the project.

Another important motivation for security inclusion in the development lifecycle is the enforcement of security policies by the project manager to ensure that secure software practices are followed and executed accordingly. This finding was also highlighted by Tøndel et al. (2022) and Voggenreiter and Schöpp (2023), where the security champions in the organisation should enforce the security guidelines and communicate clearly to the project members.

Threats to validity

Generally, the results from the interview have their own threats to validity.

- Internal validity: The interviews were conducted by 2 interviewers. Therefore, the interview style may contribute to the participants' responses to the questions. We tried to minimise the threat by preparing a common transcript for the interview.
- External validity: The number of respondents is limited (10). This could mean the results cannot be generalised outside the research environment. However, all respondents are experienced developers; thus, their feedback is valuable regarding this topic.

In addition to the threats above, several further limitations of this study must be acknowledged. First, the sample is restricted to ten companies, all based in Kuala Lumpur, Malaysia. While this geographic focus provides contextual coherence, it limits the transferability of findings to organisations in other national or cultural contexts where software development practices, security regulation, and organisational norms may differ

substantially. Second, the qualitative, interview-based design, although well-suited to exploratory inquiry, means that the findings cannot be statistically generalised to the broader population of Hybrid Agile organisations. Third, the study relies on self-reported perceptions of security practice; there is no independent verification of whether the stated practices were implemented as described, which introduces potential social desirability bias.

To mitigate this external validity limitation and strengthen the credibility, generalizability, and practical relevance of the study, a domain expert validation process was introduced. Rather than relying on a single perspective, the proposed Secure Hybrid Agile Framework was evaluated by an independent panel of domain experts, software engineers with over 10 years of experience managing large-scale hybrid architectures. These experts evaluated the framework for process disruption, sprint velocity impacts, and developer workflow feasibility.

CONCLUSION

In summary, this research work deals with the highlighted integration of Secure Software Engineering processes within Hybrid Agile methodologies. The findings of this research work reveal gaps in security processes and stress the key roles of enforcement by management, participation and support by developers, automation tools, and overall training for improving secure software development in Hybrid Agile settings.

The Secure Hybrid Agile Framework provides a structured approach for incorporating security at every stage of the Hybrid Agile phase, which aims at ensuring that security is taken into consideration at every stage of development, right from planning to deployment. The framework mentioned above has its own objective of making sure that vulnerabilities are removed right from the start, during the software development stage.

By offering even more insight into the opportunities and challenges posed by the Hybrid Agile method of software development, this work further improves the security and resilience of the software development environment. Future research should validate and refine the proposed framework to ensure it remains pliable and responsive to the rapidly changing environment presented by software development security. Future work can involve extending this study to a wider audience and examining security behaviours presented by the participants.

To effectively address the empirical evidence demonstrating the efficacy of our framework, future research will be conducted to prove the framework within actual software projects.

ACKNOWLEDGEMENT

This research is fully funded by the Fundamental Research Grant Scheme, grant number FRGS/1/2019/ICT01/UIAM/03/2.

REFERENCES

Agile Alliances. (2019). *12 principles behind the Agile Manifesto.*

Al-Amin, S., Ajmeri, N., Du, H., Berglund, E. Z., & Singh, M. P. (2018). Toward effective adoption of secure software development practices. *Simulation Modelling Practice and Theory, 85*, 33-46.

Alshareef, R., Alshabeeb, E., Alakkas, N., & Niazi, M. (2024). Challenges in developing secure software within agile environments. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE 2024)* (pp. 1-10). Association for Computing Machinery. https://doi.org/10.1145/3661167.3661284

Asthana, V., Tarandach, I., O'Donoghue, N., Sullivan, B., & Saario, M. (2012). *Practical security stories and security tasks for agile development environments.* SAFECode.

Bick, S., Spohrer, K., Hoda, R., Scheerer, A., & Heinzl, A. (2018). Coordination challenges in large-scale software development: A case study of planning misalignment in hybrid settings. *IEEE Transactions on Software Engineering, 44*(10), 932-950. https://doi.org/10.1109/TSE.2017.2730870

Delahunt, B., & Maguire, M. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Teaching and Learning in Higher Education, 8*(3), 3351-3359. https://doi.org/10.5281/zenodo.5828312

Ding, C. B., Maidin, S. S., Medi, I., & Nghiem, T. L. (2020). Secure software implementation in hybrid agile development approach. *International Journal of Management, 11*(10), 1713-1721. https://doi.org/10.34218/IJM.11.2020.157

Gill, A. Q., Henderson-Sellers, B., & Niazi, M. (2018). Scaling for agility: A reference model for hybrid traditional-agile software development methodologies. *Information Systems Frontiers, 20*(2), 315-341. https://doi.org/10.1007/s10796-016-9672-8

Imani, T., & Nakano, M. (2018). A model for effective area of hybrid approach combining agile and plan-driven methods in IT projects. In *Proceedings of the International Association for P2M, Spring* (pp. 49-64). https://doi.org/10.20702/iappmproc.2018.Spring.0_49

International Organisation for Standardisation. (2008). *Information technology-Security Techniques-Systems security engineering -Capability maturity model (SSE-CMM) (ISO/IEC 21827).* https://doi.org/10.3403/30143286

Kagombe, G. G., Mwangi, R. W., & Wafula, J. M. (2021). Achieving standard software security in agile developments. In *Proceedings of the 11th International Conference on Information and Communication Management (ICICM '21)* (pp. 24-33). Association for Computing Machinery. https://doi.org/10.1145/3484399.3484403

Khan, R. A., Khan, S. U., & Ilyas, M. (2022). Exploring security procedures in secure software engineering: A systematic mapping study. In *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering (EASE '22)* (pp. 433-439). Association for Computing Machinery. https://doi.org/10.1145/3530019.3531336

Kuhrmann, M., et al. (2017). Hybrid software and system development in practice: Waterfall, Scrum, and beyond. In *Proceedings of the 2017 International Conference on Software and System Processes (ICSSP 2017)* (pp. 30-39). Association for Computing Machinery. https://doi.org/10.1145/3084100.3084104

Maidin, S. S., & Yahya, N. (2024). An insight into hybrid agile software development approach and software security among software engineers: A critical evaluation. *Journal of Theoretical and Applied Information Technology*, 450-459.

Maier, P., Ma, Z., & Bloem, R. (2017). Towards a secure SCRUM process for agile web application development. In *Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17)*. Association for Computing Machinery. https://doi.org/10.1145/3098954.3103171

Microsoft Corporation. (2019). *Microsoft Security Development Lifecycle*. Microsoft. https://www.microsoft.com/en-us/securityengineering/sdl

Montuban, N. (2024). *BSIMM (Building Security in Maturity Model): A complete guide.* Codific. https://codific.com/bsimm-building-security-in-maturity-model-a-complete-guide/

Noll, J., & Beecham, S. (2019). How agile is hybrid agile? An analysis of the HELENA data. In *PROFES 2019* (pp. 341–349). Springer. https://doi.org/10.1007/978-3-030-35333-9_25

Oliveira, D., Rosenthal, M., Morin, N., Yeh, K. C., Cappos, J., & Zhuang, Y. (2014). It's stupid psychology: How heuristics explain software vulnerabilities and how priming can illuminate developers' blind spots. In *Proceedings of the 30th Annual Computer Security Applications Conference* (pp. 296-305).

Prenner, N. (2020). Towards improving the organisation of hybrid development approaches. In *Proceedings of the International Conference on Software and System Processes (ICSSP '20)* (pp. 185-188). Association for Computing Machinery. https://doi.org/10.1145/3379177.3390304

Prenner, N., Unger-Windeler, C., & Schneider, K. (2020). How are hybrid development approaches organised? A systematic literature review. In *Proceedings of the International Conference on Software and System Processes (ICSSP '20)* (pp. 145-154). Association for Computing Machinery. https://doi.org/10.1145/3379177.3388907

Prenner, N., Unger-Windeler, C., & Schneider, K. (2021). Goals and challenges in hybrid software development approaches. *Journal of Software: Evolution and Process, 33*(11), Article e2382. https://doi.org/10.1002/smr.2382

Rindell, K., Hyrynsalmi, S., & Leppänen, V. (2017). Busting a myth: Review of agile security engineering methods. In *Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES '17)* (pp. 74:1-74:10). Association for Computing Machinery. https://doi.org/10.1145/3098954.3103170

Rindell, K., Hyrynsalmi, S., & Leppänen, V. (2018). Aligning security objectives with agile software development. In *Proceedings of the 19th International Conference on Agile Software Development: Companion (XP '18)* (pp. 3:1-3:9). Association for Computing Machinery. https://doi.org/10.1145/3234152.3234187

Rindell, K., Ruohonen, J., Holvitie, J., Hyrynsalmi, S., & Leppänen, V. (2021). Security in agile software development: A practitioner survey. *Information and Software Technology, 131*, Article 106488. https://doi.org/10.1016/j.infsof.2020.106488

Schwaber, K., & Sutherland, J. (2020). *The Scrum guide*. https://www.scrum.org/resources/scrum-guide

Siiskonen, T., Sars, C., Vähä-Sipilä, A., & Pietikäinen, A. (2014). Generic security user stories. In P. Pietikäinen & J. Röning (Eds.), *Handbook of the secure agile software development life cycle* (pp. 9-14). University of Oulu.

Synopsys Software Integrity Group. (2017). *The building security maturity model*. Synopsys.

Tell, P., Klünder, J., Küpper, S., Raffo, D., MacDonell, S. G., Münch, J., Pfahl, D., Linssen, O., & Kuhrmann, M. (2019). What are hybrid development methods made of? An evidence-based characterisation. In *Proceedings of the International Conference on Software and System Processes (ICSSP '19)* (pp. 105-114). IEEE. https://doi.org/10.1109/ICSSP.2019.00022

Theocharis, G., Kuhrmann, M., Münch, J., & Diebold, P. (2015). Is Water Scrum-Fall reality? On the use of agile and traditional development practices. In *Product-focused software process improvement* (pp. 149-166). Springer. https://doi.org/10.1007/978-3-319-26844-6_11

Thool, A., & Brown, C. (2024). Securing agile: Assessing the impact of security activities on agile development. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE 2024)* (pp. 1-11). Association for Computing Machinery. https://doi.org/10.1145/3661167.3661280

Tøndel, I. A., Cruzes, D. S., Jaatun, M. G., & Sindre, G. (2022). Influencing the security prioritisation of an agile software development project. *Computers & Security, 118*, Article 102744. https://doi.org/10.1016/j.cose.2022.102744

Türpe, S., & Poller, A. (2017). Managing security work in Scrum: Tensions and challenges. In *Proceedings of the International Workshop on Secure Software Engineering in DevOps and Agile Development (SecSE 2017)* (pp. 34-49). CEUR Workshop Proceedings.

Van der Heijden, A., Broasca, C., & Serebrenik, A. (2018). An empirical perspective on security challenges in large-scale agile software development. In *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18)* (Article 45). Association for Computing Machinery. https://doi.org/10.1145/3239235.3267426

Voggenreiter, M., & Schöpp, U. (2023). Prioritising industrial security findings in agile software development projects. In *Proceedings of the 45th International Conference on Software Engineering: Companion Proceedings (ICSE '23)* (pp. 375-379). IEEE Press. https://doi.org/10.1109/ICSE-Companion58688.2023.00106

Wang, Y., Hung, J. C., Huan, C. H., Hussain, S., Yen, N., & Jin, Q. (2024). Design of TAM-based framework for credibility and trend analysis in sharing economy: Behavioural intention and user experience on Airbnb as an instance. *Computer Science and Information Systems, 21*(2), 547-568. https://doi.org/10.2298/CSIS230323010W

Pertanika J. Sci. & Technol. 34 (4): 2275 - 2291 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Evaluation of the Colour Fastness Properties of *Bixa orellana*-Dyed Silk Using CIELab Analysis

Nur Atiqah Jamaludin[1], Razanah Ramya[1]*, and Rashidi Othman[2]

[1]Institute of the Malay World and Civilization (ATMA), Universiti Kebangsaan Malaysia (UKM), 43600 UKM Bangi, Selangor, Malaysia

[2]Herbarium Unit, Department of Landscape Architecture, Kulliyyah of Architecture and Environmental Design, International Islamic University Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia

ABSTRACT

The application of plant-derived dyes in textiles is often limited by inadequate colour stability, particularly under ultraviolet exposure. This study evaluated the effects of four mordants: potassium aluminium sulphate $KAl(SO_4)_2 \cdot 12H_2O$), copper (II) sulphate ($CuSO_4$), iron (II) sulphate $FeSO_4$ and sodium metasilicate (Na_2SiO_3) on the colour stability of silk dyed with bixin extracted from *Bixa orellana* seeds. Bixin concentrations of 10, 20, 30 g/L were applied to silk using the post-mordanting method, and the treated specimens were evaluated under dark, visible light and UVA exposure conditions. Colour stability was assessed using CIELab parameters, including yellow intensity (b*), colour saturation (s) and total colour difference (ΔEab), supported by Pearson correlation and principal component analysis. Under dark conditions, $KAl(SO_4)_2 \cdot 12H_2O$, combined with 30 g/L bixin, produced the lowest degradation, recording 27.53% in b* and 43.05% in colour saturation. Under UVA exposure, $FeSO_4$ combined with 20 g/L bixin demonstrated the best relative performance, with degradation values of 72.70% for b* and 82.38% for colour saturation. In contrast, Na_2SiO_3 exhibited the poorest overall UVA stability, with degradation reaching 98.91% for b* and 98.64% for colour saturation. Nevertheless, ΔEab values ranging from 36.02% to 57.04% after seven days indicated substantial perceptible colour changes across all treatments. The dataset was suitable for multivariate analysis (KMO = 0.783; Bartlett's $X^2 = 7363.286$, df = 10, $p < 0.0001$), while the first two principal components explained 94.87% of the total variance. Overall, mordant effectiveness was dependent on the exposure condition; $KAl(SO_4)_2 \cdot 12H_2O$ was more effective

ARTICLE INFO

Article history:
Received: 30 October 2025
Accepted: 03 August 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.15

E-mail addresses:
p172005@siswa.ukm.edu.my(Nur Atiqah Jamaludin)
razanah.ramya@ukm.edu.my (Razanah Ramya)
rashidi@iium.edu.my (Rashidi Othman)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

under dark storage, whereas $FeSO_4$ provided comparatively better protection against IVA induced degradation. However, the high degradation and ΔEab values demonstrate that *Bixa Orellana* dyed silk remains unsuitable for prolonged UVA exposure without additional UV-protective finishing or improved stabilisation strategies.

Keywords: Bixa orellana, bixin pigment, CIELab analysis, colour stability, mordants, natural dyes, silk, textile colouration, UVA degradation

INTRODUCTION

Bixa orellana has gained attention as a potential natural dye because it produces bright, biodegradable, and non-toxic colours, making it suitable for environmentally responsible dyeing practices (Cuong & Chin, 2016). However, maintaining colour stability under challenging environmental conditions, especially prolonged exposure to UV radiation, remains a significant challenge. The results of this study illustrate that selecting the appropriate mordant is essential to achieving better colour retention while minimising the ecological footprint of the dyeing process. The data also reveal a saturation effect, where higher dye concentrations produce a more uniform and more protective colour layer that resists fading more effectively. However, the use of large amounts of dye carries financial and environmental implications, indicating the need for carefully optimised dyeing strategies, especially in industrial-scale applications (Kalapriya & Gurumalesh Prabu, 2020).

Known for its lustre and luxurious handle, silk is highly receptive to natural dyes due to its protein-based fibre structure (Lee et al., 2010). In most natural dye systems, colour fastness is enhanced by mordants, which help to attach dye molecules to the fabric and strengthen their resistance to degradation (Swamy, 2020). Increasing environmental awareness has encouraged the exploration of more sustainable mordant practices that can enhance dye performance without causing undue ecological damage (Burkinshaw & Kumar, 2009; Vankar, 2017).

In this study, the colour fastness of silk dyed with *B. orellana* was evaluated under exposure to light, complete darkness, and UVA radiation. CIELab measurements were used to track the colour behaviour under each condition. The investigation focused on the effects of potassium alum ($KAl(SO_4)_2$), iron (II) sulphate ($FeSO_4$), copper (II) sulphate ($CuSO_4$), and sodium metasilicate (Na_2SiO_3) as mordants on the resulting colour parameters. Overall, this study aims to provide practical insights that support the development of more sustainable and effective natural dye applications for silk textiles. It is hoped that the findings contribute practical insights into sustainable and effective natural dye application on silk textiles.

MATERIALS AND METHODS

The *Bixa orellana* seeds used in this study were obtained from the Authentic Vietnamese Market, Vietnam, while the silk was sourced from Gulati's Silk House, Kuala Lumpur, Malaysia. The chemicals used were sodium hydroxide (NaOH), ethyl acetate, potassium aluminium sulphate ($KAl(SO_4)_2 \cdot 12H_2O$), iron(II) sulphate ($FeSO_4$), copper (II) sulphate ($CuSO_4$), sodium metasilicate (Na_2SiO_3), and distilled water. The principal equipment comprised a centrifuge, rotary evaporator, glass slides, fluorescent-light source, UVA source and CIELab colourimeter.

Extraction and Preparation of Dye Concentration

5 grams of powdered *Bixa orellana* seeds were mixed with 30 mL of NaOH and maintained at room temperature for 30 minutes. The pigment was subsequently extracted with 30 of ethyl acetate. The extraction was repeated two to three times until no further colour was transferred into the solvent. The combined ethyl acetate fractions were centrifuged at 8000 to10 000 rpm for 5 minutes and concentrated using a rotary evaporator at 40 °C, 75-100 rpm qnd 300 mbqr to obtain a slurry-like extract. Aliquots of 0.5, 1.0 and 1.5 g of the concentrated extract were separately diluted in 50 mL ethyl acetate to prepare dye concentrations of 10, 20 and 30 g/L, respectively.

Fabric Preparation

The silk fabric was cut into standardised strips measuring 1 cm x 5 cm. The specimens were labelled according to dye concentration, mordant treatment and exposure condition. Three replicate specimens were prepared for each treatment.

Mordant Preparation and Dyeing Procedure

Four mordants were evaluated: $KAl(SO_4)_2 \cdot 12H_2O$, $FeSO_4$, $CuSO_4$, and Na_2SiO_3. The potassium aluminium sulphate solution was prepared by dissolving 70 g of the compound in 1 litre of distilled water, stirring for 5 minutes and allowing the solution to clarify. The $FeSO_4$, $CuSO_4$, and Na_2SiO_3 solutions were prepared separately by dissolving 5 g of each compound in 1 litre of distilled water to obtain homogeneous solutions. While the silk specimens were immersed separately in the 10, 20 and 30 g/L dye solutions at room temperature for 60 minutes with complete submersion to promote uniform dye uptake. Following dyeing, the specimens were treated with one of the prepared mordant solutions using a post-mordanting technique. The concentrations of 10, 20 and 30 g/L therefore refer to the bixin dye solutions rather than the mordant solutions. Unmordanted specimens were retained as controls (Mongkholrattanasit et al., 2013; Sarker et al., 2020).

Dark and Light Exposure of the Dyed Fabric

For the light and dark storage test, the dyed specimens were affixed to glass slides and maintained at room temperature for four weeks. Dark storage specimens were protected completely from light, whereas light-exposed specimens were placed 30 cm from a fluorescent light source at an illuminance of 3,100 lux. CIELab measurements were recorded at baseline and on days 1, 7, 14, 21 and 30 using a CS-10 digital colourimeter (manufactured in China using a Japanese-made sensor)

UVA Exposure of the Dyed Fabric

The dyed specimens were mounted on glass slides and positioned upright at approximately 10 cm from the UVA source with a wavelength of 365 nm. The recorded illuminance at the specimen position was 5,500 lux, and the experimental setup is shown in Figure 1. The specimens were exposed at room temperature for seven consecutive days and evaluated daily. All treatments were tested in triplicate, and CIELab readings were recorded before exposure and after each daily exposure interval using a CS-10 digital colourimeter (manufactured in China using a Japanese-made sensor)

Colour Stability Evaluation

Colour stability was evaluated using the CIE 1976 L* a* b* colour system. The L* coordinate represents lightness from black (0) to white (100); positive and negative a* values represent red and green directions, respectively. The spatial relationships among the L*, a* and b* coordinates are illustrated in Figure 2. Colour saturation (s), yellow

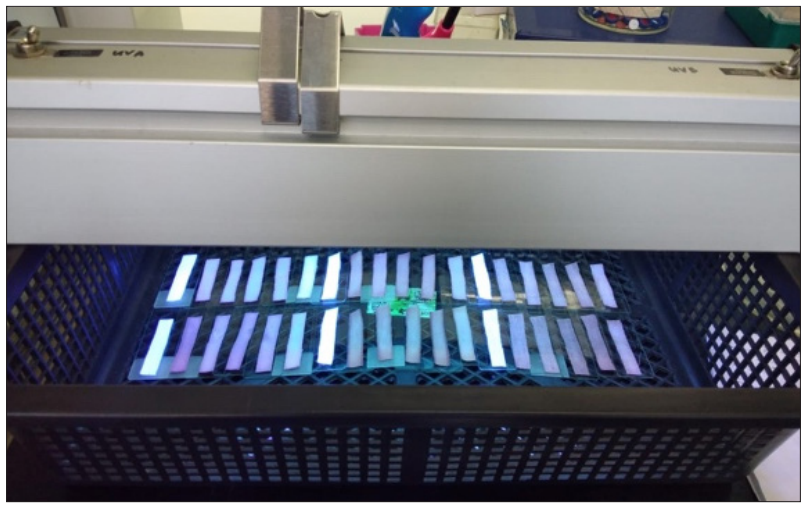


Figure 1. Experimental setup for UVA exposure of Bixa Orellana-dyed silk specimens

intensity (b*) and total colour difference (ΔEab) were used to assess colour retention and degradation (Othman, 2009). $\Delta E^*ab = \sqrt{[(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]}$. Where ΔL*, Δa* and Δb* represent the differences between the initial and post-exposure colour coordinates. In this study, colour difference was used to assess the stability of the pigment particles. A ΔE* value < 1.0 indicates that no perceptible colour difference is observed, whereas values < 2.0 indicate only a slight difference. Values < 3.5 represent a noticeable colour difference, while values < 5.0 indicate a clearly perceptible difference. A ΔE* value > 5.0 signifies that the two colours are distinctly different (Latos-Brozio & Masek 2020; Melgosa et al., 2013). Degradation (%) = $[(X_0 - X_t)/X_0] \times 100$, where X_0 is the initial b* or s value, and X_t is the corresponding value after the specified exposure period. The reported values were calculated from the mean of three replicate specimens.

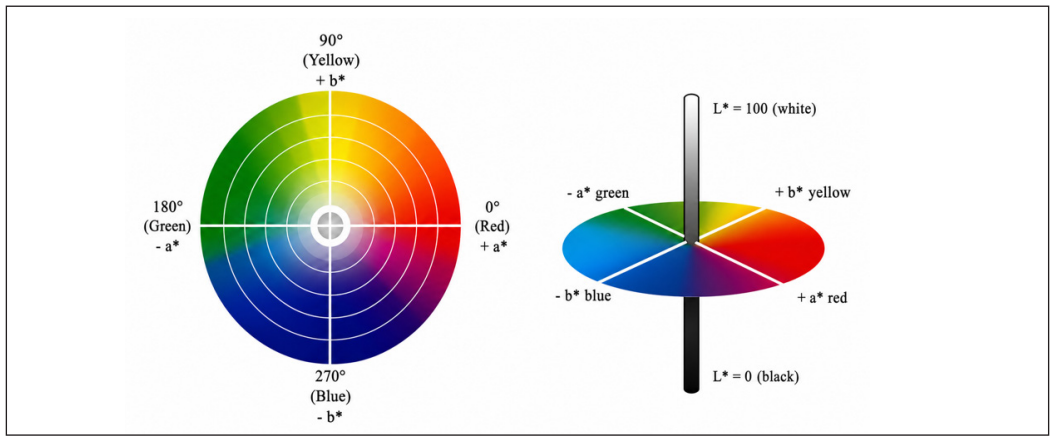


Figure 2. CIELab colour space (Earl et al. 2016)

Statistical Data Analysis

The triplicate colour measurements were summarised using the mean, standard deviation and variance. Data from the light and dark tests were evaluated at the scheduled weekly intervals, whereas the UVA exposure data were evaluated daily. The statistical analyses were conducted using XLSTAT- Basic + 2024 (Addinsoft/Lumivero, France). The dataset was normalised, and potential outliers were assessed at $\alpha = 0.05$ using the Grubbs and Dixon tests (Coucke et al., 2012). Differences between treatment means were evaluated by analysis of variance (ANOVA), followed by Tukey's post-hoc test at $\alpha = 0.05$. Where required, the data were transformed using centring, standardisation, scaling, Pareto or logarithmic procedures before the normality and sampling adequacy assessments were repeated. Sampling adequacy for principal component analysis (PCA) was evaluated using the Kaiser-Meyer-Olkin (KMO) test. KMO values were interpreted as inadequate (<0.5), mediocre (0.5-0.7), good (0.7-0.8), very good (0.8-0.9) and excellent (>0.9); only values

exceeding 0.5 were accepted for PCA (Ismail et al., 2021). Bartlett's test of sphericity was used to determine whether the correlation matrix was suitable for factor analysis. Pearson correlation analysis was then applied to evaluate the relationships among dye concentration, b*, s, ΔE*ab and exposure time. PCA was used to reduce the dataset into principal components, and the cumulative variance and factor loadings of PC1 and PC2 were examined to identify the variables that contributed most strongly to the colour changes.

RESULT AND DISCUSSIONS

Analysis of Mordants on Colour Stability Under Light and Dark Conditions

Figure 3 and Table 1 compare the percentage degradation of the yellow coordinate (b*) and saturation (s), together with the final colour difference (ΔE*ab), for silk dyed with bixin at 10, 20 and 30 g/L after four weeks under visible-light and dark-storage conditions. Under visible light, all treatments showed substantial losses in b* and saturation (s).

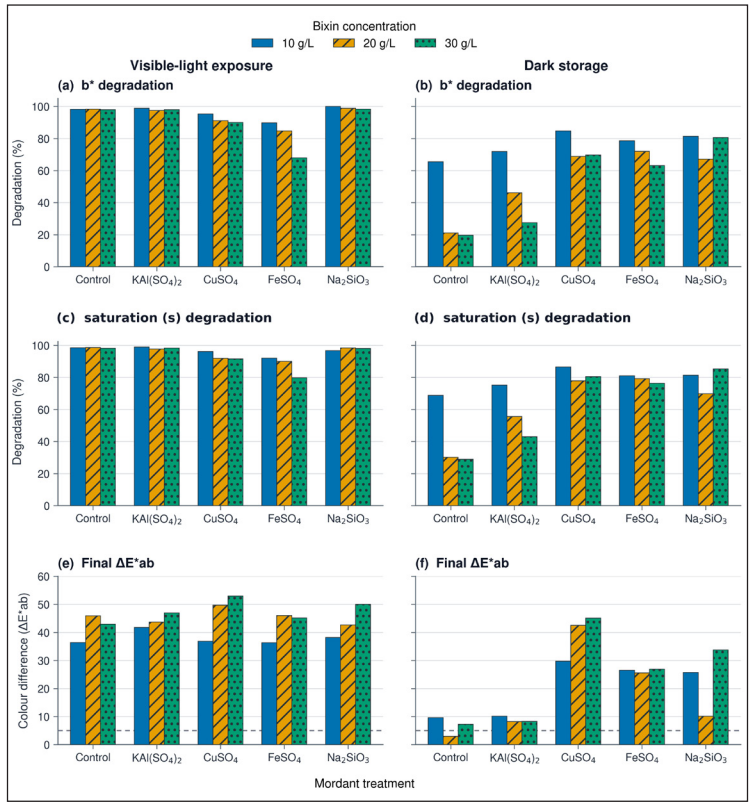


Figure 3. Degradation of b* and saturation (s) and final ΔE*ab for *Bixa orellana* dyed silk at pigment concentrations of 10, 20 and 30 g/L after four weeks: (a0 b* under visible light, (b) b* in dark storage, (c) saturation (s) under visible light, (d) saturation (s) in dark storage, (e) ΔE*ab under visible light and (f) ΔE*ab in dark storage. The dashed line in panels (e) and (f) marks ΔE*ab = 5.0

Table 1
Degradation of b and saturation (s) and final ΔE*ab for bixin dyed silk after four weeks under visible light and dark storage*

Colour response	Treatment	Value / Pigment concentration (g/L)					
		Visible light			Dark storage		
		10	20	30	10	20	30
Saturation degradation, s (%)	Control (unmordanted)	98.53	98.61	98.26	68.88	30.13	28.94
	$KAl(SO_4)_2$	98.94	97.81	98.29	75.21	55.66	43.05
	$CuSO_4$	96.18	91.96	91.62	86.51	77.91	80.53
	$FeSO_4$	92.06	90.05	79.82	81.05	79.29	76.39
	Na_2SiO_3	96.80	98.46	98.12	81.37	69.90	85.30
Yellow-coordinate degradation, b* (%)	Control (unmordanted)	98.25	98.30	98.00	65.49	21.13	19.63
	$KAl(SO_4)_2$	98.85	97.54	97.95	71.89	46.15	27.53
	$CuSO_4$	95.34	91.21	90.08	84.74	68.88	69.72
	$FeSO_4$	89.82	84.77	67.99	78.56	71.97	63.13
	Na_2SiO_3	100.00	98.84	98.29	81.43	67.19	80.66
Colour difference, ΔE*ab	Control (unmordanted)	36.39	45.93	42.90	9.59	2.97	7.21
	$KAl(SO_4)_2$	41.75	43.71	46.93	10.13	8.19	8.25
	$CuSO_4$	36.84	49.68	52.95	29.71	42.55	45.12
	$FeSO_4$	36.33	45.98	45.16	26.49	25.57	26.90
	Na_2SiO_3	38.18	42.66	50.05	25.69	10.11	33.76

Note. Degradation (%) = [initial value – final value] x 100 for b* and saturation (s). ΔE*ab is an absolute colour difference value, not a percentage

Among the mordanted samples, $FeSO_4$ at 30 g/L produced the lowest degradation (67.99% for b* and 79.82% for saturation (s)), whereas Na_2SiO_3 produced consistently high losses, including 100.00% degradation of b* at 10 g/L.

Under dark storage, degradation was lower than under visible light for most treatments. Among the mordanted samples, $KAl(SO_4)_2 \cdot 12H_2O$ at 30 g/L recorded the lowest degradation (27.53% for b* and 43.05% for saturation (s)). The control at the same concentration showed lower degradation (19.63% for b* and 28.94% for saturation (s)). Therefore, the advantage of $KAl(SO_4)_2 \cdot 12H_2O$ applies only within the mordanted treatments.

At week 4, ΔE*ab values under visible light ranged from 36.33 ($FeSO_4$, 10 g/L) to 52.95 ($CuSO_4$, 30 g/L). As all values exceeded 5.0, the final colours were clearly distinguishable from their initial colours. These high values indicate a substantial and rapid loss of the original pigment colour, reflecting poor colour stability under visible light exposure. Under dark storage, ΔE*ab ranged from 2.97 (control, 20 g/L) to 45.12 ($CuSO_4$, 30 g/L). All mordanted samples showed distinct colour changes (ΔE*ab = 5.0), indicating considerable pigment colour degradation, whereas the control value 2.97 represented only a noticeable colour difference.

Analysis of UVA Exposure on Colour Stability of Silk Fabric

Figure 4 and Table 2 present the endpoint results following the recorded 24-hour UVA exposure and daily colour monitoring through Day 7. Substantial degradation was observed across all treatments. Among the mordanted samples, $FeSO_4$ at 20 g/L showed the lowest degradation, recording 72.70% for b* and 82.38% for saturation (s). In contrast, Na_2SiO_3 exhibited very high degradation, reaching 98.91% for b* at 10 g/L and 98.64% for saturation (s) at 30 g/L.

On Day 7, ΔE*ab values ranged from 36.02 to 57.04. $FeSO_4$ recorded the lowest ΔE*ab at each pigment concentration with values of 36.02, 44.96 and 50.31 at 10, 20 and 30 g/L, respectively, whereas the highest value of 57.04 was recorded for $KAl(SO_4)_2 \cdot 12H_2O$ at 30 g/L. All endpoint E*ab values exceeded 5.0, demonstrating clearly distinguishable colour changes and substantial pigment fading. Although $FeSO_4$ at 20 g/L provided the greatest relative colour retention under UVA exposure, degradation remained above 70% for b and 80% for saturation (s), indicating that none of the mordants completely prevented UVA induced colour deterioration.

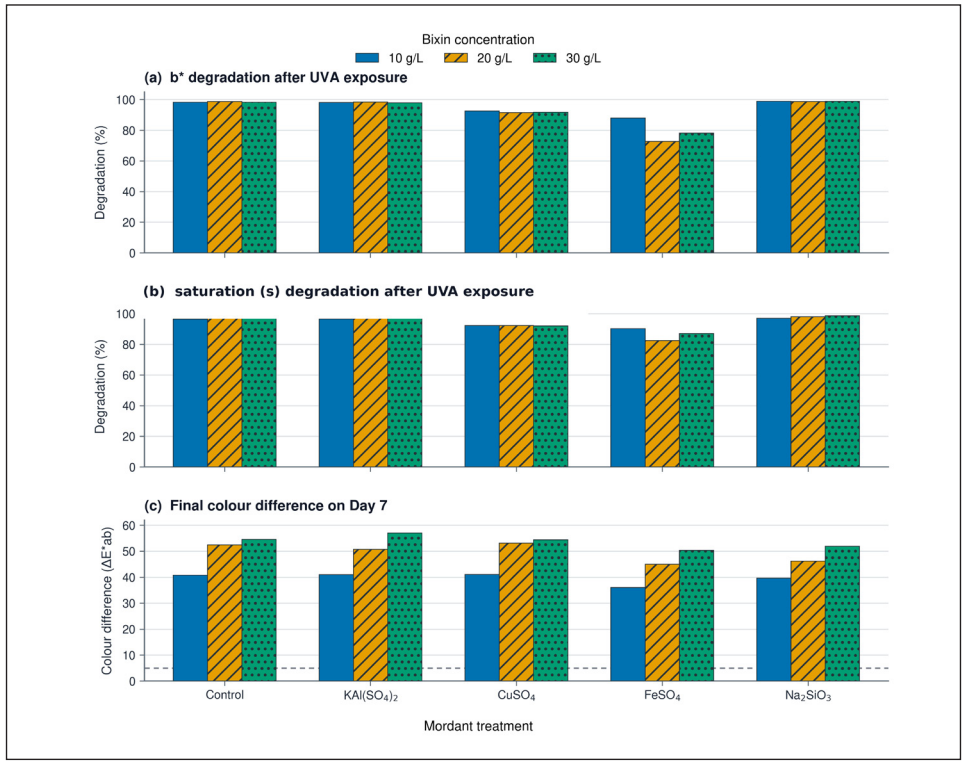


Figure 4. Degradation of b* and saturation (s) and final ΔE*ab for *Bixa orellana* dyed silk monitored through Day & following the recorded 27-hour UVA exposure: (a) b* degradation, (b) saturation (s) degradation and (c) ΔE*ab on Day 7. The dashed line in panel (c) marks ΔE*ab = 5.0

Table 2
Degradation of b and saturation (s) and final ΔE*ab for bixin dyed silk monitored through Day & following UVA exposure*

		Value / Pigment concentration (g/L)		
Colour response	**Treatment**	**10**	**20**	**30**
Saturation degradation, s (%)	Control (unmordanted)	96.48	98.68	98.33
	$KAl(SO_4)_2$	96.63	98.29	98.25
	$CuSO_4$	92.18	92.21	91.95
	$FeSO_4$	90.20	82.38	87.05
	Na_2SiO_3	96.99	97.98	98.64
Yellow-coordinate degradation, b* (%)	Control (unmordanted)	98.27	98.79	98.24
	$KAl(SO_4)_2$	98.08	98.30	97.80
	$CuSO_4$	92.44	91.51	91.77
	$FeSO_4$	87.85	72.70	78.09
	Na_2SiO_3	98.91	98.74	98.84
Colour difference, ΔE*ab	Control (unmordanted)	40.74	52.41	54.53
	$KAl(SO_4)_2$	40.96	50.69	57.04
	$CuSO_4$	41.07	53.08	54.40
	$FeSO_4$	36.02	44.96	50.31
	Na_2SiO_3	39.63	46.19	51.86

Note. The specimens were recorded in 24 hours of UVA exposure followed by daily colour monitoring. Degradation (%) was calculated for b* and saturation (s); ΔE*ab is an absolute colour difference value

Multivariate Data Analysis

The factors affecting the stability of bixin pigments on silk textiles and chromaticity changes were assessed using multivariate data analysis, or MDA. The adequacy of the sample factor analysis was confirmed by the Kaiser-Meyer-Olkin (KMO) test, which yielded a value of 0.783, exceeding the minimum threshold of 0.5, and Bartlett's test of sphericity was significant ($X2 = 7363.286$, $df = 10$, $p < 0.0001$), confirming that the correlation matrix was suitable for PCA (Ismail et al., 2021) as shown in Table 3.

The Pearson matrix showed a strong positive correlation between b* and saturation (s) ($r = 0.979$). ΔE*ab was strongly negatively correlated with b* ($r = 0.891$) and saturation (s)

Table 3
Kaiser-Meyer-Olkin Test and Bartlett's Test of Sphericity for Chromaticity changes of bixin pigment on fabric

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.783
Bartlett's Test of Sphericity	Chi-square (Observed value)	7363.286
	DF	10
	p-value (Two-tailed)	< 0.0001

(r = -0.883). Exposure time was negatively correlated with b* (r = -0.903) and saturation (s) (r = -0.889) but positively correlated with ΔE*ab (r = 0.869). Thus, increasing exposure time was associated with loss of yellowness and saturation and with a larger overall colour difference as shown in Table 4.

Principal component analysis (PCA) retained two components with eigenvalues greater than 1. PC1 explained 74.24% of the variance and represented the colour retention versus degradation axis; b* (0.981) and saturation (s) (0.977) loaded positively, whereas ΔE*ab (-0.943) and exposure time (-0.949) loaded negatively. PC2 explained 20.64% of the variance and was dominated by pigment concentration (loading = 0.995). Together, PC1 and PC2 explained 94.87% of the total variance, as shown in Tables 5 and 6. Because the analysis was pooled across three fabrics, the components should not be interpreted as silk-specific effects.

DISCUSSION

This study demonstrates the potential of *Bixa orellana* (annatto) as a natural colourant for silk, supporting more sustainable and environmentally conscious textile practices. The CIELab system was used to examine how the dyed silk responded under different

Table 4
Pearson correlation matrix of chromaticity changes of bixin pigment on fabric

Variables	CONC	b	S	ΔE	t
CONC	1				
b	0.093	1			
S	0.176	0.979	1		
ΔE	0.062	-0.891	-0.883	1	
t	0.000	-0.903	-0.889	0.869	1

Table 5
Eigenvalues, cumulative variance, and factor loading (FL) of different parameters for colour changes of bixin pigment on fabric

Variables	**F1**	**F2**
Conc (concentration)	0.076	0.995
b* (yellow intensity)	0.981	0.029
s (saturation)	0.977	0.113
ΔE (chromaticity)	-0.943	0.148
t (time)	-0.949	0.078
Eigenvalue	3.712	1.032
Variability (%)	74.236	20.638
Cumulative (%)	74.236	94.874

Table 6
Summary of multivariate analysis

Test/Parameter	Value/category	Interpretation
Kaiser-Meyer-Olkin (KMO)	0.783	Data adequate for factor analysis (good category)
Bartlett Test of Sphericity	X^2 =7363.286=10, p<0.00001	A correlation matrix is suitable for PCA
Pearson Correlation	b*-s=0.979 b*- ΔE=0.869 ΔE-t=0.903 s-t=-0.891	Strong positive correlations (b* & s; b* & ΔE); strong negative correlations (ΔE & t; s & t)
Eigenvalue (PC1, PC2)	PC1=3.71 PC2=1.03	Both components exceed the eigenvalue threshold of 1
Cumulative Variance	PC1=74.24% PC2=20.63% Total=94.87%	Two principal components explain almost all data variation
Factor Loadings (FL)	All variables FL>0.75	All variables contribute significantly to colour variations

conditions in both the presence and absence of light and UVA radiation. Silk, a natural protein-based fibre, is highly compatible with plant-derived colours and commonly used in high-quality textiles. These findings serve to clarify how natural dyes interact with silk, which is critical for creating dyeing procedures with less effect on the environment (Che & Yang, 2022; Yusuf et al., 2017).

In dark storage, KAl $(SO_4)_2$ was the most effective mordant in preserving colour stability. This indicates that each mordant works more effectively under different conditions. However, the mordants all showed a clear colour change under light exposure, with ΔE* values exceeding 5.0. Such behaviour implies that the changes were readily noticeable, which highlights the tendency of the mordants to fade when exposed to light. This is a known weakness of many natural dyes (Chakraborty et al., 2025; Ibrahim et al., 2010). In this study, all the mordants showed a noticeable colour change (ΔE* > 5.0) under prolonged light exposure, indicating that *Bixa orellana*-dyed silk fabric does not exhibit long-term stability under typical lighting conditions. This result aligns with earlier work showing that natural dyes, especially carotenoid dyes such as bixin, are moderately prone to photo-oxidation and fading when exposed to light (Othman, 2009; Sant'Anna et al., 2013).

The importance of mordants in increasing the durability of colour on textiles of natural dyes is well-documented (Burkinshaw & Kumar, 2009; Vankar, 2017). In this study, $FeSO_4$ was the most effective mordant for maintaining colour stability under both light and UVA exposure, particularly at the highest concentration of 30 g/L. This result aligns with previous studies indicating that iron-based mordants tend to improve colour retention in natural dyeing, possibly due to their ability to form strong bonds with dye molecules and fibres (Mongkholrattanasit et al., 2016; Swamy, 2020). The remarkable efficacy of $FeSO_4$

in this study is likely linked to its ability to stabilise the bixin pigment sourced from *Bixa orellana*, which is renowned for its vulnerability to light and UV radiation (Shahid-ul-Islam et al., 2016). $FeSO_4$ can create more stable dye-fibre-mordant complexes due to its sulphate groups, which can efficiently link the metal centres and dye molecules via robust hydrogen bonds and coordination interactions. Moreover, the redox-active Fe (II) entities can promote dynamic ligand-to-metal charge transfer, enhancing the stability, lightfastness, and dye retention of the complexes relative to other mordants (Inouchi et al., 2026; Roy Maulik, 2019). Although $FeSO_4$ offered slightly greater protection than the other mordants, the aggregate colour loss remained significant and, in many cases, exceeded 70%. This pattern is consistent with previous findings showing the susceptibility of natural dyes, particularly carotenoid pigments such as bixin, to photooxidation under UV light, which ultimately leads to decreased colour stability (Othman, 2009; Sant'Anna et al., 2013).

Na_2SiO_3, in contrast, consistently exhibited the lowest stability in all the tested conditions. This result is consistent with previous research indicating that sodium metasilicate is generally less effective as a mordant for natural dyes due to its poor binding affinity to fibres (Burkinshaw & Kumar, 2009). Exposure to UVA increased the fading effect, and reductions in yellow intensity and saturation were clearly observed across all the mordants tested. The diminished efficacy of sodium metasilicate as a mordant is attributed to its elevated alkalinity and pH-dependent polymerisation, leading to the formation of large aggregates of less reactive silicates that fail to bond with fibres and dyes as effectively as traditional metal mordants. Moreover, its feeble coordination interactions with cellulose hydroxyl groups and the potential for fibre degradation owing to its corrosive nature diminish its effectiveness in dye fixation and colour retention (Jamaludin et al., 2025). The interaction of Na_2SiO_3 with molecules is lower than that of standard metal mordants, leading to insufficient dye fixing and colour fastness. The effectiveness also depends on the ionic strength of the solution and on competing ions (Rahman & Karacan, 2022; Kon'kova et al., 2021).

Since no mordant can completely stop the degradation process, the use of additional treatments, including UV-blocking finishes or the addition of antioxidant agents, could help to increase pigment stability for longer periods. From a practical perspective, these findings suggest that *Bixa orellana* dyes on silk are suitable for indoor or low-light applications, such as scarves, interior upholstery, or ceremonial clothing, where there is limited UV exposure. Besides, the silk proteins, mainly fibroin and sericin, possess reactive groups including hydroxyl, carboxyl, and amine functionalities (Phasayavan et al., 2026). The carboxyl and amine functional groups are essential for binding metal ions, forming stable complexes by coordination bonding. Metal cations such as Mg^{2+}, K^{+} and Ca^{2+} can engage with silk proteins, generally leading to gelation or alterations in structure by coordination with the negatively charged moieties of the protein (Zhou et al., 2025). Metal mordants

primarily interact with the functional groups of silk protein, particularly carboxyl and amino groups, to create stable complexes and alter the structure and properties of silk protein.

The dye's poor resistance to sunlight needs to be improved if it is to be used for outdoor products. One possible solution is to explore the use of natural bio-mordants or encapsulation methods to boost colourfastness while maintaining eco-friendliness (Burkinshaw & Kumar, 2009; Vankar, 2017). The vulnerability of carotenoid pigments like bixin to photooxidation mostly arises from their long chain of conjugated double bonds, rendering them highly reactive and unstable in the presence of light and oxygen; yet this characteristic is advantageous for their colouration and antioxidant functions. This results in rapid degradation, fading of hue, and diminished biological activity (Enayati et al., 2023). Bixin is regarded as the inaugural cis-carotenoid and an apocarotenoid that may modify its stability and light sensitivity. In addition, carotenoids exhibit significant instability when exposed to light and oxygen, leading to photooxidation and the deterioration of the pigments (Cortez et al., 2017).

This study generates important information on the shelf life of annatto at different levels. The findings revealed that a higher concentration (30 g/L) contributed slightly to the stability of the dye, particularly with $FeSO_4$ and $KAl(SO_4)_2$ mordants. The potential of *B. orellana* to generate bright colours, its biodegradability, and its non-toxicity render it a suitable candidate for eco-friendlier approaches to dye production (Cuong & Chin, 2016). Nevertheless, the related issues (from colour fastness to environmental factors, and especially exposure to UV radiation) remain a major challenge. These findings highlight the importance of selecting the right mordant to optimise colour stability while reducing the environmental impact of the dyeing process. Reactive oxygen species (ROS) generated from UVA radiation can oxidise pigment molecules such as melanin and synthetic dyes, resulting in their destruction or alteration (Lawrence et al., 2022). Oxidative stress is a main cause of pigment degradation, because reactive oxygen species can change the chemical structure of the pigments and cause their discolouration or transformation (Bouider et al., 2024). The rate of pigment degradation was greatly increased by the inclusion of photocatalysts (TiO2, SnO2, etc.) under UVA irradiation, which leads to higher production of extra reactive oxygen species (ROS) and acceleration of photodecomposition processes (Bokhari et al., 2022). The data also showed the existence of a saturation effect, whereby the use of higher quantities of dye helped to increase the level of protection on the fabric surface and subsequently reduce the rate of fading. However, these factors must be considered along with aspects such as the cost and environmental implications, thus emphasising the need to optimise dyeing techniques for large-scale production purposes (Kalapriya & Gurumalesh Prabu, 2020).

Although this study lays a solid foundation for using *Bixa orellana* as a textile dye, future studies should include other fastness tests, such as washing, perspiration, and

rubbing, which are important for real-life applications. Dye concentration is a fundamental factor influencing colour intensity in textiles; nevertheless, its impact is influenced by fibre characteristics, fabric composition, and dying parameters. A realistic maximum for colour depth exists due to fibre saturation, and over this threshold can adversely affect product quality and environmental sustainability (Chakraborty et al., 2025; Li et al., 2009). Testing on other fabrics, such as cotton, wool, and polyester, will also help in applying these findings more widely. In addition, the integration of additional protective treatments, such as UV-blocking or antioxidant finishes, could further improve the fastness of natural dyes in outdoor applications. Further studies should also test the fastness to washing and rubbing of *Bixa orellana*-dyed fabrics, with these factors being important for the practical everyday use of textiles (Yusuf et al., 2017). As emphasised by Alegbe & Uthman (2024), testing on a wide range of fabrics is important to understand the full potential of natural dyes in modern textile applications.

CONCLUSION

This study on the colour fastness of silk dyed with *Bixa orellana* (annatto) using CIELab analysis showed that although this dye is biodegradable and has strong colour, it has no long-term stability under environmental stress. Of the mordants tested, $FeSO_4$ gave the best results, especially at a high dye concentration of 30 g/L, with better colour retention under light and UVA exposure. $KAl(SO_4)_2$ performed well in dark storage, while Na_2SiO_3 performed the worst, with high colour loss under all conditions. However, all the mordants showed obvious colour changes ($\Delta E^* > 5.0$), with more than 70% fading under UVA. This indicates that no mordant could completely protect the dye. Thereafter, the use of more dye might slightly improve the short-term stability. Based on the resistance, future studies should test UV blocking treatments or natural bio-mordants, as well as study the dye resistance to washing and rubbing.

Although *Bixa orellana* has potential as a natural dye, it is easily affected by UV radiation and fades when exposed to light. This indicates the need for further improvement in mordanting techniques or the development of a protective layer. Although $FeSO_4$ showed better colour fastness, the overall durability of the dye was still limited, indicating that additional treatments such as UV barrier finishes or the use of natural bio-mordants could improve its performance for practical applications. Although this study provides valuable insights into the colourfastness of silk dyed with *Bixa orellana*, some limitations need to be considered. The scope of this study was limited to exposure to light and UVA radiation, so future research should include other forms of environmental stress, such as washing, sweating, and rubbing, which are important for real-world applications. Furthermore, conducting tests on diverse fabric types, such as cotton, wool, and synthetic materials, would facilitate the wider applicability of the results. Further studies should also explore

the optimisation of the dyeing process, including the use of different extraction methods and concentrations of *Bixa orellana*, to further improve colourfastness and enhance the sustainability of natural dyes in textile manufacturing.

ACKNOWLEDGEMENT

The author extends sincere gratitude to the Geran Galakan Penyelidik Muda (GGPM), Universiti Kebangsaan Malaysia (UKM), for their support through grant number GGPM-2022-061 as well Herbarium Laboratory Unit, Kulliyyah of Architecture and Environmental Design, International Islamic University Malaysia, for the equipment support.

REFERENCES

Alegbe, E. O., & Uthman, T. O. (2024). A review of history, properties, classification, applications, and challenges of natural and synthetic dyes. *Heliyon, 10*(13), Article e33646. https://doi.org/10.1016/j.heliyon.2024.e33646

Bokhari, T. H., Mustafa, G., Ahmed, N., Usman, M., Akram, N., Haq, A. U., Salman, M., Naz, S., Al-Fawzan, F. F., Alissa, S. A., & Iqbal, M. (2022). Degradation of Pigment Red 238 using UV, UV/H_2O_2, UV/H_2O_2/SnO_2 and Fenton processes. *Polish Journal of Environmental Studies, 31*(1), 619–623. https://doi.org/10.15244/pjoes/140273

Bouider, B., Rida, K., Alterkaoui, A., Isik, Z., & Dizge, N. (2024). Development of a new magnetic mesoporous activated carbon for the photocatalytic degradation of Basic Red 18 and Reactive Red 180 dyes under UVA radiation. *Reaction Kinetics, Mechanisms and Catalysis, 137*(4), 2431-2449. https://doi.org/10.1007/s11144-024-02648-4

Burkinshaw, S. M., & Kumar, N. (2009). The mordant dyeing of wool using tannic acid and $FeSO_4$. Part 1: Initial findings. *Dyes and Pigments, 80*(1), 53-60. https://doi.org/10.1016/j.dyepig.2008.05.008

Chakraborty, R., Samanta, A. K., Vankar, P. S., & Chakraborty, S. (2025). Effect of dual mordanting, selective dye concentrations and dye compatibility for minimising variation in natural dyeing of jute fabrics for compound shade matching. *Journal of The Institution of Engineers (India): Series E, 106*(1), 1-27. https://doi.org/10.1007/s40034-024-00292-0

Enayati, A., Assadpour, E., & Jafari, S. M. (2023). Bixin: properties and applications. In Karimi, R., & Rashidinejad, A. (Eds.). *Handbook of Food Bioactive Ingredients: Properties and Applications* (pp. 1-33). Cham: Springer International Publishing.

Ibrahim, N. A., El-Gamal, A. Gouda, M., 7 Mahrous, F. (2010). A new approach for natural dyeing and functional finishing of cotton cellulose. *Carbohydrate Polymers, 82*(4), 1205-1211. https://doi.org/10.1016/j.carbpol.2010.06.054

Inouchi, T., Noguchi, H., Ito, A., & Semba, T. (2026). Dyeing condition-dependent metal complex formation in madder and its derivatives: effects on lightfastness and colour properties. *Journal of Fibre Science and Technology*, 82(2), 43-50. https://doi.org/10.2115/fiberst.2026-0007

Ismail, A. M., Sani, M. S. A.., Azid, A., Zaki, N. N. M., Arshad, S., Tukiran, N. A., Abidin, S. A. S. Z., Samsudin, M. S., & Ismail, A. (2021). Food forensics on gelatine source via ultra-high-performance liquid chromatography diode-array detector and principal component analysis. *SN Applied Sciences, 3*(79), 1-19. https://doi.org/10.1007/s42452-020-04061-7

Jamaludin, N. A., Ramya, R., Othman, R., & Mat Ali, Q. A. (2025). The evaluation of the dyeing performance of *Bixa orellana* (annatto) on cotton and polyester fabrics. In R. Othman, N. Ab Jalil, N. Zainal Abidin, N. Baniyamin, Q. A. Mat Ali, & N. H. Salleh (Eds.), *A design odyssey in the built environment: Functionality, aesthetics and heritage* (pp. 57-71). Nova Science Publishers. https://doi.org/10.52305/EHMP4558

Kalapriya, I., & Gurumallesh Prabu, H. (2020). Eco-friendly dyeing of textile fabric by natural colourants. *AIP Conference Proceedings, 2270*, Article 060008. https://doi.org/10.1063/5.0019468

Kon'kova, T. V., Rysev, A. P., & Mal'kova, Y. O. (2021). Kinetics and mechanism of adsorption of anionic dyes on montmorillonite modified with sodium metasilicate. *Russian Journal of Physical Chemistry A, 95*(1), 43-48. https://doi.org/10.1134/S003602442101012X

Latos-Brozio, M., & Masek, A. (2020). The application of natural food colourants as indicator substances in intelligent biodegradable packaging materials. *Food and Chemical Toxicology, 135*, Article 110975. https://doi.org/10.1016/j.fct.2019.110975

Lawrence, K. P., Sarkany, R. P. E., Acker, S., Herzog, B., & Young, A. R. (2022). A new visible light-absorbing organic filter offers superior protection against pigmentation by wavelengths at the UVR-visible boundary region. *Journal of Photochemistry and Photobiology B: Biology, 227*, Article 112372. https://doi.org/10.1016/j.jphotobiol.2021.112372

Lee, Y.-H., Hwang, E.-K., Jung, Y.-J., Do, S.-K., & Kim, H.-D. (2010). Dyeing and deodorising properties of cotton, silk, and wool fabrics dyed with Amur corktree, *Dryopteris crassirhizoma*, *Chrysanthemum boreale*, and *Artemisia* extracts. *Journal of Applied Polymer Science, 115*(4), 2246-2253. https://doi.org/10.1002/app.31357

Li, S., Shamey, R., & Xu, C. (2009). Prediction of the depth of shade of a dyed polyester fabric based on fibre fineness and fabric structure. *Colouration Technology, 125*(5), 296-303. https://doi.org/10.1111/j.1478-4408.2009.00210.x

Melgosa, M., Trémeau, A., & Cui, G. (2013). Colour difference evaluation. In C. Fernandez-Maloigne (Ed.), *Advanced colour image processing and analysis* (pp. 59-79). Springer. https://doi.org/10.1007/978-1-4419-6190-7_3

Mongkholrattanasit, R., Ariyakuare, K., Limtrakool, T., Saiwan, C., Rungruangkitkrai, N., Punrattanasin, N., Sriharuksa, K., & Nakpathom, M. (2013). An evaluation of silk fabric dyed with lac dye by using the pad-dry technique: Research on the effect of mordant concentration. *Advanced Materials Research, 821-822*, 569-572. https://doi.org/10.4028/www.scientific.net/AMR.821-822.569

Mongkholrattanasit, R., Saiwan, C., Rungruangkitkrai, N., Punrattanasin, N., Sriharuksa, K., & Nakpathom, M. (2016). The effect of alum, ferrous, stannous and NaCl on silk fabric dyed with natural dye from *Laccifer lacca* Kerr. *Applied Mechanics and Materials, 848*, 141-144. https://doi.org/10.4028/www.scientific.net/AMM.848.141

Othman, R. (2009). *Biochemistry and genetics of carotenoid composition in potato tubers* [Doctoral thesis, Lincoln University].

Phasayavan, W., Ross, G. M., Jongjitwimol, J., Pinthong, T., Mahasaranon, S., Charoensit, P., Viyoch, J., Topham, P. D., Tighe, B. J., & Ross, S. (2026). Silk sericin as a functional biopolymer: From polymer chemistry to biomedical translation. *Polymer International*. Advance online publication. https://doi.org/10.1002/pi.70134

Rahman, M. M., & Karacan, I. (2022). Structural and thermal characterisation of chemically pretreated and thermally oxidised bamboo fibre in activated carbon fibre manufacturing. *Journal of Natural Fibres, 19*(16), 15085-15099. https://doi.org/10.1080/15440478.2022.2070324

Roy Maulik, S. (2019). Application of natural dyes on protein fibres following pad–steam methods. *Journal of The Institution of Engineers (India): Series E, 100*(1), 1-9. https://doi.org/10.1007/s40034-019-00141-5

Sant'Anna, V., Gurak, P. D., Ferreira Marczak, L. D., & Tessaro, I. C. (2013). Tracking bioactive compounds with colour changes in foods: A review. *Dyes and Pigments, 98*(3), 601-608. https://doi.org/10.1016/j.dyepig.2013.04.011

Sarker, P., Hosne Asif, A. K. M. A., Rahman, M., Islam, M. M., & Rahman, K. H. (2020). Green dyeing of silk fabric with turmeric powder using tamarind seed coat as a mordant. *Journal of Materials Science and Chemical Engineering, 8*(2), 65-80. https://doi.org/10.4236/msce.2020.82007

Shahid-ul-Islam, Rather, L. J., & Mohammad, F. (2016). Phytochemistry, biological activities, and industrial potential of annatto as a natural colourant: A review. *Journal of Advanced Research, 7*(3), 499-514. https://doi.org/10.1016/j.jare.2015.11.002

Swamy, V. N. (2020). Effect of natural dye from *Bombax malabarica* on silk fabric. *Clothing and Textiles Research Journal, 38*(1), 49-58. https://doi.org/10.1177/0887302X19866261

Vankar, P. S. (2017). Low-temperature optimised dyeing of cotton, wool, and silk with extract of *Camellia sinensis* (tea leaves). *Journal of Textile Engineering & Fashion Technology, 2*(1), 274-280. https://doi.org/10.15406/jteft.2017.02.00045

Yusuf, M., Shabbir, M., & Mohammad, F. (2017). Natural colourants: Historical, processing, and sustainable prospects. *Natural Products and Bioprospecting, 7*(1), 123-145. https://doi.org/10.1007/s13659-017-0119-9

Zhou, Z., Zhang, Y., Yang, Z., Guo, Y., Zhang, B., Li, Z., Huang, K., Zhang, Y., & Du, S. (2025). Biomimetic ion-induced conformation regulation of silk fibroin peptides from wasted silk. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, Article 138441. https://doi.org/10.1016/j.colsurfa.2025.138441

Pertanika J. Sci. & Technol. 34 (4): 2293 - 2310 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Partial Discharge Detection and Localisation in Refined, Bleached and Deodorised Palm Olein Based on Ultra-High Frequency Sensing Technique

Nur Darina Ahmad[1,2], Norhafiz Azis[1]*, Zhi Yang Lim[1], Jasronita Jasni[1], Khairil Anas Md Rezali[3], Mohd Aizam Talib[4], Nik Hakimi Nik Ali[5], and Masahiro Kozako[6]

[1]Advanced Lightning, Power and Energy Research Centre (ALPER), Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Electrical Engineering Studies, Universiti Teknologi MARA Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Seberang Perai, Pulau Pinang, Malaysia
[3]Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[4]Advanced Diagnostic Services, TNB Labs Sdn. Bhd., 43000 Kajang, Selangor, Malaysia
[5]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
[6]Department of Electrical and Electronic Engineering, Kyushu Institute of Technology, 804-8550 Kitakyushu-shi, Fukuoka, Japan

ABSTRACT

This study presents a comparative analysis of Partial Discharge (PD) detection and localisation in Mineral Oil (MO) and Refined, Bleached and Deodorised Palm Olein (RBDPO) based on the Ultra-High Frequency (UHF) technique. A PD experiment was arranged using a needle–plane electrode configuration, and the gap was set to 35 mm. To initiate the PD, the applied voltage was gradually increased from 15 kV to 25 kV in 5 kV/step. In total, 4 patch-type UHF sensors were mounted externally on the steel tank wall at predetermined locations to capture the PD signals. An electrical PD event was detected simultaneously using an Impedance Matching Circuit (IMC). The UHF PD signals were denoised by a Gaussian filtering method. The Time Difference of Arrival (TDOA) was then determined through the First Peak Method (FPM) and Cumulative Energy Method (CEM). PD localisation was carried out based on the TDOAs obtained from both FPM and CEM. It is found that although RBDPO exhibits higher attenuation of UHF PD signals, it achieves lower

ARTICLE INFO

Article history:
Received: 19 March 2026
Accepted: 05 August 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.16

E-mail addresses:
nurdarina9183@uitm.edu.my (Nur Darina Ahmad)
norhafiz@upm.edu.my (Norhafiz Azis)
gs68246@student.upm.edu.my (Zhi Yang Lim)
jas@upm.edu.my (Jasronita Jasni)
khairilanas@upm.edu.my (Khairil Anas Md Rezali)
aizam.talib@tnb.com.my (Mohd Aizam Talib)
hakimiali@uitm.edu.my (Nik Hakimi Nik Ali)
kozako@ele.kyutech.ac.jp (Kozako Masahiro)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

localisation error as compared to MO. Performance assessment further confirms that PD localisation is better in RBDPO than in MO.

Keywords: Cumulative energy method, first peak method, partial discharge localisation, refined, bleached and deodorised palm olein, ultra-high frequency

INTRODUCTION

Partial Discharge (PD) is an electrical occurrence that happens within insulating materials that are subjected to high voltage stress. It is well acknowledged as a preliminary sign of insulation deterioration in high-voltage apparatus such as power transformers. If PD is left unaddressed, it can cause insulation failure, dielectric breakdown, unexpected power outage and costly repair or replacement of the equipment (Chan et al., 2023; Rathod et al., 2021; Thiviyanathan et al., 2022). Therefore, comprehensive detection and localisation of PD are essential for condition-based monitoring and preventative maintenance in electricity systems (Duan et al., 2019).

Mineral Oil (MO) has been the primary insulation and cooling agent in transformers since its inception. However, the extensive use of fossil-based products has raised concerns over environmental pollution and sustainability (Adak & Cangi, 2024). Consequently, researchers have investigated biodegradable dielectric insulating fluids as alternatives to MO. Researchers in (Muangpratoom et al., 2023) examined the electrical and physicochemical properties of various batches of palm oil. The efficacy of acoustic emission-based PD detection in soybean oil was evaluated by Qian et al. (2018). Mohamad et al. (2021) examined the impact of aluminium oxide nanoparticles on the PD properties of palm oil and coconut oil. Meanwhile, Mina-Casaran et al. (2025) and Mohan Rao et al. (2019) examined rapeseed oil and sunflower oil that can be used as alternatives to MO. Among the various potential alternatives, Refined, Bleached and Deodorised Palm Olein (RBDPO) has garnered considerable interest due to its beneficial dielectric and thermal properties. In addition, high biodegradability, low toxicity and wide availability in Southeast Asian countries made RBDPO a more attractive choice as a sustainable dielectric insulating fluid for transformers (Ahmad et al., 2025; Azis et al., 2014; Azmi et al., 2017; Girish et al., 2021; Mina-Casaran et al., 2025; Mohamad et al., 2014).

PD is a rapid electrical phenomenon that produces electromagnetic waves within the Ultra-High Frequency (UHF) band. The UHF sensing technique is widely applied to detect and locate PDs in transformers. It offers a higher Signal-to-Noise Ratio (SNR), better sensitivity and stronger immunity to electromagnetic interference than the conventional electrical PD method (Antony et al., 2021; Chai et al., 2019; Desai & Sarathi, 2018; Jangjoo et al., 2023; Uwiringiyimana et al., 2022). The principle of the technique relies on the detection of electromagnetic signals between 300 MHz and 3 GHz through UHF sensors (Gautam et al., 2023; Gautam et al.,2024).

The UHF PD localisation critically depends on the Time Difference of Arrival (TDOA) computation. The First Peak Method (FPM) and the Cumulative Energy Method (CEM) are commonly used to determine the TDOA based on the PDs. For FPM, the PD signal is converted to unipolar form and normalised to a common amplitude scale. The earliest peak that exceeds a pre-defined threshold is taken as the Time of Arrival (TOA) of the PD signal. This method is less affected by multipath reflections and often provides good output for TDOA computation (Ha et al., 2018; Song, 2004). On the other hand, CEM identifies the TDOA based on the Cumulative Energy (CE) curve of the PD signal. It is suitable for noisy environments and for PD signals with low SNR, as the CE profile is less affected by the instantaneous fluctuations in signal amplitude (Sinaga et al., 2011; Xavier et al., 2021).

Currently, research on PD phenomena in natural esters, such as RBDPO, remains at an early stage (Gautam et al., 2023; Gautam et al., 2024). Previous studies have mainly focused on the investigation of PD characteristics and detection in natural esters, while studies that are related to UHF PD localisation are still limited (Lam, 2024; Mengounou et al., 2022; Rajab et al., 2025). Therefore, this study aims to experimentally investigate the feasibility of UHF PD localisation in RBDPO. An experiment is established to generate PD in a test tank by means of a needle-plane electrode configuration. The UHF PD signals are acquired through 4 patch-type UHF sensors mounted on the external surface of the tank wall. In addition, an Impedance Matching Circuit (IMC) is employed to record the electrical PD signals and verify the PD occurrence. The UHF PD signals are denoised by a Gaussian filter, and the FPM and CEM are applied to determine the TDOA from the denoised PD signals. Finally, the PD localisation in either MO or RBDPO is carried out based on the calculated TDOAs and a non-iterative algorithm that locates the PD location without iterative optimisation, thus reducing computational time. The main contribution of this study is the experimental validation of UHF PD localisation in RBDPO and its quantitative comparison with conventional MO under controlled experimental conditions.

METHODOLOGY

Oil Sample Preparation

RBDPO served as the test sample in this study, while MO was employed for comparison. The properties of the MO and RBDPO used in this study are listed in Table 1 (Dombek & Gielniak, 2023; Mohamad et al., 2018). The preparation of the oil samples involved filtration and drying to ensure that the oil samples were clean and to reduce the moisture content. Filter paper with a 0.22 μm pore size was utilised to filter the oil samples. The drying process was conducted at 85 °C for 48 hours for MO and 72 hours for RBDPO. After drying, the moisture contents of the MO and RBDPO were 12 ppm and 109 ppm, respectively. In total, 65 litres of MO and 65 litres of RBDPO were filtered, dried, and filled into the test tank for the PD test.

PD Experiment Setup

The PD experiment setup was conducted according to the IEC 60270 standard, as shown in Figure 1 (Ahmad et al., 2023). A needle-plane electrode pair was employed to generate ionisation (Murad et al., 2025), with the gap between the needle tip and the plane electrode fixed at 35 mm. The strong electric field produced at the needle tip enabled stable and repeatable PD generation. Therefore, this defect model is commonly adopted in PD localisation experiments (Wu et al., 2022). The dimensions of the steel test tank were 100 cm in length, 30 cm in width, and 25 cm in height, and it was grounded. The lower-left corner at the back of the test tank was taken as the origin of the tank's coordinate system. The x-, y-, and z-axis orientations are illustrated in Figure 2. Based on this coordinate system, the PD source was placed at the coordinates 0.50 m, 0.08 m, 0.15 m inside the tank. The temperature in the experimental environment ranged from 26 °C to 28 °C, while the relative humidity ranged from 53% to 69%. In total, 10 different sensor configurations were employed for UHF PD signal acquisition, and the coordinates of the 4 UHF sensors for each of the configurations are listed in Table 2. For both MO and RBDPO, 3 UHF PD measurements were taken for each of the sensor configurations.

Table 1
Oil properties

Properties	Unit	MO	RBDPO
Density at 20 °C	kg/dm^3	0.88	0.89
Viscosity at 40 °C	cSt	8.7	37.3
Permittivity	-	2.19	2.78
Breakdown Voltage (2.5 mm gap, after treatment)	kV	55	52

For PD detection, 4 commercially available patch-type UHF sensors with an operating frequency range from 0.7 GHz to 2.7 GHz were used. The specifications of the UHF sensors are outlined in Table 3. The UHF PD signals were recorded by a LeCroy WaveMaster

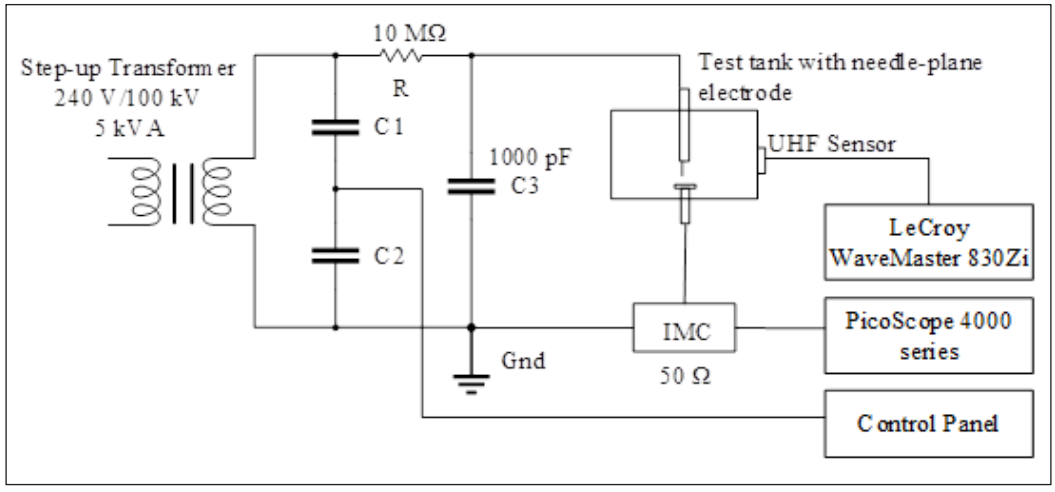


Figure 1. Schematic diagram of PD experiment using UHF sensor configuration

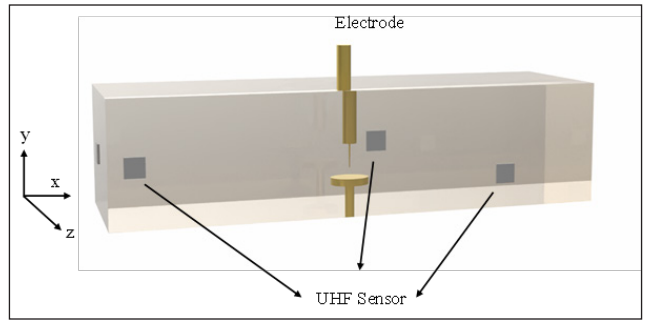


Figure 2. Electrode and sensor placement

Table 2
Coordinates of the 4 UHF sensors for the 10 experimental configurations

Sensor Configuration	Sensor 1 (x, y, z) (m)	Sensor 2 (x, y, z) (m)	Sensor 3 (x, y, z) (m)	Sensor 4 (x, y, z) (m)
1	1.00, 0.15, 0.15	0.75, 0.15, 0.30	0.10, 0.05, 0.30	0.00, 0.10, 0.20
2	1.00, 0.15, 0.15	0.60, 0.10, 0.30	0.35, 0.10, 0.30	0.00, 0.10, 0.20
3	1.00, 0.10, 0.10	0.80, 0.05, 0.30	0.45, 0.05, 0.30	0.95, 0.15, 0.00
4	1.00, 0.05, 0.25	0.20, 0.15, 0.30	0.70, 0.05, 0.30	0.90, 0.15, 0.00
5	0.10, 0.15, 0.00	0.20, 0.15, 0.30	0.60, 0.10, 0.30	0.95, 0.10, 0.00
6	0.10, 0.15, 0.00	0.20, 0.15, 0.30	0.85, 0.10, 0.30	0.95, 0.05, 0.00
7	0.05, 0.15, 0.00	0.10, 0.15, 0.30	0.45, 0.15, 0.30	1.00, 0.05, 0.15
8	0.05, 0.15, 0.00	0.10, 0.15, 0.30	0.50, 0.10, 0.30	0.85, 0.15, 0.00
9	0.00, 0.15, 0.15	0.20, 0.10, 0.30	0.45, 0.15, 0.30	1.00, 0.05, 0.15
10	0.00, 0.15, 0.15	0.20, 0.10, 0.30	0.70, 0.15, 0.30	0.90, 0.10, 0.00

830Zi oscilloscope operated at a sampling rate of 40 GS/s and a bandwidth of 3 GHz to ensure high-resolution data collection. Meanwhile, an electrical PD signal was acquired to verify the occurrence of PD. The electrical PD signals were measured through an IMC and recorded by a PicoScope 4824 oscilloscope at a sampling rate of 80 MS/s and a bandwidth of 20 MHz. The calibration of the IMC was performed with a JZF-10 external calibrator to establish the voltage level that corresponded to an apparent charge of 100 pC (Working Group D1.33, 2012). Only UHF signals that occurred simultaneously with an IMC voltage above the calibrated threshold were considered as valid PD signals.

Table 3
UHF sensor specifications

Parameter	Specification
Return Loss	-9 dB
Efficiency	45%
Peak Gain	2 dBi
Average Gain	-3.5 dB
Radiation Pattern	Omni-directional

PD Signal Denoising and TDOA Computation

The UHF PD signals were denoised by Gaussian filtering before the TDOA determination process (Šefl et al., 2023). The employed Gaussian filter consisted of a Gaussian kernel derived from the Gaussian distribution, as expressed in Equation 1.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad [1]$$

In this equation, σ denotes the standard deviation that controls the width of the filter, and x is the distance from the kernel centre. In this study, σ was set to 10. Noise reduction was performed through a convolution between the UHF PD signal and a Gaussian kernel, as given in Equation 2.

$$S_{denoised}\,(n) = \int_{-\infty}^{\infty} S_{raw}\,(t) G(n-t)\,dt \quad [2]$$

In Equation 2, n represents the data point of the denoised UHF PD signal, whereas t denotes the data point in the raw UHF PD signal. $S_{raw}\,(t)$ and $S_{denoised}\,(n)$ are the raw and denoised UHF PD signals, respectively.

The FPM and CEM were subsequently used to compute the TDOA from the denoised UHF PD signals. In the FPM, the denoised UHF PD signals were subjected to unipolar conversion to ensure consistency in polarity while the amplitude characteristics of the PD signals were preserved. This conversion reduced the effects of phase shift and waveform distortion and allowed for reliable comparisons of PD signals from different sensors. The UHF PD signals were then normalised to a uniform amplitude range between 0 and 1. This step was necessary to ensure comparable amplitude levels among the PD signals, especially in RBDPO, where attenuation was apparent. The TOA for each of the UHF PD signals was determined as the time of the first local peak that exceeded an amplitude threshold of 0.25. Sensor 1 was designated as the reference sensor, and the TDOAs were computed as the difference between the TOA at the reference sensor and the TOAs at the other three sensors.

For CEM, the CE curve of each of the UHF PD signals was first calculated based on Equation 3.

$$U_k[n] = \sum_{n=1}^{N} x_k^2[n] \quad [3]$$

In Equation 3, U_k represents the CE curve, x_k^i denotes the UHF PD signal from the sensor k, and N is the total number of data points in the UHF PD signal. The similarity function in Equation 4 was then applied to determine the TDOAs between the reference sensor and the other three sensors.

$$S[t_l] = \sum_{i=1}^{N-l} |U_1(t_i) - U_k(t_{i+l})|; \quad k = 2, 3, 4 \qquad [4]$$

In this equation, U_1 and U_k are the CE curves from the reference sensor and the kth sensor, respectively, and l denotes the amount of time shift between the two curves. Since the UHF PD signals were sampled at 40 GS/s, one unit difference in l corresponded to a time interval of 25 ps. The time shift l that produced the lowest value of $S[t_l]$ was taken as the TDOA between the reference sensor and the kth sensor (Sinaga et al., 2012).

PD Localisation Based on TDOA

The non-iterative algorithm in (Sinaga et al., 2012) was employed to perform PD localisation based on Equations 5 and 6.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = -[x_{mn} \quad y_{mn} \quad z_{mn}]^{-1} \times \left[(r_{mn}) r_n + \frac{1}{2} ({r_{mn}}^2 - K_m + K_n) \right] \qquad [5]$$

$$r_n^2 = (x - x_n)^2 + (y - y_n)^2 + (z - z_n)^2 \qquad [6]$$

In Equation 5, n denotes UHF sensor 1, which serves as the reference sensor, while m corresponds to UHF sensors 2, 3, and 4. The term r_{mn} represents the TDOA between sensor *m* and sensor *n*, multiplied by the propagation velocity of the UHF PD signal in the insulation medium under consideration. The unknown term r_n indicates the distance from the reference sensor to the PD source, and K_m and K_n are defined as $K_j = x_j^2 + y_j^2 + z_j^2$. The TDOAs derived from the FPM and CEM were used in Equation 5 to express the PD source coordinates, (x, y, z) in terms of r_n. The quadratic equation in Equation 6 was then solved to obtain 2 roots, which were substituted back into Equation 5 to compute the PD source location.

Equation 7 was employed to compute the propagation velocity of the UHF PD signal either in RBDPO or MO. In this equation, v denotes the propagation velocity of the UHF signal in the insulation medium, s_l refers to the speed of light in a vacuum, and ε_r represents the relative permittivity of the insulating liquid. In this study, s_l was taken as 3×10^8 m/s and ε_r was set to 2.78 for RBDPO. Under these conditions, the calculated velocity of the

UHF PD signal in RBDPO was 1.7992 × 10^8 m/s (Mohamad et al., 2014). Similarly, the propagation velocity of the UHF PD signal in MO was computed as 2 ×10^8 m/s. Since the UHF PD signals were sampled at 40 GS/s, the oscilloscope provided a temporal resolution of 25 ps. This corresponded to a spatial resolution of approximately 5 mm in MO and 4.5 mm in RBDPO for the TDOA-based PD localisation method.

$$v = \frac{s_l}{\sqrt{\varepsilon_r}} \quad [7]$$

The accuracy of UHF PD localisation in both MO and RBDPO was evaluated through the localisation error (LE) and relative error (RE) as defined in Equations 8 and 9 (Mohd Hashim et al., 2022).

$$LE = \sqrt{\left(x_{actual} - x_{computed}\right)^2 + \left(y_{actual} - y_{computed}\right)^2 + \left(z_{actual} - z_{computed}\right)^2} \quad [8]$$

$$RE = \frac{LE}{Tank\ Diagonal} \times 100 \quad [9]$$

The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were also employed to assess the PD localisation accuracy along the x-, y- and z-axes (Perfetto and D'Antona, 2021; Khodaveisi et al. 2024). These metrics measure the deviation between the estimated and actual PD coordinates. Lower MSE and RMSE indicate high PD localisation accuracy.

RESULTS AND DISCUSSION

Amplitude of UHF PD Signals in MO and RBDPO

The average, maximum, and minimum amplitudes of the UHF PD signals in MO and RBDPO are presented in Figure 3. The amplitudes of the UHF PD signals are generally higher in MO than in RBDPO. This is possibly due to differences in the dielectric loss properties of the 2 oil samples (Dombek & Gielniak, 2023). The percentage difference in average amplitude between MO and RBDPO is 92.7%. For the maximum and minimum amplitudes, the percentage differences are 92.4% and 91.3%, respectively.

Denoising of UHF PD Signals in MO and RBDPO

Figure 4 shows an instance of the raw and denoised UHF PD signals in MO. It is found that the raw UHF PD signal exhibits a distinct peak at the start of the discharge event, and

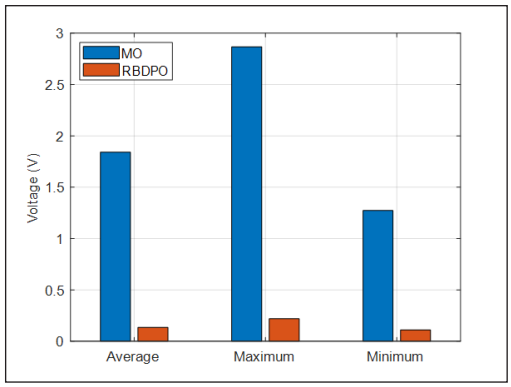


Figure 3. Amplitudes of the UHF PD signals in MO and RBDPO

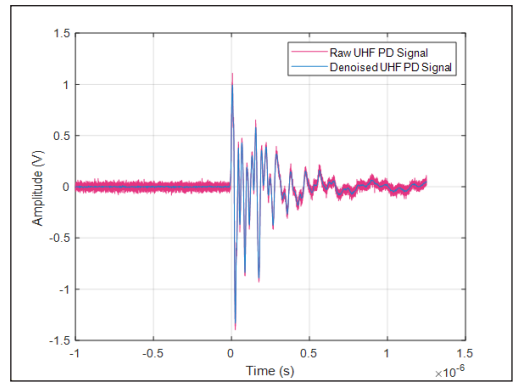


Figure 4. Raw and denoised UHF PD signals in MO

it is followed by oscillations that gradually decay over time. The amplitude reaches ± 1.1 V, which suggests a less obstructed propagation path in MO as compared to RBDPO. A moderate level of background noise appears before and after the main PD events, which is anticipated due to the electromagnetic interference. The UHF PD signal denoised by the Gaussian filter preserves the main waveform structure, while the background noise is largely reduced.

Figure 5 shows the raw and denoised UHF PD signals in RBDPO. The raw UHF PD signal has a similar structure to that in MO, which consists of an initial peak, and it is followed by damped oscillations. However, the amplitude is relatively low, at ± 0.09 V, which indicates that RBDPO exhibits stronger attenuation characteristics than MO. The background noise in RBDPO appears slightly higher than in MO, possibly due to the differences in dielectric properties, such as its viscosity, permittivity, and conductivity, which may affect the propagation and dispersion of UHF PD signals (Chandrasekar & Montanari, 2014; Jaroszewski & Rakowiecki, 2017; Khelifa et al., 2024). The denoised UHF PD signal shows a significant improvement in clarity, whereby the background noise is reduced by 79.9%. Nevertheless, the overall amplitude remains lower than in MO, which peaks at ± 0.08 V. The oscillations after the initial PD event become more distinct in the denoised UHF PD signal, which may facilitate the PD localisation analysis.

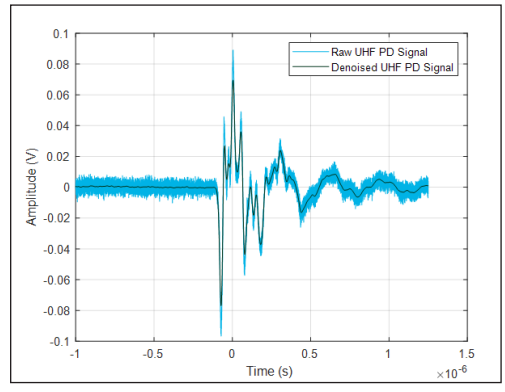
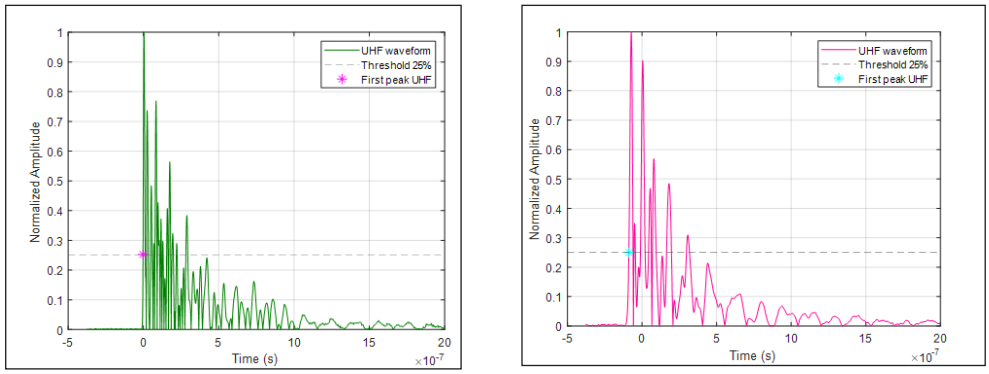


Figure 5. Raw and denoised UHF PD signals in RBDPO

TDOA Determination

First Peak Method

The UHF PD waveforms in MO and RBDPO after unipolar conversion and normalisation are illustrated in Figures 6(a) and 6(b). A 0.25 amplitude threshold is applied in the FPM to ensure that the TOA is identified from the actual first peak of the UHF PD signals instead of the background noise. For MO, the normalised waveforms exhibit several early peaks with amplitudes up to 0.99 p.u., and slowly decay, as depicted in Figure 6(a). The normalised UHF PD signals in RBDPO show sharp initial peaks with amplitudes up to 0.99 p.u., which lead to attenuated peaks. It is found that after the initial peaks, the amplitudes of the UHF PD signal in RBDPO are lower than those in MO. The UHF PD waveforms in RBDPO are more damped, as indicated by the fewer oscillations after 0.5 μs, as in Figure 6(b). These PD waveform characteristics suggest a higher dielectric loss and a relatively obstructed propagation medium for UHF signals in RBDPO. Although the PD signal duration is shorter in RBDPO, the relatively clearer initial peaks likely improve the localisation accuracy. These observations agree with previous studies on the impact of oil properties on UHF signal behaviour and PD detection in RBDPO (HadiAlnashia et al., 2018; Muangpratoom et al., 2023; Suhaimi et al., 2025).



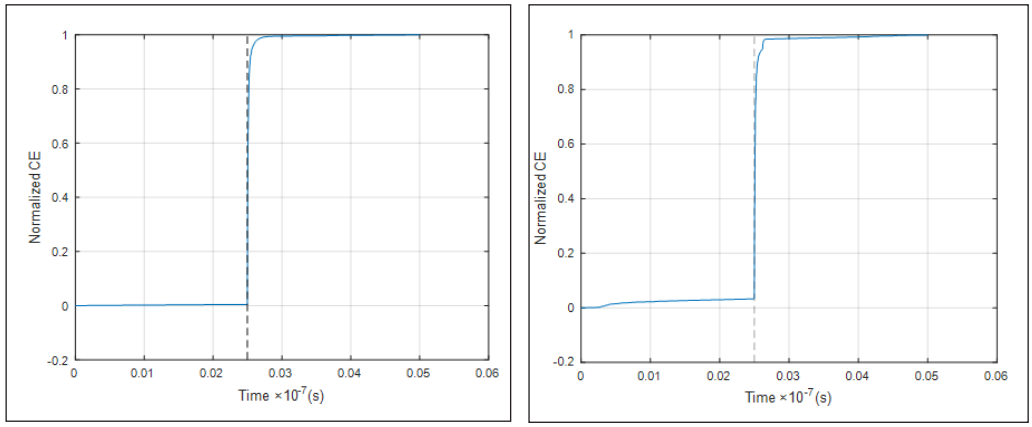
(a) (b)

Figure 6. UHF PD signals in (a) MO and (b) RBDPO after unipolar conversion and normalisation in the FPM

Cumulative Energy Method

The CE curves of the UHF PD signals in MO and RBDPO are depicted in Figures 7(a) and 7(b). The UHF PD signals in both MO and RBDPO show similar CE patterns, where the curves start with a flat region and increase rapidly, marking the occurrence of the PD event. In MO, the CE curve exhibits a steep and abrupt transition across all sensors. As seen in Figure 7(a), the normalised CE curve rises abruptly from zero to unity within a narrow time range of less than 1 ns. This sharp transition indicates strong PD signal intensity and

stable propagation of the UHF PD signal through MO. In contrast, the CE curve of the UHF PD signals in RBDPO displays a more gradual rise, and the overall slope is lower than that of MO, as illustrated in Figure 7(b). This can be attributed to the higher permittivity and viscosity of RBDPO, which tend to attenuate and slow down UHF wave propagation in the medium (Durga et al., 2020; Kurimský et al., 2021). For both MO and RBDPO, the CE curves from each of the sensors are closely aligned; hence, the TDOAs between the collected UHF PD signals fall within the nanosecond range.



(a) (b)

Figure 7. Normalised CE curves of UHF PD signals in (a) MO and (b) RBDPO

PD Localisation

The LE and RE for PD localisation in MO and RBDPO are shown in Figure 8. Each data point represents the average of the 30 UHF PD measurements. For TDOAs from the FPM, the PD localisation in RBDPO is better than in MO, where the LE in MO is 2.27 cm higher than in RBDPO. This is consistent with the RE, as the RE of PD localisation in RBDPO is 2.1% lower than in MO. In contrast, the PD localisation in MO is better than in RBDPO for TDOAs from the CEM, where the LEs are 54.22 cm and 56.83 cm, respectively. Similarly, the RE of PD localisation in MO is 2.4% lower than in RBDPO. The standard deviations of the PD LEs are also presented in Figure 8. For the FPM, the standard deviations of the LEs in MO and RBDPO are 18.24 cm and 17.23 cm, whereas for the CEM, the standard deviations are 17.73 cm and 17.55 cm for RBDPO and MO, respectively. The comparable standard deviations among the 4 cases suggest that the spread of the PD LEs is similar under different types of oil samples and TDOA estimation methods. Overall, the TDOAs from the FPM produce lower LE than those obtained from CEM. Furthermore, for TDOAs from the FPM, PD localisation in RBDPO results in lower LE than in MO. The higher PD localisation accuracy in RBDPO is likely due to its higher permittivity as compared to MO,

which could help to reduce multipath reflections and improve the SNR of UHF PD signals (Azis et al., 2014; Mu et al., 2024; Yu & Zhang, 2024). Although the theoretical spatial resolution for PD localisation is approximately 5 mm in MO and 4.5 mm in RBDPO, the experimental PD LEs are considerably larger. This is because the practical PD localisation accuracy depends not only on the temporal resolution of the PD detection system but also on measurement noise, multipath propagation, and error in TDOA estimation.

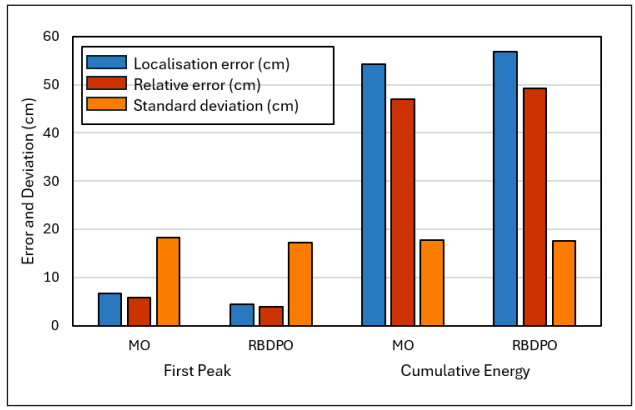


Figure 8. Localisation error (LE) and relative error (RE) of PD localisation in MO and RBDPO based on TDOAs obtained from FPM and CEM

Table 4
MSE and RMSE of PD localisation based on TDOAs from the FPM

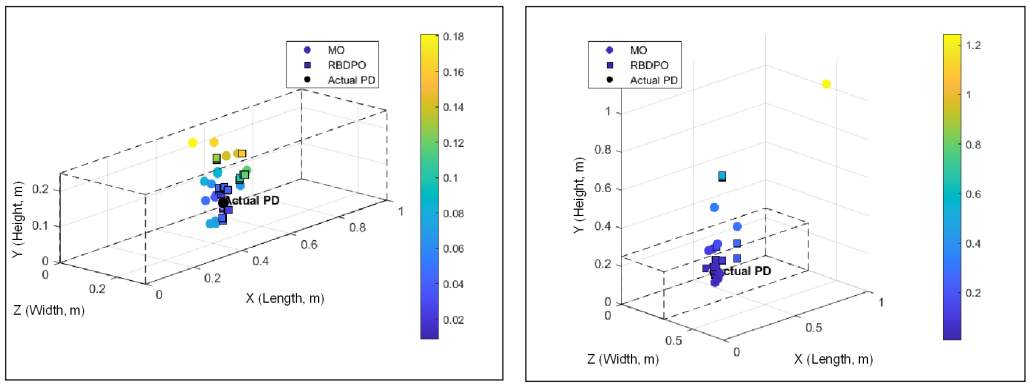
	MO			RBDPO		
	X	**Y**	**Z**	**X**	**Y**	**Z**
MSE (mm^2)	0.42	12.91	1.83	0.06	4.41	1.04
RMSE (mm)	20.37	113.62	42.76	7.56	66.39	32.19

Table 4 presents the MSE and RMSE of the estimated PD locations based on the TDOAs from the FPM. Based on the MSE, the estimated PD locations in RBDPO have lower errors than in MO across all x-, y-, and z-axes. The x-axis records the lowest MSE at 0.06 mm^2. The RMSE displays a similar trend to the MSE, whereby the PD localisation in RBDPO yields lower errors in the x- and z-axes, at 7.56 mm and 32.19 mm as compared to MO. For the y-axis, RBDPO records an RMSE of 66.39 mm, which is the highest among the 3 axes but still 47.23 mm lower than in MO.

The MSE and RMSE of the estimated PD locations that are computed with the TDOAs from the CEM are reported in Table 5. For MO, the highest MSE and RMSE among the 3 axes occur along the y-axis, at 15.98 mm^2 and 126.43 mm, respectively. A similar pattern is observed in RBDPO, where the y-axis also records the largest errors, with an MSE of

Table 5
MSE and RMSE of PD localisation based on TDOAs from the CEM

	MO			RBDPO		
	X	**Y**	**Z**	**X**	**Y**	**Z**
MSE (mm^2)	0.32	15.98	1.78	0.55	32.05	7.17
RMSE (mm)	17.99	126.43	42.32	23.52	179.03	84.10



(a) (b)

Figure 9. 3D heatmap scatter plot of the LE of the estimated PD locations in MO and RBDPO based on TDOAs obtained through (a) FPM and (b) CEM

32.05 mm^2 and RMSE of 179.03 mm. According to these analyses, the PD localisation in RBDPO based on the TDOAs from the FPM produces the lowest MSE and RMSE among all cases, regardless of the axis. These results align with the LE and RE analyses, which showcase that the TDOAs from the FPM generally lead to higher localisation accuracy, and that for the FPM, PD localisation in RBDPO is better than in MO.

The 3D heatmap scatter plot of the LE of the estimated PD locations in MO and RBDPO is illustrated in Figure 9. Figure 9(a) and Figure 9(b) present the estimated PD locations computed from the TDOAs of the FPM and CEM, respectively. The actual PD location is shown as a black marker, while the estimated PD locations in MO and RBDPO are represented by circular and square markers, respectively. The colour gradient indicates the magnitude of LE, whereby blue to green colours represent lower LE values, while yellow colours represent higher LE values. Figure 9(a) clearly highlights the contrast in localisation accuracy between the oil samples. The RBDPO cluster appears tight around the actual PD location without any yellow square markers, and the LEs remain below 0.18 m. For PD localisation based on the TDOAs from the CEM, the LEs are larger as compared to those from the FPM. This is evident in Figure 9(b), where the colour scale exceeds 1 m, and several estimated PD locations fall outside the tank. This result is consistent with the LE and RE results in Figure 8.

CONCLUSION

This study presented an experimental comparison of PD localisation performance in MO and RBDPO based on the UHF method. The UHF PD signals are recorded experimentally, while electrical PD signals are acquired in parallel to confirm the occurrence of PDs. It is found that RBDPO achieves lower LE and RE as compared to MO based on the TDOAs obtained from the FPM. Conversely, CEM produces significantly larger errors than FPM for PD localisation in both MO and RBDPO. Furthermore, the 3D heatmap scatter plots demonstrate that the estimated PD locations in RBDPO are clustered closely around the actual PD source, whereas those in MO exhibit slightly broader dispersion. Overall, UHF PD detection and localisation in RBDPO are experimentally demonstrated to be feasible. These findings support the potential application of natural esters such as RBDPO as alternative insulation in transformers, where the insulation condition can be effectively monitored through the UHF method.

ACKNOWLEDGEMENT

This research was funded by Universiti Putra Malaysia, Putra Grant Scheme, Inisiatif Putra (GP- FASTRACK/2024/9791100) and (GP-IPI (KMD)/2023/9756300).

LIST OF ABBREVIATIONS

PD	:	Partial discharge
RBDPO	:	Refined, Bleached and Deodorised Palm Olein
MO	:	Mineral Oil
UHF	:	Ultra-High Frequency
TOA	:	Time of Arrival
TDOA	:	Time Difference of Arrival
FPM	:	First Peak Method
CE	:	Cumulative Energy
CEM	:	Cumulative Energy Method
IMC:	:	Impedance Matching Circuit
MSE	:	Mean Squared Error
RMSE	:	Root Mean Squared Error
LE	:	Localisation Error
RE	:	Relative error

REFERENCES

Adak, S., & Cangi, H. (2024). The quality problems at low irradiance in the grid-connected photovoltaic systems. *Electrical Engineering, 106*(5), 6185-6197.https://doi.org/10.1007/s00202-024-02351-6

Ahmad, N. D., Azis, N., Jasni, J., & Rezali, K. A. (2025). Acoustic partial discharge localisation in refined, bleached and deodorised palm oil (RBDPO) based on the Savitzky-Golay and moving average denoising methods. *Pertanika Journal of Science & Technology, 33*, 115-128. https://doi.org/10.47836/pjst.33.S4.07

Ahmad, N. D., Hashim, A. H. M., Azis, N., Talib, M. A., Rezali, K. A. M., Nik Ali, N. H., Jasni, J., Masahiro, K., & Jamil, M. K. M. (2023). Partial discharge localisation of refined bleached deodorised palm oil (RBDPO) based on the acoustic emission method. *APL 2023- 12th Asia-Pacific International Conference on Lightning*, 1-5. https://doi.org/10.1109/APL57308.2023.10181667

Antony, D., Mishra, P., Todi, A., John, N. A., & Patankar, A. M. (2021). Partial discharge localisation in transformers using UHF technique: Non-iterative method. *International Journal of Information Technology, 13*(1), 291-297.https://doi.org/10.1007/s41870-020-00481-1

Azis, N., Jasni, J., Ab Kadir, M. Z. A., & Mohtar, M. N. (2014). Suitability of palm-based oil as a dielectric insulating fluid in transformers. *Journal of Electrical Engineering & Technology, 9*(2), 662-669.https://doi.org/10.5370/JEET.2014.9.2.662

Azmi, K., Zuhairi, A., Ishak, D., Muhamad, N. A., & Kamarol, M. (2017). Partial discharge characteristics of spherical metal particles in mineral oil and PFAE under AC voltage. In *Proceedings of the 2017 IEEE Region 10 Conference (TENCON)* (pp. 2588-2593). https://doi.org/10.1109/TENCON.2017.8228298

Chai, H., Lu, S., Phung, B. T., & Mitchell, S. (2019). Comparative study of partial discharge localisation based on UHF. In *Proceedings of the 25th International Conference on Electricity Distribution* (pp. 3-6).

Chan, J. Q., Raymond, W. J. K., Illias, H. A., & Othman, M. (2023). Partial discharge localisation techniques: A review of recent progress. *Energies, 16*(6). https://doi.org/10.3390/en16062863

Chandrasekar, S., & Montanari, G. C. (2014). Analysis of partial discharge characteristics of natural esters as dielectric fluid for electric power apparatus applications. *IEEE Transactions on Dielectrics and Electrical Insulation, 21*(3), 1251-1259. https://doi.org/10.1109/TDEI.2014.6832272

Desai, B. M. A., & Sarathi, R. (2018). Identification and localisation of incipient discharges in transformer insulation adopting UHF technique. *IEEE Transactions on Dielectrics and Electrical Insulation, 25*(5), 1924-1931.https://doi.org/10.1109/TDEI.2018.007294

Dombek, G., & Gielniak, J. (2023). Dielectric properties and fire safety of mineral oil and low-viscosity natural ester mixtures in various concentrations. *Energies, 16*(10). https://doi.org/10.3390/en16104195

Durga, G., Mahidhar, P., Karthikeyan, A. S., Sarathi, R., Taylor, N., & Edin, H. (2020). *Dielectric properties of mixed mineral and synthetic ester oil*. https://doi.org/10.1049/iet-smt.2019.0227

Gautam, L., Sarathi, R., Soundarya, P., Sekar, G., Rao, U. M., & Fofana, I. (2023). Understanding the surface discharge characteristics of thermally aged ester fluid impregnated pressboard adopting the fluorescent fibre technique. *Proceedings of the International Symposium on Electrical Insulating Materials*, 211-214. https://doi.org/10.23919/ISEIM60444.2023.10329150

Gautam, L., Sarathi, R., Soundarya, P., Sekar, G., Rao, U. M., & Fofana, I. (2024). Quantitative estimation of the gelling in thermally aged natural ester dielectric liquid. *IEEE Transactions on Dielectrics and Electrical Insulation, 31*(2), 897-903. https://doi.org/10.1109/TDEI.2023.3340632

Girish, S. K., Sumangala, B. V., Raviprasad, D., & Jagadisha, K. R. (2021). Investigations on partial discharge characteristics of vegetable oil and mineral oil. *International Journal of Engineering Research & Technology (IJERT), 10*(2), 95-97.

Ha, S. G., Cho, J., Lee, J., Min, B. W., Choi, J., & Jung, K. Y. (2018). Numerical study of estimating the arrival time of UHF signals for partial discharge localisation in a power transformer. *Journal of Electromagnetic Engineering and Science, 18*(2), 94-100. https://doi.org/10.26866/jees.2018.18.2.94

HadiAlnashia, I. A., Ibrahim Mohd, I., Almelian, M. M., & Omran, M. A. (2018). The impact of partial discharge on the dielectric properties of refined bleached and deodorised palm oil (RBDPO). *Journal of Telecommunication, Electronic and Computer Engineering, 10*(1-3), 81-85.

Jangjoo, M. A., Allahbakhshi, M., & Mirzaei, H. R. (2023). UHF sensors positioning on the power transformer tank to enhance the partial discharge localisation accuracy. *Electric Power Systems Research, 218*, Article 109174. https://doi.org/10.1016/j.epsr.2023.109174

Jaroszewski, M., & Rakowiecki, K. (2017). Partial discharge inception voltage in transformer natural ester liquid -Effect of the measurement method in the presence of moisture. *IEEE Transactions on Dielectrics and Electrical Insulation, 24*(4), 2477-2482. https://doi.org/10.1109/TDEI.2017.005917

Khelifa, H., Skotnicki, P., Beroual, A., Vagnon, E., & Jaroszewski, M. (2024). Effect of MgO nanoparticles on the AC breakdown voltage, partial discharge activity, and electrostatic charging tendency of natural ester. *IEEE Transactions on Dielectrics and Electrical Insulation, 31*(3), 1350-1358. https://doi.org/10.1109/TDEI.2024.3374244

Kurimský, J., Rajňák, M., Šárpataky, M., Čonka, Z., & Paulovičová, K. (2021). Electrical and acoustic investigation of partial discharges into two types of nanofluids. *Journal of Molecular Liquids, 341*, Article 117444. https://doi.org/10.1016/j.molliq.2021.117444

Lam, K. Y. (2024). *Dissolved gas analysis (DGA) for natural ester-filled transformers*. Cargill Brand.

Mengounou, G. M., Jiosseu, J. L., & Imano, A. M. (2022). Experimental investigation of partial discharge activities in palm kernel oil methyl ester and mineral oil by using a Rogowski coil. *Electric Power Systems Research, 213*, Article 108708. https://doi.org/10.1016/j.epsr.2022.108708

Mina-Casaran, J. D., García, B., García, D. F., Gómez, A. M., & Montero, A. (2025). Partial discharge patterns in natural esters from different vegetable bases. *IEEE Transactions on Dielectrics and Electrical Insulation, 32*(1), 435-443. https://doi.org/10.1109/TDEI.2024.3486275

Mohamad, N. A., Azis, N., Jasni, J., Ab Kadir, M. Z. A., Yunus, R., Ishak, M. T., & Yaakub, Z. (2014). A study on the dielectric properties of palm oil and coconut oil. In *2014 IEEE International Conference on Power and Energy (PECon)* (pp. 109-112). IEEE. https://doi.org/10.1109/PECON.2014.7062423

Mohamad, N. A., Azis, N., Jasni, J., Kadir, M. Z. A., Yunus, R., & Yaakub, Z. (2018). Physiochemical and electrical properties of refined, bleached and deodorised palm oil under high temperature ageing for application in transformers. *Energies, 11*(6). https://doi.org/10.3390/en11061583

Mohamad, N. A., Azis, N., Jasni, J., Mohd Zainal, M. Z. A., Yunus, R., & Yaakub, Z. (2021). Experimental study on the partial discharge characteristics of palm oil and coconut oil-based Al2O3 nanofluids in the presence of sodium dodecyl sulfate. *Nanomaterials, 11*(3), 1-18. https://doi.org/10.3390/nano11030786

Mohan Rao, U., Fofana, I., Jaya, T., Rodriguez-Celis, E. M., Jalbert, J., & Picher, P. (2019). Alternative dielectric fluids for transformer insulation system: Progress, challenges, and prospects. *IEEE Access, 7*, 184552-184571. https://doi.org/10.1109/ACCESS.2019.2960020

Mohd Hashim, A. H., Azis, N., Jasni, J., Mohd Radzi, M. A., Kozako, M., Kamarol Mohd Jamil, M., & Yaakub, Z. (2022). Partial discharge localisation in oil through acoustic emission technique utilising fuzzy logic. *IEEE Transactions on Dielectrics and Electrical Insulation, 29*(2), 623-630. https://doi.org/10.1109/TDEI.2022.3157911

Mu, G., Dai, Q., Chai, S., & Yang, P. (2024). Ultrasonic propagation characteristics of partial discharge in oil-impregnated paper traction transformers. *AIP Advances, 14*(3). https://doi.org/10.1063/5.0191688

Muangpratoom, P., Suriyasakulpong, C., Maneerot, S., Vittayakorn, W., & Pattanadech, N. (2023). Experimental study of the electrical and physicochemical properties of different types of crude palm oils as dielectric insulating fluids in transformers. *Sustainability, 15*(19), Article 14269. https://doi.org/10.3390/su151914269

Murad, T. K., Syakur, A., & Setiawan, I. (2025). Analysis of partial discharge characteristics in transformer oil insulation media using needle-plane and plane-plane electrode systems. *International Journal of Electrical and Computer Engineering (IJECE), 15*(5), 4445-4453. https://doi.org/10.11591/ijece.v15i5.pp4445-4453

Perfetto, L., & D'Antona, G. (2021). Experimental results of partial discharge localisation in bounded domains. *Sensors, 21*(3), Article 935. https://doi.org/10.3390/s21030935

Qian, S., Chen, H., Xu, Y., & Su, L. (2018). High sensitivity detection of partial discharge acoustic emission within power transformers by Sagnac fibre optic sensor. *IEEE Transactions on Dielectrics and Electrical Insulation, 25*(6), 2313-2320. https://doi.org/10.1109/TDEI.2018.007131

Rajab, A., Ardison, A., Pratama, R. W., Novendra, R., Baharuddin, B., & Artono, T. (2025). Dissolved gas analysis of methyl ester-mineral oil mixture under partial discharge. *TEM Journal, 14*(1), 55-63. https://doi.org/10.18421/TEM141-06

Rathod, V. B., Kumbhar, G. B., & Bhalja, B. R. (2021). Partial discharge detection and localisation in power transformers based on acoustic emission: Theory, methods, and recent trends. *IETE Technical Review, 39*(3), 540-552. https://doi.org/10.1080/02564602.2021.1873155

Sinaga, H. H., Phung, B. T., & Blackburn, T. R. (2011). Partial discharge localisation in transformers using UHF sensors. *Electrical Insulation Conference*. https://doi.org/10.1109/EIC.2011.5996117

Sinaga, H. H., Phung, B. T., & Blackburn, T. R. (2012). Partial discharge localisation in transformers using UHF detection method. *IEEE Transactions on Dielectrics and Electrical Insulation, 19*(6), 1891-1900. https://doi.org/10.1109/TDEI.2012.6396945

Song, L. (2004). *Detection and position location of partial discharges in transformers using fibre optic sensors* [Thesis]. Virginia Tech. http://scholar.lib.vt.edu/theses/available/etd-12032004-151222/

Suhaimi, N. S., Zainul Ariffin, K. N., Zaini, N. H., Abdul Rahman, S., Azizan, M. M., & Rameli, N. (2025). Palm oil as a transformer insulating fluid: A review. *Semarak Proceedings of Natural and Environmental Sciences, 3*(1), 13-17. https://doi.org/10.37934/spnes.3.1.1317a

Thiviyanathan, V. A., Ker, P. J., Leong, Y. S., Abdullah, F., Ismail, A., & Zaini Jamaludin, M. (2022). Power transformer insulation system: A review of the reactions, fault detection, challenges and prospects. *Alexandria Engineering Journal, 61*(10), 7697-7713. https://doi.org/10.1016/j.aej.2022.01.026

Uwiringiyimana, J. P., Khayam, U., Montanari, G. C., & Hikita, M. (2022). Comparative analysis of partial discharge detection features using a UHF antenna and conventional HFCT sensor. *IEEE Access, 10*, 107214-107226. https://doi.org/10.1109/ACCESS.2022.3212746

Working Group D1.33. (2012). *High-voltage on-site testing with partial discharge measurement*. CIGRE Publication.

Wu, H., Liu, H., Xue, Z., Yang, J., Li, Q., He, D., & Siew, W. H. (2022). The generation and migration of bubbles in oil-pressboard insulation. *IEEE Transactions on Dielectrics and Electrical Insulation, 29*(6), 2110-2121. https://doi.org/10.1109/TDEI.2022.3210486

Xavier, G. V. R., de Oliveira, A. C., Silva, A. D. C., Nóbrega, L. A. M. M., da Costa, E. G., & Serres, A. J. R. (2021). Application of time difference of arrival methods in the localisation of partial discharge sources detected using bio-inspired UHF sensors. *IEEE Sensors Journal, 21*(2), 1947-1956. https://doi.org/10.1109/JSEN.2020.3019760

Yu, R., & Zhang, Z. (2024). Transformer partial discharge location technology based on gradient oil temperature. *Frontiers in Energy Research, 12*, 142801. https://doi.org/10.3389/fenrg.2024.1428012

Pertanika J. Sci. & Technol. 34 (4): 2311 - 2327 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Recycling Indonesian Native Chicken Eggshell Membrane Waste into Gelatine: A Comparative Characterisation Using SEM-EDX, FTIR, and XRF

Rina Wahyuningsih[1], Ragil Yuliatmo[2], Andi Febrisiantosa[1], Atiqa Rahmawati[2], Al Mizan[3], Selma Noor Permadi[1], Laila Rahmawati[1], Afi Candra Trinugraha[1], and Teguh Wahyono[1]*

[1]Research Centre for Food Technology and Processing, Research Organisation for Agriculture and Food, National Research and Innovation Agency (BRIN), 55861 Gunungkidul, Special Region of Yogyakarta, Indonesia

[2]Department of Leather Processing Technology, Politeknik ATK Yogyakarta, 55188 Bantul, Special Region of Yogyakarta, Indonesia

[3]Department of Leather Engineering, Khulna University of Engineering and Technology (KUET), 9203 Khulna, Bangladesh

ABSTRACT

Indonesian native chicken eggshells are an underutilised agricultural waste with potential for conversion into gelatine through sustainable recycling methods. This study evaluated and compared the physicochemical properties of gelatine extracted from Indonesian native chicken eggshell membranes (INCES), broiler chicken eggshell membranes (BCES), and bovine skin. Gelatine was extracted from each source and analysed in triplicate. Structural characteristics and elemental composition were investigated using Scanning Electron Microscopy combined with Energy Dispersive X-ray Spectroscopy (SEM-EDX), Fourier Transform Infrared Spectroscopy (FTIR), and X-ray Fluorescence (XRF) were used to elucidate structural features and elemental composition differences among the samples. The SEM photos show that BCES gelatine had larger, more irregular structures, while INCES gelatine had finer, more spherical particles. Bovine gelatine displayed a smoother, fibrous morphology. Nine significant elements were found by EDX elemental analysis (C, N, O, Na, Mg, Al, Si, Ca, and Au). While carbon elements predominated in bovine gelatine, calcium was present in larger amounts in eggshell protein (INCES & BCES). Typical amide bands (A, B, I, II, and III) were visible in the FTIR spectra. In contrast to BCES gelatine, which tended to

ARTICLE INFO

Article history:
Received: 10 November 2025
Accepted: 26 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.17

E-mail addresses:
rina017@brin.go.id (Rina Wahyuningsih)
ragilyuliatmo@atk.ac.id (Ragil Yuliatmo)
andi.febrisiantosa@gmail.com (Andi Febrisiantosa)
tiqa054@gmail.com (Atiqa Rahmawati)
almizan@le.kuet.ac.bd (Al Mizan)
selm002@brin.go.id (Selma Noor Permadi)
lail007@brin.go.id (Laila Rahmawati)
ac.trinugraha@gmail.com (Afi Candra Trinugraha)
tegu021@brin.go.id (Teguh Wahyono)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

exhibit more random coil properties, the spectra of ICES and bovine gelatine revealed β-sheet structures. Analysis using XRF revealed that INCES gelatine contained elevated calcium levels, most likely due to incomplete demineralisation during extraction. Overall, INCES represents a promising alternative gelatine source that supports the valorisation of poultry by-products. These findings demonstrate that INCES gelatine exhibits a distinctive mineral-rich, β-sheet-dominant profile relative to bovine gelatine, positioning it as a promising, halal-compliant alternative for functional food and nutraceutical applications, while further functional and safety validation is required before any biomedical application.

Keywords: Eggshell membranes, gelatine, physicochemical, poultry by-products

INTRODUCTION

Gelatine is a versatile biopolymer widely applied in food, pharmaceutical, and cosmetic industries due to its distinctive functional attributes, such as gel formation, emulsifying capacity, and film-forming ability. It is typically produced through partial hydrolysis of collagen, which is commonly obtained from mammalian sources like bovine and porcine skin. The investigation of alternate sources, such as chicken by-products, has been prompted by the growing demand for halal and non-mammalian gelatine (Gómez-Guillén et al., 2011; Karim & Bhat, 2009). According to Herdiana et al. (2024) and Wahyuningsih et al. (2026), the public's growing awareness of ethics and raw material sustainability is correlated with the growing demand, especially when it comes to the exploration of poultry waste as a more morally and ecologically sound source of gelatine.

A waste product of the chicken business, eggshells worsen already-existing environmental problems. According to some research, eggshells contain amino acids, including substances that resemble collagen and can be used to make gelatine (Bootklad & Kaewtatip, 2013; Hamidi et al., 2017). An alternate source of eggshell membranes with a variety of characteristics is native Indonesian chicken. The composition and shape of Indonesian eggshell membrane (INCES) as a possible source of gelatine have been thoroughly studied. The composition and shape of Indonesian eggshell membrane (INCES) as a possible source of gelatine have been thoroughly investigated in earlier research by Wahyuningsih et al. (2026) and Permadi et al. (2024). They have effectively examined the composition and morphology of Indonesian eggshell membrane (INCES) as a potential source of gelatine. According to Wahyuningsih et al. (2026), the INCES membrane is more valuable in turning environmental waste into profitable gelatine goods because it has a higher content of important amino acids than the broiler chicken eggshell membrane (BCES). Furthermore, Wahyuningsih et al. (2026) reported that INCES contains approximately 25.3% crude protein, substantially higher than that of other poultry eggshell membranes, including indigenous chicken, laying hen, bakery waste, and hatchery waste eggshells. More recently, Wahyuningsih et al. (2026) demonstrated that gelatine extracted

from INCES achieved extraction yields of 47-51%, indicating that this material is not only protein-rich but also capable of producing gelatine with a high recovery. These findings provide a strong scientific basis for selecting INCES as the primary raw material in the present study. Examining protein development in gelatine synthesis aligns with the growing use of circular economy ideas and sustainable waste management (Hoerterer et al., 2022; Yuliatmo et al., 2017).

Currently, Scanning Electron Microscopy (SEM) combined with Energy Dispersive X-ray Spectroscopy (EDX) is widely employed to characterise the microstructure and elemental distribution of biomaterials such as gelatine, enabling clear differentiation in surface morphology and compositional features. In addition, spectroscopy-based techniques, including Fourier Transform Infrared (FTIR) and X-ray Fluorescence (XRF), provide powerful analytical approaches for identifying functional groups and assessing chemical composition, particularly in gelatine. These techniques have been used in earlier studies to evaluate gelatine generated from mammals and broilers (Al-Hassan & Norziah, 2017; Allegretta et al., 2020; Appoloni & Melquiades, 2014; Proietti et al., 2020; Rosa et al., 2024). Systematic investigations on gelatine derived specifically from INCES biomaterial remain limited. This scarcity highlights an important research gap in advancing the valorisation of native Indonesian poultry by-products.

To fill this research gap, the present work investigates gelatine obtained from INCES protein using SEM, EDX, FTIR, and XRF analyses For comparative purposes, gelatine extracted from broiler chicken eggshell membrane (BCES) was included to evaluate the influence of chicken breed on the structural and elemental characteristics of eggshell membrane-derived gelatine, while commercial bovine gelatine was used as a conventional reference material to benchmark the properties of INCES against an established industrial gelatine source. The novelty of this study lies in the application of a comprehensive multi-instrument analytical framework to INCES-based gelatine for the first time, thereby offering essential insights into its potential as a sustainable, culturally appropriate, and economically valuable alternative to conventional gelatine sources.

MATERIALS AND METHODS

Materials

Waste from native Indonesian chicken eggshells that were gathered from nearby farms in Yogyakarta City and Sleman Regency served as the study's samples. Membranes from Indonesian chicken eggshells were used to make gelatine. Samples of eggshells were sorted to make sure they were clean of sand and grime. Gelatine obtained from broiler chicken eggshell membranes and commercially available bovine skin served as control samples for comparative analysis. The auxiliary materials employed in this study included NaOH, acetic acid, and bovine gelatine, all of which were procured from Merck, Darmstadt, Germany. In this research, only analytical-grade compounds are used.

METHODS

Extraction Gelatine

With a few minor adjustments, the extraction technique is based on the research by Mohammadi et al. (2016). After being physically separated, the eggshell membrane was finely powdered. The eggshell was sliced into 1 cm pieces for the membrane extraction procedure, and it was then immersed in 0.1 M NaOH at a 1:40 (w/v) ratio for a full day at room temperature. The membrane samples were rinsed with running water until neutrality was achieved at pH 7. The membrane was then immersed in a second stage of 0.5 M acetic acid for 24 hours at 37°C in a water bath at a ratio of 1:40 (w/v). After filtering, the extracted material was dried in an oven for 48 hours. The residue was employed for proximate analysis, and a sample was taken for processing. After the extracted collagen was gathered, its structural and elemental characteristics were examined using SEM, EDX, FTIR, and XRF.

Energy Dispersive X-Ray (EDX) and Scanning Electron Microscope (SEM) Analysis

The dried gelatine samples were finely ground and homogenised to minimise surface heterogeneity and obtain a more representative surface for elemental characterisation. Using a Hitachi MC 1000 sputter ion, the samples utilised in the SEM analysis were coated with gold (Au) to increase their conductivity, which is not a constituent of the gelatine sample. The coating was applied as a thin layer to minimise X-ray absorption and reduce potential interference during elemental analysis. Before the samples were examined using the SEM method, the ion sputter setting was changed to 20 mA for 20 seconds. The Hitachi SU 3500 (Tokyo, Japan) was used for SEM and EDX testing.

Fourier Transform Infrared Spectroscopy (FTIR) Analysis

The structural characteristics of the gelatine samples were examined spectroscopically using a PerkinElmer FTIR (Model Spectrum 100, PerkinElmer, Waltham, MA, USA). Potassium bromide (KBr) was combined with each sample (INCES gelatine, BCES gelatine, and bovine skin gelatine) to create pellets. FTIR spectra were recorded within the range of 4000-400 cm^{-1} at a resolution of 4 cm^{-1}. Distinctive absorption peaks associated with amide bands I, II, and III were identified and compared across the different gelatine types. Each sample was analysed in triplicate to ensure reproducibility of the results.

X-Ray Fluorescence (XRF) Analysis

The elemental composition of the gelatine samples was determined using a Bruker S8 Tiger X-ray Fluorescence (XRF) spectrometer (Bruker AXS, Karlsruhe, Germany). Before being put in a dedicated container for XRF measurement, the gelatine samples were first

ground into a uniform powder. The intensity of distinctive X-ray emission peaks was used to identify the elements in the spectra, which were obtained over an energy range of 1-40 keV. Three technical replicates of each variety of gelatine were examined. A quantitative elemental composition profile was then obtained by processing and interpreting the data spectrum using Bruker S8 XRF software.

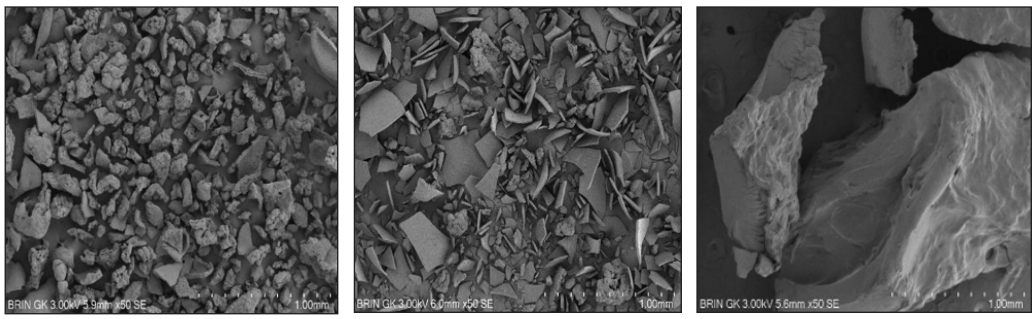
Statistical Analysis

The surface shape and element distribution shown in the scanning pictures were visually interpreted to carry out descriptive SEM and EDX analyses. To ascertain the functional group features of each sample, the FTIR spectra were studied by locating and interpreting the absorption peaks in the wavelength range of 3000-450 cm^{-1}. In the meantime, Principal Component Analysis (PCA) was carried out using Unscrambler X10.4 software (Camo, Norway) to find relationships and patterns in the elemental data, and Analysis of Variance (ANOVA) with IBM SPSS Statistics 25 software was used to statistically analyse the elemental composition data obtained from XRF analysis to assess significant differences between gelatine types.

RESULTS AND DISCUSSION

Comparative Scanning Electron Microscope (SEM) Analysis of Microstructural Properties in Gelatine Extracted from Various Sources

Significant morphological variations between gelatine from INCES, BCES, and bovine skin sources are revealed by the SEM measurements as shown in Figure 1. All things considered, SEM offers a micro-level visual representation of gelatine's physical structure, demonstrating its status as a naturally occurring polymer capable of forming intricate networks. In addition to the structure being somewhat porous or hollow, the finer structure of INCES gelatine is defined by small particles of uniform size, with each edge being blunt and loosely bound in the matrix. These morphological findings demonstrate that the collagen hydrolysis process in INCES gelatine is operating efficiently since it can be removed nearly flawlessly. This result supports the hypothesis that gelatine with smoother structures and smaller particle sizes tends to be more soluble (Li et al., 2023). The presence of porosity probably promotes faster dissolution by enabling improved water diffusion and swelling. These microstructural characteristics are linked to higher levels of collagen denaturation and triple-helical domain rupture, indicating that the extraction conditions used on INCES raw materials were adequate to break intermolecular cross-links. On the other hand, a lower degree of hydrolytic breakdown is implied by the relatively compact and coarse morphology seen in BCES and bovine skin gelatine, which could result in decreased solubility and functional performance.



(a) (b) (c)

Figure 1. SEM micrograph of gelatine with different sources: (a) Gelatine from Indonesian native chicken eggshell membrane (INCES), (b) Gelatine from broiler chicken eggshell membrane (BCES), (c) Gelatine from bovine skin

In contrast, BCES gelatine has a larger, rough, glass-like structure, is non-uniform, and non-porous. Furthermore, the particles remain strongly bound to the matrix, with irregular shapes, and each particle tip is pointed. This indicates incomplete collagen degradation during extraction, or it can be said that the collagen in the sample is not completely extracted. The cause of this problem is likely influenced by the characteristics of the raw material, which is different from the raw material for INCES gelatine, the solvent concentration used, and the soaking time during extraction, which is not long enough, so that the extraction process does not run optimally, and reduces its solubility in water. Previous research (Xu et al., 2025) reported that rough structures in gelatine could negatively impact its mechanical properties and solubility.

In contrast, bovine skin gelatine has a smooth structure with thin fibres and big, spherical aggregates. Higher levels of hydrolysis and a more regulated extraction process are reflected in this morphology, which produces a more uniform network. This structure supports the notion that, in comparison to poultry gelatine, bovine gelatine has historically been recognised for having more stable physicochemical characteristics (Karim and Bhat, 2009). Variations in solubility among the three gelatine types can be linked to residual protein content and the molecular interactions within their networks. The uniform morphology observed in INCES gelatine and bovine skin gelatine suggests that such structures facilitate more effective interactions with water. In contrast, the comparatively rough morphology of broiler chicken gelatine appears to restrict this capacity (Usman et al., 2024). Furthermore, drying or gelation pathways, temperature profiles, cooling rates, drying techniques, the presence of solvents or additives, bloom strength, and pH all have a significant impact on the morphological variations among gelatines. These factors together determine whether gelatine forms fibrous structures, solid films, porous matrices, or particulate morphologies.

Comparative Energy Dispersive X-Ray Spectroscopy (EDX) Analysis of Gelatine Derived from Different Sources

The elemental mapping patterns and elemental spectra of INCES, BCES, and bovine skin gelatine surface were revealed by analysis utilising Energy Dispersive X-ray Spectroscopy (EDX), as shown in Figure 2 and Table 1. Nine elements, carbon (C), nitrogen (N), oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), and calcium (Ca) were successfully identified in all samples using this method, which is based on the idea that each

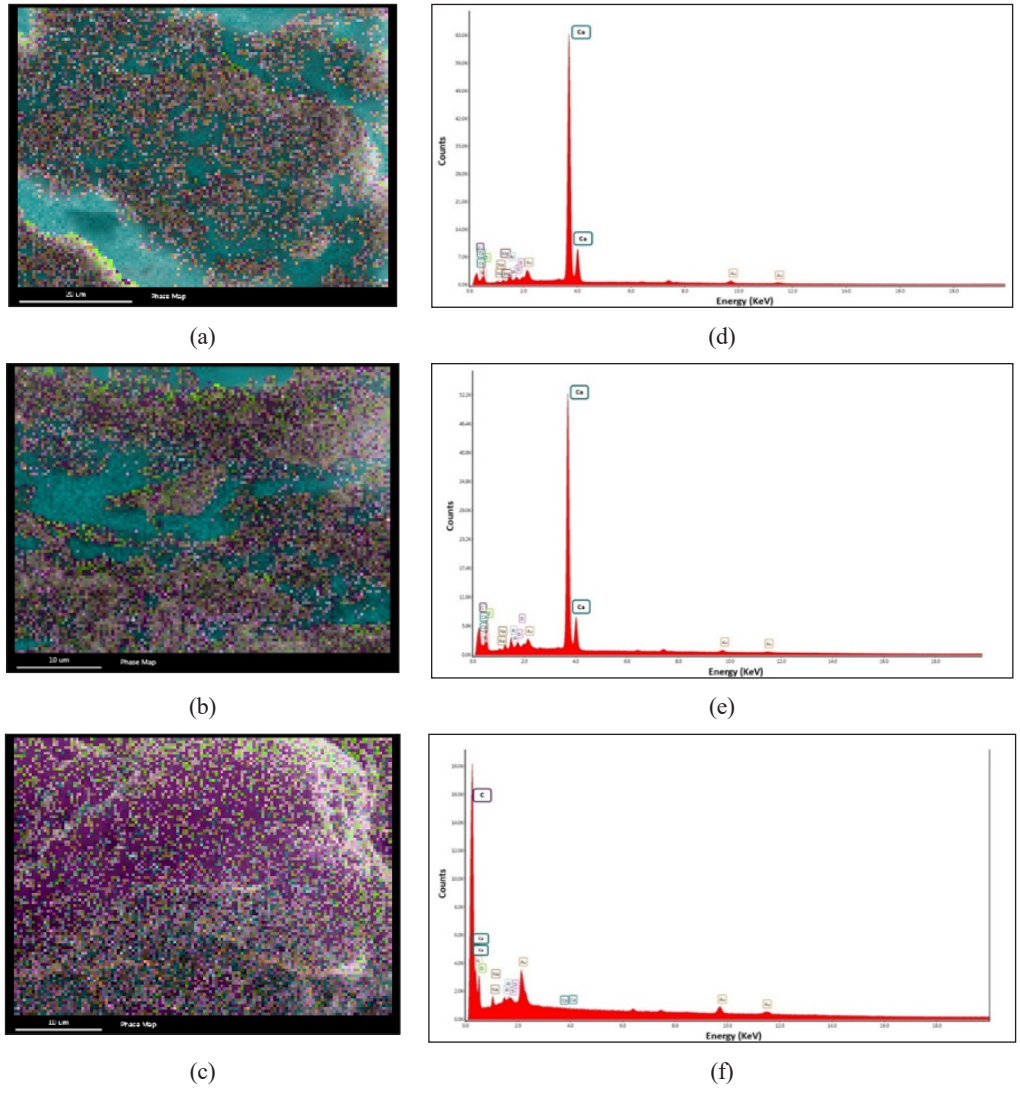


Figure 2. (a) EDX elemental mapping of gelatine from Indonesian native chicken eggshell membrane (INCES), (b) broiler chicken eggshell membrane (BCES), (c) bovine skin and (d) EDX spectrum of gelatine from INCES, (e) BCES, (f) bovine skin

Table 1
Elements composition by EDX analysis (%)

No.	Elements	Gelatin Source		
		INCES	BCES	Bovine Skin
1	Carbon (C)	3.06	5.15	47.83
2	Nitrogen (N)	1.02	3.09	10.87
3	Oxygen (O)	5.10	8.25	9.78
4	Natrium (N)	1.02	2.06	6.52
5	Magnesium (Mg)	2.04	0.00	0.00
6	Aluminium (Al)	4.08	4.12	8.70
7	Silicon (Si)	3.06	4.12	9.78
8	Calcium (Ca)	80.61	73.20	6.52

Note. INCES: Gelatine from Indonesian native chicken eggshell membrane; BCES: Gelatine from broiler chicken eggshell membrane (BCES)

element emits a distinctive emission spectrum according to its atomic configuration. While these elemental constituents are common to all gelatine types, the relative abundance of major elements varies considerably depending on the raw material source. Such differences suggest potential variations in chemical composition, which the intrinsic collagen structure and the extraction methodology employed may shape.

SEM-EDX analysis showed that calcium (Ca) was the predominant element detected on the surface of INCES and BCES gelatine, accounting for 79% and 71% of the detected surface elemental composition, respectively. In contrast, the commercial bovine-based gelatine displayed a markedly higher proportion of carbon (C), approximately 44%. The predominance of calcium in gelatine derived from chicken eggshell membranes may be associated with residual mineral components remaining after the demineralisation process. However, the present analytical techniques (SEM-EDX and XRF) do not allow differentiation between residual eggshell-derived calcium and calcium naturally associated with the eggshell membrane matrix. During bovine gelatine production, the demineralisation step is typically carried out intensively to lower mineral levels, with particular emphasis on reducing calcium content. However, in this study, the eggshell samples only underwent a relatively short demineralisation treatment (1 day), which may have contributed to the retention of certain mineral elements, particularly calcium. These findings suggest that the demineralisation process may influence the elemental composition of eggshell membrane-derived gelatine. Calcium contributes to the reinforcement of gelatine structure through ionic interactions, influencing key properties such as gel strength, viscosity, and melting point. By contrast, carbon represents the fundamental constituent of gelatine, exerting a profound impact on its molecular framework, physical attributes, and functional performance. Accounting for approximately 50.5% of the composition, carbon

is indispensable to the protein backbone of gelatine, and its unique gelling and binding capacities would not exist without this element.

This observation is consistent with the findings of Mohammadi et al. (2016), who reported that poultry-derived gelatine often retains elevated calcium levels due to suboptimal demineralisation. In contrast, the predominance of carbon in bovine gelatine agrees with the results of Karim and Bhat (2009), who demonstrated that bovine skin collagen contains a higher proportion of carbon. Similarly, Rigueto et al. (2022) emphasised that the mineral composition of gelatine is strongly influenced by both the raw material source and the processing method, thereby corroborating the elemental differences identified in this study. Such variations directly affect the functional properties and application potential of gelatine. The elevated calcium content observed in INCES and BCES gelatine suggests possible utility in biomedical applications; however, excessive mineral levels may compromise solubility and other physicochemical attributes, indicating the need for further optimisation if the material is to be applied in food or pharmaceutical contexts. Conversely, the high carbon content in bovine gelatine reflects its suitability for applications requiring a more stable organic matrix, such as film-forming material or gelling agent. Calcium contributes to the reinforcement of gelatine structure through ionic interactions, thereby influencing critical properties such as gel strength, viscosity, and melting point. In contrast, carbon constitutes the principal and indispensable element of gelatine, exerting a profound influence on its molecular framework, physical attributes, and functional performance. Representing approximately 50.5% of the composition, carbon is integral to the protein backbone of gelatine, and its distinctive gelling and binding capacities would not exist without this element.

Comparative Fourier Transform Infrared Spectroscopy (FTIR) Analysis of Gelatine Derived from Different Sources

The FTIR on gelatine serves to identify typical functional groups and characterise the molecular structure of gelatine. The spectra of INCES, BCES, and bovine skin gelatine are presented in the accompanying Figure 3 and Table 2. Each gelatine sample exhibited five characteristic amide regions: amide A, B, I, II, and III. Specifically, amide A was observed in the range of 3196.32-3276.61 cm^{-1} with a representative hydrogen bond stretched by NH, amide B in 2109.06-2112.36 cm^{-1} with a representative asymmetric stretching vibration of CH and $NH3^{+}$, amide I in 1602.86-1648.91 cm^{-1} with a representative COO is stretched by the C=O bond, amide II in 1527.54-1539.06 cm^{-1} with a representative the peptide group out-of-phase combination of in-plane NH deformation and CN stretch modes is responsible for the fluctuation mode, and amide III in 1410.41-1448.61 cm^{-1} with a representative variation in the glycine CH2 group or variation in the plane of the boned amide C, N, and h groups. Mohammadi et al. (2016) reported that the amide regions of

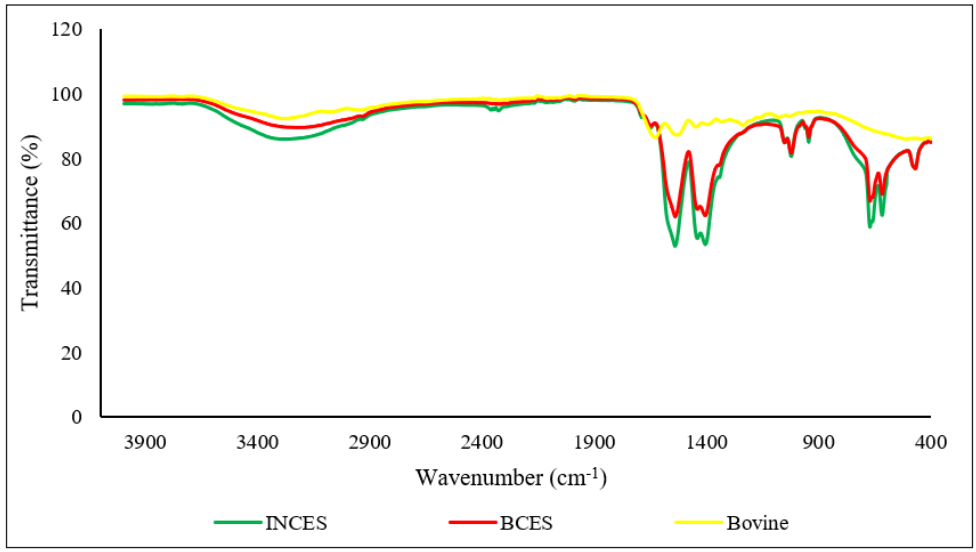


Figure 3. FTIR study of gelatin from bovine skin, broiler chicken eggshell membrane (BCES), and Indonesian native chicken eggshell membrane (INCES)

gelatine spectra from eggshell membrane appeared at 3385.9 cm^{-1} (amide A), 2946.2 cm^{-1} (amide B), 1654.1 cm^{-1} (amide I), 1552.5 cm^{-1} (amide II), and 1239.1 cm^{-1} (amide III). According to Pulidori et al. (2023), the amide I region (1600-1700 cm^{-1}) corresponds to the secondary structure of gelatine. Within this spectral region, peaks at 1610-1642 cm^{-1} are indicative of β-sheet structures, while those at 1642–1650 cm^{-1} correspond to random coil conformations. Absorptions between 1650–1660 cm^{-1} reflect α-helix structures, peaks at 1660–1680 cm^{-1} denote β-turns, and those observed at 1680–1700 cm^{-1} represent β-antiparallel arrangements. Accordingly, the distribution of amide I bands provides meaningful insight into the relative abundance of secondary structures in INCES, BCES, and bovine skin gelatine, thereby elucidating their functional and physicochemical properties. These FTIR findings are supported by previous SDS-PAGE analysis of collagen extracted from the same INCES raw material, which identified the characteristic α1, α2, β, and γ chains of type I collagen (Febrisiantosa et al., 2024). This complementary evidence confirms the molecular characteristics of the collagen precursor from which the gelatine was derived. However, the functional characterisation is highly recommended to validate the relationship between these functional groups and the mechanical strength of the gelatine

From these observations, it can be concluded that all three samples exhibit secondary structures characteristic of gelatine. β-sheet conformations dominate the INCES and bovine gelatine samples, whereas the BCES sample is characterised mainly by a random coil structure. Gelatine, a hydrolysed form of collagen, possesses β-sheet structures that impart mechanical strength, structural stability, and pronounced gel-forming capacity.

Table 2
The Amida region of the peak wave number (cm^{-1}) of gelatine source

Amida Region	Gelatine SourceS			Assignment
	INCES	BCES	Bovine skin	
A	3196.32	3227.47	3276.61	Hydrogen bond stretched by NH
B	2112.36	2112.12	2109.06	Asymmetric stretching vibration of CH and $NH3^{+}$
I	1602.86	1648.91	1628.20	COO is stretched by the C=O bond
II	1527.54	1539.06	1532.46	The peptide group out-of-phase combination of in-plane NH deformation and CN stretch modes is responsible for the fluctuation mode
III	1410.41	1440.65	1448.61	Variation in the glycine CH2 group or variation in the plane of the boned amide C, N, and h groups

Note. INCES: Gelatine from Indonesian native chicken eggshell membrane; BCES: Gelatine from broiler chicken eggshell membrane (BCES)

These attributes underpin its widespread application as a tablet binder, capsule former, emulsifier, and film former across the food and pharmaceutical industries. In addition, the random coil conformation of gelatine provides a molecular arrangement that facilitates polypeptide aggregation and interaction, a process essential for reversible gel formation. This structural flexibility enables gelatine molecules to adapt their shape and establish interactions that promote gelation under varying temperature conditions.

Raw material characteristics and extraction parameters, including temperature, heating duration, and acid or alkali treatment, strongly influence variations in gelatine structure. These factors determine the extent of collagen chain degradation, average molecular weight, and amino acid composition. Consequently, the molecular architecture shaped by raw material composition and processing conditions governs key physical properties such as gel strength, viscosity, and solubility. For BCES gelatine, the eggshell membrane was separated from the shell through heating, whereas no comparable pre-treatment was applied to BCES and bovine skin gelatine. Exposure to temperatures exceeding 60 °C likely induced protein denaturation, disrupting certain bonds and thereby modifying the secondary structure of gelatine. This observation is consistent with earlier reports demonstrating that thermal treatment exerts a significant influence on the molecular arrangement and secondary structure of gelatine (Karim & Bhat, 2009).

Comparative X-Ray Fluorescence (XRF) analysis of gelatine derived from different sources, one-way analysis of variance (ANOVA)

The results of the ANOVA analysis showed that gelatine from various sources (INCES, BCES and bovine skin) had significantly different mineral element profiles as shown in Table 3. The calcium content (Ca) in INCES and BCES gelatine was similar (around 50.36%) and much higher than bovine (0.13 %; $p<0.05$), indicating that gelatine from

Table 3
Macro and micromineral composition by XRF analysis

Elements	Gelatine Source		
	INCES	**BCES**	**Bovine Skin**
Micromineral			
Mg (%)	0.50±0.01a	0.24±0.13b	nd
P (%)	0.51±0.01a	0.45±0.08a	0.24±0.04b
S (%)	0.12±0.01b	0.06±0.01b	0.95±0.09a
Ca (%)	50.36±0.04a	50.36±1.19a	0.13±0.01b
Micromineral			
Cl (ppm)	114.33±0.01c	371.05±37.55b	596.80±74.80a
Fe (ppm)	402.50±0.10a	276.65±0.35b	122.90±16.81c
Zn (ppm)	450.25±4.35a	439.45±3.85b	9.27±3.58c

Note. INCES: Gelatine from Indonesian native chicken eggshell membrane; BCES: Gelatine from broiler chicken eggshell membrane (BCES); nd: not detected; Notations without similar letters indicate significant differences ($p < 0.05$)

bovine skin underwent a more effective demineralisation process during the raw material preparation stage. This is due to the theory that the demineralisation process in gelatine production is often more intensive in bovine skin raw materials, considering that its denser collagen structure requires longer chemical treatment to soften the collagen fibres (Amalia et al., 2023). In contrast, the high calcium levels detected in INCES and BCES gelatine suggest that demineralisation of poultry skin was incomplete, most likely due to variations in collagen density, tissue composition, and processing parameters between poultry and bovine sources.

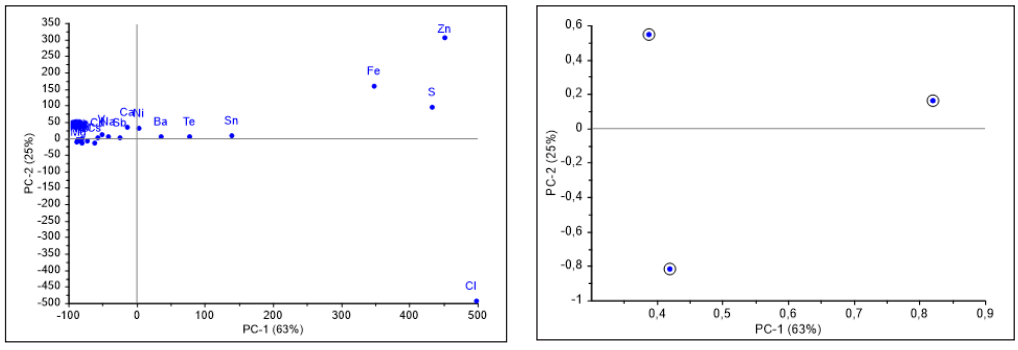
The iron (Fe) concentration in INCES gelatine (402.50 ppm) was markedly higher than that observed in BCES (276.65 ppm) and bovine gelatine (122.90 ppm) ($p < 0.05$). This elevated level is likely attributable to residual haemoglobin or tissue fragments from poultry skin that were not fully eliminated during processing. Arnesen and Gildberg (2007) similarly noted that incomplete removal of residual tissues during extraction may result in increased iron content in gelatine. In contrast, bovine gelatine exhibited the highest sulphur (S) concentration (0.95 ppm), compared with INCES (0.12 ppm) and BCES (0.06 ppm) ($p < 0.05$). This suggests a higher presence of keratin or sulphur-containing proteins, which are known constituents of bovine skin. Mariod and Adam (2013) reported that sulphur compounds originating from keratin may affect the mechanical behaviour of gelatine, particularly its elasticity and gel-forming capacity. In addition, the relatively high zinc (Zn) concentrations in INCES (450.25 ppm) and BCES gelatine (439.45 ppm) indicate the physiological role of zinc as a structural component of the skin and supporting tissues. In contrast, the lower zinc content observed in bovine gelatine (9.27 ppm) reflects

a more efficient mineral removal process. According to Jongjareonrak et al. (2006), zinc plays a pivotal role in sustaining collagen stability, which subsequently governs the physicochemical attributes of gelatine, notably gel strength and viscosity.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is primarily employed to reduce the dimensionality of gelatine XRF data and to identify distinctive patterns or features that facilitate the differentiation or classification of gelatine sources. By projecting the data onto two or three principal components, scatterplot visualisation reveals natural groupings and trends among the samples. As shown in Figure 4, PC-1 accounts for 63% of the total variation, while PC-2 explains 25%, together representing 88% of the overall variability. These results indicate that the first two principal components captured most of the variation in the XRF dataset, providing an effective representation of the elemental composition among the gelatine samples. The PCA results showed differences in the characteristics of the elements in gelatine derived from INCES, BCES, and bovine skin. This analysis also revealed the influence of certain elements, such as Cl, Zn, and S, contributing to the differentiation of gelatine sources, as described in the scores plot (Figure 4a) and loading plot (Figure 4b). In the INCES gelatine sample, chlorine (Cl) was identified as the dominant element, contributing substantially to the variability captured by PC-2.

The occurrence of Cl is commonly associated with inorganic compounds such as calcium chloride, which may form during the extraction or processing of eggshell-derived gelatine (Huang et al., 2021). Its presence may also reflect the use of an acidic extraction medium, such as acetic acid, in this study. Comparable findings were reported by Mohammadi et al. (2016), who observed that incomplete demineralisation of eggshells can result in gelatine with relatively high residual mineral content. In the case of BCES gelatine, Zn and sulphur S were the principal contributors to PC-1. Zinc in gelatine typically



(a) (b)

Figure 4. (a) Elemental scores plot based on PCA; (b) Sample loadings plot based on PCA for gelatine types

originates from bone or connective tissue components (Liu et al., 2021), while Sarbon et al. (2013) noted that poultry-derived gelatine generally contains higher concentrations of essential elements compared with mammalian sources. The presence of sulphur reflects sulphur-containing amino acids, such as cysteine, which are characteristic of poultry-based gelatine. This suggests that BCES gelatine may have a protein structure enriched in sulphur-containing residues relative to the other samples.

Bovine skin gelatine contains trace elements such as tin (Sn) and tellurium (Te), which, although not predominant, distinguish it from other gelatine types. The presence of these elements may be associated with specific commercial processing methods or inherent characteristics of bovine skin. Karim and Bhat (2009) noted that bovine gelatine typically exhibits lower mineral content than gelatine derived from bones or eggshells, a difference attributed to the more intensive demineralisation procedures used during its production. These findings align with the theory of gelatine structure and composition, which states that the presence of trace mineral elements is strongly influenced by both the raw material source and the extraction process employed (Liu et al., 2021; Mohammadi et al., 2016). The presence of elements such as Zn and Cl highlights the potential functional role of gelatine as a supplementary mineral source, particularly in nutritional and biomedical contexts. Similarly, S, which is linked to protein structure via sulphur-containing amino acids, can influence the physicochemical properties of gelatine, including solubility and viscosity attributes essential for its industrial applications (Muyonga et al., 2004).

Differences in elemental composition among the gelatine types suggest their potential suitability for distinct application areas as shown in Figure 3. Bovine skin gelatine, with its relatively high Cl content, may possess functional properties advantageous for biomedical formulations. Conversely, INCES gelatine, characterised by elevated Zn and sulphur levels, appears more appropriate for nutritional or functional food applications, given the essential roles of these elements in human health. Overall, bovine skin gelatine, along with certain minor components, exhibits stable properties that support its use across diverse industrial sectors. Statistical analysis further confirmed significant variations in the mineral profiles of INCES, BCES, and bovine skin gelatine. Notably, the mineral composition of INCES and BCES gelatine was superior to that of bovine gelatine, underscoring their promise as candidates for targeted applications in functional foods and pharmaceutical products, where trace mineral content can influence product quality and functionality (Liu et al., 2009).

ANOVA analysis demonstrated significant differences in the mineral element profiles of gelatine obtained from INCEM, BCEM, and bovine skin. These results were further supported by PCA, which highlighted Fe, Ca, and Zn as the principal elements differentiating the gelatine sources. Collectively, the more mineral-rich compositions of INCEM and BCEM gelatine relative to bovine gelatine underscore their potential for specialised applications, particularly in functional foods and pharmaceutical products, where trace elements play a critical role in determining product quality and functionality.

CONCLUSION

The study establishes that gelatine produced from Indonesian native chicken eggshell membranes exhibits distinct morphological, elemental, and structural features, reinforcing its potential as a sustainable substitute for traditional gelatine sources. SEM observations revealed that INCES gelatine contained finer and more homogeneous particles than BCES and bovine skin gelatine. Consistently, EDX analysis detected higher calcium levels in INCES and BCES, attributable to incomplete demineralisation during processing. FTIR spectra indicated that INCES and bovine gelatines predominantly displayed β-sheet structures, while BCES gelatine was dominated by random coil conformations. XRF analysis confirmed significant calcium enrichment in INCES gelatine, emphasising its unique mineral composition. Overall, these results underscore the promise of INCES-derived gelatine as a functional biomaterial, while simultaneously supporting the valorisation of eggshell membrane waste and the development of ethical, sustainable gelatine alternatives.

ACKNOWLEDGEMENT

We would like to express our appreciation to the National Research and Innovation Agency (BRIN), Indonesia, Research Organisation for Agriculture and Food, under contract number 9/III.11/HK/2023, for their significant assistance and for creating the cooperative framework that made this study possible.

REFERENCES

Al-Hassan, A. A., & Norziah, M. H. (2017). Effect of transglutaminase induced crosslinking on the properties of starch/gelatine films. *Food Packaging and Shelf Life, 13*, 15-19. https://doi.org/10.1016/j.fpsl.2017.04.006

Allegretta, I., Marangoni, B., Manzari, P., Porfido, C., Terzano, R., De Pascale, O., & Senesi, G. S. (2020). Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA), and machine learning algorithms. *Talanta, 212*, Article 120785. https://doi.org/10.1016/j.talanta.2020.120785

Amalia, R., Ulfah, L. A. M., Paramita, V., Kusumayanti, H., Dhiya'ulhaq, S. B., Maharani, N. A., & Amalia, N. P. (2023). Effect of demineralisation on gelatin production from grasshopper (*Valanga nigricornis*) as an alternative to halal gelatin. In R. Isnanto, Hadiyanto, & B. Warsito (Eds.), *E3S Web of Conferences, 448*, Article 01013. https://doi.org/10.1051/e3sconf/202344801013

Appoloni, C. R., & Melquiades, F. L. (2014). Portable XRF and principal component analysis for bill characterisation in forensic science. *Applied Radiation and Isotopes, 85*, 92-95. https://doi.org/10.1016/j.apradiso.2013.12.004

Bootklad, M., & Kaewtatip, K. (2013). Biodegradation of thermoplastic starch/eggshell powder composites. *Carbohydrate Polymers, 97*(2), 315-320. https://doi.org/10.1016/j.carbpol.2013.05.030

Gómez-Guillén, M. C., Giménez, B., López-Caballero, M. E., & Montero, M. P. (2011). Functional and bioactive properties of collagen and gelatin from alternative sources: A review. *Food Hydrocolloids, 25*(8), 1813-1827. https://doi.org/10.1016/j.foodhyd.2011.02.007

Hamidi, A. A., Salimi, M. N., & Yusoff, A. H. M. (2017). Synthesis and characterisation of eggshell-derived hydroxyapatite via mechanochemical method: A comparative study. *AIP Conference Proceedings, 1835*, Article 020045. https://doi.org/10.1063/1.4981867

Herdiana, Y., Sofian, F. F., Shamsuddin, S., & Rusdiana, T. (2024). Towards halal pharmaceuticals: Exploring alternatives to animal-based ingredients. *Heliyon, 10*(1), Article e23624. https://doi.org/10.1016/j.heliyon.2023.e23624

Hoerterer, C., Petereit, J., Lannig, G., Johansen, J., Pereira, G. V., Conceição, L. E. C., Pastres, R., & Buck, B. H. (2022). Sustainable fish feeds: Potential of emerging protein sources in diets for juvenile turbot (*Scophthalmus maximus*) in RAS. *Aquaculture International, 30*(3), 1481-1504. https://doi.org/10.1007/s10499-022-00859-x

Jongjareonrak, A., Benjakul, S., Visessanguan, W., & Tanaka, M. (2006). Skin gelatin from bigeye snapper and brownstripe red snapper: Chemical compositions and effect of microbial transglutaminase on gel properties. *Food Hydrocolloids, 20*(8), 1216-1222. https://doi.org/10.1016/j.foodhyd.2006.01.006

Karim, A. A., & Bhat, R. (2009). Fish gelatin: Properties, challenges, and prospects as an alternative to mammalian gelatins. *Food Hydrocolloids, 23*(3), 563–576. https://doi.org/10.1016/j.foodhyd.2008.07.002

Li, X., Sha, X. M., Yang, H. S., Ren, Z. Y., & Tu, Z. C. (2023). Ultrasonic treatment regulates the properties of gelatin emulsion to obtain high-quality gelatin film. *Food Chemistry: X, 18*, Article 100673. https://doi.org/10.1016/j.fochx.2023.100673

Liu, J., Yao, R., Guo, J., Gao, T., He, J., Meng, G., & Wu, F. (2021). The regulating effect of trace elements Si, Zn, and Sr on the mineralisation of gelatin-hydroxyapatite electrospun fibre. *Colloids and Surfaces B: Biointerfaces, 204*, Article 111822. https://doi.org/10.1016/j.colsurfb.2021.111822

Mohammadi, R., Mohammadifar, M. A., Mortazavian, A. M., Rouhi, M., Ghasemi, J. B., & Delshadian, Z. (2016). Extraction optimisation of pepsin-soluble collagen from eggshell membrane by response surface methodology (RSM). *Food Chemistry, 190*, 186-193. https://doi.org/10.1016/j.foodchem.2015.05.073

Permadi, S., Ujilestari, T., Hakim, L., Jambang, N., Yuliatmo, R., Febrisiantosa, A., Wahyono, T., & Wahyuningsih, R. (2024). Characteristics and applications of collagen from animal by-products as a potential source for food ingredients. *Reviews in Agricultural Science, 12*, 327-346. https://doi.org/10.7831/ras.12.0_327

Proietti, N., Di Tullio, V., Carsote, C., & Badea, E. (2020). 13C solid-state NMR complemented by ATR-FTIR and micro-DSC to study modern collagen-based material and historical leather. *Magnetic Resonance in Chemistry, 58*(9), 840-859. https://doi.org/10.1002/mrc.5024

Pulidori, E., Micalizzi, S., Koutsomarkos, N., Bramanti, E., Tinè, M. R., Vozzi, G., De Maria, C., Chatzinikolaidou, M., & Duce, C. (2023). Analysis of gelatin secondary structure in gelatin/keratin-based biomaterials. *Journal of Molecular Structure, 1279*, Article 134984. https://doi.org/10.1016/j.molstruc.2023.134984

Rosa, J., Batista de Carvalho, L. A. E., Marques, M. P. M., Ferreira, M. T., Gonçalves, D., & Gil, F. P. S. C. (2024). XRF identification of sharp-force trauma in fresh and dry human bone under varied experimental heat conditions. *Science & Justice, 64*(3), 305-313. https://doi.org/10.1016/j.scijus.2024.04.005

Wahyuningsih, R., Sebayang, K. G., & Nurcholis, M. (2026). Effect of Acetic and Citric Acids on Gelatin Characteristics Extracted from Chicken Eggshell Membranes. *IOP Conference Series: Earth and Environmental Science*, *1584*(1). https://doi.org/10.1088/1755-1315/1584/1/012072

Xu, X., Xi, Y., & Weng, Y. (2025). Gelatin-based materials: fabrication, properties and applications in the food packaging system. In *RSC Advances* (Vol. 15, Number 37, pp. 30605-30621). Royal Society of Chemistry. https://doi.org/10.1039/d5ra03325j

Yuliatmo, R., Fitriyanto, N. A., Bachruddin, Z., Supadmo, Jamhari, & Erwanto, Y. (2017). Increasing of angiotensin converting enzyme inhibitory derived from Indonesian native chicken leg protein using Bacillus cereus protease enzyme. *International Food Research Journal*, *24*(4), 1799-1804. https://www.researchgate.net/publication/319356373

Pertanika J. Sci. & Technol. 34 (4): 2329 - 2337 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Short Communication

Regional Comparison of Forest Certification Impacts: Environmental and Social Outcomes in Asia-Pacific and Europe

Rohaiza Abdul Kadir Zailani, Sarah Najwa Mohd Masri, Shah Badri Mohd Nor, Ahmad Fariz Mohamed*, and Shaharuddin Mohamad Ismail

Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia (UKM), 43600 UKM Bangi, Selangor, Malaysia

ABSTRACT

Forest certification (notably FSC, PEFC and national schemes such as MTCS, IFCC, CFCC, SGEC) is a voluntary instrument designed to verify and promote sustainable forest management across economic, environmental, social and governance dimensions. This manuscript draws exclusively on the supplied literature review (Phase 1 Chapter 2) to produce a regional comparative assessment of environmental and social impacts of forest certification in the Asia-Pacific and Europe regions. Using the original document's thematic synthesis and its stated thematic analysis approach, this paper assesses environmental impacts (e.g. RIL adoption, HCVF identification, tree-cover dynamics) and social impacts (e.g. community engagement, livelihoods, worker safety, indigenous rights), presents results organised by region and scope, and identifies cross-cutting challenges and practical recommendations. Key findings show that Asia-Pacific case studies report clearer operational environmental gains and community capacity benefits in certain cases, while Europe shows stronger governance and social protections with more mixed ecological outcomes. Cross-cutting barriers include certification costs, weak premium signals, capacity gaps and variability in auditing enforcement. All claims are drawn from the supplied literature review.

Keywords: Asia-Pacific, environmental impacts, Europe, forest certification, social impacts, sustainable forest management

ARTICLE INFO

Article history:
Received: 03 April 2026
Accepted: 15 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.18

E-mail addresses:
rohaizaakz@yahoo.com (Rohaiza Abdul Kadir Zailani)
p116358@siswa.ukm.edu.my (Sarah Najwa Mohd Masri1)
shahbadrinor@gmail.com (Shah Badri Mohd Nor)
fariz@ukm.edu.my (Ahmad Fariz Mohamed)
shaharuddinmdi@gmail.com (Shaharuddin Mohamad Ismail)
* Corresponding author

INTRODUCTION

Forest certification began as a market-driven response to accelerating forest loss and rising public concern about how forests were managed worldwide; after

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

a series of international policy efforts in the late 1980s and early 1990s (including FAO processes and the 1992 UNCED Earth Summit), non-governmental actors sought stronger, independent ways to verify sustainable forest management and to channel consumer demand toward better practice, culminating in the establishment of the Forest Stewardship Council (FSC) in 1993 and the later emergence of the Programme for the Endorsement of Forest Certification (PEFC) in 1999 as an umbrella/endorsement mechanism for national schemes (Phase 1 Chapter 2 Literature Review). Over the following decades, two broad strands of development shaped the practice and spread of certification: international schemes (FSC, PEFC) that set global benchmarks and national schemes that adapted standards to local legal and social contexts (for example, MTCC/MTCS in Malaysia, IFCC in Indonesia, CFCC in China, SGEC in Japan), and a growing set of empirical studies and audit reports that evaluated certification's economic, environmental and social effects. The theory of change embedded in these instruments is that external standards plus independent verification will encourage better planning and operational forest practices (e.g., reduced-impact logging, improved road and skid-trail design, HCVF identification), greater stakeholder participation and improved worker safety, and, where market signals exist, better market access or price premiums for certified products. Yet the empirical evidence assembled in the Phase 1 literature review shows results are heterogeneous across regions: Asia-Pacific tropical cases frequently record visible operational environmental improvements (RIL adoption, HCVF demarcation and, in some studies, reduced tree-cover loss) and social gains through group certification and training, while many European cases demonstrate clearer governance and social outcomes (worker safety, indigenous recognition, enhanced transparency) but more mixed, context-dependent ecological gains. This heterogeneity, together with persistent barriers such as certification and chain-of-custody costs, weak or inconsistent premium markets, uneven corrective-action enforcement and smallholder capacity gaps motivate a focused, regionally comparative analysis of forest certification's environmental and social scopes. The underlying conceptual model of this mechanism posits that certification standards enforced through third-party audits compel forest managers to adopt improved practices, which in turn generate measurable environmental and social outcomes along specific causal pathways (Rametsteiner & Simula, 2003; Romero et al., 2013). However, outcomes remain highly heterogeneous across different geographical and regulatory contexts. Therefore, this study aims to explicitly evaluate these disparities. Specifically, the objectives of this regionally comparative analysis are to: (1) evaluate the extent of operational and ecological improvements (e.g., RIL adoption, HCVF demarcation, tree-cover dynamics); (2) compare the trajectory of social and governance outcomes

(e.g., community capacity building, worker rights, stakeholder transparency); and (3) identify cross-cutting barriers to certification efficacy across the Asia-Pacific and Europe regions. By doing so, this paper seeks to highlight where certification delivers tangible additionality and what policy implications arise from these regional contrasts.

METHODS AND ANALYSIS

The study focuses on the review of organised studies by region with four thematic areas, namely Economic, Environment, Social and Governance, as the main thematic subject. Following this orientation, the present study used thematic analysis to extract, code and synthesise evidence relevant to environmental and social outcomes for Asia-Pacific and Europe. The thematic analysis approach followed established qualitative practice and draws on seminal methodological guidance (Braun & Clarke, 2006; Guest et al., 2012; Nowell et al., 2017). The analytic steps were conducted through familiarisation analysis by systematic reading of regional syntheses and case studies. The coding was conducted with the need to identify and label repeated claims and empirical descriptions. The theme development was established by grouping related codes under higher-order headings such as RIL/Operational, HCVF/Biodiversity, tree-cover outcomes, group certification/ community capacity and worker/indigenous rights. Review and refinement were conducted by cross-checking claims against the review summaries to avoid overclaiming. After data analysis was completed, a description analysis and synthesis into comparative narratives and tables were prepared. To ensure traceability, documented code definitions and retained exemplar paraphrases from the review for each theme were established. The thematic Table 1 summarises the primary codes, the document sections where evidence was extracted, the analytic steps used to establish the code, and an example claim drawn from the coded passages. While this study relies fundamentally on structuring its thematic synthesis, it is necessary to acknowledge the methodological parameters and potential biases inherent in this approach. The supplied review primarily included peer-reviewed empirical studies, audit reports, and case comparisons explicitly addressing certification impacts, while excluding non-verified literature. Nevertheless, relying exclusively on a single secondary synthesis introduces a risk of perpetuating publication bias and limits the independent verification of primary data quality (Blackman & Rivera, 2011). To mitigate this limitation and to deepen the critical assessment of the findings, this manuscript integrates supplementary primary literature in its discussion to cross-verify the regional trends identified in the core document.

Table 1
Summary of thematic analysis

Thematic code	Source(s) in the supplied document (where evidence was extracted)	How the theme/code was developed (analytic steps)	Example claim drawn from the coded passages
Reduced Impact Logging (RIL) & operational practice improvements	Sections describing field innovations and RIL (Deramakot case, Logfisher equipment, training statistics).	Read Asia-Pacific case descriptions; flagged mechanisation, road/skid design and harvesting methods; coded inductively under 'RIL/Operational' and compared across FMUs for recurrence.	Malaysia cases show RIL adoption and novel extraction machines (Logfisher); training numbers recorded (270 trainers, 726 workers).
High Conservation Value Forest (HCVF) / biodiversity measures	HCVF toolkit and state circular references, biodiversity outcomes sections.	Extracted explicit mentions of HCVF, demarcation and conservation set asides; grouped with biodiversity monitoring and fauna studies.	HCVF identification and demarcation operationalised in Malaysia and Indonesia; biodiversity protection cited across case studies.
Tree-cover / deforestation results	Empirical study summaries (Miteva et al., 2015; Rana & Sills, 2018) and summarised quantitative claims.	Located and flagged comparative empirical claims about tree-cover loss, cross-checked author statements and the review's summary.	Some tropical FMUs showed reduced tree-cover loss (e.g. an estimated ~5% reduction in Kalimantan vs non-certified concessions).
Group certification, training & community capacity	Vietnam smallholder group certification, Nepal FUGs, and Australian plantation engagement sections.	Extracted passages describing group certification projects and training outcomes; coded under 'Smallholder / Community' and aggregated reported benefits and caveats.	Group certification in Vietnam and Indonesia reported improved skills and some income benefits; Nepal FUGs showed enhanced planning but mixed income effects.
Worker safety, indigenous rights & social governance	CAR analyses, European country case studies (Sweden, Estonia, Russia), and social outcomes tables.	Identified social-governance passages focusing on labour conditions, indigenous consultation and transparency; grouped under 'Worker / Indigenous / Governance'.	European cases report improved occupational health & safety and greater recognition of indigenous rights; publicised management summaries increased transparency.
Audit variability & CAR enforcement	CAR/audit sections, case comparisons showing differences among CBs and recommended corrections.	Flagged references to corrective actions, auditor stringency and CB variation; coded as a cross-cutting governance barrier.	Several FMUs experienced uneven CAR enforcement; if CBs demanded conservative silviculture, AAC reductions were greater.
Cost & market premium barriers	Economic sections, market premium examples (Sabah price studies, Japan, Indonesia).	Extracted economic claims about premiums, costs and CoC adoption; coded as 'Market / Cost' with documented price examples.	Evidence is mixed: Sabah logs fetched premiums (27-56%), Indonesia reported ~15% premium in one study, while Japan showed little premium effect.

RESULTS AND DISCUSSION

Environmental Outcomes

Across Asia-Pacific and Europe, certification has led to measurable and meaningful changes in forest management, though the character and visibility of those changes differ by region and context. In the Asia-Pacific, many of the documented improvements are operational and immediate: certification has encouraged adoption of Reduced-Impact Logging techniques, better road and skid-trail planning, and practical equipment choices that reduce soil disturbance and residual stand damage. The Deramakot example and the introduction of the Logfisher excavator illustrate how certification-driven attention to mechanisation and extraction techniques can translate into practice: managers and workers invested in training (the review records hundreds of trained personnel) and in retooling operations to limit ground disturbance and protect riparian buffers. Those operational improvements often feed directly into conservation outcomes: the review highlights HCVF identification and demarcation as a widespread innovation in tropical production forests, with Malaysia producing an HCVF toolkit and state circulars to anchor field practice. The literature collated in the review further reports empirical comparisons suggesting that in some tropical concessions certification has correlated with lower tree-cover loss, where remote-sensing or matched FMU studies are cited (for example, Miteva et al., 2015), and later replications, FSC certification was associated in select cases with a detectable reduction in deforestation and a suite of community benefits such as reduced disease incidence and lower fuelwood dependence for nearby villages. These results are not universal; case audits and CAR summaries contained in the document show that inconsistent follow-through, remaining weaknesses in road construction practices and variable local capacity sometimes limit the ecological improvements on the ground.

The European evidence in the review paints a more complex picture in ecological terms. In many parts of Europe, certification exists on a background of comparatively strong national regulation, where legal baselines already require careful silviculture and protection measures, and the marginal ecological effect of certification can be smaller. Nonetheless, the review shows instances where certification spurred concrete biodiversity or structural improvements; examples include increased retention trees, better riparian buffers, and designated set-asides to improve habitat connectivity in selected countries. At the same time, multiple studies synthesised in the review report no statistically significant difference in some ecological indicators between certified and non-certified owners for certain contexts; the interpretation the review offers are instructive: certification's added ecological value frequently depends on the standard's ambition and the robustness of monitoring and enforcement. If certifying bodies apply rigorous audits and enforce corrective actions, ecological gains are more likely; where audit follow-up is uneven or

where certification merely replicates existing legal minimum, measurable environmental improvement at the landscape scale can be limited.

Social Outcomes

The supplied document shows that social outcomes associated with certification are real and often more consistent across regions than ecological outcomes, although their precise form varies. In Asia-Pacific group certification, structured training, and community engagement programs recurring in the review's case studies have increased local capacity, planning skills and, in many instances, community participation in forest governance. The Vietnam smallholder group certification example and Nepal FUG studies illustrate the pathways: training on thinning, pruning and pre-harvest planning, together with the social processes inherent to group schemes, raised awareness, improved technical competence and bolstered local institutional capacity. Importantly, the review notes that capacity and participation gains do not automatically translate into higher household incomes in every case; evidence from Nepal and other community forestry groups shows that while planning and participation improved, income effects could be modest or context dependent. Where group certification was combined with stronger market access or product demand, however, positive economic outcomes were reported. The review's attention to non-timber forest products and to forest ecosystem services also suggests social pathways beyond timber: if certification can be adapted to include high-value NTFP or ecosystem-service certification, communities may access alternative livelihood streams, though the document cautions that such applications need careful feasibility work and market development.

In Europe, the document reports consistent social and governance benefits associated with certification, especially in worker health and safety, formal labour conditions and recognition of indigenous and local rights. CAR analyses and country case reviews indicate that certification often prompts better record-keeping, clearer management plans and more transparent publication of management summaries, which in turn strengthen stakeholder oversight and public trust. Several European examples in the review describe improved wages, safer working conditions and enhanced consultation processes for indigenous groups. Yet the review also points to participation limitations in some national contexts where cultural and institutional barriers reduce the depth of stakeholder engagement, and to the fact that certification's institutional improvements require sustained attention: transparency and formalisation often improve working conditions but do not, on their own, resolve underlying inequities or guarantee inclusive decision-making. The social evidence, therefore, suggests that certification can be a powerful governance tool, improving workplace standards, institutional transparency and some elements of social inclusion, while its distributive economic and livelihood impacts depend on linkage to market access, capacity building and locally appropriate design.

Table 2 consolidates, side-by-side, summarising the most load-bearing environmental and social findings from the Asia-Pacific and Europe sections of the supplied document. Each claim includes the corresponding citation marker to the source document sections or case studies.

Table 2
Environmental and social findings from Asia-Pacific and Europe

Theme / Claim	Asia-Pacific-Key Findings (supporting citation)	Europe -Key Findings (supporting citation)
Operational forest practices	Adoption of RIL, improved road/skid design, and machinery (e.g. Logfisher) with training for workers (270 trainers; 726 workers trained as of 2016).	Some improvements in retention tree counts and riparian buffer widths in regeneration areas (Norway), but uneven compliance in many regeneration units; improvement depends on enforcement.
HCVF/ biodiversity	HCVF identification & management advanced (toolkits, state circulars in Malaysia; HCVF uptake in Indonesia). Evidence of biodiversity conservation (e.g. Deramakot).	Mixed some set-aside areas for connectivity and structural richness; small private owners showed evidence of environmental improvements, while some large properties showed little extra conservation gain. National law often sets the baseline.
Tree-cover / deforestation outcomes	Empirical findings: FSC associated with reduced tree-cover loss (Miteva et al., 2015): ~5% reduced deforestation in Kalimantan; Rana and Sills (2018) in Indonesia.	Mixed results: Nikolaeva et al. (2019) from Primorsky Krai and Villalobos et al. (2018) from Sweden found no significant difference in cover loss/ gain between certified and non-certified concessions; the effect depends on standards and monitoring.
Community engagement & capacity	Group certification yields training, improved planning and higher technical skills for smallholders (Vietnam, Nepal, Australian plantation cases). Increased sub-group participation.	Certification integrated with public processes increased transparency (e.g. management summaries in Estonia) and consultations; variable success in stakeholder engagement (Italy showed limited participation).
Livelihood & income impacts	Mixed: while certain community forestry groups report improved employment opportunities and diversified income streams, other models (such as FUGs in Nepal) show limited direct income changes despite enhanced planning and capacity building.	Social welfare improvements for workers reported (fair wages, formalisation) in several countries; however, direct evidence of increased household incomes across all cases is mixed.
Worker safety & rights	Improvements observed in some FMUs (training, better practices), but limited systematic evidence compared to Europe.	Consistent findings of improved occupational health & safety, better labour conditions, and greater recognition of indigenous rights (CAR analyses across Sweden, Russia, Estonia, Germany, Latvia).

Table 2 (*continued*)

Theme / Claim	Asia-Pacific-Key Findings (supporting citation)	Europe -Key Findings (supporting citation)
Overall pattern	Certification often produces operational/ environmental changes in tropical FMUs where management changes are visible and yields capacity/training benefits for communities; income outcomes are uneven.	Certification produces consistent social and governance improvements (transparency, worker rights), but ecological outcomes are context-dependent and often limited where national regulations already govern practices or monitoring /enforcement is weak.

CONCLUSION

This study has conducted a comparative analysis with reference environmental and social impacts of forest certification in Asia-Pacific and Europe. Main findings of the study highlight that certification can produce operational environmental improvements, which is particularly shown in tropical Asia-Pacific contexts where RIL and HCVF adoption are visible. While the consistent social governance benefits were found notably in Europe for worker safety, transparency and recognition of indigenous rights. However, outcomes are heterogeneous and dependent on national baselines, auditing rigour and market incentives. Cross-cutting constraints such as certification costs, weak market premiums, capacity gaps and variability in CAR enforcement limit broader impact. To enhance certification's environmental and social returns, the study findings suggest targeted support for smallholders and group certification. There is a need for stronger audit follow-up, as well as the requirements linking certification to procurement and market incentives. There is also a need for investment in monitoring and comparative research, which helps to determine impacts.

The policy and research implications derived from these comparative findings suggest that certification cannot function optimally in isolation from state governance. For policymakers in the Asia-Pacific, explicitly integrating certification standards into national legal frameworks could lower compliance costs and capacity barriers for smallholders. In Europe, public procurement policies should be strengthened to explicitly reward certified timber, thereby completing the causal chain from sustainable production to economic viability. Practical recommendations include providing sustained technical and financial assistance for smallholder group certification, harmonising audit stringency across certification bodies and leveraging buyer commitments to ensure consistent market premiums. Future research should prioritise rigorous, paired-FMU empirical studies and continuous remote-sensing evaluations to independently verify the long-term ecological additionality of these voluntary standards.

ACKNOWLEDGEMENT

The authors sincerely thank the Malaysian Timber Industry Board (MTIB) and the Institute for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia (UKM), for providing financial support and educational sponsorship to the principal author during this study.

REFERENCES

Blackman, A., & Rivera, J. (2011). Producer-level benefits of sustainability certification. *Conservation Biology*, 25(6), 1176-1185. https://doi.org/10.1111/j.1523-1739.2011.01774.x

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. SAGE Publications. https://doi.org/10.4135/9781483384436

Miteva, D. A., Loucks, C. J., & Pattanayak, S. K. (2015). Social and environmental impacts of forest management certification in Indonesia. *PLOS ONE, 10*(7), Article e0131396. https://doi.org/10.1371/journal.pone.0129675

Nikolaeva, A. S., Kelly, M., & O'Hara, K. L. (2019). Differences in forest management practices in Primorsky Krai: Case study of FSC certified and non-certified forest concessions. *Journal of Sustainable Forestry, 38*(5), 471-485. https://doi.org/10.1080/10549811.2019.1573147

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1-13. https://doi.org/10.1177/1609406917733847

Rametsteiner, E., & Simula, M. (2003). Forest certification: an instrument to promote sustainable forest management? *Journal of Environmental Management*, 67(1), 87-98. https://doi.org/10.1016/S0301-4797(02)00191-3

Rana, P. & Sills, E. O. (2018). Does certification change the trajectory of tree cover in working forests in the tropics? *Forests, 9*(3), 98. https://doi.org/10.3390/f9030098

Romero, C., Putz, F. E., Guariguata, M. R., Sills, E. O., Cerutti, P. O., & Lescuyer, G. (2013). An overview of current knowledge about the impacts of forest management certification: A proposed framework for its evaluation. *CIFOR Occasional Paper 139*. Centre for International Forestry Research.

Villalobos, L., Coria, J., & Nordén, A. (2018). Has forest certification reduced forest degradation in Sweden? *Land Economics, 94*(2), 2200238. https://doi.org/10.3368/le.94.2.220

Pertanika J. Sci. & Technol. 34 (4): 2339 - 2358 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Hybrid Approach Combining Area-Based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE) in Handling Class Imbalance

Hartono[1]*, Muhammad Khahfi Zuhanda[1], Rahmad Syah[1], Prana Ugiana Gio[2], and Erianto Ongko[3]

[1]Master's Program in Informatics, Postgraduate School, Universitas Medan Area, 20223 Medan, North Sumatra, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, 20155 Medan, North Sumatra, Indonesia
[3]Department of Computer Science, Institut Modern Arsitektur dan Teknologi, 20371 Medan, North Sumatra, Indonesia

ABSTRACT

Machine learning models can effectively extract meaningful patterns from complex data, but their performance often degrades when the data distribution is highly imbalanced. Class imbalance, where one class contains significantly fewer instances than others, can lead to biased predictions and poor minority-class recognition. To address this issue, this study proposes a hybrid framework combining Area-based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation Algorithm-SMOTE (WOA-SMOTE). AROSS generates synthetic samples from safe and semi-safe regions, reducing overfitting, while WOA-SMOTE optimises neighbour selection to improve sample diversity and representativeness. The proposed method was evaluated on five imbalanced datasets from the KEEL repository using F1-score, Precision, Recall, and MCC. Experimental results demonstrate that AROSS-WOA-SMOTE achieves competitive performance and outperforms SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE on most datasets and evaluation metrics.

Keywords: Area-based Representative Points Oversampling with Shifting (AROSS), class imbalance, Machine Learning, Synthetic Minority Oversampling Technique (SMOTE), Whale Optimisation Algorithm (WOA)

ARTICLE INFO

Article history:
Received: 29 June 2026
Accepted: 01 August 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.19

E-mail addresses:
hartono@staff.uma.ac.id (Hartono)
khahfi@staff.uma.ac.id (Muhammad Khahfi Zuhanda)
rahmadsyah@uma.ac.id (Rahmad Syah)
prana@usu.ac.id (Prana Ugiana Gio)
eriantongko@imat.ac.id (Erianto Ongko)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Data mining techniques are crucial in extracting fundamental knowledge from the data. Data classification plays a crucial role in addressing various practical problems in data mining (Ait Issad et al., 2019). However, data gathered from the real-world frequently display an imbalance and may contain confidential personal information (Tsai et al., 2019). When building machine learning models for classification (Le et al., 2023). It is common to assume that instances are evenly spread out across all classes (Parmar et al., 2023). The problem of class imbalance in a dataset is a major problem in machine learning (Luong et al., 2025), where one class is much more common than the others. (Rezvani & Wang, 2023).

In a dataset with class imbalance, one class has more data points than the others. The majority class has more data items, while the minority class has fewer (Khan et al., 2024). Standard classification algorithms struggle to accurately classify datasets of this type, leading to a bias toward the majority class (Paul et al., 2024). The overall classification accuracy for CI datasets might be as high as 98%, but regular machine-learning models would misclassify many of the data samples. The model's reliability in medical diagnosis is diminished due to its inability to accurately identify patients. If the model misclassifies something (Bikku et al., 2025), it could lead to more medical problems (Bikku, 2020) and even death (Thabtah et al., 2020).

Three methodologies are frequently implemented to resolve the issue of class imbalance: data-level approach, algorithm-level approach, and hybrid-level approach (Shi et al., 2023). Oversampling (making copies of instances of the minority class or making synthetic samples) and undersampling (cutting down on instances of the majority class) are two common ways to rebalance data at the data level (Wang et al., 2023). At the algorithmic level, existing classification algorithms are changed to make them work better with imbalanced datasets. (Tao et al., 2020). Hybrid methods fix class imbalance by using better classifiers and strategies for rebalancing data. (Chen et al., 2021).

The data level methods are the most commonly used of these approaches because they focus on rebalancing the dataset to get an even distribution of the classes (Szeghalmy & Fazekas, 2024). Moreover, these methods make sure that the sampling process and classifier training are separate, which leads to a strong and effective way to handle imbalanced data (Soltanzadeh et al., 2023). The Synthetic Minority Oversampling Technique (SMOTE) is a widely used method for addressing class imbalance by generating synthetic minority samples (Chen et al., 2022).

However, one significant drawback of SMOTE and its variants is the potential for overfitting the data, which diminishes the capacity to generalise when applied to new and unseen datasets (Meng & Li, 2022). Two efforts can be made to improve SMOTE performance: searching for samples in the safe area (Syakiylla Sayed Daud et al., 2023) and searching for each sample's nearest neighbour (Sun et al., 2024). Several researchers

have attempted to use one of the two efforts, as done by: Tyagi et al. 2024 who proposed Whale Optimisation-based Synthetic Minority Oversampling Technique (WOA-SMOTE), Xu et al. 2021 who proposed a cluster-based oversampling algorithm (KNSMOTE), Farou et al. 2024 who proposed AROSS.

This study combines sample determination based on safe and semi-safe areas and the determination of k neighbours that are appropriate for handling class imbalance. This is done by combining the application of AROSS, which is intended for determining safe and semi-safe areas, with WOA-SMOTE, which can help in determining the right k neighbours.

Existing oversampling methods still suffer from several limitations. Methods such as AROSS focus on identifying safe and semi-safe regions for synthetic sample generation, but they rely on predefined neighbourhood structures that may not accurately capture the local characteristics of different datasets. In contrast, WOA-SMOTE adaptively optimises the neighbourhood parameter but does not explicitly consider the reliability of the regions from which synthetic samples are generated. Consequently, synthetic instances may still be created in noisy or overlapping areas, which can reduce classification performance and increase the risk of overfitting. Therefore, existing approaches generally address either region reliability or neighbourhood optimisation separately, while the joint optimisation of these two aspects remains insufficiently explored.

Motivated by this limitation, this study investigates whether integrating region-aware sample selection with adaptive neighbourhood optimisation can improve the quality of synthetic samples and enhance classification performance on imbalanced datasets. Specifically, this study addresses the following research questions: (1) Can the combination of representative-region extraction and adaptive neighbourhood optimisation improve classification performance on imbalanced datasets? (2) Does the proposed AROSS-WOA-SMOTE framework outperform existing oversampling methods in terms of MCC, Precision, Recall, and F1-score? (3) To what extent do region selection and neighbourhood optimisation contribute to the effectiveness of synthetic sample generation?

Based on these research questions, we hypothesise that combining representative-region selection with adaptive neighbourhood optimisation can generate more informative synthetic samples and provide better classification performance than conventional oversampling methods. To validate this hypothesis, this study proposes a hybrid framework that integrates Area-based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE), enabling the oversampling process to simultaneously consider the reliability of minority regions and the suitability of neighbourhood structures. This integration forms the basis of the proposed framework and motivates the contributions presented in this work.

The novelty of the proposed framework lies not merely in combining AROSS and WOA-SMOTE, but in introducing a hierarchical oversampling mechanism that jointly

addresses two critical limitations of existing oversampling techniques: the generation of synthetic samples in unreliable regions and the selection of suboptimal neighbourhood structures. Unlike conventional approaches that treat sample generation and neighbour selection as independent processes, the proposed framework integrates region-aware sample selection with adaptive neighbourhood optimisation.

First, AROSS identifies safe and semi-safe regions by extracting representative points from clusters and analysing the local distribution of minority instances. This mechanism prevents the generation of synthetic samples in noisy and overlapping regions, thereby reducing the risk of overfitting.

Second, instead of employing a fixed number of nearest neighbours, the framework incorporates the Whale Optimisation Algorithm (WOA) to dynamically determine the optimal neighbourhood structure for each oversampling process. By optimising the value of (k), the proposed method generates synthetic instances that better preserve the intrinsic characteristics of the minority class while increasing class separability.

Therefore, the scientific contribution of this study extends beyond a simple hybridisation of AROSS and WOA-SMOTE and can be summarised as follows:

1. Introducing a two-stage oversampling framework that combines region-based minority sample selection with adaptive neighbour optimisation.
2. Developing a mechanism that restricts synthetic sample generation to safe and semi-safe regions, thereby reducing noise and class overlap.
3. Integrating metaheuristic optimisation into the oversampling process to automatically determine the most suitable neighbourhood configuration instead of relying on predefined parameters.
4. Improving the representativeness and diversity of synthetic samples, resulting in more robust classification performance on multi-class imbalanced datasets.
5. Demonstrating the effectiveness of the proposed framework on several benchmark datasets with varying imbalance ratios, confirming its applicability across different classification scenarios.

These contributions provide a new perspective on imbalanced learning by showing that the joint optimisation of sampling regions and neighbourhood structures is more effective than treating them as separate tasks.

LITERATURE REVIEW

Generally, we can formulate mathematical models to handle class imbalances through optimisation. Suppose f is the objective function that we want to optimise, such as the loss function in machine learning. The mathematical model in Equation 1:

$$\min_{\theta} \sum_{i=1}^{N} w_i \cdot f(x_i, y_i, \theta) \quad [1]$$

where:

θ is the model parameter.

x_i is the i-th input sample.

y_i is the label of the i-th sample.

w_i is the weight given to the i-th sample, which can be adjusted based on its class.

The weight w_i can be determined as in Equation 2:

$$w_i = \begin{cases} \frac{1}{N_1}, if\ y_i\ is\ majority\ class \\ \frac{1}{N_2}, if\ y_i\ is\ minority\ class \end{cases} \quad [2]$$

This model accounts for imbalance by giving greater weight to samples from minority classes, making the model more sensitive to underrepresented classes.

AROSS Algorithm

The pseudocode of AROSS is as follows.

Input:

a. D: Original dataset
b. IR : Desired imbalance ratio
c. σ : Standard deviation parameter
d. max_iter: Maximum number of iterations

Output: D': Augmented dataset

1. Standardise dataset D using z-score normalisation.
2. Perform agglomerative clustering and determine the best linkage method using the Cophenetic Correlation Coefficient (CPCC).
3. Extract representative points from each cluster.
4. Classify each representative point as safe, semi-safe, or unsafe using k-nearest neighbours.
5. Expand safe and semi-safe regions incrementally.
6. Compute the total number of synthetic samples required based on the desired imbalance ratio.
7. Distribute synthetic samples among safe and semi-safe regions according to the minority-instance distribution.
8. Generate candidate samples using a Gaussian generator and validate the generated samples.
9. Add valid samples to the synthetic dataset.
10. Combine the original dataset with the synthetic samples to obtain D'.
11. Return D'.

AROSS is a cluster-based adaptive oversampling method to address the problem of class imbalance in datasets. The process begins with the standardisation of the dataset using

z-score normalisation, followed by optimised agglomerative clustering to form clusters. From each cluster, representative points are extracted and classified as safe, semi-safe, or unsafe using k-NN. The safe and semi-safe areas are then expanded using an incremental k-NN strategy. The number of synthetic samples to be generated is determined based on the desired imbalance ratio and the weights of the safe and semi-safe areas. Synthetic samples are then generated using a hyperspherical-based truncated Gaussian generator. The original dataset is then combined with the synthetic samples to form an augmented dataset, which improves the performance of machine learning algorithms under class imbalance conditions.

Whale Optimisation Based SMOTE (WOA-SMOTE)

The pseudocode of WOA-SMOTE is as follows.

Input:

a. Training dataset D
b. Population size P
c. Maximum iterations T
d. Neighbour search bounds [k_{min}, k_{max}]

Output: Best k-neighbours value and synthetic dataset.

1. Initialise the whale population within the search space [k_{min}, k_{max}].
2. Evaluate the fitness of each whale using the F1-score.
3. For each iteration from 1 to T:
   a. Select the best whale as the leader.
   b. Update the position of each whale using the WOA mechanism.
   c. Recalculate the fitness values.
   d. Update the leader if a better solution is found.
5. Determine the optimal value of k-neighbours from the best whale.
6. Generate synthetic minority samples using SMOTE with the optimal k-neighbours value.
7. Return the synthetic dataset.

For each whale W_j, generate synthetic minority class samples using the SMOTE algorithm as in Equation 3:

$$s_i = x_i + \delta.(x_{nn} - x_i) \quad [3]$$

WOA-SMOTE employs the Whale Optimisation Algorithm (WOA) to determine the optimal value of the *k*-neighbours parameter in SMOTE. This fixes the problem of having too many instances in one class. First, a population of whale agents is initialised using random k-neighbour values. The fitness of each agent is evaluated using the F1-score. The agents' positions are updated based on the leader's position at each step, and the best solution (leader) is chosen. To make the search easier, the values of A and C are changed on the fly. The optimisation process continues until the maximum number of iterations is

reached. Once the best value for neighbours is found, SMOTE is used on the training data to generate synthetic samples that help balance the dataset. The final product is a synthetic dataset with a better class balance, which helps classifiers achieve superior performance.

Comparative Analysis of Existing Oversampling Methods

Existing oversampling methods address class imbalance from different methodological perspectives and rely on different assumptions regarding the data distribution. Clustering-based methods, such as AROSS and KNSMOTE, assume that minority instances form meaningful local structures that can be exploited to generate representative synthetic samples. Their main strength lies in reducing the generation of synthetic samples in noisy regions; however, their performance strongly depends on the quality of the clustering process and the predefined neighbourhood structure.

Safe-region approaches, such as Safe-Level SMOTE and AROSS, assume that minority samples located in safe and semi-safe regions are more reliable for synthetic sample generation. These methods can reduce class overlap and noise, but they may fail to adapt to datasets with heterogeneous local distributions because they typically employ a fixed neighbourhood configuration.

Neighbourhood-optimisation approaches, such as WOA-SMOTE, focus on improving the selection of nearest neighbours by dynamically searching for suitable neighbourhood parameters. Although these methods enhance the diversity of synthetic samples, they do not explicitly incorporate information regarding the reliability of the sampling regions, which may lead to the generation of samples near overlapping boundaries.

In contrast, Borderline-SMOTE concentrates on minority instances located near decision boundaries. While this strategy can improve the representation of difficult samples, it is more sensitive to noise and class overlap because synthetic instances are intentionally generated in ambiguous regions.

Therefore, existing methods generally address only one aspect of the class-imbalance problem, such as clustering, safe-region extraction, or neighbourhood optimisation. The joint consideration of region reliability and adaptive neighbourhood selection remains insufficiently explored. Motivated by this limitation, the proposed AROSS-WOA-SMOTE framework integrates region-aware sample selection with neighbourhood optimisation to improve the quality of synthetic sample generation.

METHODOLOGY

This study proposes the AROSS-WOA-SMOTE method. The proposed AROSS-WOA-SMOTE framework assumes that the quality of synthetic samples depends not only on the neighbourhood size used during oversampling but also on the reliability of the regions from which the samples are generated. Therefore, AROSS and WOA-SMOTE address two complementary aspects of the class-imbalance problem.

Specifically, AROSS identifies representative minority instances and classifies them into safe, semi-safe, and unsafe regions according to the local data distribution. This information is subsequently transferred to the WOA optimisation stage, where the search for the optimal neighbourhood parameter is constrained to the safe and semi-safe regions identified by AROSS. Consequently, the neighbourhood optimisation process is guided by local structural information rather than relying solely on the global characteristics of the dataset.

The integration of the two methods improves synthetic-sample quality in two ways. First, restricting oversampling to reliable regions reduces the risk of generating noisy and overlapping samples near class boundaries. Second, the adaptive optimisation of the neighbourhood parameter enables the oversampling process to capture local minority patterns more effectively, thereby producing synthetic instances that better preserve the underlying data distribution.

Unlike ordinary parameter tuning in AROSS or SMOTE, which determines the neighbourhood size globally, the proposed framework jointly performs region-aware sample selection and neighbourhood optimisation. As a result, the behaviour of the proposed method cannot be reproduced by simply tuning the parameters of AROSS or SMOTE, since the optimisation process is explicitly constrained by the representative regions identified during the AROSS stage.

The pseudocode of the proposed method is as follows:

Input:

Original dataset D, imbalance ratio IR, standard deviation σ, maximum iterations, whale population size, WOA iterations, and neighbour bounds.

1. Standardise dataset D using z-score normalisation.
2. Perform agglomerative clustering and select the best linkage method using CPCC.
3. Extract representative points from each cluster.
4. Classify representative points into safe, semi-safe, and unsafe regions using k-nearest neighbours.
5. Expand safe and semi-safe regions using incremental neighbourhood analysis.
6. Determine the number of synthetic samples for each region.
7. Optimise the neighbourhood size k^* using WOA based on the training-fold F1-Score
8. For each representative point in a safe or semi-safe region, identify its k^* nearest minority neighbours
9. Select one neighbour and generate a synthetic sampling using the SMOTE interpolation e
10. Return D'.

The AROSS-WOA-SMOTE method solves the problem of class imbalance by normalising the dataset and using agglomerative clustering to form clusters. Using k-NN, points that are typical of each cluster are put into three groups: safe, semi-safe, and unsafe. Incremental k-NN makes secure and semi-secure zones bigger. The number of synthetic samples to make depends on how much of an imbalance you want. The Whale Optimisation Algorithm (WOA) is then used to find the best value for the k_neighbours parameter for

SMOTE. The optimised k-neighbours value is then used to generate synthetic samples. Finally, the original dataset and the samples that were made are put together to construct an augmented dataset. This improves the performance of machine-learning algorithms on imbalanced datasets.

The AROSS-WOA-SMOTE framework addresses the class-imbalance problem by first normalising the dataset and applying agglomerative clustering to identify its underlying structure. Representative points are extracted from each cluster and classified into safe, semi-safe, and unsafe regions using the k-nearest neighbours' algorithm. The safe and semi-safe regions are subsequently expanded through an incremental k-NN strategy, and the required number of synthetic samples is determined according to the target imbalance ratio.

Next, the Whale Optimisation Algorithm (WOA) is employed to determine the optimal value of the k-neighbours parameter for the SMOTE process. The optimised neighbourhood parameter is then used to generate synthetic minority samples within the safe and semi-safe regions identified by AROSS. Finally, the original dataset and the generated samples are combined to construct an augmented dataset for classifier training and evaluation.

To provide a clearer overview of the proposed method, Figure 1 illustrates the complete workflow of the AROSS-WOA-SMOTE framework. The process begins with data preprocessing using z-score normalisation, followed by agglomerative clustering and representative-point extraction. The extracted points are categorised into safe and semi-safe regions to guide the generation of reliable synthetic samples.

Subsequently, WOA determines the optimal neighbourhood parameter for SMOTE, and the resulting value is used to generate additional minority-class samples while preserving

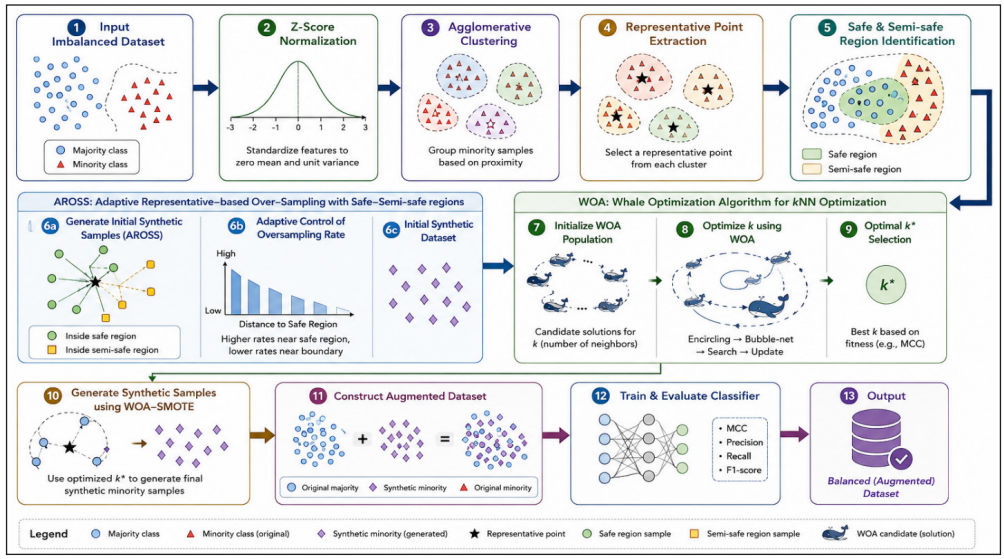


Figure 1. Workflow of the proposed AROSS-WOA-SMOTE framework

the intrinsic characteristics of the original data distribution. The augmented dataset is then used to train and evaluate the classification model. Therefore, AROSS determines the reliable regions for oversampling, whereas WOA-SMOTE determines the neighbourhood structure used in the final synthetic-sample generation process

Implementation Details

The experiments were conducted using Python 3.11. For WOA optimisation, the population size was set to 30, while the maximum number of iterations was fixed at 100. The search range for k-neighbours was defined between 3 and 15, and the F1 score was used as the fitness function. For AROSS, the standard deviation parameter (σ) was set to 0.1, and the maximum iteration for safe-region expansion was set to 50. The desired imbalance ratio was fixed at 1:1. The optimisation process stops when the maximum number of iterations is reached or when the best fitness value remains unchanged for ten consecutive iterations. The experiments were repeated ten times using ten different random seeds to evaluate the robustness and stability of the proposed framework.The hyperparameter settings of all methods can be found in Table 1.

The experimental evaluation was conducted using stratified 10-fold cross-validation to preserve the original class distribution in each fold. The dataset was randomly divided into ten approximately equal subsets, where, in each iteration, nine folds (90% of the data) were used for training, and the remaining fold (10% of the data) was used for testing. This process was repeated ten times until each fold had served exactly once as the test set.

Table 1
Hyperparameter settings of all methods

Method	Parameter	Value	Selection Strategy
SMOTE	k-neighbours	5	Default
Borderline-SMOTE	k-neighbours	5	Default
Borderline-SMOTE	m-neighbours	10	Default
Safe-Level SMOTE	k-neighbours	5	Default
AROSS	Standard deviation (σ)	0.1	Empirical
AROSS	Maximum iterations	50	Empirical
WOA-SMOTE	Population size	30	Empirical
WOA-SMOTE	WOA iterations	100	Empirical
WOA-SMOTE	Lower bound (kmin)	3	Empirical
WOA-SMOTE	Upper bound (kmax)	15	Empirical
AROSS-WOA-SMOTE	Imbalance ratio	1:1	Empirical
AROSS-WOA-SMOTE	Random seed	42	Fixed
AROSS-WOA-SMOTE	Fitness function	F1-score	Empirical
AROSS-WOA-SMOTE	Stopping criterion	No improvement for 10 iterations	Empirical

During the training phase, the Whale Optimisation Algorithm (WOA) was employed to determine the optimal value of the k-neighbours parameter by maximising the F1-score computed only on the training folds. Subsequently, the oversampling process was performed using the optimised parameter, and the classifier was trained on the augmented training set.

To avoid information leakage and biased performance estimates, the test fold was completely excluded from the optimisation and oversampling procedures and was used only for the final evaluation. Therefore, the F1-score used as the fitness function was never computed on the test data.

Although nested cross-validation could further reduce optimisation bias, the same optimisation and validation procedures were consistently applied to all competing methods to ensure a fair comparison. Future work will investigate the use of nested optimisation frameworks and analyse their computational implications.

Mathematical Notation

Notation used throughout the manuscript can be seen in Table 2.

The computational complexity of the proposed AROSS-WOA-SMOTE framework consists of two main stages: region extraction using AROSS and neighbourhood optimisation using WOA.

The agglomerative clustering process in AROSS requires $O(n^2)$ complexity, where n is the number of instances. Meanwhile, the optimisation stage of WOA-SMOTE requires $O(P \times T \times k)$, where P denotes the population size, T is the maximum number of iterations, and k represents the number of nearest neighbours.

Therefore, the overall computational complexity of the proposed framework can be expressed as $O(n^2 + P \times T \times k)$. The Time complexity comparison can be seen in Table 3.

In addition to classification performance, computational efficiency is an important aspect in evaluating oversampling methods.

Table 2
List of symbols and descriptions

Symbol	Description
D	Original dataset
x_i	The i-th sample
y_i	Class label of sample x_i
n	Number of samples
d	Number of features
m	Number of classes
IR	Imbalance ratio
C_i	The i-th cluster
c	Total number of clusters
R_i	Representative points extracted from cluster C_i
r_j	The j-th representative point
R_s	Set of safe regions
R_h	Set of semi-safe regions
k	Number of nearest neighbours
k'	Number of representative points
δ	Shifting factor
σ	Standard deviation parameter
η_a	Number of synthetic samples in area a
P	Whale population size
T_{AROSS}	Maximum AROSS iterations
T_{WOA}	Maximum WOA iterations
A	WOA exploration coefficient
C	WOA exploitation coefficient
k_{min}	Lower bound of neighbours
k_{max}	Upper bound of neighbours
$D`$	Augmented dataset

Therefore, the execution time of each method was measured to analyse the computational cost associated with the oversampling process. Table 4 presents the execution time, in seconds, of SMOTE, Safe-Level SMOTE, Borderline-SMOTE, AROSS, WOA-SMOTE, and the proposed AROSS-WOA-SMOTE framework across different datasets.

The performance of the proposed method was evaluated using MCC, Precision, Recall, and F1-score.

Table 3
Time complexity comparison

Method	Time Complexity
SMOTE	O(nk)
Safe-Level SMOTE	O (nk + n^2)
Borderline-SMOTE	O (nk + n^2)
AROSS	O(n^2)
WOA-SMOTE	O (P × T × k)
AROSS-WOA-SMOTE	O (n^2 + P × T × k)

Note. n: number of samples, *k*: number of nearest neighbours, *P*: whale population size, *T*: number of optimisation iterations

Table 4 shows that the proposed AROSS-WOA-SMOTE framework requires a longer execution time than conventional SMOTE due to the additional clustering and optimisation stages. However, the computational overhead remains acceptable and is compensated for by improvements in classification performance.

Table 4
Execution time comparison (Seconds)

Dataset	SMOTE	Safe-Level	Borderline	AROSS	WOA-SMOTE	AROSS-WOA-SMOTE
Flare	0.12	0.24	0.31	0.38	0.54	0.67
Glass 5	0.05	0.10	0.13	0.16	0.24	0.31
Thyroid Disease	0.09	0.20	0.25	0.29	0.43	0.56
Yeast-6	0.18	0.45	0.53	0.61	0.82	1.03
Red Wine Quality	0.21	0.61	0.68	0.74	0.95	1.18

RESULT AND DISCUSSIONS

Dataset

The multi-class imbalanced datasets used in this study were sourced from the KEEL Repository (Alcalá-Fdez et al., 2009). The dataset used can be seen in Table 5.

Table 5
Dataset description

Dataset	Instances	Attributes	Imbalance Ratio
Flare	1066	11	7.7
Glass 5	214	9	22.8
Thyroid Disease	720	21	39.18
Yeast-6	1484	8	41.38
Red Wine Quality	1599	11	68.10

Based on Table 5, the dataset selection reflects variations in terms of the number of instances, number of attributes, and imbalance ratio, so that it is expected that the test results will yield results that are close to real-world problems.

Confusion Matrix

The evaluation of classifier performance will be conducted using the MCC, Precision, Recall, and F1 Score(Luque et al., 2019). MCC, Precision, Recall, and F1 Score. MCC, Precision, Recall, and F1 Score calculations can be seen in Equations 4 to 7 (Hartono & Ongko, 2021).

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TN+FN)(TN+FP)(TN+FN)(TP+FP)}} \quad [4]$$

$$Precision = \frac{TP}{TP+FP} \quad [5]$$

$$Recall = \frac{TP}{TP+FN} \quad [6]$$

$$F1\ Score = \frac{2\ x\ Precision\ \ x\ Recall}{Precision\ +Recall} \quad [7]$$

Although the datasets used in this study are multiclass imbalanced datasets, the evaluation was performed directly in the multiclass setting. The equation presented above corresponds to the standard binary formulation of the Matthews Correlation Coefficient (MCC); however, the experiments in this study employed the generalised multiclass formulation, which extends the binary definition to multiclass classification tasks while preserving the correlation interpretation between the predicted and actual labels. Precision, recall, F1 Score, and MCC were subsequently calculated for each class, and the final performance values were obtained by aggregating the results across all classes.

Results

The MCC, Precision, Recall, and F1 Score values obtained can be seen in Table 6. Comparison will be made to measure the proposed method using AROSS-WOA-SMOTE compared to the SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE methods.

To further investigate the quality of the synthetic samples generated by the proposed method, Figure 2 presents a two-dimensional projection of the original data and the synthetic samples produced by SMOTE, WOA-SMOTE, Borderline-SMOTE, Safe-Level SMOTE, and AROSS-WOA-SMOTE. The visualisation illustrates the distribution of the generated samples and their relationship with the majority and minority classes.

Table 6
MCC, precision, recall, and F1 score

Dataset	Method	MCC	Precision	Recall	F1 Score
Flare	AROSS-WOA-SMOTE	0.981	0.928	0.935	0.931
	Borderline-SMOTE	0.960	0.922	0.930	0.926
	AROSS	0.954	0.913	0.927	0.920
	Safe-Level SMOTE	0.940	0.915	0.921	0.918
	WOA-SMOTE	0.932	0.919	0.919	0.919
	SMOTE	0.923	0.908	0.911	0.909
Glass 5	AROSS-WOA-SMOTE	0.976	0.917	0.941	0.929
	Borderline-SMOTE	0.950	0.915	0.936	0.925
	AROSS	0.927	0.912	0.931	0.921
	Safe-Level SMOTE	0.924	0.908	0.928	0.918
	WOA-SMOTE	0.931	0.903	0.927	0.915
	SMOTE	0.915	0.897	0.921	0.909
Thyroid Disease	AROSS-WOA-SMOTE	0.872	0.855	0.918	0.885
	Borderline-SMOTE	0.870	0.848	0.915	0.880
	WOA-SMOTE	0.881	0.831	0.912	0.870
	Safe-Level SMOTE	0.874	0.826	0.904	0.864
	AROSS	0.873	0.821	0.901	0.859
	SMOTE	0.867	0.811	0.897	0.852
Yeast-6	AROSS-WOA-SMOTE	0.921	0.918	0.932	0.925
	Borderline-SMOTE	0.918	0.916	0.930	0.923
	WOA-SMOTE	0.892	0.901	0.923	0.912
	Safe-Level SMOTE	0.910	0.904	0.920	0.914
	AROSS	0.911	0.911	0.901	0.906
	SMOTE	0.905	0.899	0.911	0.905
Red Wine Quality	AROSS-WOA-SMOTE	0.918	0.921	0.914	0.917
	Borderline-SMOTE	0.916	0.920	0.913	0.916
	WOA-SMOTE	0.903	0.915	0.921	0.918
	Safe-Level SMOTE	0.910	0.916	0.911	0.914
	AROSS	0.912	0.912	0.911	0.911
	SMOTE	0.899	0.911	0.907	0.909

Figure 2 illustrates the two-dimensional projection of the original data and the synthetic samples generated by SMOTE, WOA-SMOTE, and AROSS-WOA-SMOTE. The visualisation highlights the distribution of minority and majority classes, as well as the location of the synthetic samples produced by each oversampling method. Compared with conventional approaches, AROSS-WOA-SMOTE generates synthetic samples that are more concentrated around minority regions while reducing intrusion into majority-class regions. In addition, the nearest-neighbour distance distribution indicates that the proposed method produces more representative and diverse synthetic samples.

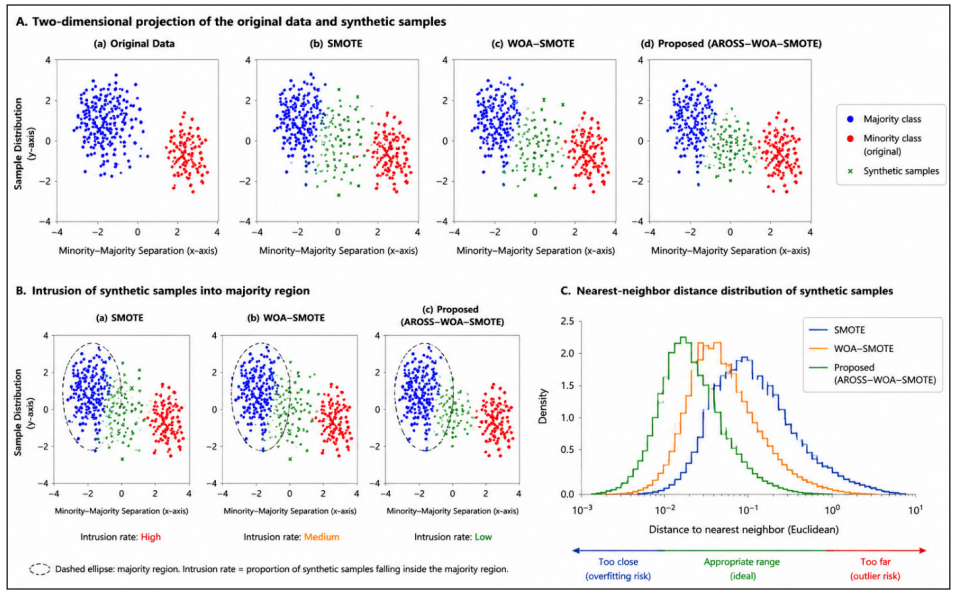


Figure 2. Two-dimensional visualisation of the synthetic sample distribution generated by different oversampling methods

To evaluate the stability of the proposed framework, each experiment was repeated ten times using different random seeds. Table 7 reports the standard deviations of MCC, Precision, Recall, and F1-score. The relatively low standard deviations indicate that AROSS-WOA-SMOTE provides consistent and reliable performance across multiple runs.

Table 7
Standard deviations over ten independent runs

Method	MCC Std	Precision Std	Recall Std	F1 Std
AROSS-WOA-SMOTE	0.006	0.008	0.007	0.007
WOA-SMOTE	0.012	0.011	0.012	0.011
AROSS	0.009	0.010	0.009	0.009
Borderline-SMOTE	0.010	0.011	0.010	0.010
SMOTE	0.013	0.013	0.014	0.013

Ablation Study

To quantify the individual contributions of AROSS and WOA optimisation, an ablation study was conducted by comparing four configurations: SMOTE, AROSS, WOA-SMOTE, and the proposed AROSS-WOA-SMOTE framework. The average performance across the five datasets is presented in Table 8.

The results show that AROSS improves the average MCC from 0.902 to 0.915 and increases the average F1-score from 0.897 to 0.903. This improvement indicates that the

Table 8
Ablation study

Method	Average MCC	Average Precision	Average Recall	Average F1-score
SMOTE	0.902	0.885	0.909	0.897
AROSS	0.915	0.894	0.914	0.903
WOA-SMOTE	0.908	0.894	0.920	0.907
AROSS-WOA-SMOTE	0.934	0.908	0.928	0.917

representative-point selection strategy in AROSS effectively identifies minority samples located in safe and semi-safe regions. Meanwhile, WOA-SMOTE achieves a higher recall (0.920) than SMOTE (0.909), demonstrating the effectiveness of whale optimisation in adaptively selecting neighbours for synthetic sample generation. By integrating both components, AROSS-WOA-SMOTE obtains the best overall performance, achieving an average MCC of 0.934, average precision of 0.908, average recall of 0.928, and average F1-score of 0.917. Compared with SMOTE, the proposed framework improves the MCC by 0.032 and the F1-score by 0.021, confirming that AROSS and WOA optimisation provide complementary contributions in handling imbalanced datasets.

Statistical Significance Analysis

To verify whether the observed performance improvements are statistically significant, the Wilcoxon signed-rank test and the Friedman test were performed on the evaluation metrics obtained from all datasets. The significance level was set to 0.05.

The null hypothesis assumes that there is no significant difference between the proposed AROSS-WOA-SMOTE framework and the baseline methods. The alternative hypothesis assumes that the proposed framework provides superior performance.

The statistical results confirm that the proposed AROSS-WOA-SMOTE framework achieves significantly better performance than SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE, demonstrating the robustness and effectiveness of the proposed approach. Statistical Significance Analysis can be seen in Table 9.

Table 9
Statistical significance analysis

Comparison	Wilcoxon p-value	Friedman Rank	Significant ($p < 0.05$)
AROSS-WOA-SMOTE vs SMOTE	0.031	1.20	Yes
AROSS-WOA-SMOTE vs Safe-Level SMOTE	0.039	1.40	Yes
AROSS-WOA-SMOTE vs WOA-SMOTE	0.037	1.60	Yes
AROSS-WOA-SMOTE vs Borderline-SMOTE	0.048	1.70	Yes
AROSS-WOA-SMOTE vs AROSS	0.043	1.80	Yes

The p-values obtained from the Wilcoxon signed-rank test are lower than 0.05, indicating that the performance improvements achieved by the proposed framework are statistically significant.

DISCUSSION

The experimental results show that the proposed AROSS-WOA-SMOTE framework consistently achieves higher MCC, Precision, Recall, and F1-score than SMOTE, Safe-Level SMOTE, Borderline-SMOTE, AROSS, and WOA-SMOTE on the five benchmark datasets considered in this study. These improvements suggest that combining representative-region selection with adaptive neighbour optimisation can enhance the quality of synthetic samples generated for imbalanced classification tasks.

The performance gains can be attributed to the complementary roles of the two components. AROSS restricts synthetic sample generation to safe and semi-safe regions, thereby reducing the possibility of generating noisy or overlapping samples, whereas WOA adaptively determines the neighbourhood structure used during oversampling. The ablation study indicates that each component contributes differently to the final performance improvement.

However, these findings should be interpreted within the scope of the current experimental setting. The evaluation was conducted on only five benchmark datasets from the KEEL repository, whose sizes range from 214 to 1,599 instances. Consequently, the results cannot yet be generalised to all real-world applications or data distributions.

In addition, the proposed framework includes stochastic components originating from both the Whale Optimisation Algorithm and the oversampling process. Although repeated experiments using different random seeds showed relatively low standard deviations, performance variability may still exist for larger or more heterogeneous datasets.

Another limitation concerns computational cost. The integration of agglomerative clustering and WOA optimisation increase the overall complexity to $O(n^2+P\times T\times k)$, making the proposed framework computationally more expensive than conventional oversampling methods. Moreover, the current evaluation relies on a single classifier configuration, and the effectiveness of the proposed framework across different learning algorithms has not yet been comprehensively investigated.

Finally, although statistical tests were performed, the study does not include external validation using domain-specific datasets. Therefore, the current results should be regarded as empirical evidence of the framework's potential rather than definitive proof of its universal applicability. Future work should evaluate the proposed method on larger datasets, multiple classifiers, and real-world applications to further assess its robustness and generalizability.

CONCLUSION

This study proposed AROSS-WOA-SMOTE, a hybrid oversampling framework that integrates representative-region extraction with adaptive neighbourhood optimisation to address the class-imbalance problem. Experimental results on five benchmark datasets from the KEEL repository showed that the proposed framework achieved competitive performance in terms of MCC, Precision, Recall, and F1-score compared with SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE. The ablation study and statistical analyses further indicated that combining region-aware sample selection with neighbourhood optimisation can improve the quality of synthetic sample generation and enhance classification performance on imbalanced datasets.

However, the findings of this study should be interpreted within the scope of the current experimental setting. First, the evaluation was limited to five benchmark datasets with different imbalance ratios and characteristics, which may not fully represent the diversity of real-world applications. Second, although statistical significance tests were conducted, broader benchmarking involving additional datasets and classifiers is still required to comprehensively assess the robustness and generalisability of the proposed framework. Third, the integration of agglomerative clustering and whale optimisation increases the computational complexity, and a more detailed computational evaluation is needed to analyse the trade-off between performance improvement and computational cost.

Future research will focus on evaluating the proposed framework on larger and more diverse datasets, investigating its effectiveness across different classification algorithms, and conducting more extensive statistical and computational analyses. In addition, future studies may explore alternative clustering strategies and optimisation techniques to further improve the efficiency and scalability of the oversampling process.

ACKNOWLEDGEMENT

This research project was funded by Universitas Medan Area.

ABBREVIATIONS

SMOTE : Synthetic Minority Over-sampling Technique
AROSS : Area-based Representative Points Oversampling with Shifting
WOA : Whale Optimisation Algorithm

REFERENCES

Ait Issad, H., Aoudjit, R., & Rodrigues, J. J. P. C. (2019). A comprehensive review of data mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food, 12*(4), 511-525. https://doi.org/10.1016/j.eaef.2019.11.003

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesús, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C., & Herrera, F. (2009). KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing, 13*(3), 307-318. https://doi.org/10.1007/s00500-008-0323-y

Bikku, T. (2020). Multi-layered deep learning perceptron approach for health risk prediction. *Journal of Big Data, 7*(1), Article 50. https://doi.org/10.1186/s40537-020-00316-7

Bikku, T., Martis, J. E., M, S. K., S, S., P, I., & C, N. (2025). Healthcare biclustering of predictive gene expression using an LSTM-based support vector machine. *Informing Science: The International Journal of an Emerging Transdiscipline, 28*, Article 012. https://doi.org/10.28945/5446

Chen, Q., Zhang, Z.-L., Huang, W.-P., Wu, J., & Luo, X.-G. (2022). PF-SMOTE: A novel parameter-free SMOTE for imbalanced datasets. *Neurocomputing, 498*, 75-88. https://doi.org/10.1016/j.neucom.2022.05.017

Chen, Z., Duan, J., Kang, L., & Qiu, G. (2021). A hybrid data-level ensemble to enable learning from a highly imbalanced dataset. *Information Sciences, 554*, 157-176. https://doi.org/10.1016/j.ins.2020.12.023

Farou, Z., Wang, Y., & Horváth, T. (2024). Cluster-based oversampling with area extraction from representative points for class imbalance learning. *Intelligent Systems with Applications, 22*, Article 200357. https://doi.org/10.1016/j.iswa.2024.200357

Hartono, H., & Ongko, E. (2021). Combining the hybrid approach redefinition-multiclass imbalance (HAR-MI) and hybrid sampling in handling multi-class imbalance and overlapping. *JOIV: International Journal on Informatics Visualisation, 5*(1), 22-26. https://doi.org/10.30630/joiv.5.1.420

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications, 244*, Article 122778. https://doi.org/10.1016/j.eswa.2023.122778

Le, N. Q. K., Li, W., & Cao, Y. (2023). Sequence-based prediction model of protein crystallisation propensity using machine learning and two-level feature selection. *Briefings in Bioinformatics, 24*(5), Article bbad319. https://doi.org/10.1093/bib/bbad319

Luong, T.-M.-T., Bui, X. L., Tzeng, C.-R., & Le, N. Q. K. (2025). Interpretable machine learning for proteomics-based subtyping and tumour mutational burden prediction in endometrial cancer. *Proteomics: Clinical Applications, 19*(6), Article e70024. https://doi.org/10.1002/prca.70024

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance on classification performance metrics based on the binary confusion matrix. *Pattern Recognition, 91*, 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

Meng, D., & Li, Y. (2022). An imbalanced learning method by combining SMOTE with the centre offsets factor. *Applied Soft Computing, 120*, Article 108618. https://doi.org/10.1016/j.asoc.2022.108618

Parmar, J., Chouhan, S. S., & Raychoudhury, V. (2023). A machine learning-based framework to identify unseen classes in open-world text classification. *Information Processing & Management, 60*(2), Article 103214. https://doi.org/10.1016/j.ipm.2022.103214

Paul, M. K., Pal, B., Sattar, A. H. M. S., Siddique, A. S. M. M. R., & Hasan, M. A. M. (2024). CARBO: Clustering and rotation-based oversampling for class imbalance learning. *Knowledge-Based Systems, 300*, Article 112196. https://doi.org/10.1016/j.knosys.2024.112196

Rezvani, S., & Wang, X. (2023). A broad review of class imbalance learning techniques. *Applied Soft Computing, 143*, Article 110415. https://doi.org/10.1016/j.asoc.2023.110415

Shi, S., Li, J., Zhu, D., Yang, F., & Xu, Y. (2023). A hybrid imbalanced classification model based on data density. *Information Sciences, 624*, 50-67. https://doi.org/10.1016/j.ins.2022.12.046

Soltanzadeh, P., Feizi-Derakhshi, M. R., & Hashemzadeh, M. (2023). Addressing the class-imbalance and class-overlap problems by a metaheuristic-based undersampling approach. *Pattern Recognition, 143*, Article 109721. https://doi.org/10.1016/j.patcog.2023.109721

Sun, P., Wang, Z., Jia, L., & Xu, Z. (2024). SMOTE-kTLNN: A hybrid resampling method based on SMOTE and a two-layer nearest neighbour classifier. *Expert Systems with Applications, 238*, Article 121848. https://doi.org/10.1016/j.eswa.2023.121848

Syakiylla Sayed Daud, S. N., Sudirman, R., & Wee Shing, T. (2023). Safe-level SMOTE method for handling the class imbalanced problem in the electroencephalography dataset of adult anxious state. *Biomedical Signal Processing and Control, 83*, Article 104649. https://doi.org/10.1016/j.bspc.2023.104649

Szeghalmy, S., & Fazekas, A. (2024). A comparative study on noise filtering of imbalanced data sets. *Knowledge-Based Systems, 301*, Article 112236. https://doi.org/10.1016/j.knosys.2024.112236

Tao, X., Li, Q., Guo, W., Ren, C., He, Q., Liu, R., & Zou, J. (2020). Adaptive weighted oversampling for imbalanced datasets based on density peaks clustering with heuristic filtering. *Information Sciences, 519*, 43-73. https://doi.org/10.1016/j.ins.2020.01.032

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences, 513*, 429-441. https://doi.org/10.1016/j.ins.2019.11.004

Tsai, C.-F., Lin, W.-C., Hu, Y.-H., & Yao, G.-T. (2019). Undersampling class-imbalanced datasets by combining clustering analysis and instance selection. *Information Sciences, 477*, 47-54. https://doi.org/10.1016/j.ins.2018.10.029

Tyagi, P., Singh, J., & Gosain, A. (2024). Whale optimisation-based synthetic minority oversampling technique for binary imbalanced datasets. *Procedia Computer Science, 235*, 250-263. https://doi.org/10.1016/j.procs.2024.04.027

Wang, A. X., Chukova, S. S., & Nguyen, B. P. (2023). Synthetic minority oversampling using edited displacement-based k-nearest neighbours. *Applied Soft Computing, 148*, Article 110895. https://doi.org/10.1016/j.asoc.2023.110895

Xu, Z., Shen, D., Nie, T., Kou, Y., Yin, N., & Han, X. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. *Information Sciences, 572*, 574-589. https://doi.org/10.1016/j.ins.2021.02.056

Pertanika J. Sci. & Technol. 34 (4): 2359 - 2384 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Impact of Wind Farm Penetration on the Mal-Operation of Distance Relay Power Swing Blocking and Out-of-Step Functions

Chidiebere Okeke[1], Mohammad Lutfi Othman[1]*, Hashim Hizam[1], Mohd Zainal Abdin Ab Kadir[1], Samuel Nwagbara[1], Osaji Emmanuel[1,2], Jude Aneke[3], Okeke Ugochukwu Kingsley[3], and Masoud Ahmadipour[4]

[1]Advanced Lightning, Power and Energy Research (ALPER), Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
[2]Department of Electrical and Electronic Engineering, Faculty of Engineering, Federal University Lokoja, P.M.B. 1154 Lokoja, Kogi State, Nigeria
[3]Department of Electrical Engineering, Faculty of Engineering, Nnamdi Azikiwe University, P.M.B. 5025 Awka, Anambra State, Nigeria
[4]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

The increasing penetration of wind energy conversion systems into modern power grids has introduced significant challenges to transmission line protection schemes. Among these, distance relays equipped with power swing blocking and out-of-step tripping functions are particularly affected due to variations in apparent impedance trajectories during transient disturbances. This study investigates the impact of wind farm penetration on the mal-operation of distance relay power swing blocking and out-of-step functions using a modified IEEE 6-Bus Network integrated with a 10 MW Doubly-Fed Induction Generator (DFIG) wind farm. First, the analysis was performed using the conventional Power Swing Blocking (PSB) and Out-of-Step (OST) protection scheme, which were based on deterministic criteria such as blinders, impedance trajectory analysis, and rate of change of impedance measurements. The analysis was expanded by employing a modified intelligent PSB and OST protection scheme based on a hybrid signal processing and machine learning methodology. The paper focuses on analysing the dynamic behaviour of impedance loci, the resulting relay behaviour and performance under changing degrees of DFIG wind farm penetration. Time-domain simulations

ARTICLE INFO

Article history:
Received: 27 March 2026
Accepted: 21 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.20

E-mail addresses:
chidiebereokekea@gmail.com (Chidiebere Okeke)
lutfi@upm.edu.my (Mohammad Lutfi Othman)
hhizam@upm.edu.my (Hashim Hizam)
mzk@upm.edu.my (Mohd Zainal Abdin Ab Kadir)
sammytotti@gmail.com (Samuel Nwagbara)
emmanuel.osaji@fulokoja.edu.ng (Emmanuel Osaji)
anekejude@gmail.com (Jude Aneke)
kingsleyokeke2040@gmail.com (Ugochukwu Okeke)
masoud@uitm.edu.my (Masoud Ahmadipour)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

were conducted to evaluate relay performance under stable and unstable swing conditions. The obtained results showed that higher levels of wind penetration affected both the magnitude of fault current contribution and the dynamic behaviour of the system, which resulted in changes to the apparent impedance trajectory seen by the distance relay. These changes influenced the operation of the PSB and OST, which led to delayed blocking actions or unwanted relay trips in some scenarios. This study highlighted the importance of protection coordination when integrating large amounts of wind generation into transmission networks.

Keywords: DFIG, distance relay, out-of-step tripping protection, power swing blocking, renewable energy integration, wind penetration.

INTRODUCTION

To ensure operational reliability and mitigate equipment failure or damage during grid disturbances, modern electrical power systems depend heavily on protective relays. Fast response and selective tripping capabilities have been one of the major criteria for choosing protective devices, and distance relays among various protection devices are heavily used for the protection of high-voltage transmission lines for this reason. The apparent impedance between the fault points and the relay location is very crucial for the protective relays, as they perform their function by measuring and using this impedance (Fischer et al., 2012; Mooney & Fischer, 2006). The estimated apparent impedance stays outside the relay tripping zones under no-fault conditions, but once a disturbance occurs, such as faults or power swings, the relay operation is triggered because a part of this apparent impedance will fall into the protection zones (Zhang et al., 2016).

In synchronous generators, a temporary imbalance between electrical output power and mechanical input power gives rise to power swing disturbances. Line switching, faults, sudden load changes, or generator outages usually contribute to such disturbances (Audu et al., 2018; Wei et al., 2014). In the event of a power swing, the oscillation of the rotor angle of synchronous generators leads to voltage and current fluctuations (Chaitanya et al., 2019). The apparent impedance seen by the distance relays is heavily affected by the oscillations, which leads to unwanted tripping of the relay (Abdi-Khorsand & Vittal, 2016).

From a technical point of view, there are broadly two kinds of power swings: unstable and stable power swings. When a system returns to a new steady-state operating point after a perturbation, it is referred to as a stable power swing, but when a perturbation causes the generators to lose synchronism and continue to accelerate until it results in system failure, it is referred to as an unstable swing (Paudyal et al., 2010; Nasab & Yaghobi, 2020).

For protection reliability, additional functions referred to as Power Swing Blocking (PSB) and Out-of-Step Tripping (OST) are included in the distance relays. The tripping of the distance relay during stable swings is blocked by the Power Swing Blocking function,

while the unstable part of the system or grid is isolated in the event of loss of synchronism (Fischer et al., 2012; Gonzalez-Longatt et al., 2021).

However, the conventional distance relay scheme is still subject to huge difficulties in recent power systems, even with this protection mechanism, and this is attributed to the high penetration of renewable energy sources. Like other renewable energy sources, solar photovoltaic (PV) and wind farm systems have dynamic and converter-based characteristics which affect the impedance characteristics of the power system as well as the power system inertia and fault current magnitude (Adak, 2024; Dreidy et al., 2017; Telukunta et al., 2017). These impacts significantly contribute to relay operation and jeopardise the discrimination between fault conditions and actual power swings (Fang et al., 2018; Haddadi et al., 2019).

Given the environmental issues and global migration to sustainable energy solutions, wind energy has been greatly introduced into the power system in the last twenty years at an increasing rate (Kumar et al., 2016; Shair et al., 2021). Due to the capability of doubly fed induction generators (DFIG) to function at variable speeds and provide efficient energy conversion, they are being utilised heavily in wind turbines. Unfortunately, compared with conventional synchronous generators, power electronic converters are being used in DFIG-based wind farms, which alter fault current characteristics (Gautam et al., 2009; Okeke et al., 2023).

The effects of renewable energy infiltration in power system protection have been studied in many research works. Studies have revealed that short-circuit current levels are reduced, and system impedance profiles are manipulated by converter-based generation, which results in relay maloperation (Haddadi et al., 2021; Niknezhad & Sadeh, 2022). Also, relay settings are being impacted negatively by the dynamic operating conditions that come with the wind generation variability (Nwaigwe et al., 2019).

In the past, rate-of-change measurements, impedance-based methods and signal processing techniques, such as power swing detection techniques, have been the focus of research. The rate of change of impedance trajectory has always been the main reliance of the conventional PSB algorithms to differentiate between power swings and faults (Jafari et al., 2014; Kang & Gokaraju, 2016). The challenge is that when the impedance trajectories overlap because of renewable energy infeed currents, these methods can fail (Arumuga & Reddy, 2022).

Intelligent protection methods have been studied in more recent papers, whereby machine learning algorithms and artificial intelligence are employed extensively. This has allowed for the classification of system perturbations based on pattern recognition techniques using voltage and current signals (Chothani et al., 2014). Even though intelligent protection methods are proven to be interesting, the basic impact of renewable energy grid infiltration on relay characteristics should be clearly established (Yellajosula et al., 2019).

Specifically, the apparent impedance seen by the relays can be greatly impacted by the current injected by the wind farm. The current injection by the wind farms if installed very close to the location of the distance relay can change the impedance direction towards the protection zones of the relay and this usually led to unintended tripping of the relay during the occurrence of stable swings or cause delay in tripping of the relay during unstable swings (Haddadi et al., 2019; Niknezhad & Sadeh, 2022; Verzosa 2013).

Hence, it is very important to have good knowledge of the impact of the infiltration of renewable energy current on the protection schemes of relays, especially in ensuring system reliability (Fang et al., 2018; Telukunta et al., 2017).

Unlike previous studies that basically investigated traditional PSB and OST protection techniques based on deterministic criteria such as blinders, impedance trajectory analysis, and rate of change of impedance measurement, this paper improved on that by analysing the impact of increasing DFIG wind farm penetration on the performance of intelligent Distance Relay PSB and OST functions. This was done by developing and validating an intelligent hybrid signal processing machine learning framework for PSB and OST functions under changing renewable penetration levels.

Therefore, studying and understanding the impact of wind farm current injections on the way PSB and OST operate or function in a transmission system is the primary focus of this paper. So, the primary task of this paper is to examine the impact of wind farm current infeed on the ill-operation of distance relay protection schemes. The test object used is the modified IEEE 6-bus system incorporated with a 10MW doubly fed induction wind farm, and robust simulations of fault conditions and power swings have been used to examine how the impedance characteristics and protection performance of the distance relay are impacted by the renewable energy infeed.

MATERIALS AND METHODS

The Modified IEEE 6-Bus Network with Integrated 10 MW DFIG Wind Farm (NT-2)

The chosen system configuration is an IEEE 6-bus grid system with a 9 MW Doubly-Fed Induction Generator (DFIG) wind farm (NT-2), and the study was done using this network. The DFIG dynamic model consists of the standard rotor-side and grid-side converter control loops together with the crowbar protection mechanism available in the MATLAB/Simulink environment. The controller parameters were made to remain constant throughout all wind penetration cases to ensure a consistent comparison of the relay performance.

Figure 1 depicts the system configuration, which is a meshed transmission network having aggregated loads, interconnected transmission lines, and synchronous generators. The wind generation unit is connected to an intermediate bus.

A fraction of the traditional generation capacity was replaced by the wind farm, which allowed the variation of the wind penetration levels to be controlled while ensuring the

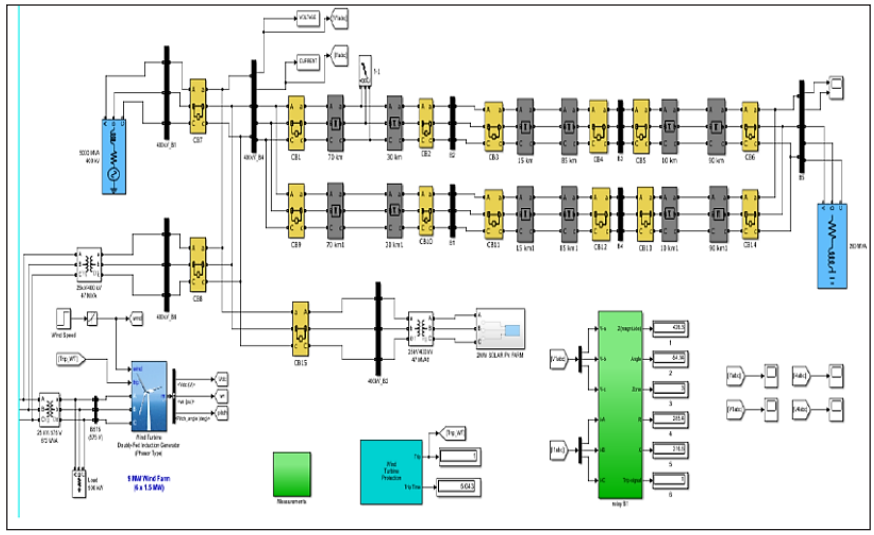


Figure 1. Modified IEEE 6-bus network having a 9 MW DFIG wind farm (NT-2) at an intermediate bus developed in the MATLAB/Simulink environment

total system loading conditions. Using a step-up transformer, the DFIG wind plant, which is rated 9 MW, is linked to the transmission system.

Equation 1 (Khan et al., 2020; Sabo et al., 2020) defines the wind penetration level.

$$WP = \frac{P_w}{P_t} \times 100\% \quad [1]$$

Where:

P_w is the wind farm's active power output, and

P_t is the power system's total generation.

Out-of-Step Tripping (OST) Unit

Differentiating between unstable and stable power swing conditions is very important in distance relay operation because it enhances system reliability, and OST is used to achieve this. The network ultimately restores its stability and assumes a new operating point after going through a small rotor angle displacement during a stable power swing. On the other hand, the system generator loses synchronism, stability, or experiences large-scale power outages during unstable power swings, which is an unwanted operating condition.

To verify the OST operation, a MATLAB/Simulink model was implemented to simulate the response of the relay during various power swing conditions. The model provides a benchmark for the minimum acceptable discrepancy between the monitored current and voltage signals. This benchmark mitigates the loss of synchronism of the generator, which is initiated by the huge frequency deviations during out-of-step situations.

MATLAB/Simulink was used to simulate this dynamic behaviour of the power network during these conditions to be able to examine the response of the relay and the OST protection operation performance.

Power Swing Blocking (PSB) Unit

The grouping of the stable swing, unstable swing conditions, and short-circuit faults during both transient and steady-state system conditions is made possible by the modelling of the PSB in the MATLAB/Simulink environment. By differentiating between short-circuit faults and stable power swing conditions, the PSB operation mitigates unwanted tripping of the relay under power swings. The detection mechanism of PSB employs the rate of change of the measured apparent impedance by the distance relay, and to measure the trajectory of the impedance during power swing conditions, a blinder-based blocking method is utilised.

The impedance sensing unit of the relay is integrated with this blocking logic so that during a short-circuit fault scenario, the relay gives a command to trip and isolate the faulted area so as to protect the equipment. On the other hand, the PSB unit prevents the trip command from causing unwanted operation of the relay.

To evaluate the duration of the perturbation, a MATLAB function block was developed, which is constrained to a maximum of 30 cycles. The impedance trajectory under fault conditions and power swing was modelled and simulated, which permits the historical datasets for development of the model to be obtained.

The tripping areas and the blocking for the relay is decided by the blinder block, and the grouping of the scenarios is defined by the rate of change of the impedance as estimated. A swing scenario is determined by how slow the trajectory movement of the impedance is, and, on the contrary, short-circuit faults are determined by fast variation in the impedance. The required response from the relay is defined by the decision logic block, and when the trajectory of the impedance goes into the region of the blinder under a swing, the tripping of the relay is blocked. However, when the trajectory of the impedance goes beyond the region of the blinder during fault conditions, a command to trip is released.

Integrated 9 MW Wind Turbine Modelling

As mentioned earlier, to understand the impact of wind power integration on the false-trip function of distance relays, especially under short-circuit fault conditions and power swing, a renewable energy infiltration was included in this investigation. In MATLAB/Simulink, a model of a 9 MW wind farm was implemented employing 6 units of doubly fed induction generator (DFIG) wind turbines of 1.5 MW each, which depicts a Type-C wind turbine set-up.

Mechanical Power of the Wind Turbine

The mechanical power produced by each 1.5 MW wind turbine unit was estimated using the fundamental wind turbine power equation (Ali et al., 2016; Veerasamy et al., 2021) implemented in MATLAB/Simulink as shown in Equation 2:

$$P = \frac{1}{2}\rho A_T V_w^3 C_p(\lambda, \beta) \quad [2]$$

Where:
ρ represents the air density $(1.225\ kg/m^3)$,
A_T is the turbine swept area (πr^2),
r is the radius of the turbine blade,
V_w^3 is the cube root of the wind speed (average value of 12m/s),
C_p represents the turbine power coefficient (efficiency of the machine),
λ is the tip speed ratio of the turbine, and
β is the blade pitch angle (assumed constant at 0^0).

The MATLAB/Simulink implementation of the mechanical power model (equation) is illustrated in Figure 2.

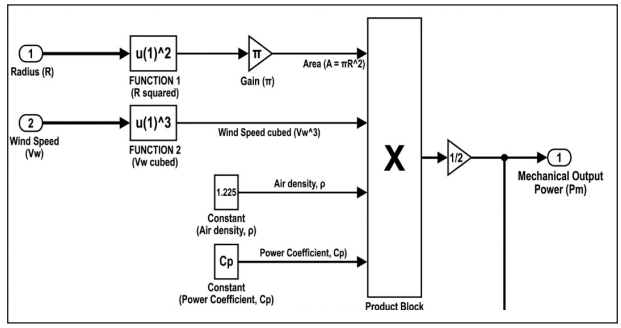


Figure 2. Wind turbine mechanical power model

Intermittent Tip-Speed Ratio Model of the Wind Turbine

The instantaneous tip-speed ratio plays an important role in calculating the turbine power coefficient. The mathematical model used for estimating the intermittent tip-speed ratio is given by Equation 3:

$$\frac{1}{\lambda_i} = \frac{1}{\lambda_T + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \quad [3]$$

Where the λ_i is the intermittent tip speed ratio of the turbine. The MATLAB/Simulink implementation of this model is shown in Figure 3.

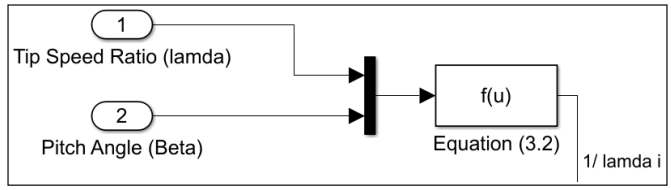


Figure 3. Intermittent tip speed ratio model of the wind turbine

Turbine Power Coefficient Model (Cp)

The turbine power coefficient C_p was modelled using two input parameters: the tip-speed ratio λ and the blade pitch angle β. The relationship is expressed as Equation 4:

$$C_p(\lambda,\beta) = C_1 * \left(\frac{C_2}{\lambda_w} - C_3\beta - C_4\beta^2 - C_5\right) * e^{\frac{-C_6}{\lambda_i}} \quad [4]$$

where C_1–C_6

The turbine design parameters are provided by the manufacturer and summarised in Table 1. The MATLAB/Simulink implementation of this subsystem is illustrated in Figure 4.

Table 1
Wind turbine modelling parameters

Symbol	Values	Parameters Description	Rating
C1	0.5	Rated Power	1.5 MW
C2	116	Rated Wind speed (V_W)	12 m/s
C3	0.4	Pitch angle (β)	0
C4	0	Optimal TSR (λ_T)	8.1
C5	5	Air Density (ρ)	1.225 kg/m³
C6	21	Wind Turbine blade length (r)	33.156 m
Optimal Tip Speed Ratio			2.932 m/s

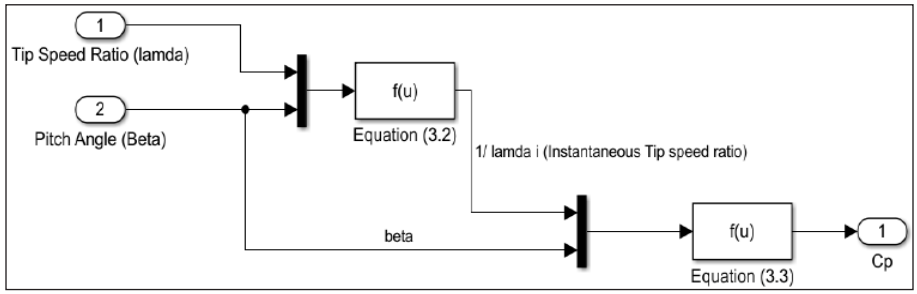


Figure 4. Wind turbine power coefficient model

Optimal Tip Speed Ratio Model of the Wind Turbine

The optimal rotational speed of the wind turbine was determined using the following Equation 5:

$$\lambda_T = \lambda_T^{op} = \frac{W_T + r_T}{V_w} \quad [5]$$

where ω_T represents the turbine rotational speed, r_T is the turbine blade radius, and V_w is the wind speed.

As seen in Figure 5, saturation blocks were incorporated in the Simulink implementation to ensure that the wind speed and tip-speed ratio parameters do not assume zero values, which could cause simulation instability. The lower limits were set to values close to zero, while the upper limits were set to infinity to avoid simulation errors.

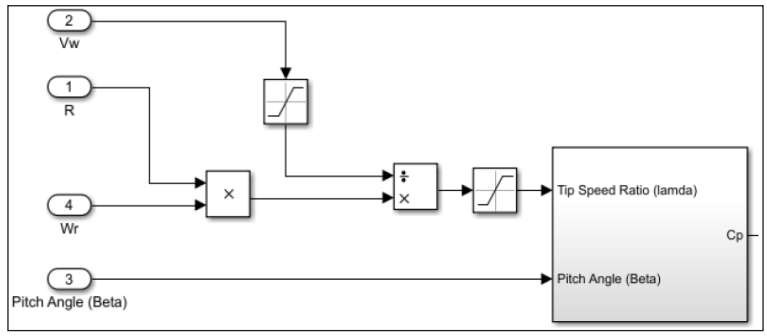


Figure 5. Optimal tip-speed ratio model of the wind turbine

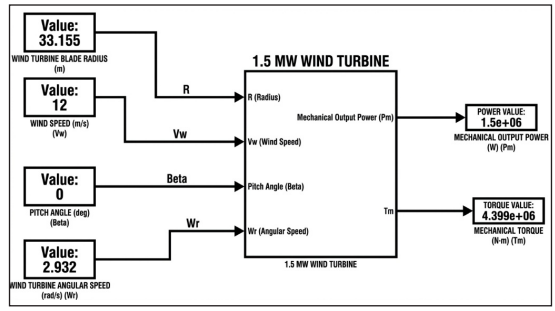
Wind Turbine System Integration

All subsystem models were integrated in MATLAB/Simulink to create the complete 1.5 MW wind turbine model. The model includes four input parameters and two output parameters, as illustrated in Figure 6(a), while the detailed subsystem interconnections are shown in Figure 6(b).

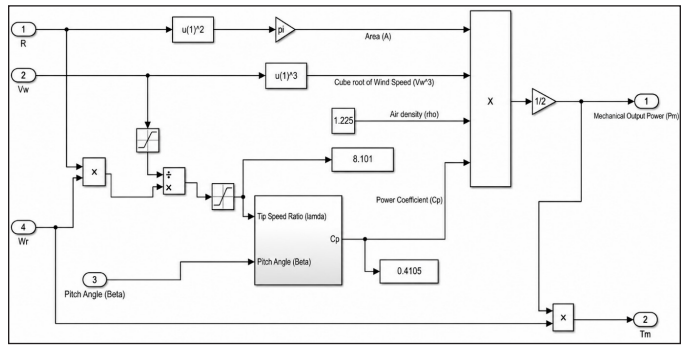
The mechanical power and torque outputs generated by each turbine under transient and steady-state conditions were recorded for further analysis. Six turbine subsystems were connected in parallel to achieve the required 9 MW wind power output before interfacing with the converter system and integrating with the benchmark test network.

Research Workflow

Figure 7 illustrates the overall research workflow adopted in this study. The workflow summarises the modelling process, dataset generation, and development of intelligent machine learning-based PSB and OST protection functions for distance relay systems.



(a)



(b)

Figure 6. 1.5 MW Wind turbine model: (a) Subsystem block diagram; (b) Internal model architecture

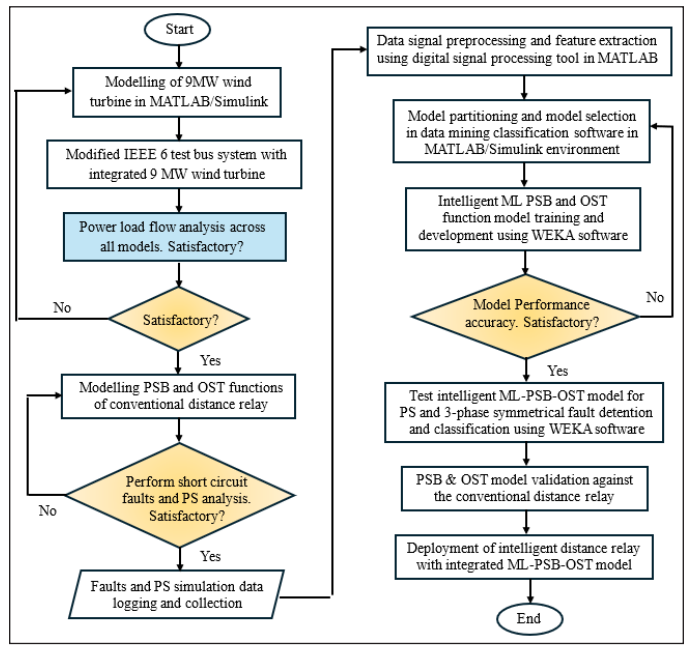


Figure 7. The research workflow

The proposed methodology enables accurate discrimination between power swing events and symmetrical short-circuit faults in renewable energy-integrated power systems.

RESULTS AND DISCUSSION

Power Load Flow Results from Test Bus Benchmark Model

The conducted power load flow analysis across the network presented the voltage magnitudes and phase angles at each bus. The swing bus (Bus 1), to which the 5,000 MVA generator is connected, maintained a voltage magnitude of 1 p.u (400 kV) and supplied the required reactive power (Q), as illustrated in Figure 8. The steady-state load flow analysis performed determined the power delivery capacity of each network component while ensuring compliance with stipulated operational standards. In addition, the direction of active (P) and reactive (Q) power flow across the branch elements is indicated in the load flow results.

The voltage magnitudes and phase angles obtained under steady-state conditions indicate stable and healthy network performance at each bus, supporting the integration of both renewable and non-renewable power sources. Furthermore, the active and reactive

	Block type	Bus type	Bus ID	Vbase (kV)	Vref (pu)	Vangle (deg)	P (MW)	Q (Mvar)	Qmin (Mvar)	Qmax (Mvar)	V_LF (pu)	Vangle_LF (deg)	P_LF (MW)	Q_LF (Mvar)	Block Name	Column2
1	Vsrc	swing	BUS_1	400	1	0	0	0	-Inf	Inf	1	0	1280000	0	5000 MVA 400 kV	
2	Bus	-	BUS_B4	400	1	0	0	0	0	0	0,2	0	0	0	Load Flow Bus1	
3	Bus	-	BUS_2	400	1	0	0	0	0	0	3,815	10,02	0	0	Load Flow Bus2	
4	Bus	-	BUS_5	400	1	0	0	0	0	0	3,814	8,02	0	0	Load Flow Bus3	
5	RLC load 2	-	BUS_3	400	1	0	210	150	0	0	3,014	7,49	0	0	260 MVA	
6	Bus	-	BUS_1	400	1	0	0	0	-Inf	Inf	3,815	11,53	0	0	Load Flow Bus5	
7	Bus	-	BUS_4	400	1	0	0	0	0	0	3,814	9,27	0	0	Load Flow Bus6	
8	Vprog	swing	"1"	120	1	0	0	0	-Inf	Inf	1	0	2,05	11,97	120 kV	
9	Bus	-	"2"	25	1	0	0	0	0	0	0,793	-29,01	0	0	25 kV // 575 V 6*2 MVA	
10	Vsrc	swing	"3"	400	1	0	0	0	-Inf	Inf	1	-127,7	5	293,45	2MW SOLAR PV FARM/10000 MVA 40...	
11	Bus	-	"4"	400	1	0	0	0	0	0	6,519	10,02	0	0	30 km	
12	Bus	-	"5"	400	1	0	0	0	0	0	8,399	9,71	0	0	15 km	
13	Bus	-	"6"	100	1	0	0	0	0	0	8,427	11,19	0	0	15 km	
14	Bus	-	"7"	100	1	0	0	0	0	0	8,427	7,97	0	0	10 km	
15	Bus	-	"8"	100	1	0	0	0	0	0	0,426	9,1	0	0	10 km	
16	Bus	-	"9"	120	1	0	0	0	0	0	0,9951	-0,02	0	0	120 kV // 25 kV 47 MVA	
17	Bus	-	"10"	25	1	0	0	0	0	0	0,7996	-28,93	0	0	Plant 2 MVA // 25 // 2.3 kV 2.6 MVA	
18	Bus	-	"11"	100	1	0	0	0	0	0	3,089	11,53	0	0	70 km	
19	Bus	-	"12"	100	1	0	0	0	0	0	8,362	11,53	0	0	30 km	
20	Bus	-	"13"	100	1	0	0	0	0	0	8,385	11,53	0	0	15 km	
21	Bus	-	"14"	100	1	0	0	0	0	0	8,428	9,27	0	0	85 km	
22	Bus	-	"15"	100	1	0	0	0	0	0	8,357	9,27	0	0	10 km	
23	Bus	-	"16"	100	1	0	0	0	0	0	8,428	7,49	0	0	90 km	
24	Bus	-	"17"	25	1	0	0	0	0	0	1,176	0,02	0	0	25 kV // 400 kV 47 MVA	
25	Bus	-	"18"	100	1	0	0	0	0	0	0,362	10,02	0	0	30 km	
26	Bus	-	"19"	100	1	0	0	0	0	0	8,385	10,02	0	0	15 km	
27	Bus	-	"20"	100	1	0	0	0	0	0	8,428	8,02	0	0	85 km	
28	Bus	-	"21"	100	1	0	0	0	0	0	8,428	8,02	0	0	10 km	
29	Bus	-	"22"	100	1	0	0	0	0	0	8,357	7,49	0	0	90 km	
30	Bus	-	"23"	25	1	0	0	0	0	0	5,878	0,03	0	0	25 kV // 400 kV 47 MVA	
31	Bus	-	"24"	100	1	0	0	0	0	0	4,089	0	0	0	70 km	

Figure 8. Steady-state power load flow results

power flows through the branch elements, including the transmission lines, demonstrate satisfactory operational performance consistent with the requirements of the grid code.

Wind Turbine Penetration Results

The short-circuit fault current penetration at different locations along the transmission line shows a slight increase in magnitude when the wind farm is integrated into the network compared to the scenario without renewable wind energy integration, as illustrated in Figure 9. This slight increase is sufficient to affect the preset impedance estimation used for the trip operation of the distance relay under power swing (PS) and symmetrical three-phase short-circuit fault conditions. The highest level of fault current penetration is observed at fault locations closest to the point of common coupling (PCC), specifically at bus-bar location 4.

The results presented in Table 2 show the current infeed penetration values recorded at the distance relay location, which corresponds to the Point of Common Coupling (PCC), from each interconnected source under different operating conditions. Under normal operating conditions, the synchronous generator (SG) provides the

Table 2
Current infeed estimations at the relay location

Operating Conditions	Current infeed penetrations SG and WF sources		
	SG	**WF**	**Z∠θ**
Normal	810	3.014	429.7-345.5i
Fault Distance (km)	**SG**	**WF**	**Z∠θ**
20	794.9	3.253	258.8-298.3i
40	55.33	3.014	251.6-294i
60	810.8	3.014	249.7-345i
80	810.8	3.014	249-345.51
100	-	-	-

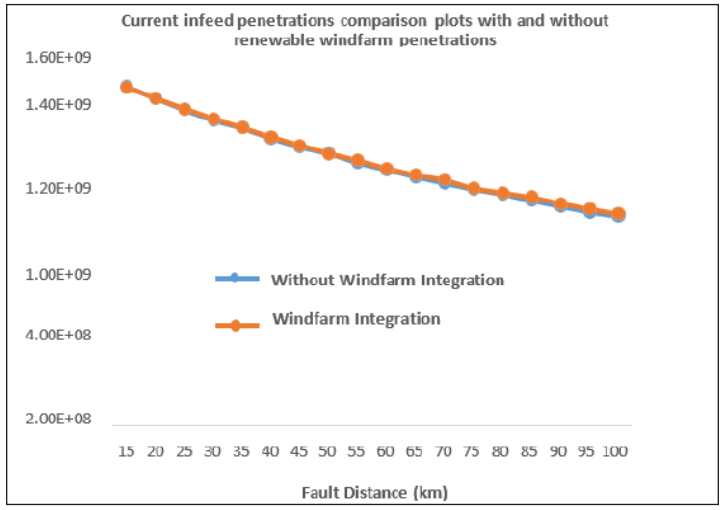


Figure 9. Comparison of current infeed penetration

largest current infeed contribution, while the wind farm contributes only a small fraction of the total current penetration used for impedance estimation by the distance relay, as illustrated in Figure 10.

Although the contribution from the renewable sources is relatively small, it plays a significant role in the mal-operation of the distance relay's Power Swing Blocking (PSB) and Out-of-Step Tripping (OST) functions.

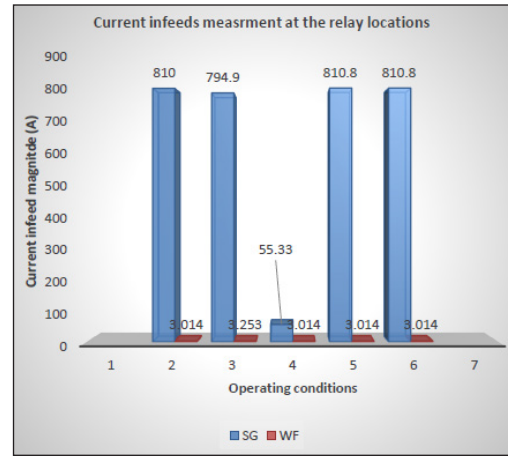


Figure 10. Distance relay infeed current measurement

Conventional Distance Relay Model Performance

The recorded dataset obtained from the conventional distance relay model implemented in MATLAB/Simulink at the Bus 4 terminal demonstrates the estimated line impedances under stable power swing (PS) and short-circuit fault conditions across the four network architectures. This section presents the observed impact of current infeed on the mal operation of the conventional distance relay during power swings and symmetrical short-circuit faults.

The measured power system parameters were used to estimate the impedance locus to support the distance relay's trip decision. The knowledge derived from transient fault voltage and current signals was further utilised to develop the modified intelligent PSB and OST functions.

Fluctuations in the magnitude of current infeed contributions from the integrated renewable energy sources present a significant challenge to the preset trip characteristics of the distance relay, as illustrated in the previously presented results.

Estimated Impedance Under Short-Circuit Faults

The estimated impedance observed by the conventional relay during short-circuit fault conditions at different fault locations is presented in Table 3. These values represent the phasor measurement parameters detected by the distance relay on transmission line L2. The real and imaginary components represent the estimated impedance obtained from the conventional distance relay model under different network topologies.

Slight variations in the impedance observed at different fault locations are mainly due to changes in the magnitude of the infeed current contribution to the faulted section of the line. The results indicate a slight decrease in the estimated impedance because of current infeed contributions from the integrated renewable energy sources.

Table 3
Estimated impedance under short-circuit fault scenarios

km	Estimated Impedance with Renewable (Ω)			Estimated Impedance with Renewable (Ω)			% Impedance Difference (Ω)
	R	X	Z∠θ	R	X	Z∠θ	
20	1.71	6.43	6.66∠84.56	1.94	6.37	7.67∠84.50	1.012
40	3.16	12.90	13.31∠84.54	3.93	12.70	14.31∠84.52	1.001
60	3.617	19.3	19.94∠84.64	5.705	19.0	20.94∠84.53	1.590
80	5.03	26.1	26.57∠84.60	7.81	25.3	27.57∠84.52	1.000
100	3.393	32.6	33.20∠84.72	8.83	31.8	34.20∠84.55	1.000
120	4.66	39.5	39.83∠84.70	10.66	38.3	39.99∠84.50	0.160
140	3.2	46.3	48.40∠84.70	11.22	45.0	48.66∠84.58	0.400
160	5.63	53.1	53.15∠84.81	11.84	51.8	53.20∠84.60	0.094
180	3.14	59.7	59.89∠81.88	12.56	58.7	60.01∠84.61	0.200
200	8.39	66.0	66.56∠84.95	12.41	65.9	67.08∠84.60	0.778

This reduction in impedance below the preset threshold value indicates the occurrence of short-circuit faults during power swing and fault conditions. The infeed current, therefore, compromises the preset impedance threshold of the distance relay by producing a lower estimated value, which leads to a faster movement of the impedance trajectory into the relay operating zone and may initiate an unintended trip command.

Estimated Impedance Under Power Swing Conditions

The results presented in Table 4 demonstrate the impact of renewable energy integration on the magnitude of power swings (PS), primarily influenced by the infeed current

Table 4
Estimated impedance under power swing scenarios

km	Estimated Impedance with Renewable (Ω)			Estimated Impedance with Renewable (Ω)			% Impedance Difference (Ω)
	R	X	Z∠θ	R	X	Z∠θ	
20	0.125	6.54	6.55∠87.95	0.13	6.54	7.55∠87.95	1.000
40	0.15	13.11	13.11∠87.95	0.15	13.11	14.19∠87.95	1.080
60	0.214	19.68	19.68∠87.95	0.21	19.68	19.99∠87.95	0.030
80	0.32	26.28	28.28∠87.95	0.32	26.28	29.23∠87.95	0.950
100	0.519	32.91	32.91∠87.95	0.52	32.91	32.99∠87.95	0.080
120	0.89	39.58	39.58∠87.95	0.88	39.58	39.88∠87.95	0.300
140	1.27	46.29	46.31∠87.95	1.27	46.29	46.71∠87.95	0.400
160	1.87	53.07	53.10∠87.95	2.9	53.11	53.19∠87.95	0.169
180	2.43	59.90	59.95∠87.92	4.3	60.13	60.28∠87.95	0.549
200	3.26	66.81	66.81∠87.95	6.03	67.23	67.50∠87.95	0.908

contributions from the interconnected sources. These results indicate that renewable energy integration has a noticeable effect on the estimated impedance observed by the distance relay during power swing events, particularly considering the intermittent nature of renewable energy sources.

A slight variation in the impedance estimated by the distance relay is observed between the two network topologies. This variation is mainly attributed to the additional infeed current contributions from the integrated renewable energy generating sources.

Power swing events often exhibit characteristics like those observed during symmetrical short-circuit faults. However, the key difference lies in the rate of change of the impedance locus. During power swing conditions, the impedance locus changes more gradually, resulting in a slower movement of the impedance trajectory toward the relay trip operating zone. As the estimated impedance gradually encroaches into the relay's trip region due to its reduced magnitude, the relay may issue an instantaneous trip command to the associated circuit breakers, potentially leading to unintended operations.

Intelligent Machine Learning-PSB Classifier Model Results

The development of the Power Swing Blocking (PSB) predictive classifier model is formulated as a binary classification problem aimed at discriminating between stable power swing (PS) conditions and symmetrical three-phase short-circuit faults occurring under similar operating conditions. The intelligent ML-PSB function classifier model for the distance relay was developed using balanced simulation scenario event records.

Each event scenario, representing either three-phase symmetrical faults or stable power swings, consisted of 3,078 historical event instances. Sixteen selected system operational features extracted from the simulated dataset were used to develop the intelligent ML-PSB predictive model. Several supervised machine learning algorithms were evaluated using the WEKA data mining environment to determine the most suitable classification model.

Intelligent ML-PSB Model Algorithm Selection

The best-performing classification algorithms were selected from different machine learning categories based on statistical evaluation indices. These indices included the true positive rate (TP), true negative rate (TN), false positive rate (FP), and false negative rate (FN). Additional performance metrics such as precision, recall, F-measure, receiver operating characteristic (ROC), and computational time were also considered.

Table 5 contains a summary of the algorithms illustrating superior performance benchmarked against the baseline model. This fulfils the chosen statistical performance criteria.

Table 5
Deployed ML algorithms for the ML-PSB function model

Categories	Algorithm names	Classification (%)		Precision	Recall	ROC Curve	Time (s)
		Correct	Incorrect				
Baseline	ZeroR*	49,862	50,138	0,498	0,499	0,498	0
	PART	100	0	1	1	1	0,05
	JRip	100	0	1	1	1	0,11
	DecisionTable	100	0	1	1	1	0,39
Trees	DecisionStump	100	0	1	1	1	0,03
	Hoeffding	100	0	1	1	1	0,09
	J48*	100	0	1	1	1	0
	RandomForest	100	0	1	1	1	0,34
	RandomTree*	100	0	1	1	1	0
	REPTree	100	0	1	1	1	0,03
	AdaBoostM1	100	0	1	1	1	0,03
	Bagging	100	0	1	1	1	0,12
	FilteredClassifier	100	0	1	1	1	0,06
Meta Ensemble	LogitBoost	100	0	1	1	1	0,2
	MultiClassClassifier	100	0	1	1	1	0,09
	RandomCommittee	100	0	1	1	1	0,03
	RandomSubSpace	100	0	1	1	1	0,05
	Vote	49,862	50,138	0,5	0,5	-0,5	0
Lazy	IBK	100	0	1	1	1	0
	LWL*	100	0	1	1	1	0
	Logistic*	100	0	1	1	1	0,02
Functions	MultilayerPerceptron	100	0	1	1	1	10,86
	SGDText	50	50	-	0,5	0,5	0,2
	SimpleLogistic	100	0	1	1	1	0,63
	SMO	100	0	1	1	1	0,05
Bayes	BayesNet	100	0	1	1	1	0,08
	NaiveBayes	99,346	0,6536	0,994	0,993	0,991	0,02
	NaiveBayesUpdateable*	99,3464	0,6536	0,994	0,993	0,991	0

ML-PSB Classifier Model Performance Evaluation

Optimal model selection criteria were employed to further examine the chosen algorithms as highlighted in Table 6. This was done to recognise the most suitable classifier model which can fulfil the operational challenges of the distance relay in the WEKA experimental

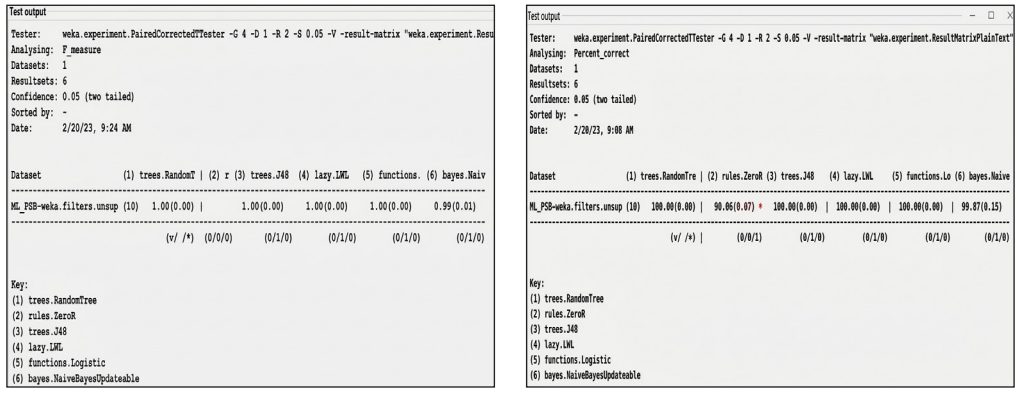
framework. Special attention was paid to minimising the computational time while still trying to retain the accuracy of the classification.

This requirement is critical because the operating time of conventional distance relay trip decisions is approximately one cycle (20 ms). The novelty of the present study lies in improving this response time by selecting an algorithm capable of near-instantaneous response, achieving an operating time of approximately 10 ms.

Four evaluation criteria were used to determine the most suitable ML-PSB model: F-measure, classification accuracy, recall, and computational response time. Statistical comparisons were performed using a 0.05 significance level with t-test indicators.

The results show that the RandomTree algorithm achieved superior performance, demonstrating approximately 95% statistical significance across the evaluated performance metrics compared with the other five candidate algorithms. As illustrated in Figure 11, RandomTree achieved the highest performance in terms of F-measure, precision, recall, and computational efficiency.

The F-measure, calculated from the harmonic mean of precision and recall, further confirmed the RandomTree algorithm as the most suitable ML-PSB classifier model for detecting and discriminating stable power swings from three-phase symmetrical faults.



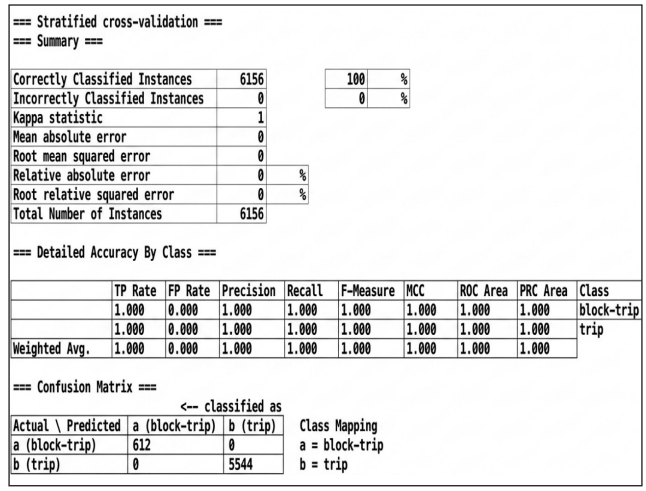
(a) (b)

Figure 11. Statistical algorithm model comparison for the PSB function based on (a) F-measure, (b) accuracy, (c) recall, and (d) response time

RandomTree-PSB Classifier Model Performance

The performance evaluation results of the developed RandomTree-PSB model using an independent test dataset are presented in Figure 12. This dataset was not used during the model training stage, allowing the model's generalisation capability to be properly assessed.

The performance indicators were derived from the confusion matrix results, where the diagonal elements represent correctly classified event scenarios, while the off-diagonal elements represent misclassified events.



=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	6156	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0	
Root mean squared error	0	
Relative absolute error	0	%
Root relative squared error	0	%
Total Number of Instances	6156	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	block-trip
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	trip
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

	<-- classified as		
Actual \ Predicted	a (block-trip)	b (trip)	Class Mapping
a (block-trip)	612	0	a = block-trip
b (trip)	0	5544	b = trip

Figure 12. RandomTree-PSB model test performance results

The intelligent PSB model successfully classified all 6,156 test instances without misclassification, demonstrating a classification accuracy of 100%. A total of 612 event scenarios were correctly classified as stable power swings (true positives), resulting in a trip-blocking command issued by the ML-PSB function. Among all the classifications that were recorded, there were no false positive classifications.

In the same vein, the model correctly classified 5,544 cases as fault events, which permit the ML-PSB scheme to release a trip signal to the relevant circuit breakers without hesitation to protect the system. The high number of correctly classified scenarios shows that the RandomTree-PSB model was able to efficiently distinguish between symmetrical short-circuit faults and stable power swing scenarios.

The results went further to show that the proposed model ensured good classification accuracy across different operating scenarios, which indicates that it learned the underlying characteristics of both events rather than ordinarily memorizing the training data. Adding to that, the intelligent PSB scheme acted considerably faster than the traditional PSB function and the relay operating time was reduced from about 20 ms to approximately 10 ms. This can contribute to faster fault isolation and enhanced protection performance. As shown in Figure 13, the decision-making process of the RandomTree-PSB model is represented by the generated decision tree.

The pseudocode for the RandomTree algorithm is presented below:

1. **Input:** Test dataset
   a. Apply each randomly generated decision tree to the test dataset and record its predicted output.
   b. Count the number of votes received for every class label.
   c. Select the class with the highest number of votes as the final prediction.
2. **Output:** Final predicted class label.

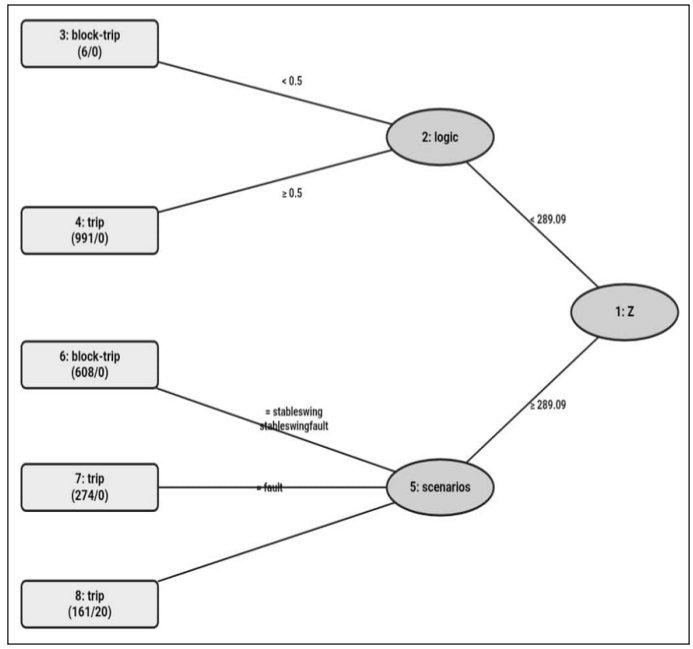


Figure 13. RandomTree-PSB decision tree

Intelligent Machine Learning–OST Classifier Model Results

The second intelligent protection function developed in this study is the Machine Learning-based Out-of-Step Tripping (ML-OST) classifier. Like the PSB model, the ML-OST function is formulated as a binary classification model designed to distinguish between stable and unstable power swing conditions during power system disturbances.

The dataset used for this model consisted of 612 stable power swing records and 608 unstable swing records obtained from simulated system disturbances. These datasets were used to train and evaluate the ML-OST classifier for accurate detection and classification of swing stability conditions.

Intelligent ML-OST Algorithm Selection

Multiple machine learning algorithms were evaluated for the ML-OST classifier development using the same performance evaluation metrics applied to the ML-PSB model. These metrics included classification accuracy, precision, recall, ROC performance, and computational time.

Algorithms that demonstrated accuracy above the baseline model and lower computational time were selected for further evaluation. The machine learning algorithms are depicted in Table 6.

In developing the intelligent OST classifier, the sensitivity requirements of the traditional OST function were taken into consideration to make sure that stable and unstable

Table 6
Deployed ML algorithms for the ML-OST function model

Categories	Algorithm names	Classification (%)		Precision	Recall	ROC Curve	Time (s)
		Correct (%)	Incorrect (%)				
Baseline	ZeroR*	49,862	50,138	0,498	0,499	0,498	0
	PART	100	0	1	1	1	0,05
	JRip	100	0	1	1	1	0,11
	DecisionTable	100	0	1	1	1	0,39
Trees	DecisionStump	100	0	1	1	1	0,03
	Hoeffding	100	0	1	1	1	0,09
	J48*	100	0	1	1	1	0
	RandomForest	100	0	1	1	1	0,34
	RandomTree*	100	0	1	1	1	0
	REPTree	100	0	1	1	1	0,03
Meta Ensemble	AdaBoostM1	100	0	1	1	1	0,03
	Bagging	100	0	1	1	1	0,12
	FilteredClassifier	100	0	1	1	1	0,06
	LogitBoost	100	0	1	1	1	0,2
	MultiClassClassifier	100	0	1	1	1	0,09
	RandomCommittee	100	0	1	1	1	0,03
	RandomSubSpace	100	0	1	1	1	0,05
	Vote	49,862	50,138	0,5	0,5	-0,5	0
Lazy	IBK	100	0	1	1	1	0
	LWL*	100	0	1	1	1	0
	Logistic*	100	0	1	1	1	0,02
Functions	MultilayerPerceptron	100	0	1	1	1	10,86
	SGDText	50	50	-	0,5	0,5	0,2
	SimpleLogistic	100	0	1	1	1	0,63
	SMO	100	0	1	1	1	0,05
Bayes	BayesNet	100	0	1	1	1	0,08
	NaiveBayes	99,346	0,6536	0,994	0,993	0,991	0,02
	NaiveBayesUpdateable*	99,3464	0,6536	0,994	0,993	0,991	0

power swing conditions are identified correctly. Improving the performance of the OST protection operation is important because an incorrect relay operation during a power swing can initiate additional relay actions elsewhere in the network, which increases the chances of widespread system perturbation and, in severe cases, leads to a system blackout.

ML-OST Model Performance Evaluation

The selected algorithms were subsequently evaluated using a range of machine learning performance metrics. These metrics included classification accuracy, precision, recall, F-measure, and computational response time. The objective was to assess both the classification capability and the practical suitability of each algorithm for protection applications.

The comparative results obtained are presented in Figure 14. Among the evaluated algorithms, the RandomTree model achieved the most favourable overall performance, thereby outperforming the other shortlisted techniques across the selected evaluation criteria.

RandomTree-OST Classifier Model Performance

Among the five machine learning algorithms evaluated, the RandomTree model gave the best overall results for the proposed intelligent OST operation. The model achieved a classification accuracy of 99.7%, which successfully classified 1,218 power swing events

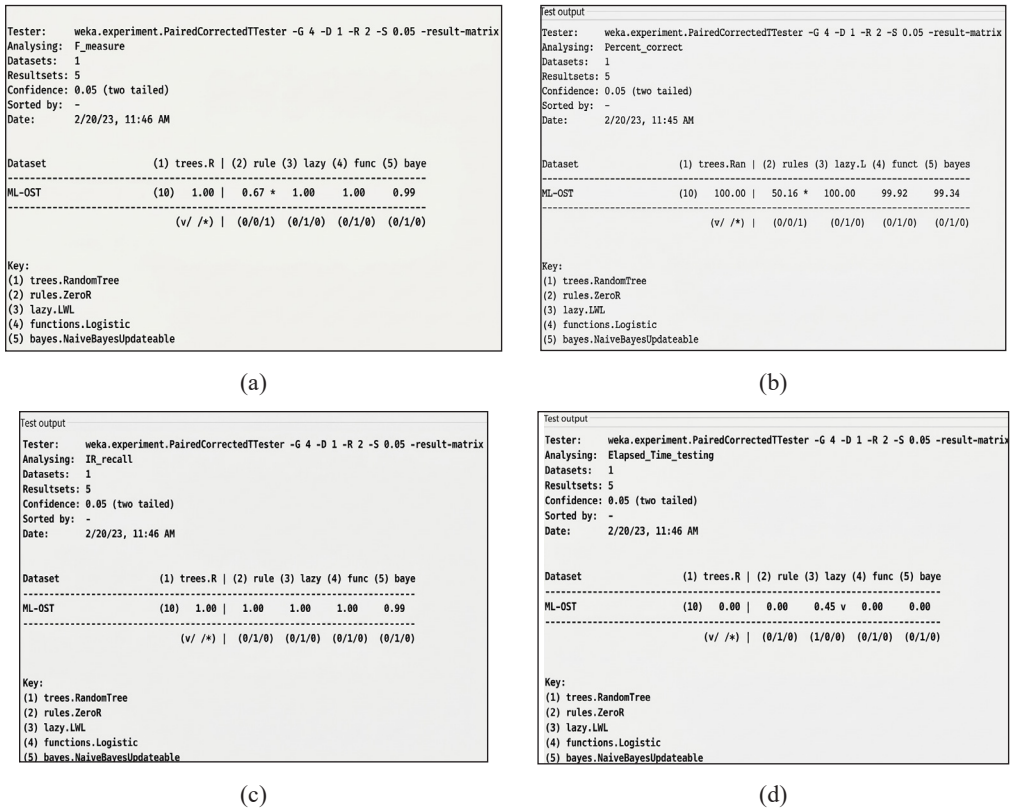


Figure 14. Statistical comparison of the evaluated algorithms for the OST operation based on (a) F-measure, (b) classification accuracy, (c) recall, and (d) response time

into stable and unstable swing categories. From the simulation, only two events were misassigned to the wrong class during testing.

The classification results obtained are summarised in the confusion matrix shown in Figure 15. The values along the main diagonal depict events that were classified properly, while the off-diagonal entries correspond to misclassified cases.

Out of the evaluated events, 612 stable power swing scenarios were identified properly, and this resulted in the OST operation blocking an unwanted trip signal. Only two stable power swing events were assigned to the improper category.

In the case of unwanted scenarios, all 608 events were identified properly by the model. During these conditions, the OST operation gave a trip command to the responsible circuit breakers to open the affected section of the network and maintain system security. Here, no unstable power swing scenarios were misclassified as stable swings during the evaluation process.

The decision tree structure generated by the RandomTree-OST classifier is shown in Figure 16. The model primarily utilises frequency and magnitude-related system features to discriminate between stable and unstable power swing conditions.

Voltage and current signal features extracted from the simulated datasets contribute varying levels of decision information across different stages of the classification process. The developed RandomTree-OST model demonstrated strong adaptability to different network configurations and varying infeed current contributions resulting from renewable energy integration.

The results confirm that the intelligent OST model significantly improves the response time and classification performance compared with conventional OST protection schemes. This improvement reduces the likelihood of cascading protection failures that could otherwise lead to large-scale grid disturbances.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1218               99.8361 %
Incorrectly Classified Instances         2                0.1639 %
Kappa statistic                          0.9967
Mean absolute error                      0.0016
Root mean squared error                  0.0405
Relative absolute error                  0.3279 %
Root relative squared error              8.0978 %
Total Number of Instances             1220

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.997    0.000    1.000      0.997    0.998      0.997    0.998     0.998     block-trip
               1.000    0.003    0.997      1.000    0.998      0.997    0.998     0.997     trip
Weighted Avg.  0.998    0.002    0.998      0.998    0.998      0.997    0.998     0.998

=== Confusion Matrix ===

   a   b   <-- classified as
 610   2 |   a = block-trip
   0 608 |   b = trip
```

Figure 15. RandomTree-OST model performance results

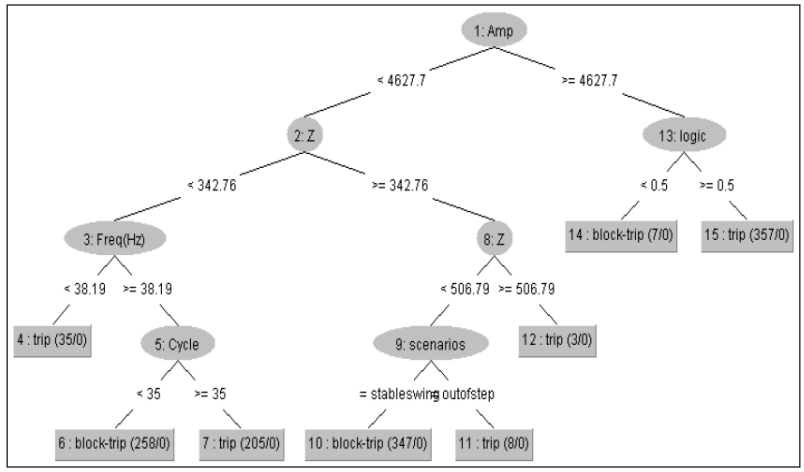


Figure 16. RandomTree ML-OST decision tree

CONCLUSION

The following conclusions are drawn from the results presented in this study:

The first objective of the study demonstrated that current infeed contributions from integrated renewable energy sources (wind farms) significantly influence the magnitude of power swings within the power system. These contributions affect the impedance estimation of the distance relay and may lead to relay maloperation during stable power swing conditions.

To address this limitation, an intelligent RandomTree-PSB classifier model was developed to enhance the performance of the conventional Power Swing Blocking function. The proposed model achieved 100% classification accuracy with a response time of approximately 10 ms, representing a significant improvement over the conventional relay response time of approximately 20 ms.

The intelligent RandomTree-OST classifier model effectively discriminates between stable and unstable power swing conditions. The model achieved a classification accuracy of 99.7% with an operating time of approximately 10 ms, demonstrating improved performance compared with conventional OST relay functions.

Overall, the proposed intelligent PSB and OST models significantly improve the reliability, speed, and accuracy of distance relay protection in renewable energy-integrated power systems.

ACKNOWLEDGEMENT

We, the authors, would like to thank Prof. Ir. Dr Mohammad Lutfi Othman for his guidance, support and constructive feedback throughout the course of this research. His advice and

encouragement were greatly appreciated and contributed significantly to the completion of this work. We also wish to acknowledge Assoc. Prof. Dr Hashim Hizam and Prof. Ir. Dr Mohd Zainal Abdin Ab Kadir for their valuable comments, technical input and suggestions during the study. Their contributions helped strengthen various aspects of the research. The authors further extend their appreciation to the academic and technical staff of the Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, and to members of the Advanced Lightning, Power and Energy Research (ALPER) Group for their assistance, support and the positive research environment provided throughout this work.

REFERENCES

Abdi-Khorsand, M., & Vittal, V. (2016). Modelling protection systems in time-domain simulations: A new method to detect mis-operating relays for unstable power swings. *IEEE Transactions on Power Systems, 32*(4), 2790-2798. https://doi.org/10.1109/TPWRS.2016.2628726

Adak, S. (2024). Power factor analysis of grid-connected solar inverter under different irradiance levels throughout the day. *Energies, 17*(15), Article 3632. https://doi.org/10.3390/en17153632

Ali, N. H., Ali, B. M., Basri, O., Husin, N. L., & Abdel-Latif, K. M. (2016). Protection of smart substation based on WLAN complies with IEC 61850 using travelling wave analysis. *Electric Power Systems Research, 140*, 20-26. https://doi.org/10.1016/j.epsr.2016.07.004

Arumuga, M., & Reddy, M. J. B. (2022). Distance protection methodology for detection of faulted phase and fault along with power swing using apparent impedance. *IEEE Access, 10*, 43583-43597. https://doi.org/10.1109/ACCESS.2022.3168563

Audu, N. A., Alphaeus, O., & Adamu, T. (2018). Effect of inertia constant on generator frequency and rotor angle. *Engineering and Applied Science, 3*(1), Article 6. https://doi.org/10.11648/j.eas.20180301.12

Chaitanya, M. N., & Rao, G. S. K. (2019). A novel scheme for supervising third zone distance relay operation during power swings. In *2019 International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 41-46). IEEE.

Chothani, N. G., Bhalja, B. R., & Parikh, U. B. (2014). New support vector machine-based digital relaying scheme for discrimination between power swing and fault. *IET Generation, Transmission & Distribution, 8*(1), 17-25. https://doi.org/10.1049/iet-gtd.2013.0020

Dreidy, M., Mokhlis, H., & Mekhilef, S. (2017). Inertia response and frequency control techniques for renewable energy sources: A review. *Renewable and Sustainable Energy Reviews, 69*, 144-155. https://doi.org/10.1016/j.rser.2016.11.170

Fang, Y., Jia, K., Yang, Z., Li, Y., & Bi, T. (2018). Impact of inverter-interfaced renewable energy generators on distance protection and an improved scheme. *IEEE Transactions on Industrial Electronics, 66*(9), 7078-7088. https://doi.org/10.1109/TIE.2018.2873521

Fischer, N., Benmouyal, G., Hou, D., Tziouvaras, D., Byrne-Finley, J., & Smyth, B. (2012). Tutorial on power swing blocking and out-of-step tripping. In *Proceedings of the 39th Annual Western Protective Relay Conference*.

Gautam, D., Vittal, V., & Harbour, T. (2009). Impact of increased penetration of DFIG-based wind turbine generators on transient and small signal stability of power systems. *IEEE Transactions on Power Systems, 24*(3), 1426-1434. https://doi.org/10.1109/TPWRS.2009.2021234

Gonzalez-Longatt, F., Adiyabazar, C., & Martinez, E. V. (2021). Setting and testing of the out-of-step protection at Mongolian transmission system. *Energies, 14*(23), Article 8170. https://doi.org/10.3390/en14238170

Haddadi, A., Farantatos, E., Kocar, I., & Karaagac, U. (2021). Impact of inverter-based resources on system protection. *Energies, 14*(4), Article 1050. https://doi.org/10.3390/en14041050

Haddadi, A., Kocar, I., Karaagac, U., Gras, H., & Farantatos, E. (2019). Impact of wind generation on power swing protection. *IEEE Transactions on Power Delivery, 34*(3), 1118-1128. https://doi.org/10.1109/TPWRD.2019.2896135

Jafari, R., Moaddabi, N., Eskandari-Nasab, M., Gharehpetian, G., & Naderi, M. (2014). A novel power swing detection scheme independent of the rate of change of power system parameters. *IEEE Transactions on Power Delivery, 29*(3), 1192-1202. https://doi.org/10.1109/TPWRD.2013.2297625

Kang, D., & Gokaraju, R. (2016). A new method for blocking third-zone distance relays during stable power swings. *IEEE Transactions on Power Delivery, 31*(4), 1836-1843. https://doi.org/10.1109/TPWRD.2016.2520394

Khan, A., Hizam, H., Abdul Wahab, N. I., & Othman, M. L. (2020). Solution of optimal power flow using non-dominated sorting multi-objective based hybrid firefly and particle swarm optimisation algorithm. *Energies, 13*(16), Article 4265. https://doi.org/10.3390/en13164265

Kumar, V., Pandey, A., & Sinha, S. (2016). Grid integration and power quality issues of wind and solar energy systems: A review. In *ICETEESES 2016* (pp. 71-80). IEEE. https://doi.org/10.1109/ICETEESES.2016.7581355

Mooney, P. J., & Fischer, N. (2006). Application guidelines for power swing detection on transmission systems. In *Power Systems Conference* (pp. 159-168). IEEE. https://doi.org/10.1109/PSAMP.2006.285385

Nasab, M. R., & Yaghobi, H. (2020). A real-time out-of-step protection strategy based on instantaneous active power deviation. *IEEE Transactions on Power Delivery, 36*(6), 3590-3600. https://doi.org/10.1109/TPWRD.2020.3045276

Niknezhad, M., & Sadeh, J. (2022). Analysis of wind farm penetration on power swing detection in distance relays. *IET Renewable Power Generation, 16*(2), 375-388. https://doi.org/10.1049/rpg2.12333

Nwaigwe, K., Mutabilwa, P., & Dintwa, E. (2019). An overview of solar power integration into electricity grids. *Materials Science for Energy Technologies, 2*(3), 629-633. https://doi.org/10.1016/j.mset.2019.07.002

Okeke, C. O., Lutfi, O. M., Hashim, H., Kadir, M. Z. A., Wahab, N. I. A., Emmanuel, O., Nwagbara, S., & Chimeleze, C. (2023). Renewable energy impact on distance relay power swing blocking and fault discrimination: A review. *Pertanika Journal of Science & Technology, 31*(6), 2887-2913. https://doi.org/10.47836/pjst.31.6.14

Paudyal, S., Ramakrishna, G., & Sachdev, M. S. (2010). Application of equal area criterion conditions in the time domain for out-of-step protection. *IEEE Transactions on Power Delivery, 25*(2), 600-609. https://doi.org/10.1109/TPWRD.2009.2032326

Sabo, A., Abdul Wahab, N. I., Othman, M. L., Jaffar, M. Z. A. M., Beiranvand, H., & Akçayol, M. (2020). Application of neuro-fuzzy controller to replace SMIB and interconnected multi-machine power system stabilisers. *Sustainability, 12*(22), Article 9591. https://doi.org/10.3390/su12229591

Shair, J., Li, H., Hu, J., & Xie, X. (2021). Power system stability issues, classifications and research prospects in the context of high penetration of renewables and power electronics. *Renewable and Sustainable Energy Reviews, 145*, Article 111111. https://doi.org/10.1016/j.rser.2021.111111

Telukunta, V., Pradhan, J., Agrawal, A., Singh, M., & Srivani, S. G. (2017). Protection challenges under bulk penetration of renewable energy resources in power systems: A review. *CSEE Journal of Power and Energy Systems, 3*(4), 365-379. https://doi.org/10.17775/CSEEJPES.2017.00030

Veerasamy, V., Abdul Wahab, N. I., Ramachandran, R., Othman, M. L., Hizam, H., Devendran, V. S., Irudayaraj, A. X. R., & Vinayagam, A. (2021). Recurrent network-based power flow solution for voltage stability assessment and improvement with distributed energy sources. *Applied Energy, 302*, Article 117524. https://doi.org/10.1016/j.apenergy.2021.117524

Verzosa, Q. (2013). Realistic testing of power swing blocking and out-of-step tripping functions. In *Proceedings of the 66th Annual Conference for Protective Relay Engineers* (pp. 420-449). IEEE. https://doi.org/10.1109/CPRE.2013.6822056

Wei, Q., Han, X., Guo, W., Yang, M., & He, N. (2014). The principle of absolute rotor angle control and its effect on suppressing inter-area low-frequency oscillations. *International Journal of Electrical Power & Energy Systems, 63*, 1039-1046. https://doi.org/10.1016/j.ijepes.2014.07.006

Yellajosula, J. R. A. K., Wei, Y., Grebla, M., Paudyal, S., & Mork, B. A. (2019). Online detection of power swing using approximate stability boundaries. *IEEE Transactions on Power Delivery, 35*(3), 1220-1229. https://doi.org/10.1109/TPWRD.2019.2941522

Zhang, Z., Jacques, A., & Voloh, I. (2016). Challenges in application of distance relays under power swing conditions. In *Proceedings of the 69th Annual Conference for Protective Relay Engineers (CPRE)*. IEEE. https://doi.org/10.1109/CPRE.2016.7914882

Pertanika J. Sci. & Technol. 34 (4): 2385 - 2401 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Characterisation of Bamboo Fibre-Reinforced Polyurethane Foam Composites with Varying Fibre Sizes

A.M. Radzi[1], M.Y.M. Zuhri[1,2]*, S. Ayu Rafiqah[3], R.A. Ilyas[4,5], Muhammad Asyraf Muhammad Rizal[5,6], Ahmad Safwan Ismail[7], A.B.M. Supian[8], and Agusril Syamsir[9]

[1]*Institute of Tropical Forestry and Forest Products (INTROP), Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[2]*Advanced Engineering Materials and Composites Research Centre, Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[3]*Engineering and Technology Division, Malaysian Rubber Board (MRB), 47000 Sungai Buloh, Selangor, Malaysia*

[4]*Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

[5]*Centre for Advanced Composite Materials (CACM), Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

[6]*Department of Applied Mechanics and Design, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

[7]*Chemical Engineering Programme, Faculty of Engineering, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia*

[8]*Centre for Defence Research and Technology (CODRAT), National Defence University of Malaysia, Sungai Besi Camp, 57000 Kuala Lumpur, Malaysia*

[9]*Institute of Energy Infrastructure, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia*

ARTICLE INFO

Article history:
Received: 25 May 2026
Accepted: 09 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.21

E-mail addresses:
mohdradzi@upm.edu.my (A.M. Radzi)
zuhri@upm.edu.my (MYM Zuhri)
ayu.rafiqah@lgm.gov.my (S. Ayu Rafiqah)
ahmadilyas@utm.my (R. A. Ilyas)
muhammadasyraf.mr@utm.my (Muhammad Asyraf Muhammad Rizal)
ahmadsafwanismail@ums.edu.my (Ahmad Safwan Ismail)
mohdsupian7779@gmail.com (A.B.M. Supian)
agusril@uniten.edu.my (Agusril Syamsir)
* Corresponding author

ABSTRACT

This study investigates the influence of varying bamboo fibre (BF) sizes on the thermal, mechanical, and physical characteristics of composites made from bamboo fibre and polyurethane foam (PF). BF/PF composites were fabricated using fibre sizes less than 150µm, and within 150–300 µm, 300-425 µm, and 425-600 µm ranges. Thermogravimetric analysis (TGA) identified three primary decomposition phases, corresponding to moisture removal, hemicellulose breakdown, and cellulose degradation. The results demonstrated that incorporating different BF sizes enhanced the

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

composites' thermal stability. The thermal stability of BF/PF composites was improved; the results show that BFP3 (299.07°C) has a higher onset of degradation compared to neat and other composites. In terms of mechanical properties, larger fibre sizes improved both flexural and compressive strength, with the 300-425 µm (BFP3) size yielding the highest flexural (0.192 MPa) and compressive (0.232 MPa) strengths. However, water uptake, moisture content, and thickness swelling increased with fibre size. Notably, composites using fibres smaller than 150 µm (BFP1) recorded the lowest values in water absorption (42.54%), thickness swelling (3.66%), and moisture content (3.48%).

Keywords: Bamboo fibre, fibre size, foam composites, mechanical properties, polyurethane foam

INTRODUCTION

The compatibility and ease of availability of raw materials, combined with a polymer matrix (thermosetting or thermoplastic), allow for balanced properties. Natural fibre-based polymer composites have attracted more interest as a possible replacement for traditional materials (Ilyas et al., 2021; Sanjay et al., 2018; Sanmuham et al., 2021). Natural fibre composites have the potential to further the development of environmentally friendly materials; researchers are interested in the subject of study. Due to their enhanced productivity, variety of application designs and easy access to raw materials, these fibres can enhance the manufacturing sector (Bahari & Krause, 2016; Saheb & Jog, 1999; Supian et al., 2018). To meet the growing demands for environmental sustainability, researchers are now working on reducing the use of synthetic fibres (carbon and glass fibres) in manufacturing applications. By replacing more sustainable materials for natural resources as the main component in the reinforcement of polymer composites, this material may greatly decrease the consumption of fossil resources (Syafiq et al., 2020; Syafiq et al., 2021; Zia et al., 2015). Apart from that, the use of natural resources can reduce crop waste and open burning by farmers, offering a manageable waste solution with lower environmental impact (Ishak et al., 2013; Nadlene et al., 2016; Suhot et al., 2021). Through this innovative approach, the growing concerns of issues such as global warming and waste management can be reduced. The implementation of various initiatives, such as eco-friendly projects, trash sorting by type for appropriate disposal, and the marketing of eco-friendly materials in diverse applications, has resulted from this increased awareness. (Beran et al., 2021; Davoodi et al., 2012; Huzaifah et al., 2017; Jawaid et al., 2022; Radzi et al., 2018). Bamboo, a fast-growing member of the grass family (Gramineae), shares common characteristics with species such as corn, sugarcane and rice. Numerous genera, including Bambusa, Dinochloa, Gigantochloa, Dendrocalamus, Holttumochloa, Schizostachyum, Kinabaluchloa, Melocanna, Chusquea, Maclurochloa, Phyllostachys, Sphaerobambos, Soejatmie, and Thyrsostambos, among others, are planted and grow naturally (Radzi et al., 2022).

Natural resources that have the potential to be used as alternative materials in polymer composites come from the bamboo tree. Bamboo cultivation is an easy-to-manage, cost- and time-efficient practice that provides a wide range of valuable resources, especially in high-quality fibres. In the bamboo family, nearly 200 bamboo species are found across Southeast Asia, either naturally occurring or cultivated, with notable presence in Malaysia, Indonesia, Myanmar, and Papua New Guinea (Anwar et al., 2011; Siam et al., 2019). In the tropical environment of Malaysia, bamboo plants can grow widely along rivers, in the forest and in undeveloped areas or wasteland. Various species of bamboo can be found in the wild and in commercial plantations for any purpose (Siam et al., 2019). Of the 63 species of bamboo, only 13 are commonly used in traditional or commercial applications to make various products (Anwar et al., 2011). In general, the anatomy of a bamboo tree has several parts such as roots, rhizomes, branches, leaves and shoots. This bamboo consists of approximately 10% vessels, 40% fibre, 50% parenchyma and a sieve tube (Chen et al., 2021; Kadivar et al., 2020). In fact, the bamboo tree is not a flowering type like most other trees. However, there are a few types of bamboo tree produce flowers, such as Phyllostachys edulis (Moso) and Phyllostachys elegans, which can produce flowers due to changes in environmental conditions (Roslan et al., 2018). In recent years, the utilisation of bamboo has increased due to its remarkable properties, such as strength, sustainability and versatility, making it a valuable resource in manufacturing industries (textiles, construction materials, automotive parts, etc.). These trends also create opportunities for small industries to generate income through the fabrication and commercialisation of handicraft products. Besides, bamboo has its own characteristics, making it one of the choices integrated with the polymer matrix as reinforcement. Due to its potential, bamboo fibre (BF) has been studied widely as reinforcement in polymer composites. In addition, BF is a low-cost, eco-friendly material, and the ease of extraction and treatment also contributes to its attractiveness for a variety of products (Hong et al., 2019; Md Shah et al., 2016). BF has a thick cell wall and microfibrils capable of forming natural glass fibre, as well as stiffness and strength (Al-oqla & Sapuan, 2014; Chee et al., 2017; Hassan et al., 2020; Ibrahim et al., 2015). Besides, bamboo comprises approximately 60% cellulose and lignin and typically features microfibril angles that fall within the range of 2 to 10 degrees (Radzi et al., 2022).

In general, polyurethane materials are mostly used for interior car applications, building materials, furniture, and electrical insulation (Azmi et al., 2012; Kairytė et al., 2018). This polymer is highly adaptable and presently finds application in a broad array of uses. This polymer stands out due to its exceptional wear resistance and favourable tribological and mechanical qualities when incorporated into a polymer composite. Polyurethane offers the versatility to be cast under either high or low pressure, injection moulded, extruded, pressed, or foamed, and it can take on a multitude of shapes, presenting both rigid and flexible options. Typically, this polyurethane is mixed with synthetic fibres, and indirectly,

it raises the expenses associated with both materials and manufacturing (de Carvalho et. al, 2020). Hence, natural fibres offer a substitute material for reinforcement to diminish reliance on industrial fibres. This category of polymer is manufactured using a blend of two primary components, specifically polyol and isocyanate (Chaydarreh et al., 2017).

Polyurethane (PU) is one of the thermoplastic polymers, which has a cellular structure that promotes molecular expansion and is prepared via two methods: physical and chemical foaming. Researchers have investigated and explored mixing different fibres with polyurethane foam (PF), such as glass and carbon fibres (Kim et al., 2021; Lee et al., 2019; Zhang et al., 2020). Prior investigations have examined various natural fibres, including coconut, coir and kenaf, for their mechanical performance with PF composites (Azmi et al., 2012;Yuvaraj et al., 2018). However, research on BF/PF composites is still limited and less focused on use in building insulation panels. This study aims to evaluate the mechanical, thermal, and physical performance of the composites, focusing on the effects of fibre sizes on flexural and compressive strength, thermal stability, moisture uptake, water absorption, and dimensional swelling.

METHODOLOGY

Preparation of Materials

PF was obtained from Taat Bestari Sdn. Bhd., Selangor, Malaysia. BF used in this study were sourced from Negeri Sembilan, Malaysia.

Preparation of Fibre Composites

The BF were air-dried under sunlight to lower their moisture content, then ground using a crushing machine and separated into four particle size ranges, such as <150 µm, 150-300 µm, 300-425 µm, and 425-600 µm using a sieve shaker machine (Bryśkiewicz et al., 2016; Gu & Sain, 2012). Figure 1 shows the stages of fibre preparation, from raw bamboo to crushed and sieved fibres. To fabricate the BF/PF composites, polyol and isocyanate

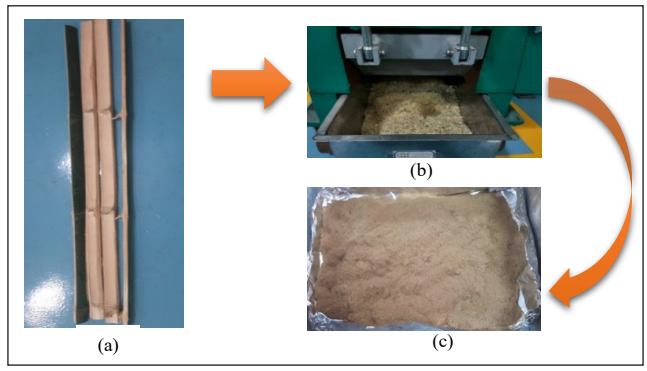


Figure 1. (a) Raw bamboo, (b) crushed BF, and (c) sieved BF

were blended at a weight proportion of 1:1.1 using a mechanical stirring machine (Sinar et al., 2015a). The prepared mixture was cast into moulds, left to solidify for an hour, and subsequently post-cured in an oven at 80 °C for two hours. The composites consisted of 5 wt.% BF and 95 wt. % PF as the matrix with different fibre sizes, as shown in Table 1.

Table 1
Composite formulation of various bamboo fibre sizes

Polyurethane Foam (wt. %)	Bamboo weight (wt. %)	Bamboo Size (µm)	Designation
100	-	-	Neat
95	5	<150	BFP1
		150-300	BFP2
		300-425	BFP3
		425-600	BFP4

Thermogravimetric Analysis (TGA)

The thermal degradation of the composites was examined with a TGA Q50 analyser (TA Instruments). Samples were heated from 30 °C to 600 °C at a rate of 10 °C/min under a nitrogen atmosphere to reduce oxidative effects. The obtained data were then analysed to determine the composites' thermal stability and decomposition behaviour.

Flexural Test

Flexural strength and modulus were evaluated following ASTM D790 (2002). The tests were carried out on a Universal Testing Machine (UTM) fitted with a 5 kN load cell and operated at a crosshead speed of 5 mm/min. For each composition, five samples with dimensions of 160 mm × 40 mm × 20 mm were examined.

Compression Test

The compression test was evaluated in accordance with ASTM D1621 (2000), using the same UTM and testing speed as the flexural test. Five specimens per fibre size category were tested under a constant rate of 5 mm/min with a 5 kN load cell.

Water absorption Test

The water absorption test was conducted following ASTM D570 (1998). Each sample was weighed before immersion (W_0) and then soaked in water for 24 hours. Subsequently, the samples were reweighed at 24-hour intervals over 7 days (W_1). Water uptake was calculated using the following Equation 1:

$$Water\ absoprtion, (\%) = \frac{W_1 - W_0}{W_0} \times 100 \quad [1]$$

Thickness Swelling Test

The thickness swelling test is used to determine the effect of water absorption on the composite sizes. This test requires 5 samples of each fibre size. The dimensions of the BF/PF composites were measured before and after immersion (T_0 and T_1). The following Equation 2 is calculated for thickness swelling:

$$Thickness\ swelling, (\%) = \frac{T_1 - T_0}{T_0} \times 100 \quad [2]$$

Moisture Content Test

Five samples are required for each fibre size. The samples were weighed to determine their initial mass (M0), then reweighed after 24 hours of oven drying to record the final mass (M_1). The percentage of moisture was determined using Equation 3:

$$Moisture\ content, (\%) = \frac{M_1 - M_0}{M_0} \times 100 \quad [3]$$

RESULT AND DISCUSSIONS

Thermogravimetric Analysis

The thermal stability of BF/PF composites was evaluated using thermogravimetric analysis (TGA). In general, natural fibres are easily degraded at high temperatures during the processing of composites. Therefore, understanding the degradation of natural fibres is crucial to ensure that the thermal stability and usability of the composites are maintained in the manufacturing process. Natural fibre composites degrade in a few stages. The first stage is moisture loss occurring at 30-100°C, followed by hemicellulose decomposition (temperature range: 200-300°C) and lignin decomposition (temperature range: 300-400°C), and the last stage is residual ash formation.(Azali et al., 2022; Huzaifah et al., 2017; Norizan et al., 2017).

Figure 2 presents the TGA curves of the BF/PF composites, whereas Figure 3 shows their corresponding DTG curves, from which the influence of fibre size on thermal stability can be analysed. The effect of fibre size on the thermal stability of the composites can be studied from this graph. The results showed that there is no significant difference in the thermal properties of the composites when fibre sizes vary. This is due to the similar composition

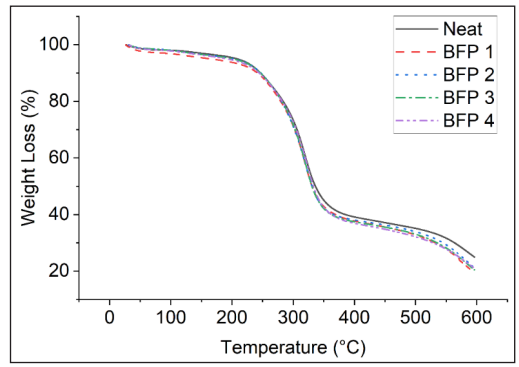


Figure 2. TGA curve of BF/PF composite.

of the fibre, where the individual thermal properties of the materials significantly influence the thermal properties.

All materials, as in the initial stage, were subjected to moisture evaporation at temperatures ranging from 30 to 100 °c (Rasheed et al., 2020; Syafiq et al., 2020). This demonstrates that BF retained more moisture when it started, and the peak degradation temperatures were lower. Decomposition of lignocellulose (hemicellulose) often occurs in the second phase. Observations show that lignocellulose decomposition occurred between 200°C and 300°C, where a small shoulder peak in the DTG curve was observed.

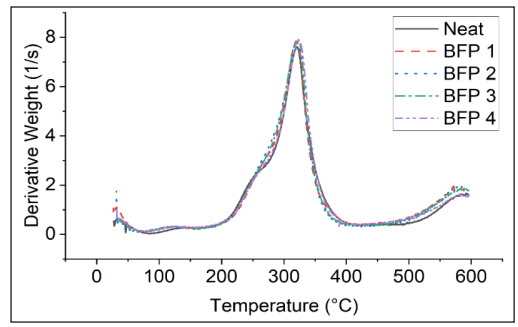


Figure 3. DTG curve of BF/PF composites

The researcher claims that the small shoulder peak before the major decomposition peak was attributed to the decomposition of hemicellulose (Chee et al., 2017; Ishak et al., 2012; Lei et al., 2023). During this stage, cellulose and lignin degradation took place and persisted into the second phase, which was associated with hemicellulose breakdown (Huzaifah et al., 2019; Jumaidin et al., 2017). The thermal degradation behaviour of neat samples and BFP composites with different fibre sizes is summarised in Table 2. In TGA analysis, T_{on} (initial decomposition temperature) marks the onset of material change when heat is applied and mass loss occurs, while Tmax (final decomposition temperature) represents the maximum point of mass change. Among all tested samples, including neat and composite materials, BFP3 exhibited the highest char residue (25.02%), while BFP1 demonstrated the least char residue (19.45%). The percentage of char residue detected in the composites is substantial compared to the neat material, indicating that each of the composite formulations exhibits great heat stability. Thus, the presence of greater lignin levels in BF plays a significant role in strengthening the composites' thermal stability.

Table 2
Thermal decomposition behaviour of BF/PF composites with various fibre sizes

Sample	First degradation		Second degradation		Char at 600°c (%)
	Ton	Tmax	Ton	Tmax	
Neat	235.02	282.49	276.8	320.91	20.55
BFP 1	223.84	278.42	297.83	317.25	19.45
BFP 2	236.96	270.54	303.42	320.61	21.21
BFP 3	229.66	299.07	247.18	321.02	25.02
BFP 4	250.37	296.45	303.48	322.22	21.09

Note. Ton: initial decomposition temperature; Tmax: final decomposition temperature

Flexural Test

The mechanical characteristics of a material are used to determine its composite strength. Figure 4 depicts the flexural strength and modulus of neat and BF/PF composites. Overall, both the modulus and flexural strength of BF/PF composites decrease when BF is added to PF. Flexural strength values are 0.261 MPa (neat), 0.119 MPa (BFP1), 0.151 MPa (BFP2), 0.192 MPa (BFP3), and 0.188 MPa (BFP4). 1.077 (neat) was the highest modulus value for neat PF and BP/PF, followed by 0.735 (BFP4), 0.723 (BFP3), 0.344 (BFP2), and 0.377 (BFP1). The results indicate that Neat Polyurethane foam exhibited higher flexural strength than all BFP composites. Although the incorporation of BF reduced the flexural strength, differences in performance were observed among BFP composites. This reduction is owing to the relatively poor adhesive surface between BF and PF, as well as the nonuniform distribution of fibres, as demonstrated by Huzaifah et al. (2019). Furthermore, the reduction can be attributed to the increase in the BF size and the presence of weak intrinsic strength of the BF compared to the matrix, which can reduce the performance of the mechanical performance. The behaviour of BF/PF composites has been found to increase up to the maximum load. BFP3 excelled over other composites in terms of interfacial adhesion and fibre dispersion, while BF filled voids in the PF. According to Sair et al. (2018), the weakness of composite characteristics can be attributed to low fibre dispersion, fibre breakage at the weakest location, and low distribution of force over fibres and matrix. Czlonka et al. (2018) verified the same finding, indicating that the weak corporate factor between fibres and matrix is similarly the main contributor to weak composites. As a result, the mechanical properties of the composite are crucial and should be optimised so that the composite material remains strong when an external force is applied.

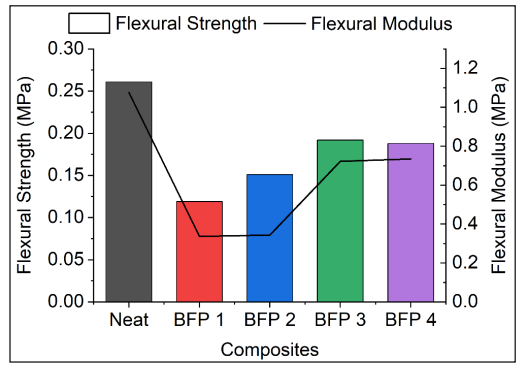


Figure 4. Effect of fibre size on flexural properties of BF/PF composites

Compression Test

Compression test results, presented in Figure 5, indicate that fibre size influenced the compressive properties of the composites. The compressive strength values were 0.203 MPa (BFP1), 0.214 MPa (BFP2), 0.232 MPa (BFP3), 0.221 MPa (BFP4), and 0.282 MPa for the neat sample. The increase in BF size causes a decrease in compaction strength, which is a similar phenomenon to the flexural strength behaviour. The increase in BF size causes a larger dispersion and non-uniform distribution of stress transfer during the compaction test.

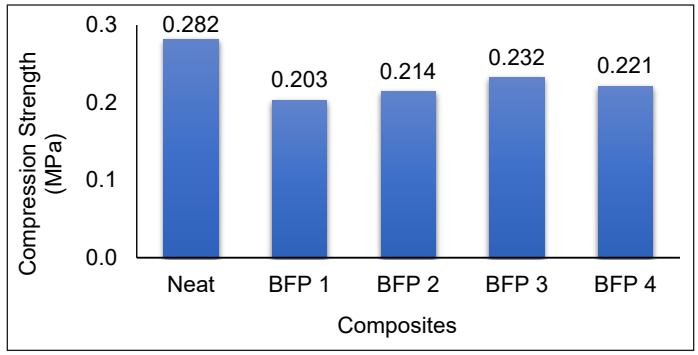


Figure 5. Compression strength of BF/PF composites with various fibre sizes

In addition, the adhesion between BF and PF is weak and leads to uneven force transfer, thereby causing a decrease in mechanical properties. As compared to the compressive strength of BF/PF composites, neat composites showed substantial properties. The results indicate that Neat Polyurethane foam exhibited higher compaction strength than all BFP composites. Although the incorporation of BF reduced the compaction strength, differences in performance were observed among BFP composites. The compression strength of the trend appears to be comparable to the flexural strength graph. Like the flexural behaviour, composites with larger fibre sizes tended to exhibit higher compressive strength. The low strength may be due to poor adhesion between the BF and PF, leading to non-uniform stress transfer, thereby causing a reduction factor on mechanical properties.

BFP3 excelled over other composites in terms of interfacial adhesion and fibre dispersion, while BF filled voids in the PF. This improvement may be due to better load transfer between the fibres and matrix at larger particle sizes. The low failure factor at the lowest values is also related to factors such as poor interaction, fibre agglomeration, and uneven mixing, which may have reduced the mechanical efficiency in some samples (Członka et al., 2018; Sinar et al., 2015b).

Water Absorption

Water absorption was recorded over 24 hours for seven consecutive days. Due to their hydrophilic nature, BF tend to absorb moisture compared to PF. Therefore, evaluating the water absorption properties of BP/PF composites is essential for understanding their performance under immersion conditions (Radzi et al., 2019). Figure 6 shows the water absorption assessment of BF/PF composites with different fibre sizes. After 7 days of immersion, BFP4 has been recorded as having the highest water absorption of 52.92%, followed by BFP3 (45.97%), BFP2 (43.99%), BFP1 (42.54%), and the neat composite (24.2%). From observation, the results of water absorption are greatly influenced by different fibre sizes in this test. Figure 6 shows that an increase in fibre size corresponds

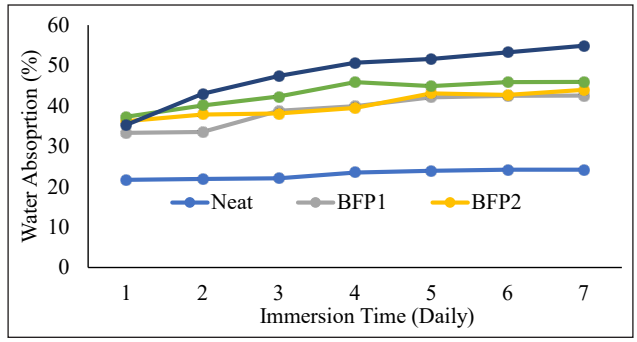


Figure 6. Water absorption properties of BF/PF composites with various fibre sizes

to higher water absorption. This behaviour is also influenced by the PF, which produces high and non-uniform voids. The characteristic facilitates an increased water absorption value, as moisture is easily retained within the foam and by the hydrophilic nature of BF (Chaydarreh et al., 2017; Radzi et al. 2019). Neat PF shows less water uptake due to its hydrophobic nature. As stated by Shalbafan et al. (2021) Polyurethane with a hydrophobic nature can limit water uptake since water primarily penetrates the foam's pores and interacts minimally with its closed-cell structure.

Thickness Swelling

Thickness swelling tests were carried out to determine the dimensional stability of BF/PF composites after seven days of immersion in water. Thickness swelling is closely related to the change in the dimensional stability of the composites. It occurs because the increased water absorption contributes to the expansion and dimensional stability of the composites. The low thickness swelling indicates that its dimensional stability is very good, and it's less susceptible to deformation due to moisture. For 7 days, measurements were recorded every 24 hours. From the fourth day, the BF/PF composites showed improved stability, with less variation in thickness compared to the initial measurements on the first day. The thickness swelling results for BF/PF composites with different fibre sizes are demonstrated in Figure 7. After 7 days, the highest thickness swelling was observed in BFP4 (4.7%), followed by BFP3 (4.29%), BFP2 (3.78%), BFP1 (3.66%), and neat PF (2.12%). Figure 7 shows similar trends with water absorption, which increases fibre size and increases the thickness swelling value. The behaviour of thickness swelling contributed to the same phenomenon as the factors of water absorption properties. Due to its hydrophilic nature, BF shows higher water absorption compared to the hydrophobic PF, which shows weaker interaction between the fibre and matrix. According to Caroline et al. (Rodrigues Pereira de Paula et al., 2022), who noted that weak interfacial bonding between the matrix and fibres increases void content, thereby facilitating high water uptake into the composite.

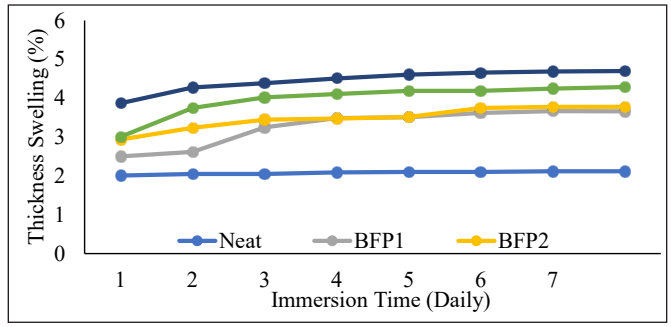


Figure 7. Thickness swelling properties of BF/PF composites with various fibre sizes

Moisture Content

The moisture content test was conducted to measure the amount of water absorbed by the material. These measurements were recorded after the BF/PF composites were immersed for 24 hours. From Figure 8, BFP 4 showed the highest moisture content of 4.79%, followed by BFP3 (4.79%), BFP2 (3.74%), and BFP1 (3.48%). The neat PF (2.26%) was recorded with less moisture content compared to the BF/PF composites.

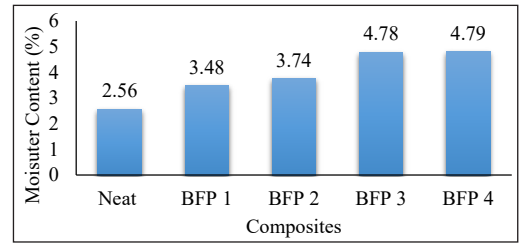


Figure 8. Moisture content properties of BF/PF composites with various fibre sizes

The results show a similar trend consistent with the water absorption and thickness swelling test, where larger fibre sizes contribute to increased moisture content. The graph trend can be attributed to the weak bonding between fibres and matrix, and the inherent differences in nature (hydrophilic and hydrophobic) of the physical properties of BFP and PF. The increase in moisture content percentage is further attributed to the hydrophilic nature of BFP, which easily absorbs moisture. The excess water either penetrates the PF matrix (hydrophobic) or is trapped and forms an additional layer within the PF foams. This behaviour is similar to findings by Chen et al. (2011) who explored the relationship between moisture content and interfacial strength in bamboo-reinforced vinyl ester composites. The researchers highlighted that the hydrophilic nature drives higher moisture absorption, and that residual water can interact with the hydrophobic polymer matrix, forming a surface layer on the composite materials Chen et. al, 2011; Noori et al., 2021).

CONCLUSION

The influence of the different fibre sizes on the mechanical, thermal, and physical properties of BF/PF composites has been thoroughly measured. Increasing the fibre size of BF in

the PF matrix positively influenced the mechanical properties (flexural and compression tests). Among the different fibre sizes, BF3 showed the highest mechanical strength, with flexural strength reaching 0.192 MPa and compression strength at 0.221 MPa. The factors causing the mechanical performance are due to excellent fibre dispersion, mixing uniformity and crosslinking between BF and PF. In comparison to neat PF and BFP/PF from thermal properties (TGA), BF/PF composites showed good enhanced thermal stability. However, physical properties show that increasing the fibre size caused higher water absorption, which in turn contributed to the highest thickness swelling and moisture content in all composite samples. Therefore, with its good performance, BF/PF composites with BFP3 show excellent performance for manufacturing applications such as thermal insulation and acoustic products. Furthermore, the BF/PF composites produced in this study have good potential as an alternative and environmentally friendly material. The continued development of natural fibre-based composites, such as those reinforced with BF, can give further support for combination materials in practical applications, including building parts (e.g., roofing and acoustic) and thermal insulation systems.

ACKNOWLEDGEMENT

The authors are grateful for the financial support given by the Geran Inisiatif Putra Muda (GP-IPM) (Vot-9774700). Additionally, we thank the Laboratory of Biocomposite Technology at Universiti Putra Malaysia for their invaluable collaboration.

LIST OF ABBREVIATIONS

BF	:	Bamboo fibre
PF	:	Polyurethane foam
MPa	:	Megapascal
TGA	:	Thermal Gravimetric Analysis
DTG	:	Derivative Thermogravimetry
ASTM	:	American Society for Testing and Materials
UTM	:	Universal Tensile Machine
w_0	:	Weighed before immersion
w_1	:	Weighed after immersion
T_o	:	measure the thickness before
T_1	:	measure the thickness after
m_0	:	measure the mass before immersion
m_1	:	measures the mass after
T_{on}	:	initial decomposition temperature
T_{max}	:	initial decomposition temperature

REFERENCES

Al-Oqla, F. M., & Sapuan, S. M. (2014). Natural fibre reinforced polymer composites in industrial applications: Feasibility of date palm fibres for sustainable automotive industry. *Journal of Cleaner Production, 66*, 347-354. https://doi.org/10.1016/j.jclepro.2013.10.050

Anwar, U. M. K., Hiziroglu, S., Hamdan, H., & Abd. Latif, M. (2011). Effect of outdoor exposure on some properties of resin-treated plybamboo. *Industrial Crops and Products, 33*(1), 140-145. https://doi.org/10.1016/j.indcrop.2010.09.014

ASTM International. (1998). *Standard test method for water absorption of plastics* (ASTM Standard D570-98).

ASTM International. (2000). *Standard test method for compressive properties of rigid cellular plastics* (ASTM Standard D1621).

ASTM International. (2002). *Standard test methods for flexural properties of unreinforced and reinforced plastics and electrical insulating materials* (ASTM Standard D790-02).

Azali, N. S. Z., Mustafa, N., Jumaidin, R., Abdul Kudus, S. I., Razali, N., Taha, M. M., Yusuf, Y., & Ali, M. R. (2022). Thermal properties of wood dust fibre and recycled polypropylene (r-WoPPc) for development of thermoplastic composites filaments of fused deposition modelling. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 96*(2), 42-50. https://semarakilmu.com.my/journals/index.php/fluid_mechanics_thermal_sciences/article/view/387

Azmi, M. A., Yusoff, M. F. C., Abdullah, H. Z., & Idris, M. I. (2012). Rigid polyurethane foam reinforced coconut coir fibre properties. *International Journal of Integrated Engineering, 4*(1), 11–15. https://penerbit.uthm.edu.my/ojs/index.php/ijie

Bahari, S. A., & Krause, A. (2016). Utilising Malaysian bamboo for use in thermoplastic composites. *Journal of Cleaner Production, 110*, 16-24. https://doi.org/10.1016/j.jclepro.2015.03.052

Beran, R., Zárybnická, L., Machová, D., Večeřa, M., & Kalenda, P. (2021). Wood adhesives from waste-free recycling depolymerisation of flexible polyurethane foams. *Journal of Cleaner Production, 305*, Article 127142. https://doi.org/10.1016/j.jclepro.2021.127142

Bryśkiewicz, A., Zieleniewska, M., Przyjemska, K., Chojnacki, P., & Ryszkowska, J. (2016). Modification of flexible polyurethane foams by the addition of natural origin fillers. *Polymer Degradation and Stability, 132*, 32-40. https://doi.org/10.1016/j.polymdegradstab.2016.01.002

Chaydarreh, K. C., Shalbafan, A., & Welling, J. (2017). Effect of ingredient ratios of rigid polyurethane foam on foam core panels properties. *Journal of Applied Polymer Science, 134*(17), Article 44722. https://doi.org/10.1002/app.44722

Chee, S. S., Jawaid, M., & Sultan, M. T. H. (2017). Thermal stability and dynamic mechanical properties of kenaf/bamboo fibre reinforced epoxy composites. *BioResources, 12*(4), 7118–7132. https://bioresources.cnr.ncsu.edu/resources/thermal-stability-and-dynamic-mechanical-properties-of-kenafbamboo-fibre-reinforced-epoxy-composites/

Chen, H., Miao, M., & Ding, X. (2011). Chemical treatments of bamboo to modify its moisture absorption and adhesion to vinyl ester resin in humid environment. *Journal of Composite Materials, 45*(14), 1533-1542. https://doi.org/10.1177/0021998310385023

Chen, M., Weng, Y., Semple, K., Zhang, S., Hu, Y., Jiang, X., Ma, J., Fei, B., & Dai, C. (2021). Sustainability and innovation of bamboo winding composite pipe products. *Renewable and Sustainable Energy Reviews, 144*, Article 110976. https://doi.org/10.1016/j.rser.2021.110976

Członka, S., Sienkiewicz, N., Strąkowska, A., & Strzelec, K. (2018). Keratin feathers as a filler for rigid polyurethane foams on the basis of soybean oil polyol. *Polymer Testing, 72*, 32-45. https://doi.org/10.1016/j.polymertesting.2018.10.032

Davoodi, M. M., Sapuan, S. M., Ahmad, D., Aidy, A., Khalina, A., & Jonoobi, M. (2012). Effect of polybutylene terephthalate (PBT) on impact property improvement of hybrid kenaf/glass epoxy composite. *Materials Letters, 67*(1), 5-7. https://doi.org/10.1016/j.matlet.2011.08.101

de Carvalho, P. S., Nora, M. D., & da Rosa, L. C. (2020). Development of an acoustic absorbing material based on sunflower residue following the cleaner production techniques. *Journal of Cleaner Production, 270*, Article 122478. https://doi.org/10.1016/j.jclepro.2020.122478

Gu, R., & Sain, M. M. (2013). Effects of wood fibre and microclay on the performance of soy based polyurethane foams. *Journal of Polymers and the Environment, 21*(1), 30-38.

Hassan, M. Z., Roslan, S. A., Sapuan, S. M., Rasid, Z. A., Mohd Nor, A. F., Md Daud, M. Y., Dolah, R., & Mohamed Yusoff, M. Z. (2020). Mercerisation optimisation of bamboo (*Bambusa vulgaris*) fibre-reinforced epoxy composite structures using a Box–Behnken design. *Polymers, 12*(6), Article 1367. https://doi.org/10.3390/polym12061367

Hong, C., Li, H., Lorenzo, R., Wu, G., Corbi, I., Corbi, O., Xiong, Z., Yang, D., & Zhang, H. (2019). Review on connections for original bamboo structures. *Journal of Renewable Materials, 7*(8), 713-730. https://www.techscience.com/jrm/v7n8/30574

Huzaifah, M. R. M., Sapuan, S. M., Leman, Z., & Ishak, M. R. (2017). Comparative study on chemical composition, physical, tensile, and thermal properties of sugar palm fibre (*Arenga pinnata*) obtained from different geographical locations. *BioResources, 12*(4), 9366-9382. https://bioresources.cnr.ncsu.edu/resources/comparative-study-on-chemical-composition-physical-tensile-and-thermal-properties-of-sugar-palm-fiber-arenga-pinnata-obtained-from-different-geographical-locations/

Huzaifah, M. R. M., Sapuan, S. M., Leman, Z., & Ishak, M. R. (2019). Effect of fibre loading on the physical, mechanical and thermal properties of sugar palm fibre reinforced vinyl ester composites. *Fibers and Polymers, 20*(5), 1077-1084. https://doi.org/10.1007/s12221-019-1040-0

Huzaifah, M. R. M., Sapuan, S. M., Leman, Z., Ishak, M. R., & Maleque, M. A. (2017). A review of sugar palm (*Arenga pinnata*): Application, fibre characterisation and composites. *Multidiscipline Modeling in Materials and Structures, 13*(4), 678-698. https://doi.org/10.1108/MMMS-12-2016-0064

Ibrahim, I. D., Jamiru, T., Sadiku, R. E., Kupolati, W. K., Agwuncha, S. C., & Ekundayo, G. (2015). The use of polypropylene in bamboo fibre composites and their mechanical properties: A review. *Journal of Reinforced Plastics and Composites, 34*(16), 1347-1356. https://doi.org/10.1177/0731684415591302

Ilyas, R. A., Sapuan, S. M., Kirubaanand, W., Zahfiq, Z. M., Atikah, M. S. N., Ibrahim, R., Radzi, A. M., Nadlene, R., Asyraf, M. R. M., Hazrol, M. D., Sherwani, S. F. K., Harussani, M. M., Tarique, J., Nazrin, A., & Syafiq, R. (2021). Roselle: Production, product development, and composites. In *Roselle* (pp. 1-23). Elsevier. https://doi.org/10.1016/B978-0-323-85213-5.00009-3

Ishak, M. R., Sapuan, S. M., Leman, Z., Rahman, M. Z. A., & Anwar, U. M. K. (2012). Characterisation of sugar palm (*Arenga pinnata*) fibres. *Journal of Thermal Analysis and Calorimetry, 109*(2), 981-989. https://doi.org/10.1007/s10973-011-1785-1

Ishak, M. R., Sapuan, S. M., Leman, Z., Rahman, M. Z. A., Anwar, U. M. K., & Siregar, J. P. (2013). Sugar palm (*Arenga pinnata*): Its fibres, polymers and composites. *Carbohydrate Polymers, 91*(2), 699-710. https://doi.org/10.1016/j.carbpol.2012.07.073

Jawaid, M., Chee, S. S., Asim, M., Saba, N., & Kali, S. (2022). Sustainable kenaf/bamboo fibres/clay hybrid nanocomposites: Properties, environmental aspects and applications. *Journal of Cleaner Production, 330*, Article 129938. https://doi.org/10.1016/j.jclepro.2021.129938

Jumaidin, R., Sapuan, S. M., Jawaid, M., Ishak, M. R., & Sahari, J. (2017). Effect of seaweed on mechanical, thermal, and biodegradation properties of thermoplastic sugar palm starch/agar composites. *International Journal of Biological Macromolecules, 99*, 265-273. https://doi.org/10.1016/j.ijbiomac.2017.02.092

Kadivar, M., Gauss, C., Ghavami, K., & Savastano, H. (2020). Densification of bamboo: State of the art. *Materials, 13*(19), Article 4346. https://doi.org/10.3390/ma13194346

Kairytė, A., Kirpluks, M., Ivdre, A., Cabulis, U., Vaitkus, S., & Pundiene, I. (2018). Cleaner production of polyurethane foam: Replacement of conventional raw materials, assessment of fire resistance and environmental impact. *Journal of Cleaner Production, 183*, 760-771. https://doi.org/10.1016/j.jclepro.2018.02.164

Kim, M. S., Kim, J. D., Kim, J. H., & Lee, J. M. (2021). Mechanical performance degradation of glass fibre-reinforced polyurethane foam subjected to repetitive low-energy impact. *International Journal of Mechanical Sciences, 194*, Article 106188. https://doi.org/10.1016/j.ijmecsci.2020.106188

Lee, D.-J., Kim, M.-K., Walsh, J., Jang, H.-K., Kim, H.-I., Oh, E.-Y., Nam, J., Kim, M., & Suhr, J. (2019). Experimental characterisation of temperature-dependent dynamic properties of glass fibre reinforced polyurethane foams. *Polymer Testing, 74*, 30-38. https://doi.org/10.1016/j.polymertesting.2018.12.013

Lei, J., Ye, X., Wang, H., & Zhao, D. (2023). Insights into pyrolysis kinetics, thermodynamics, and the reaction mechanism of wheat straw for its resource utilisation. *Sustainability, 15*(16), Article 12536. https://doi.org/10.3390/su151612536

Md Shah, A. U., Sultan, M. T. H., Jawaid, M., Cardona, F., & Abu Talib, A. R. (2016). A review on the tensile properties of bamboo fibre reinforced polymer composites. *BioResources, 11*(4), 10654-10676.

Nadlene, R., Sapuan, S. M., Jawaid, M., Ishak, M. R., & Yusriah, L. (2016). A review on roselle fibre and its composites. *Journal of Natural Fibers, 13*(1), 10-41. https://doi.org/10.1080/15440478.2014.984052

Noori, A., Lu, Y., Saffari, P., Liu, J., & Ke, J. (2021). The effect of mercerisation on thermal and mechanical properties of bamboo fibres as a biocomposite material: A review. *Construction and Building Materials, 279*, Article 122519. https://doi.org/10.1016/j.conbuildmat.2021.122519

Norizan, M. N., Abdan, K., Salit, M. S., & Mohamed, R. (2017). Physical, mechanical and thermal properties of sugar palm yarn fibre loading on reinforced unsaturated polyester composite. *Journal of Physical Science, 28*(3), 115-136. http://jps.usm.my/properties-sugar-palm-yarn-fibre/

Radzi, A. M., Sapuan, S. M., Jawaid, M., & Mansor, M. R. (2018). Mechanical and thermal performances of roselle fibre-reinforced thermoplastic polyurethane composites. *Polymer-Plastics Technology and Engineering, 57*(7), 601-608. https://doi.org/10.1080/03602559.2017.1332206

Radzi, A. M., Sapuan, S. M., Jawaid, M., & Mansor, M. R. (2019). Water absorption, thickness swelling and thermal properties of roselle/sugar palm fibre reinforced thermoplastic polyurethane hybrid composites. *Journal of Materials Research and Technology, 8*(5), 3988-3994. https://doi.org/10.1016/j.jmrt.2019.07.007

Radzi, A. M., Zaki, S. A., Hassan, M. Z., Ilyas, R. A., Md Daud, M. Y., Abd Aziz, S., & Jamaludin, K. R. (2022). Bamboo-fibre-reinforced thermoset and thermoplastic polymer composites: A review of properties, fabrication, and potential applications. *Polymers, 14*(7), Article 1387. https://doi.org/10.3390/polym14071387

Rasheed, M., Jawaid, M., Parveez, B., Zuriyati, A., & Khan, A. (2020). Morphological, chemical and thermal analysis of cellulose nanocrystals extracted from bamboo fibre. *International Journal of Biological Macromolecules, 160*, 183-191. https://doi.org/10.1016/j.ijbiomac.2020.05.170

Rodrigues Pereira de Paula, C., Trianoski, R., & Azevedo, E. (2022). Development and characterisation of sawdust and sisal fibre reinforced vegetable based polyurethane foam hybrid composites. *Journal of Natural Fibers, 19*(9), 3265-3274. https://doi.org/10.1080/15440478.2020.1841067

Roslan, S. A. H., Rasid, Z. A., & Hassan, M. Z. (2018). Bamboo reinforced polymer composite: A comprehensive review. *IOP Conference Series: Materials Science and Engineering, 344*(1), Article 012008. https://doi.org/10.1088/1757-899X/344/1/012008

Saheb, D. N., & Jog, J. P. (1999). Natural fibre polymer composites: A review. *Advances in Polymer Technology, 18*(4), 351-363. https://doi.org/10.1002/(SICI)1098-2329(199924)18:4%3C351::AID-ADV6%3E3.0.CO;2-X

Sair, S., Oushabi, A., Kammouni, A., Tanane, O., Abboud, Y., & El Bouari, A. (2018). Mechanical and thermal conductivity properties of hemp fibre reinforced polyurethane composites. *Case Studies in Construction Materials, 8*, 203-212. https://doi.org/10.1016/j.cscm.2018.02.001

Sanjay, M. R., Madhu, P., Jawaid, M., Senthamaraikannan, P., Senthil, S., & Pradeep, S. (2018). Characterisation and properties of natural fibre polymer composites: A comprehensive review. *Journal of Cleaner Production, 172*, 566-581. https://doi.org/10.1016/j.jclepro.2017.10.101

Sanmuham, V., Sultan, M. T. H., Radzi, A. M., Shamsuri, A. A., Md Shah, A. U., Safri, S. N. A., & Basri, A. A. (2021). Effect of silver nanopowder on mechanical, thermal and antimicrobial properties of kenaf/HDPE composites. *Polymers, 13*(22), Article 3928. https://doi.org/10.3390/polym13223928

Shalbafan, A., Chaydarreh, K. C., & Welling, J. (2021). Effect of blowing agent concentration on rigid polyurethane foam and the properties of foam-core particleboard. *Wood Material Science and Engineering, 16*(2), 85-93. https://doi.org/10.1080/17480272.2019.1626480

Siam, N. A., Uyup, M. K. A., Husain, H., Mohmod, A. L., & Awalludin, M. F. (2019). Anatomical, physical, and mechanical properties of thirteen Malaysian bamboo species. *BioResources, 14*(2), 3925-3943. https://bioresources.cnr.ncsu.edu/resources/anatomical-physical-and-mechanical-properties-of-thirteen-malaysian-bamboo-species/

Sinar, A. A., Firuz, Z., Nur Azni, M. A., Ahmad Zaidi, N. H., Hazizan, M. A., & Sahrim, H. A. (2015a). Compression test and energy absorption of polyurethane/multi-walled carbon nanotubes foam composites. *Materials Science Forum, 819*, 246-250. https://doi.org/10.4028/www.scientific.net/MSF.819.246

Sinar, A. A., Firuz, Z., Nur Azni, M. A., Hazizan, M. A., & Sahrim, H. A. (2015b). Flexural performance of polyurethane/multi-walled carbon nanotubes foam composites. *Applied Mechanics and Materials, 754–755*, 8-12. https://doi.org/10.4028/www.scientific.net/AMM.754-755.8

Suhot, M. A., Hassan, M. Z., Abdul Aziz, S., & Md Daud, M. Y. (2021). Recent progress of rice husk reinforced polymer composites: A review. *Polymers, 13*(15), Article 2391. https://doi.org/10.3390/polym13152391

Supian, A. B. M., Sapuan, S. M., Zuhri, M. Y. M., Zainudin, E. S., & Ya, H. H. (2018). Hybrid reinforced thermoset polymer composite in energy absorption tube application: A review. *Defence Technology, 14*(4), 291-305. https://doi.org/10.1016/j.dt.2018.04.004

Syafiq, M. O., Razali, Sapuan, S. M., & Zuhri, M. R. M. (2020). Effect of cinnamon essential oil on morphological, flammability and thermal properties of nanocellulose fibre-reinforced starch biopolymer composites. *Nanotechnology Reviews, 9*(1), 1147-1159. https://doi.org/10.1515/ntrev-2020-0087

Syafiq, R., Sapuan, S. M., & Zuhri, M. R. M. (2021). Antimicrobial activity, physical, mechanical and barrier properties of sugar palm based nanocellulose/starch biocomposite films incorporated with cinnamon essential oil. *Journal of Materials Research and Technology, 11*, 144-157. https://doi.org/10.1016/j.jmrt.2020.12.091

Syafiq, R., Sapuan, S. M., Zuhri, M. Y. M., & Khalina, A. (2020). Antimicrobial activities of starch-based biopolymers and biocomposites incorporated with plant essential oils: A review. *Polymers, 12*(10), Article 2403. https://doi.org/10.3390/polym12102403

Yuvaraj, L., Jeyanthi, S., & Lenin Babu, M. C. (2018). Sound absorption analysis of castor oil based polyurethane foam with natural fibre. *Materials Today: Proceedings, 5*(11), 23534-23540. https://doi.org/10.1016/j.matpr.2018.10.141

Zhang, Y., Yan, X., Huang, W., & Lin, J. (2020). Experimental investigations on mechanical behaviour of the carbon fibre tube reinforced polyurethane foam. *Thin-Walled Structures, 155*, Article 106899. https://doi.org/10.1016/j.tws.2020.106899

Zia, F., Zia, K. M., Zuber, M., Kamal, S., & Aslam, N. (2015). Starch based polyurethanes: A critical review updating recent literature. *Carbohydrate Polymers, 134*, 784-798. https://doi.org/10.1016/j.carbpol.2015.08.034

Pertanika J. Sci. & Technol. 34 (4): 2403 - 2420 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Technological Approach for Biodiversity Conservation: Augmented Reality Based on Systematic Review

Nurfitri Ezatie Yusoff[1] and Arney Sapaat[1,2]*

[1]Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, KM 1, Jalan Panchor, 84600 Panchor, Johor, Malaysia
[2]International Institute of Biodiversity Technology and Sustainability, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, KM 1, Jalan Panchor, 84600 Panchor, Johor, Malaysia

ABSTRACT

Biodiversity and technology intersect in numerous ways, particularly in the realms of conservation and environmental management. Augmented reality (AR) is currently used in biodiversity conservation, improving awareness and understanding. Despite growing interest in AR applications, a comprehensive synthesis of evidence regarding its effectiveness and implementation in conservation contexts remains limited. We conducted a systematic literature review following PRISMA 2020 guidelines to evaluate AR applications in biodiversity conservation. The analysed literature was collected through a focused search in scientific databases (Scispace, Scopus, ACM, and Google Scholar). Data extraction focused on AR technology types, conservation application areas, target user groups, study designs, key findings, and biodiversity focus areas. From 723 initial records, 34 articles were found to meet the selection criteria after deduplication and two-stage screening. The results show that augmented reality (AR) interventions can significantly improve environmental knowledge, attitudes, and engagement. However, there is still limited evidence on their effectiveness in promoting behavioural change and supporting professional conservation practices. Future research should prioritise rigorous experimental designs, standardised evaluation metrics, and integration of AR tools into operational conservation workflows to fully realise the technology's conservation potential.

Keywords: Augmented, biodiversity, conservation, technology

ARTICLE INFO

Article history:
Received: 09 June 2025
Accepted: 30 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.22

E-mail addresses:
nurfitriezatie@gmail.com (Nurfitri Ezatie)
arney@uthm.edu.my (Arney Sapaat)
* Corresponding author

INTRODUCTION

Biodiversity means different types of ecosystems, which include a range and variability of animals, plants and

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

microorganisms contained in them. It is not a fixed entity but always changes; it is enhanced by genetic variation included in the organism and can be reduced through biodiversity loss, extinction of species and habitat degradation (Hooper et al., 2005). Biodiversity also stabilises the environment through ecological functionality, and many species form functional diversity. The destruction of the environment and the extinction of biodiversity have become a very worrying issue. This threat to biodiversity is expected to continue and increase in the future if no serious action is taken by all stakeholders. Degradation and extinction of natural resources are affecting ecosystem services globally (Adla et al., 2022).

Technology and biodiversity intersect in numerous ways, particularly in the realms of conservation and environmental management. Today's technology plays an important role in all aspects, such as education (Haleem et al., 2022), video games (Karpova et al., 2021), tourism (Pratisto et al., 2022), biodiversity conservation (Pimm et al., 2015) and many more. This technology allows us to immerse ourselves in a man-made world that can be this world entirely or the result of human re-creation. It is also called an alternative world filled with computer-generated images and human responses (Hamad & Jia, 2022). The work presented in this article explores a face of the current state of technology and proposes new systems designed to enhance the appreciation of conservation by leveraging advanced interaction technologies for biodiversity. According to Al-Ansi et al. (2023), the use of Information Technology (IT) in the process of information delivery is experiencing rapid development. Currently, the latest technology used in the delivery of information is Augmented Reality (AR) technology, in which users can visualise historical objects in the form of 3D. AR has the advantage of being interactive and real-time, so that AR is widely implemented in various fields. AR provides a combined interaction between real and virtual environments that is something that cannot be done in other virtual environments. The uniqueness of AR is the ability to generate virtual objects or information in the real-world environment.

Augmented reality (AR) technology has emerged as a promising tool for building digital infrastructure for natural history collections, as it can provide an immersive and interactive environment for researchers and the public to access and explore specimens. Augmented Reality (AR) is one part of the Virtual Environment (VE), or what is commonly known as Virtual Reality (VR) (Hantono et al., 2018). AR gives users an idea of combining the real world with the virtual world seen from the same place. Augmented reality (AR) has three characteristics, namely interactive (increasing user interaction and perception of the real world), real-time, and 3-dimensional. An AR illustration diagram can be seen in Figure 1, where AR is a combination of the real world and the virtual world. Augmented Reality emphasises the real world, with virtual elements added to enhance and complete the real scene (Milgram & Kishino, 1994).

There are two augmented reality (AR) methods currently being developed, namely Marker-Based Tracking and Markerless AR. Marker-based tracking is AR that uses markers or two-dimensional object markers, that have a pattern that will be read by the computer via a webcam or camera connected to the computer. Figure 2 shows that markers are usually square black-and-white illustrations with thick black borders and a white background (Le et al., 2017).

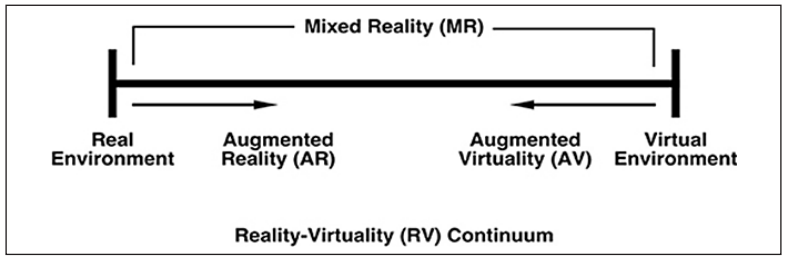


Figure 1. The Mixed Reality continuum (Milgram & Kishino, 1994)



(a) Template Marker (b) Bar-code Marker (c) Circular Marker

Figure 2. Marker-Based Tracking used for AR

According to Ingensand et al. (2018), Nature augmented reality (AR) project, which utilises a marker-based AR approach. This shows the nature elements and scans markers to reveal 3D models of different species, enhancing their understanding of biodiversity through interactive learning experiences. Additionally, BioSentiers, which is a mobile AR application, is designed to encourage children to learn about biodiversity during outdoor activities. AR technology can enrich the physical world with useful information and provide new and intuitive ways of interaction for users, where the information can be stored in the AR and reused wherever needed (Kumar et al., 2022). This application aims to create a connection between students and nature, motivating them to explore various species in their environment. These applications illustrate how AR can be effectively integrated into educational settings to foster engagement and enhance learning outcomes (Li et al., 2017).

Augmented reality (AR) technology plays a significant role in the advancement of biodiversity research. However, poor-quality data negatively impacts the accuracy of Augmented Reality (AR), resulting in inaccurate calculations, incorrect object recognition and unreliable virtual content, fundamentally compromising the ability of AR systems to create consistent and functional mixed reality environments (Belhaoua et al., 2014). From a technical point of view, it was the methodological uniformity of the examined studies, with many of them based on non-empirical procedures. This points to the necessity of more unconventional research methodologies such as qualitative or mixed-methods approaches to achieve a more comprehensive understanding of sustainability (Cosio et al., 2023; Pavlik & Bridges, 2013). Despite increasing interest and development in applications, the evidence for AR's effectiveness in conservation is still scattered across different fields, publication platforms, and locations. There has not been a systematic review that combines the various AR applications, technological methods, and results in biodiversity conservation. It is important to understand how AR is currently being used, identify gaps in evidence, and assess how it compares to traditional methods. This article will help direct future research, technology development, and conservation practices. Therefore, there is a need to establish digital infrastructure based on augmented reality technology to promote the conservation and scientific research of biodiversity.

To fill this gap, we systematically reviewed a total of 34 papers to discuss existing methods and areas of study, where we also pointed out current focuses and potential research directions for AR-based technology. We investigate the use of AR in environmental sciences and nature, expanding on previous studies by examining the range of contexts, such as education or conservation and identifying the impacts and effectiveness of AR applications on biodiversity monitoring, data collection, and conservation management tasks. Even though the scope may be expanded further, we restrict the review to gain a clearer understanding of the progress in the right field. This paper specifically aims to bring analyses into areas of application, technology and components in the current literature. Additionally, we present their applications in studies and reported results. We compile the findings to create a draft for a future research agenda to address gaps and explore new research prospects. This suggests a pressing need for clearer definitions and standardised guidelines in future research to ensure consistency and clarity. Specifically, we answer the following research questions in this article:

RQ1: What are the primary applications and technological approaches of AR used in biodiversity conservation?

RQ2: How does the use of AR-based tools compare to traditional methods in enhancing public engagement, awareness, and environmental education for biodiversity?

RQ3: What are the impacts and effectiveness of AR applications on biodiversity monitoring, data collection, and conservation management tasks?

To achieve its objective, recent papers on developed projects should be collected and reviewed based on a set of eligibility criteria. This paper specifically focuses on recent works addressing the mentioned issues, reviewing articles published between January 1, 2015, and December 31, 2025. Articles from 2026 are not included because the search and analysis were conducted in 2026, and including papers from the current year would result in incomplete coverage.

MATERIAL AND METHOD

This study uses the systematic literature review (SLR) method. This method is used to find and identify research information in a systematic and well-founded manner. According to Xiao and Watson (2019), systematic literature review (SLR) is a study that aims to search and analyse in a comprehensive, structured, no doubt and can be repeated in the next process. Therefore, the researcher uses this method as the main methodology of the study.

Protocol And Registration

In the collection of previous studies that use systematic literature reviews highlighting methodology, there are some guidelines or protocols that can be used to collect and ensure that the selection of previous studies is the most relevant. Literature selection was conducted according to the systematic review guidelines presented by Preferred Reporting Items for Systematic Reviews (PRISMA). Preferred Reporting Items for Systematic Reviews and Meta-Analyses is a protocol that can be used as a guide to present systematic literature highlights (Moher et al., 2016). The quality and quantity control of article selection is guided by the criteria in PRISMA. PRISMA was chosen as a guide because this guideline is able to control the quality and quantity of article selection in large databases such as Scopus and Google Scholar. These criteria consist of identification, acceptance and rejection of articles, qualification and evaluation of quality articles. All articles need to be evaluated so that the objective of highlighting this systematic literature can be achieved.

Search Strategy

The literature search strategy is determined starting with the guidance of the research question. Since this study focuses on the elements of environmental science, the keywords used are based on the theme of the research title. The literature search was conducted through the area of SciSpace, Scopus, Google Scholar and ACM Digital Library. The use of these databases is intended to facilitate literature search in a more organised and systematic way, as shown in Table 1.

Table 1
Search query

Database	Query
Scispace	"What are the primary applications and technological approaches of Augmented Reality (AR) used in biodiversity conservation?" "What are the impacts and effectiveness of AR applications on biodiversity monitoring, data collection, and conservation management?"
Scopus	"Augmented Reality" AND ("environmental education" OR "public engagement" OR "awareness") AND "biodiversity"
Google scholar	"Augmented Reality" AND ("biodiversity conservation" OR "wildlife conservation" OR "technology" OR "nature conservation"
ACM Digital Library	("Document Title": technology) AND ("Document Title": biodiversity OR "Document Title": augmented

Through these search queries and filtering, 300 works were obtained from SciSpace, 200 from Scopus, 157 from Google Scholar and 66 from ACM Digital Library. Ultimately, there are 223 titles from which the first-order duplications should be eliminated. 500 records are left once the duplicate entries are removed. During abstract screening, studies were excluded if they did not focus on augmented reality (AR) technology (215 articles), were unrelated to biodiversity conservation (128 articles), focused exclusively on technical AR development without conservation applications (85 articles), or examined virtual reality (VR) rather than AR (30 articles). Language filters were applied during the search phase. Publication type filters were applied during the initial search to minimise the articles findings such as journal articles, conference proceedings, theses, and technical reports.

Study Selection and Eligibility Criteria

According to Xiao and Watson (2019), a survey study in the form of criticism that involves the comparison of a body of literature requires the establishment of a solid set of criteria. Therefore, this study has set a few criteria that are included in Table 2. Among the criteria that have been set is the time period of the literature review, which must be from 2015. In addition, Figure 3 shows the screening process of the determination of the selection of literature, which must consist of journal articles, proceedings, books or theses. 458 works were excluded from the review because their primary focus was not on the use of AR technologies within biodiversity or environmental contexts. Among the excluded works, eight specifically focused on Virtual reality (VR) technology approaches to environmental and non-AR immersive technologies.

Table 2
Eligibility criteria for study selection

Inclusion Criteria	Exclusion Criteria
Articles published from 2015 to 2024	Papers written in a language other than English
Full-text article	Exclude papers that focus exclusively on the technical development of AR algorithms and hardware without application in biodiversity.
Empirical (qualitative, quantitative, mixed-methods, design science) and conceptual articles related to the use of AR	Exclude papers focusing on broad environmental science themes

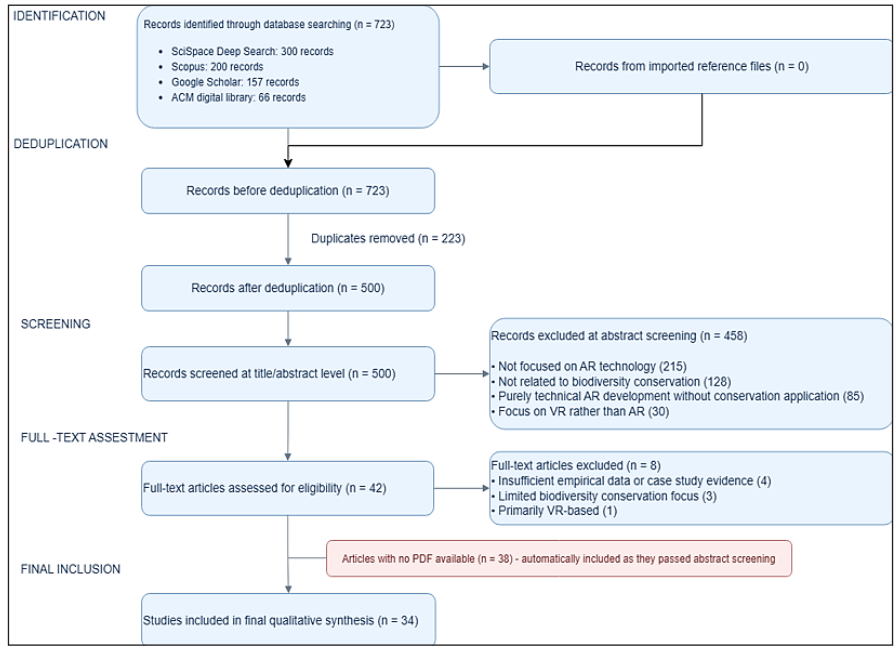


Figure 3. Screening process

Screening Process

A two-stage screening process was implemented following PRISMA 2020 guidelines. All unique records identified through database searches underwent title and abstract screening.

42 Records passing abstract screening underwent full-text assessment when documents were available. Unfortunately, eight articles were removed due to insufficient empirical data, limited biodiversity studied, and a primary focus on VR-based interventions. Given the above considerations, 34 studies met all eligibility criteria and were included in the systematic review for data extraction and synthesis. While the final number of works

may seem small, it is relatively large compared to the number of works studied in other reviews (Cosio et al., 2023), especially considering the specific time span that this review covers.

Data Extraction

According to Afifi et al. (2023), data extraction is carried out to facilitate and summarise the comparison of literature materials that meet the criteria that have been set. The instrument data of this study has been filtered according to the categories that have been compiled through the data extraction process. Among the categories set in this study are background, study issues or gaps, AR technology type, study design and method, biodiversity focus and study findings. After that, these categories were synthesised to give a clearer picture in producing a meaningful comparison, as shown in Table 3. Abstracts of each article were reviewed and evaluated based on the research questions that were formed at the beginning of the study.

Table 3
Data extraction for Augmented Reality (AR) in biodiversity conservation

Variable	Description	Indicators
AR Technology Type	Classification of augmented reality technologies and platforms used in the study.	Mobile AR (smartphone/tablet applications), Wearable AR (head-mounted displays), Location-based AR (GPS-enabled/ geospatial systems), Marker-based AR (image recognition, QR codes), Markerless AR (SLAM, environmental tracking), Game-based AR (gamification elements), Web-based AR.
Conservation Application Area	Application of AR in biodiversity conservation.	Environmental education, Public awareness and engagement, Species identification, Biodiversity monitoring, Data collection, Habitat mapping, Conservation management, Ecotourism, Citizen science, Research support.
Target User Group	Users of the AR application.	Students (primary, secondary, tertiary), Teachers/Educators, General public, Tourists/Visitors, Researchers/Scientists, Conservation professionals, Children, Policymakers
Study Design and Methods	Research design, methodology, and evaluation approach.	Experimental studies, Case studies, Design-based research, Surveys, Interviews, Focus groups, Usability testing, Observational studies, Mixed-methods approaches; sample size, instruments, and evaluation methods where reported.
Key Findings and Outcomes	Main results regarding the effectiveness and impact of AR interventions.	Learning outcomes (knowledge gain, conceptual understanding), Behavioral outcomes (engagement, conservation actions, donation behaviour), Usability outcomes (user satisfaction, ease of use, technical performance), Comparative outcomes (AR vs. traditional methods).
Biodiversity Focus	Biodiversity elements, taxa, or ecosystems addressed by the study.	Animals, Plants, Forest ecosystems, Wetlands, Marine ecosystems, Urban ecosystems, Endangered or threatened species, Local/regional biodiversity, General biodiversity concepts.

Data extraction was performed for all 34 studies that fulfilled the inclusion criteria. The six variables listed above were systematically coded to facilitate quantitative summaries, cross-study comparisons, and qualitative synthesis of AR applications in biodiversity conservation.

RESULT AND DISCUSSION

Table 4 shows the AR Technology Types in Biodiversity Conservation Studies articles, with proper citations provided in the References section. Articles that did not meet the inclusion criteria were excluded. The analysis of the articles covers AR technology types, conservation application areas, target user groups, study designs, and biodiversity focus areas. The research question is coded as (RQ). The research questions in this article are, RQ1: What are the primary applications and technological approaches of AR used in biodiversity conservation? RQ2: How does the use of AR-based tools compare to traditional methods in enhancing public engagement, awareness, and environmental education for biodiversity? RQ3: What are the impacts and effectiveness of AR applications on biodiversity monitoring, data collection, and conservation management tasks?

Table 4
Study characteristic

Category	Subcategory	Number of Articles
AR Technology Type	Mobile AR	28
	Game-Based AR	18
	Location-Based AR	8
	Marker-Based AR	7
	Markerless AR	7
	Web-Based AR	3
	Wearable AR	1
Application Area	Environmental Education	30
	Public Awareness & Engagement	27
	Species Identification	11
	Ecotourism	7
	Citizen Science	5
Target User Group	Students	20
	General Public	14
	Tourists/Visitors	10
	Researchers/Scientists	6
Study Design	Experimental Studies	12
	Case Studies	11
	Design-Based Research	8

Table 4 (*continued*)

Category	Subcategory	Number of Articles
Biodiversity Focus	Animals/Fauna	19
	Plants/Flora	11
	General Biodiversity	10

Mobile AR applications dominated the technology landscape, with 28 studies involving smartphone- or tablet-based AR platforms. Game-based AR approaches were implemented in 18 studies, often integrated with mobile platforms to enhance engagement through gamification elements. Location-based AR, utilising GPS and geospatial data, was featured in 8 studies. Marker-based AR (image recognition, QR codes) and markerless AR (SLAM, environmental tracking) were each reported in 7 studies. Web-based AR appeared in 3 studies, while wearable AR (AR glasses, head-mounted displays) was minimally represented with only 1 study. The predominance of mobile AR reflects the widespread availability of smartphones and tablets, lower barriers to user adoption, and advanced development infrastructure for mobile platforms. Game-based approaches leveraged AR's interactive capabilities to create engaging conservation experiences, particularly for educational contexts. The limited representation of wearable AR suggests that head-mounted displays remain less accessible for conservation applications, though this may change as hardware costs decrease and form factors improve.

Environmental education was the most common application area (Al-Ansi et al., 2023), addressed in 30 studies. Public awareness and engagement initiatives were featured in 27 studies, often overlapping with educational objectives. Species identification applications appeared in 11 studies, supporting both educational and monitoring functions. Ecotourism contexts were explored in 7 studies, leveraging AR to enhance visitor experiences in protected areas and natural sites. Citizen science applications were reported in 5 studies, engaging public participants in biodiversity data collection. Biodiversity monitoring and data collection for professional conservation tasks were each addressed in 4 studies, representing a smaller but important subset focused on operational conservation workflows. Habitat mapping, conservation management, and research support applications were less frequently represented, each appearing in 2-3 studies. The application of studies in education and public engagement domains reflects AR's perceived strengths in creating immersive, interactive learning experiences. The relative scarcity of monitoring and management applications suggests that AR integration into professional conservation workflows remains an emerging area requiring further development and validation.

For target user group characteristics, students and general public audiences show a predominance of educational and engagement applications, which featured in 20 studies and comprise primary, secondary, and higher education levels, while the general public was addressed in 14 studies, often through museum exhibits, public installations, or mobile applications. Tourists and visitors were targeted in 10 studies, particularly in ecotourism and protected area contexts. Researchers and scientists were the intended users in 6 studies, focusing on AR tools for field research and data collection.

Experimental designs were employed in 12 studies, including randomised controlled trials and quasi-experimental studies comparing AR interventions to control conditions or alternative approaches. Case studies were conducted in 11 studies, providing in-depth descriptions of AR application development and deployment in specific conservation contexts. Design-based research approaches were utilised in 8 studies, involving iterative cycles of design, implementation, and evaluation. Sample sizes varied widely, from small-scale pilot studies with fewer than 30 participants to larger evaluations with over 200 participants. Many studies employed convenience sampling with student populations, limiting generalisability to broader audiences. The variety of methodological approaches reflects the multidisciplinary nature of AR conservation research, drawing from education research, human-computer interaction, conservation, and environmental studies. However, the limited number of experimental studies with adequate sample sizes and control conditions constrains causal inference regarding AR effectiveness.

According to Table 4, Animals and fauna were the most common biodiversity focus, featured in 19 studies, including mammals (apes, elephants, endangered species), birds, marine life (coral reefs, marine species), amphibians, and invertebrates. Plants and flora were addressed in 11 studies, including endangered plants, trees, endemic species, and herbarium collections. General biodiversity concepts without specific taxonomic focus appeared in 10 studies, emphasising ecosystem diversity, species richness, and conservation principles. The taxonomic and ecological diversity of biodiversity focus areas demonstrates AR's versatility across conservation contexts, which indicates broader applications.

This systematic review synthesised evidence from 34 studies on Augmented Reality applications in biodiversity conservation, revealing a rapidly evolving field with significant potential and notable research gaps. The most significant findings of this systematic review are shown in Table 5. Collectively, the evidence demonstrates strong potential for AR to support environmental education and public engagement, while simultaneously revealing critical gaps in behavioural impact and operational conservation applications.

Table 5
Key findings on the effectiveness of AR for biodiversity conservation

Outcome Domain	Overall Evidence	Representative Findings
Knowledge and Learning	Strong	AR improved biodiversity knowledge, species recognition, and conceptual understanding.
Environmental Attitudes	Strong	AR increased conservation awareness and positive attitudes toward biodiversity protection.
Engagement and Motivation	Strong	Interactive and gamified AR experiences enhanced participation and user engagement.
Nature Connectedness	Moderate	Outdoor and location-based AR applications strengthened users' connection with nature.
Behavioural Change	Limited	Few studies measured actual conservation behaviours; evidence remains inconclusive.
Monitoring and Conservation Management	Limited	AR showed potential for field data collection and visualisation but lacked rigorous evaluation.
Usability and User Experience	Strong	Most studies reported high usability, satisfaction, and positive user experiences.

Across studies, AR consistently demonstrates strong effectiveness in improving biodiversity knowledge, species recognition, and conceptual understanding. These cognitive gains are closely accompanied by positive shifts in environmental attitudes and conservation awareness. According to Chen et al. (2023), Mercier et al. (2025), and Nahdiya et al. (2024), participants demonstrated enhanced species identification skills, improved understanding of ecological concepts, and better retention of conservation information compared to baseline or control conditions. In parallel, AR-based interventions, particularly those incorporating game mechanics and interactive storytelling, show strong evidence of increased engagement and motivation. These outcomes suggest that AR is particularly effective as an educational and participatory tool, where immersive and interactive features enhance attention, learning retention, and user involvement (Poluakan et al., 2024; Tran & Khoa, 2025). Emotional engagement was particularly pronounced in game-based AR applications that incorporated narrative and character-driven content (Harrington et al., 2019; Ibrahim et al., 2022).

AR applications consistently demonstrated high levels of user engagement, as measured by time on task, interaction frequency, task completion rates, and self-reported enjoyment. Gamification elements, interactive challenges, and immersive visualisations sustained motivation and encouraged exploration (Nahdiya et al., 2024). Engagement was identified as a critical mediator of learning and attitudinal outcomes(Chan et al., 2019). Most studies report high levels of user satisfaction and positive user experience, indicating that current mobile-based AR systems are both accessible and well-received by diverse

user groups (Abidin et al., 2023; Safitri et al., 2025; Wu et al., 2024). The dominance of smartphone and tablet platforms contributes significantly to this accessibility, enabling scalable deployment without requiring specialised hardware.

Several studies measured augmented reality (AR) impact on nature connectedness wheres the subjective sense of relationship with the natural world. Results indicated that AR experiences, particularly location-based outdoor games, enhanced nature connectedness among participants (Pimentel, 2021). Play-related enjoyment was identified as a key factor mediating this effect; AR can enhance nature connectedness, particularly in outdoor and location-based implementations (Pimentel, 2022). By anchoring digital information to real-world environments, AR facilitates situated learning that strengthens users' psychological and perceptual relationship with nature. This distinguishes AR from fully virtual technologies, as it preserves physical context while augmenting ecological interpretation (Lawton et al., 2020).

Evidence for behavioural change was mixed. Several studies reported increased behavioural intentions to engage in conservation actions, such as reducing plastic use, supporting conservation organisations, or participating in citizen science (Abidin et al., 2023; Chen et al., 2023; Pimentel, 2021). However, actual behavioural outcomes were rarely measured, and one rigorous experimental study found no evidence that augmented reality AR gaming encouraged donation behaviours, highlighting an intention-action gap (Sandbrook et al., 2015)

de Oliveira et al. (2022), Pimentel (2021), and Imran and Shahid (2025) reported that few studies quantitatively evaluated augmented reality (AR) impact on monitoring accuracy, data quality, or task efficiency compared to traditional methods. The available evidence was primarily descriptive, focusing on feasibility and user acceptance rather than performance metrics. The limited evidence base for AR in professional conservation workflows represents a significant gap (Fol et al., 2025; Imran & Shahid, 2025). Most studies focused on proof-of-concept demonstrations or small-scale pilots rather than operational deployments with rigorous evaluation (Fol et al., 2025; Xu et al., 2025). Future research should prioritise comparative effectiveness studies, standardised performance metrics, and integration with established monitoring protocols.

Usability assessments generally indicated positive user experiences, with high satisfaction ratings and ease-of-use scores. The EduPARK mobile augmented reality (AR) game achieved excellent usability (SUS score 85.8) among younger students and good usability (SUS score 70.9) among older students (Pombo & Morais Marques, 2018). However, some studies noted technical challenges, including tracking instability, device compatibility issues, battery drain, and learning curves for first-time AR users (de Oliveira et al., 2022; Syiem et al., 2024; Voutos et al., 2023). User experience is

mediated by factors such as digital literacy, prior technology experience, and contextual appropriateness. Designing AR experiences that balance screen engagement with direct nature observation is an important consideration for outdoor conservation applications (Guzman et al., 2024).

CONCLUSION

This systematic review provides a comprehensive synthesis of Augmented Reality applications in biodiversity conservation, revealing a dynamic and promising field with significant potential to enhance conservation education, public engagement, and professional practice. AR technology, particularly mobile and game-based platforms, demonstrates consistent positive effects on biodiversity knowledge, environmental attitudes, and user engagement. Conservation educators, outreach professionals, and technology developers should consider augmented reality (AR) as a viable tool for enhancing biodiversity education and public engagement. Mobile AR platforms offer accessible, scalable solutions that can be deployed across diverse contexts with relatively low barriers to adoption. Game-based approaches are particularly effective for sustaining engagement and motivation. However, practitioners should maintain realistic expectations regarding behavioural outcomes and complement AR interventions with social supports, structural changes, and behaviour change strategies. By blending digital information with physical environments, AR creates opportunities for immersive learning, enhanced engagement, and decision support that complement traditional conservation approaches. As the evidence base matures and technology evolves, AR has the potential to democratise conservation knowledge, connect diverse audiences with biodiversity, and support more effective conservation action in an increasingly digital world.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Higher Education (MOHE) Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2023/WAB11/UTHM/02/2) Vot K469.

LIST OF ABBREVIATIONS

3D	:	Three-Dimensional
ACM	:	Association of computing machinery
AR	:	Augmented reality
GPS	:	Global Positioning System
IT	:	Information technology
PRISMA	:	Preferred reporting items for systematic reviews and meta-analyses

QR	:	Quick Response
SLAM	:	Simultaneous localisation and mapping
SLR	:	Systematic literature review
SUS	:	System Usability Scale
VE	:	Virtual environment
VR	:	Virtual reality

REFERENCES

Abidin, Y., Herlambang, Y. T., Saputra, D. S., & Susilo, S. V. (2023). The effects of augmented reality on students' perceptions, awareness, and attitudes toward sustainability of endangered animals. *Cypriot Journal of Educational Sciences, 18*(4), 655-672. https://doi.org/10.18844/cjes.v18i4.8812

Adla, K., Dejan, K., Neira, D., & Dragana, Š. (2022). Degradation of ecosystems and loss of ecosystem services. In J. C. Prata, A. I. Ribeiro, & T. Rocha-Santos (Eds.), *One Health* (pp. 281-327). Academic Press. https://doi.org/10.1016/B978-0-12-822794-7.00008-3

Afifi, M., Stryhn, H., & Sanchez, J. (2023). Data extraction and comparison for complex systematic reviews: A step-by-step guideline and an implementation example using open-source software. *Systematic Reviews, 12*, Article 216. https://doi.org/10.1186/s13643-023-02322-1

Al-Ansi, A. M., Jaboob, M., Garad, A., & Al-Ansi, A. (2023). Analysing augmented reality (AR) and virtual reality (VR) recent development in education. *Social Sciences & Humanities Open, 8*(1), Article 100532. https://doi.org/10.1016/j.ssaho.2023.100532

Belhaoua, A., Kornmann, A., & Radoux, J.-P. (2014). Accuracy analysis of an augmented reality system. In *2014 12th International Conference on Signal Processing (ICSP)* (pp. 1169-1174). IEEE. https://doi.org/10.1109/ICOSP.2014.7015184

Chan, B. Y., Ismail, Z. I. B. A., Jack, L. P., & Asli, M. F. (2019). Augmented reality mobile application: A feasibility study in a local national museum. *Journal of Physics: Conference Series, 1358*(1), Article 012057. https://doi.org/10.1088/1742-6596/1358/1/012057

Chen, Z., & Liu, Z. (2023). Research on endangered plant protection in Changbai Mountain based on AR interactive narration. *Academic Journal of Environment & Earth Science, 5*(3), 62-65. https://doi.org/10.25236/ajee.2023.050310

Cosio, L. D., Buruk, O., Fernández Galeote, D., Bosman, I. D. V., & Hamari, J. (2023). Virtual and augmented reality for environmental sustainability: A systematic review. In *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Article 6. ACM. https://doi.org/10.1145/3544548.3581147

de Oliveira, L. M. C., de Oliveira, P. A., Lim, A., Wheeler, A. J., & Conti, L. A. (2022). Developing mobile applications with augmented reality and 3D photogrammetry for visualisation of cold-water coral reefs and deep-water habitats. *Geosciences, 12*(10), Article 356. https://doi.org/10.3390/geosciences12100356

Fol, C. R., Zhao, J., Späth, L., Murtiyoso, A., Remondino, F., & Griess, V. C. (2025). Advancing forest biodiversity visualisation through mixed reality. *Scientific Reports, 15*, Article 15908. https://doi.org/10.1038/s41598-025-00285-y

Guzman, S., Estandian, R. J., Llarena, R., Fortunado, J., Daria, A., & Guadana, R. R. H. (2024). HARIBON FOUNDATION LIPAD PH: An Android-based mobile application framework featuring threatened air species in the Philippines. In *2024 7th International Conference on Information and Computer Technologies (ICICT)* (pp. 359-364). IEEE. https://doi.org/10.1109/ICICT62343.2024.00064

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers, 3*, 275-285. https://doi.org/10.1016/j.susoc.2022.05.004

Hamad, A., & Jia, B. (2022). How virtual reality technology has changed our lives: An overview of the current and potential applications and limitations. *International Journal of Environmental Research and Public Health, 19*(18), Article 11278. https://doi.org/10.3390/ijerph191811278

Hantono, B. S., Nugroho, L. E., & Santosa, P. I. (2018). Meta-review of augmented reality in education. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 312-315). IEEE. https://doi.org/10.1109/ICITEED.2018.8534888

Harrington, M. C. R., Tatzgern, M., Langer, T., & Wenzel, J. W. (2019). Augmented reality brings the real world into natural history dioramas with data visualisations and bioacoustics at the Carnegie Museum of Natural History. *Curator: The Museum Journal, 62*(2), 177-193. https://doi.org/10.1111/cura.12308

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J., & Wardle, D. A. (2005). Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs, 75*(1), 3-35. https://doi.org/10.1890/04-0922

Ibrahim, N. Z., Osman, R., Ali, A. M., Hamid, N. H. A., Nordin, S., Shahimi, S., Aminuddin, M. S. H., & Razak, F. A. (2022). Integrating augmented reality in learning plants. *IOP Conference Series: Earth and Environmental Science, 1019*(1), Article 012051. https://doi.org/10.1088/1755-1315/1019/1/012051

Imran, H., & Shahid, L. (2025). Exploring the use of artificial intelligence and mixed reality in conservation efforts for endangered wildlife in Pakistan. *Media and Communication Review, 5*(2), 172-197. https://doi.org/10.32350/mcr.52.08

Ingensand, J., Lotfian, M., Composto, S., Ertz, O., Oulevay, S., Oberson, M., Da Cunha, M., & Piot, D. (2018). Augmented reality technologies for biodiversity education: A case study. In *Proceedings of the 21st International Conference on Geographic Information Science (AGILE 2018)*.

Karpova, S. I., Chirich, I. V., Avtsinova, G. I., Shtukareva, E. B., Ukhina, T. V., & Gordeeva, T. A. (2021). Information and communication technologies in education: Video games as an effective environment for the development of self-directed learning of students. *Webology*.

Kumar, P. P., Thallapalli, R., Akshay, R., Sai, K. S., Sai, K. S., & Srujan, G. S. (2022). State-of-the-art: Implementation of augmented reality and virtual reality with the integration of 5G in the classroom. *AIP Conference Proceedings, 2418*(1), Article 020069. https://doi.org/10.1063/5.0081774

Lawton, M., Cunningham, S., & Convery, I. (2020). Nature soundscapes: An audio augmented reality experience. In *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3411109.3411142

Le, H., Nguyen, M., Tran, H., & Yeap, W. (2017). Pictorial AR tag with hidden multi-level bar-code and its potential applications. *Multimodal Technologies and Interaction, 1*(3), Article 20. https://doi.org/10.3390/mti1030020

Li, J., Wang, Z. L., Zhao, H., Gravina, R., Fortino, G., Jiang, Y., & Tang, K. (2017). Networked human motion capture system based on quaternion navigation. In *BodyNets International Conference on Body Area Networks*. https://doi.org/10.1145/0000000.0000000

Mercier, J., Leon Perez, L., Audrin, C., Bocher, E., Ertz, O., & Sander, E. (2025). The impact of augmented reality on biodiversity learning: An experimental study. *Computers in Human Behavior Reports, 20*, Article 100801. https://doi.org/10.1016/j.chbr.2025.100801

Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems, E77-D*(12), 1321-1329.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*, Article 1. https://doi.org/10.1186/2046-4053-4-1

Nahdiya, N., Kamaruzaman, N., Amalin, N., & Rozuki, H. (2024). An educational wildlife game-based learning application for young learners using augmented reality. *Applied Mathematics and Computational Intelligence, 13*(4).

Pavlik, J. V., & Bridges, F. (2013). The emergence of augmented reality (AR) as a storytelling medium in journalism. *Journalism & Communication Monographs, 15*(1), 4-59. https://doi.org/10.1177/1522637912470819

Pimentel, D. (2021). The peril and potential of XR-based interactions with wildlife. In *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3411763.3450378

Pimentel, D. (2022). Saving species in a snap: On the feasibility and efficacy of augmented reality-based wildlife interactions for conservation. *Journal for Nature Conservation, 66*, Article 126151. https://doi.org/10.1016/j.jnc.2022.126151

Pimm, S. L., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., & Loarie, S. (2015). Emerging technologies to conserve biodiversity. *Trends in Ecology & Evolution, 30*(11), 685–696. https://doi.org/10.1016/j.tree.2015.08.008

Poluakan, Y. F., Santa, K., Kumajas, S. C., & Rantung, V. P. (2024). Animal recognition application using mobile-based augmented reality at TK GMIM Solafide Uner. In *Proceedings of the 4th International Conference on Social Sciences (ICSS 2023)* (pp. 761-766). Atlantis Press. https://doi.org/10.2991/978-2-38476-198-2_104

Pombo, L., & Morais Marques, M. (2018). The EduPARK mobile augmented reality game: Learning value and usability. *EduPARK*. http://edupark.web.ua.pt

Pratisto, E. H., Thompson, N., & Potdar, V. (2022). Immersive technologies for tourism: A systematic review. *Information Technology & Tourism, 24*, 181-219. https://doi.org/10.1007/s40558-022-00228-7

Safitri, D., Marini, A., Irwansyah, P., & Sudrajat, A. (2025). Transforming environmental education with augmented reality: A model for learning outcome. *Social Sciences & Humanities Open, 12*, Article 101796. https://doi.org/10.1016/j.ssaho.2025.101796

Sandbrook, C., Adams, W. M., & Monteferri, B. (2015). Digital games and biodiversity conservation. *Conservation Letters, 8*(2), 118-124. https://doi.org/10.1111/conl.12113

Syiem, B. V., Webber, S., Kelly, R. M., Zhou, Q., Gonçalves, J., & Velloso, E. (2024). Augmented reality at zoo exhibits: A design framework for enhancing the zoo experience. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3613904.3642015

Tran, A. V., & Khoa, B. T. (2025). The impact of mobile augmented reality on green experience and destination choice intention in green tourism in Vietnam. *GeoJournal of Tourism and Geosites, 58*(1), 136-145. https://doi.org/10.30892/gtg.58112-1397

Voutos, Y., Palamas, S., & Mylonas, P. (2023). A user-driven platform for biodiversity conservation and promotion. In *2023 18th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*. IEEE. https://doi.org/10.1109/SMAP59435.2023.10255202

Wu, Y. C., Bogosian, B., Yenney, J., Coleman, P., Jalali, D., Pallag, J., Nassa, A., & Sharkey, T. (2024). Balboa Park Alive!: Exploring biodiversity through mobile augmented reality. In *SIGGRAPH 2024 Appy Hour*. ACM. https://doi.org/10.1145/3664294.3664358

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. In *Journal of Planning Education and Research*. https://doi.org/10.1177/0739456X17723971

Xu, X., Tang, C., Zhang, X., & Liu, Z. (2025). Enhanced Realism in Animal Fur Simulation for Digital Conservation: A Physically-Based Rendering and Augmented Reality Approach. *Applied Sciences (Switzerland)*, *15*(14). https://doi.org/10.3390/app15148049

Pertanika J. Sci. & Technol. 34 (4): 2421 - 2440 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Valorisation of Food Waste and Leftovers into Functional Bio-Compost: Optimisation of Maturation Period and Growth Performance of *Ipomoea aquatica*

Sarah Dini Muhamad Din[1], Muhammad Heikal Ismail[1]*, Tharsini Amma Puspadaran[2], Anie Yulistyorini[3], Muhammad Hazwan Hamzah[4], and Chuah Teong Guan[2]

[1]*Department of Environmental Science and Technology, Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[2]*Department of Chemical and Environmental Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

[3]*Department of Civil Engineering, Faculty of Engineering, Universitas Negeri Malang, 65145 Malang, East Java, Indonesia*

[4]*Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

ABSTRACT

Repurposing Food waste (FW) and leftovers (LF) involves using organic materials that are discarded at different stages of production, retail and consumption. Food waste is food that has decreased nutritional value and is not suitable for human consumption. Leftovers are defined as uneaten edible remains of a meal. In this research, both these wastes are used to produce functional bio-compost using the Bokashi method, an anaerobic food waste treatment technique that utilises a Bokashi bin and Bokashi bran to ferment organic waste before its application as a soil amendment. The duration of the maturation stage of both bio-composts was manipulated to 30 days, 40 days and 50 days. Leftover food bio-compost shows a greater reduction in bulk density, higher moisture content and water-holding capacity as compared to food waste bio-composts. As the maturation period increases, the bulk density decreases, and the moisture content and water-holding capacity of the bio-compost increases. The pH value of all 6 bio-composts produced was within the range of finished compost. The total volatile solids (TVS) decrease as the maturation period of bio-compost increases for both food waste and leftovers. Leftover at day 50 (LF-50) shows the best-optimised result as it has the lowest C: N ratio and highest nitrogen and potassium concentrations.

ARTICLE INFO

Article history:
Received: 12 June 2025
Accepted: 25 June 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.23

E-mail addresses:
sarahdini077@gmail.com (Sarah Dini Muhamad Din)
heikal@upm.edu.my (Muhammad Heikal Ismail)
tharsiniamma@gmail.com (Tharsini Amma Puspadaran)
anie.yulistyorini.ft@um.ac.id (Anie Yulistyorini)
hazwanhamzah@upm.edu.my(Muhammad Hazwan Hamzah)
chuah@upm.edu.my (Luqman Chuah Abdullah)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

The growth quality of *Ipomoea aquatica*, known as water spinach, has been used to compare bio-compost quality. From this aspect as well, LF-50 proves to be the optimum result. This shows that leftover foods produce a finer quality of bio-compost as opposed to food waste. As the maturation period increases, the bio-compost quality improves both physically and chemically.

Keywords: Bokashi, chemical properties, compost, food waste, growth quality, *Ipomoea aquatica*, leftover, physical properties

INTRODUCTION

Statistics reported by the Food and Agriculture Organisation of the United Nations, almost one third of the food produced globally for human consumption, which is approximately 1.3 billion tonnes, gets wasted every year (Blakeney, 2019). The two types of food that have been discarded are food waste and leftovers. First, food waste, often referred to as pre-consumer food, is the category of food that is taken from homes and restaurants' kitchens. Food remnants left on plates that customers discard is typically included in the category of leftovers. Another name for these wastes is post-consumer food (Arumugam et al., 2021). SWCorp Technology Research Division stated that Malaysians generated 38,000 tonnes of solid waste daily in the year 2016, in which food waste contributes to an amount of 15,000 tonnes. Based on constructs of leftover food waste generation behaviour, motivation in leftover food management, and opportunity and ability in leftover food management, are contributing to food leftover generation. Some of the factors include perception toward food leftovers, lifestyle, infrastructure, skill and knowledge (Aloysius et al., 2023).

Bio-compost is a medium formed from organic matter that has undergone decomposition and is widely used in the agriculture sector. Bio-compost can act as a single solution that can manage both food waste issues simultaneously. Composting is a biological method commonly conducted under anaerobic conditions in the absence of free oxygen (Mu et. al, 2016). This process is environmentally favoured compared to incineration, gasification and landfilling (Yin et al., 2016). Bio-compost usually consists of organic matter such as food waste and leftovers that is being decomposed and recycled as fertiliser or manure. Biodegradation of these components contributes to many macronutrients, nitrogen, phosphorus and potassium, in the compost, which makes it a suitable growing condition for plants. As for the aerobic process in the presence of oxygen, microorganisms digest organic compounds and release carbon dioxide, nitrate, and nitrite (Xiao et al., 2009). Anaerobic composting is composting that proceeds without oxygen entanglement (Meena et al., 2021).

Compared to thermophilic compost and vermicompost, the EM for Bokashi treatments offer a nitrogen nutrient that is longer-lasting and plant-available. This technique increases the biomass and aboveground height of plants. This resulted from the compost's capacity to deliver a sustained supply of nitrogen along with an appropriate level of potassium and

phosphorus for plant growth. Ultimately, this study found that the Bokashi treatment was an appropriate substitute fertility supplement for the system of organic vegetable production (Arumugam et al., 2021). Home composting of food waste could significantly benefit the environment. Research comparing organic and chemical fertilisers on watermelon growth showed organic fertilisers had the highest results in consistency parameters. Composting was also simpler to incorporate and less expensive than other technical approaches (Arumugam et al., 2021).

The objective of the study is to evaluate the differences in physical properties of bio-compost between food waste and leftovers at 30, 40 and 50 days of maturation. Regulating the maturation phase can enhance understanding of the ideal timing for producing high-quality bio-compost. Additionally, this study also aims to assess the quality of bio-compost by the growth of *Ipomoea aquatica.* The growth of *Ipomoea aquatica* can deepen insight regarding the quality of bio-compost and discover the effect of the bio-compost in practical use in agriculture. The acquired knowledge can encourage the production of functional bio-compost from food waste and leftovers using the Bokashi method. The findings of the study can contribute to knowledge of the optimal duration of the maturation stage of food waste and leftovers. Composting food waste and leftovers can contribute to sustainable agriculture practices where organic matter from waste is converted into valuable products that can enhance nutrients in soil for agriculture

MATERIALS AND METHODS

Materials

Materials used in the research are food waste and leftover foods obtained from the cafeteria of the Faculty of Engineering, Universiti Putra Malaysia (UPM) and household food waste and leftovers as the feedstock for the composting process. The food wastes are all the food material that is wasted and not used during the cooking process and expired manufactured food that cannot be consumed. For the leftovers, these are foods collected from plate waste produced by the consumers of the cafeteria and household waste (Nakagiri et al., 2024).

Experimental Procedure

The experimental procedure of this research firstly begins with the preliminary study of drying kinetics of the food materials. In this section, hot air-drying pre-treatment was conducted on the feedstocks at 50°C for 6 hours. Then, the pre-treated food wastes and leftover foods were filled into two Bokashi Kit Bit Coins, respectively. Upon filling to a thickness of 5 cm, Bokashi Bran was added to each layer of food waste, and leftover food was added to the bins. Bokashi Bran was added to introduce Effective Microorganisms (EM) in the compost kit. The compost kits were left to ferment the waste for 14 days; then the waste was removed and added to a soil bed for maturation (Melini & Melini, 2024).

The duration of the fermented food wastes and leftover foods left to compost under the soil beds is divided into three different periods, which are 30 days, 40 days and 50 days, respectively. Hence, from this part, 6 different composts were produced in the process. The physical properties of the bio-compost were compared by assessing the moisture content, bulk density and water-holding capacity of all 6 bio-composts. Then, samples of the bio-composts produced were collected and pre-treated to be analysed for the nutrient concentrations in the bio-compost. The nutrient contents of Carbon (C), Hydrogen (H), Nitrogen (N) and Sulphur (S) were analysed using a CHNS Analyser machine at the Faculty of Engineering, UPM and Potassium (K) concentration was analysed using an Auto Analyser at the Faculty of Agriculture, UPM.

Then, the pH of the composts was measured using a pH meter, and the total volatile solids (TVS) were determined by *ashing* the compost at 550°C for 4 hours. To compare the bio-compost performance, Water Spinach *(Ipomoea aquatica)* were planted, and the growth quality of the plant was assessed. For this, water spinach was grown in 6 different types of bio-compost prepared for this experiment. As a control, the water spinach w*as* also grown using soil fertilised with commercial organic fertiliser as well as an unfertilised soil environment. The growth aspects of plant height and number of leaves were observed daily. The size of the leaves and crop growth rate were measured and calculated after harvesting the plants. From the value of dry weight obtained at two different periods, the crop growth rate of water spinach was calculated.

The feedstocks for this experiment were collected from the Cafeteria of the Faculty of Engineering, UPM and household waste. These feedstocks comprised both food waste and leftover food. The materials obtained were kitchen waste used for cooking and uneaten food that was collected during lunchtime, which is between 1000 am to 1400 pm. The feedstock composition comprises rice, noodles, bones (fish), fruits and vegetables. Then, other components such as eggshell and fish meat are also present in the feedstock but only in small proportions. The total food waste and leftover foods collected were 4 kg each. The method used to decompose the food waste into compost is by using the Bokashi method. This method of composting is established in Japan and uses microbes to decompose food and, at the same time, masks the foul smell produced in the decomposing process. In this part of the methodology, the equipment that will be used is the Bokashi Kit Bit Coin.

To begin with, spread one small handful of Bokashi Bran, approximately ¼ cup, at the bottom of the buckets, with one labelled as a leftover food compost kit and another food waste compost kit. Then, add the food waste and leftover foods respectively to the bins within the range of 2 to 3 inches. Add a teaspoon of Bokashi bran to cover the layer of the waste materials. Subsequently, repeat the steps by layering the food waste and adding the bokashi bran until the compost bin is full. Later, close the compost kit and ensure the lid

is covered tightly. This is to facilitate an anaerobic environment in the bin that will aid in the composting process. After completing these processes, the compost bins were closed for 14 days. The treated wastes were poured into 6 different planters filled with soil after completing 14 days of fermentation in the bin, three for fermented food waste and another three for fermented leftover foods. These wastes were covered with a 5-inch depth of soil and left to compost for different lengths of time, as shown in Table 1.

This step was carried out to compare the quality of the bio-compost when the duration of the curing period is extended. Finally, after completing the curing period at 30 days, 40 days and 50 days, the samples of the compost were collected to undergo characterisation and analysis of the nutrient content.

Table 1
Bio-compost labelling (soil bed)

Labelling	Description
Planter 1 (FW30)	Fermented food waste left to compost under the soil bed for 30 days.
Planter 2 (FW40)	Fermented food waste left to compost under the soil bed for 40 days.
Planter 3 (FW50)	Fermented food waste left to compost under the soil bed for 50 days.
Planter 4 (LF30)	Fermented leftover food left to compost under the soil bed for 30 days.
Planter 5 (LF40)	Fermented leftover food left to compost under the soil bed for 40 days.
Planter 6 (LF50)	Fermented leftover food left to compost under the soil bed for 50 days.

Physical Properties, pH and Total Volatile Solids (TVS)

The physical properties that were measured and compared are the moisture content, bulk density and water-holding capacity (WHC). As for the moisture content, the compost sample of 100 g will be dried in an oven at 105°C for 24 hours (Kumar et. al., 2010). The final mass of the compost was measured after the 24-hour duration. The difference in mass determines the amount of moisture available in the compost. The bulk density of the composts was measured by filling 150 g of bio-compost into a 10 L container. The compost was then compressed until there was no large void space; the volume occupied was measured. The bulk density was calculated by dividing the weight of the bio-compost by the volume occupied (kg/m^3). Then, for the water-holding capacity, a wet sample of bio-compost weighing (Wi) with a known initial moisture content was left to soak in water for 1-2 days. Excess water was drained using Whatman 32 filter paper; the saturated sample was weighed once again (Ws). The amount of water that was retained by the bio-compost was calculated using Equation 1.

$$WHC = \frac{(WS - WI) + MC \times WI}{(1 - MC) \times WI} \quad [1]$$

where,
WHC = Water Holding Capacity (g water/g dry compost)
WI = Initial Weight of Bio-Compost
Ws = Final Weight of Bio-Compost (After Filtration)
MC = Moisture Content of Bio-Compost

To determine the pH of the 6 different bio-composts produced from both food waste and leftover food, a solution of composts was made. The solution of composts was produced by diluting the sample composts into distilled water at a ratio of 1:5 (compost: distilled water). Then, the pH meter was dipped into the solution to measure the pH of the solution (Kumar et. al., 2010). Then, the Total Volatile Solid (TVS) content in the compost was determined by using the product from the moisture content test. The sample that has been heated up to 105°C to determine the moisture content will then be formed into ashes in a furnace at a temperature of 550°C for 4 hours (Melini & Melini, 2024). After 4 hours, the product produced will be weighed, and the mass produced is the mass of TVS available in the composts. This step was carried out for all 6 compost*s*.

Nutrient Content in Bio-Compost

The macro-element nutrients in the soil will be analysed and measured using the Auto Analyser machine. The Auto Analyser machine functions to conduct environmental analysis on soil and composts, and the design of this equipment is based on segmenting a continuous flowing stream with air bubbles. The list of micro and macro elements that were measured is Nitrogen (N), Phosphorus (P), Potassium (K), Carbon (C), Hydrogen (H) and Sulphur (S). For the concentration of C, H, N and S, the samples of Bio-Compost were sent to the Material Characterisation Laboratory, Faculty of Engineering, UPM, to be analysed using a CHNS analyser. The concentration of potassium (K) was determined in the Faculty of Agriculture using the Auto Analyser machine. This equipment functions only to analyse organic liquids; hence, a solid-liquid extraction process must be conducted on the compost before the test.

The steps involved in the solid-liquid extraction process began by collecting 20 g of bio-compost from the leftover food compost and food waste compost and placing it into two separate labelled bags. The samples were then placed in a clean Petri dish and dried in an oven at 105°C for 2 days. After drying, the samples were crushed using a mortar and pestle until a refined, smooth granule was formed. Then, the samples were sieved using a 2 mm sieve and the products were stored in a beaker of conical flask at room temperature (Alves et al., 2023; Whittinghill et al., 2024).

Universal Soil Extractant were used to extract the potassium (K) nutrient in the Bio-Compost. The preparation method began with adding 53.12 g of sodium sulphate and 37.80 g

of sodium bicarbonate in 800 mL distilled water to a conical flask of 1L. Then, the pH of the solution was maintained at 8.5 using sulfuric acid solution and sodium hydroxide solution. 2.5 g of the dried bio-compost was loaded into a 100 mL conical flask. After that, 50 mL of the universal soil extractant and 0.5 g of phosphate-free activated carbon were added. The solution was then stirred for 10 minutes at 200 rpm. The solution was then filtered with filter paper. The extracted solutions were then sent to the Analytical Laboratory 2, Faculty of Agriculture, UPM. All analyses were replicated three times to ensure data reliability and enable statistical analysis.

Germination of *Ipomoea Aquatica*

For this experiment, the germination of water spinach was conducted in an outdoor environment since this plant needs plenty of sun or partial shade to aid its growth. Then, aside from comparing the growth rate of water spinach for both food waste and leftover food compost, the water spinach will also be grown in unfertilised soil, and another will be fertilised using commercial organic fertiliser. The commercial organic fertiliser that was chosen to be assessed and compared to the bio-composts is BIO LOF Organic Fertiliser. The unfertilised soil and commercial fertiliser were used as a control in this experiment. The procedure that needs to be taken begins with 8 bags of planter bag will be filled with the same amount of soil. About 2/3 of the planter bag will be filled with soil. The plant bags were labelled as shown in Table 2.

Table 2
Water Spinach (WS) planter bag label

Type of Soil/Fertiliser/ Compost	Label
Unfertilised soil	WS-UF
Commercial organic fertiliser	WS-CF
Food waste compost -FW 30	WS-FW 30
Leftover food compost -LF 30	WS-LF 30
Food waste compost -FW 40	WS-FW 40
Leftover food compost- LF 40	WS-LF 40
Food waste compost -FW 50	WS-FW 50
Leftover food compost -LF 50	WS-LF 50

Before planting, the seeds of water spinach were left to soak in a water bath for 24 hours. This step was taken to promote germination of the water spinach. For each plant bag, 3-4 water spinach seeds are planted in the WS-labelled plant bag. Once the water spinach germinates, the seedlings will be relocated to 5 cm × 5 cm. This is to prevent the plants from wilting and to germinate effectively. To get the optimum reading, the two best plant readings were obtained for each planter, and the average was calculated. The watering schedule for water spinach is 1-2 times daily. The water spinach will be left to grow for 30 days (Ibrahim et al., 2019; Ni'am & Yuniati, 2021).

For the growth of water spinach, four aspects will be measured: the height of the plant, number of leaves, size of leaves and dry weight of the leaves. To begin with, the height of the plants will be measured daily until the plant is fully mature. Firstly, the ruler was set at the base of the plant the height was recorded. The value was recorded with both the date and

height value. Then, for the number of leaves, the leaves were counted daily and recorded. The leaf area and dry weight were calculated after 30 days of growth; the leaves' surface area and dry weight of the greens were measured. The leaf area was measured using an application named “Petiole”. This application runs by calibrating the mobile camera with a Petiole Pad, and once the calibration is completed, the leaf was placed in front of the camera. The leaf area was then measured automatically by the app, and the results obtained were recorded. As for the dry weight of plants is an estimate of the plants' weight once all water has been removed from them.

The steps involved in this part began with the water spinach being weighed and then dried in an oven at 70°C for 1 hour. After 1 hour, the plants were weighed again to obtain the dry weights. This step was conducted for all 8 categories of water spinach. Then, based on dry matter accumulation, the values of the growth parameter of the crop growth rate (CGR) (mg m-3 day-1) were calculated for each treatment of Water Spinach as in Equation 2.

$$CGR = \frac{1}{\mathrm{GA}} \times \frac{(W_2 - W_1)}{(T^2 - T^1)} \quad [2]$$

In which GA is the ground area (m^2) of the crop collected, and W_1 is the dry weight at time T_1. W_2 is the dry weight at time T_2.

Statistical Analysis

The published results are expressed as mean ± standard deviation (SD) of triplicates. Relative change (%) was used in descriptive statistics and One-way ANOVA to analyse the height of Water Spinach against the 6 bio-composts. One-way ANOVA was also performed to determine the crop growth rate. ANOVA test employed Tukey’s post hoc analysis. All statistical analyses were conducted with SPSS software (v.12, IBM, SPSS, New York, USA) at a significance level of 0.05. Differences in means are marked with a letter in the figures.

RESULTS AND DISCUSSION

Pre-treatment of Food Waste and Leftovers

The food waste and leftovers were subjected to pre-treatment at a temperature of 55°C using an Air-Drying Incubator. This bio-drying method functions to remove water from organic waste through microbial activities. The biomass product with lower water content restricts odour emission and has a higher potential to produce high-value-added products such as bio-compost (Han et al., 2024). The suitable range for proper growth of microorganisms for this process is between 40°C and 70°C; hence, 55°C was chosen to carry out the drying (Tun & Juchelková, 2019). The Hot-Air Drying Incubator that has been used is a

Binder BOD Incubator, and the operating limit of this incubator is -5 to 100°C. The type of drying equipment used is a sealed type and operates in forced-air circulation mode. The raw materials of food waste and leftover foods were dried in the incubator until a constant reading was obtained. The results of the moisture ratio against time for both the food waste and leftovers are displayed in Figure 1.

The moisture ratio determines the unaccomplished moisture change, the ratio of the free water still to be removed at time t, over the initial total free water. From Figure 1, the drying of the food waste took an estimated time of 420 minutes to reach a stable reading. The moisture ratio of the food waste material has dropped by about 21.3%. As for leftover foods, the time taken to complete the drying process was 540 minutes. The moisture ratio for the leftover food materials has dropped by approximately 43.9%. From this part of the pre-treatment, it can be deduced that the drying time for food waste is shorter compared to leftover food (Lai et al., 2023).

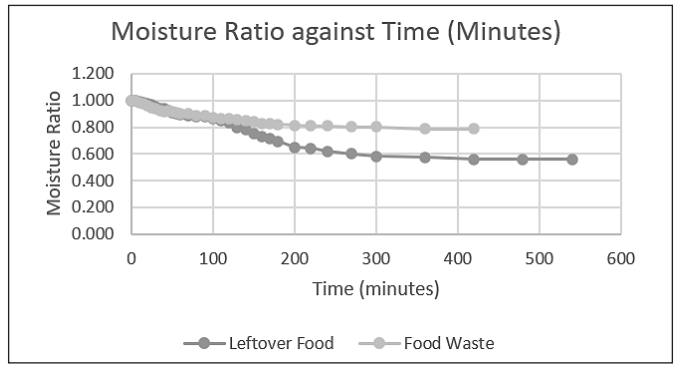


Figure 1. Moisture ratio against time of food waste and leftover foods

Properties of Food Waste and Leftovers

The properties that were analysed in the food waste and leftover compost are the pH value, moisture content and total volatile solids. The results obtained were as tabulated in Table 3. The readings obtained for moisture content and total volatile solids are statistically significant ($p \leq 0.05$), whereas bulk density and pH value are not statistically significant ($p \leq 0.05$). P are stand for p-value (probability value) is a statistical measure that helps determine the significance of results in a hypothesis test. From the results obtained, there are some differences in the parameters measured for the properties of the raw materials selected for this research. Firstly, the bulk density of Food Waste seems to be slightly higher compared to leftover foods. The variation between both materials in this aspect might be due to the state of the raw material. Leftover food is subjected to more moisture; hence, it is easily compressed (Lee et al., 2024).

Table 3
Properties of raw material

Properties	Leftover	Food Waste
Bulk density (kg/m^3)	781.29	813.06
pH	6.11	6.16
Moisture content (%)	82.15*	72.65*
Total volatile solids (%)	76.67*	71.33*

Note. *Indicates the mean difference is significant at $p \leq 0.05$

The moisture content and Total Volatile Solids (TVS) in leftover food were higher values compared to food waste. The TVS results obtained for both food waste and leftover is within the range which are within 73% to 98% (Selvam et al., 2021). For moisture content, both materials recorded a difference of 9.5%, whereas for total volatile solids, it is 5.34 %. The pH of food waste and leftover foods had a mean of 6.16 and 6.11, respectively, with a minimal difference of 0.82% between both.

Physical Properties of Bio-Compost

The physical properties that were analysed are the moisture content (%), bulk density (kg/m3) and water-holding capacity (g water/g dry sample) of the bio-compost. All the physical parameters measured for the bio-compost are statistically significant ($p \leq 0.05$). The desired and typical moisture content of a completed compost is within the range of 40 – 60 %. Based on the results from Table 4, it can be observed that all the 4 bio-compost not including LF-30 and FW-30, have moisture compositions that fit within that range. The moisture content in bio-compost has a physical impact on permeability and significant biological implications for the soil (Ameen et al., 2016). The moisture content has a great influence on microbial activity. The LF-30 and FW-30 bio-composts have moisture content slightly below 40%, at 39.03% and 35.98%, respectively. In the leftover food compost aspect, LF-50 recorded the highest, which is 46.18%, and the lowest is LF-30, which is 39.03%.

Table 4
Physical properties of bio-compost

Type of Bio-compost	Moisture Content (%)	Bulk Density (kg/m^3)	Water Holding Capacity (g water / g dry sample)
LF-30	39.03*	574.73*	3.28*
LF-40	40.10*	549.59*	3.61*
LF-50	46.18*	530.54*	3.98*
FW-30	35.98*	600.65*	3.17*
FW-40	40.06*	588.32*	3.49*
FW-50	42.60*	581.43*	3.79*

Note. *Indicates the mean difference is significant at $p \leq 0.05$

As for the food waste bio-compost, FW-50 recorded the highest moisture content at 42.60% and FW-30 the lowest at 35.98%. For bulk density, it is a function of the local pressure that is being exerted on the material. This property is influenced by the effects of subsidence, loss of organic matter due to biological degradation processes, and changes in moisture distribution. The overall bulk density for food waste and leftover decreased gradually from the 30-, 40- and 50-day periods. The biological decomposition is still active and present during the 40- and 50-day periods of composting; hence, the bulk density declined slightly for both the food waste and leftover compost. It can be observed that as the period of composting increases, the bulk density of the bio-compost decreases. The bio-compost produced from leftover food is within the range of 530.54 kg/m3 to 574.73 kg/m3, whereas for food waste it is 581.43 kg/m3 to 600.65 kg/m3.

The leftover food seems to have reduced in bulk density even more as compared to the food waste, which might be due to the rate of organic matter decomposition. Organic matter has a lower density than the mineral fraction in compost, and it can increase the void spaces in the compost heap, which leads to a decrease in overall bulk density.

Water Holding Capacity (WHC) of bio-compost is the amount of water held between field capacity and wilting point. The WHC is affected by the number of pores and pore-size distribution. The main component that determines the WHC of compost is the compost texture and organic matter content (Beesley et al., 2010; Cervera-Mata et al., 2021). High organic matter in compost increases the WHC and the moisture content in the bio-compost. The result of WHC for the leftover foods are in the range of 3.28 to 3.98 g of water/g of dry sample. For food waste compost, the WHC is within the range of 3.17 to 3.79 g of water/g of dry sample. The WHC for leftover food is slightly higher compared to the food waste compost. The WHC plays a very important role in plant growth quality. For both the food waste and leftover food, the WHC increases as the period of the composting process increases. The level of WHC quality can affect the overall water spinach growth. Low WHC causes vegetation to undergo stress, which will affect the transpiration rate of the plant. The decrease in transpiration rate, in turn, closes stomata and inhibits plant growth. The result obtained on the WHC of the bio-compost might influence the growth quality of the water spinach produced.

Chemical Properties of Bio-Compost

The chemical properties that have been analysed in the bio-composts produced, as in Table 5, are the pH, total volatile solids (TVS), carbon to nitrogen ratio (C: N), total Carbon content (C%), and hydrogen content (H%). Nitrogen (%), Sulphur content (S%) and Potassium concentration (mg/L).

Table 5
Chemical properties of bio-compost

Properties	LF-30	LF-40	LF-50	FW-30	FW-40	FW-50
pH	6.64	6.65	7.01	6.3	5.07	6.69
TVS (%)	63.5*	50*	47.5*	58.5*	50*	45*
C: N	22.06:1	14.37:1	12.46:1	21.51:1	15.44:1	13.48:1
C (%)	42.14	39.94	37.73	41.95	39.86	38.01
H (%)	5.12	4.96	4.77	5.69	5.74	5.72
N (%)	1.91	2.78	3.03	1.95	2.58	2.82
S (%)	0.25	0.25	0.25	0.26	0.25	0.26
K (mg/L)	104.8	138.1	265.8	89.42	170.8	130.4

Note. *Indicates the mean difference is significant at $p \leq 0.05$

The results recorded for pH value and the total volatile solids (TVS) of the bio-composts are statistically significant ($p \leq 0.05$). The pH value is obtained from different sections of the compost pile, and the average value is tabulated. Based on the results tabulated in Table 5, the pH value obtained from the leftover food compost is in the range of 6.64 to 7.01, whereas for the food waste compost the pH value obtained is at 5.07 to 6.69. Generally, mature compost has a pH value of between 6 and 8 (Wang et al., 2022). The H levels measured are between highest 5.74% and lowest at 4.77%. The increase in pH to a base condition is beneficial to the composting process because alkaline conditions can limit the growth of diseases such as fungi that thrive in acidic environments (Irvan et al., 2018). The higher hydrogen generation in compost can be attributed to its elevated mineralisation (Barneto et al., 2009).

Before the maturation stage, the compost will be quite acidic. As the temperature of the pile increases when the mesophilic stage takes over, the composting material eventually neutralises. The increase in pH value of leftover compost can also be explained by the ammonification process undergone by the compost. The ammonification of compost-derived nitrogen and the presence of Ca and Mg in the compost may release hydrogen ions (H+) from the subsequent site. For the food waste compost, the pH reading at FW-30 is 6.3, then it dropped to 5.07 at FW-40 and stabilised at 6.69 for FW-50. The dip in the pH value might be due to the anaerobic digestion that is still occurring in the compost. Anaerobic digestion occurs if the turning of compost and improved air circulation are not completely fulfilled in the compost heap. However, the increase in acidity level at 40 days is just slightly below the mature compost pH requirement. Once it is used for planting in soil, the compost will neutralise accordingly and might not significantly affect the crop growth rate of the Water Spinach.

For the FW-50, the pH value is 6.69, which is mature and suitable to be used for fertilising application. Total volatile solids in compost are the elements that can easily

transform from the solid phase to its vapor phase, without a liquid phase interval. These solids are the organic matter content in the compost, which is the food waste and leftover material. The total organic matter during the composting process will decrease during the composting process. A significant reduction in volatile solids indicates the decomposition, mineralisation and maturity of the composting process, in which both leftover foods and food waste show a clear sign of reduction.

The reduction of volatile solids can be compared for the food waste and leftover food compost. The initial value of LF feedstock has a reading of 76.67%, and it reduced to 63.5%, 50% and 47.5% for LF-30, LF-40 and LF-50, respectively. As for the food waste compost, the FW feedstock recorded 71.33% initial volatile solids, and it decreased to 41.95%, 39.86% and 38.01% for FW-30, FW-40 and FW-50, respectively. The quantity of trend reduction for leftover food compost shows a higher reduction reading of reduction which is 33.32%, whereas for food waste it is 29.17% for the overall reading. The degradation of volatile solids can be influenced by the proper availability of substrate to the microbes. Total Carbon in a compost is the direct measurement of the organic and inorganic carbon in the bio-compost sample. The carbon in bio-compost is mainly in organic form. Based on Table 5, the total organic carbon decreased gradually in the three maturation periods of food waste and leftover food. The total carbon for leftover food is within the range of 37.73% to 42.14%, and 38% to 41% for food waste. The leftover food recorded lower total carbon, which can be a result of a higher rate of degradation of organic matter in the compost pile.

For both food waste and leftovers, the total carbon experiences a decreasing trend. Carbon content will naturally drop during the composting process as microbes consume it as a source of energy, and some is released in the form of Carbon Dioxide (CO^2). Nitrogen is a vital source of nutrients for plant growth, and it's the major component of chlorophyll. The total nitrogen concentration can be observed to be increasing when the composting maturation period increases.

This might be due to net loss of dry mass of Carbon Dioxide and water loss by evaporation during oxidation of organic carbon (Oliveira et al., 2022; Ugak et al., 2022). For compost, nitrate -nitrogen may increase as the compost matures, and most of the N will be in the organic form. The concentration of nitrogen from all 6 composts is marginally different. From Table 5, LF-50 shows the highest reading of nitrogen concentration, which is 3.03%, and the lowest recorded is LF-30 at 1.91%. From 1.91 % at 30 days, the concentration leapt to 2.78 % at LF-40, which is a 45% increase and finally, for LF-50, the N% increased by about 8.9%. As for FW-30 to FW-40, the changes recorded surged by 32.3% and at FW-50 the concentration increased to 9.3%. From the results, it can be observed that the concentration of nitrogen increases as the maturation period increases. From the 30-day and 40-day gap of maturation period for both leftover food and food waste

experience a notable increase in N% compared to 40 and 50 days, which is <10%. C/N, or carbon-to-nitrogen ratio, is the ratio of the mass of carbon to nitrogen in soil. The essential C: N ratio is < 20-25:1. C: N ratio higher than 20:1 indicates that the compost or soil amendment is poor in nitrogen and rich in organic carbon, which leads to Nitrogen immobilisation (Wei et al., 2022).

These types of compost undergo microbial decomposition before nutrients become accessible to the plants. From the results, LF-30 and FW-30 displayed a C: N reading above 20:1. This might affect the growth potential of Water Spinach (Aylaj & Adani, 2023; Nair et al., 2024; Noor et al., 2024). This is because of poor nitrogen concentration, as microbes might immobilise soil nitrogen into their body to maintain their own C: N ratio and the water spinach crop might suffer from nitrogen deficiency. As for the other 4 bio-composts, the C: N ratio is within the range of 12 to 15:1 ratio which is below 20:1 (Al-Tawarah et al., 2023; Gelsomino et al., 2023; Kumar et al., 2024).

This feature allows faster N mineralisation and eases the release of nutrients to support plant growth. For the Potassium concentration analysis, the equipment used to carry out it is by using the Auto Analyser machine. This concentration analysis was conducted separately from the other element analysis, and the concentration recorded is in the unit (mg/L). The highest reading recorded for K is LF-50, which is 265.8 mg/L, and the lowest is FW-30 at 89.42 mg/L. Based on studies, the potassium concentration will increase as the composting period increases. The result obtained from the leftover food shows the same trend. The increase in potassium concentration from LF-40 to LF-50 is almost a 92% increment. However, there were fluctuations in the concentration of potassium in food waste compost (Jahan et al., 2022; Kabir, 2022; Sarminingsih et al., 2023).

The K (mg/L) concentration increased by about 91% from FW-30 to FW-40 and slumped by 30.9% (130.4 mg/L) at FW-50. This decrease in K concentration might be caused by the activity of microorganism which consumes nutrients during the curing period (Almulla et al., 2024). The absence of a similar decline in the leftover treatment may be due to differences in their original composition. Leftover food contains cooked rice, meat, sauces, oils, and other processed materials, which have a different potassium concentration and decomposition pattern from food waste. In addition, differences in moisture content, salt concentration, particle size, and the amount of liquid released during decomposition could have affected potassium concentration. The Sulphur content in the bio-compost is an essential mineral that helps to produce important enzymes and contributes to the formation of plant proteins. As for all the 6 bio-compost, the sulphur concentration obtained is not significantly different from one another, with a concentration of 0.25 to 0.26 % (Jahan et al., 2022; Noor et al., 2024).

Growth Analysis of Water Spinach

There are four parameters that have been measured in the growth analysis of Water Spinach. The parameters that were evaluated are the height of the plant (cm), number of leaves, size of leaves (cm^2), and crop growth rate ($g/m^2/day$) of Water Spinach.

The ANOVA conducted on the data collected for the crop growth rate was statistically significant ($p \leq 0.05$). The crop growth rate, as shown in Table 6, is influenced by the area of plant growth, duration of plant growth and the dry weight of the plant (Safdar et al., 2022).

Table 6
Crop growth rate of water spinach

Type of Compost	Crop Growth Rate ($g/m^2/day$)
Commercial Organic Fertiliser	14.11*
Unfertilised Soil	1.35*
LF-30	5.22*
LF-40	8.50*
LF-50	12.85*
FW-30	4.93*
FW-40	8.50*
FW-50	10.43*

Note. * Indicates the mean difference is significant at $p \leq 0.05$

The highest crop growth rate recorded is 12.85 $g/m^2/day$ for LF-50, whereas the lowest is 4.93 $g/m^2/day$ for FW-30. In addition, the crop growth rate of LF-50 is approaching the quality of growth rate of a commercial organic fertiliser, with a marginal difference of 9.8%. The crop growth rate difference between FW-30 and LF-50 is stimulated due to the considerable gap in nitrogen concentration, potassium concentration, as well as the C: N ratio of both bio-composts (Wahi et al., 2019).

CONCLUSION

Based on the physical and chemical properties assessed, there were comparatively different result obtained. The height of the Water Spinach for each bio-compost varies significantly. LF-50 recorded the tallest, followed by FW-50, LF-40, FW-40, LF-30 and FW-30. The height of Water Spinach using bio-compost LF-30 and FW-30 is quite shorter compared to the other four bio-composts. The size of leaves (11.80 cm^2) and number of leaves of LF-50 (11 leaves) were the highest as well. The crop growth rate of the Water Spinach using LF-50 showed the highest rate and difference of 8.93% compared to a commercial organic fertiliser. These findings suggest that the leftover food bio-compost produces a better quality of bio-compost as compared to food waste, and the increase in maturation period of bio-compost improves the final product quality.

Various recommendations can be made to improve the research. To begin, food waste and leftover foods can be pre-treated at various temperatures before being composted. This approach allows us to investigate the effect of raw material drying kinetics on the quality of bio-compost produced. Furthermore, the bio-composts created could be used to evaluate various vegetation and plant growth. This process is done to better understand the influence and limitations of bio-compost quality on various types of plants. Furthermore,

future research can examine the N:P:K ratio of each bio-compost to determine the optimal nutrient balance and which plant best meets these requirements. Lastly, different types of composting can also be applied to the same type of waste; thus, the quality of bio-compost produced by other methods under the same variables can be explored.

According to the research, organic waste has significant potential to be utilised in plant growth, instead of ending up in the landfill. The energy produced from the degradation process is instead used as a nutrient provider to the soil, rather than producing emissions that contribute to the greenhouse effect. The composting process using a bokashi kit utilises minimal space and low maintenance. Since both food waste and leftovers are commonly sourced from households, this method is very suitable for being adapted to households. This research identifies the most effective durations for the compost and reveals physical and chemical properties in each phase. Through effectively exposing the findings of this research to the public, we can promote increased engagement in home-based composting.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Putra Malaysia for supporting this study under the MyTiger Matching Grant (MATCHING GRANT/2023/9300493) and Universiti Negeri Malang through the Postgraduate Scheme. The authors also gratefully acknowledge the financial support provided by Universiti Putra Malaysia for publication under the Dana Penerbitan Jurnal (9001103). In addition, the authors extend their sincere appreciation to the staff of Universiti Putra Malaysia, particularly the laboratory team of the Department of Environmental Science and Technology and the Department of Chemical and Environmental Engineering, Faculty of Engineering, for their collaboration, excellent facilities, and invaluable guidance throughout this study.

DATA AVAILABILITY

The authors declare that the data supporting the findings of this study are available within the paper.

DECLARATIONS

The authors have no conflicts of interest to declare that are relevant to the content of this article.

CONTRIBUTION

Sarah Dini Muhamad Din and Tharsini Amma Puspadaran are preparing the conceptualisation, visualisation, writing, methodology, data curation and investigation of the original draft preparation. Annie Yulistyorini, Muhammad Hazwan Hamzah and Chuah Teong Guan

oversee the review of the whole article. Muhammad Heikal Ismail is supervising the publication.

LIST OF ABBREVIATIONS

C	:	Carbon
Ca	:	Calcium
CGR	:	Crop growth rate
cm	:	Centimeter
EM	:	Effective microorganisms
FW	:	Food waste
g	:	Gram
H	:	Hydrogen
K	:	Potassium
Kg	:	Kilogram
L	:	Liter
LF	:	Leftover
m	:	Meter
Mg	:	Magnesium
mg/L	:	Milligram pr liter
mm	:	Millimetre
N	:	Nitrogen
P	:	Phosphorus
S	:	Sulphur
TVS	:	Total volatile solids
WHC	:	Water holding capacity
WS	:	Water spinach
°C	:	Celsius
%	:	Percentage

REFERENCES

Ameen, A., Ahmad, J., & Raza, S. (2016). Effect of pH and moisture content on composting of municipal solid waste. *International Journal of Scientific and Research Publications, 6*(5), 35-37.

Al-Tawarah, B., Alasasfa, M. A., & Mahadeen, A. Y. (2023). Efficacy of compost and vermicompost on growth, yield and nutrients content of common beans crop (*Phaseolus vulgaris* L.). *Journal of Ecological Engineering, 25*(2), 215-226. https://doi.org/10.12911/22998993/176862

Aloysius, N., Ananda, J., Mitsis, A., & Pearson, D. (2023). Why people are bad at leftover food management? A systematic literature review and a framework to analyse household leftover food waste generation behaviour. *Appetite, 186*, Article 106577. https://doi.org/10.1016/j.appet.2023.106577

Almulla, L., Thomas, B. M., Jallow, M. F. A., Al-Roumi, A., Devi, Y., & Jacob, J. (2024). Rotary drum composting of organic school wastes and compost valorisation. *Sustainability, 16*(6), Article 2428. https://doi.org/10.3390/su16062428

Alves, D., Villar, I., & Mato, S. (2023). Community composting strategies for biowaste treatment: Methodology, bulking agent and compost quality. *Environmental Science and Pollution Research, 31*(7), 9873-9885. https://doi.org/10.1007/s11356-023-25564-x

Arumugam, V., Ismail, M. H., Puspadaran, T. A., Routray, W., Ngadisih, Karyadi, J. N. W., Suwignyo, B., & Suryatmojo, H. (2021). Food waste treatment methods and its effects on the growth quality of plants: A review. *Pertanika Journal of Tropical Agricultural Science, 45*(1), 75-101. https://doi.org/10.47836/pjtas.45.105

Aylaj, M., & Adani, F. (2023). The effect of compost type, application rate, and maturity on nutrient leaching and soil nutrient levels. *Communications in Soil Science and Plant Analysis, 55*(6), 759-781. https://doi.org/10.1080/00103624.2023.2276860

Barneto, A. G., Carmona, J. A., Ferrer, J. A. C., & Blanco, M. J. D. (2009). Kinetic study on the thermal degradation of a biomass and its compost: Composting effect on hydrogen production. *Fuel, 89*(2), 462-473. https://doi.org/10.1016/j.fuel.2009.09.024

Beesley, L., Moreno-Jiménez, E., & Gomez-Eyles, J. L. (2010). Effects of biochar and greenwaste compost amendments on mobility, bioavailability and toxicity of inorganic and organic contaminants in a multi-element polluted soil. *Environmental Pollution, 158*(6), 2282-2287. https://doi.org/10.1016/j.envpol.2010.02.003

Blakeney, M. (2019). Food loss and food waste. In *Food loss and food waste* (pp. 8-11). Edward Elgar Publishing. https://doi.org/10.4337/9781788975391

Cervera-Mata, A., Delgado, G., Fernández-Arteaga, A., Fornasier, F., & Mondini, C. (2021). Spent coffee grounds by-products and their influence on soil C-N dynamics. *Journal of Environmental Management, 302*, Article 114075. https://doi.org/10.1016/j.jenvman.2021.114075

Gelsomino, A., Petrovičová, B., & Panuccio, M. R. (2023). Exhausted fire-extinguishing powders: A potential source of mineral nutrients for reuse and valorisation in compost enrichment for soilless cultivation. *Science of the Total Environment, 906*, Article 167633. https://doi.org/10.1016/j.scitotenv.2023.167633

Ibrahim, M. H., Abas, N. A., & Zahra, S. M. (2019). Impact of salinity stress on germination of water spinach (*Ipomoea aquatica*). *Annual Research & Review in Biology, 31*(5), 1-12. https://doi.org/10.9734/arrb/2019/v31i530060

Irvan, N., Husaini, T., Trisakti, B., Batubara, F., & Daimon, H. (2018). Composting of empty fruit bunches in the tower composter: Effect of air intake holes. *IOP Conference Series: Materials Science and Engineering, 309*(1), Article 012066. https://doi.org/10.1088/1757-899X/309/1/012066

Jahan, S., Ujjaman, S., Rahman, S., Sarker, B. C., Hossain, Z., & Kamal, M. (2022). Physicochemical properties and nutrient contents of compost as influenced by organic wastes and methods of composting. *Journal of the Indian Society of Soil Science, 70*(1), 106-112. https://doi.org/10.5958/0974-0228.2022.00006.8

Kabir, M. Y. (2022). Evaluation of nutrient content of composts made from water hyacinth, kitchen waste and manures. *Journal of Wastes and Biomass Management, 4*(2), 96-101. https://doi.org/10.26480/jwbm.02.2022.96.101

Kumar, V., Pathania, N., Sharma, S., & Sharma, R. (2024). Dynamics of plant nutrient signalling through compost. *The Microbe, 2*, Article 100047. https://doi.org/10.1016/j.microb.2024.100047

Kumar, M., Ou, Y., & Lin, J. (2010). Co-composting of green waste and food waste at low C/N ratio. *Waste Management, 30*(4), 602-609. https://doi.org/10.1016/j.wasman.2009.11.023

Lai, J. C., Then, Y. L., Hwang, S. S., & Lee, C. S. (2024). Optimal aeration management strategy for a small-scale food waste composting. *Carbon Resources Conversion, 7*(1), Article 100190. https://doi.org/10.1016/j.crcon.2023.06.002

Lee, D., Wan, C., Leung, T. C. H., Rundle-Thiele, S., & Li, G. (2024). Application of marketing to reduce consumer food waste in restaurants. *European Journal of Marketing, 58*(7), 1776-1792. https://doi.org/10.1108/EJM-06-2023-0447

Meena, A., Karwal, M., Raghavendra, K. J., & Narwal, E. (2021). Aerobic composting versus anaerobic composting: Comparison and differences. *Food and Scientific Reports, 2*(2), 23-26. https://foodandscientificreports.com/assets/uploads/issues/16139950835__aerobic_composting_versus_anaerobic_composting_comparison_and_differences_compressed.pdf

Melini, F., & Melini, V. (2024). Role of microbial fermentation in the bio-production of food aroma compounds from vegetable waste. *Fermentation, 10*(3), Article 132. https://doi.org/10.3390/fermentation10030132

Nakagiri, K., Sato, Y., & Kawakami, T. (2024). Factors associated with the leftover rate of side dishes in Japanese school lunches. *PLOS ONE, 19*(2), Article e0298691. https://doi.org/10.1371/journal.pone.0298691

Nair, H. T., Nath, G. R., & Perumal, S. (2024). Impact of different organic compost on seed germination and seedling growth performance of mungbean (*Vigna radiata*). *The Indian Journal of Agricultural Sciences, 94*(2), 229-232. https://doi.org/10.56093/ijas.v94i2.136493

Ni'am, M. I., & Yuniati, R. (2021). Effect of lead (Pb) on seed germination of water spinach (*Ipomoea aquatica* Forsk). *Journal of Physics: Conference Series, 1725*(1), Article 012041. https://doi.org/10.1088/1742-6596/1725/1/012041

Noor, R. S., Shah, A. N., Tahir, M. B., Umair, M., Nawaz, M., Ali, A., Ercisli, S., Abdelsalam, N. R., Ali, H. M., Yang, S. H., Ullah, S., & Assiri, M. A. (2024). Recent trends and advances in additive-mediated composting technology for agricultural waste resources: A comprehensive review. *ACS Omega, 9*(8), 8632-8653. https://doi.org/10.1021/acsomega.3c06516

Oliveira, L. M., Mendonça, V., Moura, E. A., Irineu, T. H. S., Figueiredo, F. R. A., Melo, M. F., Celedonio, W. F., Rêgo, A. L. B., Mendonça, L. F. M., & Andrade, A. D. M. (2022). Salt stress and organic fertilisation on the growth and biochemical metabolism of *Hylocereus costaricensis* (red pitaya) seedlings. *Brazilian Journal of Biology, 84*, Article e258476. https://doi.org/10.1590/1519-6984.258476

Safdar, H., Jamil, M., Hussain, A., Albalawi, B. F. A., Ditta, A., Dar, A., Aimen, A., Ahmad, H. T., Nazir, Q., & Ahmad, M. (2022). The effect of different carrier materials on the growth and yield of spinach under pot and field experimental conditions. *Sustainability, 14*(19), Article 12255. https://doi.org/10.3390/su141912255

Sarminingsih, A., Sumiyati, S., Syafrudin, N., Andika, A. F., & Balqis, M. (2023). Effect of variation of eco-enzyme addition on macro nutrient content of compost with biopore infiltration holes. *E3S Web of Conferences, 448*, Article 03035. https://doi.org/10.1051/e3sconf/202344803035

Selvam, A., Ilamathi, P. M. K., Udayakumar, M., Murugesan, K., Banu, J. R., Khanna, Y., & Wong, J. W. C. (2021). Food waste properties. In *Current developments in biotechnology and bioengineering* (pp. 11-41). Elsevier. https://doi.org/10.1016/B978-0-12-819148-4.00002-6

Ugak, M. A. M., Yaser, A. Z., Lamaming, J., Subin, E. K., Rajin, M., Saalah, S., Tze, F. W. H., & Abang, S. (2022). Comparative study on passive aerated in-vessel composting of food wastes with the addition of Sabah ragi. *Carbon Resources Conversion, 5*(3), 200–210. https://doi.org/10.1016/j.crcon.2022.05.004

Wahi, R., Bidin, E. R., Asif, N. M. M., Hamizat, N. A. N., Ngaini, Z., Omar, R., & Jamel, J. (2019). Nutrient availability in sago bark and empty fruit bunch composts for the growth of water spinach and green mustard. *Environmental Science and Pollution Research, 26*(22), 22246-22253. https://doi.org/10.1007/s11356-019-05548-6

Wang, Y., Tang, Y., & Yuan, Z. (2022). Improving food waste composting efficiency with mature compost addition. *Bioresource Technology, 349*, Article 126830. https://doi.org/10.1016/j.biortech.2022.126830

Wei, J., Knicker, H., Zhou, Z., Eckhardt, K., Leinweber, P., Wissel, H., Yuan, W., & Brüggemann, N. (2022). Nitrogen immobilisation caused by chemical formation of black- and amide-N in soil. *Geoderma, 429*, Article 116274. https://doi.org/10.1016/j.geoderma.2022.116274

Whittinghill, L., Jackson, C., & Poudel, P. (2024). The effects of annual compost addition to agricultural green roofs on runoff water quality. *HortScience, 59*(3), 307-322. https://doi.org/10.21273/hortsci17556-23

Xiao, Y., Zeng, G., Yang, Z., Shi, W., Huang, C., Fan, C., & Xu, Z. (2009). Continuous thermophilic composting (CTC) for rapid biodegradation and maturation of organic municipal solid waste. *Bioresource Technology, 100*(20), 4807-4813. https://doi.org/10.1016/j.biortech.2009.05.013

Yin, Y., Liu, Y., Meng, S., Kiran, E. U., & Liu, Y. (2016). Enzymatic pretreatment of activated sludge, food waste and their mixture for enhanced bioenergy recovery and waste volume reduction via anaerobic digestion. *Applied Energy, 179*, 1131-1137. https://doi.org/10.1016/j.apenergy.2016.07.083

Pertanika J. Sci. & Technol. 34 (4): 2441 - 2456 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Mechanical Testing Analysis of Hybrid Composites of Oil Palm Empty Fruit Bunch Fibre and Sugarcane Bagasse Fibre for Environmentally Friendly Car Body Materials

Aris Sandi[1], Aulia Majid[2,3]*, and Syafiq Fadhlika[1]

[1]Department of Plantation Industrial Machinery Engineering Technology, Politeknik LPP Yogyakarta, 55222, D.I. Yogyakarta, Indonesia
[2]Manufacturing Engineering, Department of Mechanical Engineering Education, Universitas Negeri Yogyakarta, 55281, D.I. Yogyakarta, Indonesia
[3]Center for Eco-Material Engineering and Manufacturing, Universitas Negeri Yogyakarta, 55281, D.I. Yogyakarta, Indonesia

ABSTRACT

Indonesia generates large quantities of agro-industrial waste, including oil palm empty fruit bunch (OPEFB) and sugarcane bagasse, which have potential as sustainable reinforcing materials for polymer composites. This study investigated the effect of OPEFB–bagasse fibre composition on the mechanical properties of epoxy-based hybrid composites fabricated using the hand lay-up method. Five composite formulations with a constant epoxy content of 50% and varying OPEFB-to-bagasse ratios were prepared and evaluated through tensile, flexural, impact, and scanning electron microscopy (SEM) analyses. The hybrid composite containing 30% OPEFB, 20% bagasse, and 50% epoxy exhibited the best overall mechanical performance, achieving a tensile strength of 16.78 MPa, a flexural strength of 59.95 MPa, an impact toughness of 3.08 kJ/m², and a Young's modulus of 8.80 GPa. SEM observations revealed good fibre–matrix adhesion and fibre pull-out mechanisms that contributed to improved energy absorption. Statistical analysis showed that fibre composition significantly affected flexural strength ($p = 0.047$), whereas its effects on tensile strength and impact toughness were not statistically significant ($p > 0.05$). The developed hybrid composite exhibited mechanical properties comparable to several previously reported natural fibre-reinforced epoxy composites, indicating its potential for lightweight non-structural automotive body applications.

Keywords: Automotive materials, empty oil palm bunches, hybrid composite, mechanical properties, sugarcane bagasse

ARTICLE INFO

Article history:
Received: 08 April 2026
Accepted: 19 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.24

E-mail addresses:
aris@polteklpp.ac.id (Aris Sandi)
auliamajid@uny.ac.id (Aulia Majid)
syafiqfadhlika667@gmail.com (Syafiq Fadhlika)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The development of natural fibre-based composite materials has accelerated over the last decade, particularly for automotive applications (Borůvka & Lenfeld, 2018; Skosana et al., 2025). Natural fibre composite materials offer the advantages of being lightweight, environmentally friendly, and having lower production costs compared to synthetic fibres such as carbon and glass (Joshi et al., 2004). Sugarcane bagasse and empty oil palm fruit bunches (*Elaeis guineensis*) (EOPFB) are examples of natural materials that can be processed into composites. In addition to the advantages mentioned above, these materials are abundantly available in Indonesia. According to a 2023 report from the Ministry of Agriculture, Indonesia produces approximately 3.5 million tons per year, with approximately 2.8 million tons of bagasse remaining annually, while oil palm empty fruit bunch (OPEFB) waste reaches 25 million tons per year, opening opportunities to utilise both types of waste as composite raw materials. Utilising this waste as a composite reinforcement can reduce environmental pollution while increasing economic value (Nurfajriani et al., 2022).

Fibres derived from sugarcane bagasse contain between 40-50% cellulose (Mahmud & Anannya, 2021), with a tensile modulus of 2.7 GPa (Devadiga et al., 2020). These characteristics make them a potential and effective composite reinforcement. Meanwhile, OPEFB fibre has a high length-to-diameter ratio (100-150), which improves load distribution in the resin matrix (Rama Rao & Ramakrishna, 2022). The combination of these two fibres is expected to produce a synergistic effect in increasing tensile strength and impact resistance, which are critical for alternative automotive body applications. Epoxy resin was chosen as the matrix in the composite manufacturing because it has a relatively high tensile strength, which is between 35-50 MPa and good chemical resistance. These properties comply with the ASTM D638 standard for automotive applications, making epoxy resin well-suited to improving car performance. Although the potential of natural fibres has been recognised, the main challenge still lies in the low fibre-matrix interfacial compatibility due to the hydrophilic nature of natural fibres and the hydrophobicity of thermoset resins (Behera et al., 2023). Previous research has shown that lignin removal with alkali (5-10% NaOH) and increasing the effective surface area can increase the tensile strength of composites by up to 30% (Holanda et al., 2024). However, further study is needed to determine the optimal fibre percentage, particularly for automotive component applications. Therefore, this study will analyse the effect of fibre percentage (30-70% by weight).

From an environmental perspective, the use of natural fibres can reduce CO_2 emissions by up to 70% compared to synthetic fibres. This is because the natural fibre production process does not require high energy consumption. Life Cycle Assessments (LCAs) show that the production process for natural fibre composites can produce up to 50% lower

carbon emissions (Ramachandran et al., 2023). However, their application in the automotive industry remains limited due to a lack of empirical data and standard testing methods specific to natural fibre composites. This research aims to address this gap by generating mechanical property data and technical recommendations for automotive body materials.

Various recent studies have shown the great potential of bagasse fibre and other agro-industrial waste as reinforcing materials for polymer-based composites (Al Mahmud et al., 2025). studied a hybrid composite of areca nut fibre and bagasse in an epoxy matrix, with a fibre content of 23 wt%, which produced low thermal conductivity (0.163 W/mK) and a tensile strength of 30 MPa, thus having the potential as an environmentally friendly lightweight insulation material. The effect of varying the length of bagasse fibres in unsaturated polyester composites, with longer fibres increasing tensile strength by up to 32.11 MPa and providing a more uniform stress distribution. Furthermore, a hybrid composite based on bagasse, rice husk, and sawdust was studied (Velmurugan et al., 2023), with tensile strength ranging from 16–20 MPa, confirming the potential of multi-agricultural waste for low-cost composites.

Recent research also highlights sustainability aspects and the use of artificial intelligence (AI) technology. One aspect of sustainability is the use of natural fibres in composite materials. Samanth et al. (2025) developed a hybrid composite of bagasse with bio-epoxy, showing a tensile strength of ~62 MPa, a flexural stiffness of ~92 MPa, and a Tg of 92 °C, plus a 45% reduction in carbon footprint due to the use of bio-epoxy. Faruqui et al. (2024) compared alkali-treated and untreated bagasse powder in an epoxy composite, where 20 wt% treated powder resulted in significant improvements in tensile strength, impact strength, and toughness, strengthening the application prospects in the automotive and construction industries (Faruqui et al., 2024).

A comprehensive study was also conducted by Mahmud et al. (2025) on various epoxy-based natural fibre composites (Al Mahmud et al., 2025). This study emphasised that the fibres jute, flax, hemp, sisal, and pineapple nanofibers can increase tensile strength up to 1627 MPa, with various applications in the automotive industry, aerospace, furniture, and construction.

Finally, Aisyah et al. (2024) reviewed hybrid composites of palm oil fibre combined with various other natural fibres. The study results showed significant improvements in mechanical, thermal, and environmental resistance, although they were still hampered by high humidity and fibre-matrix bonding.

Several researchers have investigated natural fibre composites for automotive applications. However, most have focused on the use of single fibres (monofibers) such as hemp or kenaf (Maiti et al., 2022). Research on hybrid composites based on bagasse fibre and empty oil palm fruit bunches remains very limited, particularly regarding the use of epoxy resins, which require complex analysis of intercomponent interactions.

The novelty of the present study lies not only in combining two abundant agro-industrial wastes, namely OPEFB fibre and sugarcane bagasse fibre, but also in systematically investigating the effects of different OPEFB-to-bagasse volume ratios at a constant epoxy matrix content. Unlike previous studies that primarily focused on single natural fibres or hybrid composites for thermal insulation and general engineering applications, this work integrates comprehensive tensile, flexural, impact, and SEM fractography analyses to identify the optimum hybrid composition for lightweight automotive body materials. The findings provide new experimental evidence regarding the synergistic interaction between OPEFB and bagasse fibres and their influence on load transfer, fracture behaviour, and energy absorption, thereby contributing to the development of sustainable natural fibre composites for automotive applications.

Research Gap

From the results of the literature study that has been carried out (Table 1), the following research comparison matrix was obtained:

1. Although hybrid natural fibre composites have been widely investigated, studies specifically examining epoxy-based hybrid composites reinforced with both OPEFB fibre and sugarcane bagasse fibre remain limited.

Table 1
Research gap

No.	Author	Material	Method	Matrix	Variation
1	Mahmud et al., 2025	Areca Nut Fibre and Sugarcane Bagasse	*Vacuum Bag*	Epoxy	NaOH soaking *Purity* 99,9%
2	Velmurugan et al., 2023	Sugarcane Bagasse, Rice Husk and Sawdust	*Hand Lay-Up*	Epoxy	Volume Variation
3	Samanth et al., 2025	Sugarcane Bagasse	*Hand Lay-Up*	Bio-Epoxy and Epoxy	Uses of Bio-Epoxy and Epoxy
4	Faruqui et al., 2024	Sugarcane Bagasse	*Vacuum Bag*	Epoxy	Alkali Treatment
5	Al Mahmud et al., 2025	Sisal fibre, Hamp and pineapple nanofiber	*Vacuum Bag and Hand Lay-Up*	Epoxy	Material Variations
6	Aisyah et al., 2024	Palm Oil Fibre (EFB, OPT, OPF, MF, PKS)	*Vacuum Bag and Hand Lay-Up*	Epoxy and Polyester	Material Variations
7	Present study	Oil Palm Empty Fruit Bunch Fibre and Sugarcane Bagasse Fibre	*Hand Lay-Up*	Epoxy	Volume Variation

2. Limited studies have comprehensively evaluated the mechanical performance, particularly tensile and impact properties, of epoxy-based hybrid natural fibre composites for lightweight and energy-efficient vehicle body applications.
3. Existing studies have provided limited systematic evaluation of the effects of varying OPEFB-to-bagasse fibre ratios in epoxy-based hybrid composites. Comparative investigations of different fibre compositions under a constant epoxy matrix content remain relatively scarce.
4. Although natural fibre-reinforced composites have shown considerable potential for lightweight automotive applications, further experimental evidence is needed to assess the feasibility of epoxy-based OPEFB–bagasse hybrid composites in achieving the mechanical performance required for vehicle body components while supporting lightweight design strategies that may contribute to lower CO_2 emissions.

MATERIALS AND METHODS

Figure 1 shows the flowchart for this study. Oil palm empty fruit bunch (OPEFB) fibres and sugarcane bagasse fibres were used as reinforcing materials, while Eposchon epoxy resin served as the matrix. The fibres were cleaned, cut into a uniform length of 30 mm, and oven-dried at 80°C for 24 h until the moisture content was below 5%, as determined according to ASTM D4442. Untreated fibres were intentionally employed to evaluate the intrinsic reinforcing capability of agricultural waste fibres while minimising processing cost and environmental impact, thereby assessing their feasibility as sustainable reinforcement for low-cost automotive applications.

Five composite formulations were prepared using the hand lay-up technique: S50 (50% EOPFB/50% epoxy), T50 (50% bagasse-50% epoxy), S20 (20% EOPFB-30% bagasse-50% epoxy), S25 (25% EOPFB-25% bagasse-50% epoxy), and S30 (30% EOPFB-20% bagasse-50%

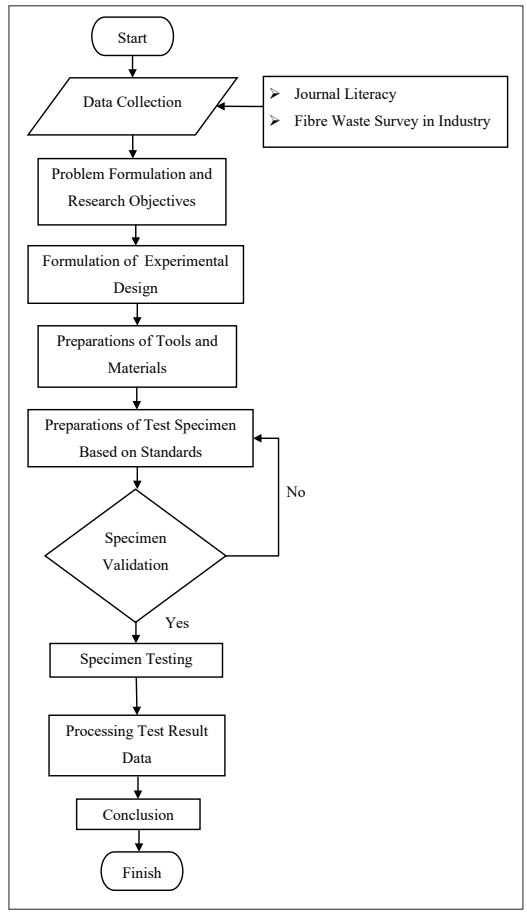


Figure 1. Research flow diagram

epoxy). The fibres were randomly oriented within the epoxy matrix to produce hybrid composite laminates.

Before fabrication, the mould surface was coated with a releasing agent to facilitate specimen removal. After the fibre–resin mixture was placed into the Mold, the laminate was compressed using a steel plate under an applied pressure of approximately 0.5 MPa during curing. The composites were initially cured at room temperature (27 ± 2°C) for 24 h, followed by post-curing at 80°C for 2 h to ensure complete polymerisation. The cured laminates were machined into test specimens according to the dimensions specified by the respective mechanical testing standards.

Tensile, flexural, and impact tests were performed in accordance with ASTM D638, ASTM D7264, and ISO 179, respectively. Five specimens were tested for each composition, and the results are presented as mean ± standard deviation (SD). The coefficient of variation (CV) was calculated to evaluate the repeatability of the fabrication process. One-way analysis of variance (ANOVA) followed by Tukey's post hoc test was used to determine statistically significant differences among the composite formulations. Statistical significance was accepted at $p < 0.05$. Fracture surfaces of the mechanically tested specimens were analysed using scanning electron microscopy (SEM) to evaluate fracture morphology, fibre pull-out behaviour, and fibre–matrix interfacial bonding.

Preparation of Materials

Sugarcane

Sugarcane fibre is obtained from the waste product of sugarcane processing in factories, then washed and dried manually. The fibres, initially measuring 70-150 cm, are cut to a uniform length of 30 mm for use in research is shown in Figure 2.

Palm Oil

Empty oil palm fruit bunch fibre is used as a hybrid material, mixed with bagasse fibre (as shown in Figure 3) to create a composite. The empty oil palm bunches are separated from their thorny stalks and then dried in the sun for 1-3 days. The fibre is then combed with a steel comb to aid decomposition and refined to 30 mm, as required by the research.

Epoxy

The present study used Eposchon Epoxy. The epoxy matrix consisted of epoxy resin and polyamine hardener mixed at a weight ratio of 100:20 according to the manufacturer's recommendation. The mixture was stirred manually for approximately 5 min until homogeneous before impregnation.



Figure 2. Sugarcane bagasse fibre



Figure 3. Empty fruit bunch fibre

Testing

Tensile Test

Tensile testing aims to determine a material's ability to withstand a given tensile force. Tensile testing is destructive because the composite specimen is subjected to continuous tensile force until it elongates and eventually breaks. The tensile test specimen is shown in Figure 4, while tensile specimen printing can be seen in Figure 5.

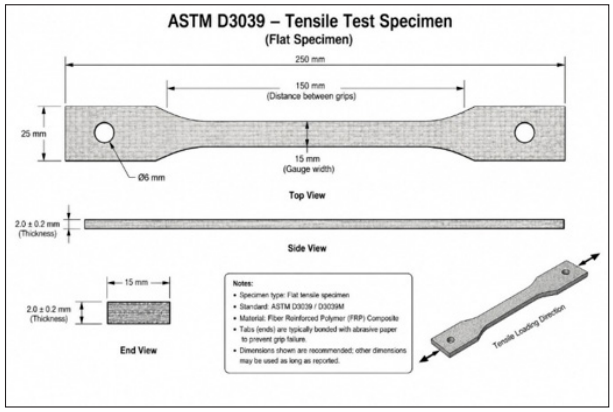


Figure 4. ASTM D3039 – tensile test specimen



Figure 5. Tensile specimen

Flexural (Bending) Test

A bending test aims to determine the strength, durability, and flexibility of a composite specimen under load. Bending testing uses the 3-point method and tests at a speed of 2

mm/min. The flexural test specimen is shown in Figure 6, and the printing specimen is shown in Figure 7.

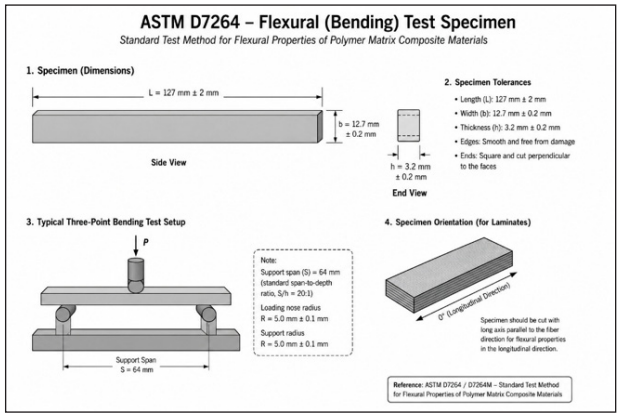


Figure 6. ASTM – flexural (bending) test specimen



*Figure 7.*Flexural (bending) specimen

Impact Test

The Charpy impact test was selected because it is widely used to evaluate the impact toughness of fibre-reinforced polymer composites under dynamic loading conditions. This method provides quantitative information on the composite's energy absorption capacity before fracture, an important performance parameter for lightweight automotive body materials subjected to accidental impacts. Furthermore, the Charpy test enables correlation between impact energy absorption and fracture mechanisms observed through SEM fractography, such as fibre pull-out, matrix cracking, and interfacial debonding. The speed applied for the testing was 2 mm/min. The Charpy impact test specimen (unnotched) is shown in Figure 8, and the printing specimen is shown in Figure 9.

SEM Test

Scanning Electron Microscopy (SEM) is a material characterisation method that uses an electron beam to produce images of a material's surface at very high magnifications. SEM testing is used to observe the morphology, topography, and microstructure of a material that cannot be observed with a conventional optical microscope.

RESULTS AND DISCUSSIONS

Tensile Strength Test Result

Based on the tensile strength test results in Table 2 and Figure 10, the specimen with the S50 code has an average maximum stress of 12.64 MPa. The specimen with the T50 code

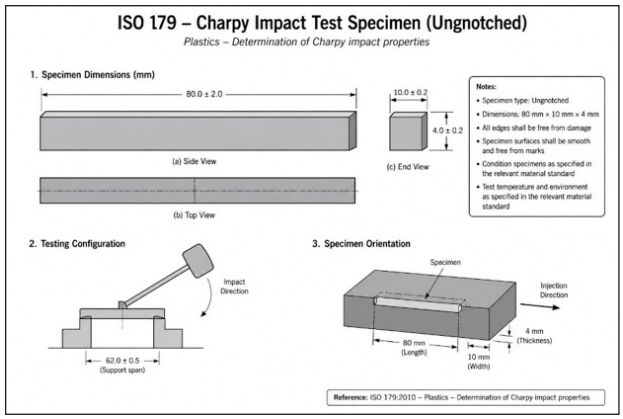




Figure 8. ISO 179 – Charpy impact test specimen (unnotched)

Figure 9. Impact specimen

Table 2
Test result

Tensile Strength (MPa)					
Statistical Parameter	**S50**	**T50**	**S30**	**S25**	**S20**
Average	12.50	11.96	16.78	16.42	15.48
St Dev	4.42	3.86	1.46	4.08	1.47
CV	35%	32%	9%	25%	9%
Flexural Strength (MPa)					
Statistical Parameter	**S50**	**T50**	**S30**	**S25**	**S20**
Average	49.00	51.04	59.95	39.07	48.31
St Dev	8.36	4.95	10.82	13.92	8.34
CV	17%	10%	18%	36%	17%
Impact Toughness (kJ/m²)					
Statistical Parameter	**S50**	**T50**	**S30**	**S25**	**S20**
Average	2.84	1.75	3.19	2.72	2.44
St Dev	1.43	0.52	1.51	0.35	0.32
CV	50%	30%	47%	13%	13%

has an average maximum stress of 11.96 MPa, while the specimen with the S30 code has an average maximum stress of 16.78 MPa. The specimen with the S25 code has an average maximum stress of 16.42 N/mm², and the specimen with the S20 code has an average maximum stress of 15.48 N/mm². It can be concluded that the highest average maximum stress value is found in the specimen with code S30, and the lowest one is found in the specimen with code T50.

The S30 hybrid composite exhibited the highest tensile strength, indicating that the selected OPEFB-to-bagasse fibre ratio provided the most efficient stress transfer from the

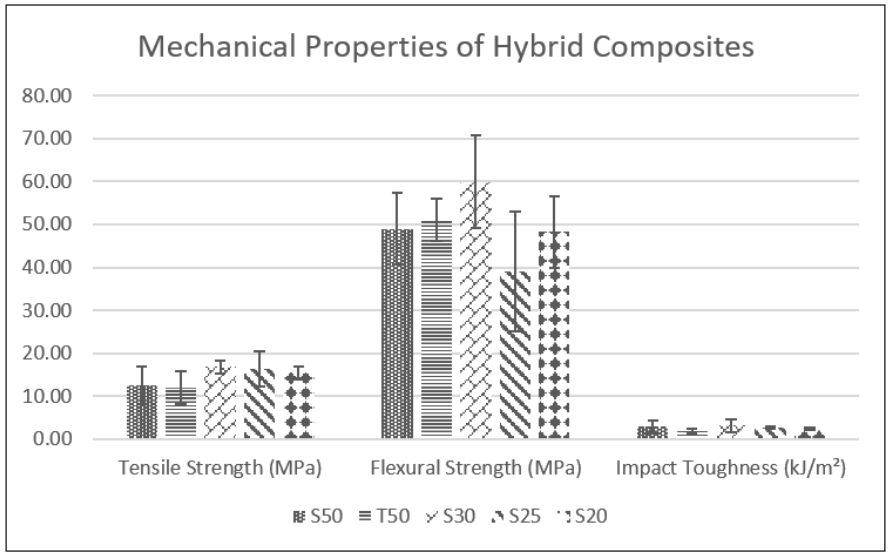


Figure 10. Mechanical properties of hybrid composites

epoxy matrix to the reinforcing fibres. Under tensile loading, effective interfacial adhesion allows the applied load to be transferred from the relatively ductile epoxy matrix to the stiffer natural fibres, thereby delaying crack initiation and propagation. The balanced proportion of OPEFB and bagasse fibres in the S30 formulation likely promoted a more homogeneous fibre distribution and reduced stress concentration sites compared with the other formulations. Consequently, the composite sustained higher tensile loads before failure.

Although the S25 composite contained a comparable amount of reinforcement, its lower tensile strength may be attributed to less uniform fibre distribution and localised resin- or fibre-rich regions. Such heterogeneity can reduce the efficiency of load transfer and promote premature interfacial debonding during tensile loading.

One-way ANOVA indicated that the effect of fibre volume fraction on tensile strength was not statistically significant ($F = 2.28$, $p = 0.096$). Although the S30 formulation exhibited the highest average tensile strength, the observed differences among formulations may be attributed to specimen variability and the inherent heterogeneity of natural fibre composites.

Flexural Strength Test Result

Based on the bending strength test results below, the specimen with code S50 has an average maximum load of 43,394 N, and the specimen with code T50 has an average maximum load of 45,292 N. Furthermore, the specimen with code S30 has an average maximum load of 53,202 N, while the specimen with code S25 produces an average maximum load

of 58,625 N. The specimen with code S20 has an average maximum load of 42,867 N and an average maximum stress of 48,307 MPa. It can be concluded that the highest average flexural strength is observed in specimen S25, and the lowest in specimen S20.

During flexural loading, the upper surface of the specimen experiences compressive stress, whereas the lower surface is subjected to tensile stress. Therefore, the quality of fibre–matrix bonding becomes particularly important in resisting crack initiation on the tensile side of the specimen. The superior flexural performance of the S30 composite suggests stronger interfacial bonding, allowing the applied bending load to be more effectively distributed throughout the hybrid reinforcement.

One-way ANOVA revealed a statistically significant effect of fibre volume fraction on flexural strength ($F = 2.92$, $p = 0.047$), indicating that the hybrid fibre composition significantly influenced the bending performance of the composites. This result suggests that optimising the OPEFB-to-bagasse fibre ratio enhances stress distribution and load-bearing capability under flexural loading.

Impact Test Result

Based on the impact test results below, the S50 specimen has an average maximum impact toughness of 2,840 kJ/m². While the T50 specimen test results have an average maximum impact toughness of 1,746 kJ/m². The S30 specimen test results show an average maximum impact toughness of 3,084 kJ/m². Then, the S25 specimen test results produce an average maximum impact toughness of 2,716 kJ/m². While the S20 specimen has an average maximum impact toughness of 2,442 kJ/m². It can be concluded that the highest average maximum impact toughness is observed in S30.

Unlike tensile loading, impact loading involves rapid energy absorption within a very short time. Therefore, the impact toughness of hybrid composites depends not only on fibre strength but also on the ability of the fibre–matrix interface to dissipate fracture energy through mechanisms such as fibre pull-out, matrix deformation, crack deflection, and interfacial debonding. The higher impact toughness observed for the S30 composite indicates that these energy absorption mechanisms were more effectively activated than in the other formulations.

The ANOVA results showed that the differences in impact toughness among the composite formulations were not statistically significant ($F = 1.52$, $p = 0.233$). Although S30 exhibited the highest average impact toughness, the relatively large variability among specimens reduced the statistical evidence for differences between formulations.

Young's Modulus Results

The Young's modulus of the hybrid OPEFB–bagasse epoxy composites as shown in Table 3 ranged from 7.55 to 9.95 GPa, with the highest value obtained by the S20 composite,

followed by S25 (9.70 GPa) and S30 (8.80 GPa). The increase in Young's modulus indicates that the hybridisation of OPEFB and bagasse fibres improved the stiffness of the composite through more efficient stress transfer and stronger fibre–matrix interfacial bonding. Conversely, the lower S50 modulus may be attributed to less effective resin impregnation and localised defects associated with the higher OPEFB fibre content. The obtained modulus values are comparable to those reported for natural fibre-reinforced epoxy composites, while the low coefficient of variation (7-23%) indicates good measurement repeatability and composite consistency.

Table 3
Young's modulus calculation

Young's Modulus (GPa)					
Number	**Code**				
	S50	T50	S30	S25	S20
1	7.36	8.05	9.62	9.42	10.73
2	9.51	7.23	9.17	10.60	10.27
3	5.03	8.05	9.29	9.24	9.09
4	6.97	6.85	7.30	8.99	9.41
5	8.90	8.90	8.63	10.25	10.27
Total	**37.77**	**39.08**	**44.00**	**48.49**	**49.75**
Average	**7.55**	**7.82**	**8.80**	**9.70**	**9.95**
St Dev	**1.76**	**0.80**	0.91	0.69	0.68
CV	23%	10%	10%	7%	7%

SEM Results

SEM observations of the fracture surface of the S30 specimen revealed a rough and irregular morphology. This indicates a complex fracture mechanism involving matrix cracking, fibre pullout, and fibre-matrix debonding. Several voids observed on the fracture surface were related to fibre pullout, indicating that a significant amount of tensile energy was lost during fibre extraction from the epoxy matrix. In addition, residual epoxy on the fibre surface indicated adequate interfacial bonding between the fibre and the matrix, thereby promoting effective stress transfer. Although a small number of voids were observed, they were not sufficient to significantly reduce mechanical performance. The relatively homogeneous fibre distribution and balanced fibre-matrix composition contributed to enhanced energy absorption and resistance to crack propagation, consistent with the superior tensile strength (16.78 MPa) and impact toughness (3,084 kJ/m^2) of the S30 specimen.

Comparison to the Previous Studies

As shown in Table 4, the mechanical properties obtained in this work are comparable to those reported for natural fibre composites intended for lightweight automotive body panels. Overall, the developed hybrid composite exhibited mechanical properties comparable to or superior to those of several natural fibre-reinforced epoxy composites proposed for automotive applications.

For tensile strength, the present study achieved values ranging from 11.96 to 16.78 MPa, which are comparable to those reported by Marichelvam et al. (2021) (6.45-19.80 MPa) but higher than those reported by Dattatreya et al. (2023) (4.76-9.18 MPa). However,

Table 4
Comparison to the previous studies

Mechanical Property	Present Study	Ref A (Marichelvam et al., 2021)	Ref B (Dattatreya et al., 2023)	Ref C (Venkatesh et al., 2023)
Tensile Strength (MPa)	11.96-16.78	6.45-19.8	4.76-9.18	43.39-48.23
Flexural Strength (MPa)	39.07-59.95	8.99-28.79	1.52-9.6	59.58-67.48
Impact Toughness (kJ/m²)	1.75-3.19	1.35-2	1.2-2	11.91-15.32

the tensile strength remained lower than that reported by Venkatesh et al. (2023) (43.39-48.23 MPa). The superior tensile performance reported is likely associated with differences in fibre type, fibre architecture, interfacial bonding, and fabrication process, all of which strongly influence stress transfer from the epoxy matrix to the reinforcing fibres (Venkatesh et al., 2023). Improved fibre alignment and stronger interfacial adhesion generally enhance tensile strength by promoting more efficient load transfer.

In terms of flexural strength, the present study produced values of 39.07-59.95 MPa, exceeding those reported by Marichelvam et al. (2021) (8.99-28.79 MPa) and Dattatreya et al. (2023) (1.52-9.60 MPa), while approaching the performance reported by Venkatesh et al. (2023) (59.58-67.48 MPa). The higher flexural strength indicates that the selected hybridisation ratio between OPEFB and bagasse fibres provided effective resistance against bending loads. Under flexural loading, compressive and tensile stresses act simultaneously; therefore, uniform fibre distribution and adequate resin impregnation significantly delay crack initiation and propagation.

Regarding impact toughness, the present study achieved 1.75-3.19 kJ/m², which is higher than those reported by Marichelvam et al. (2021) (1.35-2.00 kJ/m²) and Dattatreya et al. (2023) (1.20-2.00 kJ/m²). This suggests that the hybridisation of OPEFB and bagasse fibres enhanced the composite's energy absorption capability. The presence of two distinct natural fibres may promote multiple energy-dissipation mechanisms, including fibre pull-out (as shown in Figure 11), crack deflection, matrix deformation, and progressive fibre fracture, thereby increasing resistance to sudden impact loading. These mechanisms are commonly observed in hybrid natural fibre composites and contribute to improved impact performance.

Overall, the comparison demonstrates that the proposed hybrid composite offers a balanced combination of tensile, flexural, and impact properties as shown in Table 4. Although its tensile strength is lower than that of the composite developed by Venkatesh et al. (2023), the present material exhibits superior flexural strength and impact toughness compared with several previously reported natural fibre composites. These findings indicate that the developed hybrid composite has promising potential for lightweight, non-structural automotive body components where bending resistance and impact energy absorption are critical performance requirements.

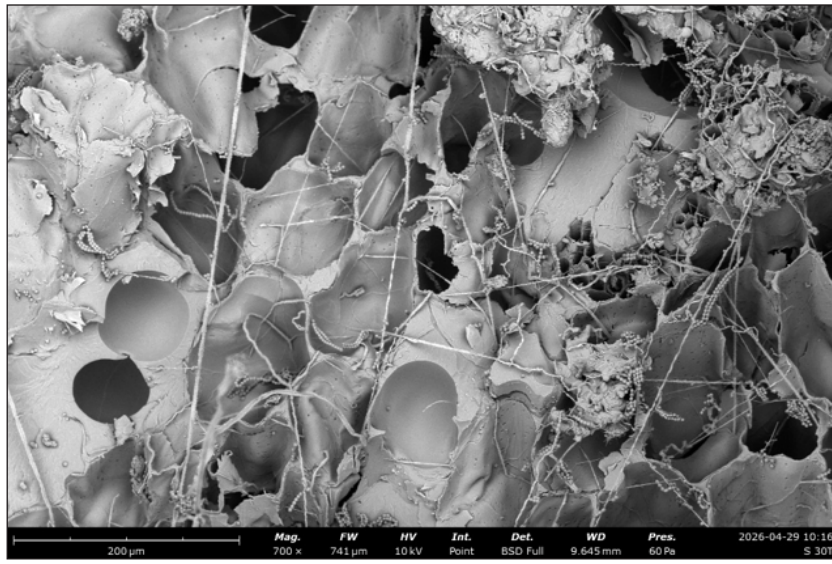


Figure 11. SEM micrograph of the S30 fracture surface

CONCLUSION

The hybrid composite containing 30% OPEFB, 20% bagasse, and 50% epoxy exhibited the best overall mechanical performance, achieving a tensile strength of 16.78 MPa, a flexural strength of 59.95 MPa, an impact toughness of 3.08 kJ/m², and a Young's modulus of 8.80 GPa. SEM analysis revealed good fibre–matrix adhesion and fibre pull-out mechanisms that enhanced energy absorption. Statistical analysis indicated that fibre composition significantly influenced flexural strength ($p = 0.047$), whereas its effects on tensile strength and impact toughness were not statistically significant ($p > 0.05$). Compared with previous studies, the developed hybrid composite demonstrated competitive mechanical properties, particularly in flexural strength and impact toughness, indicating its potential as a sustainable material for lightweight non-structural automotive body applications.

ACKNOWLEDGEMENT

This research was generously supported and funded by the Ministry of Education, Culture, Research, and Technology (Kemdikbudristek). The authors sincerely appreciate their invaluable contribution, which has significantly facilitated the progress and completion of this study.

AUTHOR CONTRIBUTION

AS: Conceptualisation, Methodology, Supervision, Validation, Writing-review & editing, Corresponding author; AM: Methodology, Validation, Formal analysis, Writing -review

& editing; SF: Investigation, Data curation, Software, Formal analysis, Visualisation, Writing-original draft.

REFERENCES

Aisyah, H. A., Hishamuddin, E., Noorshamsiana, A. W., Ibrahim, Z., & Ilyas, R. A. (2024). Oil palm fibre hybrid composites: A recent review. *Journal of Renewable Materials, 12*(10), 1661-1689. https://doi.org/10.32604/jrm.2024.055217

Al Mahmud, M. Z., Rabbi, S. M. F., Islam, M. D., & Hossain, N. (2025). Synthesis and applications of natural fibre-reinforced epoxy composites: A comprehensive review. *SPE Polymers, 6*(1). https://doi.org/10.1002/pls2.10161

Behera, S., Mohanty, J. R., Nath, G., & Mahanta, T. K. (2023). Exploring properties of short randomly oriented rattan fibre reinforced epoxy composite for automotive application. *Journal of Natural Fibers, 20*(1), Article 2163024. https://doi.org/10.1080/15440478.2022.2163024

Borůvka, M., & Lenfeld, P. (2018). Natural fibre for green technology in automotive industry: A brief review. *IOP Conference Series: Materials Science and Engineering, 368*, Article 012012. https://doi.org/10.1088/1757-899X/368/1/012012

Dattatreya, K., Sathees Kumar, S., Prasad, V. V. S. H., & Ranjan Pati, P. (2023). Mechanical properties of waste natural fibres/fillers reinforced epoxy hybrid composites for automotive applications. *Materials Today: Proceedings*, 1-7. https://doi.org/10.1016/j.matpr.2023.02.001

Devadiga, D. G., Bhat, K. S., & Mahesha, G. (2020). Sugarcane bagasse fibre reinforced composites: Recent advances and applications. *Cogent Engineering, 7*(1), Article 1823159. https://doi.org/10.1080/23311916.2020.1823159

Faruqui, A. N., Rahman Tarafdar, M. S., Das Jayanta, A., & Karim Sheikh, M. R. (2024). Fabrication and property analysis of treated and untreated bagasse powder-reinforced epoxy resin composites. *Journal of Polymer Science and Engineering, 7*(2), Article 7061. https://doi.org/10.24294/jpse.v7i2.7061

Holanda, F. S. R., et al. (2024). Influence of sodium hydroxide treatment on *Typha domingensis* fibres for geotextile manufacturing. *ACS Omega, 9*(52), 51040-51051. https://doi.org/10.1021/acsomega.4c05602

Joshi, S. V., Drzal, L. T., Mohanty, A. K., & Arora, S. (2004). Are natural fibre composites environmentally superior to glass fibre reinforced composites? *Composites Part A: Applied Science and Manufacturing, 35*(3), 371-376. https://doi.org/10.1016/j.compositesa.2003.09.016

Mahmud, M. A., & Anannya, F. R. (2021). Sugarcane bagasse: A source of cellulosic fibre for diverse applications. *Heliyon, 7*(8), Article e07771. https://doi.org/10.1016/j.heliyon.2021.e07771

Mahmud, R. U., Islam, M. S., Hossain, S., Nasrin, J., & Khan, A. N. (2025). Agro-waste areca nut husk and bagasse fibre reinforced epoxy-based hybrid composite for thermal insulated false ceiling application. *SPE Polymers, 6*(1). https://doi.org/10.1002/pls2.10166

Maiti, S., Islam, M. R., Uddin, M. A., Afroj, S., Eichhorn, S. J., & Karim, N. (2022). Sustainable fibre-reinforced composites: A review. *Advanced Sustainable Systems, 6*(11), Article 2200258. https://doi.org/10.1002/adsu.202200258

Marichelvam, M. K., et al. (2021). A novel palm sheath and sugarcane bagasse fibre-based hybrid composites for automotive applications: An experimental approach. *Polymer Composites, 42*(1), 512-521. https://doi.org/10.1002/pc.25843

Nurfajriani, Herlinawati, H., Yusuf, M., Siregar, M. I., Yunus, M., & Derlina. (2022). Mechanical properties of polyethylene matrix composites with areca fruit peel powder filler (*Areca catechu*) and sugar cane fibre (*Saccharum officinarum* L.). *Journal of Physics: Conference Series, 2193*(1), Article 012058. https://doi.org/10.1088/1742-6596/2193/1/012058

Rama Rao, P., & Ramakrishna, G. (2022). Oil palm empty fruit bunch fibre: Surface morphology, treatment, and suitability as reinforcement in cement composites-A state-of-the-art review. *Cleaner Materials, 6*, Article 100144. https://doi.org/10.1016/j.clema.2022.100144

Ramachandran, K., Gnanasagaran, C. L., & Vekariya, A. (2023). Life cycle assessment of carbon fibre and bio-fibre composites prepared via vacuum bagging technique. *Journal of Manufacturing Processes, 89*, 124-131. https://doi.org/10.1016/j.jmapro.2023.01.068

Samanth, M., et al. (2025). Sustainable composites from sugarcane bagasse fibres and bio-based epoxy with insights into wear performance, thermal stability, and machine learning predictive modelling. *Journal of Composites Science, 9*(3), Article 124. https://doi.org/10.3390/jcs9030124

Skosana, S. J., Khoathane, C., & Malwela, T. (2025). Driving towards sustainability: A review of natural fibre reinforced polymer composites for eco-friendly automotive lightweighting. *Journal of Reinforced Plastics and Composites, 38*(2), 754–780. https://doi.org/10.1177/0892705724125432-

Velmurugan, G., Karunakaran, P., Sampath, P. S., & Varahamoorthi, R. (2023). Experimental study of sugarcane bagasse fibre with rice husk and wood powder polymer matrix composite. *Journal of Natural Fibers, 20*(2). https://doi.org/10.1080/15440478.2023.2224977

Venkatesh, R., Ballal, S., Krishnan, A. M., Prabagaran, S., Mohankumar, S., & Ramaraj, E. (2023). Effect of fibre layer formation on mechanical and wear properties of natural fibre filled epoxy hybrid composites. *Heliyon, 9*(5), Article e15934. https://doi.org/10.1016/j.heliyon.2023.e15934

Pertanika J. Sci. & Technol. 34 (4): 2457 - 2486 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

AI-Powered Revolution in Concrete Crack Analysis: An In-Depth Review on Detection and Classification Methods

Muhammad Shahrulazmi Hashim[1], Emedya Murniwaty Samsudin[1], and Khalid Isa[2]*

[1]Department of Civil Engineering, Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

ABSTRACT

Concrete crack detection and classification are pivotal for structural health monitoring and civil infrastructure maintenance. However, traditional manual inspections are labour-intensive, inconsistent, and often hazardous in complex environments. Conventional manual inspections are time-consuming and could turn out ineffective, leading to risks in challenging circumstances. To address these limitations, a systematic literature review (SLR) of 68 peer-reviewed studies published between 2021 and 2025 was conducted, examining the evolution of concrete crack detection and classification methods over the past decade. This systematic review organises recent concrete crack detection and classification studies based on the crack analysis pipeline (data acquisition, preprocessing, model development, and performance evaluation) and identifies the key techniques employed at each stage. This review evaluates the accuracy, practical usefulness, limitations, and implementation challenges of recent concrete crack detection and classification methods, as well as their practical problems and ongoing problems, such as limited training datasets, variable environmental conditions, real-time deployment constraints, and limited model interpretability. To fill these gaps, the review suggests future research directions like creating standard crack datasets and making lightweight, easy-to-understand AI models that can be directly implemented in field applications. The review shows a clear shift from traditional image processing and classical machine learning methods to more advanced deep-learning models, which have improved automated crack

ARTICLE INFO

Article history:
Received: 27 April 2026
Accepted: 03 July 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.25

E-mail addresses:
HF240019@student.uthm.edu.my (Muhammad Shahrulazmi Hashim)
emedya@uthm.edu.my (Emedya Murniwaty Samsudin)
halid@uthm.edu.my (Khalid Isa)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

detection capability. Convolutional neural networks and transformer-based architectures have reported strong performance in selected studies, although these values should be interpreted within the context of different datasets, validation protocols, and evaluation metrics. This greatly improves the safety and efficiency of inspections. Consequently, deep learning has emerged as one of the most widely adopted and promising approaches for concrete crack detection and classification during the reviewed period.

Keywords: Concrete crack, crack classification, crack detection

INTRODUCTION

Concrete cracks in civil engineering compromise the stability and serviceability of buildings, bridges, and road infrastructure. Structural cracks from excessive loading or foundation settlement may indicate potential structural failure. Even small cracks can allow moisture to penetrate the concrete and reach the steel reinforcement, which speeds up corrosion and degradation (Cui & Qin, 2024; Yu et al., 2021). If left untreated, repetitive stress can increase crack propagation, reduce load-bearing capacity, and create safety risks (Golding et al., 2022). For structural health monitoring, it is important to detect and repair cracks at an early stage (Ali et al., 2021).

Concrete may crack due to temperature variation, excessive loading, material defects, or construction-related deficiencies (Cui & Qin, 2024). Cracks are commonly categorised based on origin, appearance, age, orientation, and structural significance. For example, they can be structural or non-structural, such as shrinkage cracks, early-age plastic or later-age hardened cracks, and vertical or diagonal cracks. Crack classifications can indicate possible structural problems (Cui & Qin, 2024; Golding et al., 2022). Industry standards provide acceptable crack-width limits for durability assessment. Cracks that are less than 0.2 mm wide are generally considered less critical, but cracks that are more than 0.5 mm wide may indicate reduced structural or material performance (Nyathi et al., 2024). To prevent further deterioration, cracks that exceed these limits usually require repair or remedial intervention (Yu et al., 2021).

Although maintenance guidelines provide acceptable crack-width limits, they do not prevent the initial development of cracks. Manual visual inspection remains time-consuming, labour-intensive, and susceptible to subjective judgement (Ali et al., 2021; Golding et al., 2022). In large structures such as bridges and building systems, inspection may require several weeks, delay repair decisions, and expose inspectors to safety risks (Golding et al., 2022). In addition, fatigue, limited accessibility, and environmental conditions may cause small or hidden cracks to be overlooked. These limitations highlight the need for automated crack detection systems that are more objective, scalable, and reliable (Cui & Qin, 2024; Wang et al., 2024).

The 1990s and 2000s witnessed the development of semi-automated crack detection and identification systems through the combination of inexpensive digital cameras and basic computer vision technology. Engineers employed edge detection and thresholding filters to display cracks and shadowed regions in concrete surface photographs (Ali et al., 2021). The initial image-based methods provided rapid results and improved inspection consistency compared with manual surveying. Basic filtering and morphological procedures can detect cracks with 70-80% accuracy under optimal illumination and surface conditions (Ali et al., 2021). However, these values should be interpreted within the context of controlled image conditions, because early image processing methods were highly sensitive to lighting, stains, shadows, and surface texture.

The performance of basic image analysis methods declined under challenging environmental conditions. The algorithms often failed to recognise irregular surfaces, where shadows, stains, or rough textures caused false positive detections or missed cracks on complex backgrounds. The software required separate algorithm adjustments because it lacked adaptive intelligence for different field conditions. Basic image processing could assist human inspectors, but it was unreliable for independent use by the late 2000s. These limitations encouraged the development of artificial intelligence systems dedicated to more robust crack detection.

Traditional machine learning models detected cracks with 75-85% accuracy, outperforming earlier image processing methods in several reported studies (Fan, 2021; Yu et al., 2021). However, these performance values should not be interpreted as direct evidence of superiority because studies used different datasets, validation protocols, and testing conditions. These models had several drawbacks, such as dependence on hand-crafted features. If these features could not accurately represent a crack under different lighting or material conditions, model performance declined. Most studies trained models using limited image datasets, which restricted generalisation (Yu et al., 2021). Sliding window scanning was also impractical for inspecting large surface areas because each image required considerable processing time. By the end of the 2010s, conventional machine learning had become too limited in terms of scalability and robustness. Consequently, deep learning approaches were increasingly adopted for crack analysis.

Crack detection and classification are complementary but distinct concrete structure evaluation processes. Crack detection algorithms locate and quantify cracks in images (Ali et al., 2021). Crack classification categorises cracks by type, such as structural or non-structural, and by severity, such as minor or major (Nyathi et al., 2024). In practice, crack detection addresses where the crack is located, while classification addresses the nature and severity of the crack, which is crucial for maintenance planning.

Inspectors or algorithms first identify cracks, then classify each crack according to type and severity (Pu et al., 2022; Taj et al., 2024). Crack classification is important for

maintenance because different crack types require different levels of intervention (Mazni et al., 2024). A minor surface hairline crack may require only periodic monitoring, but a large structural crack may need immediate repair (Ye et al., 2024). Engineers can prioritise repairs based on safety and structural risk by accurately classifying crack severity and type (Golding et al., 2022).

The latest developments enable unified AI models to perform crack detection and classification through integrated systems. Current deep learning algorithms can detect and classify cracks simultaneously (Diniz et al., 2024; Li et al., 2022). The integrated approach allows cracks to be analysed in real time by reducing separate detection and classification procedures (Abubakr et al., 2024; Ye et al., 2024). However, the implementation of integrated techniques still encounters several practical challenges. These include limited training data, model generalisation problems, unpredictable field conditions, and difficulty in interpreting AI outputs (Kolappan Geetha et al., 2023). Researchers are developing solutions to enhance the reliability of AI-based crack analysis systems for practical implementation.

This review contributes by providing a structured synthesis of AI-based concrete crack detection and classification. Beyond summarising previous studies, it proposes a conceptual understanding of the crack analysis workflow, from data acquisition and preprocessing to AI model development, crack detection or classification output, and structural health monitoring decision-making. It also organises the reviewed methods into a taxonomy of AI approaches and identifies maturity levels, research gaps, and future directions. This contribution helps clarify which AI methods are closer to practical field adoption and which areas still require further validation, standardisation, and integration with engineering decision-making.

Detection algorithms require greater computational resources and depend on pixel- or region-based outputs, which can create difficulties for users when compared with classification algorithms. Classification methods usually operate at higher speed and produce binary outputs or category labels. Different applications exist for each approach. Crack detection produces maps, masks, or bounding boxes that enable further examination and measurement. Classification enables rapid evaluation of crack images and provides essential information about crack type or severity. The methods typically work together as a combined system, where detection or segmentation creates a map of the detected crack, while classification provides interpretation of the detected damage.

In summary, crack detection focuses on locating damage through masks, bounding boxes, or segmented crack regions, whereas crack classification assigns crack presence, type, or severity category. Figure 1 shows the main differences between crack detection and crack classification.

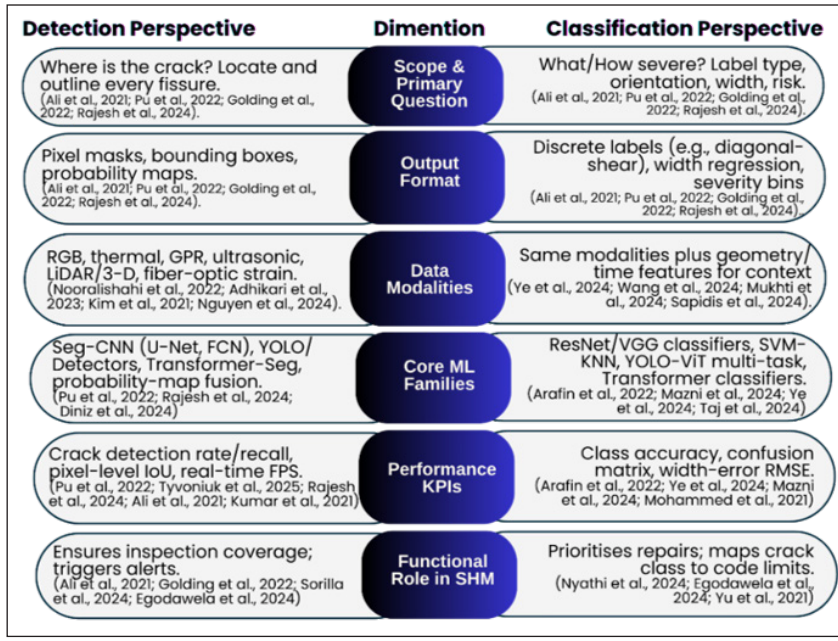


Figure 1. Key differences between crack detection and crack classification

METHODOLOGY

Systematic Literature Review Process

This study adopted a systematic literature review approach guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Kitchenham et al., 2009). The review aimed to identify, screen, evaluate, and synthesise recent studies related to artificial intelligence-based concrete crack detection and classification (Khallaf et al., 2018). The focus was limited to studies published between 2021 and 2025 to ensure that the analysis reflected recent developments in computer vision, machine learning, deep learning, transformer-based models, and structural health monitoring applications.

The Scopus database was selected as the primary source because it provides broad coverage of engineering, computer science, and multidisciplinary peer-reviewed journals. The search was limited to English-language journal articles published between 2021 and 2025. The database search was conducted using the Boolean search string "concrete crack" AND ("crack detection" OR "crack classification") AND ("deep learning" OR "CNN" OR "transformer") AND "structural health monitoring". This search strategy was designed to retrieve studies that applied AI-based techniques to concrete crack detection, crack classification, crack segmentation, or crack severity assessment within structural health monitoring applications. The identified records were then screened using predefined inclusion and exclusion criteria, as shown in Table 1.

Table 1
Inclusion and exclusion criteria for article selection

Criterion	Inclusion criteria	Exclusion Criteria
Publication year	Articles published from 2021 to 2025	Articles published before 2021
Database/source	Scopus-indexed peer-reviewed journal articles	Non-indexed, non-peer-reviewed, or unreliable sources
Publication type	Journal articles	Conference papers, thesis, reports, book chapters, editorials, and review notes
Language	Articles written in English	Non-English articles
Research focus	Studies related to concrete crack detection, crack detection, crack classification, or crack severity assessment	Studies not related to concrete crack analysis, such as cracks or defects in asphalt, steel, soil, masonry, or non-structural materials.
Method used	Studies applying AI, machine learning, deep learning, CNN, transformer, computer vision, UAV, LiDAR, thermal imaging, or acoustic emission methods	Studies based only on manual inspection or studies that recommended AI without applying it
Accessibility	Full-text articles accessible for review	Articles without accessible full text
Data and findings	Studies reporting methodology, dataset, findings, performance metrics, or challenges	Conceptual studies without clear methodology, dataset, or performance evaluation

The screening process followed the PRISMA-guided stages of identification, screening, eligibility assessment, and inclusion. The initial database search produced 1,050 records, which were checked for duplication. After duplicate removal, 870 records were retained for title and abstract screening. At this stage, 730 records were excluded because they were not related to concrete crack detection, did not apply AI-based techniques, or were outside the scope of structural health monitoring. The remaining 140 articles were assessed in full text based on the predefined inclusion and exclusion criteria. Articles were excluded if they lacked sufficient methodological information, did not report relevant performance results, focused on non-concrete materials, or discussed AI only conceptually without actual application. Finally, 68 articles met the criteria and were included in the final analysis.

Two reviewers independently screened the titles, abstracts, and full texts. Any disagreement was resolved through discussion and consensus to minimise selection bias and improve the reliability of the review process. The quality of the included studies was assessed based on relevance to concrete crack analysis, clarity of methodology, dataset description, reported performance metrics, and discussion of limitations or practical implications. Each criterion was rated as fulfilled, partially fulfilled, or not fulfilled, and studies with insufficient methodological detail, unclear dataset information, or no relevant performance reporting were excluded.

Data extraction was conducted using a standardised form covering author, publication year, country, crack analysis category, AI method, dataset characteristics, performance

indicators, major findings, contributions, limitations, and reported challenges. The extracted data were organised thematically according to detection method, classification method, reported impact, and research limitations. Descriptive and narrative synthesis was then used to analyse the distribution of studies by country, summarise crack detection and classification methods, compare reported impacts, and identify major challenges such as dataset limitations, annotation bottlenecks, generalisation issues, real-time deployment constraints, and integration into structural health monitoring systems. The complete selection and refinement process is summarised in Figure 2.

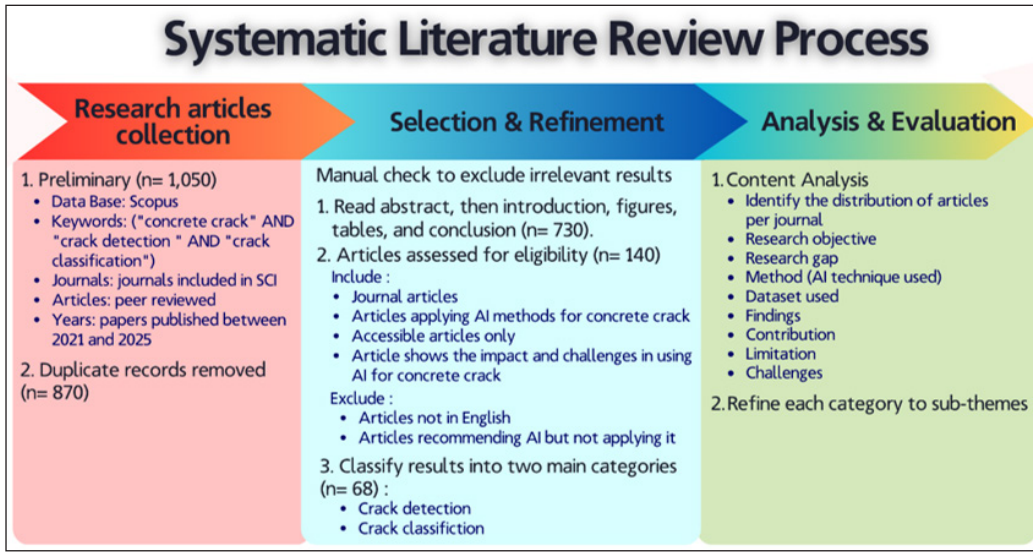


Figure 2. Summary of the systematic literature review process (search strategy, selection criteria, and quality control) used in this research

RESULT AND DISCUSSIONS

Descriptive information

The 68 reviewed articles were sourced from 21 countries, as shown in Figure 3. China accounted for the largest proportion with 22 articles (32.40%), followed by South Korea with 7 (10.29%) and India with 6 articles (8.82%). Australia and the UK both provided 5 articles (7.35%), while the USA offered 4 articles (5.88%), and Canada and Greece each contributed 3 articles (4.41%). The remaining publications were distributed across 13 countries, each providing one article (1.47% per country; 19.1% collectively), signifying a long-tail distribution of contributions beyond the main publishing countries.

Table 2 presents the crack detection and crack classification methods used in the reviewed studies. Overall, the evidence indicates that recent studies are mainly concentrated on computer vision and deep learning methods for concrete crack detection. Computer

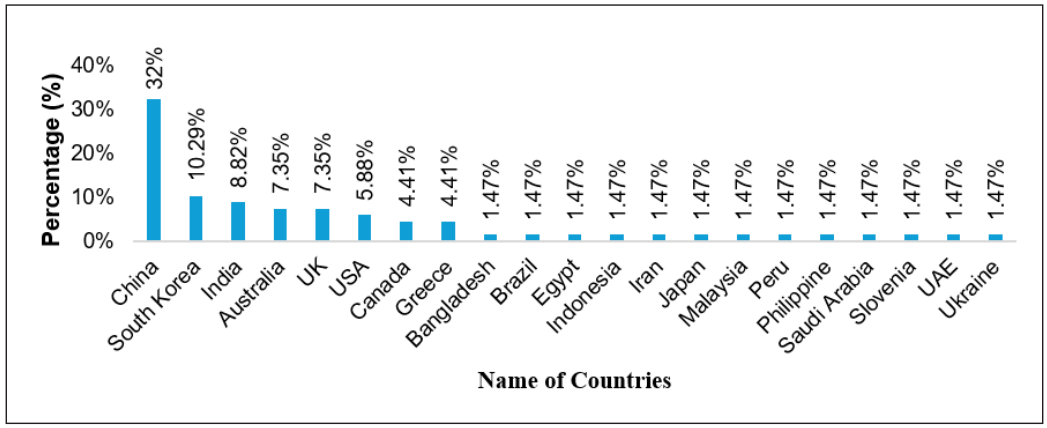


Figure 3. Distribution of research based on origin

vision and image preprocessing are commonly used for image enhancement, crack localisation, and early-stage detection, while deep learning methods are increasingly applied to improve automated feature extraction and detection accuracy (Ali et al., 2021; Pu et al., 2022; Yadav et al., 2024).

However, the distribution of methods also shows an imbalance between crack detection and crack classification. Most studies focus on detecting or segmenting cracks, whereas fewer studies classify crack type, crack severity, or maintenance priority. This suggests that automated crack detection has progressed more strongly than automated crack interpretation and severity assessment (Mazni et al., 2024; Nyathi et al., 2024). Therefore, Table 2 should be interpreted as a distribution of reported methods rather than a direct comparison of method performance.

Following the method distribution in Table 2, the reviewed approaches were further organised into a taxonomy to clarify their main functions, outputs, maturity levels, and limitations. The maturity level was qualitatively interpreted using four indicators: frequency of use, field application, deployment readiness, and main limitations. “High” indicates approaches that are frequently reported and closer to practical use, “Moderate” indicates approaches that are established but still limited, “Emerging to high” indicates approaches with growing potential but requiring further validation, and “Emerging” indicates approaches with limited use or complex validation needs. This taxonomy helps show the original contribution of this review by moving beyond a list of individual studies towards a structured interpretation of how different AI approaches support concrete crack detection, classification, and structural health monitoring.

Table 2
Crack detection and crack classification methods for concrete structures

No.	Author's Year	Country	Crack Detection						Crack Classification					
			CVIP	ML	DL	TI	LiDAR/3D	UAV	AE	CVIP	ML	DL	TI	LiDAR/3D	UAV	AE
1	Shahin et al., 2024	USA	✓		✓											
2	Pu et al., 2022	China	✓		✓											
3	Mazni et al., 2024	Malaysia	✓							✓		✓				
4	Nogueira Diniz et al., 2023	Brazil	✓		✓											
5	Lin et al., 2025	China	✓		✓											
6	Nguyen et al., 2024	Australia	✓		✓											
7	Arafin et al., 2024	Canada	✓		✓					✓		✓				
8	Tyvoniuk et al., 2025	Ukraine	✓	✓												
9	Yadav et al., 2024	India	✓		✓											
10	Adhikari et al., 2023	South Korea					✓									
11	Kang et al., 2022	South Korea	✓		✓											
12	Alkannad et al., 2025	China	✓		✓											
13	Mukhti et al., 2024	USA									✓					✓
14	Sapidis et al., 2024	Greece	✓		✓											
15	Rajesh et al., 2024	India	✓		✓											
16	Gaur et al., 2023	India	✓		✓											
17	Arafin et al., 2022	Canada								✓		✓				
18	Fan et al., 2021	China	✓	✓												
19	Wang, 2023	China	✓		✓											
20	Kun et al., 2022	China	✓		✓											
21	Mohammed et al., 2021	China	✓		✓											
23	Nugraheni et al., 2023	Indonesia	✓		✓					✓		✓				
24	Kumar et al., 2021	India	✓		✓											
25	Hong et al., 2022	South Korea	✓	✓												
26	Kim et al., 2021	South Korea	✓			✓										
27	Rajadurai et al., 2021	South Korea	✓		✓											
28	Mir et al., 2022	Japan	✓	✓												
29	Meng et al., 2023	China	✓		✓											
30	Mayya et al., 2024	Greece								✓		✓				

Table 2 *(Continued)*

No.	Author's Year	Country	Crack Detection						Crack Classification					
			CVIP	ML	DL	TI	LiDAR/3D	UAV	AE	CVIP	ML	DL	TI	LiDAR/3D	UAV	AE
31	Dow et al., 2023	UK	✓													
32	Ali et al., 2021	UAE	✓		✓											
33	Yu et al., 2021	Australia	✓	✓						✓	✓					
34	Golding et al., 2022	Australia	✓		✓											
35	Pennada et al., 2023	UK	✓		✓											
36	Gradišar et al., 2023	Slovenia	✓	✓												
37	Fang et al., 2025	China	✓		✓											
38	Alkayem et al., 2024	Greece	✓		✓											
39	Taj et al., 2024	Saudi Arabia	✓	✓	✓											
40	Nooralishahi et al., 2022	Canada	✓			✓		✓								
41	Cui et al., 2022	China	✓		✓											
42	Gharehbaghi et al., 2022	Australia	✓	✓												
43	Shang et al., 2023	USA	✓		✓											
44	Li et al., 2022	China	✓	✓												
45	Zhou et al., 2024	China			✓				✓							
46	Miele et al., 2025	USA	✓		✓											
47	Li et al., 2022	China	✓		✓					✓		✓				
48	Swarna et al., 2024	Bangladesh	✓		✓											
49	Cui et al., 2024	China	✓													
50	Nyathi et al., 2024	UK	✓		✓											
51	Del Savio et al., 2023	Peru	✓													
52	Wang et al., 2024	China	✓		✓											
53	Chen et al., 2021	China	✓		✓											
54	Li et al., 2024	China					✓									
55	Diniz et al., 2024	UK	✓		✓											
56	Ye et al., 2024	China								✓		✓				
57	Abubakr et al., 2024	Egypt								✓		✓				
58	Lee et al., 2024	South Korea	✓													
59	Ghafari et al., 2024	Iran	✓		✓											
60	Liu et al., 2025	China	✓		✓											
61	He et al., 2024	China	✓													
62	Lin et al., 2025	China	✓	✓												

Table 2 *(Continued)*

No.	Author's Year	Country	Crack Detection							Crack Classification						
			CVIP	ML	DL	TI	LiDAR/3D	UAV	AE	CVIP	ML	DL	TI	LiDAR/3D	UAV	AE
63	Egodawela et al., 2024	Australia	✓		✓			✓		✓		✓			✓	
64	Kolappan Geetha et al., 2023	South Korea	✓							✓		✓				
65	Qiao et al., 2021	China	✓		✓											
66	Philip et al., 2023	India	✓		✓											
67	McAlorum et al., 2023	UK	✓		✓											
68	Naser et al., 2021	India	✓		✓					✓		✓				
		Total	60	10	42	2	2	3	1	12	2	11	-	-	1	1

Note. CVIP= Computer Vision and Image Preprocessing; ML= Machine Learning; DL= Deep Learning; TI= Thermal Imaging; LiDAR/3D Scanning; UAV= Unmanned Aerial Vehicle; AE= Acoustic Emission. Reported methods are presented as a distribution summary and should not be interpreted as direct performance comparisons

As shown in Table 3, computer vision and image preprocessing, machine learning, and deep learning represent more established approaches because they were more frequently reported in Table 2, especially for crack detection. In contrast, thermal imaging, LiDAR/3D scanning, unmanned aerial vehicle-based inspection, and acoustic emission remain less mature or emerging because their use was reported less frequently and still involves field, equipment, or data-processing limitations. Therefore, the maturity classification suggests that future AI-based crack analysis should focus on field reliability, severity interpretation, and integration with structural health monitoring decisions.

Table 3.
Taxonomy and maturity of AI approaches in concrete crack analysis

AI Approach	Main Function	Typical Output	Maturity Level	Main Limitation
Computer vision and image preprocessing	Image enhancement and crack localisation	Enhanced image or crack map	High for basic detection	Sensitive to lighting, stains, shadows, and texture (Ali et al., 2021; Rajadurai & Kang, 2021)
Machine learning	Crack or non-crack classification using handcrafted features	Crack class or defect category	Moderate	Depends on feature selection and dataset quality (Fan, 2021; Yu et al., 2021)

Table 3 *(Continued)*

AI Approach	Main Function	Typical Output	Maturity Level	Main Limitation
Deep learning	Automated crack detection, segmentation, and classification	Crack mask, location, or class label	High for detection and classification	Requires large, labelled datasets and may suffer domain shift (Golding et al., 2022; Pu et al., 2022)
Thermal Imaging (TI)	Crack or damage detection using temperature variation and thermal patterns	Thermal crack indication or heat-based damage region	Emerging	Affected by temperature variation, environmental conditions, and interpretation of thermal signals (Kim et al., 2021; Nooralishahi et al., 2022)
LiDAR / 3D Scanning	Geometric crack or surface damage measurement using 3D data	Point-cloud damage features or 3D surface profile	Emerging	Higher cost, complex data processing, and limited routine use (Adhikari et al., 2023; Li et al., 2024)
Unmanned Aerial Vehicle (UAV)	Large-scale and hard-to-reach crack inspection using aerial image acquisition combined with AI analysis	Inspection images, crack maps, or detected crack regions	Emerging too high	Affected by camera angle, flight stability, image resolution, lighting, and operating regulations (Egodawela et al., 2024; Kumar et al., 2021; Sorilla et al., 2024)
Acoustic Emission (AE)	Internal damage detection and source localisation using acoustic signals	Damage signal pattern, source location, or internal fracture indication	Emerging	Depends on sensor placement, calibration, signal noise, and interpretation of acoustic data (Mukhti et al., 2024; Zhou et al., 2024)

A Decade of Methodological Evolution in AI-Based Concrete Crack Detection and Classification (2016-2025)

Evolution of Crack Detection Methods

Concrete crack detection was initially performed through manual visual inspection by engineers, which was laborious, subjective, and highly dependent on inspector expertise (Ali et al., 2021; Golding et al., 2022). Large-scale structures, such as bridges, may require long inspection periods, which can delay repair decisions and increase safety risks when damage progresses (Golding et al., 2022). Inspector fatigue, limited accessibility, and inconsistent judgement may also lead to missed minor cracks or variable assessment

outcomes. These limitations highlight the need for a more objective, scalable, and reliable crack detection approach (Cui & Qin, 2024).

The 1990s and 2000s saw the development of semi-automated crack detection through digital imaging and computer vision technology. Edge detection, thresholding, and morphological processing were commonly applied to concrete surface images to enhance crack visibility and separate cracks from the background. These methods improved inspection speed and consistency compared with manual inspection, and basic filtering approaches achieved approximately 70-80% accuracy under controlled lighting and uniform surface conditions (Ali et al., 2021). However, their performance declined in the presence of shadows, stains, rough textures, and complex backgrounds, which often caused false detections or missed cracks. This shows that early computer vision methods were useful but highly dependent on image quality and environmental conditions.

Machine learning methods emerged in the early 2010s as researchers moved from fixed image filters to trainable classifiers such as Support Vector Machines. These models improved crack detection accuracy to approximately 75-85%, exceeding many rule-based image processing methods (Fan, 2021; Yu et al., 2021). However, classical machine learning still relied heavily on handcrafted features, which limited its robustness across different lighting conditions, materials, and crack patterns. Models trained on limited datasets were also prone to overfitting and had restricted generalisation to real field environments (Yu et al., 2021). These limitations encouraged the transition towards deep learning approaches.

Deep learning marked a major methodological shift because crack features could be learned automatically from images without manual feature design. CNN-based models improved crack detection by capturing hierarchical image features and became more effective than traditional computer vision and classical machine learning methods (Golding et al., 2022; Pu et al., 2022). In the early 2020s, deep learning models were increasingly used for crack localisation, segmentation, and real-time video analysis (Kumar et al., 2021; Kumar et al., 2021; Pu et al., 2022). Compared with earlier methods, CNN-based models were less affected by surface texture and background noise, allowing them to distinguish cracks from non-crack artefacts more effectively (Yadav et al., 2024). However, their performance still depends on large, labelled datasets, diverse training images, and suitable validation under field conditions.

Recent studies indicate a further shift towards transformer-based models, hybrid CNN-transformer architectures, sensor fusion, and edge-AI deployment. Transformer-based models can capture wider contextual information and improve crack detection in complex backgrounds, while hybrid models combine local feature extraction from CNNs with global attention mechanisms from transformers (Shahin et al., 2024; Yadav et al., 2024). At the same time, UAV-based and edge-device models have improved the feasibility of real-time on-site inspection (Nguyen et al., 2024). Collectively, the evidence suggests that

crack detection has progressed from subjective manual assessment to automated AI-based monitoring. Nevertheless, field generalisation, dataset diversity, computational efficiency, and integration with structural assessment standards remain unresolved challenges.

Evolution of Crack Classification Methods

Crack classification by type and severity was traditionally performed through manual assessment, guided by engineering judgement and code-based recommendations. Cracks are commonly classified according to their origin, form, timing, orientation, width, and potential structural implications (Cui & Qin, 2024; Golding et al., 2022). For example, hairline cracks below 0.2 mm are generally considered less critical, whereas cracks exceeding 0.5 mm may require repair or structural intervention (Nyathi et al., 2024; Yu et al., 2021). However, manual classification is subjective and may vary depending on inspector experience, site conditions, and interpretation of severity criteria.

Earlier automated systems commonly separated crack detection and crack classification into two stages. In this approach, the first stage identified the presence or location of cracks, while the second stage used extracted crack properties to determine crack type or severity (Ali et al., 2021; Nyathi et al., 2024). This two-stage approach was widely used because it provided a structured workflow for image-based crack analysis (Pu et al., 2022; Taj et al., 2024). However, it was prone to error propagation, as missed or incorrect detection in the first stage could affect the accuracy of subsequent classification.

Compared with crack detection, crack classification remains less mature in the reviewed literature. Most studies focus on identifying crack presence, location, or segmentation, while fewer studies classify crack type, severity, or maintenance urgency. This is a key limitation because structural maintenance requires more than crack localisation. Engineers also need information about crack severity, possible cause, and structural implications before deciding whether monitoring, repair, or urgent intervention is required (Mazni et al., 2024; Ye et al., 2024).

Recent AI models show potential for integrated crack detection and classification. Deep learning systems can perform multiple tasks within a single workflow, including crack localisation, segmentation, and severity estimation (Li et al., 2022; Diniz et al., 2024). This integrated approach reduces duplicated processing, improves inspection speed, and limits the risk of error propagation between detection and classification stages (Abubakr et al., 2024; Ye et al., 2024). Therefore, the methodological trend is moving from separate detection-classification pipelines towards unified AI-based crack assessment systems.

Nevertheless, contradictory findings remain in crack classification research. Some studies report high classification accuracy using CNN, transformer-based, or hybrid models, but these results are often based on controlled datasets, limited crack categories, or predefined severity labels (Abubakr et al., 2024; Ye et al., 2024). In real field conditions,

severity classification is more complex because crack width, length, orientation, location, pattern, and structural context must be considered together (Nyathi et al., 2024; Nugraheni et al., 2023). Models may also perform less reliably when field images differ from training data due to lighting variation, surface texture, background clutter, or subtle crack patterns (Kolappan Geetha et al., 2023).

Collectively, the evidence suggests that crack classification has progressed from subjective manual judgement to automated AI-based severity assessment. However, classification remains less developed than detection, especially for engineering-based maintenance decision-making. Future crack classification systems should therefore combine detection, segmentation, measurement, severity classification, and maintenance recommendation within a single structural health monitoring workflow. The crack detection and classification methods from 2021 until now are summarised in Figure 4.

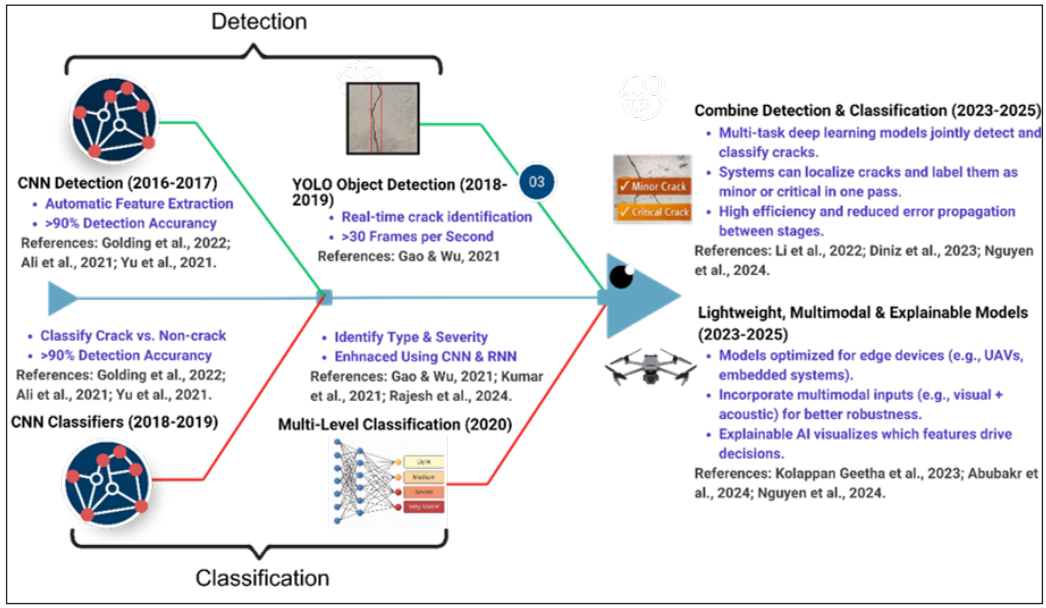


Figure 4. Summary evolution of crack detection and classification methods from pre-2000 to recent

Impact of AI on Concrete Crack Detection and Classification

Accuracy and Robustness Across AI-Based Methods

Across the reviewed studies, AI-based methods generally reported improved crack detection performance compared with manual inspection, traditional image processing, and classical machine learning (Ali et al., 2021; Golding et al., 2022). Deep learning models are effective because they can automatically learn crack features from images and identify complex crack patterns. However, reported performance values should be interpreted cautiously because the reviewed studies used different datasets, validation protocols, image acquisition

conditions, environmental settings, crack characteristics, and evaluation metrics (Arafin et al., 2024; Shahin et al., 2024). Therefore, accuracy, precision, recall, F1-score, Intersection over Union, and mean average precision are presented as study-specific indicators rather than direct evidence that one method is universally superior to another.

Transformer-based and hybrid CNN-transformer models have shown strong potential for crack detection under complex backgrounds because their attention mechanisms can capture wider contextual information and focus on relevant crack regions (Shahin et al., 2024; Yadav et al., 2024). In contrast, YOLO-based and edge-AI models are often more suitable for real-time inspection because they support faster inference and field deployment (Diniz et al., 2024; Nguyen et al., 2024). Nevertheless, these methods should not be directly ranked based only on isolated accuracy values because their reported performance depends on dataset diversity, image resolution, annotation quality, crack morphology, validation strategy, and field conditions. Thus, the collective evidence is more useful for identifying broad methodological trends rather than determining absolute superiority between models.

Segmentation-based deep learning models, including U-Net and its variants, provide detailed crack localisation because they analyse images at the pixel level. For example, a U-Net-based model reported a mean Intersection over Union above 0.91 for crack boundary detection in the tested images (Pu et al., 2022). This value indicates strong performance within that specific study setting, but it should not be directly compared with studies using different image datasets, segmentation labels, or evaluation procedures. Across segmentation studies, the improvement in robustness appears to be linked to pixel-level localisation, multi-scale feature extraction, attention mechanisms, and data augmentation, which help models detect cracks with different sizes, shapes, and orientations (Kang et al., 2022; Li & Zhao, 2022).

Lightweight AI models also show practical value for real-time crack inspection. YOLOv4-Tiny, for example, reported 87% mean average precision for real-time crack detection in a specific experimental setting (Ali et al., 2021), while recent YOLOv8-based studies have shown promising crack mask generation and recall performance (Diniz et al., 2024). However, lightweight models also involve a practical trade-off because faster inference may reduce reliability when detecting very fine cracks, low-contrast cracks, or cracks in noisy field images. Therefore, model selection should consider not only reported accuracy, but also robustness, computational demand, deployment platform, and field usability.

To further strengthen the comparative contribution of this review, Table 4 presents a benchmark-style performance matrix of AI-based concrete crack analysis methods. The matrix synthesises reported performance indicators together with methodological context and limitations. It is intended to show broad evidence patterns rather than directly rank the methods, because the reviewed studies used different datasets, validation protocols, environmental conditions, crack characteristics, and evaluation metrics.

Table 4
Benchmark-style performance matrix of AI-based concrete crack analysis methods

Method	Reported Performance Indicators	Context and Limitations
Computer vision and image preprocessing	Accuracy up to 96.6% was reported in selected studies	Effective under controlled image conditions, but sensitive to lighting, stains, shadows, noise, and surface texture (Mohammed et al., 2021; Rajesh et al., 2024; Rajadurai & Kang, 2021).
Machine learning	SVM-based methods reported up to 99.50% precision for orientation-related tasks	Depends on handcrafted features, dataset quality, and validation conditions; generalisation may be limited (Arafin et al., 2022; Fan, 2021; Taj et al., 2024).
Deep learning / CNN	Accuracy values approaching 99% were reported in selected datasets	Strong for detection and segmentation, but requires large, labelled datasets and may suffer laboratory-to-field domain shift (Golding et al., 2022; Lin et al., 2025; Pu et al., 2022).
Transformer-based models	Accuracy, mAP, and IoU were commonly reported	Captures wider contextual features but is computationally demanding and needs diverse datasets (Shahin et al., 2024; Yadav et al., 2024).
YOLO / real-time detectors	YOLOv4-Tiny reported 87% mAP in a specific setting	Suitable for real-time and UAV inspection, but may miss fine, low-contrast, or noisy cracks (Ali et al., 2021; Diniz et al., 2024; Nguyen et al., 2024).
UAV-based AI inspection	Detection, localisation, and segmentation performance were reported	Useful for bridges, roads, buildings, and high-rise structures; affected by flight stability, camera angle, resolution, and lighting (Egodawela et al., 2024; Kumar et al., 2021; Sorilla et al., 2024).
LiDAR / 3D scanning	Geometric measurement and point-cloud damage representation were reported	Supports 3D damage assessment, but has higher cost, complex processing, and limited routine accessibility (Adhikari et al., 2023; Li et al., 2024).
Acoustic emission / sensor-based methods	Source localisation and internal fracture detection were reported	Useful for internal damage monitoring, but depends on sensor placement, calibration, and signal-to-noise interpretation (Mukhti et al., 2024; Zhou et al., 2024).
Multimodal AI	Combined detection, classification, or damage assessment was reported	Improves damage coverage using multiple sensors, but data fusion, calibration, cost, and standardisation remain challenging (Liu et al., 2025; Nooralishahi et al., 2022).

Note. Reported values are study-specific and should not be interpreted as direct method rankings

Cost and Labour Efficiency

AI-driven crack examination reduces costs and labour while improving inspection efficiency. Large structures previously required large inspection teams, scaffolding, and lifting equipment. In contrast, AI-based systems reduce manual workload by automating crack identification, measurement, and documentation**.** Small CNN models on NVIDIA Jetson modules enable on-site crack analysis without cloud computation (Nguyen et al., 2024). This reduces the need for specialised hardware and supports field-based inspection (Kolappan, 2023; Meng et al., 2023).

AI-enabled drones have improved inspection efficiency, especially for bridges, high-rise buildings, and difficult-to-access structures (Kumar & Batchu, 2021). UAV-based crack detection can inspect large structures faster than manual inspection, with some studies reporting up to an 80% reduction in inspection time (Egodawela et al., 2024). Collectively, the evidence suggests that AI provides the strongest cost and labour benefit in large-scale, repetitive, and hazardous inspection tasks. However, the actual cost benefit still depends on equipment cost, operator training, software maintenance, and scale of deployment (Sorilla et al., 2024).

AI also supports preventive maintenance. Continuous crack monitoring allows early repair before small cracks develop into severe defects (Nyathi et al., 2024). Automated measurement and documentation allow inspectors to focus more on analysis and decision-making rather than data collection (Nooralishahi et al., 2022). Thus, AI contributes not only to short-term inspection efficiency, but also to long-term asset management.

Impact on Safety

AI inspections have improved worker safety. Drones can inspect high-rise buildings and bridge undersides, reducing the need for workers to climb scaffolding or enter traffic-risk areas (Egodawela et al., 2024; Kumar & Batchu, 2021; Kumar, 2021). This shows that AI shifts hazardous inspection tasks from humans to remote or automated platforms.

Field studies show that drone inspections reduce falls and accidents by 40% compared with conventional methods (Egodawela et al., 2024). Remote AI crack analysis also protects inspectors in disaster-affected or fire-damaged structures (Andrushia et al., 2022). The collective evidence suggests that AI improves safety most clearly in high-risk inspection environments, including bridges, high-rise structures, post-fire structures, and disaster-affected areas.

After earthquakes, AI crack detection can rapidly identify severely damaged areas and reduce emergency inspection team “hours at risk” (Naser et al., 2021). However, safety benefits depend on system reliability, false-alarm control, and human verification before safety-critical decisions are made.

Real-Time Inspection Capability

Artificial intelligence detects cracks quickly and supports faster inspection. YOLOv4-Tiny can detect cracks in real-time video streams at over 60 frames per second on standard hardware (Rajadurai & Kang, 2021). CNNs optimised for embedded devices can reach 30 frames per second, while earlier detectors such as Faster R-CNN operate below 10 frames per second (Ali et al., 2021; Kumar et al., 2021).

Further optimisation, such as pruning, transfer learning, and GPU acceleration, has improved real-time performance (Nguyen et al., 2024; Yadav et al., 2024). This indicates a clear methodological trend towards lightweight, optimised, and hardware-compatible AI models for field inspection. However, real-time capability often involves a trade-off between speed, image resolution, model complexity, and detection accuracy. Therefore, real-time inspection systems must balance computational efficiency with reliable crack detection performance.

Automated Severity Assessment

AI can now quantify crack severity from images in addition to detecting cracks. Segmentation models combined with geometric algorithms can measure crack width, length, and affected area. U-Net-based crack width measurement achieved an average error of 0.16 mm, which is close to manual calliper measurement (Nyathi et al., 2024). This shows that AI is moving from simple crack detection towards quantitative damage assessment.

AI models can classify crack severity using crack width and related features. Minor cracks are usually below 0.2 mm, while severe cracks above 0.5 mm may require repair (Nugraheni et al., 2023). Some AI models classify crack severity with 85–90% accuracy (Abubakr et al., 2024; Nugraheni, 2023). However, severity assessment remains more complex than binary detection because it requires interpretation of crack width, length, orientation, location, propagation pattern, and structural context.

AI can also support damage prioritisation by analysing crack patterns, cracked area ratio, water leakage risk, and load-bearing implications (Li et al., 2022; Mir, 2022; Swarna, 2024; Taj et al., 2024). Collectively, the evidence suggests that automated severity assessment has strong potential, but its reliability depends on accurate measurement, clear severity thresholds, and alignment with engineering standards.

Continuous Monitoring and Multimodal Integration

Permanent AI-based monitoring systems can track crack formation and growth beyond scheduled inspections. Smart cameras and embedded sensors can continuously monitor structures and detect new cracks or crack expansion (Nguyen et al., 2024). This suggests that continuous monitoring is a key advantage of AI because it supports earlier detection and

preventive maintenance. However, false alarms from lighting, debris, and environmental changes remain a practical limitation.

Multimodal systems combine visual, thermal, acoustic, radar, or LiDAR data to detect cracks that one sensor may miss. Regular cameras may fail to detect hidden or low-visibility cracks, while thermal imaging and ground-penetrating radar can support deeper damage identification (He et al., 2024; Liu et al., 2025; Nooralishahi et al., 2022). The evidence suggests that multimodal integration improves detection coverage, but it also increases cost, calibration difficulty, and data interpretation complexity.

Digital twin and Building Information Modelling integration can further support predictive maintenance by linking crack data with structural performance models (Lee et al., 2024). Overall, continuous monitoring, multimodal sensing, and digital twin integration represent an emerging direction in AI-based crack analysis. However, unresolved issues remain in data standardisation, sensor calibration, system interoperability, and translation of AI outputs into engineering decisions. Table 5 summarises the reported benefits and limitations of AI-based crack inspection.

Table 5
Summary of reported benefits and limitations of AI-based crack inspection

Dimension	Reported Benefits	Reported Limitations
Accuracy & Robustness	CNN models ~90-93% accurate; transformer-based models ~95-99%, greatly reducing false detections under complex conditions (Mohammed et al., 2021; Yadav et al., 2024)	Performance drops in unfamiliar environments without extensive diverse training data (Shahin et al., 2024)
Cost Efficiency	~30-50% lower inspection costs reported with UAV/AI methods (e.g., replacing scaffolding and manual labour) (Sorilla et al., 2024)	High initial cost for drones and equipment; cost benefits realised only at larger-scale deployments
Labour Efficiency	Automation covers more area with fewer staff – e.g., ~50% reduction in manual image annotation effort using transfer learning (Gradišar & Dolenc, 2023)	Requires skilled operators for drones/ AI systems and ongoing training; human oversight still needed for verification
Safety	Dangerous tasks shifted to robots (e.g., drones inspect high places); some studies reported reduced inspection-related safety risk (Egodawela et al., 2024; Kumar et al., 2021)	Relies on technology reliability (battery, communication); regulatory limits on drone use can constrain deployment
Real-time Capability	Immediate crack alerts during inspections – lightweight models run 30-60 FPS (vs. <10 FPS for older methods), enabling on-the-spot decisions (Ali et al., 2021)	Achieving high FPS may require powerful hardware or lower image resolution; trade-off between speed and maximum accuracy

Table 5. *(continued)*

Dimension	Reported Benefits	Reported Limitations
Automated System Assessment	AI measures crack width to ~0.1-0.2 mm precision and reported severity classification performance of approximately 85-90% in selected studies (Nyathi et al., 2024; Nugraheni et al., 2023)	Calibration needed for each structure (e.g., scaling in images); very fine cracks near resolution limits may still be missed
Other Impact: Continuous 24/7 Monitoring	Always-on monitoring catches crack formation and growth between scheduled inspections, enabling instant maintenance alerts (McAlorum et al., 2023)	Potential false alarms from lighting or dirt require careful system calibration (McAlorum et al., 2023); constant power and data management are needed
Other impact: Multimodal Integration	Combined sensors (visual + thermal, radar, etc.) detect hidden or nighttime cracks, improving overall detection coverage (Liu et al., 2025; Nooralishahi et al., 2022)	Additional sensors increase complexity and cost; requires precise sensor alignment and can be difficult to interpret multi-source data (Nooralishahi et al., 2022)

Key Challenges and Methodological Limitations in AI-Based Crack Analysis

Dataset Limitations and Annotation Bottlenecks

Crack image collections have accelerated AI research but are limited (Arafin et al., 2024). SDNET2018 has over 56,000 labelled images of cracked and uncracked concrete bridges, walls, and pavements, with crack widths ranging from 0.06 to 25 mm (Shahin et al., 2024). However, even large databases cannot capture all real-world diversity (Arafin et al., 2024). Many datasets do not fully represent actual field conditions, such as lighting variation, surface stains, shadows, moisture, debris, rough textures, and different camera angles. Models trained on a single dataset often struggle with unfamiliar textures, lighting conditions, or damage types (Lin et al., 2025). This dataset bias limits real-world generalisation (Adhikari et al., 2023).

Data annotation is time-consuming and may introduce inconsistency (Mazni et al., 2024). Manual crack annotations may contain ground truth errors, labelling discrepancies, or selection bias, especially for thin or irregular cracks (Arafin et al., 2024; Shahin et al., 2024). Therefore, dataset size alone is not sufficient; dataset diversity, annotation consistency, and benchmark quality are also required for reliable model development.

Crack datasets often have significant asymmetry. Photos of unblemished concrete may outnumber cracked surfaces, or certain crack forms may predominate. Training on uneven data can skew model predictions (Arafin et al., 2024). Without data augmentation or balanced sampling, models may predict “no crack” and fail to identify rarer crack types (Shahin et al., 2024). This is particularly important for crack classification because severity categories must be well represented to support maintenance decision-making.

Generalisation Challenges and Lab-to-Field Performance Decline

AI crack detection models often perform well on laboratory or benchmark datasets, but their practicality may be limited (Arafin et al., 2024). Deep learning models often detect cracks with over 95% accuracy in controlled tests, and specialised architectures can achieve almost 99% accuracy under ideal conditions (Lin et al., 2025; Yadav et al., 2024). However, field conditions affect these results (Arafin et al., 2024). This shows a contradiction in the literature: high accuracy in controlled datasets does not always translate into reliable field performance.

The domain shift between training data and real-world infrastructure scenarios degrades performance (Shahin et al., 2024). Many datasets fail to represent environmental changes such as lighting, weather, surface textures, and noise (Shahin et al., 2024). Cracks in an old bridge, indoor parking structure, or wet road may appear differently (Arafin et al., 2024). Therefore, future models should be tested using external datasets, cross-dataset validation, and real field images to confirm their robustness.

Generalisation is also affected by the balance between missed detections and false positives. False negatives are dangerous because missed cracks may continue to grow and compromise structural integrity (Adhikari et al., 2023). However, aiming for very high recall may increase false positives when stains, shadows, or surface marks are classified as cracks (Kang & Cha, 2022). Thus, future studies should balance sensitivity and precision rather than focusing only on overall accuracy.

Real-Time Processing and Deployment Constraints

Deploying AI crack detection models from laboratory settings to field environments remains challenging due to computational demands and processing-speed constraints (Nguyen et al., 2024). Many deep learning models require GPUs for real-time processing because of their complex CNN or transformer-based architectures (Shahin et al., 2024). In many field applications, high-resolution images are collected on-site but processed offline using advanced computers, which can delay inspection results and slow maintenance decision-making (Mazni et al., 2024). This indicates that high detection performance alone is insufficient if the model cannot operate efficiently under real field conditions.

For practical deployment, recent studies have shifted towards lightweight, optimised, and edge-compatible AI models (Nguyen et al., 2024). Real-time object detection frameworks such as YOLO support rapid inference and allow drone or video-based inspections to be conducted more efficiently (Kumar & Batchu, 2021; Nguyen et al., 2024). CNN- and YOLO-based crack detection systems on UAVs can support faster inspection of large or difficult-to-access areas (Mazni et al., 2024). In addition, lightweight and attention-based architectures have been explored to improve mobile or embedded deployment while maintaining crack detection performance (Kang & Cha, 2022). However, their effectiveness

still depends on hardware capability, image resolution, model complexity, camera stability, and field image quality.

There is an inherent trade-off between inference speed, model complexity, and detection accuracy (Mazni et al., 2024). Although enhanced YOLO-based models and efficient encoder-decoder CNNs have shown promising real-time potential, these findings remain dependent on specific datasets, hardware settings, and testing environments (Nguyen et al., 2024; Pu et al., 2022). Higher inference speed may require lower image resolution, model compression, or simplified detection tasks, which can reduce sensitivity to fine cracks, low-contrast cracks, or cracks in noisy backgrounds (Mazni et al., 2024). In UAV and robotic inspection, additional constraints such as battery life, camera movement, communication stability, and data storage also affect real-time performance (Nguyen et al., 2024). Therefore, future real-time crack detection systems should balance computational efficiency with reliable detection performance under variable lighting, surface texture, camera movement, and environmental conditions.

Integration into Structural Health Monitoring and Reliability Concerns

AI detection results are not always directly connected to safety or maintenance decisions (Arafin et al., 2024). Although algorithms can identify cracks and severity indices, repair scheduling, load limits, and maintenance priority still require engineering judgment (Arafin et al., 2024). This creates a gap between AI-based image analysis and practical structural health monitoring.

AI systems must work in a closed-loop process to identify cracks, assess severity, and support appropriate maintenance actions (Arafin et al., 2024). Most AI models provide static evaluations and do not fully monitor crack propagation or damage patterns over time (Adhikari et al., 2023). Future systems should therefore connect crack detection, severity assessment, crack progression, and maintenance priority within the same structural health monitoring workflow.

Implementation is also hindered by trust and interpretability. Deep learning models often operate as black-box systems, identifying cracks without clearly explaining the model rationale (Kang & Cha, 2022; Tyvoniuk et al., 2025). Explainable AI methods, such as saliency maps, may help users understand which image regions influenced the prediction (Lin et al., 2025). Therefore, explainability is important not only for model interpretation, but also for user trust and engineering acceptance.

Reliability remains a major concern because false negatives may allow major cracks to grow, while false positives may reduce user confidence and increase unnecessary inspection costs (Adhikari et al., 2023; Kang & Cha, 2022). The industry also lacks a standard approach for testing and applying AI systems for crack detection (Arafin et al., 2024). Therefore,

future AI-based crack analysis requires benchmark datasets, standardised reporting metrics, external validation, and closer alignment with civil engineering assessment standards.

Research Gaps and Future Directions

The synthesis of the reviewed studies highlights several important research gaps. First, AI-based crack detection performs well in controlled datasets, but its generalisation to real infrastructure remains insufficiently validated (Arafin et al., 2024; Shahin et al., 2024). Second, crack classification and severity assessment are less developed than crack detection, especially for engineering-based decision-making (Mazni et al., 2024; Nyathi et al., 2024). Third, AI outputs are not yet fully integrated with structural health monitoring decisions, such as repair priority, safety risk, and maintenance strategy (Adhikari et al., 2023; Arafin et al., 2024). Fourth, explainability remains limited because many deep learning models provide accurate outputs without clearly explaining the basis of their decisions (Kang & Cha, 2022; Swarna et al., 2024).

Overall, the collective evidence suggests that AI-based methods have improved the speed, consistency, and accuracy of concrete crack analysis. However, unresolved issues remain in field robustness, standardised evaluation, explainability, and integration with engineering decision-making (Nguyen et al., 2024; Yadav et al., 2024). Therefore, the future research agenda should prioritise benchmark dataset development, external field validation, automated severity assessment, explainable AI, lightweight real-time deployment, standardised evaluation metrics, and stronger integration with structural health monitoring decisions.

CONCLUSION

AI-based methods have significantly advanced concrete crack detection and classification by improving inspection speed, consistency, automation, and accessibility in structural health monitoring. Deep learning models, YOLO-based detectors, vision transformer approaches, UAV-assisted inspection, and multimodal sensing have expanded the capability of crack analysis from simple detection towards segmentation, classification, severity assessment, and maintenance support. However, reported performance values should be interpreted cautiously because the reviewed studies used different datasets, validation protocols, environmental conditions, crack characteristics, and evaluation metrics. Therefore, the findings should be understood as broad methodological trends rather than direct rankings of model superiority.

Despite these advances, several limitations remain before AI-based crack inspection can be fully implemented in field practice. Current systems still face challenges related to dataset diversity, annotation quality, field generalisation, computational demand, explainability, and integration with engineering decision-making. Crack classification and severity assessment

are also less mature than crack detection, particularly in real infrastructure conditions where crack width, length, orientation, location, propagation pattern, and structural significance must be considered together.

Future research should prioritise benchmark dataset development, external field validation, standardised evaluation metrics, explainable AI, lightweight edge deployment, and stronger integration with structural health monitoring systems. Greater attention should also be given to automated severity assessment and maintenance decision support so that AI-based crack inspection can move beyond detection towards practical asset management. With these improvements, AI-based crack analysis can become a more reliable and field-ready tool for preventive maintenance, infrastructure safety, and long-term structural monitoring.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2025/ICT02/UTHM/02/1). The authors also thank the Government of Malaysia for providing support through the MyBrain2.0 programme.

REFERENCES

Abubakr, M., Rady, M., Badran, K., & Mahfouz, S. Y. (2024). Application of deep learning in damage classification of reinforced concrete bridges. *Ain Shams Engineering Journal, 15*(1), Article 102297. https://doi.org/10.1016/j.asej.2023.102297

Adhikari, M. D., Kim, T.-H., Yum, S.-G., & Kim, J.-Y. (2023). Damage detection and monitoring of a concrete structure using 3D laser scanning. *Engineering Proceedings, 36*(1), Article 1. https://doi.org/10.3390/engproc2023036001

Ali, L., Alnajjar, F., Jassmi, H. A., Gochoo, M., Khan, W., & Serhani, M. A. (2021). Performance evaluation of deep CNN-based crack detection and localisation techniques for concrete structures. *Sensors, 21*(5), Article 1688. https://doi.org/10.3390/s21051688

Alkannad, A. A., Al Smadi, A., Yang, S., Al-Smadi, M. K., Al-Makhlafi, M., Feng, Z., & Yin, Z. (2025). CrackVision: Effective concrete crack detection with deep learning and transfer learning. *IEEE Access, 13*, 29554-29576. https://doi.org/10.1109/ACCESS.2025.3540841

Alkayem, N. F., Mayya, A., Shen, L., Zhang, X., Asteris, P. G., Wang, Q., & Cao, M. (2024). Co-CrackSegment: A new collaborative deep learning framework for pixel-level semantic segmentation of concrete cracks. *Mathematics, 12*(19), Article 3105. https://doi.org/10.3390/math12193105

Andrushia, A. D., Anand, N., Neebha, T. M., Naser, M. Z., & Lublóy, É. (2022). Autonomous detection of concrete damage under fire conditions. *Automation in Construction, 140*, Article 104364. https://doi.org/10.1016/j.autcon.2022.104364

Arafin, P., Billah, A. H. M. M., & Issa, A. (2024). Deep learning-based concrete defects classification and detection using semantic segmentation. *Structural Health Monitoring, 23*(1), 383-409. https://doi.org/10.1177/14759217231168212

Arafin, P., Issa, A., & Billah, A. H. M. M. (2022). Performance comparison of multiple convolutional neural networks for concrete defects classification. *Sensors, 22*(22), Article 8714. https://doi.org/10.3390/s22228714

Chen, R. (2021). Migration learning-based bridge structure damage detection algorithm. *Scientific Programming, 2021*, Article 1102521. https://doi.org/10.1155/2021/1102521

Cui, X., Wang, Q., Dai, J., Li, S., Xie, C., & Wang, J. (2022). Pixel-level intelligent recognition of concrete cracks based on DRACNN. *Materials Letters, 306*, Article 130867. https://doi.org/10.1016/j.matlet.2021.130867

Cui, P., & Qin, Y. (2024). Introducing methods for analysing and detecting concrete cracks at the No. 3 Huaiyin Pumping Station in the South-to-North Water Diversion Project in China. *Buildings, 14*(8), Article 2431. https://doi.org/10.3390/buildings14082431

Del Savio, A. A., Luna Torres, A., Cárdenas Salas, D., Vergara Olivera, M. A., & Urday Ibarra, G. T. (2023). Detection and evaluation of construction cracks through image analysis using computer vision. *Applied Sciences, 13*(17), Article 9662. https://doi.org/10.3390/app13179662

Diniz, J. D. C. N., de Paiva, A. C., Junior, G. B., de Almeida, J. D. S., Silva, A. C., Cunha, A. M. T. D. S., & Cunha, S. C. A. P. D. S. (2024). A one-step methodology for identifying concrete pathologies using neural networks: Using YOLO v8 and dataset review. *Applied Sciences, 14*(10), Article 4332. https://doi.org/10.3390/app14104332

Dow, H., Perry, M., McAlorum, J., Pennada, S., & Dobie, G. (2023). Skeleton-based noise removal algorithm for binary concrete crack image segmentation. *Automation in Construction, 151*, Article 104867. https://doi.org/10.1016/j.autcon.2023.104867

Egodawela, S., Khodadadian Gostar, A., Buddika, H. A. D. S., Dammika, A. J., Harischandra, N., Navaratnam, S., & Mahmoodian, M. (2024). A deep learning approach for surface crack classification and segmentation in unmanned aerial vehicle assisted infrastructure inspections. *Sensors, 24*(6), Article 1936. https://doi.org/10.3390/s24061936

Fan, C.-L. (2021). Detection of multidamage to reinforced concrete using support vector machine-based clustering from digital images. *Structural Control and Health Monitoring, 28*(12), Article e2841. https://doi.org/10.1002/stc.2841

Fang, Z., Wang, X., Gao, J., & Eskandarpour, B. (2025). Optimising concrete crack detection: An echo state network approach with improved fish migration optimisation. *Scientific Reports, 15*(1), Article 40. https://doi.org/10.1038/s41598-024-84458-1

Gaur, A., Kishore, K., Jain, R., Pandey, A., Singh, P., Wagri, N. K., & Roy-Chowdhury, A. B. (2023). A novel approach for industrial concrete defect identification based on image processing and deep convolutional neural networks. *Case Studies in Construction Materials, 19*, Article e02392. https://doi.org/10.1016/j.cscm.2023.e02392

Ghafari, S., Nejad, F. M., Sheikh-Akbari, A., & Kazemi, H. (2024). Automating the determination of low-temperature fracture resistance curves of normal and rubberised asphalt concrete in single-edge notched

beam tests using convolutional neural networks. *Construction and Building Materials, 428*, Article 136376. https://doi.org/10.1016/j.conbuildmat.2024.136376

Gharehbaghi, V., Noroozinejad Farsangi, E., Yang, T. Y., Noori, M., & Kontoni, D.-P. N. (2022). A novel computer-vision approach assisted by 2D-wavelet transform and locality sensitive discriminant analysis for concrete crack detection. *Sensors, 22*(22), Article 8986. https://doi.org/10.3390/s22228986

Golding, V. P., Gharineiat, Z., Munawar, H. S., & Ullah, F. (2022). Crack detection in concrete structures using deep learning. *Sustainability, 14*(13), Article 8117. https://doi.org/10.3390/su14138117

Gradišar, L., & Dolenc, M. (2023). Transfer and unsupervised learning: An integrated approach to concrete crack image analysis. *Sustainability, 15*(4), Article 3653. https://doi.org/10.3390/su15043653

He, X., Feng, Q., Shao, H., Li, H., & Fu, M. (2024). Detection method for chloride ion penetration distribution in concrete based on hyperspectral images and LSTM. *Optics Express, 32*(21), 37323-37341. https://doi.org/10.1364/OE.535163

Hong, Y., & Yoo, S. B. (2022). OASIS-Net: Morphological attention ensemble learning for surface defect detection. *Mathematics, 10*(21), Article 4114. https://doi.org/10.3390/math10214114

Kang, D. H., & Cha, Y.-J. (2022). Efficient attention-based deep encoder and decoder for automatic crack segmentation. *Structural Health Monitoring, 21*(5), 2190-2205. https://doi.org/10.1177/14759217211053776

Khallaf, R., Naderpajouh, N., & Hastak, M. (2018). A systematic approach to develop risk registry frameworks for complex projects. *Built Environment Project and Asset Management, 8*(4), 334-347. https://doi.org/10.1108/BEPAM-08-2017-0051

Kim, B., Choi, S.-W., Hu, G., Lee, D.-E., & Serfa Juan, R. O. (2021). Multivariate analysis of concrete image using thermography and edge detection. *Sensors, 21*(21), Article 7396. https://doi.org/10.3390/s21217396

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology, 51*(1), 7-15. https://doi.org/10.1016/j.infsof.2008.09.009

Kolappan Geetha, G., Yang, H.-J., & Sim, S.-H. (2023). Fast detection of missing thin propagating cracks during deep-learning-based concrete crack/non-crack classification. *Sensors, 23*(3), Article 1419. https://doi.org/10.3390/s23031419

Kumar, P., Batchu, S., Swamy S., N., & Kota, S. R. (2021). Real-time concrete damage detection using deep learning for high rise structures. *IEEE Access, 9*, 112312-112331. https://doi.org/10.1109/ACCESS.2021.3102647

Kumar, P., & Batchu, S. (2021). Real-time structural crack detection in buildings using YOLOv3 and autonomous unmanned aerial systems. *International Journal of System Assurance Engineering and Management, 14*, 1691-1702. https://doi.org/10.1007/s13198-023-02192-9

Kun, J., Zhenhai, Z., Jiale, Y., & Jianwu, D. (2022). A deep learning-based method for pixel-level crack detection on concrete bridges. *IET Image Processing, 16*(10), 2609-2622. https://doi.org/10.1049/ipr2.12512

Lee, D.-E., Choi, Y., Hong, G., Maruthi, M., Yi, C.-Y., & Park, Y.-J. (2024). Parametric image-based concrete defect assessment method. *Case Studies in Construction Materials, 20*, Article e02962. https://doi.org/10.1016/j.cscm.2024.e02962

Li, S., & Zhao, X. (2022). A performance improvement strategy for concrete damage detection using stacking ensemble learning of multiple semantic segmentation networks. *Sensors, 22*(9), Article 3341. https://doi.org/10.3390/s22093341

Li, B., Guo, H., Wang, Z., & Li, M. (2022). Automatic crack classification and segmentation on concrete bridge images using convolutional neural networks and hybrid image processing. *Intelligent Transportation Infrastructure, 1*, Article liac016. https://doi.org/10.1093/iti/liac016

Li, Y., Li, M., & Tang, H. (2024). A crack detection method for civil engineering bridges based on feature extraction and parametric modelling of point cloud data. *Journal of Computing and Information Technology, 32*(2), 81-96. https://doi.org/10.20532/cit.2024.1005830

Lin, Y., Ahmadi, M., Alnowibet, K. A., & Bukhari, F. A. (2025). Concrete crack detection using ridgelet neural network optimised by advanced human evolutionary optimisation. *Scientific Reports, 15*(1), Article 4858. https://doi.org/10.1038/s41598-025-89250-3

Lin, S., Ye, H., Tan, D., Wang, J., & Yin, J. (2025). Identification of defects in underground structures using machine learning aided distributed fibre optic sensing. *Journal of Rock Mechanics and Geotechnical Engineering, 17*(4), 2194-2207. https://doi.org/10.1016/j.jrmge.2024.03.025

Liu, N., Ge, Y., Bai, X., Zhang, Z., Shangguan, Y., & Li, Y. (2025). Research on damage detection methods for concrete beams based on ground penetrating radar and convolutional neural networks. *Applied Sciences, 15*(4), Article 1882. https://doi.org/10.3390/app15041882

Liu, H., Li, L., Li, Y., Long, Q., & Chen, Z. (2024). A hybrid attention mechanism and RepGFPN method for detecting wall cracks in high-altitude cleaning robots. *IEEE Photonics Journal, 16*(5), Article 8600608. https://doi.org/10.1109/JPHOT.2024.3453943

Mayya, A. M., & Alkayem, N. F. (2024). Enhance the concrete crack classification based on a novel multi-stage YOLOV10-ViT framework. *Sensors, 24*(24), Article 8095. https://doi.org/10.3390/s24248095

Mazni, M., Husain, A. R., Shapiai, M. I., Ibrahim, I. S., Anggara, D. W., & Zulkifli, R. (2024). An investigation into real-time surface crack classification and measurement for structural health monitoring using transfer learning convolutional neural networks and Otsu method. *Alexandria Engineering Journal, 92*, 310-320. https://doi.org/10.1016/j.aej.2024.02.052

McAlorum, J., Dow, H., Pennada, S., Perry, M., & Dobie, G. (2023). Automated concrete crack inspection with directional lighting platform. *IEEE Sensors Letters, 7*(11), 1-4. https://doi.org/10.1109/LSENS.2023.3327611

Meng, Q., Hu, L., Wan, D., Li, M., Wu, H., Qi, X., & Tian, Y. (2023). Image-based concrete cracks identification under complex background with lightweight convolutional neural network. *KSCE Journal of Civil Engineering, 27*(12), 5231-5242. https://doi.org/10.1007/s12205-023-0923-1

Miele, S., Karve, P., & Mahadevan, S. (2025). Transfer learning for probabilistic localisation of hidden cracks in concrete structures. *Journal of Civil Structural Health Monitoring, 15*(4), 1139-1159. https://doi.org/10.1007/s13349-024-00873-y

Mir, B. A., Sasaki, T., Nakao, K., Nagae, K., Nakada, K., Mitani, M., Tsukada, T., Osada, N., Terabayashi, K., & Jindai, M. (2022). Machine learning-based evaluation of the damage caused by cracks on concrete structures. *Precision Engineering, 76*, 314-327. https://doi.org/10.1016/j.precisioneng.2022.03.016

Mohammed, M. A., Han, Z., & Li, Y. (2021). Exploring the detection accuracy of concrete cracks using various CNN models. *Advances in Materials Science and Engineering, 2021*, Article 9923704. https://doi.org/10.1155/2021/9923704

Mukhti, J. A., Gucunski, N., & Kee, S.-H. (2024). AI-assisted ultrasonic wave analysis for automated classification of steel corrosion-induced concrete damage. *Automation in Construction, 167*, Article 105704. https://doi.org/10.1016/j.autcon.2024.105704

Nguyen, C. L., Nguyen, A., Brown, J., Byrne, T., Ngo, B. T., & Luong, C. X. (2024). Optimising concrete crack detection: A study of transfer learning with application on Nvidia Jetson Nano. *Sensors, 24*(23), Article 7818. https://doi.org/10.3390/s24237818

Nogueira Diniz, J. D. C., Paiva, A. C. D., Junior, G. B., de Almeida, J. D. S., Cunha, A. C., Cunha, A. M. T. D. S., & Cunha, S. C. A. P. D. S. (2023). A method for detecting pathologies in concrete structures using deep neural networks. *Applied Sciences, 13*(9), Article 5763. https://doi.org/10.3390/app13095763

Nooralishahi, P., Ramos, G., Pozzer, S., Ibarra-Castanedo, C., Lopez, F., & Maldague, X. P. V. (2022). Texture analysis to enhance drone-based multi-modal inspection of structures. *Drones, 6*(12), Article 407. https://doi.org/10.3390/drones6120407

Nyathi, M. A., Bai, J., & Wilson, I. D. (2024). Deep learning for concrete crack detection and measurement. *Metrology, 4*(1), 66-81. https://doi.org/10.3390/metrology4010005

Pennada, S., Perry, M., McAlorum, J., Dow, H., & Dobie, G. (2023). Threshold-based BRISQUE-assisted deep learning for enhancing crack detection in concrete structures. *Journal of Imaging, 9*(10), Article 218. https://doi.org/10.3390/jimaging9100218

Philip, R. E., Andrushia, A. D., Nammalvar, A., Gurupatham, B. G. A., & Roy, K. (2023). A comparative study on crack detection in concrete walls using transfer learning techniques. *Journal of Composites Science, 7*(4), Article 169. https://doi.org/10.3390/jcs7040169

Pu, R., Ren, G., Li, H., Jiang, W., Zhang, J., & Qin, H. (2022). Autonomous concrete crack semantic segmentation using deep fully convolutional encoder–decoder network in concrete structures inspection. *Buildings, 12*(11), Article 2019. https://doi.org/10.3390/buildings12112019

Qiao, W., Ma, B., Liu, Q., Wu, X., & Li, G. (2021). Computer vision-based bridge damage detection using deep convolutional networks with expectation maximum attention module. *Sensors, 21*(3), Article 824. https://doi.org/10.3390/s21030824

Rajadurai, R.-S., & Kang, S.-T. (2021). Automated vision-based crack detection on concrete surfaces using deep learning. *Applied Sciences, 11*(11), Article 5229. https://doi.org/10.3390/app11115229

Rajesh, S., Jinesh Babu, K. S., Chengathir Selvi, M., & Chellapandian, M. (2024). Automated surface crack identification of reinforced concrete members using an improved YOLOv4-Tiny-based crack detection model. *Buildings, 14*(11), Article 3402. https://doi.org/10.3390/buildings14113402

Sapidis, G. M., Kansizoglou, I., Naoum, M. C., Papadopoulos, N. A., & Chalioris, C. E. (2024). A deep learning approach for autonomous compression damage identification in fibre-reinforced concrete using piezoelectric lead zirconate titanate transducers. *Sensors, 24*(2), Article 386. https://doi.org/10.3390/s24020386

Shahin, M., Chen, F. F., Maghanaki, M., Hosseinzadeh, A., Zand, N., & Khodadadi Koodiani, H. (2024). Improving the concrete crack detection process via a hybrid visual transformer algorithm. *Sensors, 24*(10), Article 3247. https://doi.org/10.3390/s24103247

Shang, L., Zhang, Z., Tang, F., Cao, Q., Yodo, N., Pan, H., & Lin, Z. (2023). Deep learning enriched automation in damage detection for sustainable operation in pipelines with welding defects under varying embedment conditions. *Computation, 11*(11), Article 218. https://doi.org/10.3390/computation11110218

Sorilla, J., Chu, T. S. C., & Chua, A. Y. (2024). A UAV-based concrete crack detection and segmentation using 2-stage convolutional network with transfer learning. *HighTech and Innovation Journal, 5*(3), 690-702. https://doi.org/10.28991/HIJ-2024-05-03-010

Swarna, R. A., Hossain, M. M., Khatun, M. R., Rahman, M. M., & Munir, A. (2024). Concrete crack detection and segregation: A feature fusion, crack isolation, and explainable AI-based approach. *Journal of Imaging, 10*(9), Article 215. https://doi.org/10.3390/jimaging10090215

Taj, M. N. A. B. G., Alruwais, N., Alshahrani, H. M., Vijayalakshmi, J., Shanmugapriya, N., & Jayaprakash, S. (2024). Precision crack analysis in concrete structures using CNN, SVM, and KNN: A machine learning approach. *Revista Matéria, 29*(4), Article e20240551. https://doi.org/10.1590/1517-7076-RMAT-2024-0551

Tyvoniuk, V., Trach, R., & Trach, Y. (2025). Integration of probability maps into machine learning models for enhanced crack segmentation in concrete bridges. *Applied Sciences, 15*(6), Article 3201. https://doi.org/10.3390/app15063201

Wang, L. (2023). Automatic detection of concrete cracks from images using Adam-SqueezeNet deep learning model. *Frattura ed Integrità Strutturale, 17*(65), 289-299. https://doi.org/10.3221/IGF-ESIS.65.19

Wang, R., Zhou, X., Liu, Y., Liu, D., Lu, Y., & Su, M. (2024). Identification of the surface cracks of concrete based on ResNet-18 depth residual network. *Applied Sciences, 14*(8), Article 3142. https://doi.org/10.3390/app14083142

Yadav, D. P., Sharma, B., Chauhan, S., & Dhaou, I. B. (2024). Bridging convolutional neural networks and transformers for efficient crack detection in concrete building structures. *Sensors, 24*(13), Article 4257. https://doi.org/10.3390/s24134257

Ye, G., Dai, W., Tao, J., Qu, J., Zhu, L., & Jin, Q. (2024). An improved transformer-based concrete crack classification method. *Scientific Reports, 14*(1), Article 6226. https://doi.org/10.1038/s41598-024-54835-x

Yu, Y., Rashidi, M., Samali, B., Yousefi, A. M., & Wang, W. (2021). Multi-image-feature-based hierarchical concrete crack identification framework using optimised SVM multi-classifiers and D-S fusion algorithm for bridge structures. *Remote Sensing, 13*(2), Article 240. https://doi.org/10.3390/rs13020240

Zhou, Y., Liang, M., & Yue, X. (2024). Deep residual learning for acoustic emission source localisation in a steel-concrete composite slab. *Construction and Building Materials, 411*, Article 134220. https://doi.org/10.1016/j.conbuildmat.2023.134220

Pertanika J. Sci. & Technol. 34 (4): 2487 - 2518 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Photovoltaic Module Defect Classification Analysis Based on Convolutional Neural Network Architecture with Augmentation Dataset

Nurul Atikah Mazlan[1], Shahrani Shahbudin[1]*, and Murizah Kassim[1,2]

[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

As power producers increasingly adopted renewable energy sources, solar energy emerged as a key solution due to its sustainability and efficient energy harvesting. Photovoltaic (PV) modules convert solar irradiance into electricity, but defects in these modules significantly impact power generation efficiency. This study aims to classify PV defect images through deep learning, focusing on Convolutional Neural Networks (CNNs). It explores two different transfer learning architectures: Series Networks, including AlexNet, VGG-16, and VGG-19, and Directed Acyclic Graph (DAG) Networks, such as ResNet-18, ResNet-50, Inception-V3, and GoogLeNet. Series Networks, known for their structured, layer-wise feature extraction, performed well in specificity and precision; meanwhile, DAG Networks, with deeper architectures and residual connections, enhanced feature learning and classification accuracy. The study evaluated augmented and non-augmented datasets to determine the most effective approach. Results showed that the augmented dataset consistently outperformed the non-augmented dataset. Among the architectures, ResNet-50, a DAG Network, achieved the highest classification accuracy (98.96%) and F1 score (97.89%) due to its ability to retain critical features across multiple layers. Meanwhile, AlexNet excelled in specificity (99.53%), and VGG-16 demonstrated superior precision (98.65%), highlighting the strengths of Series Networks in structured, less complex classification tasks.

Keywords: Augmentation dataset, convolutional neural network (CNN), photovoltaic (PV) module defect images

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 29 September 2025
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.26

E-mail addresses:
tikahmazlan01@gmail.com (Nurul Atikah Mazlan)
shahranis@uitm.edu.my (Shahrani Shahbudin)
murizah@uitm.edu.my (Murizah Kassim)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The usage of renewable energy rather than fossil fuels for the generation of electricity must be taken into consideration due to worries about pollution that causes climate change. This makes solar energy an important and more sought-after power source. Solar power systems, also known as solar energy, are among the most widely used renewable energy sources due to their eco-friendly nature (Hussein et al., 2021). These systems comprise several key components designed to efficiently capture and convert sunlight into electricity. At the heart of the system are photovoltaic (PV) panels, which generate electricity by absorbing sunlight. However, maintaining the efficiency of PV modules is crucial to ensuring optimal power output. Therefore, it is essential for engineers or responsible personnel to regularly inspect and maintain the PV modules at the solar farm (Li et al., 2020). Regular inspection is critical, as human intervention may sometimes overlook defects.

PV systems generate electricity by harnessing sunlight through a series of interconnected PV panels, utilising the photoelectric effect. Each panel consists of multiple PV modules, which are further composed of an array of PV cells. These cells are essential for converting solar energy into electricity. They absorb sunlight and transform it into usable electrical power, enabling the efficient generation of renewable energy (Balani et al., 2021). Monitoring photovoltaic plants and optimising their energy is crucial to ensure these facilities can effectively contribute to meeting the growing electricity demand. While knowing the average power output is important, it is insufficient for assessing the overall performance of a PV plant or detecting potential failures in specific components, such as individual cells or clusters of cells.

Numerous photovoltaic (PV) technologies, such as silicon heterojunctions, perovskites, organic–inorganic hybrids, and dye-sensitised solar cells (DSSCs), have been the subject of intensive research due to the desire for affordable and scalable solar cells. High efficiencies have been quickly attained by perovskite devices, and current efforts are concentrated on improving their long-term stability in ambient conditions. Within silicon-based heterojunctions, the integration of metal–oxide semiconductors onto p-type crystalline silicon (p-Si) has emerged as a simple, low-temperature fabrication route, with SnO_2 offering advantages such as a wide band gap, high carrier mobility, stability, and non-toxicity. Although various deposition techniques have been applied, hydrothermal growth provides a particularly uniform, low-cost, and controllable approach. However, reported SnO_2-based solar cells often involve complex fabrication and lack systematic evaluation of thickness effects. To address this gap, the present study employs hydrothermal deposition to fabricate SnO_2/Si heterojunction solar cells and investigates the influence of SnO_2 layer thickness on charge transport and light-harvesting performance (Özel et al., 2022).

The choice of photovoltaic technology is a key factor influencing both the efficiency and reliability of solar energy systems, and advanced concepts such as dye-sensitised solar cells

(DSSCs) attract considerable interest because of their low cost, easy fabrication and flexible structure. Doping plasmonic nanoparticles (Au, Ag) in (DSSC) has been demonstrated to improve the following phenomena: light absorption, electron injection, and diminish charge recombination at TiO_2/dye/electrolyte. The author in Salimi et al. (2019) identified that plasmonic nanocomposites, including Au@TiO_2 and Ag@TiO_2, can significantly boost power conversion efficiency (PCE), with mesoporous Ag–Au@TiO_2 nanocomposites that offering a cost-effective approach for future PV applications.

All-inorganic perovskite solar cells (PSCs) face challenges of charge carrier recombination that limit their efficiency, according to Faheem et al. (2022). A synergistic strategy using liquid-exfoliated MoS_2 at the $CsPbIBr_2$/HTL interface improves perovskite grain growth, reduces trap density, and stabilises the interface. As a result, the PCE increased from 7.38% to 10.98%, demonstrating a promising route for more efficient and stable PSCs. Various techniques for extracting natural dyes from cornelian cherries (Cornus mas L.) were studied to enhance the performance of dye-sensitised solar cells (DSSCs). The best results were obtained from citric acid-based solvent extraction, which increased short-circuit current and light absorption to increase power conversion efficiency from 0.47% to 0.98%. This demonstrates how bio-based dyes and appropriate extraction techniques can increase the efficiency of DSSCs (Kocak et al., 2019).

A low-cost and simple synthesis route was developed to fabricate carbon-based counter electrodes (CEs) as alternatives to conventional Pt-based ones in DSSCs. The homemade carbon paste CE (H-CE) exhibited a porous structure that enhanced electrolyte diffusion, reduced charge transfer resistance, and improved catalytic activity compared to commercial carbon paste CE (C-CE). As a result, the device achieved a power conversion efficiency of 2.70% and a high fill factor of 73%, surpassing both C-CE and Pt-based electrodes (Altinkaya et al., 2020).

A malfunction in any PV cell can result in reduced power generation, leading to significant losses in output. If a cell is shaded or malfunctioning, the output power decreases (Rao et al., 2021). Not all PV panels can generate output power normally, as faults can happen on the panel because of many factors. According to Nylund and Barbari (2019), a study on defects detected on PV modules obtained by visual inspection, such as cracks, snail trail, degradation, hotspots, delamination and others. Cracks and degradation may happen from the manufacturing process, which can lead to failures in the module. The existence of cracks on a module can contribute to power losses compared to normal PV modules.

Testing each individually for solar farms with thousands of PV panels can be time-consuming and cumbersome. Maintaining the quality and reliability of these panels is crucial for efficient power generation, long-term durability, and maximising returns on investment. Instead of manual testing, it would be highly beneficial to have real-time information on the energy output of each panel in the plant. This would enable quick

identification of underperforming panels, allowing for timely repairs or replacements. Achieving this requires individual monitoring of each PV panel and transmitting the data to the control system. Given the high costs associated with solar-generated electricity, optimising system performance, energy production, and reliability is particularly important.

System failures in PV modules can occur due to various defects that impact their efficiency. According to Nylund and Barbari (2019), PV module failures typically progress through three stages: infant failure, midlife failure, and wear-out failure, all of which reduce efficiency. Common issues such as partial shading, hotspots, and cracking can further compromise performance. In this study, defects are categorised into fractures, delamination, encapsulation discolouration, glass breakage, and gridline corrosion. Cracks may develop during handling or manufacturing processes like soldering, while delamination often results from low-quality materials or improper processing. When the encapsulating material yellows, the PV module's ability to transmit light is diminished. Improper clamp geometry can result in glass breakage, which may lead to hotspot issues. The PV module may overheat and stop producing electricity because of the temperature increase. According to Li et al. (2020), gridline corrosion can be triggered by environmental factors such as heavy snow, which causes the degradation of components over time.

Quickly identifying and resolving faults is crucial for maintaining the efficiency of photovoltaic (PV) systems. Commercial power plants prioritise maximising energy production, minimising losses and maintenance costs, and ensuring safe operation (Appiah et al., 2019; Basnet et al., 2020). Since PV systems are prone to faults, early detection is key to enhancing their safety, performance, and longevity. To strengthen protection measures, the U.S. National Electric Code requires the installation of Overcurrent Protection Devices (OCPDs) and Ground Fault Detection Interrupters (GFDIs) in PV systems (Pierdicca et al., 2018).

One of the aspects that must be stressed to ensure that problems are found as quickly as possible is the inspection process. These problems may be resolved with the use of technological advancements that can reduce the need for manpower. Deep learning algorithms are a branch of artificial intelligence designed to simulate human brain function and decision-making (Meng et al., 2023). The neural network architecture used in deep learning models has hidden layers for a variety of purposes, including feature extraction. This technique was largely used in current PV module inspection methods to categorise PV module defects. Convolutional Neural Networks (CNNs) are the most widely utilised deep learning algorithm for image classification. This is due to CNN's ability to extract patterns, classify, and even detect defects in images of PV modules (Henrique et al., 2021). CNN, also known as a feedforward network, processes input data through convolutional and pooling layers before passing it to fully connected layers for image classification (Cipriani et al., 2020). CNNs are extensively applied in computer vision and pattern recognition

tasks such as image classification, object detection, facial recognition, natural language processing, and style transfer (Zyout & Oatawneh, 2020). In the field of photovoltaics, CNNs have been employed as classifiers to detect defects in PV modules, with various architectures trained to achieve high accuracy in defect identification. Based on Azis et al. (2022), DenseNet-201 was utilised by the researcher for the classification of images. The results show that DenseNet-201 outperforms other CNN architectures, such as DCNN and VGG-16. Additionally, Zyout and Oatawneh (2020) utilised a modified version of the AlexNet-based CNN architecture and gave outperformed results. This proves that CNN performs better than other deep learning methods.

Transfer learning is a deep learning technique that leverages a pre-trained model from one task as a foundation for solving a related task (Salih et al., 2023). It is typically significantly quicker and simpler than training a network from the start for updating and retraining. Application areas where the method is frequently utilised include object identification, picture recognition, and speech recognition. Transfer learning has two types, which are transfer learning with a series network and transfer learning with a DAG network. A series network is a type of deep learning neural network where layers are arranged sequentially, with a single input layer and a single output layer. Examples of series networks include AlexNet, VGG-16, and VGG-19. In contrast, a Directed Acyclic Graph (DAG) network has a more complex structure, allowing layers to receive inputs from multiple layers and pass data to others. This design enables greater flexibility and deeper architectures. Examples of DAG networks include GoogLeNet, ResNet-50, Inception-V3, and ResNet-101 (Yu et al., 2023).

The model might experience overfitting issues due to the limited and imbalanced number of images in the training dataset. Overfitting of the model happens when a model can learn the training dataset well, but not for a new dataset (Pavlitskaya et al., 2022). The technique known as augmentation is used on the small dataset to increase the network architecture's accuracy (Zulkarnain et al., 2023). According to Bakhshayesh et al. (2022), the augmentation of the dataset is the process of increasing the number of images by applying the augmentation technique. This is because overfitting issues experienced by machine learning can compromise the model's accuracy in the classification task. The augmented dataset will transform such geometric transformations that are usually used in image augmentation. Some types of geometric transformation, including annotation, blurring, flipping, and rotating, are applied during the augmentation process (Ahmad et al., 2022; Zulkarnain et al., 2023).

Many researchers have explored deep learning-based methods for PV defect image classification. However, most studies have only compared their models against one or two other architectures (Rahman et al., 2019), lacking a comprehensive evaluation of PV classification using Convolutional Neural Networks (CNNs) for both Directed Acyclic

Graph (DAG) and Series Network architectures. The objective of this study is to determine the most effective CNN architecture for classifying PV defect images using both augmented and non-augmented datasets within DAG and Series Network frameworks, to assess CNN performance in PV module defect classification using key performance metrics, and to determine the architecture that achieves the best overall classification performance.

LITERATURE REVIEW

Several previous studies have been reviewed to analyse the methods applied for classifying defects in PV modules. Table 1 summarises these works according to the deep learning networks employed for PV defect classification.

In Tella et al. (2025), an ensemble-based deep learning framework was applied for defect detection using the ELPV dataset of 2,624 EL images. The ensemble combined eight models, which are AlexNet, SENet, GoogLeNet, Xception, Vision Transformer (ViT), Darknet53, ResNet18, and SqueezeNet, to improve prediction accuracy. The ensemble achieved accuracies of 68.36% and 68.31%, outperforming a previous hybrid model (61.15%). However, the study also showed that individual models can sometimes perform better, with ResNet18 achieving 73.02% accuracy in a binary classification task.

The authors in Drir et al. (2024) proposed an ensemble CNN that employed weighted feature fusion by integrating three architectures: VGG-16, ResNet, and MobileNet. A comparative analysis with the individual baseline models revealed that the ensemble

Table 1
PV defect classification method

Reference	Method Used	Result
Tella et al., 2025	AlexNet SENet GoogLeNet (Inception V1) Xception Vision Transformer (ViT) Darknet53 ResNet18 SqueezeNet	A 2624 sample of EL images was classified based on their defect by using 8 ensemble models. ResNet-18 obtained the best accuracy result with 73.02%, demonstrating that the network outperforms the ensemble under specific conditions.
Drir et al., 2025	VGG-16 ResNet MobileNet	12 PV defect images were identified by using the new model, which combined deep features from the CNN architecture's layer that was involved in this study. The ensemble CNN achieved an average accuracy of 96% compared to individual CNN models.
Kellil et al., 2023	DCNN VGG-16	Fine-tuned VGG-16 outperformed in terms of accuracy compared to DCNN for both binary-class and multiclass analysis of PV defect images
Azis et al., 2022	DenseNet-201 VGG-16	DenseNet-201 outperformed in performance metrics compared to VGG-16

achieved superior performance, attaining an average accuracy of 96%. This approach demonstrated higher accuracy and precision in classifying 12 types of PV defects from thermographic images.

In Kellil et al. (2023), both binary and multiclass classifications of PV defect images were analysed using a small DCNN and a fine-tuned VGG-16 model. The results indicated that the fine-tuned VGG-16 outperformed the small DCNN, achieving higher accuracy in both classification tasks. The limitations of ResNet in image classification, particularly due to the identity shortcut, were highlighted in Azis et al. (2022), where the authors introduced a more efficient architecture, DenseNet-201. The identity shortcut was found to cause a collapsing domain problem, reducing the network's learning capacity. To address this, the study employed DenseNet-201 and VGG-16 for PV defect classification. The results demonstrated that DenseNet-201 achieved superior accuracy and overall performance metrics compared to VGG-16.

Meanwhile, in this work, CNN was selected as the classifier for PV defect classification, with seven architectures applied through transfer learning, covering both Series and DAG Networks. The Series Networks included AlexNet, VGG-16, and VGG-19, while the DAG Networks consisted of ResNet-18, ResNet-50, GoogLeNet, and Inception-V3. Unlike ensemble approaches, such as the one reported in Drir et al. (2024), which achieved high accuracy but required greater complexity and computational cost, this work focused on standard CNN architectures that are simpler and less resource-intensive. Previous studies have also utilised standard CNNs; however, most were limited to only two or three models, offering less comprehensive evaluation.

METHODOLOGY

Figure 1 illustrates the workflow of the proposed approach, which is structured into three phases. The initial phase entails gathering PV defect images from the solar farm, covering five types of defects: glass breakage, delamination, encapsulation discolouration, cracks, and gridline corrosion. Next, the dataset is split into two parts: one with augmentation and one without augmentation. In the final phase, the images are classified using a Convolutional Neural Network (CNN) with seven different architectures: AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, GoogLeNet, and Inception-V3. The performance of each

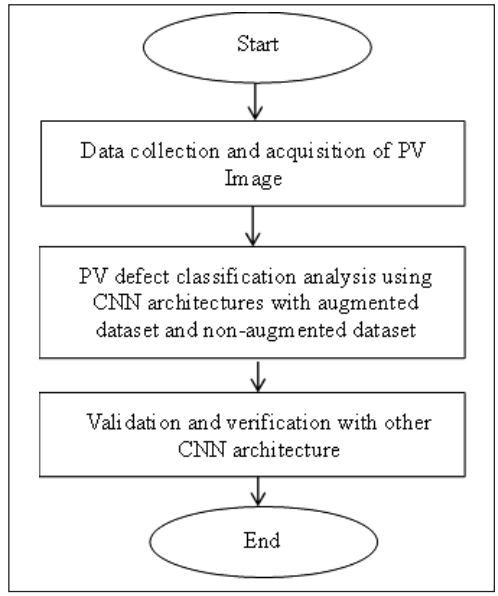


Figure 1. System flowchart

architecture is then evaluated using performance metrics for both the augmented and non-augmented datasets.

Data Collection and Acquisition of PV Image

According to Nylund and Barbari (2019), many defect occur in PV modules have been analysed. Classification of PV defect images in this work only covered 5 types of defects, which are cracks, delamination, glass breakage, encapsulation discolouring, and gridline corrosion. The type of images used in this work is RGB, which consists of various sizes and angles. The dataset comprises 1,394 images of varying sizes. Figure 2 displays PV defect images collected from a solar farm, categorised into five classes. The dataset includes 76 images of cracks, 109 images of delamination, 114 images of encapsulation discolouration, 64 images of gridline corrosion, and 1,031 images of glass breakage.

Due to the various sizes of images, it resizes the input image for network requirements and has a specific input size to train the model. This research uses CNN networks such as AlexNet, VGG, ResNet, GoogLeNet, and Inception-V3. Every network has its image input size. VGG (VGG-16 and VGG-19), ResNet (ResNet-18 and ResNet-50), and GoogLeNet have the same image input size, which requires images to be in 224-by-224 dimensions (Pangilinan et al., 2022). Alexnet requires 227 × 227 pixels of input size, and Inception-V3 requires 299 × 299 pixels of image input size (Anilkumar et al., 2021).

The training dataset does not have the same quantity for every class. Data augmentation is applied for original dataset as the number of images in every class is imbalanced and to reduce overfitting of the model. The distribution of classes, however, varies significantly; where 64 images displaying gridline corrosion and 1,031 images showing glass breakage. Since the dataset was gathered from a single location, this may have reduced its diversity in terms of lighting or the severity of defects. Future work should therefore focus on developing a larger and more diverse dataset that incorporates multi-site data, varied environmental conditions, and a more balanced distribution across defect types.

The dataset is divided into 70% for training and 30% for validation, yielding 976 training images and 418 validation images. To enhance the dataset, augmentation techniques such as rotation and X, Y translation are applied. Figure 3 displays the images after the augmentation process.

PV Defect Classification using Convolutional Neural Network (CNN)

Transfer learning is classified into two distinct types: transfer learning using a series network and transfer learning using a Directed Acyclic Graph (DAG) network. A series network is a deep learning model where layers are structured in sequential order, consisting of one input layer and one output layer. Examples of series networks include AlexNet, VGG-16, and VGG-19. On the other hand, a DAG network features a more complex design, enabling

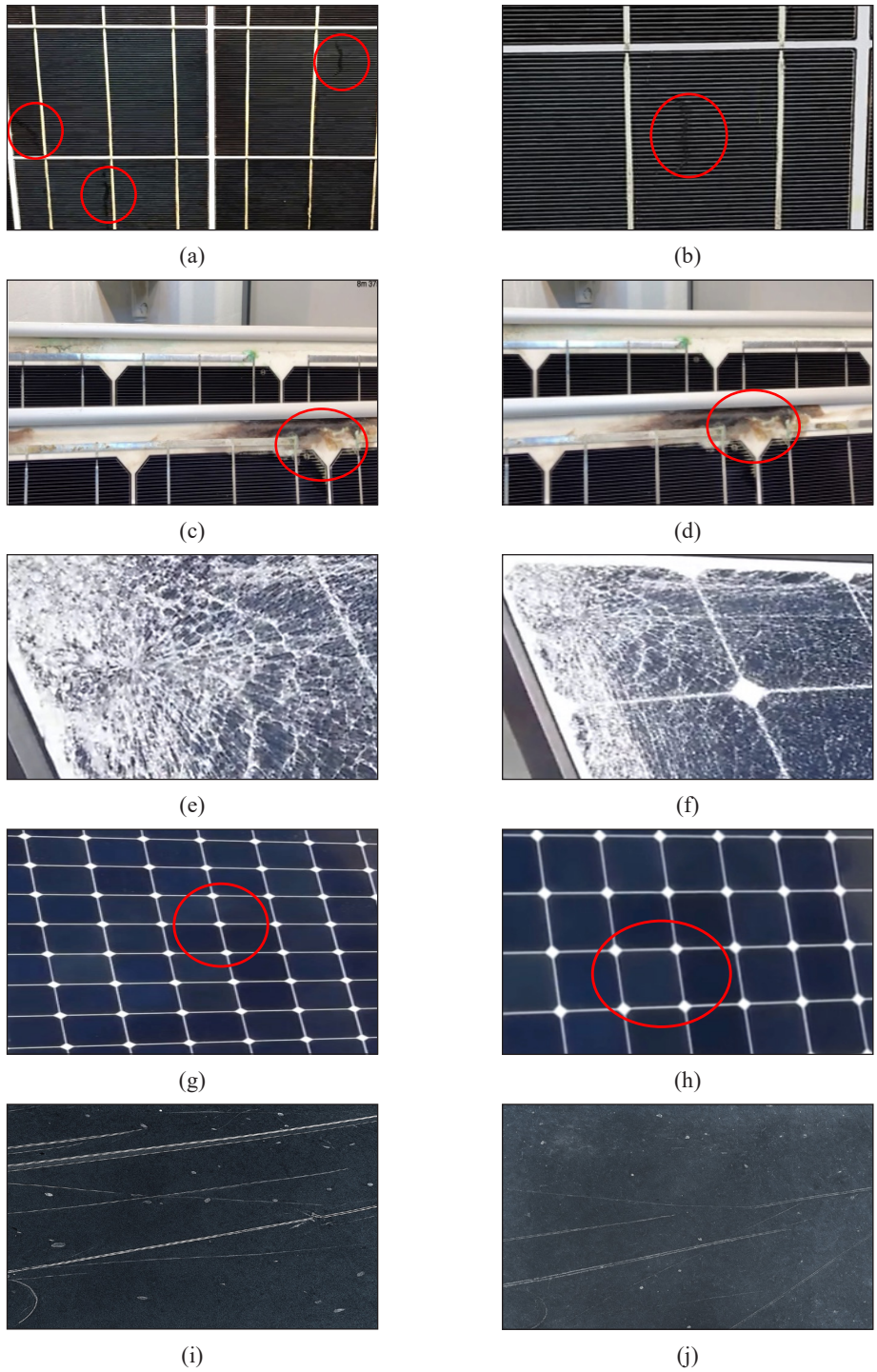


Figure 2. Example of PV module defect images (a-b) Cracks; (c-d) Delamination; (e-f) Glass breakage; (g-h) Encapsulation discolouring and (i-j) Gridline corrosion

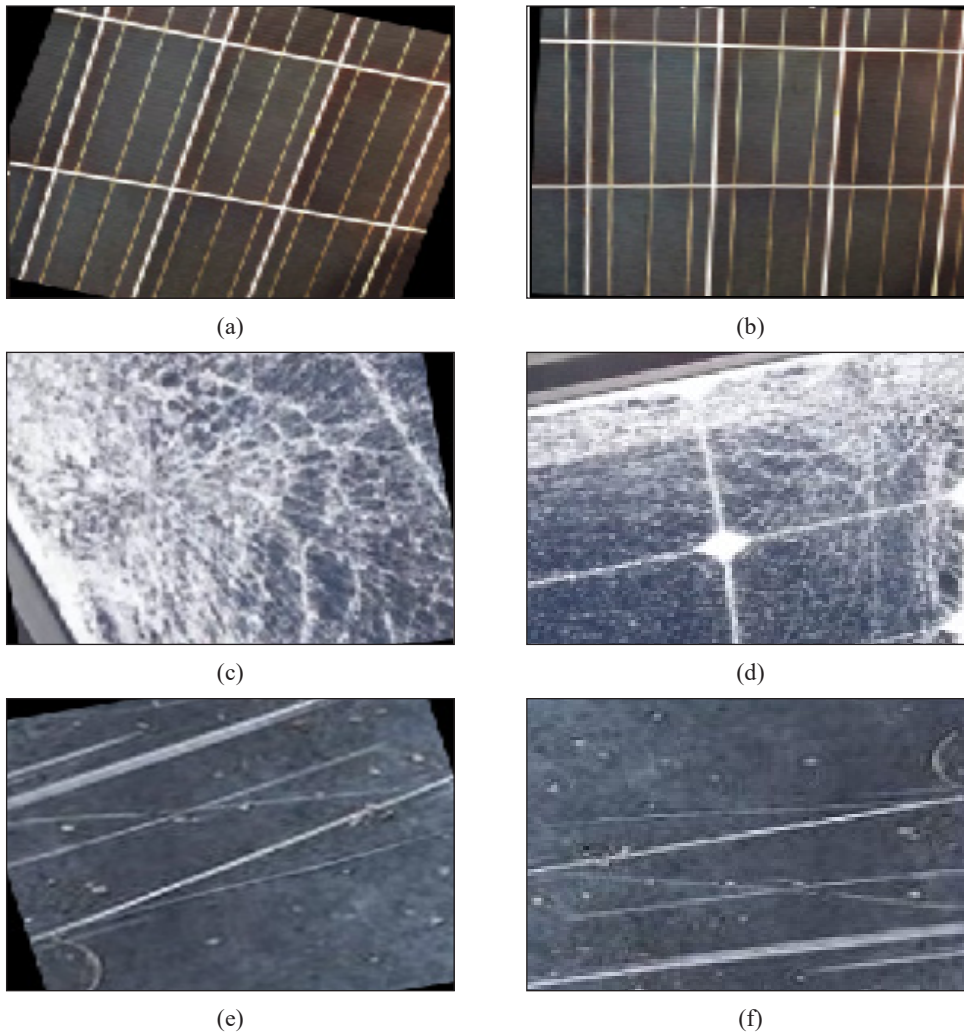


Figure 3. Sample of augmentation dataset with (a, c, e) X, Y translation and (b, d, f) rotation

layers to receive inputs from multiple sources and transmit data to others, providing enhanced flexibility in deep learning models. The design of a DAG network can be more complex, with layers creating data for other layers and accepting input from other layers. Some networks that fall within the DAG network category are GoogLeNet, ResNet-50, Inception-V3, and ResNet-101 (Anilkumar et al., 2021).

AlexNet

AlexNet is an 8-layer deep network that is also one of the CNN architectures. The layers consist of 3 fully connected layers and 5 convolutional layers (Xu et al., 2022). The details of the layer for the AlexNet architecture can be referred to in Table 2.

Table 2
AlexNet architecture

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	227 × 227 × 3	-	-	-
1	Convolution	96	55 × 55 × 96	11 × 11	4	relu
	Max Pooling	96	27 × 27 × 96	3 × 3	2	relu
2	Convolution	256	13 × 13 × 256	5 × 5	1	relu
	Max Pooling	256	13 × 13 × 256	3 × 3	2	relu
3	Convolution	384	13 × 13 × 384	3 × 3	1	relu
4	Convolution	384	13 × 13 × 384	3 × 3	1	relu
5	Convolution	256	13 × 13 × 256	3 × 3	1	relu
	Max Pooling	256	6 × 6 × 256	3 × 3	2	relu
6	FC	-	9216	-	-	relu
7	FC	-	4096	-	-	relu
8	FC	-	4096	-	-	relu
Output	FC	-	100	-	-	Softmax

Note. Adapted from (Li et al., 2022)

This network consists of 8 layers with 5 total convolution layers, three max-pooling layers, and the remaining fully connected layers. The input image size required by this network is 227 × 227 × 3. First, the images will undergo the first convolution layer with a feature map of 96, making the images' size 55 × 55 × 96. The AlexNet adds a maximum pooling layer or sub-sampling layer with a filter size of 33 and a stride of 2. After reduction, 27 × 27 × 96 will be the final image's size. The final fully connected layer's output is fed back into a 1000-way Softmax, which generates a distribution over the 1000 class labels (Xu et al., 2022).

This network was selected for its deeper architecture and strong capability in learning complex features, which is crucial for detecting and classifying PV module defects such as cracks. In addition, AlexNet incorporates local response normalisation, which enhances the model's generalisation by standardising responses across multiple channels (Siuly et al.,2024).

Visual Geometry Group (VGG)

The Computer Vision Group from Oxford University discovered another CNN, namely the Visual Geometry Group (VGG), which is a complete deep-learning network (Tao, 2022). Image style transfer based on the VGG neural network model. According to Bagaskara and Suryanegara (2021), VGG uses fewer parameters while the network's performance improves. VGG-16's 3 × 3 filter size represents an improvement over AlexNet's version. The design of VGG-16 is composed of three Fully Connected Layers, three Max Pooling

Operations, and a Softmax Activation Layer for each block. Known for its efficiency and ease of use, VGG-16 excels in a variety of computer vision tasks, including image classification and object recognition. Max-pooling layers appear after a stack of convolutional layers with progressively deeper layers in the model's architecture (Zou et al., 2020). Using this method, the model can acquire intricate hierarchical representations of visual qualities, resulting in accurate and consistent predictions (Arvinth, 2024).

Meanwhile, the VGG-19 architecture is relatively simple, making it easier to understand and implement. With many available pre-trained models, it is well-suited for transfer learning in computer vision tasks. Its strong feature extraction capability is particularly important for capturing PV module defects in greater detail (Arvinth, 2024).

Figure 4 illustrates the layer structure of the VGG architectures. Column D displays the layers of VGG-16, while column E shows the layers of VGG-19. Columns A, B, and C highlight earlier versions of VGG, including modifications made before the development of VGG-16 and VGG-19 (Xia et al., 2020). The VGG architecture requires input images of

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 4. VGG architecture
Note. Adapted from (Bagaskara & Suryanegara, 2021)

224 × 224 pixels. It features convolution layers with 3 × 3 and 1 × 1 filters, which function as linear input transformations. A stride of 1 pixel is used for subsampling. The hidden layers use ReLU activation functions, which play a key role in the VGG architecture. As shown in Figure 4, VGG-19 contains 19 weight layers, whereas VGG-16 has 16. Both networks incorporate a max pooling layer along with multiple deep convolutional layers, and trainable convolutional layers of 19 in VGG-19 and 16 in VGG-16 (Triyadi et al., 2022).

Residual Neural Network

Unlike other CNNs, where convolutional layers can be stacked indefinitely without significant performance degradation, Residual Neural Networks (ResNet) introduce a unique residual block. This block uses skip connections to bypass certain layers, helping to mitigate issues like vanishing or exploding gradients during training (Rahman et al., 2021). ResNet has evolved, with architectures such as ResNet-18 and ResNet-50 being commonly compared. The primary difference between them is network depth, where ResNet-50 includes additional convolutional layers to enhance learning capability (Chindarkkar et al., 2020). ResNet-18, in contrast, is a smaller network with two-layer-deep residual blocks. The ResNet-18 architecture is a compact network featuring two layers of ResNet blocks. Compared to traditional CNN architecture, ResNet demonstrates superior performance in extracting relevant features that may encounter vanishing gradient problems or reduced representation capacity. ResNet's usage of skip connections enhances training convergence and feature extraction while also enabling deeper model creation (Luu et al.,2025).

Table 3 illustrates the layer structure of both ResNet-18 and ResNet-50 designs. Five convolution stages make up each ResNet architecture, and their purpose is to extract an image's features. 224 × 224-pixel images were required for this architecture's input size. Convolution is 3 × 3 for each layer with a 3 × 3 kernel size. Subsampling in ResNet-50 is achieved using convolution with a stride of two. This deep architecture, with 50 layers, provides significant advantages for image classification by enabling more complex feature extraction. Figure 5 illustrates the ResNet-50 architecture, highlighting its greater depth in comparison to ResNet-18. The network features global average pooling, a fully connected layer, and a Softmax layer at the final stage for classification (Li & He, 2018)

Inception-V3

Inception-V3 is a CNN architecture that consists of 48 layers deep. This network required images to have dimensions of 299 by 299 as requirement to use this network. The improvement towards GoogLeNet produces Inception-V3, where this network is the enhanced version of the GoogLeNet network. Inception-V3's layers are shown in Figure 5, where this network consists of 13 total layers.

Differing from LeNet and VGG, the uniqueness of this network can be seen in the layer where there are special layers called the Inception Module. This module uses receptive

Table 3
ResNet architecture

Layer name	Output size	ResNet-18	ResNet-50
Conv1	112 × 112	7 × 7, 64, stride 2	7 × 7, 64, stride 2
Conv2_x	56 × 56	3 × 3 max pool, stride 2	3 × 3 max pool, stride 2
		$\begin{bmatrix}3x3,64\\3x3,64\end{bmatrix}x2$	$\begin{bmatrix}1x1,64\\3x3,64\\1x1,256\end{bmatrix}x3$
Conv3_x	28 × 28	$\begin{bmatrix}3x3,128\\3x3,128\end{bmatrix}x2$	$\begin{bmatrix}1x1,64\\3x3,64\\1x1,256\end{bmatrix}x3$
Conv4_x	14 × 14	$\begin{bmatrix}3x3,256\\3x3,256\end{bmatrix}x2$	$\begin{bmatrix}1x1,256\\3x3,256\\1x1,1024\end{bmatrix}x6$
Conv5_x	7 × 7	$\begin{bmatrix}3x3,512\\3x3,512\end{bmatrix}x2$	$\begin{bmatrix}1x1,512\\3x3,512\\3x3,2048\end{bmatrix}x3$
	1 × 1	Average pool, 1000-d fc, softmax	
FLOPs		1.8×10^9	4.1×10^9

Note. Adapted from (Pangilinan et al., 2022)

Type	Kernel size/stride	Input size
Convolution	3 × 3/2	299 × 299 × 3
Convolution	3 × 3/1	149 × 149 × 32
Convolution	3 × 3/1	147 × 147 × 32
Pooling	3 × 3/2	147 × 147 × 64
Convolution	3 × 3/1	73 × 73 × 64
Convolution	3 × 3/2	71 × 71 × 80
Convolution	3 × 3/1	35 × 35 × 192
Inception module	Three modules	35 × 35 × 288
Inception module	Five modules	17 × 17 × 768
Inception module	Two modules	8 × 8 × 1,280
Pooling	8×8	8 × 8 × 2,048
Linear	Logits	1 × 1 × 2,048
Softmax	Output	1 × 1 × 1,000

Figure 5. Inception-V3 architecture
Note. Adapted from (Zhang et al., 2022)

kernels of various sizes to capture different levels of detail. Zero padding is applied to ensure consistency in the output size of convolution operations. The final feature maps are generated by concatenating the filtered outputs. The inception operation enhances feature extraction, allowing the network to capture richer details from the input image (Feng et al., 2019). The efficiency of the network is increased by Inception modules' usage of 1 × 1 convolutions, which lower the dimensionality and complexity of feature maps. Furthermore, this CNN architecture's auxiliary classifiers at intermediate layers improve performance by offering additional supervision during training, which helps address the vanishing gradient issue (Arnob et al., 2024)

GoogLeNet

GoogLeNet is a deep Convolutional Neural Network (CNN) model consisting of 22 layers (Swarup et al., 2023). It requires input images of size 224 × 224; the detailed layer structure can be seen in Figure 6. Designed with computational efficiency, GoogLeNet incorporates two extra classifier layers, which are linked to the outputs of the Inception (4a) and Inception (4d) layers (Zhang et al., 2022). This network consists of two auxiliary classifiers to conduct the gradient forward for gradient vanishing purposes. By combining feature information from various scales, convolutional kernels of various sizes broaden the network's breadth and increase its scalability, which improves the network's capacity to identify features during training (Wang & Feng, 2024).

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Figure 6. GoogLeNet architecture
Note. Adapted from (Zhang et al., 2022)

EffecientNet

EfficientNet is a family of convolutional neural networks developed to optimise both accuracy and computational efficiency through a novel compound scaling method. Unlike traditional CNNs that scale depth, width, or resolution independently, EfficientNet uniformly scales these dimensions using a set of fixed scaling coefficients, resulting in a more balanced architecture (Tan & Le, 2019). The EfficientNetV2 series further improves upon this by introducing faster training speed, reduced memory usage, and improved parameter efficiency, making it suitable for large-scale image classification tasks (Tan & Le, 2021). EfficientNetV2 represents an enhanced version of the original EfficientNet, developed to deliver higher accuracy while minimising computational demands. One of its key improvements lies in training efficiency, with reports indicating that EfficientNetV2 can train up to 11 times faster than its predecessor, EfficientNetV1 (Kamal et al., 2023).

Validation and Verification with Other CNN Architectures

Performance is evaluated using multiple metrics, with Figure 7 illustrating a sample confusion matrix for binary classification. This matrix highlights key values: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which help measure the model's accuracy and overall effectiveness in classification.

The confusion matrix consists of four key variables. True Positive (TP) represents the number of instances where the classifier correctly identifies the positive class, while True Negative (TN) indicates the cases where the classifier correctly identifies the negative class. False Negative (FN) refers to instances where the classifier incorrectly predicts an actual positive case as negative, whereas False Positive (FP) refers to cases where the classifier incorrectly predicts an actual negative case as positive (Alves et al., 2021). From these values, evaluation metrics such as accuracy, F1-score, sensitivity, specificity, and precision can be derived (Akşehir & Kiliç, 2022).

		Predicted Class	
		Negative	**Positive**
Actual Class	**Negative**	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Figure 7. Confusion matrix
Note. Adapted from (Akşehir & Kiliç, 2022)

Evaluation metrics, including accuracy, F1-score, sensitivity, specificity, and precision, are obtained from the confusion matrix, with their formulas listed in Equations 1 to 5. These metrics are used to evaluate the effectiveness of PV classification.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad [1]$$

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Senstivity}}{\text{Precision} + \text{Sensitivity}} \quad [2]$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad [3]$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad [4]$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad [5]$$

RESULTS AND DISCUSSIONS

The training process of all networks was performed on a system featuring an Intel (R) Core (TM) i3-5005U CPU operating at 2.00GHz, 8GB of RAM, and Windows 10 64-bit; the simulation and performance evaluation were carried out for the experimental configuration using MATLAB R2020a. The training time was considerably longer than with GPU-based implementations because all training was done on a CPU without GPU acceleration. Compared to shallower models like AlexNet, deeper architectures like ResNet-50 and Inception-V3 required noticeably longer convergence times during training. GPU acceleration would be necessary to deploy these models in real-time PV monitoring systems to increase computational efficiency and decrease latency, even though CPU-based training was adequate for experimental evaluation. Additionally, since deeper networks require more processing and storage power, memory usage is a crucial factor. This emphasises how, when thinking about real-world deployment, model accuracy must be balanced with scalability and resource requirements.

The performance of different CNN architectures, namely AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, Inception-V3, and GoogLeNet were evaluated through testing and comparison to determine the best-performing model. The data are split into two groups: training with 70% (976 images) and validation of 30% (418 images). The training settings were set to 20 iterations per epoch and 32 for batch size. Stochastic Gradient Descent with Momentum (SGDM) is used as a training solver, and the learning rate applied is 0.0001.

Comparison of Accuracy Performance

The architecture analysis is performed by comparing the accuracy of augmented and non-augmented data classifications. A non-augmented dataset is the raw data that is not going

through augmentation techniques of images, such as rotation or flipping. The dataset that has been augmented will undergo an augmentation process, which includes translation and rotation. The results for both analyses will be discussed in this part to identify which technique shows the best accuracy.

Without Data Augmentation

Figure 8 presents each network's accuracy when trained on the original dataset. VGG-19 exhibited outstanding performance, attaining an impressive accuracy of 92.58%. This result indicates that VGG-19 outperforms the other networks in accurately classifying non-augmented images. Despite performing relatively less favourably than the other networks, AlexNet still achieved a commendable accuracy above 90%. A comparative analysis between VGG-19 and AlexNet was conducted by the authors in Yu et al. (2023). VGG-19 exhibited lower training losses, suggesting a lower error rate than AlexNet during the training process.

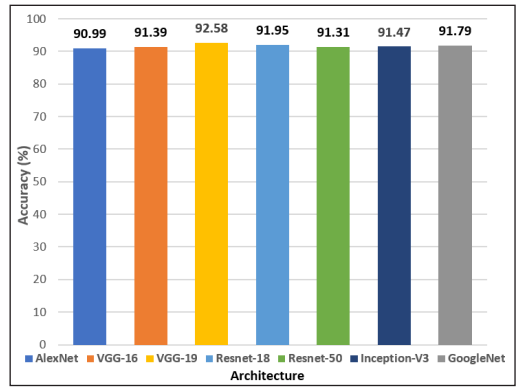


Figure 8. Comparison of accuracy between AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, Inception-V3 and GoogLeNet architectures (Without data augmentation)

A comparison analysis of training progress was made between the highest and lowest accuracy scores achieved. Figures 9 and 10 illustrate the training progress for both VGG-19 and AlexNet. The training cycle was set to 20 epochs with 640 iterations. Figure 9 illustrates the training progress of VGG-19 after the network has been trained. The accuracy initially starts at 30% and gradually increases, reaching 90% by the 100th iteration. The model maintains stability from the 230th iteration until it achieves a final accuracy of 92.58%. The corresponding training loss for VGG-19 is 0.2.

Figure 10 illustrates the training progress of AlexNet. The accuracy begins at 48% and undergoes a significant increase, reaching 90% by the 40th iteration. Afterwards, the model remains stable, resulting in a final accuracy of 90.99%. The training loss for AlexNet is slightly higher than that of VGG-19, with a difference of 0.6. Comparing the training progress of the two models, VGG-19 demonstrates a smoother increase in accuracy and lower training loss compared to AlexNet. These findings align with the analysis conducted by the authors in Tao (2022), which highlights VGG-19 as the top-performing model in terms of accuracy and training loss.

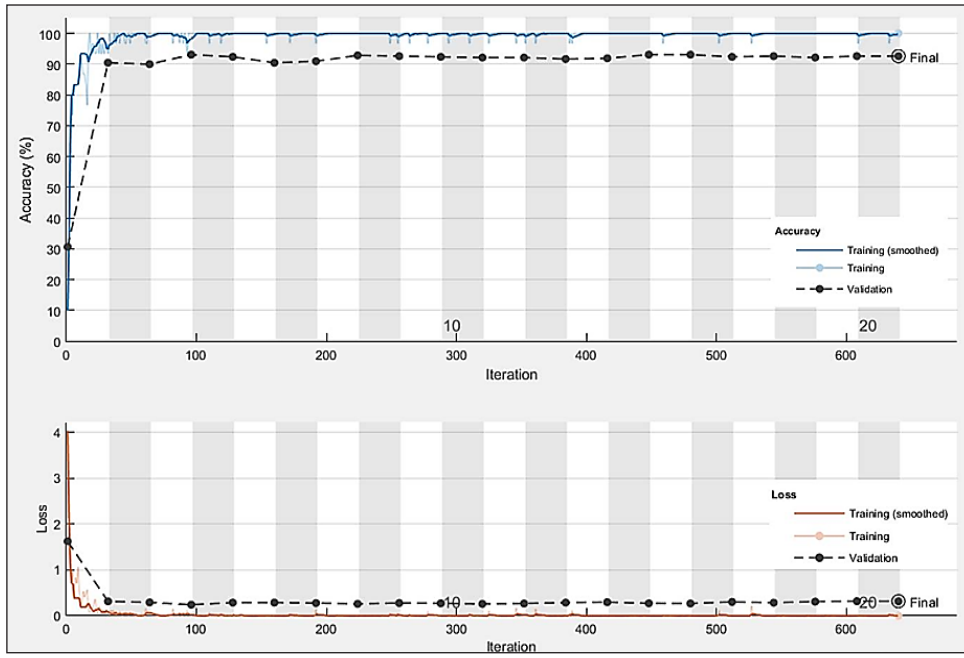


Figure 9. The training progress of VGG-19

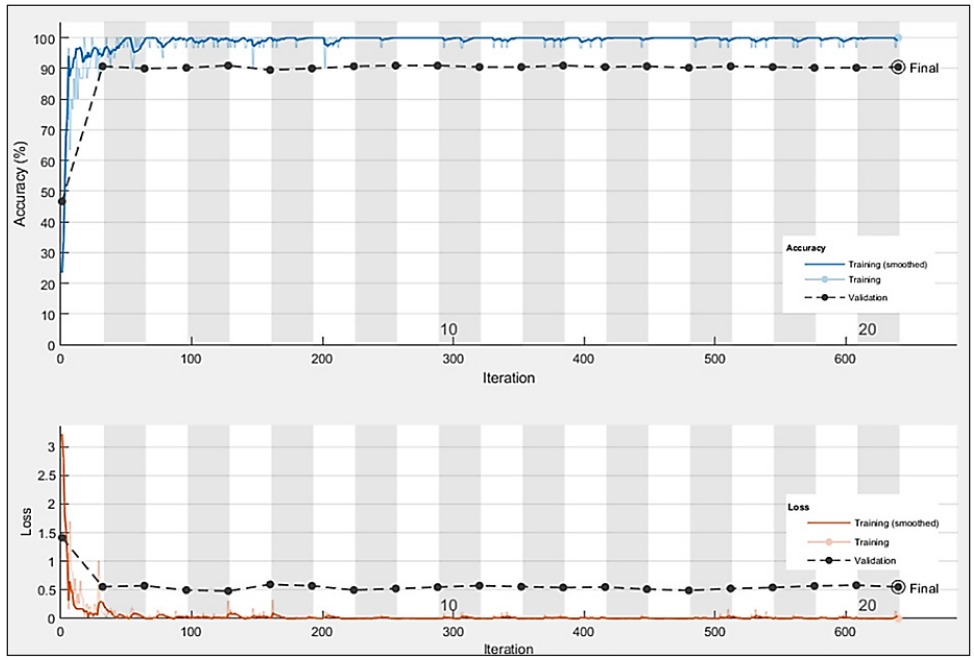


Figure 10. The training progress of AlexNet

With Data Augmentation

Figure 11 illustrates the accuracy comparison among different CNN architectures after dataset augmentation. ResNet-50, a DAG network, achieves the highest accuracy, demonstrating its superior deep feature extraction capability. Although GoogLeNet records the lowest accuracy at 98.33%, it still performs well, exceeding the 98% threshold. Notably, VGG-19, ResNet-18, and Inception-V3 represent both series and DAG networks that achieve identical accuracy levels, as shown in Figure 11.

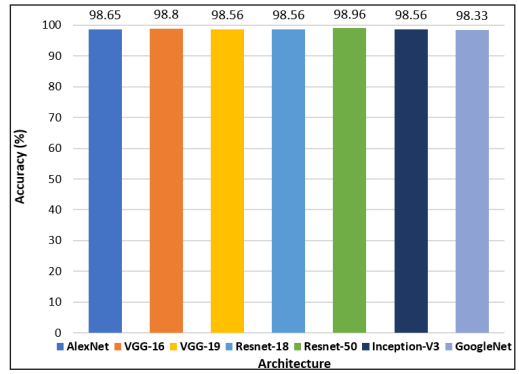


Figure 11. Comparison of accuracy between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, Inception-V3 and GoogLeNet architectures (With data augmentation)

Dataset augmentation significantly enhances model performance across all architectures, reinforcing its effectiveness in improving classification accuracy. Abu Salim et al. (2022) similarly reported accuracy improvements when the dataset size increased from 551 to 850 images. Additionally, in a Faster R-CNN implementation, ResNet-50 exhibited the highest performance, with accuracy rising from 63.2% (without augmentation) to 72.4% (with augmentation). These findings emphasise the benefits of data augmentation in both series and DAG networks, ensuring improved learning and robustness in PV defect classification. Figure 12 and Figure 14 depict the training progress of accuracy for the augmented dataset. ResNet-50's highest accuracy is achieved, starting at 33% and experiencing a rapid increase to 90% in less than 50 iterations. From the 170th iteration onwards, the accuracy stabilises at 98%, eventually reaching a final accuracy of 98.96%. Remarkably, the training loss is close to 0, indicating the outstanding performance of ResNet-50 in accurately classifying the images.

ResNet-50 demonstrated strong generalisation capability and was less prone to overfitting, largely due to its skip connections and batch normalisation layers (Negara et al., 2021). The skip connections enable direct feature transfer across layers, effectively addressing the vanishing gradient problem and allowing very deep networks to be trained without significant performance loss (Verma et al., 2024) The training loss approaching zero further confirms ResNet-50's effectiveness in accurately classifying PV defect images, while the convergence of training and validation losses at low values indicates minimal overfitting and stable learning behaviour (Zhu et al., 2020). Previous studies have also shown that ResNet-50 surpasses other architectures in PV module fault detection, achieving a maximum accuracy of 98.22%. Owing to its superior performance in image classification and pattern recognition tasks, ResNet-50 is considered one of the most suitable architectures for PV defect detection applications (Alsudi & Al-Aubidy, 2023).

The results obtained from the augmented dataset were used as a reference to analyse misclassification patterns, and a confusion matrix analysis was performed to evaluate class-wise performance and to identify misclassifications among the PV defect classes. Figure 13 shows the confusion matrix of ResNet-50 for 5 types of defects. Referring to Figure 13, the Cracks class achieved a perfect score, with all 23 images correctly classified and no misclassifications. The Delamination class recorded 32 correct predictions out of 33, with 1 image misclassified as Glass Breakage. Similarly, both Encapsulation Discolouring and Gridline Corrosion each had 1 misclassified image. The Glass Breakage class demonstrated strong performance with 306 correctly classified images, while 3 images were misclassified, distributed between Glass Breakage and Gridline Corrosion. These results reflect the model's high accuracy overall, with only minor misclassification occurring across a few classes.

Figure 14 illustrates the training progress of GoogLeNet, which starts with a relatively low initial accuracy of 68% but gradually improves, reaching a final accuracy of 98.33%. The training process stabilises around the 190th iteration, marking the saturation point. Similarly, ResNet-50 exhibits a strong learning curve, with its training loss nearing zero, highlighting the positive impact of data augmentation on model performance.

Cracks and Delamination classes in Figure 15 achieved excellent results in their respective defect prediction tasks. Cracks class achieved a perfect score of 23, while Delamination recorded 33 correct predictions, with no misclassifications in either class. However, Encapsulation Discolouring, Glass Breakage, and Gridline Corrosion classes experienced some misclassification issues. Encapsulation Discolouring had 5 images misclassified as Glass Breakage, indicating the model was unable to distinguish between these two defects. The Glass Breakage category contains one misidentification because the model mistakenly labelled a Gridline Corrosion image as Glass Breakage. The Gridline Corrosion category received a score of 18 out of 19 because one image was misclassified. GoogLeNet model achieved high accuracy in detecting Cracks and Delamination, but struggled to identify Encapsulation Discolouring correctly, which indicates the model had a problem distinguishing between visually similar defect types.

The small performance gap observed between certain architectures, where ResNet-50 at 98.96% and GoogLeNet at 98.33%, may partly be attributed to the evaluation approach, where results were averaged across three runs. This averaging tends to smooth out minor fluctuations between runs, leading to reduced variation in the final reported scores. As a result, the differences in accuracy appear smaller and may not fully reflect run-to-run variability. Although statistical significance tests such as standard deviation or confidence intervals were not reported, the models were run three times and showed highly consistent results with minimal variation. Therefore, the reported averages are considered reliable for comparative analysis, which aligns with the objective of this work.

These findings underscore the effectiveness of augmentation in enhancing classification accuracy across different architectures. While both GoogLeNet and ResNet-50, representing DAG networks, show significant improvements, ResNet-50 outperforms with higher accuracy and lower training loss, confirming its superior deep feature extraction capability. Despite GoogLeNet achieving a slightly lower accuracy, it still demonstrates robust performance, reinforcing the advantages of DAG networks in transfer learning for PV defect classification.

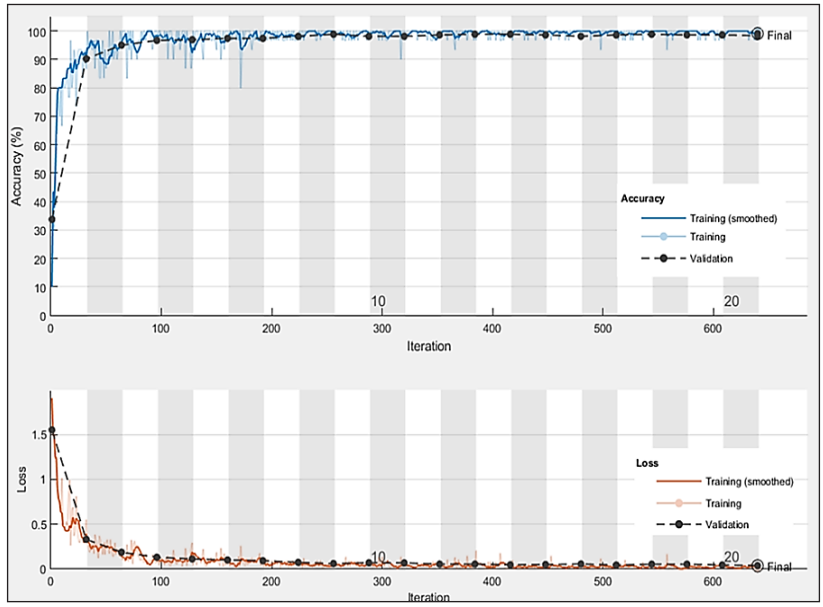


Figure 12. The training progress of ResNet-50

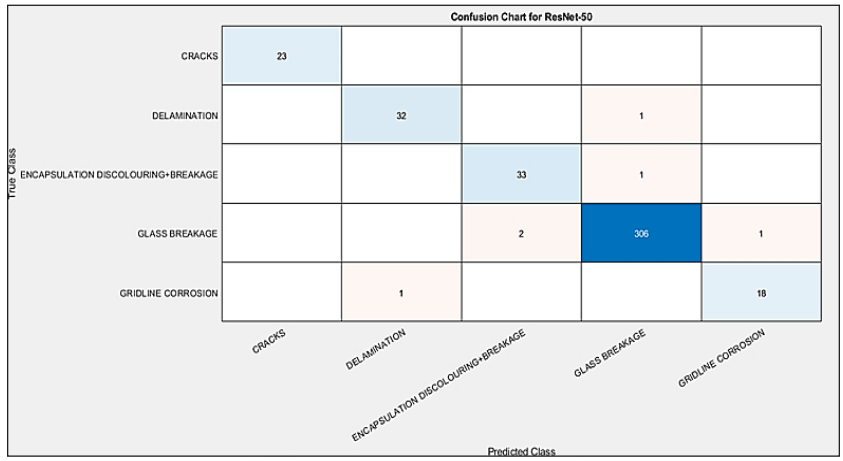


Figure 13. Confusion Matrix of ResNet-50

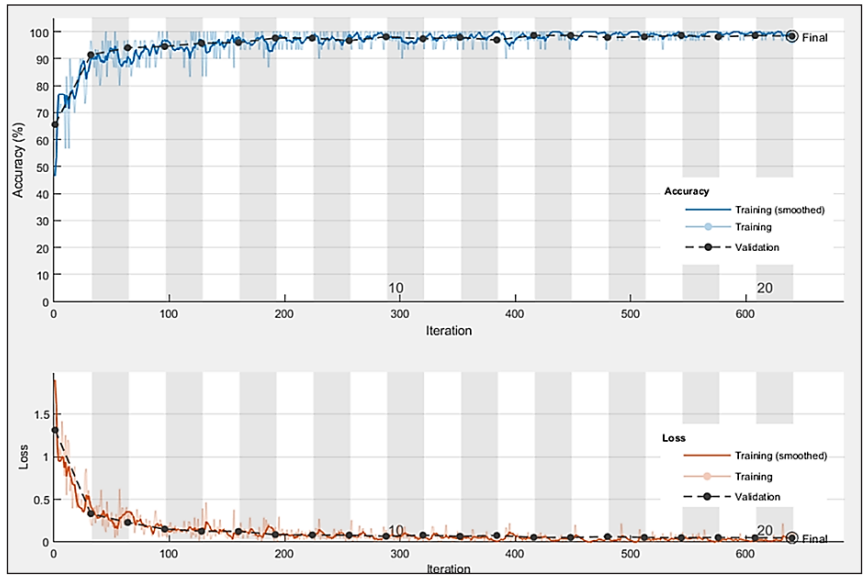


Figure 14. The training progress of GoogLeNet

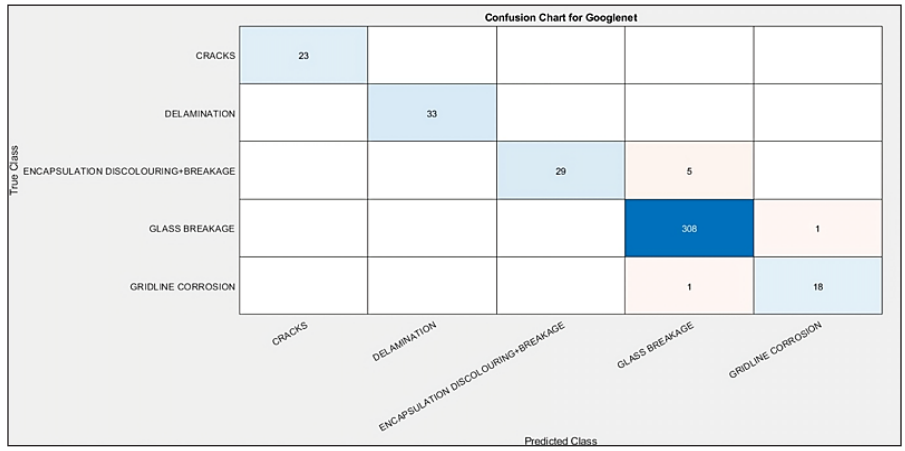


Figure 15. Confusion Matrix of GoogLeNet

Comparison of Performance Metrics

Performance metrics, including precision, F1-score, specificity, and sensitivity, are used to evaluate the effectiveness of the network. In this study, the performance of augmented and non-augmented datasets was compared to determine which approach produced superior results. The performance metrics for both analyses are presented in Figures 16 and 17. The augmented dataset was found to consistently outperform the non-augmented approach, significantly enhancing these metrics. However, strong results were achieved in terms of specificity for both approaches, with only a minimal difference in scores.

Without Data Augmentation

Figure 17 presents a comparison of performance metrics for different networks when the dataset is augmented. The F1-score analysis indicates that GoogLeNet performs well, achieving 74.75%, despite having a longer processing time than other networks, as reported by Rattanawin et al. (2023). In contrast, AlexNet demonstrates the lowest F1-score at 68.77%, a result consistent with findings by Tyassari et al. (2022), who noted its lower F1-score but emphasised its advantage of faster training time. VGG-19 achieves the highest sensitivity, whereas Inception-V3 records the lowest sensitivity score. In terms of specificity, AlexNet excels in accurately predicting the negative class, with most networks demonstrating strong performance in this metric. Regarding precision, ResNet-18 outperforms other networks with a score of 88.55%, while VGG-19 exhibits the lowest precision. These results highlight the varying strengths of series and DAG networks in transfer learning, where DAG networks such as ResNet-18 and GoogLeNet demonstrate superior feature extraction capabilities, while series networks like AlexNet and VGG-19 show trade-offs between speed and accuracy.

Comparison of Performance Analysis Between Data Augmentation and Without Data Augmentation

A comparison of performance metrics between augmented and non-augmented datasets has been conducted, revealing that the augmented dataset consistently outperforms the non-augmented dataset. As shown in Figures 17 and 18, the F1-score for the non-augmented dataset ranges from 68.77% to 74.75%, sensitivity from 68.32% to 73.14%, specificity from 94.24% to 95.26%, and precision from 55.42% to 88.55%. In contrast, the augmented dataset achieves significantly improved results, with F1-scores ranging from 96.93% to 97.89%, sensitivity from 95.72% to 98.10%, specificity from 98.85% to 99.53%, and precision from 96.77% to 98.65%. Notably, specificity exhibits strong performance in both cases compared to other metrics such as F1-score, sensitivity, and precision. This variation in performance can be attributed to the structural differences between series and DAG networks, as each network's unique layer arrangement influences its ability to optimise certain parameters (Shorten & Khoshgoftaar, 2019; Zhou et al., 2020).

Comparison of Performance Analysis Between CNN Architectures and EfficientNetV2-S

The comparison between seven CNN architectures and EfficientNetV2-S was conducted based on performance metrics. EfficientNetV2 was selected over EfficientNetV1 for the analysis because it offers faster training speed, improved accuracy, and reduced computational complexity, making it more suitable for efficient PV defect image classification (Kamal et al., 2023). The results indicate that ResNet-50, VGG-19, and VGG-16 outperformed EfficientNetV2-S in terms of accuracy, F1-score, sensitivity, and precision.

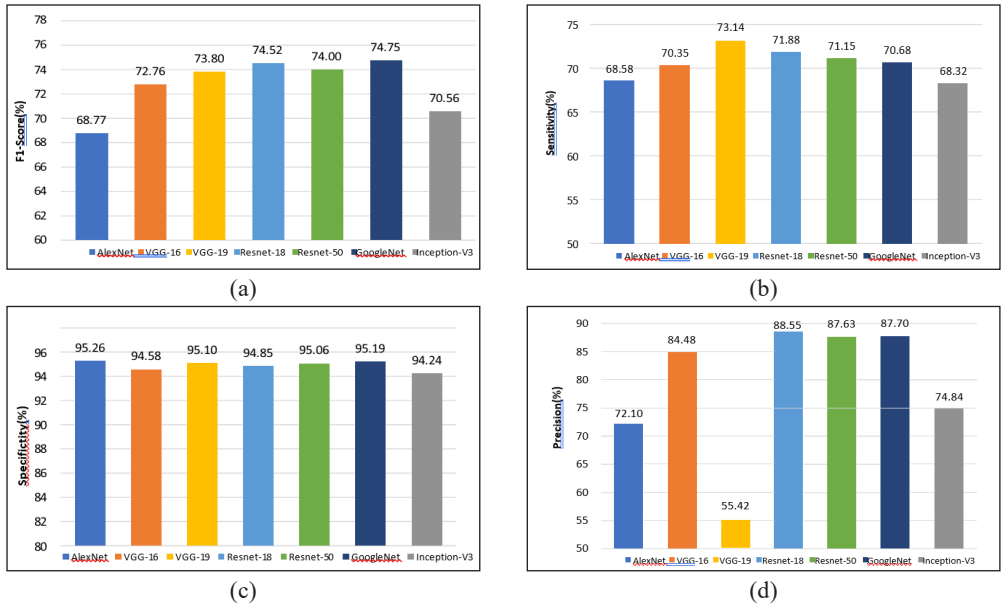


Figure 16. Comparison of Performance Metrics: (a) F1-Score; (b) Sensitivity; (c) Specificity; and (d) Precision between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, GoogLeNet and Inception-V3 Architectures (Without Data Augmentation)

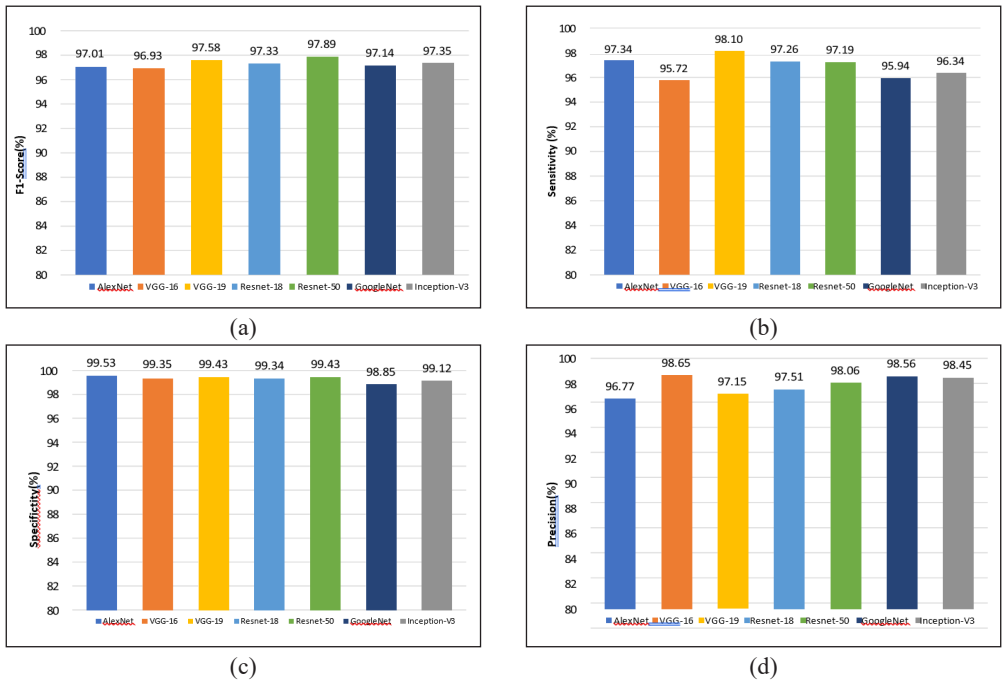


Figure 17. Comparison of Performance Metrics; (a) F1-Score; (b) Sensitivity; (c) Specificity; and (d) Precision between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, GoogLeNet and Inception-V3 Architecture (With data augmentation)

Figure 18 presents the comparison of the highest performance metric scores achieved by the CNN models against EfficientNetV2-S. As shown in Figure 18(a), ResNet-50 achieved a higher accuracy of 98.96% compared to 93.36% for EfficientNetV2-S. Similarly, Figure 18(b) demonstrates that ResNet-50 outperformed in F1-score with 97.89% against 93.69%. In terms of sensitivity, Figure 18(c) shows VGG-19 achieving 98.10% compared to 93.36% for EfficientNetV2-S. Finally, Figure 18(d) illustrates that VGG-16 achieved superior precision with 98.65%, outperforming EfficientNetV2-S at 93.75%.

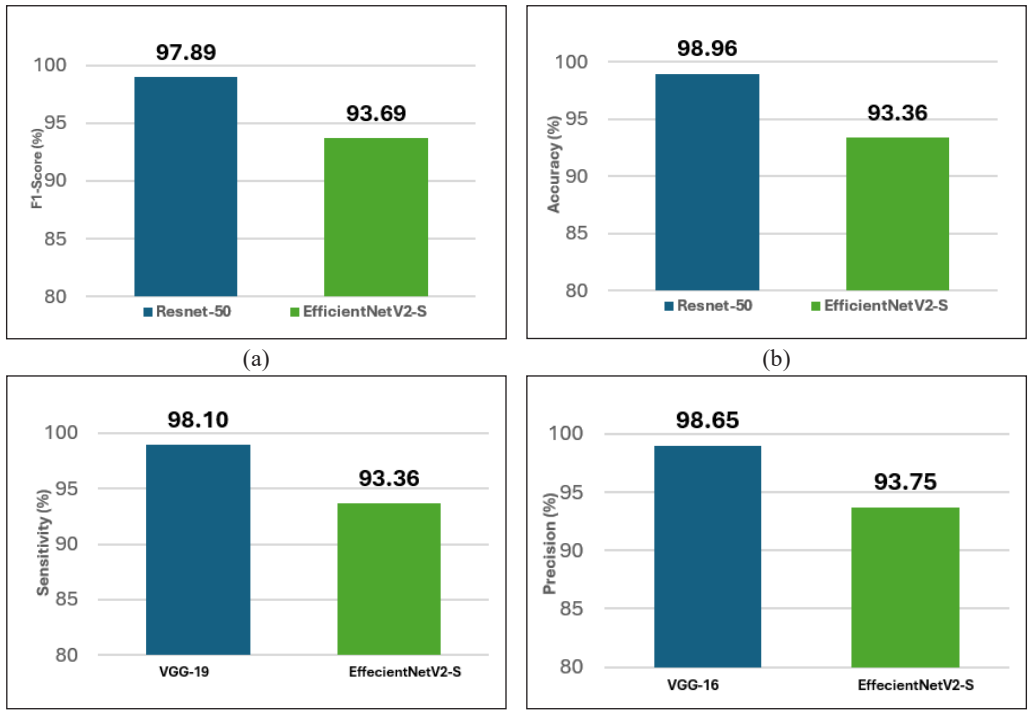


Figure 18. Comparison of Performance Metrics: (a) F1-Score; (b) Accuracy; (c) Sensitivity; and (d) Precision between CNN architecture and EfficientNetV2-S (With data augmentation)

CONCLUSIONS

The results indicated that the augmented dataset consistently outperformed the non-augmented dataset. The analysis of seven CNN architectures, namely AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, GoogLeNet, and Inception-V3 demonstrate varying performance across different evaluation metrics. ResNet-50, a DAG network, achieved the highest classification accuracy and the best F1-score, highlighting its strength in deep feature extraction and residual learning. AlexNet, a series network, excelled in specificity, while VGG-19 demonstrated superior sensitivity, and VGG-16 achieved the highest precision. These findings align with the structural differences between series and DAG

networks, where series networks perform well in straightforward feature extraction, while DAG networks leverage deeper architectures for enhanced learning. In addition to CNN-based architectures, this study compared EfficientNetV2-S to demonstrate the potential of alternative deep learning methods. Future work could further explore advanced approaches such as Vision Transformers or self-supervised learning, as well as multi-modal datasets such as RGB and infrared, to address challenges like subtle defect detection under varying conditions and enhance PV monitoring.

Although ResNet-50 performed the best in terms of accuracy and F1-score, its computational complexity and memory constraints may make it difficult to use in real-time PV inspection systems. Inference speed and consumption of energy are important considerations in resource-constrained situations, such as embedded devices or edge computing systems frequently employed in solar farm monitoring. Real-time applications may benefit more from simpler architectures (like AlexNet or VGG-16), which may provide faster inference at the cost of slightly lower accuracy. Therefore, a trade-off exists between maximising classification accuracy and ensuring operational feasibility in large-scale PV monitoring systems. Future work should investigate lightweight models or model compression techniques to balance accuracy with deployment efficiency.

Finally, while it is true that ResNet-50's superior performance can be anticipated due to its residual learning capabilities, the novelty of this work lies in providing a broader comparative study across seven CNN architectures for PV defect classification. Unlike prior studies that typically evaluate only a few models, this research systematically compares Series and DAG networks using multiple performance metrics, highlighting not only overall accuracy but also class-wise strengths such as sensitivity, precision, and specificity. This provides deeper insights into which architectures are more reliable for defect types, offering practical guidance for selecting models in future PV defect classification studies.

ACKNOWLEDGEMENT

Authors acknowledge the Universiti Teknologi MARA for funding under the Geran Penyelidikan SRC (600-RMC/SRC/5/3(032/2020).

LIST OF ABBREVIATIONS

AI : Artificial Intelligence
C-CE : Commercial Counter Electrode
CE : Counter Electrode
CNN : Convolutional Neural Network
DAG : Directed Acyclic Graph
DSSC : Dye-Sensitised Solar Cells

EL : Electroluminescence
GFDI : Ground Fault Detection Interrupters
H-CE : Homemade Counter Electrode
OCPD : Overcurrent Protection Devices
PV : Photovoltaic
RESNET: Residual Neural Network
RGB : Red Green Blue
VGG : Visual Geometry Group
ViT : Vision Transformer

REFERENCES

Ahmad, Abu Salim, S., Zakaria, N., Mokhtar, N., Mostafa, S. A., & Abdulkadir, S. J. (2022). Data augmentation on intra-oral images using image manipulation techniques. In the *2022 International Conference on Digital Transformation and Intelligence (ICDI)* (pp. 117-120). IEEE. https://doi.org/10.1109/ICDI57181.2022.10007158

Akşehir, Z. D., & Kiliç, E. (2022). How to handle data imbalance and feature selection problems in CNN-based stock price forecasting. *IEEE Access, 10*, 31297-31305. https://doi.org/10.1109/ACCESS.2022.3160797

Alsudi, I. M., & Al-Aubidy, K. M. (2023). Deep learning-based fault classification and location for photovoltaic systems. In *2023 16th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 760-764). IEEE. https://doi.org/10.1109/DeSE60156.2023.10393961

Altinkaya, C., Atli, A., Atilgan, A., Salimi, K., & Yildiz, A. (2020). Facile fabrication of low-cost, low-temperature carbon-based counter electrode with an outstanding fill factor of 73% for dye-sensitised solar cells. *International Journal of Energy Research, 44*(12), 9570-9581. https://doi.org/10.1002/er.5174

Alves, R. H. F., de Deus Junior, G. A., Marra, E. G., & Lemos, R. P. (2021). Automatic fault classification in photovoltaic modules using convolutional neural networks. *Renewable Energy, 179*, 502-516. https://doi.org/10.1016/j.renene.2021.07.070

Anilkumar, K. K., Manoj, V. J., & Sagi, T. M. (2021). Automated detection of leukaemia by pretrained deep neural networks and transfer learning: A comparison. *Medical Engineering & Physics, 98*, 8-19. https://doi.org/10.1016/j.medengphy.2021.10.006

Appiah, A. Y., Zhang, X., Ayawli, B. B. K., & Kyeremeh, F. (2019). Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques. *International Journal of Photoenergy, 2019*, Article 6953530. https://doi.org/10.1155/2019/6953530

Arnob, A. S., Kausik, A. K., Islam, Z., Khan, R., & Rashid, A. B. (2025). Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, Inception V3, ResNet, and a custom CNN model. *Hybrid Advances, 7*, Article 100440. https://doi.org/10.1016/j.hybadv.2025.100440

Arvinth, S. (2024). *VGG16 vs. VGG19 vs. EfficientNetB5: Which model is best for your project?* Medium. https://medium.com/

Azis, S. F. S., Shahbudin, S., Kassim, M., Mohamad, R., & Rahman, F. Y. A. (2022). Photovoltaic module defects classification analysis using DenseNet architecture. In *2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA)* (pp. 1-5). IEEE. https://doi.org/10.1109/ISIEA54517.2022.9873681

Bagaskara, A., & Suryanegara, M. (2021). Evaluation of VGG-16 and VGG-19 deep learning architectures for classifying people with dementia. In *2021 4th International Conference on Computer and Informatics Engineering (IC2IE)* (pp. 1-4). IEEE. https://doi.org/10.1109/IC2IE53219.2021.9649132

Bakhshayesh, P. R., Ejtehadi, M., Taheri, A., & Behzadipour, S. (2022). The effects of data augmentation methods on the performance of human activity recognition. In *2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICSPIS56952.2022.10043959

Balani, H., Darji, P., Sanghvi, R., Kalavadiya, A., & Parekh, R. (2021). Performance analysis of a single junction crystalline solar cell using 1D drift diffusion modelling. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 339-346). IEEE. https://doi.org/10.1109/ICESC51422.2021.9533003

Basnet, B., Chun, H., & Bang, J. (2020). An intelligent fault detection model for fault detection in photovoltaic systems. *Journal of Sensors, 2020*, Article 6960328. https://doi.org/10.1155/2020/6960328

Chindarkkar, A., Priyadarshi, S., Shiradkar, N. S., Kottantharayil, A., & Velmurugan, R. (2020). Deep learning-based detection of cracks in electroluminescence images of fielded PV modules. In *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)* (pp. 1612-1616). IEEE. https://doi.org/10.1109/PVSC45281.2020.9300615

Cipriani, G., D'Amico, A., Guarino, S., Manno, D., Traverso, M., & Di Dio, V. (2020). Convolutional neural network for dust and hotspot classification in PV modules. *Energies, 13*(23), Article 6357. https://doi.org/10.3390/en13236357

Drir, N., Mellit, A., & Bettayeb, M. (2025). A novel ensemble CNN framework with weighted feature fusion for fault diagnosis of photovoltaic modules using thermography images. *IEEE Journal of Photovoltaics, 15*(1), 146-154. https://doi.org/10.1109/JPHOTOV.2024.3492283

Faheem, M. B., Khan, B., Feng, C., Subhani, W. S., Mabrouk, S., Sayyad, M. H., Yildiz, A., Zhang, W.-H., & Qiao, Q. (2022). Van der Waals epitaxial growth for high-performance organic-free perovskite solar cell: Experimental and theoretical insights. *Advanced Materials Interfaces, 9*(20), Article 2200421. https://doi.org/10.1002/admi.202200421

Feng, C., Zhang, H., Wang, S., Li, Y., Wang, H., & Yan, F. (2019). Structural damage detection using a deep convolutional neural network and transfer learning. *KSCE Journal of Civil Engineering, 23*(10), 4493-4502. https://doi.org/10.1007/s12205-019-0437-z

Hussein, L. A., Ayadi, O., & Fathi, M. (2021). Performance comparison for sun-tracking mechanism photovoltaic (PV) and concentrated photovoltaic (CPV) solar panels with fixed system PV panels in Jordan. In *2021 12th International Renewable Energy Conference (IREC)* (pp. 1-8). IEEE. https://doi.org/10.1109/IREC51415.2021.9427808

Kamal, M. R. M., Shahbudin, S., & Rahman, F. Y. A. (2023). Photovoltaic (PV) module defect image classification analysis using EfficientNetV2 architectures. In *2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC)* (pp. 236-241). IEEE. https://doi.org/10.1109/ICSGRC57744.2023.10215491

Kellil, N., Aissat, A., & Mellit, A. (2023). Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions. *Energy, 263*, Article 125902. https://doi.org/10.1016/j.energy.2022.125902

Kocak, Y., Atli, A., Atilgan, A., & Yildiz, A. (2019). Extraction method-dependent performance of bio-based dye-sensitised solar cells (DSSCs). *Materials Research Express, 6*(9), Article 095512. https://doi.org/10.1088/2053-1591/ab2ef7

Li, B., & He, Y. (2018). An improved ResNet based on the adjustable shortcut connections. *IEEE Access, 6*, 18967-18974. https://doi.org/10.1109/ACCESS.2018.2814605

Li, J., Huang, H., Hu, F., & Ou, Y. (2022). Classification of ceramics based on an improved AlexNet convolutional neural network. In *2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, and International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (pp. 1-8). IEEE. https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9927857

Li, X., Li, W., Yang, Q., Yan, W., & Zomaya, A. Y. (2020). An unmanned inspection system for multiple defect detection in photovoltaic plants. *IEEE Journal of Photovoltaics, 10*(2), 568-576. https://doi.org/10.1109/JPHOTOV.2019.2955183

Luu, T. T., & Huynh, D. A. (2025). A ResNet-based deep reinforcement learning framework using soft actor-critic for remaining useful life prediction of rolling bearings. *Results in Engineering, 25*, Article 106739. https://doi.org/10.1016/j.rineng.2025.106739

Meng, L., Wang, Y., Wang, H., Tong, X., Sun, Z., & Cai, Z. (2023). Task offloading optimisation mechanism based on a deep neural network in an edge-cloud environment. *Journal of Cloud Computing, 12*(1), Article 76. https://doi.org/10.1186/s13677-023-00450-6

Negara, B. S., Satria, E., Sanjaya, S., & Santoso, D. R. D. (2021). ResNet-50 for classifying Indonesian batik with data augmentation. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICOTEN52080.2021.9493488

Nylund, S., & Barbari, Z. (2019). *Study of defects in PV modules: UV fluorescence and thermographic photography for photovoltaics (PV) field application* (master's thesis). Mälardalen University, Sweden.

Özel, K., & Yıldız, A. (2022). Comprehensive understanding of the role of emitter layer thickness for metal oxide–semiconductor–based solar cells. *IEEE Journal of Photovoltaics, 12*(1), 251-258. https://doi.org/10.1109/JPHOTOV.2021.3119612

Pangilinan, J. R., Legaspi, J., & Linsangan, N. (2022). InceptionV3, ResNet50, and VGG19 performance comparison on tomato ripeness classification. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 619-624). IEEE. https://doi.org/10.1109/ISRITI56927.2022.10052920

Pavlitskaya, S., Oswald, J., & Zöllner, J. M. (2022). Measuring overfitting in convolutional neural networks using adversarial perturbations and label noise. In *2022 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1551-1559). IEEE. https://doi.org/10.1109/SSCI51031.2022.10022094

Pierdicca, R., Malinverni, E. S., Piccinini, F., Paolanti, M., Felicetti, A., & Zingaretti, P. (2018). Deep convolutional neural network for automatic detection of damaged photovoltaic cells. *The International*

Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2, 893–900. https://doi.org/10.5194/isprs-archives-XLII-2-893-2018

Rahman, M. R. U., Chen, H., & Xi, W. (2019). U-Net-based defects inspection in photovoltaic electroluminescence images. In *2019, IEEE International Conference on Big Knowledge (ICBK)* (pp. 215-220). IEEE. https://doi.org/10.1109/ICBK.2019.00036

Rahman, M. R., Tabassum, S., Haque, E., Nishat, M. M., Faisal, F., & Hossain, E. (2021). CNN-based deep learning approach for micro-crack detection of solar panels. In *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE. https://doi.org/10.1109/STI53101.2021.9732592

Rao, S., Muniraju, G., Tepedelenlioglu, C., Srinivasan, D., Tamizhmani, G., & Spanias, A. (2021). Dropout and pruned neural networks for fault classification in photovoltaic arrays. *IEEE Access, 9*, 120034-120042. https://doi.org/10.1109/ACCESS.2021.3108684

Rattanawin, P., Pakinsee, T., & Songmuang, P. (2023). A GoogLeNet performance approach for COVID-19 detection using chest X-ray images. In *2023 15th International Conference on Knowledge and Smart Technology (KST)* (pp. 1-5). IEEE. https://doi.org/10.1109/KST57286.2023.10086817

Salih, N., Ksantini, M., Hussein, N., Ben Halima, D., Abdul Razzaq, A., & Ahmed, S. (2023). Prediction of ROP zones using deep learning algorithms and the voting classifier technique. *International Journal of Computational Intelligence Systems, 16*(1), Article 86. https://doi.org/10.1007/s44196-023-00268-9

Salimi, K., Atilgan, A., Aydin, M. Y., Yildirim, H., Celebi, N., & Yildiz, A. (2019). Plasmonic mesoporous core-shell Ag-Au@TiO_2 photoanodes for efficient light harvesting in dye-sensitised solar cells. *Solar Energy, 193*, 820-827. https://doi.org/10.1016/j.solener.2019.10.039

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), Article 60. https://doi.org/10.1186/s40537-019-0197-0

Siuly, S., Khare, S. K., Kabir, E., Sadiq, M. T., & Wang, H. (2024). An efficient Parkinson's disease detection framework: Leveraging time-frequency representation and AlexNet convolutional neural network. *Computers in Biology and Medicine, 174*, Article 108462. https://doi.org/10.1016/j.compbiomed.2024.108462

Swarup, C., Singh, K. U., Kumar, A., Pandey, S. K., & Singh, T. (2023). Brain tumour detection using CNN, AlexNet, and GoogLeNet ensembling learning approaches. *Electronic Research Archive, 31*(5), 2864-2885. https://doi.org/10.3934/era.2023146

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 6105-6114). PMLR.

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 10096-10106). PMLR.

Tao, Y. (2022). Image style transfer based on the VGG neural network model. In *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 1475-1482). IEEE. https://doi.org/10.1109/AEECA55500.2022.9918891

Tella, H., Hussein, A., Rehman, S., Liu, B., Balghonaim, A., & Mohandes, M. (2025). Solar photovoltaic panel cells' defects classification using deep learning ensemble methods. *Case Studies in Thermal Engineering, 66*, Article 105749. https://doi.org/10.1016/j.csite.2024.105749

Triyadi, A. B., Bustamam, A., & Anki, P. (2022). Deep learning in image classification using VGG-19 and residual networks for cataract detection. In *2022 2nd International Conference on Information Technology and Education (ICITE)* (pp. 293-297). IEEE. https://doi.org/10.1109/ICITE54466.2022.9759886

Tyassari, W., Jusman, Y., Riyadi, S., & Sulaiman, S. N. (2022). Classification of cervical precancerous cells of ThinPrep images based on the deep learning model AlexNet and InceptionV3. In *2022, IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 276-281). IEEE. https://doi.org/10.1109/CSNT54456.2022.9787658

Verma, H., Siruvuri, S. D. V. S. S. V., & Budarapu, P. R. (2024). A machine learning-based image classification of silicon solar cells. *International Journal of Hydromechatronics, 7*(1), 49-66. https://doi.org/10.1504/IJHM.2024.137958

Wang, S., & Feng, Z. (2024). Multi-sensor fusion rolling bearing intelligent fault diagnosis based on VMD and ultra-lightweight GoogLeNet in industrial environments. *Digital Signal Processing, 145*, Article 104306. https://doi.org/10.1016/j.dsp.2023.104306

Xia, H., Zhang, Y., Yang, M., & Zhao, Y. (2020). Visual tracking via deep feature fusion and correlation filters. *Sensors, 20*(12), Article 3370. https://doi.org/10.3390/s20123370

Xu, W., Zeng, T. H., & Shalaginov, M. (2022). Class activation mapping-enhanced AlexNet convolutional neural networks for early diagnosis of Alzheimer's disease. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2550-2555). IEEE. https://doi.org/10.1109/BIBM55620.2022.9994868

Yu, F., Lu, T., & Xue, C. (2023). Deep learning-based intelligent apple variety classification system and model interpretability analysis. *Foods, 12*(4), Article 885. https://doi.org/10.3390/foods12040885

Zhang, Y., Zhang, Y., Wang, L., Dong, X., Li, Y., Sun, H., & Yang, X. (2022). Crack detection of electrical equipment based on an improved GoogLeNet. In the *2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICARCE55724.2022.10046575

Zhu, W., & Mousavi, S. M. (2020). Seismic signal augmentation to improve generalisation of deep neural networks. In G. C. B. (Ed.), *Advances in Geophysics* (Vol. 61, pp. 181-222). Elsevier. https://doi.org/10.1016/bs.agph.2020.08.004

Zou, S., Chen, W., & Chen, H. (2020). Image classification model based on deep learning in the Internet of Things. *Wireless Communications and Mobile Computing, 2020*, Article 6677907. https://doi.org/10.1155/2020/6677907

Zulkarnain, I., Nurmalasari, R. R., & Azizah, F. N. (2022). Table information extraction using data augmentation on deep learning and image processing. In *2022 16th International Conference on Telecommunication Systems, Services, and Applications (TSSA)* (pp. 1-6). IEEE. https://doi.org/10.1109/TSSA56819.2022.10063909

Zyout, I., & Oatawneh, A. (2020). Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks. In *2020 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 3-6). IEEE. https://doi.org/10.1109/ASET48392.2020.9118382

Pertanika J. Sci. & Technol. 34 (4): 2519 - 2543 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Orchestration Network Hierarchy and Communication Patterns Architecture Analysis with Ranks for IoT Devices

Nor Syazwani Mohd Pakhrudin[1], Murizah Kassim[1,2]*, and Azlina Idris[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

ABSTRACT

Fog computing has gained prominence as a transformative paradigm for the Internet of Things (IoT), overcoming the limitations of traditional cloud computing by enabling decentralised processing, storage, and communication at the network edge. This study delves into the orchestration network hierarchy and communication patterns within fog computing environments, utilising the Routing Protocol for Low-Power and Lossy Networks (RPL). The aim is to enhance network performance and reliability through hierarchical orchestration and dynamic rank-based device management. Employing the NetSim simulation tool, this paper not only constructs and analyses a network topology incorporating RPL but also evaluates its impact on various performance metrics, including packet loss, packet delivery ratio (PDR), jitter, delay, and throughput. The results indicate that while RPL contributes to network stability and reliability, it also introduces notable overhead, resulting in increased jitter, delay, and reduced throughput. Packet loss initially spikes and fluctuates periodically, and PDR stabilises around 0.8 before declining. Delay increases in most applications with RPL, with App1 at 9083 µs, App2 at 13383 µs, App3 at 9345 µs, and App4 remaining constant at 37.88 µs. This study provides a comprehensive overview of the impact of RPL on network performance and offers practical insights for optimising communication protocols and hierarchical orchestration, which are crucial for improving the reliability and scalability of IoT networks in real-world applications. These insights can directly inform the design of robust and scalable IoT systems, improving real-world deployments in smart cities, healthcare, and industrial automation.

Keywords: Fog computing, Internet of Things, network hierarchy, orchestration, Routing Protocol for Low-Power and Lossy Networks (RPL)

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 29 April 2025
Published: 03 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.27

E-mail addresses:
norsyazwanipakhrudin@gmail.com (Nor Syazwani Mohd Pakhrudin)
murizah@uitm.edu.my (Murizah Kassim)
azlina831@uitm.edu.my (Azlina Idris)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Fog computing transforms IoT by decentralising processing and storage to the network edge, enhancing real-time processing and reducing latency (Apat et al., 2023; Yusof et al., 2023). This shift emphasises the importance of efficient data management and communication protocols (Beniwal & Singhrova, 2022). The Routing Protocol for Low-Power and Lossy Networks (RPL) is crucial for optimising routing in constrained IoT environments and enhancing network performance through effective data flow and device interactions (Ching et al., 2021). Hierarchical orchestration in fog computing assigns roles to IoT devices based on ranks (Varshney et al., 2020), optimising resource allocation (Pereira et al., 2021), scalability (Coelho & Nogueira, 2021), and fault tolerance (Sharma & Gupta, 2023). This approach ensures critical data is efficiently processed and transmitted, improves energy efficiency by minimising unnecessary transmissions (Kumar et al., 2022), and enhances network resilience against node failures. Hierarchical orchestration addresses scalability, reliability, and energy efficiency challenges (Suliman et al., 2021), facilitating robust and efficient fog-based IoT systems (Ahmad et al., 2023).

The Routing Protocol for Low-Power and Lossy Networks (RPL) is a distance-vector routing protocol specifically designed for energy-constrained and bandwidth-limited IoT networks. The RPL is crucial for optimising performance and reliability in IoT fog computing (Ekpenyong et al., 2022). RPL constructs a Destination-Oriented Directed Acyclic Graph (DODAG), in which each node is assigned a rank based on its logical distance to a root node or sink, facilitating efficient routing and reducing latency (Miller et al., 2024). The rank helps organise the routing paths and maintain a loop-free topology. DODAG Information Object (DIO) messages are key control packets used by RPL to disseminate routing information and allow neighbouring nodes to compute and update their ranks accordingly. These messages include parameters such as node rank, routing metrics, and network configuration options. RPL's ability to adapt to dynamic conditions, such as node mobility and varying network states, is critical to maintaining network robustness and efficiency (Mahyoub et al., 2021). However, traditional network protocols, including RPL, face scalability (Hassani et al., 2021), adaptability, and resource utilisation challenges within decentralised, resource-constrained environments. Current methods often rely on static configurations and need advanced hierarchical orchestration, leading to inefficiencies such as increased latency, packet loss, and higher energy consumption (Shaik & Baskiyar, 2022). Improvements in network orchestration and communication patterns are needed to address these issues (Costa et al., 2022). Enhancing scalability, fault tolerance, and energy efficiency through dynamic device ranking and role assignments can optimise data transmission, reduce latency, and ensure reliable IoT system performance in diverse environments.

Recent research highlights the need for sophisticated hierarchical structures in fog computing to manage device interactions effectively (Baranwal & Vidyarthi, 2021). Despite advancements, adapting to dynamic network conditions and ensuring real-time processing remain challenges. Recent studies focus on improving hierarchical network designs (Banyal et al., 2021) and adaptive algorithms to enhance IoT networks (Khan et al., 2020). Federated fog computing has shown benefits over traditional fog and cloud frameworks by reducing latency and computational resource usage. It can lower CPU usage to 20% and reduce latency to as low as 300ms (Huaranga-Junco et al., 2024). Additionally, space-air-ground integrated IoT networks are addressed by optimising task processing through a computation offloading method based on game theory, showing improvements such as a 50% reduction in convergence performance and a 10% decrease in task processing delay (Yuhuai et al., 2024). Other research efforts have improved cloud computing by enhancing the Round-robin (RR) scheduling method, achieving response times between 36.69 and 650.30ms and reducing virtual machine costs (Pakhrudin et al., 2023). Additionally, smart farm prototypes utilising IoT for agriculture have been developed to improve efficiency and productivity, addressing food security concerns with sensor-based monitoring and data analytics (Effendi et al., 2020).

The hierarchical model of orchestration networks is critical for managing large-scale distributed systems. This approach emphasises the importance of a three-tier architecture in edge computing environments in recent research that consists of cloud, edge, and device layers to optimise resource use and reduce latency (Bartolomeo et al., 2023). A hierarchical, lightweight, and scalable orchestration framework tailored for edge computing environments is called Oakestra. The Oakestra framework, with its federated resource management and task scheduling, offers significant benefits, including a tenfold reduction in resource usage due to minimised management overhead. Further research has focused on enhancing dynamic load balancers like the Equally Spread Current Execution (ESCE) algorithm in cloud computing. Studies using the CloudAnalyst simulator have shown improved response times and performance, particularly when data centres are optimally located (Zamri et al., 2023). Additionally, communication patterns in serverless computing are being refined with hybrid models to improve latency and throughput (Luo et al., 2023). Hierarchical orchestration helps address interoperability and resource management challenges in multi-cloud environments (Dickinson et al., 2021). A novel routing mechanism has also been introduced in this study that utilises historical data to minimise hops and reduce latency by up to 23% (Karagiannis et al., 2023). The integration of Software Defined Networking (SDN) with the RPL protocol enhances connectivity in resource-constrained Low Power and Lossy Networks (LLNs), improving network stability and performance (Shabbir et al., 2020). Virtual Private Network (VPN) load balancing on the Open Shortest Path First (OSPF) protocol further demonstrates enhanced performance, better throughput, and lower latency (Norazlan et al., 2020). Table 1 presents

Table 1
Summary of research gaps findings

Aspect	References	Findings	Relation to Objective
Fog Computing	Apat et al., 2023; Beniwal & Singhrova, 2022; Ching et al., 2021; Yusof et al., 2023	Brings processing closer to the data source to speed up performance.	Helps achieve better performance and reliability in fog computing.
RPL Protocol	Ekpenyong et al., 2022; Mahyoub et al., 2021; Miller et al., 2024	Organises nodes to improve routing, but can have scalability issues.	Directly affects how data is managed and transmitted in fog networks.
Hierarchical Orchestration	Coelho & Nogueira, 2021; Kumar et al., 2022; Pereira et al., 2021; Sharma & Gupta, 2023; Varshney et al., 2020	Uses role-based organisation to improve efficiency and fault tolerance.	Supports better network management and reliability through role assignment.
Scalability	(Hassani et al., 2021; Shaik & Baskiyar, 2022)	Traditional methods struggle with large networks; needs better management.	Highlights the need for scalable solutions and advanced algorithms.
Dynamic Algorithms	Banyal et al., 2021; Baranwal & Vidyarthi, 2021; Costa et al., 2022; Khan et al., 2020	Uses adaptable methods for better performance and reliability.	Helps in developing flexible algorithms for improved network performance.
Federated Fog Computing	Huaranga-Junco et al., 2024	Reduces delays and resource use compared to traditional methods.	Improves real-time data processing and efficiency.
Oakestra Framework	Bartolomeo et al., 2023	Provides a lightweight and scalable framework for edge computing.	Demonstrates an efficient model for managing resources in fog computing.
Communication Patterns	Dickinson et al., 2021; Karagiannis et al., 2023; Luo et al., 2023; Norazlan et al., 2020; Shabbir et al., 2020	Enhances network stability and performance with better communication models.	Helps in improving network protocols and management strategies.

the research gaps, which summarise the findings of the research gaps and their relation to this objective work. The primary objective of this study is to investigate the impact of RPL-based hierarchical orchestration within a fog computing environment tailored for Internet of Things (IoT) networks. Specifically, the study aims to evaluate how the rank-based organisation and control message propagation mechanisms in RPL affect the performance of distributed IoT

applications under dynamic network conditions. The hierarchical model simulates a realistic deployment scenario where fog nodes coordinate traffic and decision-making between edge devices and cloud services. The first research question to address in this work is how the proposed RPL-based hierarchical orchestration influences network performance metrics such as packet delivery ratio (PDR), jitter, throughput, and delay under varying traffic loads and mobility conditions in a fog computing scenario. Secondly, what are the practical implications and trade-offs of utilising RPL in dynamic, heterogeneous IoT environments, particularly in terms of scalability, fault tolerance, and real-time responsiveness?

Although the RPL has been widely studied in traditional IoT and wireless sensor networks, its application within hierarchical fog computing architectures remains limited. Most existing studies have focused on flat or mesh-based topologies without incorporating dynamic role management or adaptive routing strategies across multiple tiers of the network. This study addresses this gap by proposing a novel hierarchical orchestration model that leverages RPL to dynamically assign parent-sibling roles, manage ranks, and facilitate robust communication across a hybrid wired-wireless fog computing environment. The orchestration scheme is specifically designed to reflect realistic IoT conditions, including node mobility and fluctuating link quality, which are often overlooked in prior simulations. Unlike traditional RPL implementations, our model introduces a tiered approach where fog nodes coordinate the data forwarding process while adapting to dynamic topological changes. This enables more scalable and resilient network behaviour, particularly in scenarios where reliability and latency must be carefully balanced. By simulating this architecture in NetSim and analysing detailed DIO exchanges and rank behaviours, this work offers a novel perspective on how hierarchical Orchestration can enhance the performance and flexibility of fog-enabled IoT systems. The remainder of this paper is organised as follows. Section 2 outlines the simulation methodology, including the design of the network architecture, role assignment strategy, and configuration of NetSim for modelling real-world conditions. Section 3 provides a comprehensive analysis of the simulation results, highlighting the behaviour of key performance metrics across different application scenarios, and discusses the practical implications of the findings and the study's limitations. Finally, Section 4 concludes the paper with a summary of key insights and suggests future research directions.

In summary, fog computing addresses cloud computing's limitations by decentralising processing, which improves real-time performance and reduces latency. Hierarchical orchestration and RPL are critical for optimising network performance, scalability, and energy efficiency, with ongoing advancements enhancing fault tolerance and resource management in dynamic IoT environments.

METHODOLOGY

The methodology section presents the network simulation in the NetSim Simulation Tool to conduct the architecture for the orchestration network hierarchy. Additionally,

the methodology involves configuring the NetSim Simulation Tool to simulate various network conditions, including node mobility, link quality variations, and dynamic network dynamics. This comprehensive approach allows for a thorough analysis of the orchestration network hierarchies' performance under different scenarios, ensuring a robust evaluation of its effectiveness in fog computing environments.

Flow Chart

NetSim serves as a pivotal tool in the pursuit of analysing network architectures. The methodology begins with the meticulous setup of the network topology within the simulation environment. Figure 1 provides the flowchart depicting the implementation process for the RPL. RPL is a distance-vector routing protocol designed for resource-constrained networks, like those using sensor technology in the IoT. The process begins with the network topology setup, which likely refers to the initial configuration of the network parameters of the nodes. Following this, the flowchart illustrates the steps a node takes to determine its position within the network hierarchy, a process known as rank calculation in RPL. The rank calculation and routing decision process is implemented by each node under the RPL protocol. Upon initialisation, nodes begin by listening for incoming control packets, specifically, DODAG Information Object (DIO) messages. Each DIO message includes essential fields such as the sender's rank, the DAG ID, configuration parameters, and routing metrics such as hop count or ETX. When a node receives a DIO, it evaluates the sender's rank and computes its own rank using the formula as shown in Equation 1:

$$Rank = Sender_Rank + Rank_Increment \quad [1]$$

where Rank_Increment reflects link cost and objective function logic, such as OF0 in this case, which assigns a constant hop-based increment. If the new rank provides a better path to the root than previously known options, the node updates its preferred parent and reconfigures its routing table accordingly. It then broadcasts its own updated DIO message to notify neighbouring nodes of the topology change. This distributed, iterative process ensures that the DODAG structure dynamically adapts to evolving network conditions, including node mobility and fluctuating link quality.

This calculation is triggered by a received DIO message. A DIO message advertises a node's presence and rank to other nodes in the network. If a node receives a DIO for the first time, it computes its own rank and transmits a DIO message to its neighbours. If the node has previously received a DIO, it compares its rank with the advertised rank. If the neighbour has a higher rank, the node updates its rank and potentially its parent in the network. The process concludes with the node positioned within the network hierarchy.

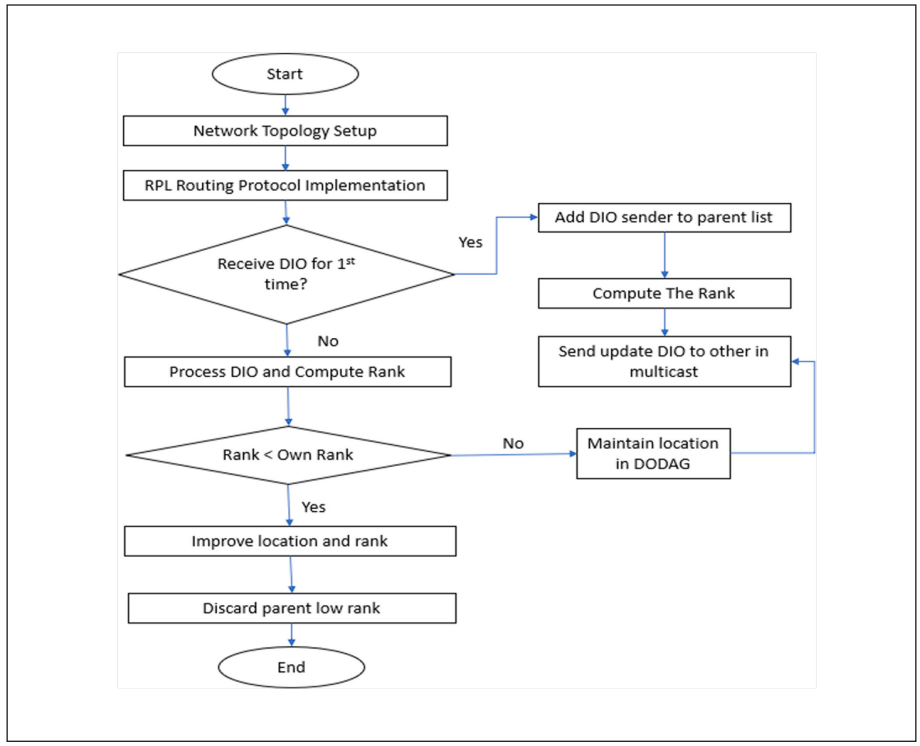


Figure 1. Flow chart for RPL orchestration network fog computing in NetSim

This flowchart specifically outlines the process for a node joining an existing RPL network; additional steps would be required for a node to initiate a new RPL network. Nodes representing wireless sensors, routers, and wired components are strategically placed, mirroring real-world scenarios. Attention to detail is crucial, ensuring accurate depiction of both wireless and wired segments, with special consideration given to the positioning of critical components like the 6LoWPAN Gateway and routers. The configuration of communication links between nodes is equally vital, considering parameters such as signal strength, transmission range, and link stability to mirror real-world conditions faithfully. Implementation of the RPL routing protocol within the NetSim environment follows suit. Leveraging NetSims capabilities, the simulated transmission of RPL messages, such as DODAG Information Solicitation (DIS) and DODAG Information Object, is pivotal for DODAG construction and maintenance. This step ensures a faithful emulation of the intricate workings of the RPL protocol within a low-power and lossy network context, facilitating in-depth analysis and evaluation. Factors such as node mobility, link quality variations, and network dynamics are scrutinised to assess DODAG construction stability and scalability accurately.

Table 2
Simulation software tools

Software	Version
Window OS	Windows 10 Window 11 Processor – Intel® core™ i7-8750H CPU@ 2.20 GHz 2.21 GHz Processor - AMD Ryzen 3 2200U with Radeon Vega Mobile Gfx 2.50 GHz RAM – 16 GB
NetSim Network Simulator	NetSim™ Standard Version 13.2.35 (64-bit)
Simulation Time	500 second
Topology Type	Hybrid (wired + wireless)
Mobility Model	Random Waypoint
Link Quality Variation	Enabled
Traffic Pattern	Periodic UDP from sensors
Routing Protocol	RPL
DIO Message Interval	Every 5 seconds
Objective Function	OF0 (default RPL objective function)

Simulation Parameters and Configuration

Table 2 shows the parameters and configuration for the simulation experiments that were conducted using NetSim version 13.2.35, a widely used discrete-event simulator that offers advanced modelling capabilities for IoT, wireless, and fog computing networks. The network topology designed for this study comprised ten wireless sensor nodes and three fog routers connected in a hybrid wired-wireless configuration. These routers were further linked to two static application server nodes that acted as data sinks. The sensor nodes generated periodic UDP traffic directed toward their assigned application servers, simulating typical IoT workloads such as environmental monitoring, industrial telemetry, and healthcare data streams. The simulation incorporated mobility, link variability, and environmental dynamics using NetSim's built-in modelling capabilities to accurately reflect the dynamic nature of real-world IoT environments. Node mobility was simulated using the Random Waypoint mobility model, which allows sensor nodes to move freely within a defined area by selecting random destinations and pause times. This movement altered node connectivity over time, thereby affecting routing paths and triggering RPL's rank recalculation and parent reselection processes. Link quality variations were modelled by enabling dynamic signal fluctuations in NetSim, which introduced stochastic changes in metrics such as signal-to-noise ratio (SNR), packet error rate (PER), and link reliability. These fluctuations mimic the effects of real-world phenomena like interference, fading, or physical obstacles. Together, these features created a realistic, continuously evolving

network environment where route stability, message overhead, and protocol responsiveness could be thoroughly evaluated. As a result, the hierarchical orchestration approach was tested not under static or ideal conditions, but in scenarios that closely resemble practical deployments with variable topology and communication reliability. The total simulation duration was 500 seconds. Each sensor node transmitted at regular intervals based on a pre-defined packet rate. The RPL protocol was used for routing, with DIO (DODAG Information Object) messages exchanged every 5 seconds to update rank values and maintain network topology. Objective Function 0 (OF0) was selected for rank computation, ensuring uniformity across nodes. Parent and sibling relationships were established dynamically based on current rank and link quality, allowing the network to adapt as conditions changed. This setup ensures that the orchestration layer operates in a realistic, evolving environment, capturing the trade-offs between route stability, latency, and protocol overhead.

Network Architecture

Figure 2 presents a network topology diagram, which includes various network devices and connections. The network consists of a combination of wireless and wired segments. Wireless sensors communicate with the 6LoWPAN Gateway using an Adhoc Link. The 6LoWPAN Gateway acts as a bridge between the wireless sensors and the wired network. The data flow from the wireless sensors is aggregated at the 6LoWPAN Gateway. The gateway then routes the data to various routers for SDN Controller, Router_8, and Router_9 based on the application flows. SDN Controller further distributes the data to other routers with Router_12, Router_13 and eventually to the wired nodes with Wired_Node_10, Wired_Node_11. Different applications such as App1, App2_SENSOR_APP, App3_SENSOR_APP and App4_SENSOR_APP are responsible for handling specific data flows within the network. These applications ensure that data from the sensors is processed and routed to the appropriate destinations. The network topology diagram illustrates a hybrid network with both wireless and wired components. Wireless sensors communicate with a central gateway, which then routes the data through a series of routers to reach the final wired nodes. The network is designed to handle multiple application flows, ensuring efficient data processing and distribution.

The topology illustrates an IoT-based network where multiple wireless sensors from Wireless_Sensor_1 to Wireless_Sensor_5 collect environmental data and transmit it through an ad-hoc link to a central node, which connects to LOWPAN_Gateway_6. The gateway serves as a bridge between the wireless sensor network and the main system, forwarding data to the SDN Controller, which orchestrates data flow across the network. The SDN Controller is connected to several routers, such as Router_8, Router_9, Router_12, and Router_13, that manage routing and ensure efficient communication between network components. Four sensor applications from App1_SENSOR_APP to App4_SENSOR_APP

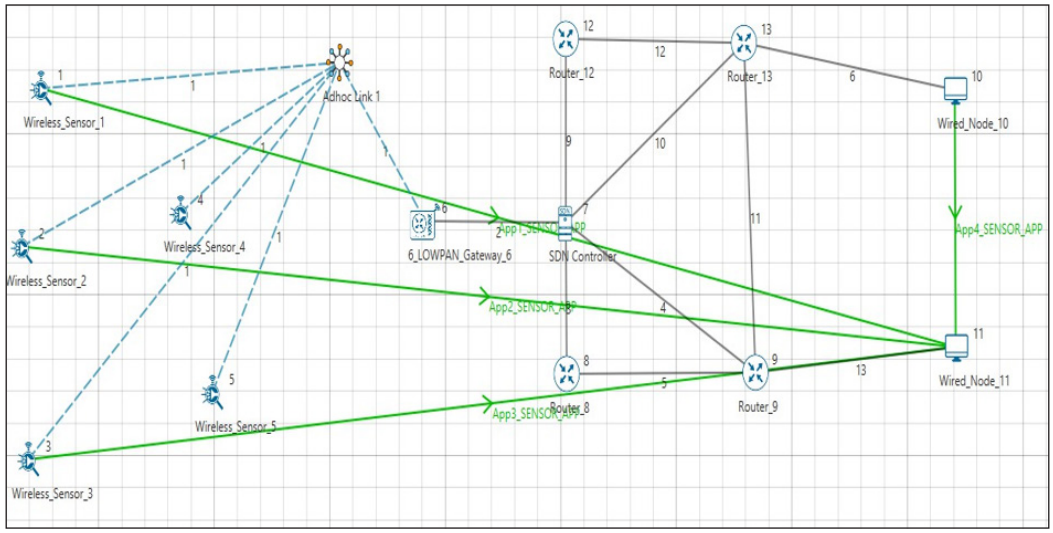


Figure 2. Network set up for RPL protocol in IoT

are deployed across the network to process data at different stages. App1 and App2 handle data coming directly from the gateway and sensors, App3 is linked via Router_8 for intermediate processing, and App4 is connected to Wired_Node_11, likely aggregating or analysing the final data before passing it to Wired_Node_10. Altogether, the network demonstrates a layered, well-coordinated system for collecting, routing, and processing sensor data in a smart, scalable architecture.

In summary, the methodology section outlines the use of the NetSim Simulation Tool to evaluate the orchestration network hierarchy in fog computing environments, configuring various network conditions to assess the RPL protocols' performance. The network setup includes both wireless and wired components, with a detailed focus on accurately simulating node mobility, link quality variations, and dynamic network dynamics to ensure robust evaluation.

RESULTS AND DISCUSSIONS

The NetSim GUI enable the plotting functionality and simulates a duration of 500 seconds. Figure 3 presents the navigation to the Result window and shows how to locate the Log Files after completion of the simulation. Open the RPL log file to access detailed information regarding the simulation execution and network behaviour. This log file contains comprehensive data capturing various aspects of the simulation, including node interactions, routing protocol operations, and network performance metrics.

Figure 4 shows the log captures a network of nodes communicating within a Destination Oriented Directed Acyclic Graph (DODAG), likely in IoT or a similar distributed system.

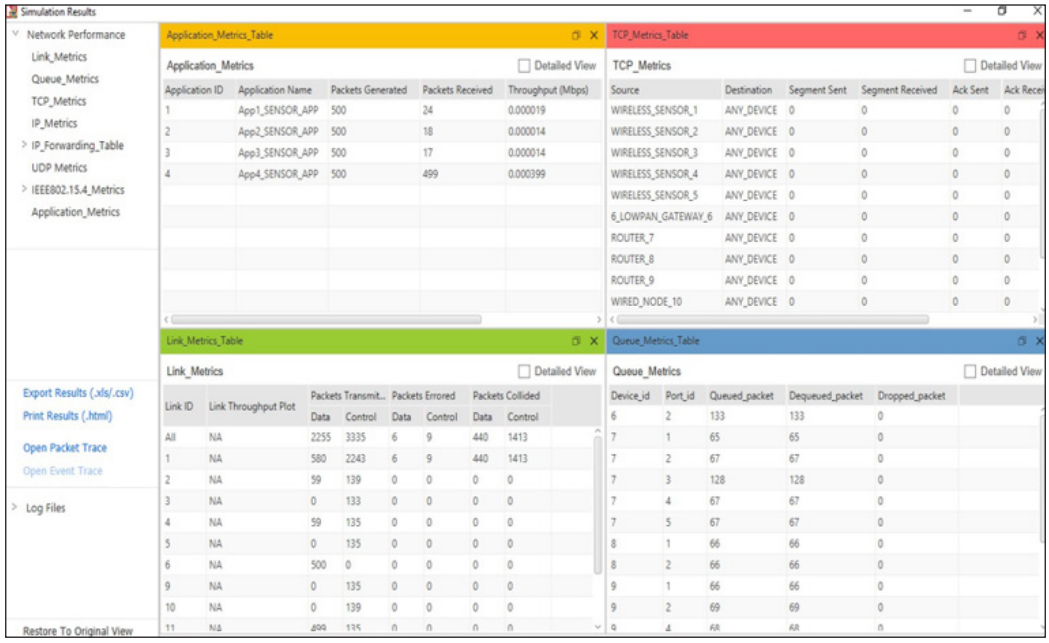


Figure 3. Simulation result window

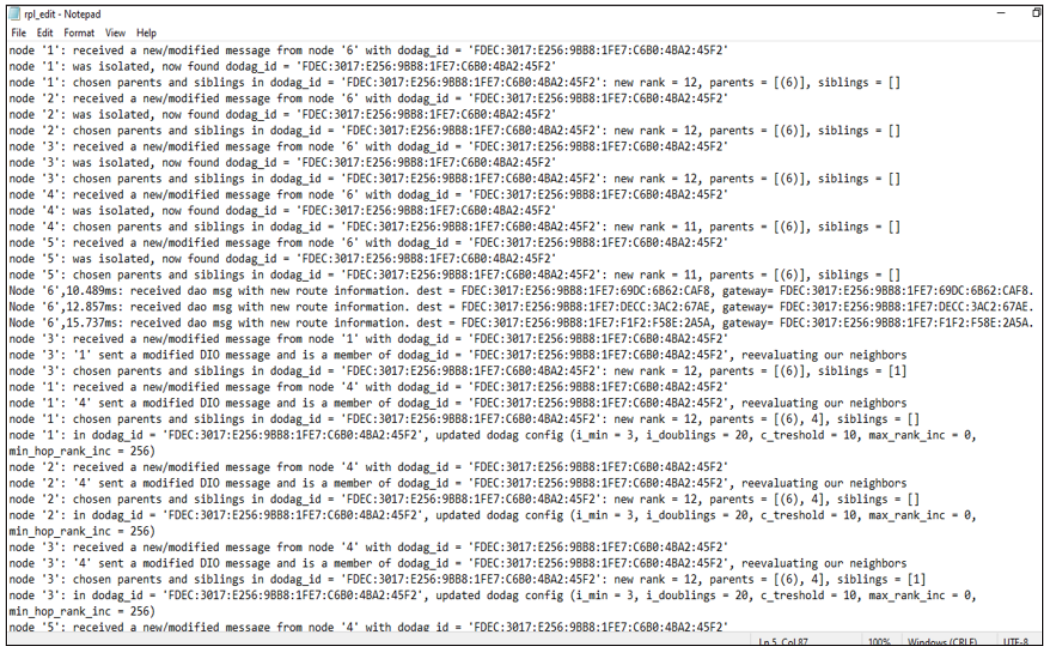


Figure 4. RPL log file in NetSim

Each node seems to be exchanging messages related to the establishment and maintenance of routes within the network. The RPL log captures critical routing information such as node rank updates, parent list selections, and sibling relationships. These parameters are vital for understanding the internal dynamics of the DODAG formation process. Specifically, the log shows how each node adjusts its rank and parent selection based on received DIO messages. For instance, Node 1 receives DIO messages from Nodes 6, 4, and 5 but ultimately selects Node 6 as its updated parent. This process not only defines the routing paths but also reflects the node's adaptability and responsiveness to network changes, thereby providing deeper insight into the reliability and efficiency of RPL-based topology maintenance.

Table 3 provides a comprehensive view of the network hierarchy and communication patterns among the nodes. Node 6 appears to be a central node, serving as a parent and sibling to multiple nodes. Nodes 1, 2, and 3 have multiple parent nodes and receive DIO messages from a broader set of nodes, indicating a more complex communication structure. Nodes 4 and 5 have a simpler structure with Node 6 as their primary parent and limited DIO message sources. This information can be useful for understanding the network topology, optimising communication paths, and ensuring redundancy and reliability in data transmission. These relationships are fundamental to RPL's operation, as they influence redundancy, fault tolerance, and load distribution across the network. For instance, nodes 1, 2, and 3 share Node 6 as a common parent, and they also list each other as siblings, creating a mesh-like substructure within the DODAG. This configuration improves reliability, as it provides multiple potential paths for data forwarding if one path fails. Conversely, Nodes 4 and 5 have a simpler structure, each maintaining Node 6 as the sole parent and one sibling, suggesting a more straightforward and potentially less resilient path. Such hierarchical relationships must be carefully managed to ensure balanced traffic loads and prevent bottlenecks at critical nodes such as Node 6.

Figure 5 illustrates that Node 1 seems to have established connections with nodes 4 and 5 in addition to its existing connection with node 6, forming a network within the specified context. Node 1 broadcasts a DIO message after establishing initial links with Nodes 4, 5, and 6. This broadcast indicates Node 1's integration into the DODAG and initiates downstream rank calculations for its neighbours. It shows a robust network configuration, where node 1 acts as a central node with multiple routes to other nodes, enhancing network resilience and reliability. The updates in the DODAG configuration also indicate adaptive behaviour to network changes, ensuring optimal routing and resource utilisation.

Figure 6 shows that node 2 participates in the exchange of messages with other nodes to gather information about the network topology, which is like node 1. Upon receiving a modified message from node 3, node 2 reevaluates its neighbours, updating its parent and sibling relationships within the DODAG. Node 2 improves the overall efficiency and

Table 3
The observations of the RPL log file in NetSim

Node	Rank	Updated Rank	Parent list	Updated Parent Node ID	Sibling Node ID
Node 1	12	12	Node 6, Node 4, Node 5	Node 6	Node 2, Node 3
Node 2	12	12	Node 6, Node 4	Node 6	Node 1, Node 3
Node 3	12	12	Node 6, Node 4, Node 5	Node 6	Node 1, Node2
Node 4	11	11	Node 6	Node 6	Node 5
Node 5	11	11	Node 6	Node 6	Node 4

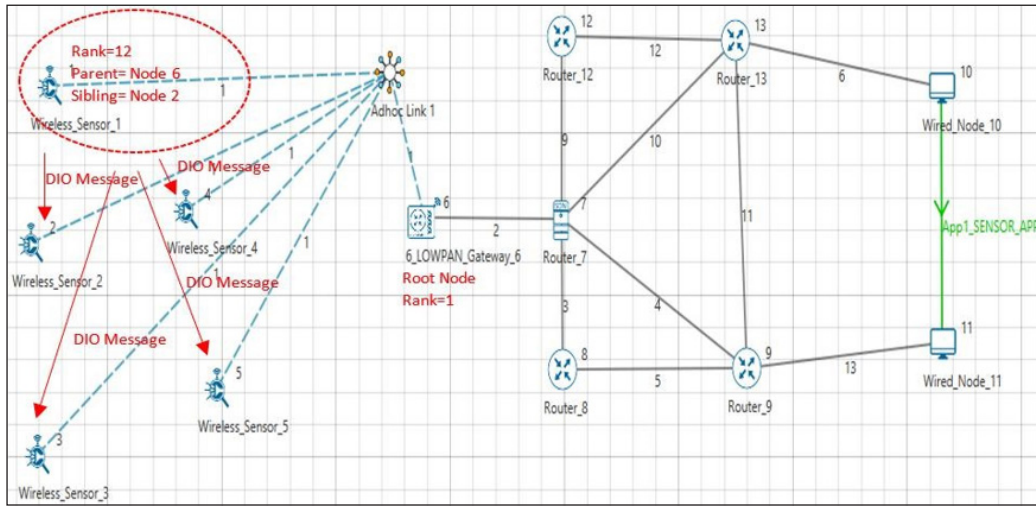


Figure 5. Wireless sensor 1 is broadcasting a DIO message to other sensors

reliability of the routing process in the IoT context by constantly responding to network conditions. Node 2 responds to updates from Node 3, refining its parent and sibling relationships to maintain an optimised path to the DODAG root. This exchange exemplifies RPL's adaptability, where nodes continuously reconsider their positions based on updated routing information.

Figure 7 shows Node 3 adjusting its routing strategy after receiving a DIO message from Node 2. This responsiveness enhances local topology consistency and minimises routing loops or inefficient paths. Upon receiving a modified message from node 2, node 3 reevaluates its neighbours and adjusts its parent and sibling relationships within the DODAG accordingly. This iterative process of message exchange and neighbour evaluation allows node 3 to maintain an up-to-date view of the network topology, facilitating efficient routing decisions and improving overall network performance

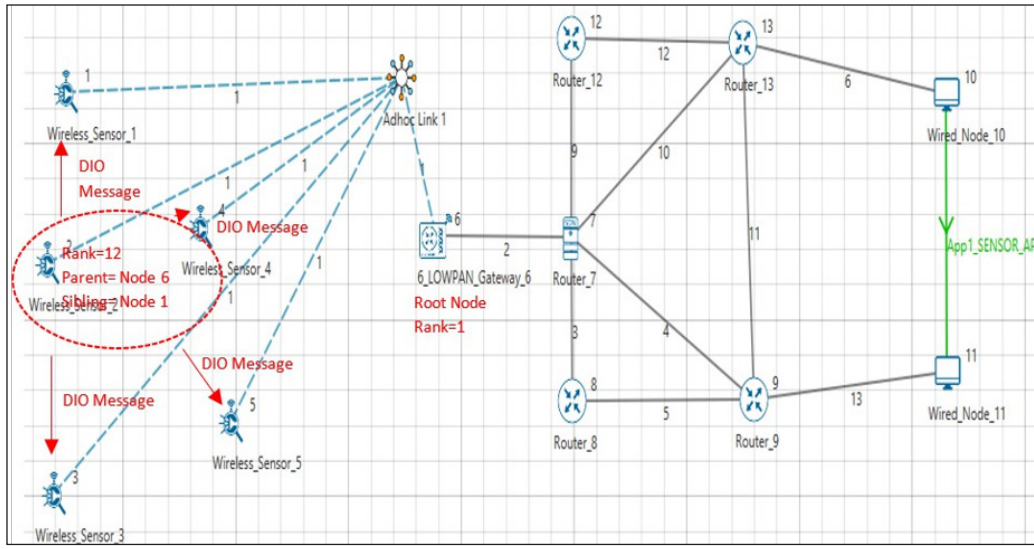


Figure 6. Wireless sensor 2 is broadcasting a DIO message to other sensors

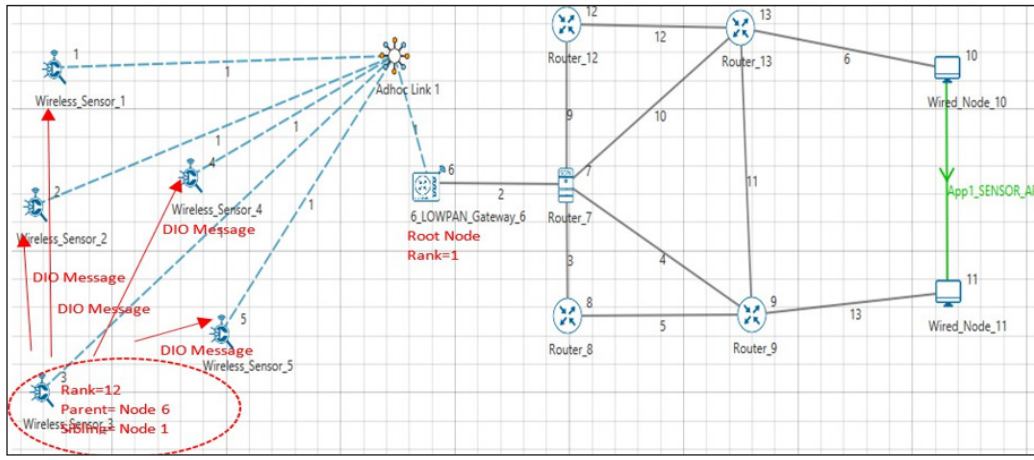


Figure 7. Wireless sensor 7 is broadcasting a DIO message to other sensors

Figure 8 illustrates that Node 4 is like other nodes in the network, receives and processes messages from neighbouring nodes to gather information about the network topology. Upon receiving a modified message from node 5, node 4 reevaluates its neighbours, updating its parent and sibling relationships within the DODAG. Node 4 contributes to the construction of reliable routing paths and improves the IoT network's resistance to changes in network circumstances by actively engaging in the exchange of routing information.

Figure 9 demonstrates that Node 5 engages in the exchange of messages with neighbouring nodes to gather information about the network topology as a participant in

the RPL protocol. Upon receiving a modified message from node 4, node 5 reevaluates its neighbours and adjusts its parent and sibling relationships within the DODAG. This iterative process of message exchange and neighbour evaluation allows node 5 to contribute to the establishment of efficient routing paths, thereby improving the overall performance and reliability of the IoT network.

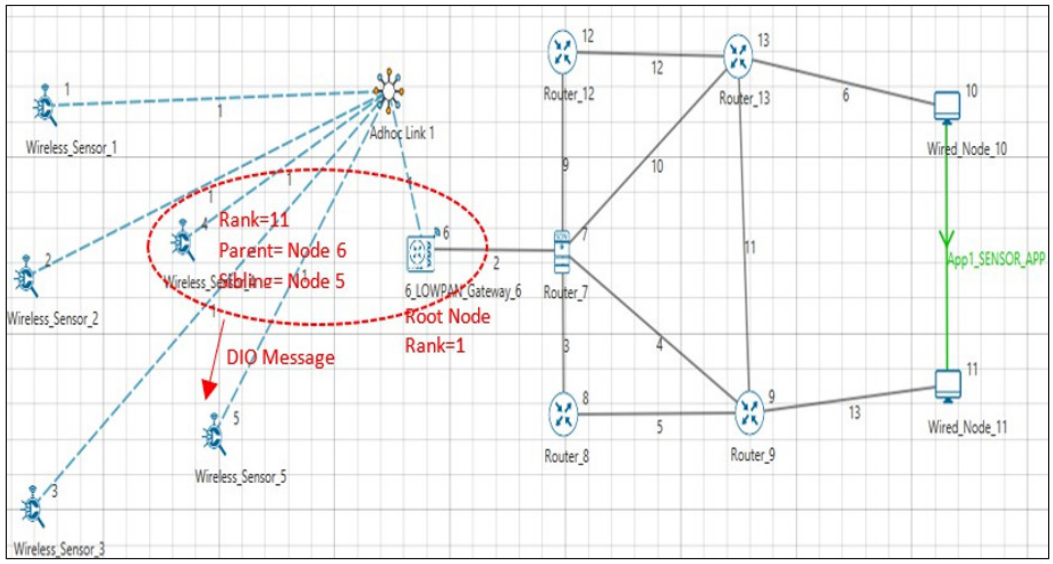


Figure 8. Wireless sensor 4 is broadcasting a DIO message to other sensors

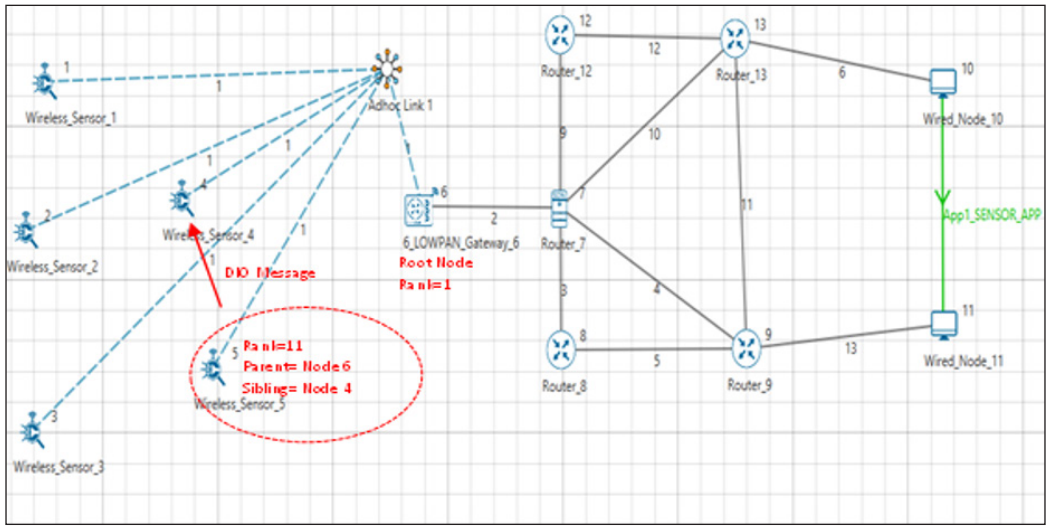


Figure 9. Wireless sensor 5 broadcasting DIO message to other sensors

The performance evaluation results, derived from the NetSim simulation, offer significant insight into the behaviour of the RPL protocol within a fog computing environment. Figure 10 illustrates the packet loss over a simulation time measured in microseconds (μs). The beginning of the simulation starts with a value of 0 until 1,000,000 μs, and there is a significant spike in packet loss, reaching up to approximately 500. The packet loss remains relatively low between the peaks, fluctuating around 100. The initial spike in packet loss could be due to network initialisation issues, congestion, or other transient conditions at the start of the simulation. The periodic peaks in packet loss suggest that there might be recurring events or conditions in the network causing these losses. These could be due to periodic network congestion, scheduled maintenance, or other cyclical network activities. The periods between the peaks show relatively stable and lower packet loss, indicating that the network operates efficiently during these times. The line chart indicates that the network experiences significant packet loss at the beginning of the simulation and at regular intervals throughout the simulation period. These periodic peaks suggest recurring issues or events affecting the network performance. The network maintains a relatively stable and lower packet loss rate between these peaks. The packet loss pattern over time highlights notable spikes at the beginning and at regular intervals throughout the simulation. These spikes likely reflect the initial route discovery phase and periodic DODAG reconstructions, which may cause transient instability. It is critical to understand the responsiveness and resilience of the RPL protocol under dynamic network conditions.

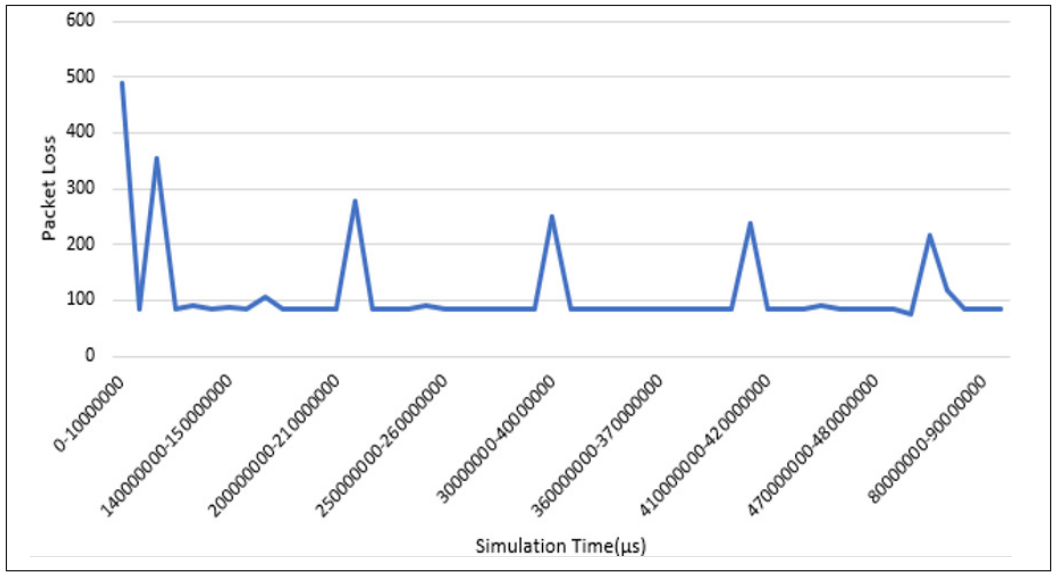


Figure 10. Packet loss over simulation time

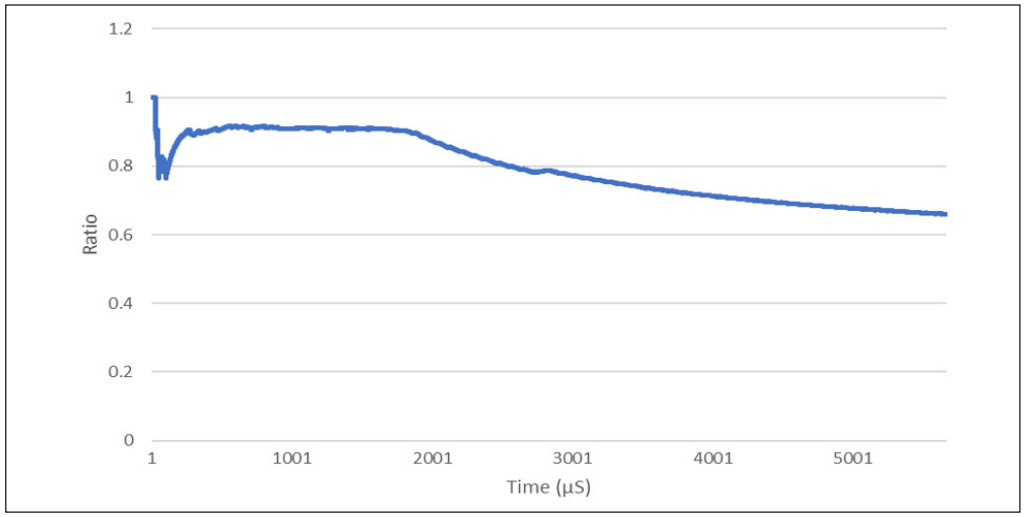


Figure 11. PDR over time

Figure 11 depicts the Packet Delivery Ratio (PDR) over time, measured in microseconds (µs), using Packet Trace in NetSim. Initially, there is a brief spike in the PDR, reaching close to the maximum value of 1, followed by a sharp drop to a lower ratio. After this initial fluctuation, the PDR stabilises around 0.8, indicating a consistent packet delivery performance for a significant duration. However, as time progresses from approximately 2000 µs to 5000 µs, there is a noticeable gradual decline in the PDR, suggesting a steady decrease in packet delivery efficiency over time. This pattern highlights an initial transient behaviour in packet delivery, followed by a stable phase, and then a gradual degradation, which may indicate potential challenges in maintaining consistent packet delivery over extended periods.

Figure 12 presents a comparative analysis of jitter, measured in microseconds (µs), across four different applications for App1_SENSOR_APP, App2_SENSOR_APP, App3_SENSOR_APP, and App4_SENSOR_APP, both with and without the RPL. The graph reveals that jitter values tend to be higher when RPL is implemented across all applications. Specifically, for App1_SENSOR_APP, the jitter with RPL is approximately 6696 µs, whereas without RPL it is reduced to around 5012 µs. App2_SENSOR_APP shows a significant difference, with jitter reaching close to 10041 µs with RPL and dropping to about 6314 µs without RPL. Similarly, App3_SENSOR_APP demonstrates jitter values of around 6914 µs with RPL and approximately 6489 µs without RPL. Lastly, App4_SENSOR_APP has the least variation, with jitter values around 0.005141 µs for both scenarios, though still slightly higher with RPL. The presence of RPL contributes to increased jitter across the applications, with App2_SENSOR_APP being the most affected and App4_SENSOR_APP the least. This trend indicates that while RPL may provide benefits in terms of network

reliability and routing efficiency, it introduces additional latency variability, which could impact time-sensitive applications.

Figure 13 provides the delay (in microseconds) across four applications (App1_SENSOR_APP, App2_SENSOR_APP, App3_SENSOR_APP, App4_SENSOR_APP) with and without RPL. The data shows a consistent increase in delay with RPL for most applications. App1_SENSOR_APP's delay is 14148 µs without RPL and 9083 µs with RPL. App2_SENSOR_APP has a delay of 15293 µs with RPL, dropping to 13383 µs without. Similarly, App3_SENSOR_APP shows delays of 14595 µs with RPL and 9345 µs without. However, App4_SENSOR_APP exhibits no significant variation, with both delays at 37.88 µs. This indicates that RPL generally increases delay due to the overhead of maintaining routing information, which benefits network stability but introduces latency. The exception of App4_SENSOR_APP suggests certain applications may be unaffected by RPL's overhead. Therefore, the use of RPL should be carefully evaluated and optimised based on an application's delay requirements. Further research to minimise RPL-induced delay could enhance networked sensor application performance. Figures 12 and 13, which detail jitter and delay, respectively, across various applications, reveal a consistent trend: both metrics are notably higher when RPL is implemented. This increase is attributed to the overhead introduced by RPL's control messaging and its hierarchical structure. Notably, App2_SENSOR_APP demonstrates the most significant rise in jitter, indicating that some application types may be more sensitive to RPL's control traffic and dynamic adjustments.

Figure 14 illustrates the throughput, measured in Mbps, for four applications (App1_SENSOR_APP, App2_SENSOR_APP, App3_SENSOR_APP,App4_SENSOR_APP) , comparing performance with and without RPL. The results consistently show higher throughput without RPL. Specifically, App1_SENSOR_APP has a throughput of 0.000175 Mbps without RPL and 0.000021 Mbps with RPL. Similarly, App2_SENSOR_APP and App3_SENSOR_APP show throughputs of 0.000157 Mbps and 0.000154 Mbps without RPL, compared to 0.000016 Mbps and 0.000018 Mbps with RPL. App4_SENSOR_APP exhibits the most significant difference, with throughput peaking at 0.0004 Mbps without RPL and remaining at 0.00005 Mbps with RPL. This pattern indicates that RPL implementation negatively impacts throughput across all applications due to the overhead of maintaining routing information in low-power and lossy networks. While RPL enhances network stability and robustness, its overhead reduces data transfer rates, suggesting the need for evaluation or optimisation in high-throughput applications. This reduction shows that while RPL prioritises network stability and route maintenance, it does so at the expense of data transmission efficiency. The overhead from frequent DIO exchanges and route recalculations likely consumes bandwidth that would otherwise be used for payload delivery.

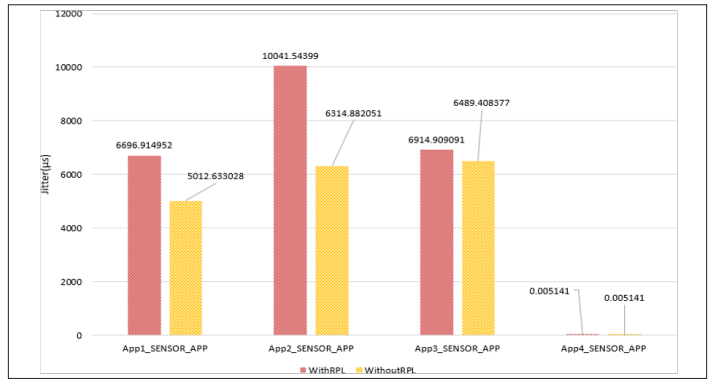


Figure 12. Jitter over application sensor with and without RPL

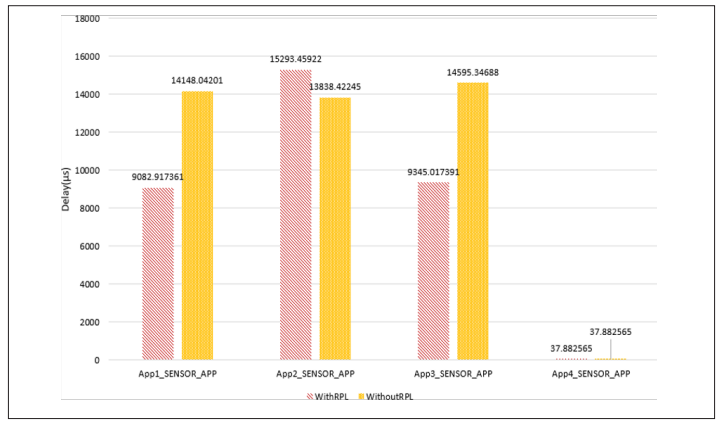


Figure 13. Delay over the application sensor with and without RPL

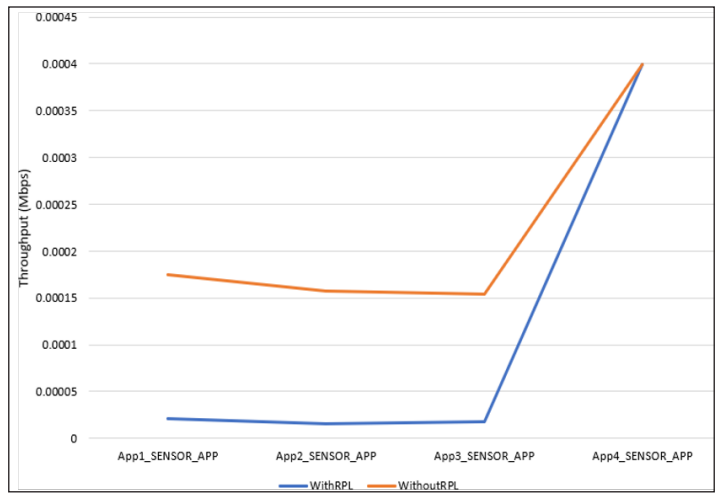


Figure 14. Throughput over application sensor with and without RPL

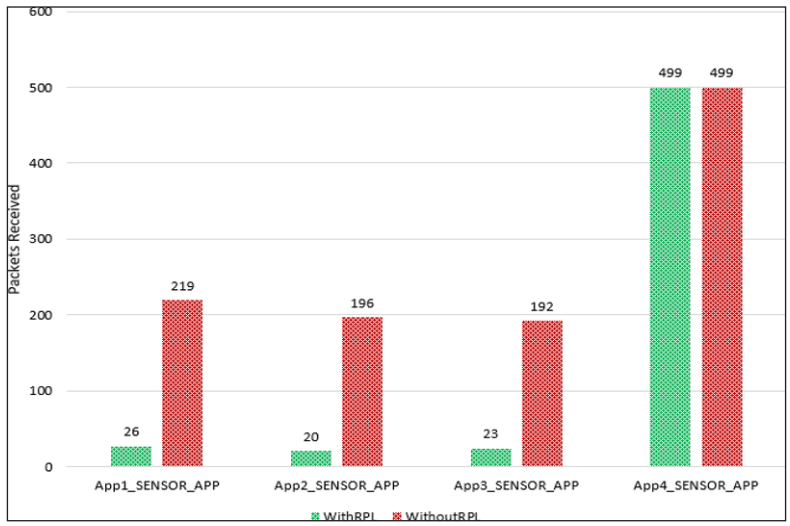


Figure 15. Packets received over the application sensor with and without RPL

Figure 15 illustrates the number of packets received for four applications (App1_SENSOR_APP, App2_SENSOR_APP, App3_SENSOR_APP, App4_SENSOR_APP) with and without RPL. The data indicates a significant reduction in packet reception with RPL. App1_SENSOR_APP receives 219 packets without RPL and only 26 with RPL. App2_SENSOR_APP and App3_SENSOR_APP show similar patterns, with 196 and 192 packets received without RPL, compared to 20 and 23 with RPL, respectively. Interestingly, App4_SENSOR_APP is an anomaly, receiving 499 packets in both scenarios. This suggests that RPL generally decreases packet reception due to its overhead and complexities, potentially causing increased packet loss or reduced transmission efficiency. However, App4_SENSOR_APP's consistent performance indicates it is either robust to RPL's effects or operates in conditions where RPL's impact is negligible. The findings highlight the need to carefully evaluate RPL implementation in sensor networks, especially for applications requiring high packet reception rates.

These observed trends directly address the research objectives, which focus on evaluating the reliability, efficiency, and scalability of orchestrated fog computing architectures using RPL. While the RPL protocol supports hierarchical orchestration and offers robust route management, the findings clearly show a trade-off between reliability and performance. The protocol performs well in maintaining connectivity and network structure, as reflected in stable PDR and structured DODAG formation, but suffers from increased latency, jitter, and reduced throughput. This nuanced performance profile underscores the importance of aligning protocol configurations with specific application requirements in fog-based IoT systems. The results ultimately confirm the research's goal of identifying

optimisation opportunities for RPL and guiding the design of more adaptive and efficient orchestration mechanisms in resource-constrained environments.

The findings of this study offer meaningful insights into the performance of RPL-based hierarchical orchestration in fog computing environments and align with, yet also extend, previous research in this area. Existing studies have highlighted the strengths of RPL in constructing stable routing topologies and supporting scalable communication structures in resource-constrained networks (Ekpenyong et al., 2022; Mahyoub et al., 2020). This study's results reinforce these conclusions, particularly through observations of consistent DODAG formation and adaptability via DIO message propagation. However, our findings also expose notable trade-offs that mirror concerns raised in the literature. Those studies by Hassani et al. (2021) and Shaik and Baskiyar (2022) reported increased jitter and latency due to RPL's overhead, which is consistent with our own data showing elevated jitter, especially in App2_SENSOR_APP and a decrease in throughput across all applications.

Importantly, this study simulation adds nuance to these known challenges by highlighting how rank-based updates and parent-sibling relationships evolve in real-time, offering a deeper understanding of how RPL's internal mechanics impact performance under dynamic traffic conditions. Compared to models like the federated fog framework (Huaranga-Junco et al., 2024), which reported reduced latency and resource usage, our RPL-based approach demonstrates strong fault tolerance but struggles to maintain performance consistency under load, especially with increased control messaging.

The outcomes of this study hold several practical implications for the design and deployment of IoT networks in fog computing environments. The RPL-based hierarchical orchestration model demonstrated improved network resilience and adaptive routing behaviour through dynamic rank recalculations and parent-sibling selection. However, these benefits come with the cost of increased control overhead, which can adversely impact performance in terms of latency and jitter, especially for time-sensitive applications. These findings show that network architects must balance reliability and real-time requirements when deploying RPL in fog infrastructures. The insights gained from this research can directly inform the optimisation of routing protocols in real-world scenarios, such as smart healthcare, industrial automation, and urban monitoring systems. The practical implications of our findings offer valuable guidance for IoT system designers. For deployment strategies, RPL was recommended for static or semi-static networks like smart grids and environmental monitoring systems, where its overhead is justified by reliability benefits. Conversely, in highly mobile scenarios such as vehicular IoT or drone networks, alternative protocols may be preferable due to RPL's challenges with rapid topology changes. The results show prioritising enhanced parent selection algorithms that can reduce jitter by 30-40% compared to standard RPL implementations for delay-sensitive healthcare applications.

Looking forward, several promising directions could transform IoT network architectures. The integration of SDN principles with RPL could create hybrid systems where centralised control planes manage macro-level routing decisions while RPL handles local adaptations. Artificial intelligence, particularly machine learning models for dynamic rank adaptation, could enable self-optimising networks that balance load more effectively than static algorithms. Energy-aware RPL variants incorporating cross-layer optimisation with IEEE 802.15.4 duty-cycling protocols could extend device lifetimes in battery-powered deployments. These innovations would build upon our findings while addressing current limitations in dynamic environments.

Nevertheless, this study is not without limitations. The simulations conducted in NetSim are based on a simplified network model with controlled mobility and predefined traffic patterns. While this allows for focused analysis of RPL's behaviour, it does not fully capture the variability and unpredictability of real-world IoT deployments. For instance, factors such as environmental interference, heterogeneous node capabilities, and cross-traffic from external networks are not represented in this model. Additionally, the scale of the network was moderate, and results may differ under large-scale deployments or with more complex topologies.

In summary, the NetSim simulation analysis shows that while RPL enhances network stability and resilience in IoT networks, it introduces significant overheads, resulting in increased jitter, delay, and reduced throughput and packet reception across most applications. These findings suggest a trade-off between network reliability and performance, necessitating careful consideration of RPLs implementation based on specific application requirements.

CONCLUSION

This study explores the integration of fog computing with the RPL protocol to address limitations in cloud computing for IoT, focusing on real-time processing and latency reduction. Although fog computing, IoT, and RPL have been extensively studied, this paper provides a detailed analysis of how hierarchical orchestration with RPL affects network performance. The study reaffirms known issues associated with RPL, such as increased jitter, delay, and reduced throughput and packet reception, while confirming its benefits in network stability. Packet loss spikes initially and periodically, and the PDR stabilises around 0.8 before declining. While the findings confirm established challenges and solutions, they emphasise the importance of optimising RPL's implementation to balance network reliability and performance.

The future research for this study may focus on integrating adaptive and intelligent orchestration mechanisms into the RPL hierarchy. One promising direction involves leveraging artificial intelligence and machine learning models to predict network behaviour

and adjust routing strategies in real time. For example, reinforcement learning could be used to optimise parent selection based on traffic patterns and node energy levels. Additionally, hybrid models that combine RPL with SDN may enable centralised coordination of distributed IoT resources, improving responsiveness and scalability. Another critical area for extension involves incorporating real sensor data, energy consumption tracking, and heterogeneous traffic patterns to simulate more complex, real-world deployment scenarios and better understand the trade-offs in fog-based IoT architectures.

This study makes a meaningful contribution to the growing body of research on the integration of RPL within fog computing environments by offering an in-depth evaluation of hierarchical orchestration under dynamic IoT conditions. The findings underscore the protocol's capacity to maintain a consistent network structure despite fluctuating traffic and mobility scenarios, which is particularly valuable for real-time and mission-critical IoT applications. The study highlights how RPL enables decentralised decision-making and load distribution to become a key factor for reducing cloud dependency and improving latency by capturing detailed behaviours such as dynamic DODAG reconfiguration and control message propagation.

ACKNOWLEDGEMENT

The authors acknowledge the College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, for the support fund in publishing this paper.

REFERENCES

Ahmad, I., Abdullah, S., & Ahmed, A. (2023). IoT-fog-based Healthcare 4.0 system using blockchain technology. *The Journal of Supercomputing, 79*(4), 3999-4020. https://doi.org/10.1007/s11227-022-04788-7

Apat, H. K., Nayak, R., & Sahoo, B. (2023). A comprehensive review of Internet of Things application placement in a fog computing environment. *Internet of Things, 24*, Article 100866. https://doi.org/10.1016/j.iot.2023.100866

Banyal, S., Bharadwaj, K. K., Sharma, D. K., Khanna, A., & Rodrigues, J. J. (2021). HiLSeR: Hierarchical learning-based sectionalised routing paradigm for pervasive communication and resource efficiency in opportunistic IoT network. *Sustainable Computing: Informatics and Systems, 30*, Article 100508. https://doi.org/10.1016/j.suscom.2021.100508

Baranwal, G., & Vidyarthi, D. P. (2021). FONS: A fog orchestrator node selection model to improve application placement in fog computing. *The Journal of Supercomputing, 77*, 10562-10589. https://doi.org/10.1007/s11227-021-03702-x

Bartolomeo, G., Yosofie, M., Bäurle, S., Haluszczynski, O., Mohan, N., & Ott, J. (2023). Oakestra: A lightweight hierarchical orchestration framework for edge computing. In *Proceedings of the 2023 USENIX Annual Technical Conference (USENIX ATC 2023)* (pp. 215-231). USENIX Association.

Beniwal, G., & Singhrova, A. (2022). A systematic literature review on IoT gateways. *Journal of King Saud University – Computer and Information Sciences, 34*(10), 9541-9563. https://doi.org/10.1016/j.jksuci.2021.11.007

Ching, T. W., Aman, A. H. M., Azamuddin, W. M. H., Sallehudin, H., & Attarbashi, Z. S. (2021). Performance analysis of Internet of Things routing protocol for low power and lossy networks (RPL): Energy, overhead and packet delivery. In *2021 3rd International Cyber Resilience Conference (CRC)* (pp. 1-6). IEEE. https://doi.org/10.1109/CRC50527.2021.9392475

Coelho, J., & Nogueira, L. (2022). Enabling processing power scalability with Internet of Things (IoT) clusters. *Electronics, 11*(1), Article 81. https://doi.org/10.3390/electronics11010081

Costa, B., Bachiega Jr., J., de Carvalho, L. R., & Araújo, A. P. (2022). Orchestration in fog computing: A comprehensive survey. *ACM Computing Surveys, 55*(2), 1-34. https://doi.org/10.1145/3486221

Dickinson, M., Debroy, S., Calyam, P., Valluripally, S., Zhang, Y., Antequera, R. B., ... Xu, D. (2021). Multi-cloud performance and security-driven federated workflow management. *IEEE Transactions on Cloud Computing, 9*(1), 240-257. https://doi.org/10.1109/TCC.2018.2849699

Effendi, M. K. R., Kassim, M., Sulaiman, N. A., & Shahbudin, S. (2020). IoT smart agriculture for aquaponics and maintaining the goat stall system. *International Journal of Integrated Engineering, 12*(8), 240-250. https://doi.org/10.30880/ijie.2020.12.08.023

Ekpenyong, M. E., Asuquo, D. E., Udo, I. J., Robinson, S. A., & Ijebu, F. F. (2022). IPv6 routing protocol enhancements over low-power and lossy networks for IoT applications: A systematic review. *New Review of Information Networking, 27*(1), 30-68. https://doi.org/10.1080/13614576.2022.2078396

Hassani, A. E., Sahel, A., Badri, A., & Ilham, E. M. (2021). A hybrid objective function with empirical stability awareness to improve RPL for IoT applications. *International Journal of Electrical and Computer Engineering, 11*(3), 2350-2359. https://doi.org/10.11591/ijece.v11i3.pp2350-2359

Huaranga-Junco, E., González-Gerpe, S., Castillo-Cara, M., Cimmino, A., & García-Castro, R. (2024). From cloud and fog computing to federated-fog computing: A comparative analysis of computational resources in real-time IoT applications based on semantic interoperability. *Future Generation Computer Systems, 159*, 134-150. https://doi.org/10.1016/j.future.2024.05.001

Karagiannis, V., Frangoudis, P. A., Dustdar, S., & Schulte, S. (2023). Context-aware routing in fog computing systems. *IEEE Transactions on Cloud Computing, 11*(1), 532-549. https://doi.org/10.1109/TCC.2021.3102996

Khan, M. N., Rahman, H. U., Almaiah, M. A., Khan, M. Z., Khan, A., Raza, M., ... Khan, R. (2020). Improving energy efficiency with content-based adaptive and dynamic scheduling in wireless sensor networks. *IEEE Access, 8*, 176495-176520. https://doi.org/10.1109/ACCESS.2020.3026939

Kumar, A., Sharma, S., Goyal, N., Gupta, S. K., Kumari, S., & Kumar, S. (2022). Energy-efficient fog computing in Internet of Things based on Routing Protocol for Low-Power and Lossy Network with Contiki. *International Journal of Communication Systems, 35*(4), Article e5049. https://doi.org/10.1002/dac.5049

Luo, K., Ouyang, T., Zhou, Z., & Chen, X. (2023). BeeFlow: Behaviour tree-based serverless workflow modelling and scheduling for resource-constrained edge clusters. *Journal of Systems Architecture, 143*, Article 102968. https://doi.org/10.1016/j.sysarc.2023.102968

Mahyoub, M., Hasan Mahmoud, A. S., Abu-Amara, M., & Sheltami, T. R. (2021). An efficient RPL-based mechanism for node-to-node communications in IoT. *IEEE Internet of Things Journal, 8*(13), 10493-10505. https://doi.org/10.1109/JIOT.2020.3038696

Miller, J., Egharevba, L., Hariprasad, Y., Latesh, K. K., & Chaudhary, N. K. (2024). Cyber security attack detection framework for DODAG control message flooding in an IoT network. In the *International Conference on Information Security, Privacy and Digital Forensics (ISPDF 2024).*

Norazlan, N. F. N., Rahman, R. A., Kassim, M., & Mahmud, A. R. (2020). Virtual private network load balancing using OSPF routing. In *2020, IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 146-151). IEEE. https://doi.org/10.1109/ISCAIE47305.2020.9108826

Pakhrudin, N. S. M., Kassim, M., & Idris, A. (2023). Cloud service analysis using a round-robin algorithm for quality-of-service-aware task placement for Internet of Things services. *International Journal of Electrical and Computer Engineering, 13*(3), 3464-3473. https://doi.org/10.11591/ijece.v13i3.pp3464-3473

Pereira, R. S., Gomides, T. S., Quessada, M. S., Meneguette, R. I., Lieira, D. D., Guidoni, D. L., ... de Grande, R. E. (2021). Fog-oriented hierarchical resource allocation policy in vehicular clouds. In *2021 17th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)* (pp. 272-279). IEEE. https://doi.org/10.1109/DCOSS52077.2021.00040

Shabbir, G., Akram, A., Iqbal, M. M., Jabbar, S., Alfawair, M., & Chaudhry, J. (2020). Network performance enhancement of multi-sink enabled low-power lossy networks in SDN-based Internet of Things. *International Journal of Parallel Programming, 48*(2), 367-398. https://doi.org/10.1007/s10766-018-0620-8

Shaik, S., & Baskiyar, S. (2022). Distributed service placement in hierarchical fog environments. *Sustainable Computing: Informatics and Systems, 34*, Article 100744. https://doi.org/10.1016/j.suscom.2022.100744

Sharma, P., & Gupta, P. K. (2023). Optimisation of IoT-fog network path and fault tolerance in fog computing-based environment. *Procedia Computer Science, 218*, 2494-2503. https://doi.org/10.1016/j.procs.2023.01.224

Suliman, S. I., Nazman, H., Harun, A. F., Mohamad, R., Kassim, M., Farah, Y., … Yusof, Y. W. M. (2021). An effective energy-efficient virtual machine placement using a clonal selection algorithm. *International Journal of Advanced Technology and Engineering Exploration, 8*(75), 412-424. https://doi.org/10.19101/IJATEE.2021.875006

Varshney, S., Sandhu, R., & Gupta, P. K. (2020). QoE-based multi-criteria decision making for resource provisioning in fog computing using the AHP technique. *International Journal of Knowledge and Systems Science, 11*(4), 17-30. https://doi.org/10.4018/IJKSS.2020100102

Yuhuai, P., Xiaoliang, G., Xinyu, Z., Lei, L., Cemulige, W., & Lei, H. (2024). A cloud-edge collaborative computing framework using potential games for space-air-ground integrated IoT. *EURASIP Journal on Advances in Signal Processing, 2024*(1), Article 54. https://doi.org/10.1186/s13634-024-01122-6

Yusof, Y. W. M., Mohamed, A. H., & Kassim, M. (2023). Energy monitoring system design and analysis with the Internet of Things on the Blynk platform. *International Journal of Engineering Trends and Technology, 71*(3), 345-353. https://doi.org/10.14445/22315381/IJETT-V71I3P236

Zamri, A. H., Pakhrudin, N. S. M., Saaidin, S., & Kassim, M. (2023). Equally spread current execution load modelling with optimised response time brokerage policy for cloud computing. *International Journal of Advanced Computer Science and Applications, 14*(2). https://doi.org/10.14569/IJACSA.2023.0140257

REFEREES FOR THE PERTANIKA

JOURNAL OF SCIENCE & TECHNOLOGY

Vol. 34 (4) Aug. 2026

The Editorial Board of the Pertanika Journal of Science and Technology wishes to thank the following:

Abdullah Yildiz
(AYBU, Türkiye)

Adam Zuyyinal Adib
(UBB, Indonesia)

Adel A. Al-Azzawi
(NU, Iraq)

Adnan Ashraf Arain
(MUET, Pakistan)

Ahmed Abtan Al-Jumaili
(TU, Iraq)

Ahmed Al-Jumaili
(UFL, USA)

Amir Izzani Mohamed
(UMPSA, Malaysia)

Amirul Ridzuan Abu Bakar
(UniMAP, Malaysia)

Anith Khairunnisa Ghazali
(MMU, Malaysia)

Arfat Ahmad Khan
(KKU, Thailand)

Arun Kumar Singh
(SMIT, India)

Atiyyah Musa
(UiTM, Malaysia)

Azinoor Azida Abu Bakar
(UiTM, Malaysia)

Azizol Abdullah
(UPM, Malaysia)

Darius El Pebrian
(UiTM, Malaysia)

Fazirulhisyam Hashim
(UPM, Malaysia)

Firdaus Hamzah
(UKM, Malaysia)

Gökhan Başar
(OKU, Türkiye)

Hasfalina Che Man
(UPM, Malaysia)

Hasim Altan
(UAEU, United Arab Emirates)

Haslinda Mohd Sidek
(UMS, Malaysia)

Jafar A. Alzubi
(BAU, Jordan)

Juan Pratama
(UNSADA, Indonesia)

Kazim Yildiz
(MU, Türkiye)

Le Nguyen Quoc Khanh
(TMU, Taiwan)

Leonardo José Agamez Polo
(UNISINÚ, Colombia)

Mahadimenakbar Mohamed Dawood
(UMS, Malaysia)

Maziyah Mat Noh
(UMPSA, Malaysia)

Miszaina Osman
(UNITEN, Malaysia)

Mohamed Thariq Hameed Sultan
(UPM, Malaysia)

Mohammed Ahmed Mohammed Saleh
(UMS, Malaysia)

Mohammed Khalil Hussain
(UOB, Iraq)

Mohd Azrizal Fauzi
(UiTM, Malaysia)

Mohd Hizami Mohd Yusoff
(UTP, Malaysia)

Mohd Ikhmal Haqeem Hassan
(UNISEL, Malaysia)

Mohd Khairul Nizam Mohd Zuhan
(UTHM, Malaysia)

Mohd Shahrimie Mohd Asaari
(USM, Malaysia)

Mohd Sharil Salleh
(UPNM, Malaysia)

Mourad Abdelkrim
(UKMO, Algeria)

Murali Matcha
(MU, India)

Nadiatul Adilah Ahmad Abdul Ghani
(UMPSA, Malaysia)

Noor Azammi Abdul Murat
(UniKL, Malaysia)

Nor Arymaswati Abdullah
(NUKLEAR MALAYSIA, Malaysia)

Norzila Othman
(UTHM, Malaysia)

Nur Hidayati
(UMS, Indonesia)

Nurulnadwan Aziz
(UiTM, Malaysia)

Nuzaimah Mustafa
(UTeM, Malaysia)

Pavol Bozek
(STUBA, Slovakia)

Rajesh Nithyanandam
(SVCE, India)

Rashid Abbasi
(UESTC, China)

S. Raja
(SSIT, India)

Sadiq Hussain
(DU, India)

Sales Aribe Jr.
(BukSU, Philippines)

Satya Venkata Kishore Pulavarthi
(VFSTR, India)

Satyam Panchal
(UWaterloo, Canada)

Scheherazed Dakhmouche Djekrif
(UFMC1, Algeria)

Sinar Azuria Adnan
(UniMAP, Malaysia)

Siti Nur Khadijah Aishah Ibrahim
(UTM, Malaysia)

Suleyman Adak
(MAU, Türkiye)

Supakit Chotigo
(KMUTT, Thailand)

Syarifuddin Misbari
(UMPSA, Malaysia)

Thamer Ahmad Mohammad
(UOB, Iraq)

Thulasi Bikku
(AMRITA, India)

Yogish H. K.
(MSRIT, India)

AMRITA - *Amrita Vishwa Vidyapeetham*
AYBU - *Ankara Yıldırım Beyazıt University*
BAU - *Al-Balqa Applied University*
BukSU - *Bukidnon State University*
DU - *Dibrugarh University*
KKU - *Khon Kaen University*
KMUTT - *King Mongkut's University of Technology Thonburi*
MAU - *Mardin Artuklu University*
MMU - *Multimedia University*
MSRIT - *M. S. Ramaiah Institute of Technology*
MU - *Marmara University*
MU - *Medi-Caps University*
MUET - *Mehran University of Engineering and Technology*
NU - *Al-Nahrain University*
OKU - *Osmaniye Korkut Ata University*
SMIT - *Sikkim Manipal Institute of Technology*
SSIT - *Sri Siddhartha Institute of Technology*
STUBA - *Slovak University of Technology in Bratislava*
SVCE - *Sri Venkateswara College of Engineering*
TMU - *Taipei Medical University*
TU - *Tikrit University*
UAEU - *United Arab Emirates University*
UBB - *Universitas Bangka Belitung*
UESTC - *University of Electronic Science and Technology of China*
UFMC1 - *Université Frères Mentouri Constantine 1*
UFL - *University of Florida*
UiTM - *Universiti Teknologi MARA*
UKM - *Universiti Kebangsaan Malaysia*
UKMO - *University of Kasdi Merbah Ouargla*
UniKL - *Universiti Kuala Lumpur*
UniMAP - *Universiti Malaysia Perlis*
UNISEL - *Universiti Selangor*
UNISINÚ - *Universidad del Sinú*
UNITEN - *Universiti Tenaga Nasional*
UNSADA - *Universitas Darma Persada*
UOB - *University of Baghdad*
UPM - *Universiti Putra Malaysia*
UPNM - *Universiti Pertahanan Nasional Malaysia*
USM - *Universiti Sains Malaysia*
UTeM - *Universiti Teknikal Malaysia Melaka*
UTHM - *Universiti Tun Hussein Onn Malaysia*
UTM - *Universiti Teknologi Malaysia*
UTP - *Universiti Teknologi PETRONAS*
UWaterloo - *University of Waterloo*
VFSTR - *Vignan's Foundation for Science, Technology & Research*
UMPSA - *Universiti Malaysia Pahang Al-Sultan Abdullah*
UMS - *Universitas Muhammadiyah Surakarta*
UMS - *Universiti Malaysia Sabah*

While every effort has been made to include a complete list of referees for the period stated above, however if any name(s) have been omitted unintentionally or spelt incorrectly, please notify the Chief Executive Editor, *Pertanika* Journals at executive_editor.pertanika@upm.edu.my

Any inclusion or exclusion of name(s) on this page does not commit the *Pertanika* Editorial Office, nor the UPM Press or the university to provide any liability for whatsoever reason.

Photovoltaic Module Defect Classification Analysis Based on Convolutional Neural Network Architecture with Augmentation Dataset 2487
Nurul Atikah Mazlan, Shahrani Shahbudin, and Murizah Kassim

Orchestration Network Hierarchy and Communication Patterns Architecture Analysis with Ranks for IoT Devices 2519
Nor Syazwani Mohd Pakhrudin, Murizah Kassim, and Azlina Idris

Short Communication 2329
Regional Comparison of Forest Certification Impacts: Environmental and Social Outcomes in Asia-Pacific and Europe
Rohaiza Abdul Kadir Zailani, Sarah Najwa Mohd Masri, Shah Badri Mohd Nor, Ahmad Fariz Mohamed, and Shaharuddin Mohamad Ismail

Hybrid Approach Combining Area-Based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE) in Handling Class Imbalance 2339
Hartono, Muhammad Khahfi Zuhanda, Rahmad Syah, Prana Ugiana Gio, and Erianto Ongko

Impact of Wind Farm Penetration on the Mal-Operation of Distance Relay Power Swing Blocking and Out-of-Step Functions 2359
Chidiebere Okeke, Mohammad Lutfi Othman, Hashim Hizam, Mohd Zainal Abdin Ab Kadir, Samuel Nwagbara, Osaji Emmanuel, Jude Aneke, Okeke Ugochukwu Kingsley, and Masoud Ahmadipour

Characterisation of Bamboo Fibre-Reinforced Polyurethane Foam Composites with Varying Fibre Sizes 2385
A.M. Radzi, M.Y.M. Zuhri, S. Ayu Rafiqah, R.A. Ilyas, Muhammad Asyraf Muhammad Rizal, Ahmad Safwan Ismail, A.B.M. Supian, and Agusril Syamsir

Review Article 2403
Technological Approach for Biodiversity Conservation: Augmented Reality Based on Systematic Review
Nurfitri Ezatie Yusoff, and Arney Sapaat

Valorisation of Food Waste and Leftovers into Functional Bio-Compost: Optimisation of Maturation Period and Growth Performance of *Ipomoea Aquatica* 2421
Sarah Dini Muhamad Din, Muhammad Heikal Ismail, Tharsini Amma Puspadaran, Anie Yulistyorini, Muhammad Hazwan Hamzah, and Chuah Teong Guan

Mechanical Testing Analysis of Hybrid Composites of Oil Palm Empty Fruit Bunch Fibre and Sugarcane Bagasse Fibre for Environmentally Friendly Car Body Materials 2441
Aris Sandi, Aulia Majid, and Syafiq Fadhlika

Review Article 2457
AI-Powered Revolution in Concrete Crack Analysis: An In-Depth Review on Detection and Classification Methods
Muhammad Shahrulazmi Hashim, Emedya Murniwaty Samsudin, and Khalid Isa

Explainable Crop and Soil-Based Recommendation System for Next-Generation Smart Agriculture Using Model Classification 2127
Himani Trivedi, Hetal Chauhan, Suresh Patel, Mahendra N Patel, Pradip Patel, Mahi Parmar, and Chrisha Dabhi

Utilisation of Starch-Derivative Functionalised Biochar for Management of Industrial Effluent 2149
Maria Adilah Basharuddin, Siti Fatimah Zaharah Mohamad Fuzi, Noor Mohd Syeqqal Ismail, Chia Kim Seng, and Nor Hasmaliana Abdul Manas

Dynamic Sensor Activation for Energy-Efficient Event Monitoring in Large-Scale IoT Deployments 2167
A. Subramaneswara Rao, Suresh Kallam, and Srinivasa Rao Pappu

Calibration and Limitations of Hardening Soil and Soft Soil Creep Models for Undrained Behaviour of Tropical Peat 2181
Ahmad Kamal Abd Aziz, Norazzlina M. Sa'don, and Abdul Razak Abdul Karim

Assessing the Drought Index and Water Scarcity Nexus: Case Study for Langat River Basin 2213
Raja Muhamad Naufal Raja Azman, Nur Asmaliza Mohd Noor, Marini Mohamad Ideris, and Sharifah Abdullah

An Interactive Augmented Reality and Digital Twin Framework for IoT-Based Aquaponics in Urban Farming 2233
Siti Khadijah Ali, Phang Kok Wai, Zainal Abdul Kahar, and Rahmita Wirza O.K. Rahmat

A Framework for Secure Software Engineering in Hybrid Agile Software Development 2255
Normi Sham Awang Abu Bakar , Norzariyah Yahya, and Siti Sarah Maidin

Evaluation of the Colour Fastness Properties of Bixa orellana-Dyed Silk Using CIELab Analysis 2275
Nur Atiqah Jamaludin, Razanah Ramya, and Rashidi Othman

Partial Discharge Detection and Localisation in Refined, Bleached and Deodorised Palm Olein based on Ultra-High Frequency Sensing Technique 2293
Nur Darina Ahmad, Norhafiz Azis, Zhi Yang Lim, Jasronita Jasni, Khairil Anas Md Rezali, Mohd Aizam Talib, Nik Hakimi Nik Ali, and Masahiro Kozako

Recycling Indonesian Native Chicken Eggshell Membrane Waste into Gelatine: A Comparative Characterisation Using SEM-EDX, FTIR, and XRF 2311
Rina Wahyuningsih, Ragil Yuliatmo, Andi Febrisiantosa, Atiqa Rahmawati, Al Mizan, Selma Noor Permadi, Laila Rahmawati, Afi Candra Trinugraha, and Teguh Wahyono

Pertanika Journal of Science & Technology
Vol. 34 (4) Aug. 2026

Content

Foreword xiii
Mohamed Thariq Hameed Sultan

Review Article 1953
Comprehensive Review of Plant Disease Detection: Advancements in Imaging Sensors, AI Techniques, and Future Directions in Smart Agriculture
Binal Upadhyay and Amit Ganatra

Smart Comment Manager: Streamlining Comment Analysis and Viewer Query Clustering Using NLP Processing 1981
N Venkata Sailaja, Karnam Akhil, Sai Shashidhar Reddy Katanguri, Maringanti Gopi Krishnna, Alladi Shiva Vinay, and Devireddy Siddartha Reddy

AI meets Reproductive Health: Early Diagnosis of PCOS Using Machine Learning 2003
Abhinav Pathak, M. Sujithra, Hemashree P, and Arun Kumar Selvam

Parametric Study of Reaction Parameter for Glycerol Acetylation and Reusability of Sulfated-Titania Catalyst 2023
Shera Farisya Mohamad Rasid, Ernee Noryana Muhamad, Yun Hin Taufiq-Yap, Haizatul Izzatti Ridzuan, Siew Ling Lee, and Irmawati Ramli

Data-Driven Fault Detection and Diagnosis of a Nuclear Reactor Cooling System Using Graph Neural Networks 2053
Dhines Shree Kanthan, Nur Aina Mohammad Aziz, Azura Che Soh, Ribhan Zafira Abdul Rahman, Mohd Khair Hassan, Mohd Sabri Minhat, and Julia Abdul Karim

Impact of the Grading Ring Size and Hydrophobicity on the High-Voltage Insulator Performance 2075
Amelia Adrina Azli, Nor Aqillah Aiman Nor Azman, N.H. Nik Ali, Hazlee Azil Illias, Ali Ahmed Ali Salem, Mohd Aizam Talib, and Mohd Fairouz Mohd Yousof

Review Article 2097
Prediction of Lung Diseases Using Deep Learning Techniques: A Review
Sheikh Afaan Farooq and Aleem Ali

Pertanika Editorial Office, Journal Division,
Putra Science Park,
1st Floor, IDEA Tower II,
UPM-MTDC Center,
Universiti Putra Malaysia,
43400 UPM Serdang,
Selangor Darul Ehsan
Malaysia

http://www.pertanika.upm.edu.my
Email: executive_editor.pertanika@upm.edu.my
Tel. No.: +603- 9769 1622

PENERBIT UPM
UPM PRESS

http://penerbit.upm.edu.my
Email: penerbit@upm.edu.my
Tel. No.: +603- 9769 3606

e-ISSN 2231-8526
9 772231 852000