Vol. 45 | Special Issue 1 | June 2026

MALAYSIAN JOURNAL OF SCIENCE

Special Issue on FAST-UTHM GLOBAL NEXUS CONFERENCE 2024: 9th International Conference on the Application of Science and Mathematics (SCIEMATHIC) 2024

MJS Guest Editors
Associate Professor Ts. Dr. Zul Ilham Zulkiflee Lubes, Universiti Malaya
Dr. Mohd Arif Mohd Sarjidan, Universiti Malaya
Dr. Loo Tee How, Universiti Malaya

Special Issue Editors
Dr. Mohd Zieauddin Kufian, Universiti Malaya
Dr. Najiah Mohd Sadali, Universiti Malaya

Guest Editors
Dr. Shakila Abdullah, Universiti Tun Hussein Onn Malaysia
Dr. Siti Amira Othman, Universiti Tun Hussein Onn Malaysia
Dr. Nurul Nadia Adnan, Universiti Tun Hussein Onn Malaysia

Published: 30-06-2026

MALAYSIAN JOURNAL OF SCIENCE

MJS

ISSN 1394-3065

Malaysian Journal of Science is indexed in Scopus, EMBASE, Compendex, GEOBASE, EMBiology, Elsevier BIOBASE, FLUIDEX ,World Textiles, CAB Abstracts, Chemical Abstracts Service Database and ASEAN Citation Index (ACI).

Special Issue on FAST-UTHM GLOBAL NEXUS CONFERENCE 2024: 9th International Conference on the Application of Science and Mathematics (SCIEMATHIC) 2024

We are honoured to present the selected papers from the 9th International Conference on the Applications of Science and Mathematics (SCIEMATHIC) 2024, held at The Everly Hotel, Putrajaya, Malaysia, on October 3–4, 2024. We trust that all participants have gained meaningful insights and enriching experiences in the fields of science and mathematics through their engagement in the conference.

The primary aim of this conference is to foster meaningful dialogue and the exchange of knowledge among experts in science, mathematics, and biodiversity, with a focus on advancing a sustainable future. Additionally, the conference seeks to empower scientists and professionals to communicate scientific and mathematical knowledge in ways that promote harmony with nature and encourage a holistic understanding for future generations. It also serves as a platform to highlight and celebrate successful examples of integrating science, mathematics, and biodiversity in support of global sustainability initiatives.

Distinguished keynote speakers, including prominent professors from Malaysia and Indonesia, were invited to share the latest developments and breakthroughs in their respective fields. This conference provides an invaluable opportunity for students and researchers to disseminate their findings and exchange expertise, while also enhancing the university's visibility and recognition at both national and international levels.

This issue features a curated selection of 12 papers submitted by contributors from academic institutions, research organizations, and industry. All submissions underwent a rigorous peer-review process conducted by the conference committee in collaboration with external reviewers. The selected papers were evaluated based on their academic merit and alignment with the conference themes. The content is organized into six core categories: mathematics and statistics, physics, chemistry, biology, engineering sciences, and artificial intelligence.

We extend our heartfelt appreciation to all contributing authors for their scholarly work. Our sincere gratitude also goes to the organizing committee, reviewers, keynote speakers, session chairs, and all participants for their steadfast support and contributions to the success of SCIEMATHIC 2024.

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

https://mjs.um.edu.my

Surface Temperature and Pressure Drop Simulation Analysis of Nanofluid Flow in Microchannel Heat Sink

Nurafiqah Mohd Alhata[1a], Amira Saryati Ameruddin[2a]*, Rosmila Abdul-Kahar[3a] and Mirza Basyir Rodhuan[4a]

Abstract: The increasing demand for high-performance electronic devices has driven the need more effective thermal management solutions. Standard air-cooling methods fall short when handling high heat flow rates. Among the most promising solutions is the Micro Channel Heat Sink (MCHS), specifically designed to address this challenge. To draw heat away from the chip, a cold liquid navigates intricate channels within the electronic system. The design of the heat sink profoundly influences crucial factors, such as microchannel area, heatsink size, channel width, height, and inlet/outlet diameters, each of which is affected differently by the heat sink's configuration. This study investigated heat transfer performance across various thicknesses, exploring nanoliquids, air, and water as cooling agents. After reviewing the setup steps, the ideal fluid performance for the given design is identified. Using COMSOL Multiphysics 5.5 software, a 3D model of a rectangular copper MCHS is created employing the finite element method. Results indicate that a 5 mm channel thickness using Aluminum oxide (Al_2O_3) nanofluids yields the most promising thickness based on the findings. Al_2O_3 exhibits outstanding performance, maintaining momentum even at lower temperatures. The influence of fluid flow velocity on the MCHS is reflected in the 5 mm channel thickness, significantly impacting the effective pressure drop. One avenue for potentially enhancing performance involves considering the aspect ratio of nanofluidic particles and exploring diverse materials such as iron and aluminum.

Keywords: Nanofluid, Heat transfer, Temperature, Pressure drop, Microchannel heat sink

1. Introduction

MCHS have emerged as pivotal components for cooling high-performance electronic devices since their inception by Tuckerman and Pease (1981). Traditional air-cooling methods struggle to handle the high heat generated by modern devices, prompting the search for more effective cooling techniques. Liquid cooling through MCHS, where chilled liquid flows through tiny channels to remove heat from the chip, has proven to be one of the most efficient solutions (Van Oevelen, 2014). These channels are built into the heat sink, which is placed on or embedded within the chip to ensure optimal cooling. MCHS are typically made from highly conductive materials, such as copper and aluminum. Copper has better thermal conductivity, while aluminum is lighter and more affordable. Many designs use both materials to balance performance, weight, and cost, ensuring efficient cooling of electronic devices.

One of the main challenges in optimizing MCHS is enhancing heat transfer efficiency. Adjusting the shape, size, and orientation of the microchannels can significantly improve heat transfer and reduce pressure drop. For example, triangular fins are more effective at transferring heat than rectangular ones, although specific designs, such as angled ribs, improve heat transfer while increasing pressure resistance (Behnampour *et al.,* 2017; Liou, Chang, & Chan, 2018). In addition, the size of the channels and the inlet/outlet ports also play a crucial role in thermal efficiency. Numerical simulations help engineers analyze these variables, allowing them to optimize designs for better performance without increasing costs (Dhaiban & Hussein, 2020).

Using advanced coolants, such as nanofluids, which are liquids containing tiny nanoparticles, has been shown to significantly enhance heat transfer rates. Nanofluids, such as Al_2O_3 mixed with water, can boost cooling efficiency by up to 8.58% compared to regular water (Elbadawy *et al.,* 2023). However, challenges like sedimentation and increased fluid viscosity must be addressed to ensure their long-term effectiveness (Rafiq, 2021). The performance of nanofluids is influenced by factors such as particle size, material type, and temperature. Preventing particle clumping and maintaining fluid stability over time are essential for maximizing their thermalbenefits. Ongoing research focuses on improving nanofluid stability and understanding their behavior using advanced simulations, particularly the Finite Element Method (FEM) and Computational Fluid Dynamics (CFD), which are key tools for enhancing MCHS designs (Nazzal *et al.,* 2021; Gao *et al.,* 2022).

This research used COMSOL Multiphysics software to run detailed simulations that model heat transfer and fluid flow in 3D microchannel designs. It is ideal for studying MCHS to improve cooling performance metrices. Building on earlier work by Gunnasegaran *et al.,* (2012), which demonstrated how

Authors information:
[a]Department of Physics, Faculty of Applied Sciences and Technology (FAST), Universiti Tun Hussein Onn Malaysia (UTHM), Pagoh Branch Campus, Pagoh Higher Education Hub, Km 1, Jalan Panchor, 84600 Muar, Johor Darul Ta'zim, MALAYSIA.
E-mail: gw220014@student.uthm.edu.my[1], amira@uthm.edu.my[2], ammaz5011@gmail.com[3], mirzabasyir95@gmail.com[4]

*Corresponding Author: amira@uthm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

nanofluids, such as Al_2O_3 can enhance heat transfer compared to air cooling, this study extends to investigation by testing different channel thicknesses. It finds that a 5 mm thickness works best for efficient cooling and examines how fluid temperature, velocity, and novel materials, such as iron and aluminum, can enhance heat transfer. The study also explores innovative methods to enhance MCHS performance through FEM simulations.

2. Materials and method

Three key parameters must be observed in the design: heat transfer performance in solids and fluids, surface temperature, and pressure. The design in this study is adapted from research conducted by Gunnasegaran and his team (Gunnasegaran *et al.,* 2012). They focused on the 3D simulation of heat conduction, convection, and laminar flow behavior with various heat sink shapes, the effect of nanoparticle volume fraction, different types of nanofluids, and nanoparticles in different base fluids. In this paper, the heat performance of various geometry thicknesses of heat sinks and different types of nanofluids is examined. The influence of nanofluids on varying thicknesses of heat sinks is the main focus. The results are presented in terms of surface temperature and pressure.

2.1 Geometry

The inlet/outlet area, channel width, channel height, channel length, heat sink size, and copper microchannel area are the parameters examined in this study. Figure 1 shows the overall design of the inlet and outlet of the copper rectangular MCHS design in 3D. The design remain the same for the 5mm and 8mm thicknesses of the heat sink channel with different types of nanofluids. The parameter details of the MCHS design used in the simulation are shown in Table 1. Figure 2(a) and Figure 2(b) show the XZ-cut plane view and YZ-cut plane view, which were used as guidelines to present the results in the next section.

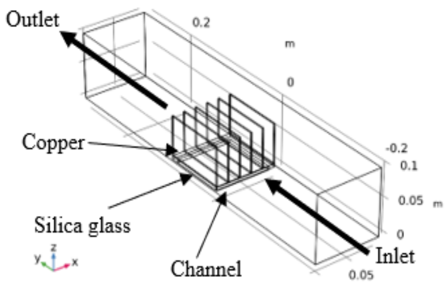


Figure 1. Copper rectangular microchannel heat sink.

Table 1. The parameters of copper rectangular microchannel heat sink.

Parameter	Dimension
Inlet/outlet area, mm	42.5 x 50
Channel width, mm	10
Channel height, mm	50
Channel length, mm	95
Microchannel area, mm^2	60 x 95

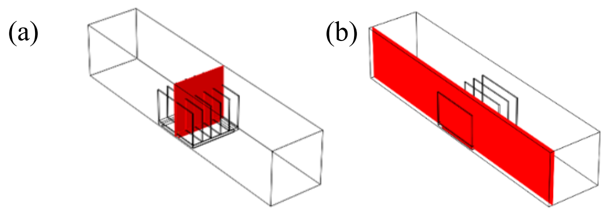


Figure 2. Microchannel heat sink (a) XZ-cut plane view and (b) YZ-cut plane view.

2.2 Nanofluids

The model design begins with an electronic chip placed at the bottom of the heat sink as a heat source made from silica glass. The copper MCHS acts as the path for nanofluids to flow and is surrounded by a domain. The physical interface is transferred from the electronic chip using nanofluids, air, and water as the fluid flows from the inlet to the outlet of the heat sink along the MCHS. As the nanofluids, air, and water flow in the rectangular channel, the thermal balance for the electronic component was optimized. In this part, the properties of all the nanofluids, air, and water were defined by inserting them into COMSOL. Thermophysical properties of solids and fluids for materials and nanofluids used in this study are shown in Tables 2 and 3.

The thermophysical properties of the solid materials and working fluids presented in Tables 2 and 3 are adapted from Patel *et al.,* (2019), Gunnasegaran *et al.,* (2012), and Mohammed *et al.,* (2010). Nanoparticle constants (Al_2O_3, SiO_2, and ZnO) and the effective nanofluid properties were determined using the conventional mixture relations described by Pak and Cho (1998) and Hamilton and Crosser (1962), validated experimentally by Sohel *et al.,* (2014) and theoretically by Malvandi and Ganji (2014). These references confirm that, for dilute nanoparticle concentrations (≤ 4 vol %), the nanofluid density remains very close to that of the base fluid, differing by less than 3 %. Base-fluid properties for water and air were taken at 293 K following Patel *et al.,* (2019), Mohammed *et al.,* (2010), and the National Institute of Standards and Technology (NIST) standard property tables. The solid-phase properties of copper and silica glass were verified against those reported by Patel *et al.,* (2019) and Gunnasegaran *et al.,* (2012) to ensure consistency with established MCHS studies.

2.3 Equations

This section includes the equations of heat transfer in solids and fluids, laminar flow, temperature, pressure drop, and boundary conditions for heat transfer in solids and fluids, and laminar flow is solved in COMSOL Multiphysics. Heat transfer in solids and fluids is defined by the heat source, heat rate, heat flux, inflow, and outflow, as describe in the COMSOL Reference Manual. Eq (1) is the equation for the heat source. Inserting a value for the heat rate P_0 in the overall heat transfer rate field as

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

Table 2. Thermophysical properties of solid.

Thermophysical properties	Copper	Silica glass
Density, ρ_s (kg/m^3)	8690	8690
Thermal Conductivity, κ (W/m·k)	400	400
Relative Permittivity, ϵ_r	1	3.9
Heat Capacity at Constant Pressure, C_p (J/kg·k)	385	385

Table 3. Thermophysical properties of fluid.

Thermophysical properties	Al_2O_3	SiO_2	ZnO	Water	Air
Density, ρ_s (kg/m^3)	3880	2200	5600	997	1.23
Thermal Conductivity, κ (W/m·k)	30	703	113	148	0.025
Relative Permittivity, ϵ_r	8.5	3.9	5	78.3	1
Heat Capacity at Constant Pressure, C_p (J/kg·k)	791	745	495	4182	1006

mentioned in Eq (2). Q_0 as the heat per unit volume, as a linear heat source, or as a heat rate and V is the total volume.

$$Q = Q_0 \quad (1)$$

$$Q = \frac{P_0}{V} \quad (2)$$

$$U = -nU_0 \quad (3)$$

In this study, the laminar flow options available in the boundary condition options for the inlet and outlet are pressure and velocity. Eq (3) shows the normal inflow velocity, where n is the boundary normal pointing out of the domain, and U_0 is the normal inflow speed that is defined by the simulation software.

Eqs (4) and (5) are the inlet boundary conditions, and Eq (6) is the outlet boundary condition (Memon *et al.,* 2020). Below are the equations for the inlet and outlet boundary conditions:

$$T_f = T_{in} = 293K \quad (4)$$

$$u = u_{in} \quad (5)$$

$$P = P_{out} = 1atm \quad (6)$$

For the above equations, T_{in} and u_{in} are the temperature and velocity at the inlet of the heat sink. T_f is the temperature field inside the fluid. P_{out} is the pressure at the outlet. The temperature of each channel, shown in Eqs (7) and (8), was used to calculate the pressure drops along the length of the heat sink channel (Gunnasegaran *et al.,* 2012).

$$\theta = \frac{T_f - T_i}{T_w - T_i} \quad (7)$$

$$\hat{p} = \frac{\Delta P}{\rho u_{in}^2} \quad (8)$$

T_f is a fluid temperature, T_i is an inlet temperature and T_w is a wall temperature. For the Eq (8), ΔP is the pressure drop of the microchannel configuration, ρ is the water density and u_{in} is the inlet water velocity (Mohammed, Gunnasegaran, & Shuaib, 2010).

2.4 Meshing and validation

The MCHS design of the entire model used an "extremely coarse" mesh, as shown in Figure 3, to reduce computational cost while still capturing essential flow and thermal behavior, featuring 6,770 triangular and 68 vertex elements for a 2 mm channel thickness, 7,164 triangular and 68 vertex elements for 5 mm, and 6,644 triangular and 68 vertex elements for 8 mm. The minimum element quality improved from 0.114 (2 mm) to 0.2626 (5 mm) and 0.2828 (8 mm), indicating better element shapes with increasing thickness. The coarse meshing was applied because the simulation focused on broad thermal and fluid patterns, where fine local details are less critical. Boundary conditions, such as inlet/outlet flow, wall heat transfer, and symmetry, ensure realistic thermal-fluid behavior while maintaining efficient computation.

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

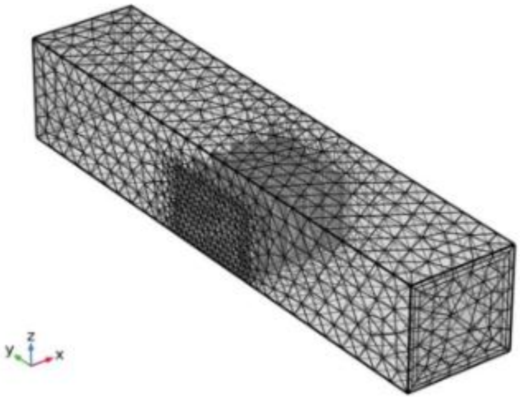


Figure 3. Meshing result by "Extremely coarse" of the model.

3. Result and discussion

This section discusses and analyzes the heat transfer performance, surface temperature, and pressure drop in a copper rectangular microchannel heat sink with varying thicknesses in Aluminum Oxide (Al2O3), Silicon Oxide (SiO2), and Zink Oxide (ZnO) nanofluids, as well as air and water. Temperature surface plots are used to present the results of temperature distributions on the boundaries of a model's geometry of the heat sink. The impact of different design conditions shows the increase or decrease in the temperature and pressure drop.

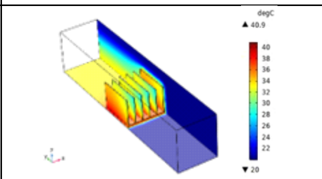
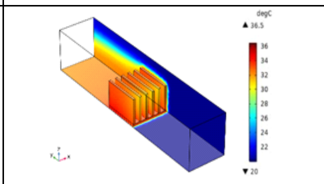
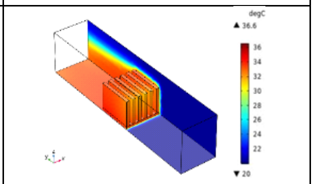
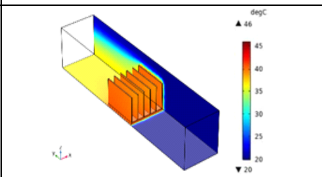
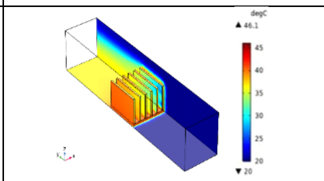
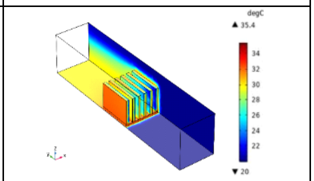
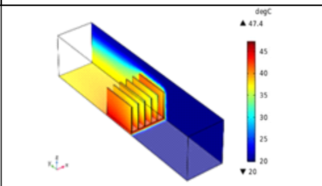
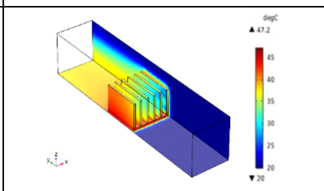
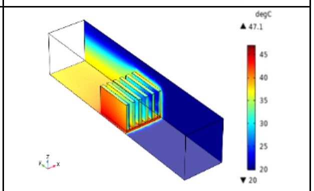
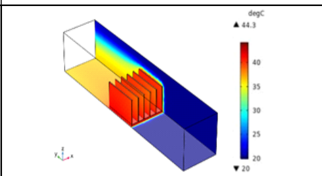
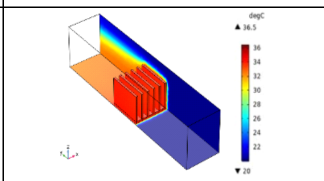
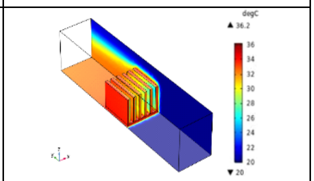
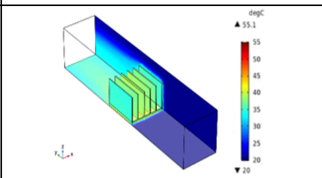
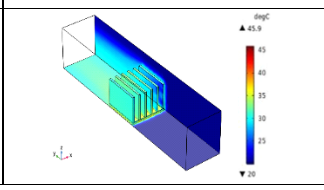
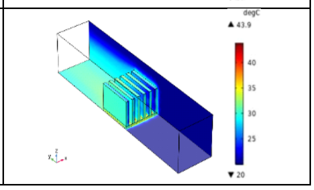
As the nanofluid passes through the heat sink channels, the heat is transferred from the electronic chip to the microchannels by conduction, convection, and radiation due to direct solid-liquid contact within the heat sink, as shown in Tables 4 and 5. The best nanofluids enhance heat removal performance at lower temperatures. The velocity of the fluid flow affects the pressure of the heat sink. According to Bernoulli's Principle, when the velocity increases, the pressure decreases. The pressure of the heat sinks changes from high to low pressure from the inlet to the outlet. The pressure drop was determined by measuring the pressure changes along the MCHS.

3.1 The effect of MCHS thickness and cooling element on the temperature and pressure surface

The copper rectangular MCHS acted as a device cooling system that enhanced the heat transfer performance while allowing nanofluids to flow along the channel. Tables 4 to 7 show the results of the temperature and pressure surface data with different nanofluids, air, and water from various views based on Figures 1, and Figure 2 (a and b).

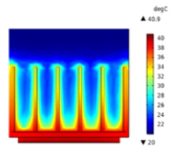
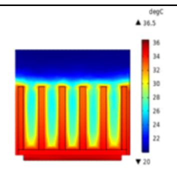
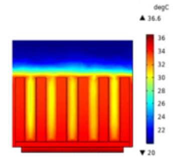
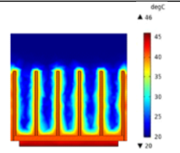
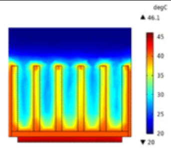
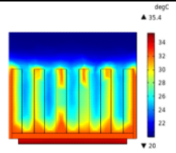
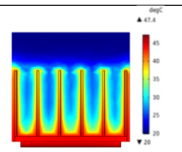
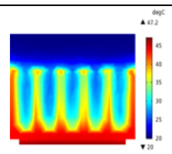
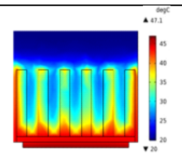
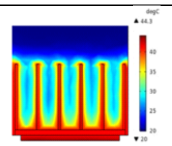
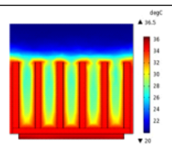
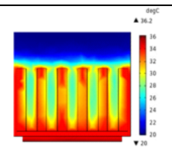
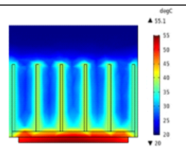
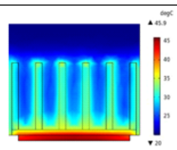
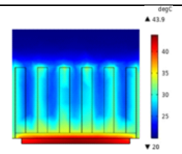
As shown in Tables 4 and 5, the 5 mm channel, utilizing Al_2O_3 nanofluid, achieves the lowest maximum temperature of 27.39˚C, a result attributed to Al_2O_3's superior thermal conductivity, which enable efficient heat dissipation from electronic components.

Table 4. Surface temperature of microchannel heat sink from 3D view.

Nanofluids	2mm	5mm	8mm
Al_2O_3			
SiO_2			
ZnO			
Air			
Water			

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

Table 5. Surface temperature of microchannel heat sink from top view.

Nanofluids	2mm	5mm	8mm
Al_2O_3			
SiO_2			
ZnO			
Air			
Water			

Water and ZnO in 2 mm and 8 mm channels also demonstrate effective cooling capabilities, with the lowest maximum temperatures of 27.44˚C and 29.18˚C, respectively. This performance is attributed to water's favourable flow properties and ZnO's thermal conductivity, which together improve heat removal and regulation in confined spaces.

Conversely, the 2 mm channel records the highest temperature when using ZnO nanofluid, reaching 39.52˚C. This suggests that ZnO's lower thermal conductivity restricts its performance in managing heat within narrower channels. Additionally, constrained fluid dynamics in smaller channels can further reduce effective heat absorption. Although the 8 mm channel offers a greater surface area for heat exchange, certain nanofluids, such as SiO_2, still show elevated temperatures (32.45˚C), indicating that larger channels do not consistently enhance cooling performance across all nanofluid types.

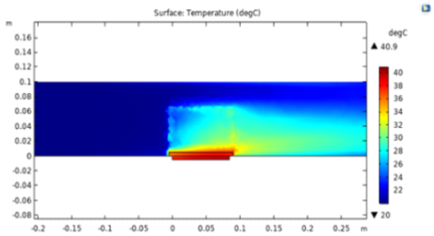
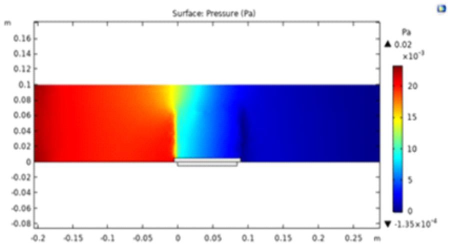
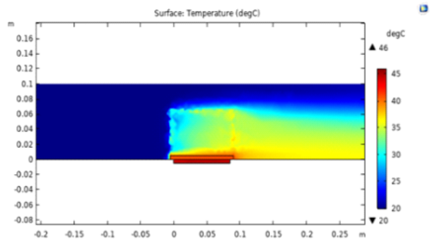
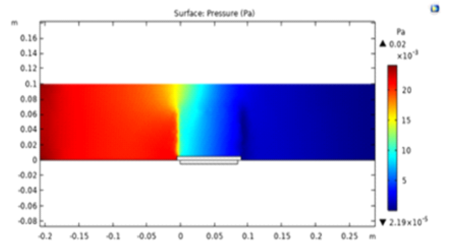
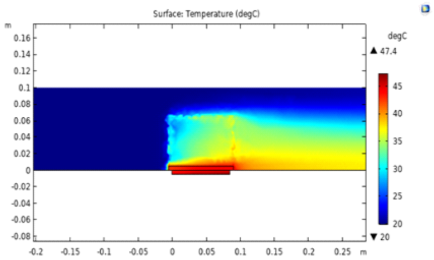
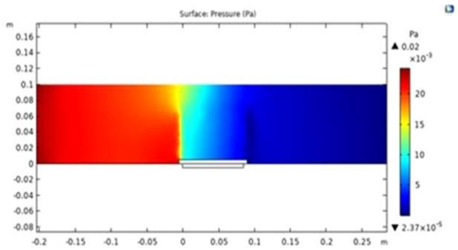
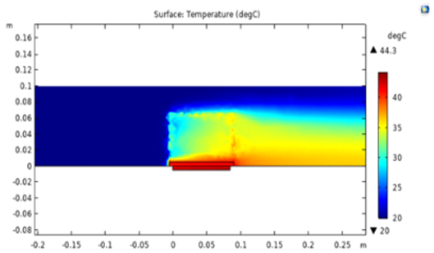
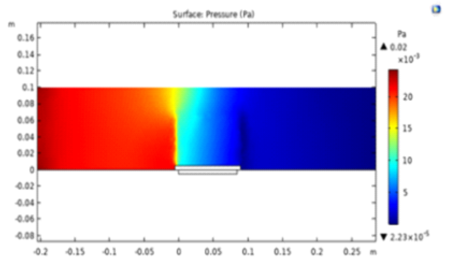
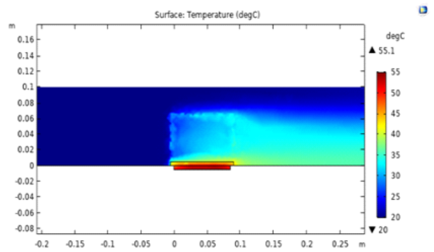
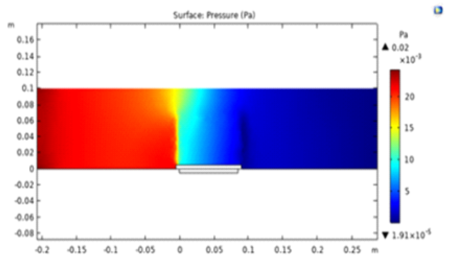
In summary, optimizing cooling performance requires careful selection of both the nanofluid and channel thickness. Nanofluids with high thermal conductivity, such as Al_2O_3, combined with appropriately sized channels, can substantially improve heat regulation in MCHS. As the fluids traverse the heat sink, temperatures decrease through conduction, convection, and radiation, with the ideal case being the lowest achievable temperature. These findings are crucial for advancing the design of heat sinks to improve the efficiency of electronic cooling systems.

Tables 6 and 7 present the results for 2 mm and 5 mm channel thicknesses from the 2D YZ-cut plane. The 8 mm channel results are shown in Table 8 for completeness. However, they are excluded from the comparative analysis as they exhibit anomalous thermal and hydraulic behavior that diverges from the consistent patterns observed for smaller channel thicknesses. The analysis focuses on comparing the 2 mm and 5 mm channels, where the highest temperatures are observed at the bottom of the heat sink above the electronic chip. The temperature profiles indicate that the 2 mm channel reaches higher maximum temperatures (40˚C to 56˚C) compared to the 5 mm channel (36˚C to 48˚C), primarily due to the thinner channel's reduced surface area for heat dissipation and increased localized heat accumulation. Although the 2 mm channel experiences intense convective heat transfer, its restricted flow velocity limits the nanofluid's heat removal capability.

Conversely, the 5 mm channel benefits from improved fluid circulation and heat transfer, resulting in a more balanced thermal environment and a lower temperature gradient. In a comparative analysis of pressure drop across various channel sizes in MCHS, it was observed that all nanofluids in the 2 mm channel outperform water, exhibiting the lowest pressure drop, which suggests superior flow behavior and reduced resistance. As channel size increases to 5 mm, both SiO_2 and ZnO nanofluids continue to maintain relatively low pressure drops, indicating their effectiveness in enhancing fluid flow, albeit with a slight increase in resistance due to larger cross-sectional area and potential turbulence. Conversely, in the 8 mm channel, all tested nanofluids experience higher pressure drops compared to the smaller channels, demonstrating the challenges of sustaining efficient flow in larger diameters. Notably, ZnO stands out as the

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

Table 6. Temperature and pressure surface of 2mm channel thickness from 2D YZ-cut plane.

Nanofluids	Temperature (℃)	Pressure (Pa)
Al_2O_3		
SiO_2		
ZnO		
Air		
Water		

only nanofluid that retains a low pressure drop in this context, indicating its unique characteristics may help maintain flow efficiency under increased channel size. Overall, these results illustrate that the 2 mm channel, particularly when using nanofluids, provides the most effective option for minimizing pressure drop.

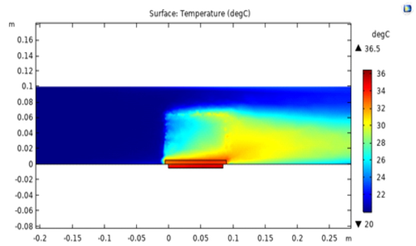
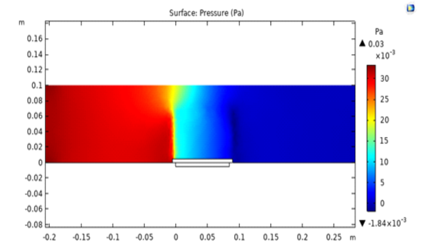
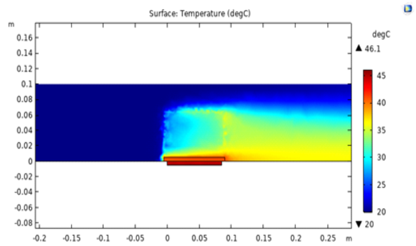
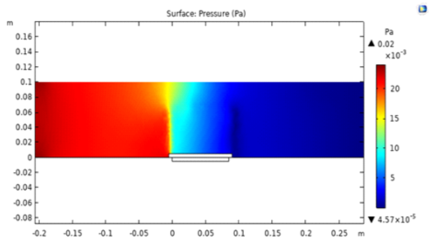
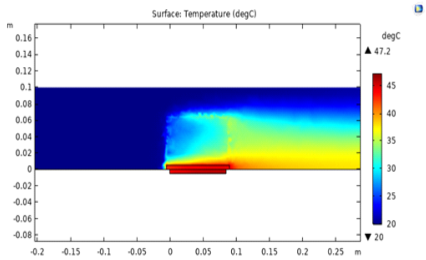
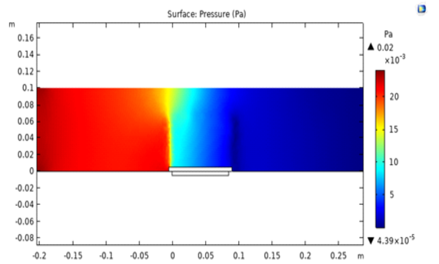
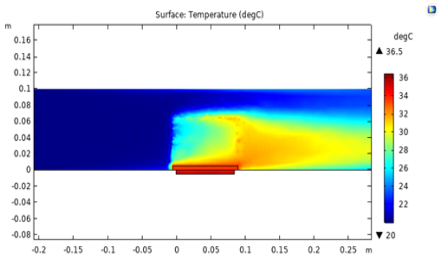
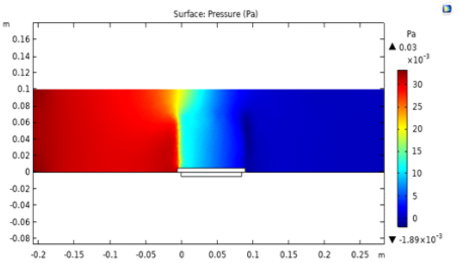
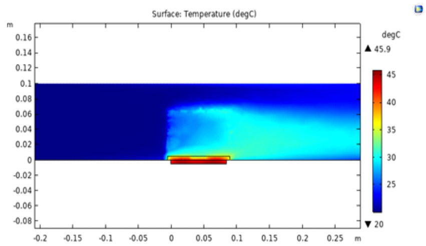
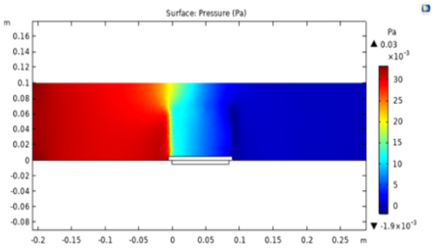
The measurements of channel thickness and coolant results were taken from -0.2 m (inlet) to 0.2 m (outlet), with the midpoint at 0 m along the x-axis. Overall, the lowest temperature was observed at the inlet, while the highest temperature occurred in the middle section. As the heat moved towards the outlet, the temperature decreased. The findings are detailed in Table 8, which analyzes the data for each channel thickness and nanofluid.

The results indicate that the 5 mm channel thickness achieves the lowest maximum temperature in the middle section, likely due to enhanced heat transfer and a more balanced flow distribution, allowing the coolant to absorb heat more effectively.

In contrast, the 2 mm channel thickness exhibits the lowest pressure drop, suggesting faster fluid velocity but potentially less time for effective heat absorption. This illustrates a trade-off between thermal efficiency and pressure loss, which is essential for optimizing heat sink designs based on specific cooling requirements.

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

Table 7. Temperature and pressure surface of 5mm channel thickness from 2D YZ-cut plane view.

Nanofluids	Temperature (℃)	Pressure (Pa)
Al_2O_3		
SiO_2		
ZnO		
Air		
Water		

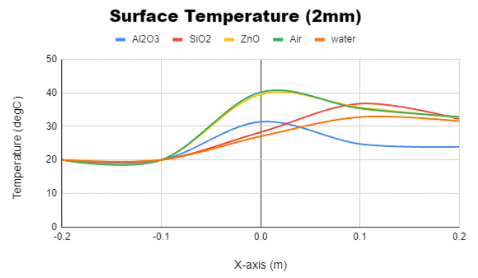
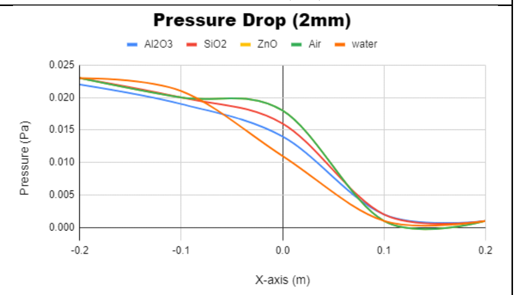
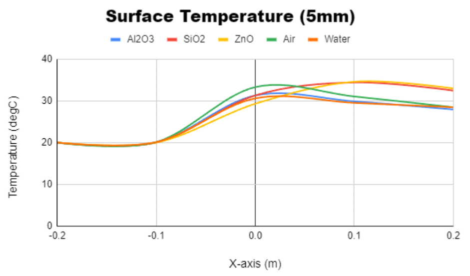
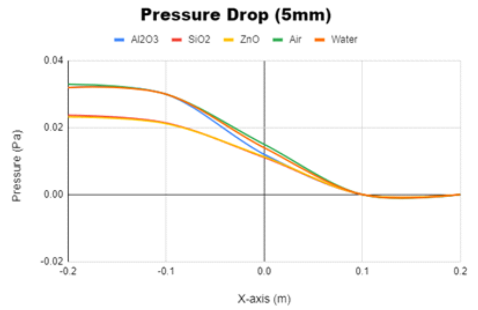
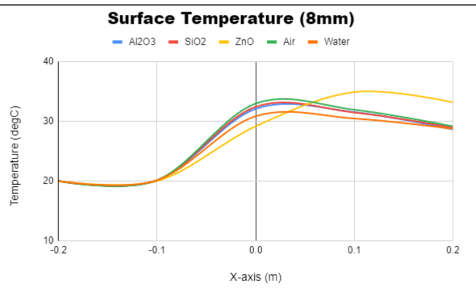
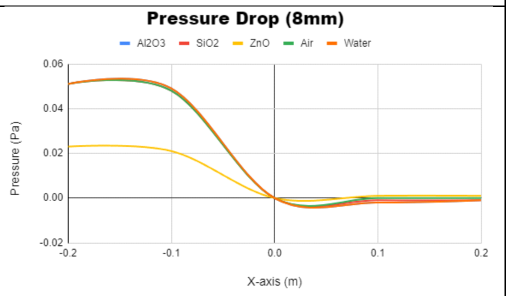
3.2 The effect of channel thickness on pressure drop

Measurements of different channel thicknesses and coolant temperature were taken from -0.2 m (inlet) to 0.2 m (outlet), with the midpoint at 0 m along the x-axis. The inlet exhibited the lowest temperature, while the maximum was observed in the middle section, gradually decreasing toward the outlet. Detailed results for each channel thickness and nanofluid are presented in Table 8.

Table 8 (a) shows that Al_2O_3 nanofluids have the lowest surface temperature and pressure drop due to their superior thermal conductivity, which improves heat transfer and reduces thermal resistance. This results in significantly lower surface temperatures compared with other nanofluids, as seen in the balanced temperature and pressure drop behavior. The high heat flux, combined with fast-moving Al_2O_3 nanofluids, raises the outlet temperature but allows effective heat dissipation along the channel. The low pressure drop observed suggests that Al_2O_3 nanofluids maintain stable flow dynamics, making them more effective at extracting heat and managing temperature in the MCHS. Other nanofluids, though subjected to the same conditions, show less efficient heat management, maintaining higher surface temperatures and larger thermal gradients, particularly near the channel outlet.

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

Table 8. Graph of surface temperature and pressure drops of different channel thicknesses in different nanofluids.

	Temperature (℃)	Pressure (Pa)
a) 2mm channel thickness		
b) 5mm channel thickness		
c) 8mm channel thickness		

In Table 8 (b), point out the graphs for surface temperature and pressure drop of a 5 mm channel thickness. Al_2O_3 continues to exhibit the lowest surface temperature, indicating better heat transfer due to its high thermal conductivity. ZnO, in contrast, shows the highest surface temperature, reflecting less efficient heat dissipation. Al_2O_3 maintains low temperatures throughout, followed by water, air, and SiO_2, with ZnO performing the least effectively. When examining the pressure drop, air experiences the highest pressure drop, while ZnO has the lowest. The higher pressure drop observed for air flow is attributed to its low viscosity and compressibility, which increase flow resistance within the microchannel, rather than its density. All fluids show a zero pressure drop after exiting the MCHS, as the channel thickness affects fluid velocity, with a larger channel increasing resistance and thus the pressure drop within the MCHS.

Table 8 (c) shows the surface temperature and pressure drop for an 8 mm channel thickness. The temperature for this case indicates that there is no significant difference compared to the 5 mm channel thickness. However, in the 8 mm channel, the pressure drop is higher for all nanofluids except ZnO. The higher pressure drop in the 8 mm channel is likely due to increased friction from the larger surface area that contacts the fluid. Although the parabolic inlet velocity profile reduces wall shear stress, the increased channel size leads to more fluid-wall interaction, increasing resistance. Nanoparticles contacting the walls further increase wall shear stress, slowing the flow and causing higher pressure drops.

4. Conclusion

The copper rectangular MCHS was successfully modeled using COMSOL Multiphysics 5.5, with a detailed analysis of temperature and pressure drop for various nanofluids across channel thicknesses of 2mm, 5mm, and 8mm. Al_2O_3 nanofluid consistently showed the lowest temperature and highest pressure drop relative to other nanofluids, with the 5mm channel emerging as the optimal design due to its balanced performance. Al_2O_3 in the 5mm channel achieved the lowest maximum temperature of 27.39˚C, attributed to its superior thermal conductivity and efficient heat transfer. The 2mm channel, while providing higher surface area for fluid interaction, recorded the highest temperature of 39.52˚C with ZnO nanofluid, demonstrating that narrower channels face difficulties with constrained fluid flow and reduced thermal efficiency.

In contrast, the 8mm channel had the lowest pressure drop, particularly with ZnO, but its larger size did not significantly improve heat transfer, as indicated by slightly higher temperatures. The air in the 5mm channel experienced the highest pressure drop, indicating that fluid dynamics have a strong influence on both heat dissipation and resistance. These findings emphasize the importance of carefully selecting both the nanofluid and channel size to optimize cooling performance. Overall, the 5mm channel showed the best heat transfer performance, offering a balanced approach to both temperature reduction and pressure drop. To further enhance microchannel heat sink efficiency, future improvements could involve optimizing the aspect ratio of Al_2O_3 nanoparticles or testing alternative materials, such as aluminum or iron, for the microchannel. This could lead to even greater cooling performance in electronic applications.

Acknowledgement

This study was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/STG02/UTHM/02/1) and Universiti Tun Hussein Onn Malaysia (UTHM) through vote K173. The authors would also like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

References

Ahmed, H. E., Salman, B. H., Kherbeet, A. S., & Ahmed, M. I. (2018). Optimization of thermal design of heat sinks: A review. *International Journal of Heat and Mass Transfer*, *118*, 129-153.

Alkasmoul, F. S., Asaker, M., Almogbel, A., & AlSuwailem, A. (2022). Combined effect of thermal and hydraulic performance of different nanofluids on their cooling efficiency in microchannel heat sink. *Case Studies in Thermal Engineering*, *30*, 101776.

Al-Rashed, A. A., Shahsavar, A., Entezari, S., Moghimi, M. A., Adio, S. A., & Nguyen, T. K. (2019). Numerical investigation of non-Newtonian water-CMC/CuO nanofluid flow in an offset strip-fin microchannel heat sink: thermal performance and thermodynamic considerations. *Applied Thermal Engineering*, *155*, 247-258.

Alsabery, A. I., Gedik, E., Chamkha, A. J., & Hashim, I. (2019). Effects of two-phase nanofluid model and localized heat source/sink on natural convection in a square cavity with a solid circular cylinder. *Computer Methods in Applied Mechanics and Engineering*, *346*, 952-981.

Behnampour, A., Akbari, O. A., Safaei, M. R., Ghavami, M., Marzban, A., Shabani, G. A. S., & Mashayekhi, R. (2017). Analysis of heat transfer and nanofluid fluid flow in microchannels with trapezoidal, rectangular and triangular shaped ribs. *Physica E: Low-Dimensional Systems and Nanostructures*, *91*, 15-31.

Cao, Y., Abbas, M., El-Shorbagy, M. A., Gepreel, K. A., Dahari, M., Badran, M. F., ... & Wae-hayee, M. (2022). Thermo-hydraulic performance in ceramic-made microchannel heat sinks with an optimum fin geometry. *Case Studies in Thermal Engineering*, *36*, 102230.

Chen, X., & Chen, Y. (2021). A review on modelling, simulation and experiment of thermal conductivity of nanofluids. *International Journal of Environmental Analytical Chemistry*, *101*(10), 1347-1362.

Choi, S. U., & Eastman, J. A. (1995). *Enhancing thermal conductivity of fluids with nanoparticles* (No. ANL/MSD/CP-84938; CONF-951135-29). Argonne National Lab. (ANL), Argonne, IL (United States).

Coşkun, T., & Çetkin, E. (2020). Heat transfer enhancement in a microchannel heat sink: nanofluids and/or micro pin fins. *Heat Transfer Engineering*, *41*(21), 1818-1828.

Dhaiban, H. T., & Hussein, M. A. (2020). The optimal design of heat sinks: A review. *Journal of applied and computational mechanics*, *6*(4), 1030-1043.

Dogonchi, A. S., & Ganji, D. D. (2016). Investigation of MHD nanofluid flow and heat transfer in a stretching/shrinking convergent/divergent channel considering thermal radiation. *Journal of Molecular Liquids*, *220*, 592-603.

Elbadawy, I., Alhajri, A., Doust, M., Almulla, Y., Fayed, M., Dinc, A., ... & Al-Kouz, W. (2023). Reliability of different nanofluids and different micro-channel configurations on the heat transfer augmentation. *Processes*, *11*(3), 652.

Gao, J., Hu, Z., Yang, Q., Liang, X., & Wu, H. (2022). Fluid flow and heat transfer in microchannel heat sinks: Modelling review and recent progress. *Thermal Science and Engineering Progress*, *29*, 101203.

Gunnasegaran, P., Shuaib, N. H., Mohammed, H. A., Jalal, M. A., & Sandhita, E. (2012). Heat transfer enhancement in microchannel heat sink using nanofluids. *Fluid Dynamics, Computational Modeling and Applications*, 287-326.

Hamilton, R. L., & Crosser, O. K. (1962). Thermal conductivity of heterogeneous two-component systems. *Industrial & Engineering chemistry fundamentals*, *1*(3), 187-191.

Iyengar, M., & Bar-Cohen, A. (1998). Optimization of vertical pin-fin heat sinks in natural convective heat transfer. In *International Heat Transfer Conference Digital Library*. Begel House Inc.

Jennings Jr, D., & Smith, S. (2020, October). Comparative Study of Effective Cooling in Microchannel Heat Sinks (MCHS) With Nanofluid and Hydrophobic Nanostructures Using Computational Fluid Dynamics. In *International Electronic Packaging Technical Conference and Exhibition* (Vol. 84041, p. V001T07A002). American Society of Mechanical Engineers.

Li, W., Xie, Z., Xi, K., Xia, S., & Ge, Y. (2021). Constructal optimization of rectangular microchannel heat sink with porous medium for entropy generation minimization. *Entropy*, *23*(11), 1528.

Li, W., Zhu, L., Ji, F., Yu, J., Jin, Y., & Wang, W. (2020, July). Optimization of Manifold Mmicrochannel Heat Sink Based on Equivalent Resistance Model. In *2020 19th IEEE Intersociety*

Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm) (pp. 497-500). IEEE.

Li, X., & Chen, J. (2019). Microembossed copper microchannel heat sink for high-density cooling in electronics. *Micro & Nano Letters*, *14*(12), 1258-1262.

Liou, T. M., Chang, S. W., & Chan, S. P. (2018). Effect of rib orientation on thermal and fluid-flow features in a two-pass parallelogram channel with abrupt entrance. *International Journal of Heat and Mass Transfer*, *116*, 152-165.

Littmarck, H. S., & Saeidi, F. (1986). Comsol. *COMSOL. Available: https://www. comsol. com/.,(accessed Aug. 1, 2022)*.

Malvandi, A., & Ganji, D. D. (2014). Brownian motion and thermophoresis effects on slip flow of alumina/water nanofluid inside a circular microchannel in the presence of a magnetic field. *International Journal of Thermal Sciences*, *84*, 196-206.

Memon, S. A., Cheema, T. A., Kim, G. M., & Park, C. W. (2020). Hydrothermal investigation of a microchannel heat sink using secondary flows in trapezoidal and parallel orientations. *Energies*, *13*(21), 5616.

Mohammed, H. A., Gunnasegaran, P., & Shuaib, N. H. (2010). Heat transfer in rectangular microchannels heat sink using nanofluids. *International Communications in Heat and Mass Transfer*, *37*(10), 1496-1503.

Muneeshwaran, M., Srinivasan, G., Muthukumar, P., & Wang, C. C. (2021). Role of hybrid-nanofluid in heat transfer enhancement–A review. *International Communications in Heat and Mass Transfer*, *125*, 105341.

Nazzal, I. T., Salem, T. K., & Al Doury, R. R. J. (2021, February). Theoretical investigation of a pin fin heat sink performance for electronic cooling using different alloys materials. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1094, No. 1, p. 012087). IOP Publishing.

Pak, B. C., & Cho, Y. I. (1998). Hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. *Experimental Heat Transfer an International Journal*, *11*(2), 151-170.

Patel, A. K., Bhuvad, S., & Rajput, S. P. S. (2019). Effects of nanofluid flow in micro channel heat sink for forced convection cooling of electronics device: a numerical simulation. *Int J Innov Technol Exploring Eng (IJITEE)*, *9*(2), 5230-5243.

Rafiq, M., Shafique, M., Azam, A., & Ateeq, M. (2021). Transformer oil-based nanofluid: The application of nanomaterials on thermal, electrical and physicochemical properties of liquid insulation-A review. *Ain Shams Engineering Journal*, *12*(1), 555-576.

Rajput, R. S. & Sahu, H. (2021). Investigation of different types of heat sink and their applications. *International Journal of Rubber Technology (IJRT), 6*(3), 105-109.

Shamsuddin, H. S., Estelle, P., Navas, J., Mohd-Ghazali, N., & Mohamad, M. (2021). Effects of surfactant and nanofluid on the performance and optimization of a microchannel heat sink. *International Journal of Heat and Mass Transfer*, *175*, 121336.

Sohel, M. R., Khaleduzzaman, S. S., Saidur, R., Hepbasli, A., Sabri, M. F. M., & Mahbubul, I. M. (2014). An experimental investigation of heat transfer enhancement of a minichannel heat sink using Al2O3–H2O nanofluid. *International Journal of Heat and Mass Transfer*, *74*, 164-172.

Song, J., Liu, F., Sui, Y., & Jing, D. (2021). Numerical studies on the hydraulic and thermal performances of trapezoidal microchannel heat sink. *International Journal of Thermal Sciences*, *161*, 106755.

Sur, A., & Gulia, V. (2022). A comprehensive review on microchannel heat exchangers, heat sink, and polymer heat exchangers: Current state of the art. *Frontiers in Heat and Mass Transfer (FHMT)*, *18*.

Tang, K., Lin, G., Guo, Y., Huang, J., Zhang, H., & Miao, J. (2023). Simulation and optimization of thermal performance in diverging/converging manifold microchannel heat sink. *International Journal of Heat and Mass Transfer*, *200*, 123495.

Tuckerman, D. B., & Pease, R. F. W. (1981). High-performance heat sinking for VLSI. *IEEE Electron device letters*, *2*(5), 126-129.

Van Oevelen, T. (2014). Optimal heat sink design for liquid cooling of electronics.

Waqas, H., Khan, S. A., Farooq, U., Muhammad, T., Alshehri, A., & Yasmin, S. (2022). Thermal transport analysis of six circular microchannel heat sink using nanofluid. *Scientific Reports*, *12*(1), 8035.

Zhang, J., Guo, W., Xie, Z., Guan, X., Qu, X., & Ge, Y. (2023). Optimization of fin layout in liquid-cooled microchannels for multi-core chips. *Case Studies in Thermal Engineering*, *41*, 102615.

.

DOI: https://doi.org/10.22452/mjs.vol45sp1.1
Malaysian Journal of Science 45 (Special Issue 1): 1-10 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

https://mjs.um.edu.my

Parameters Improvement of Gaussian Mixture Model for Vehicle Detection in Different Weather Conditions

Nor Afiqah Mohd Aris[1a] and Siti Suhana Jamaian[2a*]

Abstract: Accurate vehicle detection under different weather conditions is crucial to ensure efficient and safe traffic management systems. This paper presents a novel approach for vehicle detection in different weather scenarios, leveraging on Improved Gaussian Mixture Model (Improved GMM). Traditional methods often falter in different weather conditions, posing significant challenges to traffic management. The modified model addresses these challenges by integrating advanced features such as adaptive time-varying learning rates, exponential decay, and outlier processing. This ensures robust performance across different weather conditions, including daytime, nighttime, and rainy weather. Real-world datasets evaluate the model's efficacy, simulating various weather scenarios to assess its adaptability and reliability. Comparative analysis demonstrates the superiority of the Improved GMM over traditional techniques, exhibiting enhanced accuracy and robustness in different weather conditions while maintaining computational efficiency. The findings highlight the potential of the modified model to significantly improve vehicle detection accuracy in different weather conditions, thus contributing to the advancement of reliable traffic management systems irrespective of environmental challenges.

Keywords: Gaussian Mixture Model, Adaptive Time-varying Learning Rate, Exponential Decay, Outlier Processing, Vehicle Detection.

1. Introduction

Accurately detecting vehicles is important for traffic management systems, ensuring smooth traffic flow, safety, and overall road efficiency. In recent years, computer vision techniques have played a pivotal role in vehicle detection and tracking tasks. Researchers have explored various methods for studying vehicle detection, including traditional computer vision, machine learning, deep learning, motion-based approaches, radar-based techniques, and fusion methods. This research concentrates on traditional computer vision techniques, particularly background subtraction, emphasising mathematical contributions. Among these methods, the Gaussian Mixture Model (GMM) proposed by Stauffer and Grimson (1999) has emerged as a popular choice due to its simplicity and effectiveness in background subtraction. However, despite the progress made in the topic, many difficulties remain, making it hard to create robust and adaptable detection systems that can handle the complexities of different environmental conditions, especially those caused by various types of weather.

This study focuses on improving vehicle detection in varying weather conditions—normal day, rainy day, normal night and rainy night. To solve this issue, we propose an Improved Gaussian Mixture Model (Improved GMM) to enhance vehicle detection under these different weather conditions. Traditional GMMs struggle with a constant learning rate that does not adjust to changing data characteristics. This can result in the model either not converging or converging to a suboptimal solution. An Improved GMM addresses this by introducing an adaptive time-varying learning rate, allowing the model to adjust based on current data dynamically, leading to better performance in varying traffic densities.

Traditional GMMs also assume pixels close to the mean belong to the same cluster, which can be a limiting factor. The Improved GMM emphasises these pixels more effectively, improving the identification and separation of objects from the background. Additionally, traditional GMMs treat all observations equally, including outliers, which can distort data distribution estimates. The Improved GMM includes outlier processing to control the influence of new observations on covariance updates.

While previous research have explored various techniques to improve vehicle detection accuracy, there remains a notable gap in the literature regarding comprehensive solutions that address the complexities of different weather conditions. Existing methodologies often fail to provide consistent performance across diverse environmental scenarios, leading to inaccuracies and potential safety hazards. Therefore, this study seeks to bridge this research gap by proposing a novel approach, the Improved Gaussian Mixture Model (Improved GMM), designed explicitly to mitigate the effects of different weather conditions on vehicle detection accuracy. By integrating advanced features such as adaptive time-varying learning rates, exponential decay, and outlier processing, the modified model aims to enhance the adaptability and reliability of vehicle detection systems, thereby

Authors information:
[a]Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology (FAST), Universiti Tun Hussein Onn Malaysia (UTHM), Pagoh Branch Campus, Pagoh Higher Education Hub, Km 1, Jalan Panchor, 84600 Muar, Johor Darul Ta'zim, MALAYSIA. Email: norafiqahmohdaris@gmail.com[1], suhana@uthm.edu.my[2]

*Corresponding Author: suhana@uthm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

contributing to the advancement of traffic management systems and overall road safety.

This paper is organised as follows: Section 2 provides a comprehensive review of the relevant literature. Section 3 presents the methodology, beginning with an overview of the traditional GMM, followed by a discussion on the Improved GMM. Section 4 undertakes an in-depth analysis and discussion of the obtained results. Finally, Section 5 concludes the study by summarising key findings.

2. Literature review

2.1 Introduction to Vehicle Detection

Vehicle detection is a crucial component of modern transportation systems, with applications ranging from traffic management to autonomous driving. Over the years, significant research efforts have been dedicated to advancing vehicle detection methodologies to improve accuracy, reliability, and efficiency. Researchers in computer vision and intelligent transportation systems have shown substantial interest in traffic identification (Buch, Velastin & Orwell, 2011; Zhang et al., 2011). Various technological initiatives are underway to automate the detection and classification of cars (Saran & Sreelekha, 2015; Wang, Xiao & Gu, 2008). Image processing techniques classify vehicles based on shape, size, texture, colour, and other characteristics (Chen, Ellis & Velastin, 2011; Lai & Yung, 2000). The extracted information is then transmitted to a centralised system for regulation and monitoring (Kim & Kim, 2002; Chen et al., 2023). These advancements aim to improve the efficiency and effectiveness of traffic management systems, ensuring smoother transportation operations in response to the challenges posed by increasing vehicle numbers.

The vehicle detection concept involves identifying a vehicle's precise location on the road in real time within a specific area of interest. However, this task becomes highly complex due to the dynamic nature of the traffic scene and the real-time acquisition of images. The constantly changing traffic scene presents significant challenges in accurately detecting vehicles. Various factors further impact the difficulty of vehicle detection, including changes in lighting, weather conditions, shadows, complex background interference, camera parameters, vehicle movements, occlusions, and so on (Pang, Lam & Yung, 2004; Wang, 2010; Karangwa et al., 2023). GMM is commonly employed to generate a foreground mask to address this issue. However, false positive detections are still likely to occur. To overcome this challenge, ongoing research is exploring enhanced methods of vehicle surveillance that aim to reduce false positive detections (Husain, Maity & Yadav, 2020; Premaratne et al., 2023).

For instance, Zivkovic (2004) developed an adaptive GMM with a configurable number of Gaussian components, producing improved model adaptations to changing conditions. Similarly, Zuo et al. (2019) designed an enhanced method for noise interruption for the traditional GMM. In addition to GMM, deep learning techniques have emerged as promising alternatives for vehicle detection tasks (Braham & Droogenbroeck, 2016; Babaee, Dinh & Rigoll, 2018). For example, Yadav and Xiaogang described an innovative hybrid method that combined GMM, background subtraction, Hue-Saturation-Value (HSV) colour model, feature extraction, and neural networks (Yadav & Xiaogang, 2020). Overall, vehicle detection encompasses a diverse range of methodologies and techniques, each offering unique advantages and addressing specific challenges.

2.2 Different Weather Conditions

Weather conditions such as rain, fog, snow, and varying lighting conditions present substantial challenges for vehicle detection systems. Research by Li et al. (2023) demonstrated that heavy rainfall can reduce the accuracy of vehicle detection systems due to decreased visibility and increased noise in sensor data. Similarly, studies by Zang et al. (2019) highlighted the impact of rain, snow, fog, and hail on vehicle detection, showing a significant decrease in detection rates under the worst conditions. Additionally, low-light conditions pose challenges for detection algorithms, as demonstrated by O'Malley, Glavin, and Jones (2011).

In response to the challenges posed by different weather conditions, researchers have explored various strategies to enhance the adaptability of vehicle detection systems. One approach is the integration of weather-specific features into detection algorithms. For example, Al Machot et al. (2019) proposed a method incorporating raindrop detection to improve vehicle detection accuracy during rainy conditions. Another strategy involves the development of robust detection algorithms capable of dynamically adjusting to changing environmental conditions. Research by Rezaei, Terauchi, and Klette (2015) introduced a dynamic thresholding technique that enables vehicle detection systems to adapt to varying lighting conditions, improving detection accuracy under different weather scenarios.

Generally, vehicle detection in different weather conditions poses significant challenges for existing systems. However, research efforts focused on weather-specific adaptations and robust detection algorithms show promise in improving detection accuracy and reliability under different weather conditions.

3. Methodology

3.1 Gaussian Mixture Model in Vehicle Detection

The GMM is a widely used statistical method in computer vision and machine learning for modelling complex data distributions. It provides a flexible framework for representing data that may arise from multiple underlying probability distributions. At its core, the GMM represents the probability distribution of observed data as a weighted sum of several Gaussian distributions, also known as components or clusters. Each Gaussian component in the mixture model represents a cluster of data points in the feature space, with its mean and covariance matrix. The mixture of these components allows GMM to capture complex data patterns that a single Gaussian distribution cannot adequately model.

Many researchers have suggested using multimodal probability density functions to handle backgrounds with moving textures (like a car on a highway). For example, Stauffer and Grimson (1999) described a method where each pixel is modelled

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

using a mix of K Gaussians. This means that the likelihood of colour appearing at a particular pixel is represented as:

$$f(x_t) = \sum_{k=1}^{K} \Pi_{k,t} \cdot \Phi\left(x_t, \mu_{k,t}, \sigma_{k,t}\right) \quad (1)$$

where x_t is the pixel value; $\Phi(x_t, \mu_k, \sigma_{k,t})$ is the Gaussian component density with mean $\mu_{k,t}$ and covariance matrix $\sigma_{k,t}$; $\Pi_{k,t}$ is the weight associated with the k^{th} Gaussian component. Subsequently, $\Phi(x_t, \mu_{k,t}, \sigma_{k,t})$ is formulated as:

$$\Phi(x_t, \mu_{k,t}, \sigma_{k,t}) = 1/((2\pi)^{\frac{n}{2}}|\sigma_{k,t}|^{\frac{1}{2}})\, e^{-\frac{1}{2}(x_t-\mu_{k,t})^T \Sigma_{k,t}^{-1}(x_t-\mu_{k,t})} \quad (2)$$

$$\mu_{k,t} = (1-\beta)\mu_{k,t-1} + \beta x_t \quad (3)$$

$$\sigma_{k,t}^2 = (1-\beta)\sigma_{k,t-1}^2 + \beta(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (4)$$

$$\Pi_{k,t} = (1-\alpha)\Pi_{k,t-1} + \alpha\psi_{k,t} \quad (5)$$

where $\beta = \alpha \cdot \Phi(x_t, \mu_{k,t}, \sigma_{k,t})$, $\psi_{k,t}$ is the indicator function, and α is the constant learning rate. The $\mu_{k,t}$ and $\sigma_{k,t}$ parameters of unmatched distributions remain unchanged. At the same time, their weight is decreased according to the formula in (5) to achieve decay. If no component matches x_t, the one with the lowest weight is substituted with a Gaussian distribution having a mean of x_t, a large initial variance σ_0 and a small weight Π_0. After updating each Gaussian, the K weights Π_0 are normalised that their sum equals 1. Next, the K distributions are arranged based on a fitness value calculated as $\Pi_{k,t}$ divided by $\sigma_{k,t}$, and only the B most reliable distributions are selected to be part of the background.

$$B = \operatorname{argmin}\left(\sum_{k=1}^{b} \Pi_k > th\right) \quad (6)$$

where th is the threshold.

3.2 Improved Gaussian Mixture Model

Improved GMM refined the traditional GMM by adjusting the mean, covariance, and weight calculations, following the methodology proposed by Stauffer and Grimson (1999), leading to more accurate vehicle detection across varying weather conditions. These modifications address the limitations of traditional GMM methods by incorporating an adaptive time-varying learning rate, exponential decay, and outlier processing. As a result, these improvements enable Improved GMM to capture variations in weather accurately and enhance detection accuracy. The enhancements implemented in Improved GMM are based on improvements to the mean, covariance, and weight equations proposed by Stauffer and Grimson (1999). These upgrades include:

$$\mu_{k,t} = (1-\beta(t))\mu_{k,t-1} + \beta(t)x_t \cdot e^{-\lambda d_t} \quad (7)$$

$$\sigma_{k,t}^2 = (1-\beta(t))\sigma_{k,t-1}^2 + \beta(t)(1-\gamma)(1 - e^{-2\lambda d_t})(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (8)$$

$$\Pi_{k,t} = (1-\alpha)\Pi_{k,t-1} + \alpha e^{-\lambda d_t} \quad (9)$$

Incorporating an adaptive time-varying learning rate helps regulate the weights assigned to newly arriving data samples (Fried, Lizarralde & Leite, 2022). This adjustment allows the algorithm to swiftly adjust to changes in traffic flow, leading to more accurate vehicle detection by factoring in the distance between the pixel and the current means. This study replaced the fixed learning rate parameters with an adaptive time-varying learning rate parameter. For this study, the value of $\beta(t)$ is obtained using the Robbins-Monro stochastic approximation method, which requires solving the recursive equation:

$$\beta(t) = c/(c+t) \quad (10)$$

where c is a constant that controls the learning rate. The value of the parameter c is chosen based on prior knowledge of the data, such as possible ranges for model parameters and data distribution. It is important to note that the choice of c significantly impacts the algorithm's performance; selecting an incorrect value can result in slow convergence or instability. The iteration or t value is typically incremented by one with each iteration, representing the number of algorithm iterations or observations analysed both the initial value of t and its growth rate influence convergence speed and algorithm stability. If the initial value of t is too small, the step size may be too large, leading to instability and overestimating the optimal solution. Conversely, if the initial value of t is too high, the step size may be too small, resulting in slow convergence and the risk of getting stuck with a suboptimal solution. The growth rate of t also affects convergence speed and algorithm stability. A rapid increase in t promotes a faster convergence but may lead to instability and overestimation. On the other hand, slower growth in t results in stable behaviour but slower convergence.

The parameter λ adjusts how much each current pixel (mean, covariance, and weight parameters) contributes to the Improved GMM. When this adjustment is made, the precision of vehicle detection is increased by considering the distance between the current pixel and the current mean. Here, $\mu_{k,t-1}$ represents the previous mean, x_t is the current pixel, and d_t is the Euclidean distance between x_t and the previous mean $\mu_{k,t-1}$. Additionally, introducing the exponential decay factor $e^{-\lambda d_t}$ from (7) to (9) ensures that the contribution of each pixel decreases as its distance from the current mean increases (Bishop, 2006). This aligns with the idea that pixels closer to the mean significantly influence parameter changes more than those further away (Ruder, 2016). The additional term $1 - e^{-2\lambda d_t}$ adjusts the contribution of each pixel to the covariance parameter based on

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

its distance from the current mean. Moving from (7) to (9), $\Pi_{k,t-1}$ represents the previous weight and $e^{-\lambda d_t}$ modifies the contribution of each data point to the weight parameter relative to its distance from the current mean. This modification ensures that the contribution of each data point to the weight decreases as its distance from the current mean increases (Aboutanios, 2009).

As a pixel moves farther from the current mean, its contribution to the mean parameter decreases. This helps prevent outliers or noisy data points from significantly affecting the estimated mean (Reynolds & Rose, 1995). Pixels that are distant from the current mean may not accurately represent the underlying distribution, and thus, their contribution to the mean should be reduced. In vehicle detection, a pixel corresponding to a vehicle far from the current mean is considered an outlier or an erroneous detection caused by noise in the sensor data. By reducing the contribution of each pixel to the mean, the algorithm becomes more robust to such outliers and reduces the likelihood of false positives or misclassifications. Therefore, adjusting the contribution of each data point to the mean parameter based on its distance from the current mean improves the accuracy and robustness of the GMM algorithm for vehicle detection in real-time traffic flow analysis.

Outlier processing involves introducing a robustness parameter γ to control how much the model reacts to new pixels, making it more resistant to outliers (Stewart, 1999). This adjustment allows for better control over the influence of new observations on covariance matrix updates (Warwick & Jones, 2005). Inclusion of γ in the covariance update equation of the GMM enhances the model's robustness against outliers and noisy data. This updated equation with outlier processing can be spotted in (8). The traditional GMM treats all observations, including outliers, equally. Consequently, the covariance matrix adapts to outliers and incorporates their influence, potentially distorting the underlying data distribution estimates. In contrast, the Improved GMM, introduced with the parameter γ, allows for controlling the influence of new observations on covariance updates. By adjusting the value of γ, the weight assigned to outlier-like observations is reduced, minimising their impact on covariance estimation. When γ approaches 1, the effect of new observations is diminished, enhancing the model's resilience to outliers. Robust estimators effectively mitigate the impact of outliers or deviations from model assumptions, yielding more reliable parameter estimates (Zoubir et al., 2012). Incorporation of γ into the covariance update equation enables control over the model's responsiveness. This mechanism strikes a balance between incorporating new information and safeguarding against outliers. Adjusting γ reduces the weight given to outlier-like observations, enabling the model to prioritise reliable and representative data points (Tadjudin & Landgrebe, 2000).

Extensive experiments were conducted on real-world datasets to evaluate the effectiveness of the proposed modification. Performance of the traditional GMM formulation was then compared to the modified version, which incorporates the time-varying learning rate, exponential decay, and outlier processing.

3.3 Evaluation Metrics

A range of performance evaluation metrics were employed to assess the effectiveness of the Improved GMM for vehicle detection. These metrics provide quantitative measures of the model's accuracy, reliability, and robustness in detecting vehicles across various weather conditions. In this assessment, four categories are used to assess the accuracy of segmentation algorithms. Firstly, true positives (TP) denote pixels that the segmentation algorithm correctly identifies as positive and are labelled positive in the ground truth annotation. Conversely, false positives (FP) refer to pixels identified as positive by the algorithm but are negative according to the ground truth. False negatives (FN) represent pixels labelled as negative by the algorithm but are positive in the ground truth annotation. Lastly, true negatives (TN) indicate pixels correctly identified as negative by the segmentation algorithm, aligning with the negative label in the ground truth annotation (Zhang et al., 2015; Lin et al., 2023).

Recall (RCL) assesses how well a model identifies actual positive instances by measuring the ratio of true positives to the sum of true positives and false negatives. Precision (PRC) indicates the accuracy of the model's positive predictions by calculating the ratio of true positives to the sum of true and false positives. The F-measure (FMS), or F1 score, is the harmonic mean of precision and recall, offering a balanced view of the model's performance. It is computed as two times the product of precision and recall divided by their sum. The False Positive Rate (FPR) quantifies the proportion of negative instances incorrectly predicted as positive, calculated as the ratio of false positives to the sum of false positives and true negatives. The False Negative Rate (FNR) measures the proportion of positive instances incorrectly identified as negative, calculated as the ratio of false negatives to the sum of false negatives and true positives. Accuracy (ACY) evaluates the overall correctness of the model's predictions, calculated as the ratio of the sum of true positives and true negatives to the total number of instances. The Wrong Classifications Percentage (WCP) represents the proportion of incorrectly classified instances, calculated as the ratio of false positives and false negatives to the total number of instances. These metrics include RCL in (11), PRC in (12), FMS in (13), FPR in (14), FNR in (15), ACY in (16) and WCP in (17) are defined by Goyette et al. (2012).

$$RCL=TP/(FN+TP) \tag{11}$$

$$PRC=TP/(TP+FP) \tag{12}$$

$$FMS=(2\times RCL\times PRC)/(RCL+PRC) \tag{13}$$

$$FPR=FP/(FP+TN) \tag{14}$$

$$FNR=FN/(FN+TP) \tag{15}$$

$$ACY=(TP+TN)/(TP+FP+FN+TN) \tag{16}$$

$$WCP=(FP+FN)/(TP+FP+FN+TN) \tag{17}$$

These evaluation metrics offer comprehensive insights into the model's performance, enabling a thorough assessment of its capabilities in vehicle detection tasks. By examining these metrics, a more comprehensive insight into the strengths and weaknesses of the Improved GMM is obtained, aiding in informed decision-

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

making and identifying areas for potential enhancements in future model iterations.

4. Analysis and discussion

In this section, performance of the Improved GMM was analysed and discussed using a wide range of video sequences for vehicle detection across different weather conditions, including clear and rainy scenarios. The objective is to provide insights into the effectiveness and robustness of the model in addressing the challenges encountered in real-world traffic scenarios. This comprehensive selection enabled us to evaluate how well the proposed methods performed under different weather conditions. The video datasets were sourced from real traffic footage in Kuala Lumpur obtained from Sena Traffic Systems Sdn. Bhd. Our industrial collaborator in Malaysia generously provided these datasets, which also served as valuable case studies. Their contribution closely matched the research problem we aimed to address.

Obtaining ground truth data for the real videos posed challenges, particularly in manually annotating reference images to establish ground truth. Despite these challenges, the model demonstrated promising results, indicating its reliability in accurately detecting vehicles within real traffic flow environments. A significant aspect of the analysis involved the application of a uniform parameter set across all video sequences, chosen based on empirical study. Following Zuo et al. (2019), the parameter values proved effective in various situations, showcasing the model's versatility and robustness in handling different levels of complexity within the traffic flow. The specific parameter values used are listed in Table 1.

Table 1. Parameter Values

K	α	λ	γ	d	c
3	0.9	0.1	0.5	5	100

A comparison of traditional GMM and Improved GMM performance is conducted using these quantitative parameters, as outlined in Table 2. The optimal values are highlighted in bold.

The Improved GMM outperformed the traditional GMM significantly, achieving a recall rate of 0.8915 compared to 0.3983 for the traditional GMM. This indicates that the Improved GMM correctly identified a higher proportion of actual positive instances, showcasing its superior ability to detect vehicles in the given scenarios. PRC measures the accuracy of positive predictions made by the model. The Improved GMM achieved a slightly higher precision rate than the traditional GMM (0.6921 vs. 0.6744). This implies that the Improved GMM exhibited a better ability to minimise false positive predictions while maximising true positive identifications, thus demonstrating its enhanced accuracy in vehicle detection tasks. Meanwhile, FMS, which provides a balanced measure of the model's precision and recall, showed a notable improvement for the Improved GMM (0.7572) compared to the traditional GMM (0.4815). This indicates that the Improved GMM achieved a better balance between precision and recall, leading to more reliable and consistent performance across different evaluation scenarios.

Analysing the FPR and FNR, it is evident that the Improved GMM exhibited superior performance in both metrics, with lower values indicating fewer misclassification instances than the traditional GMM. This suggests that the Improved GMM effectively minimises false positives and negatives, thereby enhancing the overall accuracy of vehicle detection. ACY serves as a measure of the overall correctness of the model's predictions across all classes. Here, the Improved GMM achieved a significantly higher accuracy rate (0.9676) compared to the traditional GMM (0.8661), highlighting its ability to make more precise and reliable predictions across diverse scenarios. Lastly, WCP provides insights into the proportion of instances incorrectly classified by the model. The Improved GMM demonstrated a substantially lower WCP (0.0324) compared to the traditional GMM (0.1339), indicating its effectiveness in reducing misclassifications and improving overall model robustness. Overall, the comparative analysis of performance metrics clearly demonstrates superiority of the Improved GMM over the traditional GMM in vehicle detection tasks, highlighting its potential for enhancing the accuracy and reliability of traffic management systems in various real-world scenarios.

Figure 1 illustrates the F-measure comparison between the traditional GMM and the Improved GMM across different weather conditions. For this measure, the Improved GMM consistently outperformed the traditional GMM across all weather conditions. Specifically, during normal day and night scenarios, the Improved GMM achieved notably higher F-

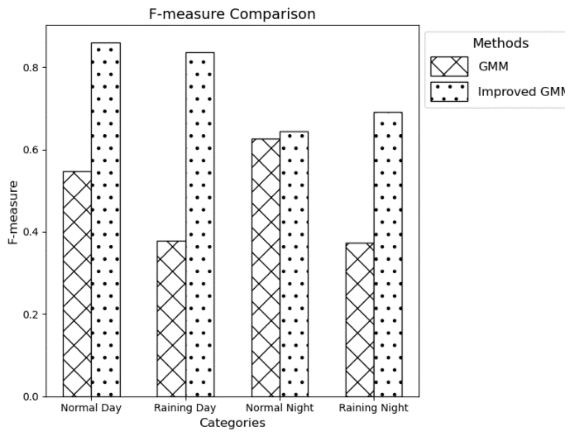


Figure 1. Comparison of F-measure Result of GMM and Improved GMM

Table 2. Summary of Evaluation Metrics of GMM and Improved GMM

Method	RCL	PRC	FMS	FPR	FNR	ACY	WCP
GMM (Stauffer & Grimson, 1999)	0.3983	0.6744	0.4815	0.0303	0.6017	0.8661	0.1339
Improved GMM	**0.8915**	**0.6921**	**0.7572**	**0.0239**	**0.1084**	**0.9676**	**0.0324**

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

measure scores compared to the traditional GMM method. This indicates that the Improved GMM model performs better in accurately identifying vehicles under typical lighting and weather conditions. Moreover, even in challenging weather conditions such as rainy days and nights where visibility is significantly reduced, the Improved GMM still demonstrated remarkable performance, with considerably higher F-measure scores compared to the traditional GMM approach. This suggests that the Improved GMM model is more robust and reliable in detecting vehicles for different weather conditions, highlighting its effectiveness in real-world applications. By leveraging robust statistical techniques and adaptive learning mechanisms, the Improved GMM proves to be a promising solution for improving the reliability of vehicle detection systems in real-world applications.

Figure 2 compares the AC results obtained from two vehicle detection methods across different weather conditions. The data illustrates remarkably high accuracy levels for both methods, with most accuracy scores surpassing 0.8. However, upon closer examination, notable contrasts in their performances become apparent. The Improved GMM consistently exhibits superior

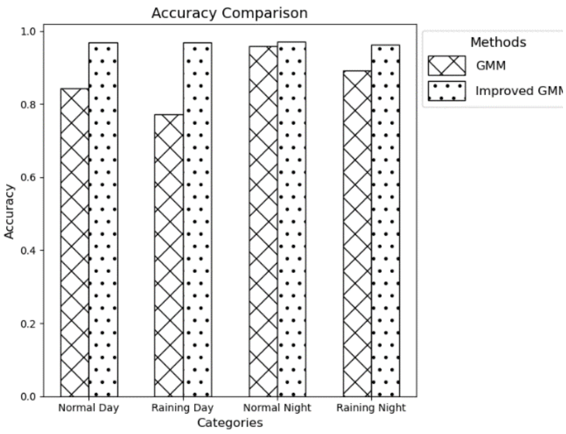


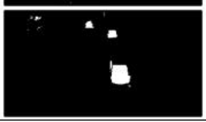
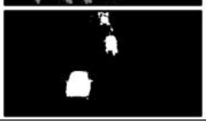
Figure 2. Comparison of Accuracy Result of GMM and Improved GMM

accuracy compared to the traditional GMM across all weather conditions. Notably, it achieves higher accuracy levels in normal weather conditions and challenging scenarios such as rainy days and nights. This indicates the robustness of the Improved GMM model to different weather conditions, emphasizing its reliability in real-world applications where accurate vehicle detection is essential. Interestingly, both methods demonstrate higher accuracy during normal nighttime and daytime conditions. This observation could be attributed to improved visibility and lighting conditions during nighttime, facilitating more accurate vehicle detection outcomes. Overall, the Improved GMM is the preferred choice due to its consistently higher accuracy levels across different weather conditions. Its superior performance highlights its potential value in critical applications such as autonomous driving systems and surveillance platforms. These findings offer valuable insights for researchers seeking to enhance the accuracy and reliability of vehicle detection systems in diverse environmental settings.

Table 3 compares vehicle detection based on real dataset, offering valuable insights into how well the traditional GMM and the Improved GMM methods detect vehicles across different weather conditions. The table displays original video frames (input) in the first column and manually created GT masks indicating vehicle locations in the second column. The third and fourth columns show the vehicle detection generated by the traditional GMM and Improved GMM methods, respectively. The aim is to evaluate how accurately the detection represents vehicle positions and shapes in the video frames. The results from Table 3 indicate that the Improved GMM produces satisfactory detection outcomes, particularly noticeable in normal daytime conditions. It captures vehicle locations and shapes accurately, showcasing its reliability in typical daytime traffic scenarios. This suggests that the Improved GMM method accurately identifies and outlines vehicles in clear, well-lit conditions, which are common during daytime traffic.

On the other hand, the traditional GMM method demonstrates comparable performance, especially in specific scenarios. While the Improved GMM method excels in normal

Table 3. Comparison of Vehicle Detection for Real Dataset

Weather Condition	Normal Day	Raining Day	Normal Night	Raining Night
Input				
Ground Truth				
GMM				
Improved GMM				

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

daytime conditions, the traditional GMM method may perform well in other situations, such as normal nighttime conditions. This suggests that the traditional GMM method may have strengths in more straightforward, well-lit scenarios, highlighting its effectiveness in specific lighting and environmental conditions. However, it is essential to acknowledge that both methods exhibit false positives and negatives to some extent, indicating that neither technique is flawless. The study emphasises the need to develop an algorithm that balances computational efficiency and performance, making it suitable for real-time applications with limited processing time. Furthermore, the increased complexity introduced by the Improved GMM algorithm requires further evaluation to understand its implications completely, particularly in scenarios with different weather conditions.

Previous research in vehicle detection has provided valuable insights into the challenges and advancements in this area. Our study builds upon and contributes to this existing body of knowledge by introducing Improved GMM as a novel approach to address the limitations of traditional vehicle detection methods, particularly in different weather conditions. While previous studies have explored various methodologies for vehicle detection, our approach differs significantly in its focus on robustness and adaptability across different weather conditions. Unlike previous methods that may specialise in specific weather scenarios or rely solely on image processing techniques, the Improved GMM integrates advanced features such as adaptive time-varying learning rates, exponential decay and outlier processing to enhance performance in challenging environments.

Methodological differences between our study and previous research may influence the interpretation of results. For example, while some studies may have utilised synthetic datasets or controlled environments for evaluation, our study employed real-world traffic footage obtained from Kuala Lumpur, providing a more realistic assessment of model performance in practical scenarios. Additionally, variations in parameter settings, dataset characteristics, and evaluation metrics may contribute to differences in reported outcomes across studies. Our findings may diverge from those of previous studies in certain aspects. For instance, while some studies may have reported higher accuracy rates under specific weather conditions using alternative methodologies, our results demonstrate the superior performance of the Improved GMM across a range of environmental scenarios. These discrepancies may stem from differences in model complexity, dataset composition, or evaluation criteria, highlighting the importance of context-specific interpretations.

While our study benefits from real-world data and a comprehensive evaluation framework, it may be limited by dataset availability, annotation accuracy, and computational resources. Conversely, our approach offers advantages in adaptability and robustness, addressing critical gaps identified in previous methodologies. Future research in vehicle detection could further explore the implications of our findings and address the remaining challenges. Areas for future investigation may include refining the Improved GMM algorithm, validating its performance in diverse geographical locations, and assessing its applicability in emerging technologies such as autonomous vehicles and intelligent transportation systems. Additionally, comparative studies with alternative approaches could provide valuable insights into different detection methods' relative strengths and weaknesses. The practical implications of our study extend to various domains, including transportation management, urban planning, and public safety. By improving the accuracy and reliability of vehicle detection systems, particularly in adverse weather conditions, our approach can enhance the efficiency of traffic management operations, reduce congestion, and mitigate risks associated with adverse weather events. Policymakers and practitioners may benefit from integrating robust detection technologies into existing infrastructure and decision-making processes to improve transportation resilience and sustainability.

5. Conclusion

In addition to the comprehensive analysis and evaluation of vehicle detection methods under different weather conditions, this study introduces a novel approach, Improved GMM, to address the inherent challenges in traditional methodologies. Unlike previous approaches that may focus on specific weather scenarios or rely solely on image processing techniques, the Improved GMM leverages advanced features such as adaptive time-varying learning rates and outlier processing to enhance adaptability and reliability across diverse environmental settings. Through extensive experimentation and evaluation using real-world datasets, our findings demonstrate the superiority of the Improved GMM over traditional methods, particularly in accurately detecting vehicles under challenging weather conditions. The Improved GMM exhibits remarkable adaptability and robustness, showcasing its potential to significantly improve the accuracy and reliability of traffic management systems in real-world scenarios. Furthermore, this study provides valuable insights into the performance of both traditional and novel methods, highlighting the importance of balancing computational efficiency and performance in real-time applications. While the traditional GMM method may excel in specific scenarios, the Improved GMM emerges as the preferred choice for its consistent and superior performance across various environmental settings.

In conclusion, this study contributes to the advancement of vehicle detection systems by introducing a novel approach that addresses the limitations of traditional methodologies. The findings offer valuable implications for researchers aiming to improve the accuracy and reliability of traffic management systems in real-world scenarios. Future research directions may involve further exploration of the Improved GMM algorithm's implications, particularly in scenarios with varying weather conditions and lighting, to enhance its effectiveness and applicability in critical applications such as autonomous driving systems and surveillance platforms.

6. Acknowledgement

This research was supported by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2021/STG06/UTHM/02/1).

7. References

Aboutanios, E. (2009). Estimation of the frequency and decay factor of a decaying exponential in noise. IEEE Transactions on Signal Processing, 58(2), 501-509.

Al Machot, F., Ali, M., Haj Mosa, A., Schwarzlmüller, C., Gutmann, M., & Kyamakya, K. (2019). Real-time raindrop detection based on cellular neural networks for ADAS. Journal of Real-Time Image Processing, 16, 931-943.

Babaee, M., Dinh, D. T., & Rigoll, G. (2018). A deep convolutional neural network for video sequence background subtraction. Pattern Recognition, 76, 635-649.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer google scholar, 2, 1122-1128.

Braham, M., & Van Droogenbroeck, M. (2016). Deep background subtraction with scene-specific convolutional neural networks. In 2016 international conference on systems, signals and image processing (IWSSIP), 1-4.

Buch, N., Velastin, S. A., & Orwell, J. (2011). A review of computer vision techniques for the analysis of urban traffic. IEEE Transactions on Intelligent Transportation Systems, 12(3), 920-939.

Chen, C., Wang, C., Liu, B., He, C., Cong, L., & Wan, S. (2023). Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 24(11), 13023-13034.

Chen, Z., Ellis, T., & Velastin, S. A. (2011). Vehicle type categorisation: A comparison of classification schemes. In 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), 74-79.

Fried, J., Lizarralde, F., & Leite, A. C. (2022). Adaptive image-based visual servoing with time-varying learning rates for uncertain robot manipulators. In 2022 American Control Conference (ACC), 3838-3843.

Goyette, N., Jodoin, P. M., Porikli, F., Konrad, J., & Ishwar, P. (2012). Changedetection. net: A new change detection benchmark dataset. In 2012 IEEE computer society conference on computer vision and pattern recognition workshops, 1-8.

Husain, A. A., Maity, T., & Yadav, R. K. (2020). Vehicle detection in intelligent transport system under a hazy environment: a survey. IET Image Processing, 14(1), 1-10.

Karangwa, J., Liu, J., & Zeng, Z. (2023). Vehicle detection for autonomous driving: A review of algorithms and datasets. IEEE Transactions on Intelligent Transportation Systems.

Kim, J. B., & Kim, H. J. (2003). Efficient region-based motion segmentation for a video monitoring system. Pattern recognition letters, 24(1-3), 113-128.

Lai, A. H., & Yung, N. H. C. (2000). Vehicle-type identification through automated virtual loop assignment and block-based direction-biased motion estimation. IEEE Transactions on Intelligent Transportation Systems, 1(2), 86-97.

Li, H., Bamminger, N., Magosi, Z. F., Feichtinger, C., Zhao, Y., Mihalj, T., ... & Eichberger, A. (2023). The effect of rainfall and illumination on automotive sensors detection performance. Sustainability, 15(9), 7260.

Lin, T. C., Chen, H. H., Chen, K. S., Chen, Y. P., & Chang, S. H. (2023). Decision-Making Model of Performance Evaluation Matrix Based on Upper Confidence Limits. Mathematics, 11(16), 3499.

O'malley, R., Glavin, M., & Jones, E. (2011). Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions. IET Intelligent Transport Systems, 5(1), 1-10.

Pang, C. C. C., Lam, W. W. L., & Yung, N. H. C. (2004). A novel method for resolving vehicle occlusion in a monocular traffic-image sequence. IEEE Transactions on Intelligent Transportation Systems, 5(3), 129-141.

Premaratne, P., Kadhim, I. J., Blacklidge, R., & Lee, M. (2023). Comprehensive review on vehicle detection, classification and counting on highways. Neurocomputing, 126627.

Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE transactions on speech and audio processing, 3(1), 72-83.

Rezaei, M., Terauchi, M., & Klette, R. (2015). Robust vehicle detection and distance estimation under challenging lighting conditions. IEEE transactions on intelligent transportation systems, 16(5), 2723-2743.

Ruder, S. (2016). An overview of gradient descent optimisation algorithms. arXiv preprint arXiv:1609.04747.

Saran, K. B., & Sreelekha, G. (2015). Traffic video surveillance: vehicle detection and classification. In 2015 International Conference on Control Communication & Computing India (ICCC), 516-521.

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149), 2, 246-25.

Stewart, C. V. (1999). Robust parameter estimation in computer vision. SIAM review, 41(3), 513-537.

Tadjudin, S., & Landgrebe, D. A. (2000). Robust parameter estimation for mixture model. IEEE Transactions on Geoscience and Remote Sensing, 38(1), 439-445.

Wang, G., Xiao, D., & Gu, J. (2008). Review on vehicle detection based on video for traffic surveillance. In 2008 IEEE international conference on automation and logistics, 2961-2966.

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

Wang, Y. (2010). Joint random field model for all-weather moving vehicle detection. IEEE Transactions on Image Processing, 19(9), 2491-2501.

Yadav, B. K., & Xiaogang, J. (2020). An algorithm for intelligent identification of moving objects in natural environment. International Journal of Machine Learning and Computing, 10(5).

Zang, S., Ding, M., Smith, D., Tyler, P., Rakotoarivelo, T., & Kaafar, M. A. (2019). The impact of adverse weather conditions on autonomous vehicles: how rain, snow, fog, and hail affect the performance of a self-driving car. IEEE vehicular technology magazine, 14(2), 103-111.

Zhang, J., Wang, F. Y., Wang, K., Lin, W. H., Xu, X., & Chen, C. (2011). Data-driven intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems, 12(4), 1624-1639.

Zhang, Y., Zhao, C., He, J., & Chen, A. (2015). Vehicles detection in complex urban traffic scenes using a nonparametric approach with confidence measurement. In 2015 International Conference and Workshop on Computing and Communication (IEMCON), 1-7.

Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, 28-31.

Zoubir, A. M., Koivunen, V., Chakhchoukh, Y., & Muma, M. (2012). Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. IEEE Signal Processing Magazine, 29(4), 61-80.

Zuo, J., Jia, Z., Yang, J., & Kasabov, N. (2019). Moving target detection based on improved Gaussian mixture background subtraction in video images. IEEE Access, 7, 152612-152623.

DOI: https://doi.org/10.22452/mjs.vol45sp1.2
Malaysian Journal of Science 45 (Special Issue 1): 11-19 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024 Special issue

Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026) https://mjs.um.edu.my

Impact of Climate Factors on Hand, Foot, and Mouth Disease in Malaysia: A Generalised Linear Models Approach

Yan Mun Lau[1a*], Wooi Chen Khoo[2b], and Yeh-Ching Low[3c]

Abstract: Hand, Foot, and Mouth Disease (HFMD) is a re-emerging infectious disease in the Asia-Pacific region, particularly in Malaysia. To the best of our knowledge, studies focusing on analyzing the relationship between climate factors and HFMD in Malaysia have not used 10 years of climate data or conducted research covering all states in the country. These existing studies have suggested that examining more data would produce more precise results. Therefore, this study employs generalized linear models (GLMs) to examine the relationship between climate factors and HFMD incidences and to forecast the HFMD cases. Data from 2009 to 2019 were analyzed using five different distributions across fourteen Malaysian states, considering daily HFMD cases, temperature, relative humidity, and rainfall. The Zero-inflated Negative Binomial model was identified as the most appropriate for analyzing and predicting HFMD cases in multiple states, providing valuable insights for policymakers and health authorities.

Keywords: Hand Foot Mouth Disease (HFMD), Zero-inflated Models, Generalised Linear Models, Climate Factors Analysis, Model Fitting and Prediction.

1. Introduction

Hand, Foot, and Mouth Disease (HFMD) is a significant public health issue in the Asia-Pacific region, primarily affecting children under five years of age. The disease is caused by enteroviruses, predominantly Coxsackievirus A16 (CVA16) and Enterovirus A71 (EV-A71), and is characterised by symptoms including painful mouth sores and a rash with blisters on the hands, feet, and buttocks (Chang et al., 2002). While the disease is generally mild, severe cases may result in complications like meningitis, encephalitis, and polio-like paralysis, which can be fatal (Puenpa et al., 2019). HFMD is transmitted directly with contaminated bodily secretions, including blister fluid and faeces, or via contaminated objects (Sun et al., 2016).

The incidence of HFMD exhibits distinct seasonal patterns, indicating that climate factors such as temperature, relative humidity, and rainfall play an important role in the spread of the disease (Luo et al., 2022; Wu et al., 2022; Kim et al., 2016; Onozuka & Hashizume, 2011). Temperature affects viral survival and transmission rates, rainfall contributes to environmental contamination and increased human contact, and relative humidity influences virus stability and vulnerability. This study focuses specifically on these three climate factors because they have consistently shown the strongest associations with HFMD outbreaks in previous research. Other climate variables, such as wind speed and air pressure, have demonstrated inconsistent or weaker relationships with HFMD cases and are thus not included in this analysis.

Motivated by the prevalence of HFMD in Malaysia in recent years, this study aims to evaluate the fit of generalized linear models (GLMs) for HFMD incidence data, identify the most suitable model, and use it to predict HFMD cases in Malaysia based on climate factors. Additionally, the study employs daily data from a span of ten years to examine which climate factors significantly influence HFMD by conducting factor analysis and evaluating the statistical significance of the variables through p-values. By identifying the best model for HFMD prediction, the findings will offer valuable insights for early warning systems and targeted public health interventions to prevent HFMD outbreaks effectively.

2. Literature review

Several studies have examined the relationship between climatic factors and HFMD incidences. Onozuka and Hashizume (2011) investigated the impact of temperature on HFMD incidence in Japan, finding that the disease exhibited a seasonal pattern with peaks during summer in temperate regions, and both spring and autumn in subtropical regions. The study indicated that within a high temperature range, the risk of HFMD increased, but extremely high temperatures led to a decline in cases, suggesting an optimal temperature range for the survival and transmission of the virus.

In Singapore, studies conducted during epidemic years from 2005 to 2007 revealed two distinct peaks in HFMD incidence. The dominant circulating strain of the virus during these outbreaks varied, with CVA16 being prevalent in 2005 and 2007, while EV-

Authors information:
[a]School of Mathematical Sciences, Sunway University, Selangor, MALAYSIA. E-mail: lau.yanmun@gmail.com[1]
[b]Institute of Actuarial Science and Data Analytics, UCSI University, MALAYSIA. E-mail: w_c_khoo@yahoo.com[2]
[c]School of Computing and Artificial Intelligence, Sunway University. E-mail: yehchingl@sunway.edu.my[3]

*Corresponding Author: lau.yanmun@gmail.com

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

A71 was prevalent in 2006 (Ang et al., 2009). The Ministry of Health in Singapore reported significant variations in the number of HFMD clusters affecting preschools and childcare centres from 2018 to 2022, further underscoring the influence of climate on HFMD spread.

Thailand experienced its first major HFMD outbreak in 2012, with over 45,000 reported cases, followed by another notable outbreak four years later. These events highlighted HFMD as a growing public health concern in Thailand (Verma et al., 2021). Similarly, in India, the first documented HFMD outbreak occurred in Kerala in 2005, attributed to EV-A71, with subsequent reports of over 1,150 cases from 2012 to 2017 (Chavan et al., 2023).

In Malaysia, HFMD incidences peak during the Southeast Monsoon season (May to August), which is typically dry, and decline during the Northeast Monsoon season (November to February), associated with heavy rainfall (Abdul Wahid et al., 2021). This seasonal variation highlights the potential influence of climate factors on HFMD transmission.

Previous studies in Malaysia have provided valuable insights into HFMD patterns and related risk factors. A study in Kota Kinabalu analysed HFMD occurrences and demographic risk factors, using logistic regression to estimate the odds ratios for severe cases. The findings indicated significant associations between severe HFMD and factors such as age and ethnicity (Chin et al., 2022). In Sarawak, researchers applied an autoregressive moving average (ARMA) model to seven years of HFMD data, achieving a 94% confidence interval for predicting HFMD cases (Sham et al., 2014). Another study in Kelantan used multiple logistic regression to examine the relationship between risk factors and HFMD incidences, identifying factors like preschool attendance and home care as significant (Wan Mohamad et al., 2021). In Selangor, a study from 2010 to 2016 performed correlation analysis and GLM to estimate the relationship between HFMD occurrences and temperature, considering a lag-time effect (Abdul Wahid et al., 2020). The research concluded that temperature could serve as an early warning indicator for HFMD outbreaks, with a 2-week lag time being significant (Wahid et al., 2021).

While previous studies have examined the relationship between climate factors and HFMD, there are key gaps that this study aims to address. First, this study analyses ten years of HFMD data. In contrast, most existing studies have only covered an average of five years, as suggested by Abdul Wahid et al. (2020), who emphasized the importance of using longer time spans for more reliable conclusions. Additionally, this study uses daily HFMD data, following the recommendation of Chen et al. (2019) that analysing daily data can yield more precise findings compared to weekly or monthly data. Furthermore, Wahid et al. (2021) highlighted the need to explore the impact of climate variables on HFMD at the state level in Malaysia, which this study addresses by assessing regional variations in climate effects.

Despite these studies, there remains a gap in understanding the impact of climate factors across all Malaysian states using statistical models like Zero-inflated models. This study aims to bridge this gap by evaluating the fit of different GLMs and identifying the most appropriate model for predicting HFMD cases based on climate factors.

3. Methodology

3.1 Data Sources

HFMD data from 2009 to 2019 were obtained from the Institute of Medical Research, Malaysia, covering daily reported cases across fourteen states. Climate data, including temperature, rainfall, and relative humidity, were acquired from the Malaysia Meteorological Department.

3.2 Generalised Linear Models

To model Hand, Foot, and Mouth Disease (HFMD) cases, several count models are considered, including the Poisson model, Conway-Maxwell Poisson (COM-Poisson) model, Zero-inflated Poisson (ZIP) model, Zero-inflated Geometric (ZIGeo) model, and Zero-inflated Negative Binomial (ZINB) model. These models address key characteristics observed in HFMD data, such as overdispersion and excess zero counts.

The Poisson model serves as a baseline and assumes that HFMD cases follow a Poisson distribution. The COM-Poisson model extends this by allowing for variable dispersion levels, making it useful for under-dispersed and over-dispersed data. However, since daily HFMD data often contain an excess number of zeros, zero-inflated models are applied. Each of these models is mathematically formulated as shown in Table 1.

3.3 Statistical Analysis

The analysis was conducted using R version 4.2.2 with the 'pscl' package. Five models were evaluated: Poisson, Conway-Maxwell Poisson (COM-Poisson), Zero-inflated Poisson (ZIP), Zero-inflated Geometric (ZIGeo), and Zero-inflated Negative Binomial (ZINB). The data were divided 80:20 for model training and testing. Model selection was based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Accuracy was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percent Error (MAPE).

4. Results

4.1 Model Fitting Performance

The model fitting performance for the five evaluated models is presented in Table 2. The Zero-inflated Negative Binomial (ZINB) model demonstrated the lowest AIC and BIC values across most states, indicating it as the most appropriate fit for analysing HFMD data. The results for selected states are summarized as in Table 2.

The model fitting results indicate that the Zero-inflated Negative Binomial (ZINB) model consistently outperformed other models across several states. In Sabah, the ZINB model had the lowest AIC (17,273.85) and BIC (17,328.52), indicating the best fit. Similarly, in Sarawak, it achieved the lowest AIC (26,647.21) and BIC (26,701.88), confirming its superior performance. In Selangor, both the Zero-inflated Geometric (ZIGeo) and ZINB models

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 1. Statistical Models for HFMD Data.

Models	**Mathematical Expression**
Poisson	$P(Y=y)=\frac{e^{-\lambda}\lambda^y}{y!}$ $\log(\lambda)=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k$
Conway-Maxwell (COM) Poisson (Conway & Maxwell, 1962)	$P(Y=y)=\frac{\lambda^y}{(y!)^v Z(\lambda,v)}$ $\lambda=E(y^v)>0$ $Z(\lambda,v)=\sum_{j=0}^{\infty}\frac{\lambda_j}{(j!)^v}$ $v\geq 0$ denotes dispersion parameter $v=1$ denotes equi-dispersion
Zero-inflated Poisson (Greene, 1994)	$P(Y=y)=\begin{cases}\pi+(1-\pi)\mathrm{e}^{-\lambda}, if\ y=0\\(1-\pi)\frac{\lambda^y\mathrm{e}^{(-\lambda)}}{y!}, if\ y>0\end{cases}$ $\lambda=\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k)$
Zero-inflated Geometric (Nanjundan & Pasha, 2015)	$P(Y=y)=\begin{cases}\pi+(1-\pi)p, if\ y=0\\(1-\pi)pq^y, if\ y>0\end{cases}$ $\lambda=\exp\left(\beta_0+\beta_1x_{m+1}+\cdots+\beta_{k-m}x_k\right)$
Zero-inflated Negative Binomial (Greene, 1994)	$P(Y=y)=\begin{cases}\pi+(1-\pi)\left(\frac{\mathrm{r}}{r+\lambda}\right)^r, if\ y=0\\(1-\pi)\left(\frac{\mathrm{r}}{r+\lambda}\right)^r\frac{\delta(r+y)}{y!\,\delta(r)}\left(\frac{\lambda}{r+\lambda}\right)^y, if\ y>0\end{cases}$ $\lambda=\exp\left(\beta_0+\beta_1x_{m+1}+\cdots+\beta_{k-m}x_k\right)$ $r=positive\ constant$

showed comparable performance, with the ZINB model marginally better in terms of AIC and BIC.

Wilayah Persekutuan also had the best fit with the ZINB model, with an AIC of 19,335.21 and BIC of 19,389.88. A similar trend was observed in Perak and Perlis, where the ZINB model showed the lowest AIC and BIC values, indicating its robustness across different states.

4.2 Factor Analysis p-values

Table 3 presents the p-values from the factor analysis for temperature (β_1), rainfall (β_2), and relative humidity (β_3) using the Zero-inflated Negative Binomial model. Only the count model coefficients (negbin with log link) are included, as they directly estimate the impact of climate factors on the expected number of HFMD cases. The zero-inflation model coefficients (binomial with logit link), which predict the probability of excess zero cases, are excluded since the primary focus of this study is on identifying significant climate factors that influence HFMD incidence rather than explaining the existence of true zero cases in the data. The significant climate factors ($p < 0.05$) influencing HFMD cases across various states are summarised below:

The results from Table 3 indicate that temperature significantly affects HFMD cases across different states in Malaysia. In most states, higher temperatures are associated with an increase in HFMD cases, with Johor (β1 = 0.2672 ***) and Melaka (β1 = 0.3136 ***) showing the strongest positive associations. This suggests that warmer conditions may create a more favourable environment for HFMD transmission.

An exception is observed in Sarawak, where temperature has a negative effect (β1 = -0.2140), implying that cooler weather may contribute to increased HFMD cases in this state. However, the p-values for all climate variables, including temperature, are unavailable (NA) in the output, indicating potential issues with model estimation. This could be due to collinearity among variables, insufficient data variability, or computational singularity, as suggested by the warning message.

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 2. Model Fitting Performance.

State	Model	AIC	BIC
Sabah	Poisson	39002.28	39026.58
	COM-Poisson	21983.75	22014.12
	Zero-inflated Poisson	30847.28	30895.88
	Zero-inflated Geometric	17495.82	17544.42
	Zero-inflated Negative Binomial	**17273.85**	**17328.52**
Sarawak	Poisson	75514.18	75538.47
	COM-Poisson	35637.58	35667.95
	Zero-inflated Poisson	72396.65	72445.24
	Zero-inflated Geometric	26665.26	26713.85
	Zero-inflated Negative Binomial	**26647.21**	**26701.88**
Selangor	Poisson	82233.46	82257.75
	COM-Poisson	33095.37	33125.74
	Zero-inflated Poisson	78530.65	78579.24
	Zero-inflated Geometric	26115.44	**26164.04**
	Zero-inflated Negative Binomial	**26111.79**	26166.46
Wilayah Persekutuan	Poisson	36931.83	36956.13
	COM-Poisson	21229.43	21259.80
	Zero-inflated Poisson	32587.58	32636.17
	Zero-inflated Geometric	19353.25	19401.85
	Zero-inflated Negative Binomial	**19335.21**	**19389.88**
Kelantan	Poisson	22823.43	22847.72
	COM-Poisson	14389.53	14419.90
	Zero-inflated Poisson	18486.72	18535.32
	Zero-inflated Geometric	13332.44	13381.04
	Zero-inflated Negative Binomial	**13122.71**	**13177.38**
Pahang	Poisson	13862.97	13887.27
	COM-Poisson	10874.74	10905.12
	Zero-inflated Poisson	11839.62	11888.21
	Zero-inflated Geometric	10405.88	10454.48
	Zero-inflated Negative Binomial	**10345.73**	**10400.40**
Terengganu	Poisson	10837.39	10861.69
	COM-Poisson	9096.156	9126.528
	Zero-inflated Poisson	9663.740	9713.335
	Zero-inflated Geometric	8922.715	8971.309
	Zero-inflated Negative Binomial	**8883.816**	**8938.485**
Kedah	Poisson	21062.40	21086.70
	COM-Poisson	15205.62	15235.99
	Zero-inflated Poisson	18919.42	18968.01
	Zero-inflated Geometric	14872.84	**14921.43**
	Zero-inflated Negative Binomial	**14870.34**	14925.00
Perak	Poisson	31047.72	31072.02
	COM-Poisson	18583.35	18613.73
	Zero-inflated Poisson	27160.56	27209.15
	Zero-inflated Geometric	17427.22	17475.81
	Zero-inflated Negative Binomial	**17369.45**	**17424.11**

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 2. Model Fitting Performance (continue).

State	Model	AIC	BIC
Perlis	Poisson	8358.494	8382.791
	COM-Poisson	7476.174	7506.546
	Zero-inflated Poisson	7728.705	7777.300
	Zero-inflated Geometric	7443.517	7492.112
	Zero-inflated Negative Binomial	**7425.571**	**7480.240**
Penang	Poisson	28363.47	28387.76
	COM-Poisson	19370.06	19400.43
	Zero-inflated Poisson	25965.96	26014.55
	Zero-inflated Geometric	**17667.38**	**17715.97**
	Zero-inflated Negative Binomial	17667.60	17722.27
Johor	Poisson	31317.07	31341.37
	COM-Poisson	20394.95	20425.32
	Zero-inflated Poisson	28221.37	28269.97
	Zero-inflated Geometric	18454.17	**18502.76**
	Zero-inflated Negative Binomial	**18448.59**	18503.26
Melaka	Poisson	20887.61	20911.90
	COM-Poisson	13986.04	14016.41
	Zero-inflated Poisson	17514.96	17563.56
	Zero-inflated Geometric	13602.45	13651.04
	Zero-inflated Negative Binomial	**13523.83**	**13578.50**
Negeri Sembilan	Poisson	20424.16	20448.46
	COM-Poisson	14133.61	14163.98
	Zero-inflated Poisson	17329.83	17378.43
	Zero-inflated Geometric	13590.85	13639.45
	Zero-inflated Negative Binomial	**13543.00**	**13597.67**

insights into potential relationships, they cannot be statistically proven. This means that no definitive conclusions can be drawn regarding the significance of climate factors in Sarawak.

Rainfall exhibits varied effects on HFMD cases depending on the state. In Sabah, Wilayah Persekutuan, Kelantan, Johor, Melaka, and Negeri Sembilan, greater rainfall is associated with increased HFMD cases, with Johor (β2 = 0.0260 ***) and Melaka (β2 = 0.0377 ***) showing significant positive correlations. Conversely, in Perak (β2 = -0.0281 ***) and Perlis (β2 = -0.0203 ***), higher rainfall appears to reduce HFMD cases.

Relative humidity has a negative effect on HFMD cases, indicating that increased humidity levels are linked with a decrease in reported infections. This effect is particularly notable in Sabah (β3 = -0.0109 ***), where significant negative coefficients were observed. However, in Perlis, higher humidity corresponds with an increase in HFMD cases (β3 = 0.0108 **).

4.3 Model Prediction Accuracy

Table 4 presents the prediction accuracy for the five models using MAE, RMSE, and MAPE. The key findings include the following. The model prediction accuracy varied among different states. The Poisson model had the lowest MAE and RMSE in Sabah and Sarawak, indicating strong predictive performance. Similarly, in Selangor, the Poisson model performed best with the lowest MAE (17.3880) and RMSE (25.2163). For Wilayah Persekutuan, the Zero-inflated Poisson (ZIP) model demonstrated slightly better prediction accuracy compared with other models. In contrast, for Kelantan, Pahang, Terengganu, Kedah, Perak, Perlis, Penang, Johor, Melaka, and Negeri Sembilan, the ZINB model consistently exhibited the best prediction accuracy, particularly for Perak, Perlis, and Penang.

5. Discussion

The analysis highlights the ZINB model as the most appropriate for fitting and predicting HFMD cases based on climate factors in Malaysia. This model is particularly effective because HFMD data often display overdispersion and excess zeros (Greene, 1994).

The negative binomial component of the model handles overdispersion by introducing a dispersion parameter, which permits the variance to be greater than the mean (Greene, 1994):

$$\frac{Var[y_i]}{\mathbb{E}[y_i]} = 1 + \alpha\mathbb{E}[y_i] > 1 \tag{1}$$

where y_i represents HFMD case counts. This property makes the negative binomial distribution more flexible than the Poisson distribution, which assumes $Var[y_i] = \mathbb{E}[y_i]$ and may not capture HFMD's fluctuating case numbers (Greene, 1994). The

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 3. Factor Analysis of Climate Variables on HFMD Cases.

State	Temperature β_1	Rainfall β_2	Relative Humidity β_3
Sabah	0.1560 ***	0.0408 ***	-0.0109 ***
Sarawak	-0.2140	-0.0571	-0.0078
Selangor	0.0923 ***	-0.0072	0.0062 ***
Wilayah Persekutuan	0.1645 ***	0.0106 *	0.0043 *
Kelantan	0.0510	0.0277 *	-0.0022
Pahang	0.2297 ***	0.0055	-0.0012
Terengganu	0.1175 **	-0.0032	0.0026
Kedah	0.0024	0.0021	0.0014
Perak	0.0306 *	-0.0281 ***	0.0072 *
Perlis	0.0968 ***	-0.0203 ***	0.0108 **
Penang	0.0120	0.0018	0.0021
Johor	0.2672 ***	0.0260 ***	-0.0037
Melaka	0.3136 ***	0.0377 ***	-0.0052
Negeri Sembilan	0.1452 ***	0.0242 **	-0.0036

Significance levels: *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$

zero-inflation component improves model performance by differentiating between true zero cases and zeros due to randomness (Greene, 1994). The model assumes that each observation follows a mixture process:

$$Prob[y_i = 0] = Prob[z_i = 0] + Prob[z_i = 1, y_i^* = 0] \\ = q_i + (1 - q_i)f(0)$$

$$Prob[y_i = k] = (1 - q_i)f(k), k = 1,2, \dots \quad (2)$$

where q_i is the probability of true zero cases, and $f(k)$ represents the probability mass function of the negative binomial distribution (Greene, 1994). This formulation allows the model to better capture regions or periods where HFMD is zero.

Combining these two components, the ZINB model offers a flexible and accurate framework for modelling HFMD data, with better predictions than traditional count models. Its ability to account for overdispersion and excess zeros makes it highly suitable for HFMD forecasting and public health planning in general.

5.1 Key Climate Factors

The results show that climate factors significantly impact HFMD incidences across different Malaysian states. Temperature plays a critical role in HFMD transmission. Optimal temperature ranges can facilitate viral survival and transmission, while extremely high or low temperatures might reduce it. States such as Sarawak, Pahang, and Kedah showed significant temperature impacts, suggesting a need for region-specific temperature monitoring. Rainfall affects HFMD cases in states such as Selangor and Wilayah Persekutuan, as heavy rainfall can influence virus transmission by affecting human behaviour, such as increased indoor activities that facilitate close contact among children. Relative humidity is significant in several states, including Sabah and Selangor, where high humidity levels might support viral persistence in the environment, enhancing transmission risks. These insights can support public health strategies by allowing for targeted interventions based on seasonal and regional climate conditions to better prevent HFMD outbreaks.

5.2 Regional Differences

The variation in significant climate factors across states demonstrates the importance of localized approaches to HFMD prevention and control. For instance, while temperature was a significant factor in northern states, rainfall and relative humidity were more influential in central and southern regions. These differences require tailored public health strategies to address HFMD risks effectively.

6. Conclusion

This study highlights the significant impact of climate factors on HFMD incidences in Malaysia and identifies the Zero-inflated Negative Binomial model as the most suitable for data analysis and prediction. The findings guide developing targeted public health interventions and can serve as a reference for other regions with comparable climatic conditions. By enhancing our understanding of the climate-HFMD relationship, this research contributes to improved disease management and prevention strategies, ultimately protecting public health. The findings offer valuable insights for policymakers and health authorities, allowing them to implement timely interventions to mitigate HFMD outbreaks. By identifying key climate factors and utilizing the ZINB model for accurate predictions, authorities can take proactive measures such as public awareness campaigns to educate communities about the influence of climate on HFMD and promote preventive measures.

For future research, the study implication can be applied to other Asia-Pacific countries with similar climates, such as Singapore, Thailand, and Brunei. The ZINB model adoption may be useful in their public health system. One can consider an alternative method such as the Cox Proportional Hazards Regression Models for impact analysis (Khoo et al., 2022).

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 4. Model Prediction Accuracy.

State	Model	MAE	RMSE	MAPE
Sabah	Poisson	5.536938	**8.703627**	**98.92037**
	COM-Poisson	4807.440	136032.7	147.0561
	Zero-inflated Poisson	5.533498	8.707670	98.92683
	Zero-inflated Geometric	**5.531052**	8.711916	98.95993
	Zero-inflated Negative Binomial	5.532531	8.713534	98.99258
Sarawak	Poisson	16.82827	23.75413	72.76203
	COM-Poisson	46.00056	753.6395	95.94758
	Zero-inflated Poisson	16.82671	23.74165	72.76128
	Zero-inflated Geometric	16.80448	23.70645	**72.59127**
	Zero-inflated Negative Binomial	**16.80379**	**23.70450**	72.59155
Selangor	Poisson	17.38799	25.21629	**82.09430**
	COM-Poisson	18.99343	29.28693	93.02014
	Zero-inflated Poisson	**17.38590**	25.21893	82.09531
	Zero-inflated Geometric	17.39162	**25.21359**	82.09465
	Zero-inflated Negative Binomial	17.39151	25.21362	82.09614
Wilayah Persekutuan	Poisson	6.042738	8.839537	84.72911
	COM-Poisson	6.352827	9.509369	90.31472
	Zero-inflated Poisson	**6.042544**	8.840781	84.76679
	Zero-inflated Geometric	6.045528	8.842220	84.86904
	Zero-inflated Negative Binomial	6.043901	**8.838485**	**84.61422**
Kelantan	Poisson	2.859967	4.614232	110.1171
	COM-Poisson	3.099292	5.711534	124.9455
	Zero-inflated Poisson	2.858334	4.612774	110.0435
	Zero-inflated Geometric	**2.857076**	**4.611442**	**109.9434**
	Zero-inflated Negative Binomial	2.862506	4.613632	23650.59
Pahang	Poisson	1.587677	2.622078	111.8821
	COM-Poisson	1.614908	2.659068	116.2612
	Zero-inflated Poisson	**1.587451**	**2.621841**	111.8816
	Zero-inflated Geometric	1.588179	2.622170	**111.8491**
	Zero-inflated Negative Binomial	1.589665	2.622696	113.4683
Terengganu	Poisson	1.294917	2.334898	128.5674
	COM-Poisson	1.313821	2.350453	134.4711
	Zero-inflated Poisson	**1.294782**	2.334745	**128.5229**
	Zero-inflated Geometric	1.294963	**2.334688**	128.5535
	Zero-inflated Negative Binomial	1.295264	2.334885	128.6255
Kedah	Poisson	2.848100	4.811941	89.24372
	COM-Poisson	2.868848	4.817015	89.48412
	Zero-inflated Poisson	2.848474	4.812269	89.21289
	Zero-inflated Geometric	2.843152	**4.805715**	**88.84789**
	Zero-inflated Negative Binomial	**2.843143**	4.806401	88.88041
Perak	Poisson	4.765443	7.027617	91.84752
	COM-Poisson	4.817328	7.249890	94.53100
	Zero-inflated Poisson	4.764963	7.029099	91.87115
	Zero-inflated Geometric	**4.761979**	7.027808	91.91114
	Zero-inflated Negative Binomial	4.762420	**7.027417**	**91.76178**

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Table 4. Model Prediction Accuracy (continue)

State	Model	MAE	RMSE	MAPE
Perlis	Poisson	0.885055	1.318105	123.8650
	COM-Poisson	0.885864	1.319426	124.5666
	Zero-inflated Poisson	0.885095	1.318152	123.8726
	Zero-inflated Geometric	0.885037	1.318042	123.8427
	Zero-inflated Negative Binomial	**0.884852**	**1.317895**	**123.5608**
Penang	Poisson	4.135669	**6.090004**	78.45892
	COM-Poisson	4.164218	6.125970	**78.11093**
	Zero-inflated Poisson	4.136250	6.091779	78.37926
	Zero-inflated Geometric	4.132330	6.091060	78.15331
	Zero-inflated Negative Binomial	**4.131520**	6.090470	78.13294
Johor	Poisson	**5.232985**	7.531515	86.55955
	COM-Poisson	5.456939	7.949198	91.26989
	Zero-inflated Poisson	5.245626	**7.523421**	85.84830
	Zero-inflated Geometric	5.245267	7.525213	85.77475
	Zero-inflated Negative Binomial	5.245637	7.524500	**85.67280**
Melaka	Poisson	2.769605	**4.312197**	**104.6849**
	COM-Poisson	2.871048	4.432487	111.0587
	Zero-inflated Poisson	**2.769525**	4.312394	104.6876
	Zero-inflated Geometric	2.769760	4.312516	104.6922
	Zero-inflated Negative Binomial	2.770196	4.312783	104.7040
Negeri Sembilan	Poisson	2.597698	4.021901	99.74504
	COM-Poisson	2.642547	4.089556	101.6317
	Zero-inflated Poisson	2.595862	4.020736	99.71958
	Zero-inflated Geometric	2.591382	**4.019127**	**99.62895**
	Zero-inflated Negative Binomial	**2.591132**	4.019977	99.72834

during high-risk periods. Improved surveillance can also help monitor climate data to predict and prepare for HFMD outbreaks, with quicker responses. Developing region-specific strategies based on significant climate factors ensures more effective and customized HFMD control measures.

7. Acknowledgement

The authors of this paper are supported by the Ministry of Higher Education grant FRGS/1/2020/STG06/SYUC/02/1. The authors would also like to express their appreciation to the Institute for Medical Research (IMR) for providing the data used in this study.

8. References

Abdul Wahid, N. A., Suhaila, J., & Rahman, H. Abd. (2021). Effect of climate factors on the incidence of hand, foot, and mouth disease in Malaysia: A generalized additive mixed model. Infectious Disease Modelling, 6, 997–1008. https://doi.org/10.1016/j.idm.2021.08.003

Abdul Wahid, N. A., Suhaila, J., & Sulekan, A. (2020). Temperature effect on HFMD transmission in Selangor, Malaysia. Sains Malaysiana, 49(10), 2587–2597. https://doi.org/10.17576/jsm-2020-4910-24

Ang, L. W., Koh, B. K., Chan, K. P., Chua, L. T., James, L., & Goh, K. T. (2009). Epidemiology and control of hand, foot and mouth disease in Singapore, 2001-2007. Annals of the Academy of Medicine, Singapore, 38(2), 106–112. https://doi.org/10.47102/annals-acadmedsg.v38n2p106

Chang, L.-Y., King, C.-C., Hsu, K.-H., Ning, H.-C., Tsao, K.-C., Li, C.-C., Huang, Y.-C., Shih, S.-R., Chiou, S.-T., Chen, P.-Y., Chang, H.-J., & Lin, T.-Y. (2002). Risk factors of enterovirus 71 infection and associated hand, foot, and mouth disease/herpangina in children during an epidemic in Taiwan. Pediatrics, 109(6). https://doi.org/10.1542/peds.109.6.e88

Chavan, N. A., Lavania, M., Shinde, P., Sahay, R., Joshi, M., Yadav, P. D., Tikute, S., Waghchaure, R., Ashok, M., Gupta, A., Mittal, M., Khan, V., Fomda, B. A., Ahmad, M., Tiwari, V. P., Pote, P., Dhongade, A. R., Mohanty, A., Mohan, K., … Bhardwaj, A. (2023). The 2022 outbreak and the pathobiology of the Coxsackie virus [hand foot and mouth disease] in India. Infection, Genetics and Evolution, 111, 105432. https://doi.org/10.1016/j.meegid.2023.105432

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

Chen, S., Liu, X., Wu, Y., Xu, G., Zhang, X., Mei, S., Zhang, Z., O'Meara, M., O'Gara, M. C., Tan, X., & Li, L. (2019). The application of meteorological data and search index data in improving the prediction of HFMD: A study of two cities in Guangdong Province, China. *Science of The Total Environment*, *652*, 1013–1021. https://doi.org/10.1016/j.scitotenv.2018.10.304

Chin, S. N., Lem, F. F., Toh, J. Y., Chee, F. T., Yew, C. W., Saptu, A. R., Sampil, N. S., Mathew, M. G., & Sharip, J. (2022). A retrospective analysis on incidence of hand, foot, and mouth disease in Kota Kinabalu, district of Sabah Malaysia. International Journal of Public Health Science (IJPHS), 11(2), 369. https://doi.org/10.11591/ijphs.v11i2.21294

Conway, R.W. & Maxwell, W.L. (1962) A queuing model with state dependent service rates. Journal of Industrial Engineering, 12, 132-136.

Greene, William H., Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models (March 1994). NYU Working Paper No. EC-94-10, Available at SSRN: https://ssrn.com/abstract=1293115

Kim, B. I., Ki, H., Park, S., Cho, E., & Chun, B. C. (2016). Effect of climatic factors on hand, foot, and mouth disease in South Korea, 2010-2013. PLOS ONE, 11(6). https://doi.org/10.1371/journal.pone.0157500

Khoo, W. C., Yeah, K. L., Hong, S. Y. (2022). Modeling unemployment duration, determinants and insurance premium pricing of Malaysia: insights from an upper middle-income developing country. SN Business & Economics 2(8), 116.

Luo, C., Qian, J., Liu, Y., Lv, Q., Ma, Y., & Yin, F. (2022). Long-term air pollution levels modify the relationships between short-term exposure to meteorological factors, air pollution and the incidence of hand, foot and mouth disease in children: A DLNM-based multicity time series study in Sichuan Province, China. BMC Public Health, 22(1). https://doi.org/10.1186/s12889-022-13890-7

Nanjundan, N.G., & Pasha, S.S. (2015). A characterization of zero-inflated geometric model. *International Journal of Mathematics Trends and Technology*, *23*(1), 71–73. https://doi.org/10.14445/22315373/ijmtt-v23p510

Onozuka, D., & Hashizume, M. (2011). The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. Science of The Total Environment, 410–411, 119–125. https://doi.org/10.1016/j.scitotenv.2011.09.055

Puenpa, J., Wanlapakorn, N., Vongpunsawad, S., & Poovorawan, Y. (2019). The history of enterovirus A71 outbreaks and molecular epidemiology in the Asia-Pacific region. Journal of Biomedical Science, 26(1). https://doi.org/10.1186/s12929-019-0573-2

Sham, N. M., Krishnarajah, I., Shitan, M., & Lye, M.-S. (2014). Time series model on hand, foot and mouth disease in Sarawak, Malaysia. Asian Pacific Journal of Tropical Disease, 4(6), 469–472. https://doi.org/10.1016/s2222-1808(14)60608-3

Sun, L., Lin, H., Lin, J., He, J., Deng, A., Kang, M., Zeng, H., Ma, W., & Zhang, Y. (2016). Evaluating the transmission routes of hand, foot, and mouth disease in Guangdong, China. American Journal of Infection Control, 44(2). https://doi.org/10.1016/j.ajic.2015.04.202

Verma, S., Razzaque, M. A., Sangtongdee, U., Arpnikanondt, C., Tassaneetrithep, B., Arthan, D., Paratthakonkun, C., & Soonthornworasiri, N. (2021). Hand, foot, and mouth disease in Thailand: A comprehensive modelling of epidemic dynamics. Computational and Mathematical Methods in Medicine, 2021, 1–15. https://doi.org/10.1155/2021/6697522

Wahid, N. A., Suhaila, J., & Rahman, H. Abd. (2021). The influence of climate factors on hand-foot-mouth disease: A five-state study in Malaysia. Universal Journal of Public Health, 9(5), 324–331. https://doi.org/10.13189/ujph.2021.090515

Wan Mohamad, W. M., Ismail, N. A., & Mohd Fuzi, N. M. (2021). Risk factors of hand, foot and mouth disease (HFMD) outbreak cases among children under five years old in North Eastern State of Malaysia. Malaysian Journal of Public Health Medicine, 21(2), 315–320. https://doi.org/10.37268/mjphm/vol.21/no.2/art.1028

Wu, Y., Wang, T., Zhao, M., Wang, S., & Shi, J. (2022). Spatiotemporal cluster patterns of hand, foot, and mouth disease at the province level in mainland China, 2011-2018. PLoS ONE, 17(8), e0270061. https://doi.org/10.21203/rs.3.rs-1601423/v1

DOI: https://doi.org/10.22452/mjs.vol45sp1.3
Malaysian Journal of Science 45 (Special Issue 1): 20-28 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

https://mjs.um.edu.my

Assessment of CO_2 Emissions from Solid Waste at Kebon Kongok Landfill Using Warm Analysis

Intan Febrianti[1a], Norhafezah Kasmuri[2a*] and Satoto Endar Nayono[3b]

Abstract: The impact of climate change is interrelated with greenhouse gas emissions. The pollutants from greenhouse gases come from carbon dioxide (CO_2) and methane gases. Therefore, estimating the CO_2 emission contributing to the greenhouse phenomenon's hazardous environment is crucial. The Waste Reduction Model (WARM), created by the Environmental Protection Agency (EPA), provides high-level comparisons of potential greenhouse gas emissions reductions, energy savings, and economic impacts when considering different materials management practices. Materials management practices in WARM include source reduction, recycling, anaerobic digestion, combustion, composting, and landfilling. It can be noted that most landfill sites do not have materials recovery facilities for recycling purposes. Thus, it is crucial to estimate the potential impact of recycling efforts, as this could lead to a significant reduction in the amount of garbage going to landfills and limit CO_2 emissions. This study examines the environmental and economic impacts of solid waste management practices at the Kebon Kongok Landfill in West Lombok, Indonesia, and the potential of adopting recycling activities. WARM can estimate greenhouse gas (GHG) emissions, energy consumption, and potential revenue from recycling activities in these two regions (Mataram and West Lombok). Raw data from the daily waste collection in the landfill from both sites were collected, and the amount was inserted into WARM. The results demonstrate the efficacy of recycling scenarios in reducing GHG emissions and conserving energy. It can be noted that Mataram's recycling efforts can result in a value of 1,149.45 $MTCO_2E$ reduction, which is more than double the 450.81 $MTCO_2E$ in West Lombok. Hence, the energy saving of 8183.95 MBTU for Mataram can be significantly achieved compared to West Lombok (3159.87 MBTU). The research highlights the economic benefits of recycling for both regions if the recycling effort is adopted on-site. In Mataram, recycling activities can yield a substantial revenue of USD 45.6494 from mixed paper alone. Meanwhile, mixed paper recycling in West Lombok can generate a significant income of USD 18.0139. These findings underscore the effectiveness of recycling initiatives in reducing GHG emissions, conserving energy, and generating economic value while emphasising the importance of tailored waste management strategies to address regional variations in waste composition and consumption patterns. Therefore, strong policymaking can be introduced to adapt recycling activities for income generation, thereby transforming waste into wealth in achieving a circular economy.

Keywords: Solid waste management, greenhouse gas emissions, recycling activities, energy consumption, waste reduction model.

1. Introduction

Population expansion and an increase in urbanisation have contributed to the excessive volume of waste generated. Moreover, improper solid waste management has emerged as a critical issue that will pose significant health risks related to environmental pollution for the inhabitants (Tesseme et al., 2022; Velis & Cook, 2021; Vinti et al., 2021). In addition, with its vast territory and high population, Indonesia has grappled with a substantial amount of solid waste production, particularly in urban areas (Abir et al., 2023). The demographic data suggests that municipal solid waste in Indonesia has steadily risen, exacerbated by rapid urbanisation and population growth (Brotosusilo & Handayani, 2020; Gunawan et al., 2020). This surge of waste production strains existing waste management infrastructure and exacerbates environmental pollution, which later contributes to significant greenhouse gas (GHG) emissions. Therefore, effective waste management strategies must be implemented to mitigate the adverse impacts of municipal solid waste generation.

Greenhouse gas emissions from the waste sector, predominantly from landfill disposal, pose significant environmental concerns, including climate change and pollution (Chen et al., 2020; de Meij, 2020). The waste sector accounts for a substantial portion of national GHG emissions in Indonesia, underscoring the urgency for effective waste management interventions to mitigate climate impacts (MEF, 2021).

In response to global climate commitments, including the Paris Agreement, Indonesia has set ambitious targets to reduce GHG emissions. This necessitates proactive measures to improve waste management practices. However, addressing municipal solid waste (MSW) challenges requires more than just ambition. It demands localised solutions tailored to specific regions, considering unique socio-economic and environmental contexts.

Authors information:
[a]Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA, Email: intanbrianti@gmail.com[1], norhafezahkasmuri@uitm.edu.my[2]
[b]Universitas Negeri Yogyakarta, E-mail: satoto.nayono@uny.ac.id[3]

*Corresponding Author: norhafezahkasmuri@uitm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

NTB Province in Indonesia is one of the regions facing a pressing waste management issue. The province is contending with mounting waste generation, which is impairing environmental degradation and posing challenges to tourism development, a vital economic sector (Hardesty et al., 2016; Landrigan et al., 2020; Dabrowska et al., 2021). The situation at the Kebon Kongok landfill, a prominent waste disposal site in NTB Province, underscores the urgency of improving waste management practices and mitigating environmental impacts.

While previous studies have explored various aspects of waste management in the Kebon Kongok landfill, there is still a gap in understanding the comprehensive GHG emissions associated with different waste management scenarios. This study aims to fill this gap by quantifying GHG emissions from MSW management in the Kebon Kongok landfill using a waste reduction model (WARM). The use of WARM analysis as a free web tool provided by the Environmental Protection Agency (EPA), United States, is to aid solid waste planners and organisations in tracking and voluntarily reporting on greenhouse gas (GHG) emissions reductions, energy savings, and economic impacts from several different waste management practices (EPA, 2018). Therefore, the potential impact of adopting recycling activities in the regions, following the results estimated by WARM in transforming waste into wealth and achieving a circular economy, is significant in reducing GHG. Moreover, this research seeks to inform evidence-based decision-making and long-term planning for sustainable waste management in NTB Province by evaluating GHG emissions and energy savings. This study's significant contribution to the broader discourse on MSW management in Indonesia underscores the importance and impact of the research. Overall, this study is expected to significantly contribute to the broader discourse on MSW management in Indonesia and offer valuable insights into mitigating GHG emissions and promoting environmental sustainability in the region.

2. Research methodology

The case study was conducted at the Kebon Kongok landfill in West Lombok Regency, Indonesia. The landfill, which covers 8.6 hectares, has been operating since 2006. Data for this study were sourced from the landfill's Regional Technical Implementation Unit and the Department of Environment and Forestry Service of NTB Province, demonstrating a collaborative effort with local authorities to ensure the credibility of the study. The raw MSW data from the Kebon Kongok Landfill has been collected for one week, from 1st August to 8th August 2023. This data calculates recycling activities using a ratio percentage (EPA, 2018). Figure 1 depicts the flowchart outlining the municipal solid waste (MSW) data analysis process conducted for this research. This flowchart illustrates the steps for analysing MSW data collected from the Kebon Kongok landfill. The data obtained from the landfill was processed to determine weight and total amounts, referencing

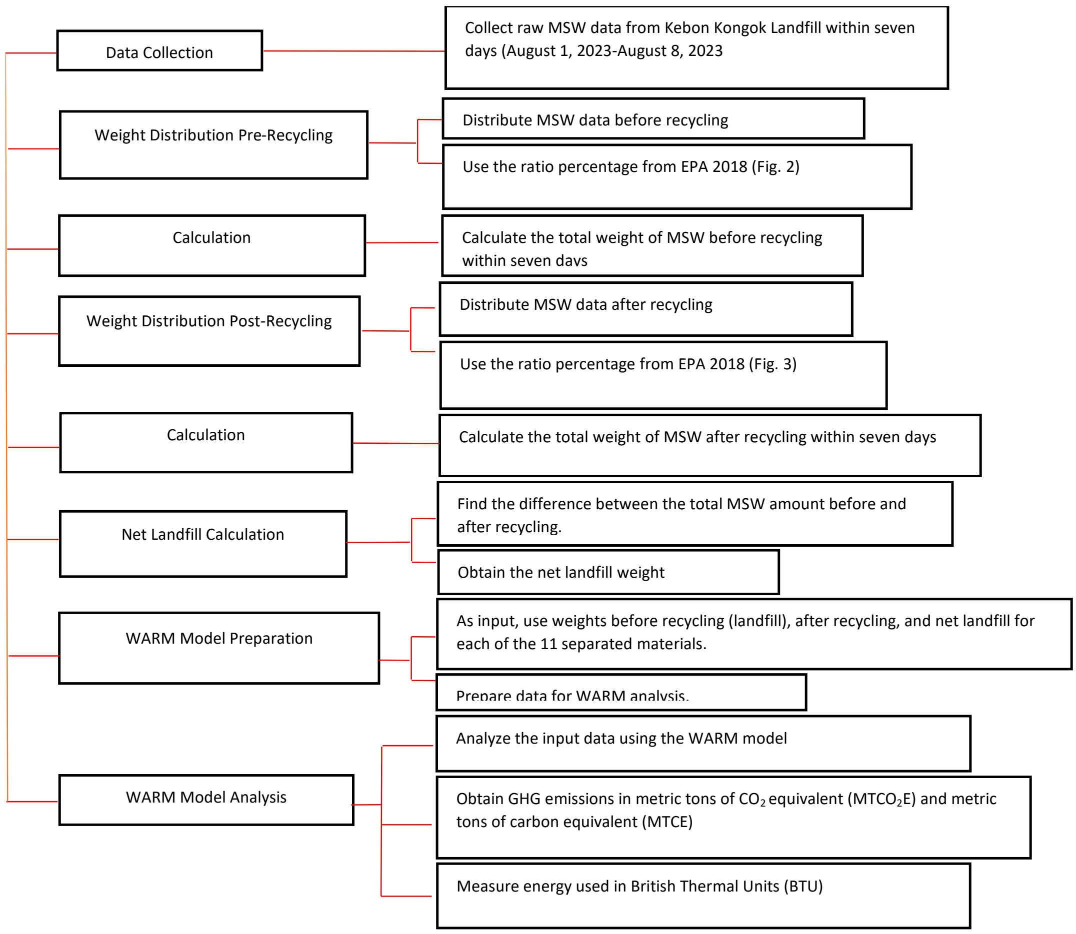


Figure 1. Municipal Solid Waste Data Analysis Flowchart

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

Figure 2 and Figure 3. Subsequently, the results obtained were assessed using WARM analysis to calculate GHG emissions and energy savings regarding carbon dioxide equivalents ($MTCO_2E$) and metric tonnes of coal equivalents (MTCE).

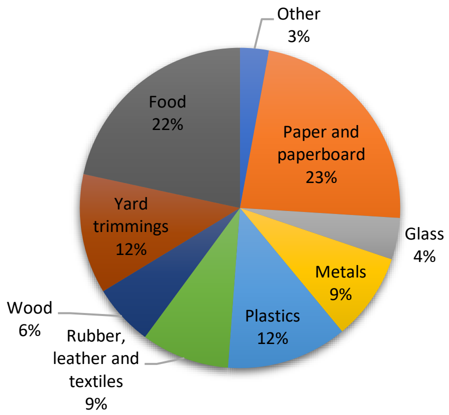


Figure 2. Total MSW generation 292.4 million tons (before recycling by material) (EPA, 2018)

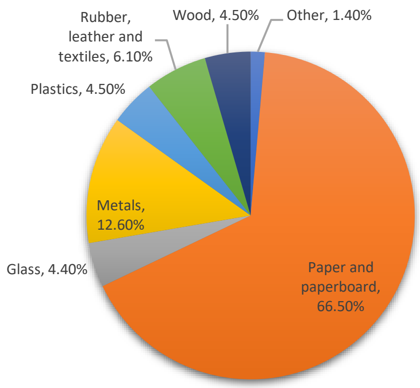


Figure 3. Total MSW recovery (69.1 million tons by material) (EPA, 2018)

2.1 Waste Reduction Model (WARM)

From the input of the weight distribution in pre-recycling and post-recycling (see Figure 1), the later steps are to assess the net landfill weight to estimate GHG emissions using the highly significant WARM model. The data analysis compared the baseline landfill scenario with alternative scenarios involving recycling, waste composition, and tonnage reduction. The WARM model, a crucial tool in the field, facilitated the calculation of GHG emissions across various material types commonly found in MSW. Key outputs from the model include:

a) Metric tonnes of carbon dioxide equivalent ($MTCO_2E$)

b) Metric tonnes of carbon equivalent (MTCE)

c) Energy units measured in million British thermal units (MBTU)

The study utilised WARM, developed by the United States Environmental Protection Agency (EPA), to assess GHG emissions associated with MSW management practices. WARM facilitated simulations of various waste management scenarios, allowing for comparisons between baseline strategies and alternative methods (EPA, 2018). The software was programmed to analyse data related to waste management practices, encompassing both GHG emissions and socio-economic factors, to evaluate the overall impact of different scenarios on global warming and socio-economic conditions. Engelmann et al. (2022) reported utilising WARM to assess a broader and more accurate range of waste management options, later enabling the analysis of alternative scenarios to identify the best solutions for sustainable MSW management, offering hope for a greener future.

2.2 WARM in this case study

In this analysis, recycling assumes that recycled materials substitute virgin inputs in manufacturing processes, with constant demand for new and recycled materials. Thus, increased recycling does not alter the overall production of materials. According to WARM, each tonne of recycled material displaces virgin material that would otherwise have been produced. The WARM analysis, a significant and impactful tool, simplifies the relationship between recycling and demand for products with recycled content by assuming no change in overall demand due to increased recycling (EPA, 2020). Therefore, in the context of Indonesian waste management conditions, the WARM analysis can significantly estimate greenhouse emissions by adopting an alternative scenario, such as recycling practices for the chosen site of study, in this case, the Kebon Kongok landfill.

2.3 Greenhouse gas emissions (GHG) and energy saving in WARM

GHG savings are determined by comparing emissions from an alternative scenario with those from a baseline scenario. The emissions of the baseline scenario cannot be easily calculated by multiplying quantity with an emission factor (EPA, 2018). It is crucial to have two scenarios for an effective application of WARM energy factors: one representing current management practices (baseline scenario) and another representing alternative practices. Analysing these scenarios allows for the calculation of energy consumption or avoidance in both cases. The difference between the alternative and baseline scenarios reveals the energy consumed or avoided due to the alternative management approach (EPA, 2018).

2.4 Income generated from recyclable materials

The income generated from recyclable materials was calculated based on their weight post-recovery, which was obtained from a comprehensive and accurate representation of the total weight of municipal solid waste (MSW) collected over a seven-day period. This data was used in conjunction with the ratios provided in the MSW Fact and Figure 2018 report (EPA, 2018) (Figures 2 & 3). The weight of each material was determined by multiplying the separation percentage ratio by the total recovery weight, as outlined in the MSW Fact and Figure Recovery guidelines (EPA, 2018). Revenue from recyclable materials was then calculated by multiplying each material's

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

Table 1. Market Price of Recyclable Generated (Rakhman et al., 2022)

Recyclable Items	Market Price (USD/Kg)
Glass	0.032
Mixed Paper	0.19
Mixed Metal	0.39
Mixed Plastic	0.16
Tyres	0.0054
Mixed Recyclable	NA

Note: NA- Not Available

composition by its respective market price (Rakhman et al., 2022). (Table 1). Table 1 comprises the net value of the price of recyclable materials, referring to the current market price in Indonesia (Rakhman et al., 2022).

3. Results and discussion

Figure 4 illustrates the total incremental GHG emissions in the form of carbon dioxide equivalents ($MTCO_2E$) for both Mataram and West Lombok, showcasing the source-segregated materials recycling developed in this research. Analysis of Figure 4 reveals the promising potential of recycling in reducing GHG emissions, as a positive net value indicates $MTCO_2E$ emissions savings. The results indicate nine materials with net $MTCO_2E$ savings from the recycled materials analysis. However, one non-recyclable material, yard trimmings destined for composting, exhibited negative values for incremental GHG emissions.

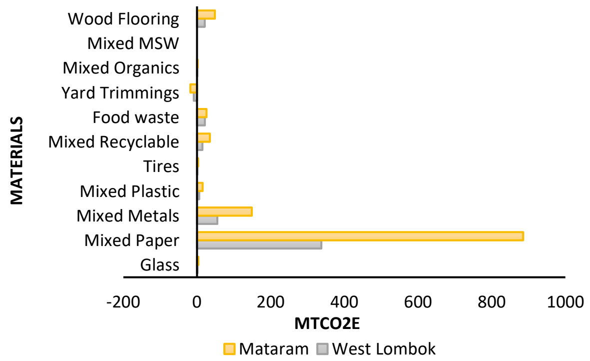


Figure 4. Total incremental $MTCO_2E$ emission

The extent of GHG savings varied significantly among Mataram's different materials. Mix paper exhibited the highest savings in total incremental GHG emissions for $MTCO_2E$ (Figure 4), with 885.46 $MTCO_2E$ for Mataram and 337.58 $MTCO_2E$ for West Lombok. Conversely, yard trimmings showed the highest incremental GHG emissions, with 18.03 MT for Mataram and 8.84 MT for West Lombok. The carbon emission results per tonne produced, depicted in Figure 5, revealed that mixed paper resulted in the highest GHG emissions savings in MTCE, with 233.31 MTCE for Mataram. Yard trimmings for Mataram recorded the highest incremental GHG emissions at 4.92 MTCE, compared to 2.41 MTCE for West Lombok. These emission variations underscore the potential impact of source reduction and recycling practices on the net GHG emissions of the materials produced, emphasising the urgency of our role in waste management.

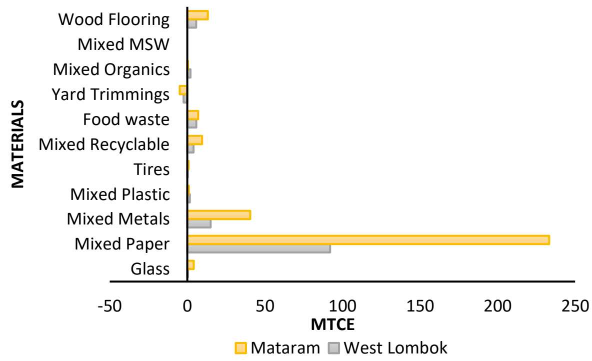


Figure 5. Total incremental of MTCE emission

The energy usage from both the baseline and alternative scenarios was assessed using the Waste Reduction Model (WARM), a widely accepted tool for evaluating the environmental impact of waste management options. Results from Mataram indicated the highest incremental energy savings, particularly from mixed paper, totalling 4903.55 MBTU. However, the maximum energy consumption was observed in the yard trimming material, with a value of 14.07 MBTU, as depicted in Figure 6. Yard trimmings were identified as contributors to increased energy consumption due to the decomposition of organic materials, which generate methane gas (CH_4). It is crucial to recognise that the atmospheric warming potential of CH_4 is 28–36 times that of CO_2 when averaged over 100 years (Chai et al., 2016). However, the recycling materials depend on the solid waste generated in both regions, and uncertainties regarding the value will later make a difference in the outcomes of the WARM analysis.

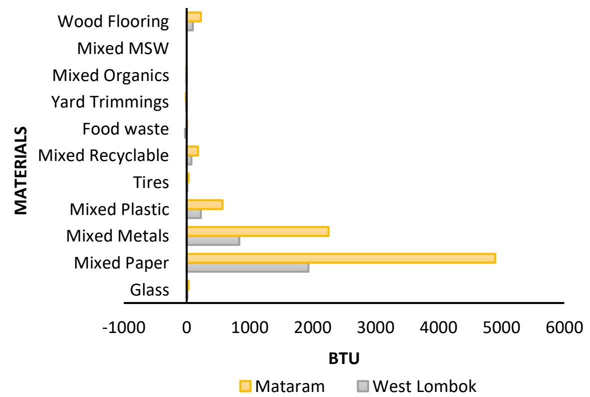


Figure 6. Total incremental energy consumption

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

Figure 7 illustrates the total incremental GHG emissions of $MTCO_2E$ for Mataram and West Lombok over seven days (1 week). Core recyclables, including glass, mixed paper, metals, plastics, tyres, and others, significantly reduced GHG emissions in both Mataram and West Lombok. In Mataram, core recyclables saved 1,091.37 tonnes of $MTCO_2E$, while West Lombok saved 416.29 $MTCO_2E$. Although other materials like food waste, yard trimmings, and wood flooring also showed reductions, their impact was smaller. Mataram saved 58.08 $MTCO_2E$, and West Lombok saved 34.52 $MTCO_2E$ by recycling these materials. Overall, Mataram achieved a total GHG emission reduction of 1,149.45 $MTCO_2E$, compared to 450.81 $MTCO_2E$ in West Lombok. It highlights the potential of recycling initiatives to impact carbon emissions substantially.

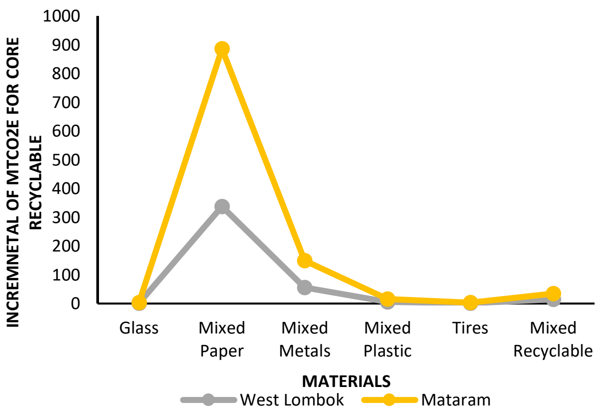


Figure 7. Total incremental of $MTCO_2E$ emission for core recyclable

The core recyclable material for MTCE emission yielded the same result (Figure 8), with the maximum saving recorded by Mataram at 289.47 MT, generating 322.53 tonnes of recycled solid waste. Like core recyclables, other materials also showed noteworthy reductions, contributing to 3096.51 MBTU in West Lombok and 7968.59 MBTU in Mataram (Figure 9). Mataram obtained the greatest total energy saving for core recyclable materials. At the same time, other materials also exhibited smaller reductions, resulting in a decrease of 63.36 MBTU in West Lombok and 215.36 MBTU in Mataram. Furthermore, the maximum energy saving was from Mataram at 8183.95 MBTU for 509.5 tonnes of total weight compared to West Lombok's 3159.87 MBTU for 232.55 tonnes.

In the context of energy-saving consumption, it is noteworthy that mixed paper exhibited a considerably more substantial value in terms of energy savings compared to GHG emissions. This discrepancy can be attributed to variations in a pivotal parameter known as the value factor. The energy factor was measured at 20.56, while the emission factor stood at 3.55 (EPA, 2023). Moreover, the GHG emissions for cardboard and office paper are lower than energy usage ratings in WARM due to its assumption of forest carbon credits for recycling paper. Forest carbon credits are a form of tradeable carbon offset that represent the avoidance of carbon emissions from deforestation or degradation of forests. Recycling reduces tree harvesting, increasing carbon

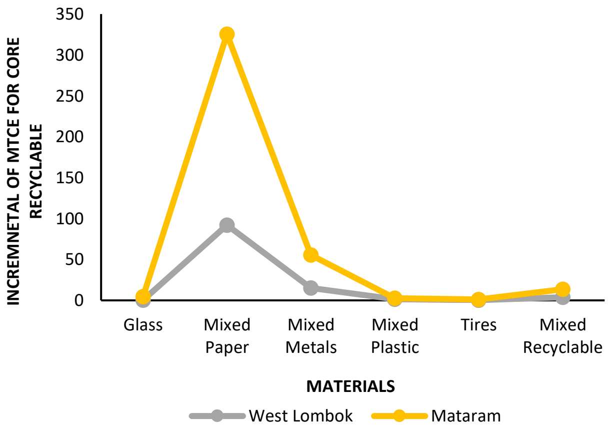


Figure 8. Total incremental MTCE emission for core recyclable

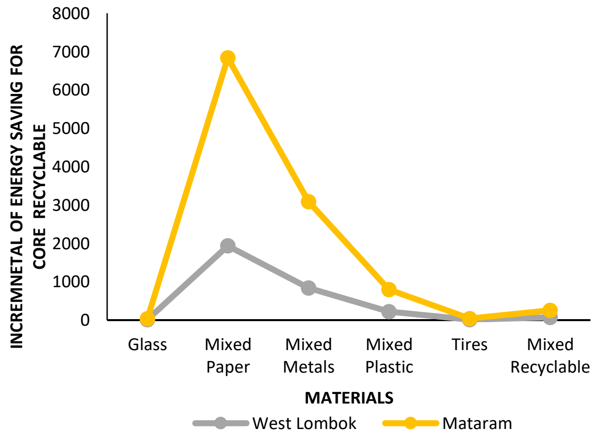


Figure 9. Total incremental energy used for recyclable core

storage in soil. Other models do not include these credits (Anshassi & Townsend, 2021).

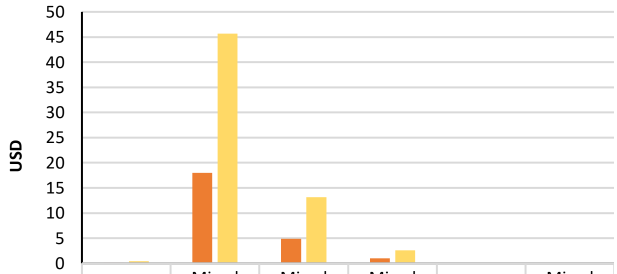
The reduction in energy consumption achieved through recycling plays a pivotal role in minimising the environmental footprint and concurrently reducing the emission of harmful pollutants. Furthermore, saving energy from recycling not only lessens dependence on fossil fuels but also leads to significant cost savings, underlining the economic benefits of sustainable practices (Mallick et al., 2023).

GHG release is closely tied to community waste generation volume. This finding is supported by a study conducted by Abu Hassan et al. (2019). Their research established that increased waste production correlates with higher greenhouse gas emissions. Consequently, actively decreasing GHG emissions becomes imperative to mitigate environmental impact, reduce pollutants, and enhance control, ultimately prioritising the well-being and health of the local population (Abu Hassan et al., 2019). Moreover, to reduce GHG emissions, there is technology in landfills for gas capture for methane gases as a mitigation strategy

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

for limiting GHG emissions. Therefore, lifecycle-based environmental assessment can be conducted to mitigate landfill gas emissions. Environmental professionals, policymakers, and stakeholders play a crucial role in implementing these measures, such as cover and gas collection, along with energy utilisation at ageing landfills. This will lead to significant environmental improvements compared to the baseline scenario with no emission mitigation action (Scheutz et al., 2023).

The estimation of revenue from recyclables, as presented in Figure 10, sheds light on the potential economic benefits of recycling initiatives in West Lombok and Mataram. In West Lombok, recyclable items include glass, mixed paper, mixed metal, mixed plastic, tyres, and mixed recyclables. The market prices per kilogram and their respective amounts in tonnes contribute to the estimated revenue. Notably, the high market price of mixed metal at USD 0.39 per kilogram results in a substantial value of USD 4.8828 for the 12.52 tonnes recycled, indicating a lucrative avenue for generating revenue. Similarly, mixed paper, with a significant amount of 94.81 tonnes and a market price of USD 0.19 per kilogram, contributes substantially to the total revenue, amounting to USD 18.0139.



	Glass	Mixed Paper	Mixed Metal	Mixed Plastic	Tires	Mixed Recyclable
West Lombok	0.1651	18.0139	4.8828	1.0192	0.0167	0
Mataram	0.3994	45.6494	13.1586	2.584	0.042	0

Figure 10. Comparison of total estimation for recycling

In Mataram, a similar analysis reveals the significant economic potential of recycling. With a market price of USD 0.032 per kilogram and 12.48 tonnes, glass contributes USD 0.3994 to the estimated revenue. The most significant contributor, mixed paper, with 240.26 tonnes and a market price of USD 0.19 per kilogram, generates a substantial value of USD 45.6494. These findings, corroborated by research in the field, such as the work of Larouche et al. (2020), underscore the economic benefits of recycling programmes. They demonstrate how strategic emphasis on high-value recyclables can significantly contribute to overall revenue, providing a promising outlook for sustainable waste management (Larouche et al., 2020).

These correlations underscore the crucial role of tailored recycling strategies that leverage market dynamics and emphasise economically valuable recyclable materials. As Ismail et al. (2018) point out, the potential revenue generated from recycling three key materials, aluminium, plastic, and paper, amounts to RM 42.50 within a short timeframe. This underscores the belief that instituting waste separation at the source not only contributes to environmental conservation but also provides a viable avenue for generating additional income for the institution. This knowledge empowers us to make informed decisions about waste management, enhancing our ability to protect the environment and our financial interests (Ismail et al., 2018).

The findings from this study of mixed paper recycling align with the findings from Samadikun et al. (2021), which specifically underscore the stable economic potential of paper waste recycling within the informal recycler sector in Purwodadi Sub-District (Samadikun et al., 2021). This stability is a reassuring factor in the volatile world of waste economics. Another study highlighted the importance of plastic recycling for Banda Aceh's economic growth. However, metals, once valuable, are becoming scarcer due to advancements in material engineering, leading to price fluctuations. In contrast, paper waste retains stable demand and price due to its consistent supply and demand pattern (Nizar et al., 2019). These implications underscore the need for further extensive consideration and detailed analysis of metal and plastic waste values in Indonesia, particularly in regions like Mataram and West Lombok, where significant revenue generation potential is observed in metals and paper, respectively.

In a different country, such as Nepal, it is suggested that the recovery of plastic materials can generate revenue, amounting to 1.38 times the cost of managing plastic waste. This potential for increased revenue is a reason for optimism in the field of waste management. It becomes achievable when the collection efficiency reaches 66.7%. Additionally, a mere 1% increase in the recovery rate and collection efficiency could cover an additional 4.64% and 2.06% of the costs of managing plastic waste, respectively (Nepal et al., 2020). However, a study conducted by Vogt et al. (2021) reported that the effective collection of used plastic waste for recycling is probably reliant on financial motivations and regulatory measures established by the government (Vogt et al., 2021). Businesses need to broaden the role of producers and encourage residents to participate actively in recycling efforts (Liu et al., 2020).

In a study by Rangga et al. (2022), it is evident that certain regions like Johor, Kedah, and Pahang in Malaysia have gained substantial profits from the sale of specific types of waste, such as plastic, paper, metal, glass, and aluminium (Rangga et al., 2022). In the United States, aluminium and plastic are reported to be major contributors to the economy after 'e-waste' recycling (EPA, 2020). 'E-waste' refers to electronic waste, such as old computers, mobile phones, and other electronic devices. Additionally, a study by Han et al. (2021) highlighted the profit potential of recycling e-waste in Japan and Korea (Han et al., 2021). Nevertheless, it is crucial to consider the costs associated with the recycling method. Genc et al. (2019) reported that the overall cost of plastic recycling through mechanical means becomes economically viable only when there is an improvement in the recovery of plastics (Genc et al., 2019). Thus, it is important to note that the profits from recycling activities can vary depending on the market price of the materials in each region and the recycling method.

It can be denoted that through proper collection of municipal solid waste, valuable sources of recycling materials such as metal can be recovered and used in new products. The latter will significantly lower the production cost and, more importantly,

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

reduce the environmental impact (Larouche et al., 2020). Moreover, material recovery facilities can be included in the landfill, as the operational costs of waste management can be diminished due to the reduction of waste materials that end up in the landfill. From the results analysis, the concept of optimal waste transformation into wealth is gaining popularity, and even though fluctuations in market prices for recyclables could affect economic feasibility, the future of waste management looks promising with the increasing demand for recyclable materials, as these valuable resources and products exist in waste streams, focusing mainly on their monetary value (Kalkanis et al. 2022).

4. Conclusion

This research was conducted to comprehensively evaluate CO_2 emissions and energy consumption associated with solid waste management practices at the Kebon Kongok Landfill in West Lombok, Indonesia. By leveraging WARM, the research investigated the environmental impact of waste management strategies and compared waste segregation methods between Mataram and West Lombok. The study discovered key insights regarding the composition and generation of MSW at the landfill. The high organic content was a notable finding, with food waste and yard trimmings particularly prominent. By applying the WARM analysis, the study demonstrated the efficacy of the recycling scenario, especially for core recyclables such as mixed paper, plastics, and metals, in reducing GHG emissions and conserving energy. It aligns with the overarching goals of sustainability. By estimating the revenue that could be generated from recyclable materials, the research underscored the potential economic benefits of recycling programmes, which can contribute significantly to overall revenue and foster a circular economy. Moreover, the study revealed distinct waste generation patterns between West Lombok and Mataram, indicating the effect of regional variations in consumption patterns and waste management practices. The research emphasised the critical role of community engagement and behavioural changes in waste reduction and initial collection. These factors are essential to improving recycling rates and minimising environmental impact. Thus, the results of the evaluation of effective recycling strategies can influence policymakers to revise regulations or provide incentives for recycling programmes, which can guide industries, businesses, and citizens toward sustainable waste management practices.

5. Acknowledgement

The authors gratefully acknowledge the Faculty of Civil Engineering and Research Management Centre (RMC), Universiti Teknologi MARA, Shah Alam, for providing the resources and financially supporting this research study.

6. References

Abir T M., Datta M and Saha S R. (2023) Assessing the Factors Influencing Effective Municipal Solid Waste Management System in Barishal Metropolitan Areas Journal of Geoscience Environmental Protection 11(1):49–66 doi: 10.4236/gep.2023.111004.

Abu Hassan M., Kasmuri N and Basri M. (2019) Assessment of CO_2 Emission and Energy Reduction on Solid Waste in Jeram Landfill Using Warm Analysis International *Journal of Integrated Engineering* 11:019 doi: 10.30880/ijie.2019.11.01.019.

Anshassi M and Townsend T G. (2021) Reviewing the underlying assumptions in waste LCA models to identify impacts on waste management decision making *Journal of Cleaner Production* 313:127913 doi: 10.1016/j.jclepro.2021.127913.

Brotosusilo A and Handayani D. (2020) Dataset on waste management behaviours of urban citizens in large cities of Indonesia *Data in Brief* 32 doi: 10.1016/j.dib.2020.106053.

Chai X., Tonjes D J and Mahajan D. (2016) Methane emissions as energy reservoir: Context, scope, causes and mitigation strategies *Progress in Energy and Combustion Science* 56:33-70 doi: 10.1016/j.pecs.2016.05.001.

Chen D M C., Bodirsky B L., Krueger T., Mishra A and Popp A. (2020) The world's growing municipal solid waste: trends and impacts *Environmental Research Letters* 15(7) doi: 10.1088/1748-9326/ab8659.

Dąbrowska J., Sobota M., Świderska M., Borowski P., Moryl A., Stodolak R., Kucharczak E., Zięba Z and Kazak J. (2021) Marine waste—Sources, fate, risks, challenges and research needs International Journal of Environmental Research & Public Health 18(2):433 doi: 10.3390/ijerph18020433.

de Meij A. (2020) Annual European Union greenhouse gas inventory 1990–2018 and inventory report 2020 *European Commission in DG Climate Action European Environment Agency.*

Engelmann P., Santos V., Rocha P., Santos G., Lourega R., Lima J and Pires M. (2022) Analysis of solid waste management scenarios using the WARM model: Case study *Journal of Cleaner Production* 345:130687-130687 doi 10.1016/j.jclepro.2022.130687.

Genc A., Zeydan Ö., and Sarac S. (2019) Cost analysis of plastic solid waste recycling in an urban district in Turkey *Waste Management & Research* 37:906-913 doi: 10.1177/0734242X19858665.

Gunawan J., Permatasari P and Tilt C. (2020) Sustainable development goal disclosures: Do they support responsible consumption and production? *Journal of Cleaner Production* 246 doi: 10.1016/j.jclepro.2019.118989.

Han J., Ijuin H., Kinohista Y., Yamada T., Yamada S and Inoue M. (2021) Sustainability Assessment of Reuse and Recycling Management Options for End-of-Life Computers-Korean and Japanese Case Study Analysis *Recycling* 6:55 doi: 10.3390/recycling6030055.

Hardesty B et al., (2016) Estimating quantities and sources of marine debris at a continental scale *Frontiers in Ecology and the Environment* 15 doi: 10.1002/fee.1447.

Ismail H., Hassan C., Jamal M., Martin J., Mat Salleh M Z., Chik W., N. Rahim, and Yusof N A U. (2018) Potential revenue from solid waste recycling practices in Faculty of Civil Engineering UiTM Johor, Pasir Gudang Campus *Journal of Applied and Fundamental Sciences* 10:2540-2550 doi: 10.4314/jfas.v10i6s.193.

Kalkanis K., Alexakis D E., Kyriakis E. et al. (2022) Transforming Waste to Wealth, Achieving Circular Economy *Circular Economy and Sustainability* 2:1541–1559 https://doi.org/10.1007/s43615-022-00225-2.

Landrigan P., Stegeman J., Fleming L., Allemand D., Anderson D., Backer L., Brucker-Davis F., Chevalier N., Corra L., Czerucka D., Dechraoui Bottein M Y., Demeneix B., Depledge M., Deheyn D., Dorman C., Fénichel P., Fisher S., Gaill F., Galgani F and Rampal P. (2020) Human health and ocean pollution *Annals of Global Health* 86(1):1-64 doi: 10.5334/aogh.2831.

Larouche F., Tedjar F., Amouzegar K., Houlachi G., Bouchard P., Demopoulos G and Zaghib K. (2020) Progress and Status of Hydrometallurgical and Direct Recycling of Li-Ion Batteries and Beyond *Materials* 13:801 doi: 10.3390/ma13030801.

Liu M., Tan S., Zhang M., He G., Chen Z., Fu Z and Luan C. (2020) Wastepaper recycling decision system based on material flow analysis and life cycle assessment: A case study of waste paper recycling from China *Journal of Environmental Management* 255:109859 doi 10.1016/j.jenvman.2019.109859.

Mallick K., Sahu A., Dubey N and Das A. (2023) Harvesting marine plastic pollutants-derived renewable energy: A comprehensive review on applied energy and sustainable approach *Journal of Environmental Management* 348:119371 doi 10.1016/j.jenvman.2023.119371.

Ministry of Environment and Forestry (MEF) Indonesia (2021) Directorate General of Climate Change *Third Biennial Update Report Under the United Nations Framework Convention on Climate Change*.

Nepal M., Bharadwaj B and Rai R. (2020) Sustainable financing for municipal solid waste management in Nepal *PLoS ONE* doi: 10.1371/journal.pone.0231933.

Nizar M., Munir E., Matseh I and Fakhrurrazi F. (2019) Examining the Economic Benefits of Urban Waste Recycle Based on Zero Waste Concepts doi: 10.2991/agc-18.2019.47.

Rakhman M., Busyairi M and Kahar A. (2022) Analisis Timbulan dan Komposisi Sampah Perumahan dan Non Perumahan Wilayah Kabupaten Kutai Kartanegara (Studi Kasus: Kecamatan Anggana) *Jurnal Teknologi Lingkungan* UNMUL 6(2):24-32 doi: 10.30872/jtlunmul.v6i2.8109.

Rangga J., Sharifah S I., Rasdi I and Karuppiah K. (2022) Waste Management Costs Reduction and the Recycling Profit Estimation from the Segregation Programme in Malaysia *Pertanika Journal of Science and Technology* 30:1457-1478 doi: 10.47836/pjst.30.2.34.

Samadikun B., Sinttia D., Rezagama A., Sumiyati S., Huboyo H., Ramadan B., Hadiwidodo M and Nabila F. (2021) The economic potential of paper waste recycling activities on the informal sector in Grobogan District a case study: Purwodadi Sub-district *IOP Conference Series: Earth and Environmental Science* 623:012077 doi: 10.1088/1755-1315/623/1/012077.

Scheutz C., Duan Z., Møller J., Kjeldsen P. (2023) Environmental assessment of landfill gas mitigation using biocover and gas collection with energy utilisation at aging landfills *Waste Management* 165:40-50.

Tesseme A T., Vinti G and Vaccari M. (2022) Pollution potential of dumping sites on surface water quality in Ethiopia using leachate and comprehensive pollution indices *Environmental Monitoring Assessment* 194(8):545 doi: 10.1007/s10661-022-10217-2.

U.S. Environmental Protection Agency (EPA). (2018) National Overview: Facts and Figures on Materials *Wastes and Recycling* United States.

U.S. Environmental Protection Agency (EPA). (2020) Documentation for greenhouse gas emission and energy factors used in the waste reduction model (WARM) United States.

U.S. Environmental Protection Agency (EPA). (2023) Documentation for greenhouse gas emission and energy factors used in the waste reduction model (WARM): Management practices chapters United States.

Velis C A and Cook E. (2021) Mismanagement of Plastic Waste through Open Burning with Emphasis on the Global South: A Systematic Review of Risks to Occupational and Public Health *Environmental Science Technology* 55(11):7186–7207 doi: 10.1021/acs.est.0c08536.

Vinti G et al., (2021) Municipal solid waste management and adverse health outcomes: A systematic review International Journal Environmental Research Public Health 18(8):1–26 doi: 10.3390/ijerph18084331.

Vogt B., Stokes K and Kumar S. (2021) Why is Recycling of Postconsumer Plastics so Challenging? *ACS Applied Polymer Materials* 3 doi: 10.1021/acsapm.1c00648.

DOI: https://doi.org/10.22452/mjs.vol45sp1.4
Malaysian Journal of Science 45 (Special Issue 1): 29-36 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024 Special issue

Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026) https://mjs.um.edu.my

Mode-Locked Based on Side-Polished Fiber Coated with Gold as a Medium for the Temperature Sensor

Noor Azura Awang[1a]* and Nik Noor Haryatul Eleena Nik Mahmud[2b]

Abstract: We present a mode-locked based on side-polished fiber (SPF) as a medium for temperature sensing. To develop a mode-locked fiber sensor (MLFS), an erbium-doped fiber and gold nanoparticles with 10 nm thickness were coated onto SPF. An MLFS pulse was generated at 38.5 MHz with a pulse interval of 0.0258 µs, matching the cavity round-trip time. The sensing part was tested under two conditions: heating (100 to 0 °C) and cooling (0 to 100 °C). The experiment revealed a linear relationship between the resonant wavelength shift and temperature variations, with sensitivities of 0.4079 and 0.3775 nm $°C^{-1}$ during heating and cooling, respectively. The SNR value increased from 43.9 to 56.3 dB, indicating stable output power precision. Moreover, the incorporation of a gold-nanoparticle coating on SPF improved sensor sensitivity and reliability while reducing the effects of external variables such as moisture and vibration. These properties also describe the sensor's strength and adaptability in demanding industrial and commercial applications. Future research will focus on assessing long-term performance under changing environmental conditions, thereby ensuring dependability across varied contexts.

Keywords: Gold nanoparticle, Fiber laser, Mode-locked, Side-polished fiber, Temperature sensor.

1. Introduction

The characteristics of mode-locked fiber laser (MDFL), such as narrow pulse width, wide spectrum, and high pump power, could lead to potential uses such as environmental sensing, material processing, and medicine (Fried et al., 2005; Li et al., 2022; Stewart et al, 2003; Zhang et al., 2002). A passive MDFL with a saturable absorber (SA) is more effective at generating ultrashort pulses than an active MDFL that employs active modulation. The use of SA changes resonator losses much faster than active electronics. Additionally, it is more economical and compact (Lu et al., 2018). A prior study indicated that pulse fiber lasers effectively achieved operation under the impact of SA, including 2D materials, topological insulators (TI), Semiconductor Saturable Absorber Mirror (SESAM), black phosphorus (BP), and transition metal dichalcogenides (TMD) within the laser cavity (Jung et al., 2014; Matsas et al., 1992; Novoselov et al., 2005; Song et al., 2017; Trushin et al., 2016). However, some SA materials have severe limitations, including low stability and performance, a low damage threshold, and a narrow operating wavelength (Keller, 2004). Recently, Guo et al. and Ahmad et al. reported their work using silver nanoparticles (SNPs) because SA can produce pulsed fiber lasers. Thus, this suggests that transition-metal exploration includes silver and gold, as SA has emerged as a new approach to ultrashort-pulse lasers due to its unique optical properties (Liao et al., 1997; Zhou et al., 1994).

Recently, fiber laser temperature sensors have attracted significant attention from researchers due to their high sensitivity, high signal-to-noise ratio (SNR), ease of manufacture, and good anti-interference performance (Madry et al., 2019; Gonzalez-Reyna et al., 2015). Numerous optical fiber sensors operating at 1.5 µm have been demonstrated, largely because of the superior quality and broad availability of components in the C-band. For example, optical fiber including fiber Bragg gratings (FBGs) (Lin et al., 2015), long-period fiber gratings (LPGs) (Viegas et al., 2009; Fu et al., 2020), Sagnac interferometers (SI) (Chen et al., 2018), and photonic crystal fibers (PCF) (Luo et al., 2021). However, some of the technologies have inherent drawbacks. Standard FBGs etched with UV lasers may exhibit reduced reflectivity and increased brittleness at high temperatures, especially above 300 °C (Wu et al., 2015; Wang et al., 2022). The Michelson interferometer with misaligned fibre requires additional adjustments or calibration during construction, which increases complexity (Cao et al., 2017), in contrast to the Fabry-Perot interferometer with air cavities, which may be insensitive to temperature variations (Lu et al., 2023).

Furthermore, new developments in the production of fibre Bragg gratings using femtosecond laser pulses have removed several constraints. Femtosecond laser-induced fibre Bragg gratings are more heat-resistant and more stable compared to their UV-inscribed equivalents. Based on the research conducted by Grobnic et al. (2006), it is possible to permanently alter the refractive index by a factor of up to 10^{-3}. Grobnic et al. (2006) have found that Type-II FBGs made of SMF-28 fibres may resist

Authors information:
[a]Faculty of Applied Sciences and Technology (FAST), Universiti Tun Hussein Onn Malaysia (UTHM), Pagoh Branch Campus, Pagoh Higher Education Hub, Km 1, Jalan Panchor, 84600 Muar, Johor Darul Ta'zim, MALAYSIA.
E-mail: norazura@uthm.edu.my[1]
[b]School of Physics, Universiti Sains Malaysia, 11800 USM, Penang, MALAYSIA.
Email: niknoorharyatuleleena@gmail.com[2]

*Corresponding Author: norazura@uthm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

temperatures as high as 1000 °C (Grobnic et al., 2006). Therefore, the use of femtosecond pulses offers a reliable solution for high-temperature sensing, thereby expanding their usefulness in very demanding conditions.

In addition to FBGs, other fiber-optic-based temperature sensors have also been studied. A biconical fiber temperature sensor has previously been investigated, which had temperature sensitivities of 0.010 and 0.0067 nm °C^{-1} for tapered lengths of 20.548 and 18.382 mm, respectively (Kieu & Mansuripur, 2006). However, these geometries have some disadvantages, including susceptibility to low temperatures and the need for complex processing techniques. In contrast, side-polished fibers (SPF) are among the most attractive alternatives because their cores are exposed, enabling strong evanescent-field interaction with external media. Because of this kind of interaction, SPF structures are very responsive to changes in the external refractive index and temperature. Previous work has shown their efficiency in improving the performance of relative humidity sensors (Huang et al., 2018), refractive index, and temperature (Teng et al., 2022). Their versatility suggests the possibility of enhancing sensing applications involving temperature and the environment using a simpler geometry than configurations such as FBGs.

In addition, SPF offers several advantages over etched, tapered, or LPFG structures, including low attenuation losses, cost-effectiveness, ease of fabrication, and mechanical robustness (Arjmand et al., 2017). To perform MDFL, the SPF has been coated with metal nanoparticles such as gold nanoparticles (NP) (Mahmud et. al., 2023). Since gold-NP shows a higher value of the third-order non-linear coefficient than that of other nanomaterials, it is a suitable material to be used as an SA for pulse generation (Liao et al., 1997).

In this study, we constructed a mode-locked fiber sensor (MLFS) incorporated into an EDFL cavity to produce pulses for temperature measurement. Compared with the femtosecond laser-inscribed FBG, this solution does not require an expensive ultrafast inscription system while still offering high temperature sensitivity and strong signal-to-noise performance enabled by mode-locking. The entire laser cavity consists of 0.5 m of EDF with an overall length of 2.81 m. The gold-NP SPF serves simultaneously as a saturable absorber and a temperature-sensitive element with sensitivities of 0.4079 (cooling) and 0.3775 nm °C^{-1} (heating). This simple, low-cost arrangement demonstrates the capability of SPF-based structures for scalable, practical optical fiber temperature sensors.

2. Experimental setup

2.1 Side-Polished Fiber-Coated Gold

The fabrication of gold thin film is shown in Figure 1 by using the Quorum DC magnetron sputtering technique. To begin the procedure, the chamber was filled with argon gas to achieve a complete vacuum. This process enabled the ionized Argon gas atoms to collide with the gold target, therefore releasing the particles into the existing plasma. Thus, the negatively charged gold atoms are moved into a positively charged substrate and then condensed, forming a thin layer. Before starting the deposition process, the density was set at 19.32 $gm^{-3}$with a thickness of 10 nm. The distances of core and cladding after deposition were 93.239 and 245.811 µm.

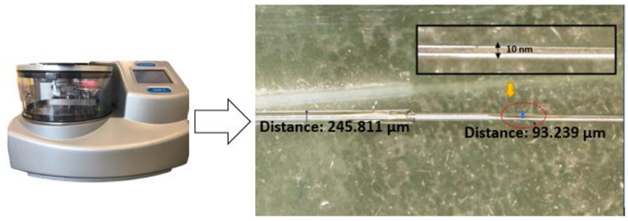


Figure 1. SPF coated with gold using a Quorum DC magnetron sputter coater.

2.2 Temperature Sensor

Figure 2 shows a single-ring cavity that includes 32.12 dBm^{-1}absorption coefficient of 0.5 m EDF with overall length cavity 2.81 m (Mahmud et al., 2022). This setup operates with a 980 nm laser diode (LD). To prevent unidirectional light transmission within the laser cavity, an isolator was used in the setup. A wavelength-division multiplexer (WDM) was used to combine and separate light into distinct wavelengths. The laser output was extracted through a 10% fiber coupler, while the remaining 90% circulated within the cavity. To generate MDFL, a gold-NP SPF was integrated in the setup, and 50% of the laser output was monitored by an optical spectrum analyser (OSA) and an oscilloscope (OSCI). The gold-NP SPF, which served as the sensor, was submerged in water, and a thermometer was inserted to measure the temperature. An MFLS was evaluated throughout both the heating and cooling cycles. The water was heated to 100 °C, and after the measurements were taken, the hot plate was turned off to initiate cooling. During heating, the MLFS output pulse varied over the temperature range of 0 to 100 °C; during cooling, it varied from 100 to 0 °C.

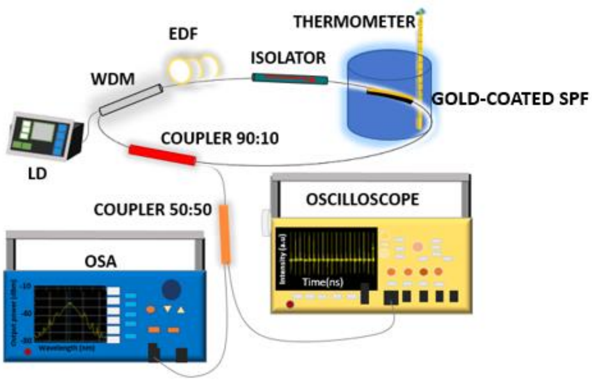


Figure 2. Configuration of mode-locked fiber sensor (MLFS).

DOI: https://doi.org/10.22452/mjs.vol45sp1.5
Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026)

3. Results and analysis

An initial measurement of the optical spectrum of the MFLS at room temperature is shown in Figure 3(a). With a pump power of 297.3 mW, the MLFS optical spectrum has a centre wavelength of 1563.1 nm and a 3-dB bandwidth of 11.2 nm. The presence of Kelly sidebands in the output spectrum was attributed to the abnormal dispersion occurring in the ring cavity (Sun et al., 2010). These sidebands arise from the periodic interference between soliton pulses and dispersive waves circulating in the cavity, confirming that pulse shaping is governed by soliton dynamics. Figure 3 (b) depicts a pulse train of the MLFS operating at a frequency of 38.5 MHz. The pulse interval was determined to be about 0.0258 µs, matching the cavity's round-trip duration exactly. Hence, verifying the mode-locked laser's basic operation.

As shown in Figure 3(c), the pulse width of the MFLS was determined by analyzing the autocorrelation trace and fitting it with a $sech^2$ profile. The blue curves represent the measured autocorrelation (ACF) signal, while the red curves refer to the theoretical fit based on the $sech^2$ temporal profile. From the fitting result, the full width at half maximum (FWHM) of the autocorrelation trace was measured to be 1.149 ps. Assuming the $sech^2$ pulse shape, a deconvolution factor of 1.543 was applied, yielding an estimated pulse duration of 0.691 ps ≈ 691 fs. The MSE of 5.313×10^{-5} indicated excellent agreement between the experimental data and the fitted curve, and hence the generated pulse closely follows a $sech^2$ profile. Furthermore, the narrow autocorrelation width and good fit quality confirm that stable ultrashort pulses were produced by the mode-locked fiber laser.

Figure 3(d) shows that the OSCI captured the RF spectrum of the MLFS, with a 500 MHz scanning span. The inserted image shows that the first-order peak aligns precisely with the laser cavity's repetition frequency, observed at 38.5 MHz. With an SNR of 44.1 dB, this finding indicates that the MLFS is operating steadily. Consistent with the basic cavity repetition rate determined from the OSCI trace, the observed trace generates the fundamental frequency at 3.86 MHz. According to Li et al., the stability and coherence of mode-locked operation were verified by the presence of a clear fundamental peak and the absence of subharmonic components (Li et al., 2023).

After the reference measurements were completed, the water was heated using a hot plate, with measurements taken across the temperature range from 0 to 100 °C. Throughout the experiment, the pump power was maintained at 297.3 mW to ensure consistent results and minimize environmental interference. Figure 4 (a) presents the output spectrum of the MLFS at 0, 40, and 100 °C, as recorded by OSA. At 0 °C, the resonant wavelength was observed at 1530.20 nm. As the temperature increased, the resonant wavelength shifted to longer wavelengths, moving from 1530.20 nm at 0 °C to 1568.70 nm at 100 °C, with a corresponding change in output power from -13.60 to -13.20 dBm. Figure 4(a) shows that 100 °C results in the largest shift among the other temperatures, due to the SPF's thermo-optic behavior (Senosianin et al., 2002). Although the refractive index was not directly measured in this work, temperature-induced changes in the refractive index of the exposed core region are well established in the literature, particularly for structures with strong evanescent-field

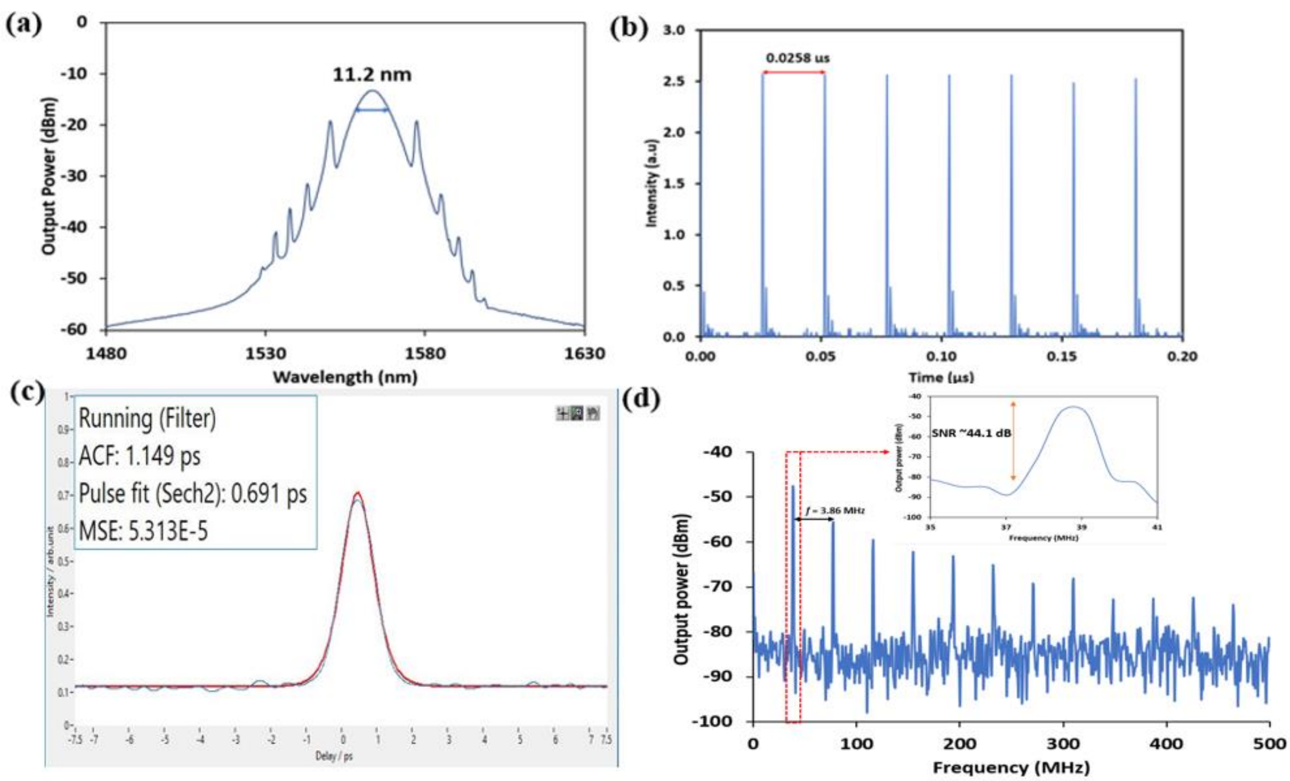


Figure 3. Performance of MLFS (a) Optical spectrum (b) Pulse train (c) Autocorrelation trace (d) Radio-frequency (RF) spectrum

DOI: https://doi.org/10.22452/mjs.vol45sp1.5
Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026)

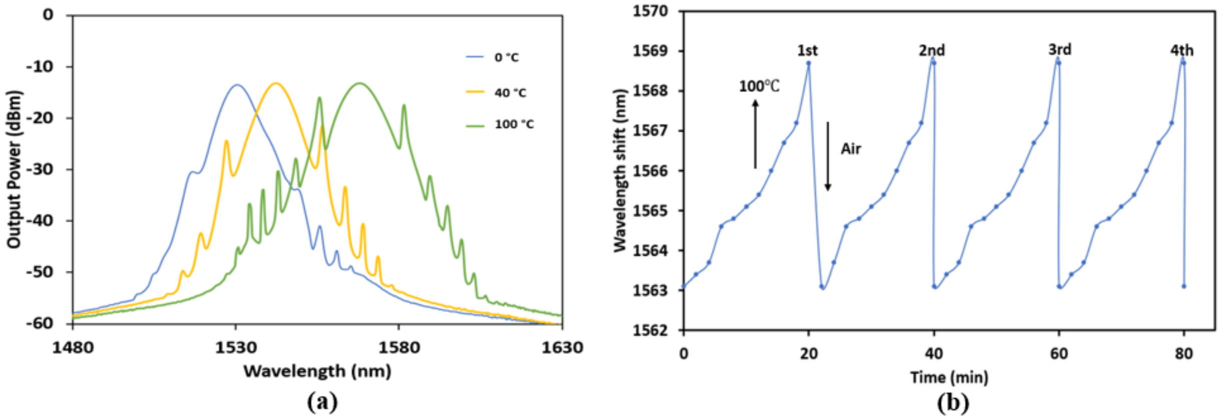


Figure 4. (a) Output spectrum of the MLFS varies at temperatures 0, 40, and 100 °C. (b) Repeatability of wavelength shifting at 100°C temperature.

interactions. Such a thermo-optic effect alters the optical path length of SPF, resulting in a red shift of the lasing wavelength at high temperatures (Zainuddin et al., 2019). Otherwise, temperature-dependent effects have been observed in plasmonic absorption and the non-linear response of the gold-NP coating (Ganeev, 2019). These changes further affect the EDF laser's intracavity gain-loss balance, leading to spectrum broadening and intensity fluctuations in the experiment, which explain the significant temperature-dependent wavelength shifts observed in this work (Jachpure & Vijaya 2024).

Additionally, the gold-coated SPF was monitored at 100 °C (indicated by the upward arrow) for 4 cycles, as illustrated in Figure 4(b). The findings reveal that wavelength shifting from 1563.1 to 1568.7 nm with a tuning range of 5.6 nm. After 20 minutes, the wavelength shift stabilized, and upon exposure to air (arrow downward), the wavelength returned to its reference value within 10 minutes. This behavior was consistent across multiple cycles, even at other temperatures, demonstrating that the sensor has good stability and a fast recovery time in air.

During heating, the OSA obtained the radio-frequency (RF) spectrum of the MLFS, as shown in Figure 5(a). The power and frequency of the MLFS's output are shown in the graph as a function of temperature. The frequency remains constant at 38.5 MHz, whereas the output power drops from -44.74 dBm to -57.21 dB as the temperature increases. At 100 °C, the SNR is 56.30 dB, up from 43.84 dB at room temperature, suggesting a higher SNR even with the pump power unchanged. This increase in SNR is due to the laser cavity being thermally stabilized and to the gold-coated SPF, which introduces improved resonance characteristics.

The temperature was lowered from 100 to 0 °C in 20 °C increments by adding ice cubes throughout the cooling process. Figure 5(b) shows that the RF spectrum produces an output power of -50.83 dBm at 100 °C and gradually increases to -44.76 dBm at 0 °C while the MLFS operates at a cooling temperature. The rate of repetition remains the same throughout the procedure. Although thermal factors throughout the experiment play a

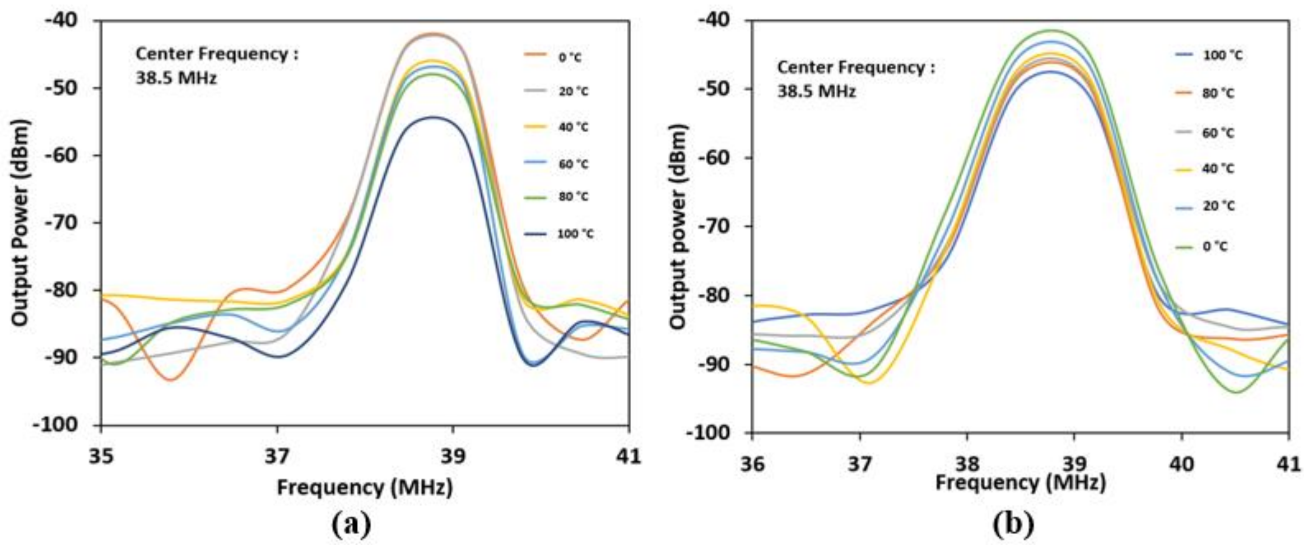


Figure 5. The graph between output power and frequency against temperature for (a) heating and (b) cooling processes.

DOI: https://doi.org/10.22452/mjs.vol45sp1.5
Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026)

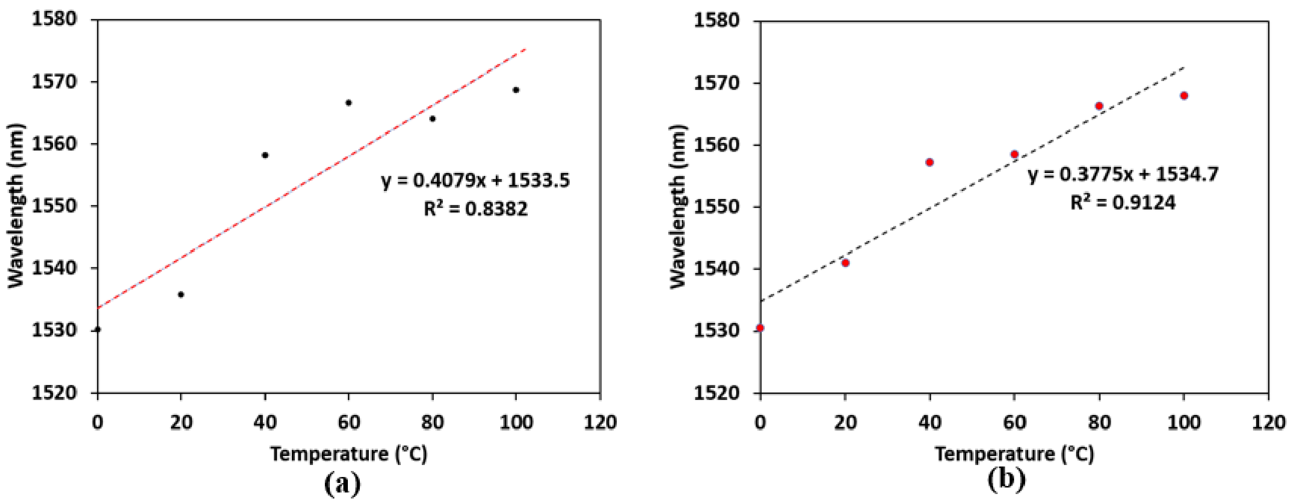


Figure 6. a) Wavelength shift of the MFLS during (a) heating and (b) cooling processes.

secondary role in determining SNR stability, the SNR declined from 50.00 to 43.90 dB during cooling, suggesting that temperature fluctuations had little effect on SNR (Fu et al., 2020). The gold-NP SPF contributes to the higher SNR observed at higher temperatures. According to Chou et al. (2019), this material has superior performance compared to others when it comes to wavelength tuning range, resonance augmentation, and sensitivity (Chou et al., 2019). Incorporating a gold-NP SPF into the laser ring cavity thus significantly enhances the MFLS SNR. To evaluate the sensor's functionality and performance in harsh environments, further research over a wide temperature range is required, even though the present findings demonstrate that the SNR remains generally constant across the observed range.

The data are fit to a linear relationship between the resonant wavelength shift and temperature variation, as shown in Figure 6(a). The graph shows that as temperature increases, the wavelength shifts toward longer wavelengths (Sharma & Gupta, 2006; Bai et al., 2019). This is due to the thermal effect, which causes the resonance wavelength to shift during heating (Chou et al., 2019). Therefore, the wavelength shifts to longer wavelengths during heating. At temperatures higher than 60 °C, there is a non-linear shift in wavelength that is caused by the interplay between the exposed core area of the SPF, the plasmonic and non-linear absorption properties of the gold nanoparticle coating (Ganeez 2019), and the EDF's temperature-dependent gain profile (Bolshtyansky et al., 2000). These factors cause wavelength changes to occur more rapidly at higher temperatures, unlike the usual simple linear thermo-optic behavior observed in FBG-based sensors.

During cooling, the MLFS temperature was slowly reduced to room temperature, then reached 0 °C after some ice cubes were added. In Figure 6(b), the resonant wavelength shifts toward shorter wavelengths as the temperature decreases from 100 to 0 °C. During the cooling process, the temperature decreased, and the resonant wavelength shifted from 1568.00 to 1530.55 nm, so the wavelength shifted by around < 10 nm. Based on Figure 6 (a-b), the temperature sensitivity of the heating and cooling conditions was about 0.4079 and 0.3775 $nm\,°C^{-1}$, indicating a hysteresis of approximately 8% (Lin et al., 2020). This level of hysteresis is considered significant because an ideal temperature sensor should exhibit responses that are almost identical during heating and cooling. The observed divergence suggests that the system does not fully return to its original spectral position after each heat cycle. This conclusion is most likely due to the gold-NP SPF expanding at different rates during heating, potentially resulting in residual mechanical stress at the interface. Each condition yields R^2 value of 0.8382 and 0.9124.

4. Conclusion

In this paper, the successful demonstration of an MLFS coated with gold-NP is presented, with its temperature characteristics thoroughly examined. The sensor exhibits temperature sensitivities of 0.4079 $nm\,°C^{-1}$during heating and 0.3775 $nm\,°C^{-1}$ during cooling, the presence of thermal hysteresis. As the temperature rises, the resonant wavelength shifts toward longer wavelengths, and the experimental data confirm a strong linear correlation between the wavelength shift and ambient temperature fluctuations. In addition, the temperature effect on output power and SNR was examined, showing that an increase in SNR due to the enhanced sensitivity of the gold-NP SPF, which improves the sensor's performance. Thus, the proposed sensor has several merits, such as ease of use, a simple fabrication process, and applicability for temperature monitoring across various applications. Future research will focus on reducing the effects of thermal hysteresis through material optimization and improved manufacturing techniques to provide long-term stability and reliability in practical applications.

5. Acknowledgement

The Ministry of Higher Education (MOHE) funded this study under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/WAB05/UTHM/02/5) (grant No: K170).

6. References

Arjmand, M., Saghafifar, H., Alijanianzadeh, M., & Soltanolkotabi, M. (2017). A sensitive tapered-fiber optic biosensor for the label-free detection of organophosphate pesticides. *Sensors and Actuators B: Chemical*, 249, 523-532.

Bai, Y., Yan, F., Feng, T., Han, W., Zhang, L., Cheng, D., ... & Wen, X. (2019). Temperature fiber sensor based on single longitudinal mode fiber laser in 2 μm band with Sagnac interferometer. *Optical Fiber Technology*, 51, 71-76.

Bolshtyansky, M., Wysocki, P., & Conti, N. (2000). Model of temperature dependence for gain shape of erbium-doped fiber amplifier. *Journal of lightwave technology*, *18*(11), 1533.

Cao, H., & Shu, X. (2017). Miniature all-fiber high temperature sensor based on Michelson interferometer formed with a novel core-mismatching fiber joint. *IEEE Sensors Journa*l, 17(11), 3341-3345.

Chen, F., Jiang, Y., Zhang, L., Jiang, L., & Wang, S. (2018). Fiber optic refractive index and magnetic field sensors based on microhole-induced inline Mach–Zehnder interferometers. *Measurement Science and Technology*, 29(4), 045103.

Chou, Y. L., Wu, C. W., Jhang, R. T., & Chiang, C. C. (2019). A novel optical fiber temperature sensor with polymer-metal alternating structure. *Optics & Laser Technology*, 115, 186-192.

Fried, N. M. (2005). Thulium fiber laser lithotripsy: An in vitro analysis of stone fragmentation using a modulated 110-watt Thulium fiber laser at 1.94 μm. *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery*, 37(1), 53-58.

Fu, X., Zhang, Y., Wang, Y., Fu, G., Jin, W., & Bi, W. (2020). A temperature sensor based on tapered few mode fiber long-period grating induced by CO2 laser and fusion tapering. *Optics & Laser Technology*, 121, 105825.

Ganeev, R. A. (2019). Characterization of the optical nonlinearities of silver and gold nanoparticles. *Optics and Spectroscopy*, *127*(3), 487-507.

Gonzalez-Reyna, M. A., Alvarado-Mendez, E., Estudillo-Ayala, J. M., Vargas-Rodriguez, E., Sosa-Morales, M. E., Sierra-Hernandez, J. M., & Rojas-Laguna, R. (2015). Laser temperature sensor based on a fiber Bragg grating. *IEEE Photonics Technology Letters*, 27(11), 1141-1144.

Grobnic, D., Smelser, C. W., Mihailov, S. J., & Walker, R. B. (2006). Long-term thermal stability tests at 1000 C of silica fibre Bragg gratings made with ultrafast laser radiation. *Measurement Science and Technology*, 17(5), 1009.

Huang, Y., Zhu, W., Li, Z., Chen, G., Chen, L., Zhou, J., & Yu, J. (2018). High-performance fibre-optic humidity sensor based on a side-polished fibre wavelength selectively coupled with graphene oxide film. *Sensors and actuators B: chemical*, 255, 57-69.

Jachpure, D., Prakash, O., & Vijaya, R. (2024). Saturable absorption in erbium-doped fiber for controlling lasing peaks and their linewidths in fiber laser. *Optical Fiber Technology*, *88*, 104022.

Jung, M., Lee, J., Koo, J., Park, J., Song, Y. W., Lee, K., & Lee, J. H. (2014). A femtosecond pulse fiber laser at 1935 nm using a bulk-structured Bi 2 Te 3 topological insulator. *Optics express*, 22(7), 7865-7874.

Kieu, K. Q., & Mansuripur, M. (2006). Biconical fiber taper sensors. *IEEE photonics technology letters*, 18(21), 2239-2241.

Keller, U. (2000, September). Ultrafast solid-state lasers. In The European Conference on Lasers and Electro-Optics (p. CMB3). *Optica Publishing Group*.

Li, X., Xu, W., Wang, Y., Zhang, X., Hui, Z., Zhang, H., & Al-Sehemi, A. G. (2022). Optical-intensity modulators with PbTe thermoelectric nanopowders for ultrafast photonics. *Applied Materials Today*, 28, 101546.

Li, J., Wu, J., Hao, Q., Huang, K., Yang, K., & Zeng, H. (2023). Wavelength and repetition rate tunable high peak power dissipative soliton resonance in an all polarization maintaining Yb-doped fiber laser. *Optics & Laser Technology*, 162, 109204.

Lu, L., Liang, Z., Wu, L., Chen, Y., Song, Y., Dhanabalan, S. C., & Zhang, H. (2018). Few-layer bismuthene: sonochemical exfoliation, non-linear optics and applications for ultrafast photonics with enhanced stability. *Laser & Photonics Reviews*, 12(1), 1700221.

Lu, Z., Liu, C., Li, C., Ren, J., & Yang, L. (2023). Ultra-High Sensitivity and Temperature-Insensitive Optical Fiber Strain Sensor Based on Dual Air Cavities. *Materials*, 16(8), 3165.

Lin, Y., Gong, Y., Wu, Y., & Wu, H. (2015). Polyimide-coated fiber Bragg grating for relative humidity sensing. *Photonic sensors*, 5, 60-66.

Lin, Q., Hu, Y., Yan, F., Hu, S., Chen, Y., Liu, G., & Chen, Z. (2020). Half-side gold-coated hetero-core fiber for highly sensitive measurement of a vector magnetic field. *Optics Letters*, 45(17), 4746-4749.

Luo, W., Li, X., Meng, J., Wang, Y., & Hong, X. (2021). Surface plasmon resonance sensor based on side-polished D-shaped photonic crystal fiber with split cladding air holes. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-11.

Liao, H. B., Xiao, R. F., Fu, J. S., Yu, P., Wong, G. K. L., & Sheng, P. (1997). Large third-order optical nonlinearity in Au: Si O 2 composite films near the percolation threshold. *Applied Physics Letters*, 70(1), 1-3.

Mądry, M., Alwis, L., Binetti, L., Pajewski, Ł., & Bereś-Pawlik, E. (2019). Simultaneous measurement of temperature and relative humidity using a dual-wavelength erbium-doped fiber ring laser sensor. *IEEE Sensors Journal*, 19(20), 9215-9220.

Matsas, V. J., Newson, T. P., Richardson, D. J., & Payne, D. N. (1992). Self-starting, passively mode-locked fibre ring soliton

DOI: https://doi.org/10.22452/mjs.vol45sp1.5
Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026)

laser exploiting non-linear polarisation rotation. *Electronics Letters*, 28(15), 1391-1393.

Mahmud, N. N. H. E. N., Awang, N. A., & Zulkefli, N. U. H. H. (2022). Supercontinuum generation of gold coated side-polished fiber based-mode-locked pulse. *Optik*, 260, 169074.

Mahmud, N. N. H. E. N., Awang, N. A., Zalkepali, N. U. H. H., Latif, A. A., Muhammad, N. A. M., & Zamri, A. Z. M. (2023). The influence of Au-NP thickness coated on side-polished fiber on the properties of mode-locked erbium-doped fiber laser. *Infrared Physics & Technology*, 130, 104616.

Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Katsnelson, M. I., Grigorieva, I. V., & Firsov, A. A. (2005). Two-dimensional gas of massless Dirac fermions in graphene. *Nature*, 438(7065), 197-200.M.

Senosiain, J., Díaz, I., Gastón, A., & Sevilla, J. (2002). High sensitivity temperature sensor based on side-polished optical fiber. *IEEE Transactions on Instrumentation and measurement*, 50(6), 1656-1660.

Sun, Z., Hasan, T., Torrisi, F., Popa, D., Privitera, G., Wang, F., & Ferrari, A. C. (2010). Graphene mode-locked ultrafast laser. *ACS nano*, 4(2), 803-810.

Song, H., Wang, Q., Zhang, Y., & Li, L. (2017). Mode-locked ytterbium-doped all-fiber lasers based on few-layer black phosphorus saturable absorbers. *Optics Communications*, 394, 157-160.

Sharma, A. K., & Gupta, B. D. (2006). Influence of temperature on the sensitivity and signal-to-noise ratio of a fiber-optic surface-plasmon resonance sensor. *Applied optics*, 45(1), 151-161.

Stewart, G., Culshaw, B., Johnstone, W., Whitenett, G., Atherton, K., & McLean, A. (2003). Optical fibre sensors and networks for environmental monitoring. *Management of Environmental Quality: An International Journal*, 14(2), 181-190.

Teng, C., Shao, P., Li, S., Li, S., Liu, H., Deng, H., & Deng, S. (2022). Double-side polished U-shape plastic optical fiber based SPR sensor for the simultaneous measurement of refractive index and temperature. *Optics communications*, 525, 128844.

Trushin, M., Kelleher, E. J., & Hasan, T. (2016). Theory of edge-state optical absorption in two-dimensional transition metal dichalcogenide flakes. *Physical Review B*, 94(15), 155301.

Viegas, D., Goicoechea, J., Santos, J. L., Araújo, F. M., Ferreira, L. A., Arregui, F. J., & Matias, I. R. (2009). Sensitivity improvement of a humidity sensor based on silica nanospheres on a long-period fiber grating. *Sensors*, 9(1), 519-527.

Wu, W., & Liu, X. (2015). Investigation on high temperature characteristics of FBG sensors. *Optik*, 126(20), 2411-2413.

Wang, X., Sun, X., Hu, Y., Zeng, L., Liu, Q., & Duan, J. A. (2022). Highly-sensitive fiber Bragg grating temperature sensors with metallic coatings. *Optik*, 262, 169337.

Zhou, H. S., Honma, I., Komiyama, H., & Haus, J. W. (1994). Controlled synthesis and quantum-size effect in gold-coated nanoparticles. *Physical Review B*, 50(16), 12052.

Zhang, Y., Jin, W., Yu, H. B., Zhang, M., Liao, Y. B., Ho, H. L., & Li, Y. H. (2002). Novel intracavity sensing network based on mode-locked fiber laser. *IEEE Photonics Technology Letters*, 14(9), 1336-1338.

Zainuddin, N. A. A. M., Ariannejad, M. M., Arasu, P. T., Harun, S. W., & Zakaria, R. (2019). Investigation of cladding thicknesses on silver SPR based side-polished optical fiber refractive-index sensor. *Results in Physics*, 13, 102255.

DOI: https://doi.org/10.22452/mjs.vol45sp1.5
Malaysian Journal of Science 45 (Special Issue 1): 37-43 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

https://mjs.um.edu.my

Predictive Lead Scoring Models: A Contemporary Review of Identified Gaps and Future Directions

Jia Yee Lee[1a]*, Chi Wee Tan[2a], Chuk Fong Ho[3b], and Noor Aida Husaini[4a]

Abstract: Lead scoring plays a pivotal role in modern marketing strategies, aiding businesses in prioritising and optimising their sales efforts. This review paper provides a contemporary analysis of lead scoring models, focusing on their development, application, and effectiveness across Machine Learning algorithms. The paper begins by introducing the concept and evolution of lead scoring, highlighting the shift in how previous researchers developed lead scoring models, from traditional Machine Learning algorithms to more advanced approaches, such as Ensemble Learning, Neural Networks, and Deep Learning algorithms. It also addresses existing gaps in current models and clarifies the misconceptions about the definition of "lead scoring". In addition, the paper reviews the methodologies used in lead scoring, including data collection, feature selection, and the algorithms applied. Finally, it examines emerging trends and challenges in lead scoring, offering an insightful overview of the current landscape and suggesting future research directions to unlock new opportunities for businesses to enhance their marketing strategies.

Keywords: Lead Scoring, Machine Learning, Artificial Neural Networks, Deep Learning, Social Media

1. Introduction

In today's highly competitive business environment, effectively managing and prioritising potential customers is essential for maximising sales efficiency and refining marketing strategies (Morgado, 2018). Lead scoring, a systematic approach to ranking prospects based on their likelihood of conversion (Rodrigues, 2020), has become crucial for businesses aiming to optimise their sales processes and resource allocation (Ceccatelli, 2023; Lindahl et al., 2017; K. K. Sharma et al., 2023; Sundman, 2021).

Traditionally, manual lead scoring relied heavily on heuristic methods, drawing on the intuition and experience of sales teams (Duncan & Elkan, 2015; Nygård & Mezei, 2020). However, with the advent of big data and advancements in Machine Learning (ML), lead scoring has evolved from a traditional approach to a predictive one. Modern predictive models yield more accurate and actionable insights beyond human capability (Wu et al., 2024) by leveraging large customer data and advanced ML algorithms, such as Neural Networks (NN) (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Espadinha-Cruz et al., 2021; Nygård & Mezei, 2020), Deep Neural Networks (DNN) (Chaudhuri et al., 2021), Feedforward Neural Networks (FNN) (Puravankara & Narendra Babu, 2020), and Recurrent Neural Network (RNN) (Giorcelli, 2019). Thus, this review aims to synthesise recent advancements and provide a road map for future research and practical applications.

Specifically, the review paper aims to achieve four key objectives. First, it seeks to classify and compare ML algorithms and feature engineering techniques used in building lead scoring models. Second, it analyses the publicly available "X Online Education" dataset to assess its applicability in this domain. Third, it evaluates the performance metrics and limitations of current lead scoring models. Finally, it identifies emerging trends and proposes actionable future research directions to advance this field.

The paper is organised as follows: Section 2 provides an overview of lead scoring models, discussing the concepts and data used in building these models. Section 3 introduces fundamental concepts of ML algorithms, which are categorised into traditional and advanced ML types. Section 4 offers an introduction to feature engineering techniques. Section 5 analyses patterns and findings from existing lead scoring-related work, focusing on trends in ML types and algorithms, insights from the "X Online Education" dataset, and current applications of performance metrics and feature engineering methods. Section 6 examines the

Authors information:
[a]Department of Computer Science and Data Science Faculty of Computing and Information Technology (FOCS), Tunku Abdul Rahman University of Management and Technology (TAR UMT), Jalan Genting Kelang, Setapak, 53300 Wilayah Persekutuan Kuala Lumpur, MALAYSIA. E-mail: jylee-wm19@student.tarc.edu.my[1], chiwee@tarc.edu.my[2], nooraida@tarc.edu.my[4]
[b]Department of Software Engineering and Technology, Faculty of Computing and Information Technology (FOCS), Tunku Abdul Rahman University of Management and Technology (TAR UMT), Jalan Genting Kelang, Setapak, 53300 Wilayah Persekutuan Kuala Lumpur, MALAYSIA. E-mail: cfho@tarc.edu.my[3]

*Corresponding Author: jylee-wm19@student.tarc.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

challenges in the field and suggests potential future research directions to guide upcoming studies. Finally, Section 7 presents the conclusions of the review.

2. Lead scoring models

2.1 Lead

A lead, also referred to as a business lead or sales lead, is an individual who has shown interest in an organisation's brand, products, or services (Waters, 2022). Sales or marketing teams often target leads as potential customers, as they are the contact persons that the teams aim to pitch to (Fahad, 2017). Leads can be ascertained through various channels, including emails, phone calls, search engine behaviours, in-person visits, forms, or surveys (Niemi, 2022).

2.2 Traditional Lead Scoring

Once leads are collected, scores that indicate whether they are potential customers or not are calculated and assigned to them (Benhaddou & Leray, 2017). Manual lead scoring, also known as traditional lead scoring, is the process of identifying potential leads based on certain criteria determined by an organisation's sales or marketing department, with the usage of a scoring matrix (Ayaz, 2023). A scoring matrix is a tool used to evaluate and rank potential leads based on criteria, such as demographic information and behavioural data.

Figure 1 provides an example of how a scoring matrix is constructed, which may vary between organisations. Each criterion is assigned a weight, and leads are scored based on how well they meet these criteria, with points allocated for specific attributes. This results in a total score that helps rank the leads in order of priority (Nygård & Mezei, 2020). Typically, lead scores categorise prospects as hot (high scores, immediate follow-up), warm (moderate scores, needs nurturing), or cold (low scores, low priority).

Activity	Points
Form/Landing Page Submission	+ 5
Submitted "Contact Me" Form	+25
Received an Email	0
Email Open	+1
Email Clickthrough	+3
Registered for Webinar	+3
Attended Webinar	+10
Downloaded a Document	+5
Visited a Landing Page	+2
Unsubscribed from Newsletter	-2
Watched a Demo	+8
Contact is a CXO	+5
Visited Trade Show Booth	+3
Visited Pricing Page	+10

Figure 1. Scoring Matrix (Autopilot, 2016).

However, manual lead scoring has its limitations. First, it risks missing predictive signals of purchasing behaviour since data is often captured by salesmen or marketers who may not always record all relevant interactions or insights. For instance, a lead might repeatedly view pricing pages or inquire during sales calls, which are not likely to be recorded in manual systems. Another example occurs when salesmen or marketers prioritise capturing immediate customer responses and overlook subtle behavioural cues, such as browsing patterns or social media interactions, which can be crucial for accurate lead scoring. This incomplete data can result in a skewed understanding of a lead's true potential, potentially leading to less effective targeting and missed opportunities. Second, manual lead scoring is time-intensive and cannot process the large data volumes needed for statistically significant insights (Pereira, 2021). Thus, these limitations have driven the adoption of predictive lead scoring, which is achieved by using ML algorithms to automate the lead scoring process (Etminan, 2021).

2.3 Modern Lead Scoring

Figure 2 illustrates the process of automated lead scoring.

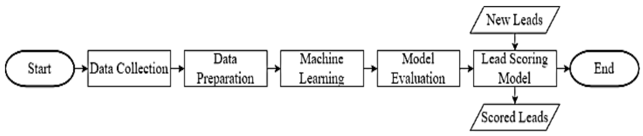


Figure 2. Process of Automated Lead Scoring.

The initial phase involves data collection, where labelled data is gathered specifically for training the predictive model. This is different from the typical data collection done in real-world situations, which often involves unlabelled data intended for lead conversion purposes. The second phase is data preparation, which includes data cleaning, transformation, and feature selection to ensure the data is suitable for modelling. The third phase involves implementing ML algorithms to learn lead conversion patterns from the training data. If available, validation data is used in this phase to fine-tune the model's parameters, ensuring that the model neither overfits nor underfits the training data. In the fourth phase, model evaluation is conducted to assess how well the model generalises to unseen data by using test data that was not included during training or validation. Ultimately, a lead scoring model is built with the capability to predict the likelihood of lead conversion by assigning scores to new leads. Ultimately, a lead scoring model is built with the capability to predict the likelihood of lead conversion by assigning scores to new leads.

2.4 Lead Scoring Datasets

The data used in predictive lead modelling can be categorised into two (2) main types: social media data and non-social media data. Table 1 summarises the characteristics of the fourteen datasets utilised across sixteen studies, with some studies employing the same datasets.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Table 1. Dataset Characteristics Table.

Characteristics	Data Type	Description	Related Work
Structured, Research records	Non-social media	Public Dataset entitled "X Online Education" obtained from websites, forms and past referrals.	(Jadli et al., 2022; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024)
Structured, Sales and Marketing records	Non-social media	Sales and Marketing data stored in Salesforce from 2018 to 2019.	(Ayaz, 2023)
Structured, Marketing records	Non-social media	B2B Dataset was collected as proprietary data from a company located in the western United States B2C Dataset was extracted from a France-based company, Criteo AI Lab's website.	(Bhatta, 2022)
Structured, Web browsing records	Non-social media	Data collected from Google Analytics API.	(Etminan, 2021)
Structured, Web browsing records	Non-social media	Web browsing data collected from an online e-commerce platform based in Germany and Belgium.	(Chaudhuri et al., 2021)
Structured, Marketing records	Non-social media	Data from a telecommunications company across channels (telemarketing, shop, website).	(Espadinha-Cruz et al., 2021)
Structured, Marketing records	Non-social media	Data from HUUB's Customer Relationship Management platforms, HubSpot.	(Pereira, 2021)
Structured, Marketing records	Social media and Non-social media	14 months of lead conversion data from different channels.	(Karthik et al., 2020)
Structured, Web browsing records	Non-social media	Real-life data from an international company includes insights on potential leads in Finland gathered from internal systems, website visits, email sends, opens, click-throughs, and form submissions.	(Nygård & Mezei, 2020)
Structured, Sales records	Non-social media	Customer's sales data over the 3 months from the grocery superstore's billing system named "Taradin Super Shop".	(Choudhury & Nur, 2019)
Structured, Web browsing records	Non-social media	An imbalanced dataset of website visitors' behaviour generated from Google Analytics 360 by Greenchoice for the entire year of 2018.	(Swelsen, 2019)
Structured, Form data	Non-social media	Self-reported information input by the user on the lead form.	(Giorcelli, 2019)
Unstructured, News records	Non-social media	News about different companies on the Internet.	(Benhaddou & Leray, 2017)

Based on Table 1, non-social media data refers to data sources that are not derived from social media platforms. This category includes, but is not limited to, data collected from websites, responses gathered through forms or surveys, and records maintained within databases. Most non-social media datasets are structured and consist of records related to research, sales, marketing, web browsing, forms, and news. Notably, only one non-social media dataset in the context of lead scoring work is unstructured, as presented in the work of Benhaddou & Leray (2017), which contains news data about different companies collected from the Internet. Outside the context of lead scoring, numerous examples of unstructured non-social media data exist, including textual clinical notes (Goh et al., 2021), call transcriptions (Vo et al., 2021), and customer feedback (De Haan & Menichelli, 2019).

Social media data, on the other hand, encompasses information generated and collected through social media platforms such as Facebook, Instagram, and Twitter. Surprisingly, the reviewed literature does not show studies that use social media data exclusively. However, one study by Karthik et al. (2020) integrates both social media and non-social media data sources, including platforms such as Facebook and Google, as well as directories and services like Board, info@tds, Justdial, and Sulekha.

To date, social media data has been widely utilised in various domains, including the prediction of future sales and sentiment analysis (Asur & Huberman, 2010; Balfagih, 2016; Kanavos et al., 2023). For instance, a study by Asur & Huberman (2010) used Twitter data to forecast box-office revenues for movies based on tweet rates, which is an example of unstructured textual data. Similarly, a study by Balfagih (2016) imported data from his Facebook friend list to identify potential employees for a marketing team, which presents semi-structured user records. More recently, a study by Kanavos et al. (2023) analysed Twitter users' posts to detect spam content, which is another instance of unstructured textual data. Overall, social media data in other

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

domains tends to appear as unstructured or semi-structured data in the form of textual or user records.

However, using social media data presents several challenges. First, the biased population spectrum is observed across different social media platforms. One review paper clearly mentioned the differences in users' demographics when comparing Twitter, Facebook, Sina Weibo, WeChat, and Instagram (Huang et al., 2022). Thus, choosing the right social media platform as the data source is crucial, as it may impact the performance of lead scoring models. Second, the same review paper also stated the challenge of investigating multilingual data (Huang et al., 2022). In real-world scenarios, social media data often face challenges in contextual interpretation. However, to the extent of current knowledge, the data employed in sixteen lead scoring-related studies focus on only one language.

2.5 The Underexplored Potential of Social Media Data

Table 2 categorises the lead scoring-related papers based on data type (social media or non-social media) and source type (public or private). It distinguishes between these categories with counts for each combination.

Table 2. Dataset Allocation Table.

Data Type	Source Type	Count
Non-social media	Public	2
	Private	11
Social media and Non-social media	Public	0
	Private	1

As indicated in Table 2, fifteen out of sixteen (93.75%) studies in this area rely on non-social media data, and only one (6.25%) previous work by Karthik et al. (2020) utilised a mix of social media and non-social media data. Therefore, to the extent of current knowledge, there is a notable lack of research on lead scoring models utilising social media data. This gap in the literature suggests that current models might be missing out valuable insights derived from social media interactions and user behaviours. There is a need for more comprehensive studies to explore how incorporating social media data can enhance the accuracy and effectiveness of lead scoring models. Addressing this gap is essential for advancing the field, as integrating social media data could improve model performance and offer a more comprehensive understanding of lead-scoring dynamics.

In addition, only two out of fourteen (14.29%) datasets are publicly available. This gap highlights the challenge that researchers face due to the lack of accessible public datasets for experimental purposes in the context of lead scoring models. Moreover, this limitation complicates the assessment of a model's generalisation as it restricts the ability to test and validate models across diverse datasets.

3. Machine learning algorithms

ML is a subset of Artificial Intelligence that encompasses a vast array of algorithms and methodologies which are designed to enable computers to learn from data and make predictions or decisions (Helm et al., 2020). ML algorithms can be classified into two (2) categories: traditional ML algorithms and advanced ML algorithms. Traditional ML algorithms are simpler, more interpretable, and rely on feature engineering, including Logistic Regression (LR) (Fernandes et al., 2021), Naive Bayes (NB) (Berrar, 2018), Support Vector Machine (SVM) (Pisner & Schnyer, 2020), Decision Tree (DT) (Song & Lu, 2015), Ensemble Learning methods (Mienye & Sun, 2022) like Random Forest (RF) (A. B. Shaik & Srinivasan, 2019) and Gradient Boosting (GB) (Bentéjac et al., 2021). In contrast, advanced ML algorithms, such as Neural Networks (NN) (Bouwmans et al., 2019) and Deep Learning (DL) (Alzubaidi et al., 2021), are more complex, less interpretable, and capable of automatic feature learning but require significant computational resources (Thompson et al., 2021).

3.1 Traditional Machine Learning Algorithms

LR is specifically used for binary classification problems, where the outcome is categorical with two possible values (Arista, 2022; Fernandes et al., 2021; Zabor et al., 2022). NB is a classification algorithm that computes the probability of an event A given evidence B based on Bayes' Theorem (Berrar, 2018; Chen et al., 2021; Ismail et al., 2020). For example, calculating the probability of a lead being converted given their behaviour or activity data. SVM strives to identify the optimal hyperplane that maximises the margin between different classes, making it robust to outliers and effective for high-dimensional or non-linear data classification (Kim et al., 2020; Pisner & Schnyer, 2020; Zhu et al., 2023). DT is a highly efficient strategy for data mining, which organises a population into branch-like segments that form an inverted tree structure with a root node, internal nodes, and leaf nodes, thereby establishing a robust classification system (Aldino & Sulistiani, 2020; Lee et al., 2022; Song & Lu, 2015). For instance, DT is effective for organising leads into high- or low-potential groups by using a tree structure. Each branch of the tree represents a decision rule based on lead attributes such as engagement, demographics, and past behaviour. The root node might split leads by engagement level, with branches further dividing them by job title. This hierarchy helps categorise leads into segments, making it easy to identify and prioritise them.

Ensemble Learning is the synergy of two or more ML algorithms, which combines their strengths to achieve superior performance beyond what each algorithm could achieve alone (Mienye & Sun, 2022; Mohammed & Kora, 2023; Zhao et al., 2020). RF is an ensemble method that constructs and intertwines multiple decision trees. RF is capable of handling large datasets and operates without the necessity for cross-validation or data preprocessing techniques such as data rescaling and transformation (Bhardwaj et al., 2022; Mohd Ali et al., 2022; A. B. Shaik & Srinivasan, 2019). GB is another ensemble ML algorithm, which combines several weak prediction models, often simple DTs, to form a stronger model, sequentially training each new model to correct the errors of previous models (Bentéjac et al., 2021; Nie et al., 2021; Sahin, 2020).

3.2 Advanced Machine Learning Algorithms

While traditional ML algorithms are effective for many tasks, advanced ML algorithms have emerged to address complex problems. NN was inspired by the structure and function of the human brain, which consists of interconnected nodes referred to

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

as neurons, and eventually arranged into layers (Bouwmans et al., 2019; Thakur & Konde, 2021; Yang & Wang, 2020). Each neuron receives input, processes it through hidden layers using weights, and produces an output. DL employs a multilayered NN to enable the simultaneous process of learning and classification, effectively addressing the issues of incorrect predictions caused by biased feature selection (Alzubaidi et al., 2021). DL architectures come in a variety of forms, each tailored to specific tasks, including DNN (Srinidhi et al., 2021), FNN (Puravankara & Narendra Babu, 2020), RNN (Shiri et al., 2023), Long Short-Term Memory (LSTM) (DiPietro & Hager, 2020), and Kolmogorov–Arnold Networks (KAN) (Liu et al., 2024).

FNN is a simple and straightforward NN that contains one or multiple hidden layers with direct connections between units and information travelling only in a forward direction (Abdel-Nasser Sharkawy, 2020; Markowska-Kaczmar & Kosturek, 2021; Wang et al., 2023). RNN can effectively handle sequential relationships due to its backward linkages from the output to the input and hidden layers (Kane & Hussain, 2023; Shiri et al., 2023), which has been applied in various civil engineering domains (Mers et al., 2022). LSTM is a subset of RNN designed to address the vanishing gradient problem with the usage of memory units and gating mechanisms to manage long-term dependencies (Nosouhian et al., 2021; Shiri et al., 2023), making it well-suited for healthcare, energy and natural language processing domains (Al-Selwi et al., 2024). DNN is a sophisticated, multilayered NN composed of several hidden layers, including input and output layers (Sarker, 2021; Vijayakumar et al., 2021). It is capable of learning complex patterns from input data without requiring manual feature engineering and has been widely used for time-series prediction and classification tasks (How et al., 2020; Sarker, 2021; Vijayakumar et al., 2021). KAN, inspired by the Kolmogorov-Arnold representation theorem, enhances model flexibility by replacing the fixed linear weights in traditional NN with learnable activation functions, which excel in handling time series forecasting (Hou et al., 2024; Liu et al., 2024).

While there is no direct comparison of the computational costs associated with the relevant studies due to the lack of such data in the cited studies, advanced ML algorithms are often believed to have a longer training time and higher computational costs compared to traditional ML algorithms (Ali et al., 2024).

4. Feature Engineering Techniques

Feature engineering involves preparing datasets to optimise a model's performance by transforming, extracting, and selecting relevant data features (Kumar et al., 2023). Feature transformation refers to the process of converting existing features into a more suitable form for modelling, which includes techniques like normalisation, scaling, and encoding (Das & Spanos, 2022; Zhuang et al., 2020). Feature extraction can be defined as a process of deriving new features from existing ones, utilising methods such as principal component analysis (PCA), latent semantic indexing (LSI), and clustering methods (Suhaidi et al., 2021). For instance, PCA can help to identify the most important features that have the largest covariance with lead outcomes (C. Zhang et al., 2018), LSI works well with textual lead data in a semantic space to capture relationships between terms and documents (Zelikovitz & Hirsh, 2001), and clustering methods group leads with similar characteristics, allowing for more effective segmentation and prioritisation based on lead potential (Bharti & Singh, 2015). Feature selection is the process of reducing data dimensionality by identifying and selecting relevant features in order to minimise overfitting, increase accuracy, and reduce training time (Kumar et al., 2023; T. Shaik et al., 2022).

5. Discussions and analysis

5.1 The Debate on Lead Scoring: Is it Truly Classification?

Table 3 categorises the lead scoring-related papers based on the ML task, distinguishing between classification and regression.

Table 3. ML Task Allocation: Classification or Regression.

ML Task	Class	Count
Classification	Binary	16
	Non-Binary	0
Regression	Continuous Scores	0

Examining Table 3, it is apparent that all studies focus on predicting binary classes using classification techniques. Despite the term lead "scoring", these models often produce binary classification outputs rather than assigning continuous scores (Ayaz, 2023; Benhaddou & Leray, 2017; Bhatta, 2022; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Espadinha-Cruz et al., 2021; Etminan, 2021; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Pereira, 2021; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Swelsen, 2019; Yim et al., 2024). This observed pattern offers a descriptive analysis of current methodological practices, where the reviewed literature consistently uses the term "scoring" to describe binary classification frameworks. However, in its typical mathematical usage, the term "scoring" generally implies continuous measurement.

In fact, lead scoring is typically a classification task to simplify the decision-making process. A possible explanation for using a binary outcome, such as "convert" or "not convert", is that it provides a clear and actionable decision point to prioritise potential leads more effectively. However, reducing lead scoring to a binary outcome may result in the loss of valuable information about how promising each lead truly is. For instance, with continuous scoring, the lead who frequently visits the pricing page could receive a higher score of 85 out of 100, reflecting a stronger likelihood of conversion compared to the lead with fewer interactions, who might receive a score of 60. This more granular approach enables decision-makers to prioritise leads more effectively. Sales teams could allocate more resources to the higher-scoring leads, potentially increasing conversion rates and improving overall sales efficiency. Continuous scoring, therefore, provides more nuanced insights into the likelihood of conversion, leading to more refined decision-making and optimised resource allocation. This discussion highlights the trade-off between simplicity and granularity in lead scoring approaches.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

5.2 Comparative Analysis of "X Online Education" Dataset

By far, the most widely used dataset for lead scoring models is the publicly available "X Online Education" dataset, which was employed in four different research papers (Jadli et al., 2022; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024). This dataset[1] is available on Kaggle, an online community that allows users to find and publish datasets. The dataset pertains to an education company, known as X Education, which offers online courses tailored for industry professionals. The company collects data through multiple channels, including forms filled out by website visitors and past referrals. The collected data is used to classify individuals into two categories: converted (those who enrol in courses) or not converted (those who do not enrol) after being approached by the sales team.

Even when using the same dataset, the findings differ across the four papers. Table 4 presents that two unique works by Jadli et al. (2022) and M. Sharma (2022) reported that RF demonstrated the highest accuracy in predicting lead scoring

Table 4. Comparison of the Accuracy Score between four papers that used the same "X Online Education" dataset. The highest results achieved by each algorithm will be bolded, while the highest results achieved by each study will be marked with an asterisk (*) at the end.

Algorithm	Yim et al. (2024)	M. Sharma (2022)	Jadli et al. (2022)	Puravankara & Narendra Babu (2020)
AdaBoost	-	-	-	**0.929**
Bagging	**0.9216**	-	-	-
Boosting	**0.9116**	-	-	-
CatBoost	-	**0.8502**	-	-
DT	-	0.7911	**0.9147**	-
Feedforward NN	-	-	-	**0.94***
GB	-	-	-	**0.933**
KNN	**0.9197**	-	0.8004	-
LR	0.9207	0.8326	0.8989	**0.923**
NB	**0.9142**	-	0.8018	-
RF	0.9222	0.8528*	**0.9302***	0.925
Stacking	**0.9233***	-	-	-
SVM	**0.9215**	-	0.7322	-
XGBoost	-	0.8517	-	**0.935**

Table 5. Comparison of Feature Engineering Techniques between four papers that used the same "X Online Education" dataset.

Feature Engineering Techniques	Yim et al. (2024)	M. Sharma (2022)	Jadli et al. (2022)	Puravankara & Narendra Babu (2020)
Check duplicates	✓	✓	✓	✓
Drop ID columns (prospect id, lead number)	✓			✓
Replace unselected value to NaN			✓	
Drop columns with null	✓ (drop if more than 40%)	✓ (drop if more than 3,000)	✓ (drop if more than 70%)	✓ (drop if more than 45%)
Replace null value	✓ (data imputation, mode)	✓ (data imputation, mode)	✓ (mode, mean, median)	✓ (mode)
Check correlations	✓			✓
Drop columns with same values		✓ (drop if more than 90%)		
Check data skewness		✓	✓	
Check outliers	✓	✓	✓	✓
Remove outliers	✓	✓		
Transform categorical to numeric	✓ (dummy variable creation)	✓ (one hot encoding)	✓ (one hot encoding)	✓ (dummy variable creation)
Scaling	✓ (standard)	✓ (min max)		✓
Normlisation	✓	✓		
K-fold	✓		✓	
Split	80/20	70/30	70/30	80/20
Feature Elimination	✓ (Recursive Feature Elimination)			

[1] X Online Education dataset link: https://www.kaggle.com/datasets/amolbhone/lead-score-case-study?select=lead+scoring+case+study.pdf

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

outcomes. Meanwhile, a study by Yim et al. (2024) diverged in the finding, identifying that Stacking, one of the Ensemble Learning algorithms, was the most effective. Moreover, a study by Puravankara and Narendra Babu (2020) reveals that FNN performs the best in predicting lead conversion. Overall, the best model, according to performance metrics, was produced by Puravankara and Narendra Babu (2020) with the usage of the FNN algorithm.

The variation in the performance findings highlights the influence of different factors on model performance, such as data preprocessing techniques, specific implementation and fine-tuning of the algorithms. Consistent data preprocessing techniques are crucial, as they help ensure the quality and reliability of the data, thereby reducing bias and enhancing the model's generalisability across different datasets. Looking at Table 5, the work of Puravankara and Narendra Babu (2020) has fewer data preprocessing steps when compared to other works. This study did not replace unselected values with NaN values, drop columns with identical values, check for data skewness, remove outliers, apply normalisation, perform k-fold cross-validation, or conduct feature elimination. It might imply that too many data preprocessing techniques are unnecessary for handling a lead scoring dataset.

For comparison purposes, LR and RF are selected since all four studies implemented these algorithms. Among the studies using both LR and RF algorithms, the work of Puravankara and Narendra Babu (2020) achieved the highest performance, followed by the work of Yim et al. (2024). The common data preprocessing techniques used by these two studies but not by the other two include dropping ID columns, removing columns that have 40% to 45% null values, using mode to replace null value, checking correlations, not checking for data skewness, using dummy variable creation to transform categorical values into numeric values, and splitting the data into training and test sets with a ratio of 80:20. These techniques are likely contributed to their superior results compared to other studies that did not use these methods.

5.2 Current Trend on Predictive Lead Scoring Models

This subsection focuses on trends related to the application of algorithms and performance metrics. It will first discuss the trend between traditional and advanced ML algorithms. Next, the subsection will delve into each algorithm, including both traditional and advanced ones.

Figure 3 illustrates the trends of ML types, including traditional and advanced ML algorithms, from 2013 to 2024. Based on Figure 3, it is evident that traditional ML algorithms play a significant role in the development of lead scoring models and are the preferred choice for researchers. Despite the effectiveness of advanced ML algorithms, traditional ML algorithms remain popular. A great deal of previous research works (Ayaz, 2023; Balfagih, 2016; Bhatta, 2022; Choudhury & Nur, 2019; D'Haen & Van Den Poel, 2013; Espadinha-Cruz et al., 2021; Etminan, 2021; Jadli et al., 2022; Nygård & Mezei, 2020; Puravankara & Narendra Babu, 2020; M. Sharma, 2022; Yim et al., 2024) have focused on implementing traditional ML algorithms. This preference may be attributed to their ease of implementation, interpretability, and lower computational requirements (G. Zhang, 2023).

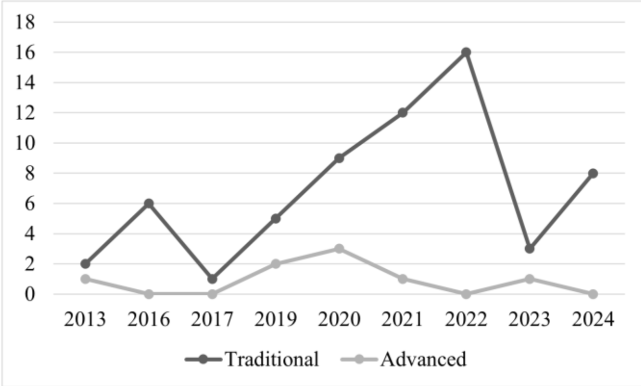


Figure 3. Trends of ML Types in Lead Scoring Models.

However, traditional ML algorithms are not necessarily the best-performing ones, despite their popularity. Up to now, several studies have indicated that advanced ML algorithms achieved superior performance compared to traditional ML algorithms in building predictive lead scoring models (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020). Indeed, advanced ML algorithms excel in handling large datasets (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020), which involve a high degree of complexity, manage a high volume of variables, and capture complex non-linear relationships between features and targets. However, their "black box" nature poses challenges in understanding, explaining, and interpreting the models, which may contribute to their lower popularity (Koeppe et al., 2022). This mixed performance, along with the higher complexity and computational requirements of advanced algorithms, contributes to the ongoing preference for traditional ML methods.

In some cases, advanced ML algorithms are not always the best-performing algorithms. The work by D'Haen and Van Den Poel (2013) concluded that DT outperformed NN and LR by initiating a feedback loop to optimise and stabilise the model. Nygård and Mezei (2020) found that RF outperformed LR, DT, and NN in terms of overall performance. Although NN had higher accuracy and specificity, RF had higher AUC and sensitivity. Furthermore, Espadinha-Cruz et al. (2021) stated that Ensemble Learning was the best-performing algorithm. Thus, the three individual studies make it difficult to conclusively state that advanced ML algorithms are more effective overall.

The findings from the datasets used in these studies indicate that NN tend to outperform in scenarios involving larger datasets (Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; Puravankara & Narendra Babu, 2020). Conversely, studies with smaller datasets often find traditional ML algorithms to be more effective (D'Haen & Van Den Poel, 2013; Nygård & Mezei, 2020). In contrast, the study by Espadinha-Cruz et al. (2021) demonstrates that an ensemble approach, combining regression with PCA and NN using varying numbers of neurons in the hidden layer, outperforms a standalone NN.

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Figure 4 provides a comparative synthesis, delineating the algorithms used in lead scoring models across the sixteen studies. The examined papers underscore a diverse range of ML algorithms, spanning from traditional ML to Ensemble Learning and advanced ML algorithms.

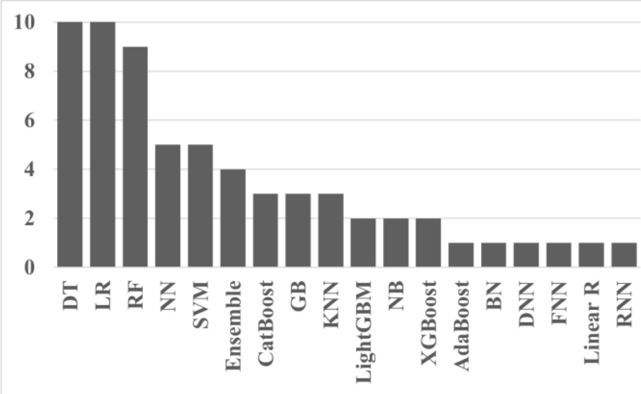


Figure 4. Trends of ML Algorithms Used in Lead Scoring Models.

Figure 4 proves that DT, LR, and RF are among the most frequently implemented algorithms in lead scoring models. Perhaps, the popularity of LR can likely be attributed to its effectiveness in solving linear relationships between features and targets, which aligns well with the binary outputs of lead scoring models. On the other hand, DT is prone to overfitting but is advantageous for handling non-linear relationships between features. RF, an ensemble method that combines multiple DTs, tends to offer improved performance over DT.

5.3 Performance Metrics in Existing Lead Scoring Models

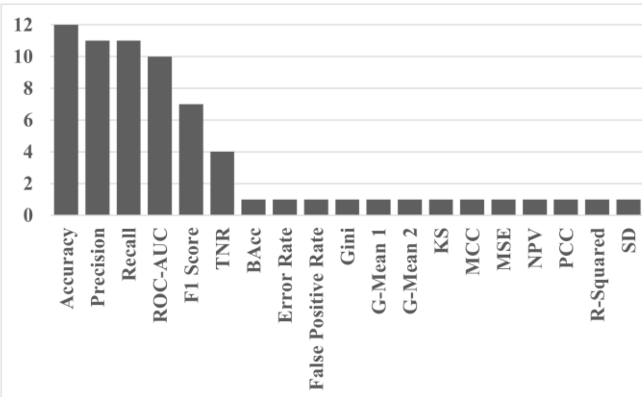


Figure 5. Trends of Performance Metrics Used in Lead Scoring Models.

Figure 5 shows that, among 19 performance metrics utilised across sixteen studies, accuracy is the most frequently used metric (17.65%). Since the studies did not employ a common performance metric for model evaluation, this paper will focus on the most commonly used metric, which is accuracy. The highest accuracy score reported is 99.41%, which is achieved by the work of Choudhury & Nur (2019). However, due to the variation in datasets used across the studies, direct performance comparisons should only be made when the same dataset is applied.

5.4 Feature Engineering Techniques in Existing Lead Scoring Models

Figure 6 compares the different approaches employed by sixteen studies to handle feature engineering.

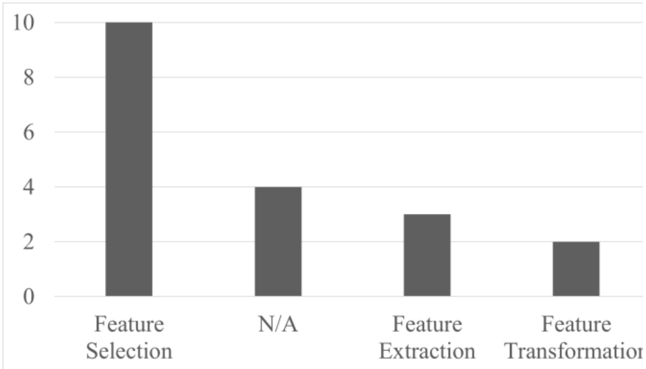


Figure 6. Trends of Feature Engineering Techniques Used in Lead Scoring Models.

From Figure 6, twelve out of sixteen (75%) studies utilised one or more feature engineering techniques. Among these, feature selection emerged as the most popular method, as it effectively reduces the dataset's dimensionality and avoids overfitting issues by choosing relevant features.

6. Challenges and future work

Despite the promising results reported in the studies, several challenges or limitations were noted and are classified in Table 6, which will be discussed in the following subsections.

Table 6. Challenges/Limitations faced by Related Work.

Challenges/Limitations Faced	Related Work
Dataset issues	(Ayaz, 2023; Benhaddou & Leray, 2017; Bhatta, 2022; Espadinha-Cruz et al., 2021; Etminan, 2021; Giorcelli, 2019; Pereira, 2021; Puravankara & Narendra Babu, 2020; Swelsen, 2019; Yim et al., 2024)
Hyperparameter Tuning issues	(Benhaddou & Leray, 2017; Bhatta, 2022; Choudhury & Nur, 2019; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Swelsen, 2019; Yim et al., 2024)
Performance issues	(Ayaz, 2023; Chaudhuri et al., 2021; Choudhury & Nur, 2019; D'Haen & Van Den Poel, 2013; Jadli et al., 2022; Karthik et al., 2020; Pereira, 2021)

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

6.1 Dataset Issues

One recurring issue is dataset issues, including imbalanced data (Benhaddou & Leray, 2017; Etminan, 2021; Swelsen, 2019; Yim et al., 2024), limited size (Benhaddou & Leray, 2017; Etminan, 2021; Giorcelli, 2019), and low quality of data (Pereira, 2021), which may eventually lead to biased predictions (Ayaz, 2023; Yim et al., 2024). Other reported dataset issues include a lack of relationships between datasets (Puravankara & Narendra Babu, 2020), the use of real customer data instead of lead data (Swelsen, 2019), and insufficient data visibility (Bhatta, 2022).

A prior study by Pereira (2021) struggled to achieve high model performance due to low-quality data with only 56% accuracy, 71% sensitivity, 49% specificity, and 42% precision. The low performance of the model was attributed to poor data quality, characterised by a significant proportion of missing values, incomplete data, and inappropriate data formats. Similarly, another study by Ayaz (2023) emphasised that the effectiveness of lead scoring models heavily relies on the quality of the data used. The possible reasons for these dataset issues are problems during the data collection phase, such as a limited data collection period, inadequate data validation processes, and the absence of relevant data. Such issues will lead to imbalances, a limited dataset size, and low data quality, collectively contributing to biased or unreliable model predictions. These studies underscore the critical importance of high-quality data for the success of building a lead scoring model.

Suggestions for future work include thoroughly planning the data collection phase to avoid insufficient and incomplete data. Besides that, Generative Adversarial Networks, a DL model that generates synthetic data mimicking real data, might help solve dataset issues (Akkem et al., 2024). GAN consists of two components: a generator, which produces synthetic data by sampling from a random prior noise distribution, and a discriminator, which evaluates whether the data produced by the generator is real or fake (Navidan et al., 2021). Thus, GAN can address issues such as data imbalance, limited dataset size and low data quality by generating more high-quality lead examples.

6.2 Hyperparameter Tuning Issues

Hyperparameter tuning issues pose significant challenges or limitations. Several studies have mentioned using default parameter values (Benhaddou & Leray, 2017; Choudhury & Nur, 2019; Giorcelli, 2019; Jadli et al., 2022; Karthik et al., 2020; Nygård & Mezei, 2020; Swelsen, 2019; Yim et al., 2024), indicating that hyperparameter tuning was not implemented. As a result, the reported outcomes may not represent the model's optimal performance. A study by Bhatta (2022) mentioned the trade-off between comprehensive hyperparameter optimisation and limited computing resources, which led to the usage of a smaller grid search. This issue can lead to missed opportunities for finding the optimal parameter values, resulting in models that are less effective than they could be with more extensive tuning.

Future research should focus on using advanced hyperparameter optimisation techniques to maximise model performance. Additionally, there is a need to develop more efficient and automated methods for hyperparameter tuning to save time and computational resources. For instance, the Hyperband algorithm, combined with the Bayesian optimisation method, which leverages the strengths of both standalone methods, often outperforms other approaches such as Random Search, as well as each method, and demonstrates a greater margin of improvement, specifically in more complex optimisation problems (Wang Blippar et al., 2018). This approach would enable the optimisation process of lead scoring models to allocate resources to the most promising hyperparameters dynamically. By leveraging the Hyperband algorithm with the Bayesian optimisation method, the search can be refined based on past trials, which significantly speeds up the tuning process and reduces computational costs.

6.3 Performance Issues

Many studies have identified performance issues as one of the challenges or limitations faced. These issues include not addressing the possibility of overfitting in high-performance models (Choudhury & Nur, 2019; Jadli et al., 2022; Karthik et al., 2020), models not being fully tested in real-time environments (Chaudhuri et al., 2021; D'Haen & Van Den Poel, 2013; Pereira, 2021), and the trade-off between accuracy and interpretability (Ayaz, 2023).

The possibility of overfitting in high-performance models indicates that the model may be too closely tailored to the training data, thereby leading to poor performance on new, unseen data. For instance, if a lead scoring model is overfitted, it may misclassify potentially high-value leads as low-value simply because their characteristics do not match the training data. Conversely, it might incorrectly classify low-value leads as high-value, leading to wasted resources on leads that are unlikely to convert. This undermines the generalisability of the model, making it less reliable for practical applications. The lack of testing the models in real-time environments raises questions about the practical applicability of these models, as a model may perform well in a controlled setting but not in a dynamic, real-time situation. While complex models using advanced ML algorithms often yield higher accuracy, they do so at the cost of interpretability. Models that lack interpretability may not earn users' trust as they cannot discern the factors driving the model's outputs.

To date, DL algorithms have yet to be widely integrated into lead scoring models. In fact, despite being ten times slower than Multi-layer Perceptrons, KAN, a newly introduced NN algorithm, offers better accuracy for science-related tasks, making it a promising alternative for future work where accuracy may outweigh computational speed (Liu et al., 2024). Aside from that, Hyperbolic Deep Learning (HNN), which employs hyperbolic space rather than traditional Euclidean space, warrants further analysis. The work by Ganea et al. (2018) indicates that HNN often outperforms its Euclidean variants in most settings. HNN has the potential to significantly improve model accuracy, particularly in tasks involving complex hierarchical or relational data (Ganea et al., 2018), as hyperbolic space is well-suited for representing tree-like structures (Chami et al., 2020). Furthermore, HNN's ability to handle large and complex datasets could make it a powerful tool

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

for improving accuracy in lead scoring models, especially as data becomes more abundant and intricate in future work.

7. Conclusion

This review paper briefly introduces the concepts of lead scoring, ML algorithms, and feature engineering techniques. It also identifies common characteristics of datasets used in lead scoring models, such as reliance on non-social media data and the lack of accessible public datasets. Besides that, this review paper includes a classification table to summarise the lead scoring-related papers based on the ML task, either classification or regression. Apart from that, a comprehensive table provides an overview of different studies utilising the same "X Online Education" dataset. Despite using the same dataset, the results varied due to differences in data preprocessing and feature selection techniques employed by the authors.

This review paper summarises key findings on trends in lead scoring models, highlighting the popularity of traditional ML algorithms, particularly DT, LR, and RF algorithms. Furthermore, this review paper justifies the trends in feature engineering techniques used in lead scoring models.

The paper continues to suggest future directions for building lead scoring models, such as planning the data collection phase thoroughly, using advanced hyperparameter optimisation techniques, and implementing DL algorithms. Nevertheless, more research is required in order to conclude whether the suggested solutions are able to enhance the performance of lead scoring models.

In brief, this review paper plays a pivotal role in advancing the field of lead scoring by offering valuable insights for both academics and practitioners seeking to refine their strategies. By optimising lead scoring, businesses can streamline their sales processes, prioritise high-potential leads, and ultimately enhance overall efficiency and profitability, driving economic growth and job creation. As the landscape of data and technology continues to evolve, ongoing research is essential to refine these strategies further and unlock even greater potential for business success.

8. Acknowledgement

This research was supported by Tunku Abdul Rahman University of Management and Technology (TAR UMT) through the grant UC/I/G2022-00098.

9. References

Abdel-Nasser Sharkawy. (2020). Principle of Neural Network and Its Main Types: Review. *Journal of Advances in Applied & Computational Mathematics*, *7*, 8–19. https://doi.org/10.15377/2409-5761.2020.07.2

Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Engineering Applications of Artificial Intelligence*, *131*, 107881. https://doi.org/10.1016/J.ENGAPPAI.2024.107881

Aldino, A. A., & Sulistiani, H. (2020). Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department Of Information System, Universitas Teknokrat Indonesia). *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, *7*(1), 40–50. https://doi.org/10.21107/EDUTIC.V7I1.8849

Ali, A., Jayaraman, R., Azar, E., & Maalouf, M. (2024). A comparative analysis of machine learning and statistical methods for evaluating building performance: A systematic review and future benchmarking framework. *Building and Environment*, *252*, 111268. https://doi.org/https://doi.org/10.1016/j.buildenv.2024.111268

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., Muneer, A., Sumiea, E. H., Alqushaibi, A., & Ragab, M. G. (2024). RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *Journal of King Saud University - Computer and Information Sciences*, *36*(5), 102068. https://doi.org/https://doi.org/10.1016/j.jksuci.2024.102068

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Arista, A. (2022). Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19. *Jurnal Dan Penelitian Teknik Informatika*, *7*(1). https://doi.org/10.33395/sinkron.v7i1.11243

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*. https://doi.org/10.1109/wi-iat.2010.63

Autopilot. (2016, January 12). *Lead Scoring is Broken. Here's What to Do Instead.* https://medium.com/marketing-on-autopilot/lead-scoring-is-broken-here-s-what-to-do-instead-194a0696b8a3

Ayaz, S. B. (2023). *Lead Scoring with Machine Learning*.

Balfagih, A. (2016). *Direct Selling Business Lead Prediction by Social Media Data Mining*. https://DalSpace.library.dal.ca//handle/10222/71412

Benhaddou, Y., & Leray, P. (2017). Customer relationship management and small data - Application of Bayesian network elicitation techniques for building a lead scoring model. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, *2017-October*, 251–255. https://doi.org/10.1109/AICCSA.2017.51

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Berrar, D. (2018). *Bayes' Theorem and Naive Bayes Classifier*. https://doi.org/10.1016/B978-0-12-809633-8.20473-1

Bhardwaj, P., Tiwari, P., Olejar, K., Parr, W., & Kulasiri, D. (2022). A machine learning application in wine quality prediction. *Machine Learning with Applications*, *8*, 100261. https://doi.org/https://doi.org/10.1016/j.mlwa.2022.100261

Bharti, K. K., & Singh, P. K. (2015). Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Systems with Applications*, *42*(6), 3105–3114. https://doi.org/https://doi.org/10.1016/j.eswa.2014.11.038

Bhatta, I. (2022). *Optimizing Marketing Channel Attribution for B2B and B2C with Machine Learning Based Lead Scoring Model*.

Bouwmans, T., Javed, S., Sultana, M., & Jung, S. K. (2019). Deep neural network concepts for background subtraction:A systematic review and comparative evaluation. *Neural Networks*, *117*, 8–66. https://doi.org/10.1016/J.NEUNET.2019.04.024

Ceccatelli, J. (2023). *The impact of lead scoring on CRM and the sales process*. https://www.repository.utl.pt/handle/10400.5/28214

Chami, I., Gu, A., Chatziafratis, V., & Ré, C. (2020). From Trees to Continuous Embeddings and Back: Hyperbolic Hierarchical Clustering. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 15065–15076). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/ac10ec1ace51b2d973cd87973a98d3ab-Paper.pdf

Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*, *149*, 113622. https://doi.org/10.1016/J.DSS.2021.113622

Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *Eurasip Journal on Advances in Signal Processing*, *2021*(1), 1–12. https://doi.org/10.1186/S13634-021-00742-6/TABLES/5

Choudhury, A. M., & Nur, K. (2019). A Machine Learning Approach to Identify Potential Customer Based on Purchase Behavior. *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, 242–247. https://doi.org/10.1109/ICREST.2019.8644458

Das, H. P., & Spanos, C. J. (2022). Improved dequantization and normalization methods for tabular data pre-processing in smart buildings. *BuildSys 2022 - Proceedings of the 2022 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 168–177. https://doi.org/10.1145/3563357.3564072

De Haan, E., & Menichelli, E. (2019). *The Incremental Value of Unstructured Data in Predicting Customer Churn*.

D'Haen, J., & Van Den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management*, *42*(4), 544–551. https://doi.org/10.1016/J.INDMARMAN.2013.03.006

DiPietro, R., & Hager, G. D. (2020). Chapter 21 - Deep learning: RNNs and LSTM. In S. K. Zhou, D. Rueckert, & G. Fichtinger (Eds.), *Handbook of Medical Image Computing and Computer Assisted Intervention* (pp. 503–519). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Duncan, B., & Elkan, C. (2015). Probabilistic modeling of a sales funnel to prioritize leads. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2015-August*, 1751–1758. https://doi.org/10.1145/2783258.2788578

Espadinha-Cruz, P., Fernandes, A., & Grilo, A. (2021). Lead management optimization using data mining: A case in the telecommunications sector. *Computers & Industrial Engineering*, *154*, 107122. https://doi.org/10.1016/J.CIE.2021.107122

Etminan, A. (2021). *Prediction of Lead Conversion With Imbalanced Data : A method based on Predictive Lead Scoring*. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-176433

Fahad, K. M. W. (2017). *A study on business development strategies of LEADS Impacto Ltd*. http://dspace.bracu.ac.bd:8080/xmlui/handle/10361/8694

Fernandes, A. A. T., Filho, D. B. F., da Rocha, E. C., & da Silva Nascimento, W. (2021). Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política*, *28*(74), 006. https://doi.org/10.1590/1678-987320287406EN

Ganea, O.-E., Bécigneul, G., & Hofmann, T. (2018). Hyperbolic Neural Networks. *Advances in Neural Information Processing Systems*, *31*.

Giorcelli, G. (2019). *Variable-sized input, character-level recurrent neural networks in lead generation: predicting close rates from raw user inputs*. https://arxiv.org/abs/1901.05115v1

Goh, K. H., Wang, L., Yeow, A. Y. K., Poh, H., Li, K., Yeow, J. J. L., & Tan, G. Y. H. (2021). Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature Communications*, *12*(1), 711. https://doi.org/10.1038/s41467-021-20910-4

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Hou, Y., zhang, D., Wu, J., & Feng, X. (2024). *A Comprehensive Survey on Kolmogorov Arnold Networks (KAN)*. https://arxiv.org/abs/2407.11075v3

How, D. N. T., Hannan, M. A., Lipu, M. S. H., Sahari, K. S. M., Ker, P. J., & Muttaqi, K. M. (2020). State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach. *IEEE Transactions on Industry Applications*, *56*(5), 5565–5574. https://doi.org/10.1109/TIA.2020.3004294

Huang, X., Wang, S., Zhang, M., Hu, T., Hohl, A., She, B., Gong, X., Li, J., Liu, X., Gruebner, O., Liu, R., Li, X., Liu, Z., Ye, X., & Li, Z. (2022). Social media mining under the COVID-19 context: Progress, challenges, and opportunities. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102967. https://doi.org/https://doi.org/10.1016/j.jag.2022.102967

Ismail, M., Hassan, N., & Saleh Bafjaish, S. (2020). Comparative Analysis of Naive Bayesian Techniques in Health-Related For Classification Task. *Journal of Soft Computing and Data Mining*, *1*(2), 1–10. https://doi.org/10.30880/jscdm.2020.01.02.001

Jadli, A., Hamim, M., Hain, M., & Hasbaoui, A. (2022). *TOWARD A SMART LEAD SCORING SYSTEM USING MACHINE LEARNING*. https://doi.org/10.21817/indjcse/2022/v13i2/221302098

Kanavos, A., Antonopoulos, N., Karamitsos, I., & Mylonas, P. (2023). A Comparative Analysis of Tweet Analysis Algorithms Using Natural Language Processing and Machine Learning Models. *2023 18th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2023*. https://doi.org/10.1109/SMAP59435.2023.10255184

Kane, A., & Hussain, A. (2023). Artificial Neural Networks: An Overview. *Mesopotamian Journal of Computer Science*, *2023*(5), 124–133. https://doi.org/10.58496/MJCSC/2023/015

Karthik, P. V. V., Rao, P. G. P., & Narsimlu, M. (2020). Case Study: Lead Conversion of Digital Marketing Data Using Predictive Analytics. *Lecture Notes in Electrical Engineering*, *601*, 510–519. https://doi.org/10.1007/978-981-15-1420-3_54

Kim, D., Kang, S., & Cho, S. (2020). Expected margin–based pattern selection for support vector machines. *Expert Systems with Applications*, *139*, 112865. https://doi.org/10.1016/J.ESWA.2019.112865

Koeppe, A., Bamer, F., Selzer, M., Nestler, B., & Markert, B. (2022). Explainable Artificial Intelligence for Mechanics: Physics-Explaining Neural Networks for Constitutive Models. *Frontiers in Materials*, *8*, 824958. https://doi.org/10.3389/FMATS.2021.824958/BIBTEX

Kumar, K. P., Unal, A., Pillai, V. J., Murthy, H., & Niranjanamurthy, M. (2023). *Data engineering and data science: concepts and applications*.

Lee, C. S., Yeng, P., Cheang, S., & Moslehpour, M. (2022). *Predictive Analytics in Business Analytics: Decision Tree*.

Lindahl, E., Skolan, K., Datavetenskap, F., & Kommunikation, O. (2017). *A qualitative examination of lead scoring in B2B marketing automation, with a recommendation for its practice.* https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-213432

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). *KAN: Kolmogorov-Arnold Networks*. https://arxiv.org/abs/2404.19756v2

Markowska-Kaczmar, U., & Kosturek, M. (2021). Extreme learning machine versus classical feedforward network. *Neural Computing and Applications*, *33*(22), 15121–15144. https://doi.org/10.1007/s00521-021-06402-y

Mers, M., Yang, Z., Hsieh, Y.-A., & Tsai, Y. (James). (2022). Recurrent Neural Networks for Pavement Performance Forecasting: Review and Model Performance Comparison. *Transportation Research Record*, *2677*(1), 610–624. https://doi.org/10.1177/03611981221100521

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, *10*, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/J.JKSUCI.2023.01.014

Mohd Ali, N. F., Mohd Sadullah, A. F., Abdul Majeed, A. P. P., Mohd Razman, M. A., & Musa, R. M. (2022). The identification of significant features towards travel mode choice and its prediction via optimised random forest classifier: An evaluation for active commuting behavior. *Journal of Transport & Health*, *25*, 101362. https://doi.org/10.1016/J.JTH.2022.101362

Morgado, A. V. (2018). The Value of Customer References to Potential Customers in Business Markets. *Journal of Creating Value*, *4*(1), 132–154. https://doi.org/10.1177/2394964318771799

Navidan, H., Moshiri, P. F., Nabati, M., Shahbazian, R., Ghorashi, S. A., Shah-Mansouri, V., & Windridge, D. (2021). Generative Adversarial Networks (GANs) in networking: A comprehensive survey & evaluation. *Computer Networks*, *194*, 108149. https://doi.org/https://doi.org/10.1016/j.comnet.2021.108149

Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, *7*, 1246–1255. https://doi.org/10.1016/J.EGYR.2021.02.006

Niemi, E. (2022). *Sales Lead Utilization with Distribution Partners in B2B : Creating an Analytical Model for Sales Lead Assessment*. https://trepo.tuni.fi/handle/10024/136382

Nosouhian, S., Nosouhian, F., & Khoshouei, A. K. (2021). A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU. *Preprints*. https://doi.org/10.20944/preprints202107.0252.v1

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Nygård, R., & Mezei, J. (2020). Automating Lead Scoring with Machine Learning: An Experimental Study. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-January*, 1439–1448. https://doi.org/10.24251/HICSS.2020.177

Pereira, R. M. M. (2021). *Building a predictive lead scoring model for contact prioritization : the case of HUUB*. https://repositorio.ucp.pt/handle/10400.14/34877

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. *Machine Learning: Methods and Applications to Brain Disorders*, 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Puravankara, R., & Narendra Babu, C. (2020). Lead Forecasting using LSTM based Deep Learning Architecture for Sentiment Analysis. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 159–164. https://doi.org/10.1109/ICOIACT50329.2020.9332092

Rodrigues, J. D. da S. S. (2020). *Automated lead scoring system: a case study of a Portuguese startup*. Universidade do Porto (Portugal).

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, *2*(7), 1–17. https://doi.org/10.1007/S42452-020-3060-1/TABLES/1

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Shaik, A. B., & Srinivasan, S. (2019). A Brief Survey on Random Forest Ensembles in Classification Model. *Lecture Notes in Networks and Systems*, *56*, 253–260. https://doi.org/10.1007/978-981-13-2354-6_27

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis. *IEEE Access*, *10*, 56720–56739. https://doi.org/10.1109/ACCESS.2022.3177752

Sharma, K. K., Tomar, M., & Tadimarri, A. (2023). Optimizing Sales Funnel Efficiency: Deep Learning Techniques for Lead Scoring. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (Online)*, *2*, 261–274. https://doi.org/10.60087/jklst.vol2.n2.p274

Sharma, M. (2022). *Identifying Factors Contributing to Lead Conversion Using Machine Learning to Gain Business Insights MSc Research Project MSCDADJAN22*.

Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). *A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU*. https://arxiv.org/abs/2305.17473v2

Song, Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*, 130–135. https://api.semanticscholar.org/CorpusID:18242585

Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, *67*, 101813. https://doi.org/https://doi.org/10.1016/j.media.2020.101813

Suhaidi, M., Kadir, R. A., & Tiun, S. (2021). A REVIEW OF FEATURE EXTRACTION METHODS ON MACHINE LEARNING. *JOURNAL INFORMATION AND TECHNOLOGY MANAGEMENT (JISTM)*, *6*(22), 51–59. https://doi.org/10.35631/JISTM.622005

Sundman, S.-M. (2021). *Customer Acquisition Development Plan : Improving Lead Generation*. http://www.theseus.fi/handle/10024/356250

Swelsen, C. W. J. M. (2019). *Proposing a generic online lead scoring model for a B2C market*.

Thakur, A., & Konde, A. (2021). Fundamentals of Neural Networks. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *9*, 2321–9653. www.ijraset.com

Thompson, N., Greenewald, K., Lee, K., & Manso, G. F. (2021). *The Computational Limits of Deep Learning*.

Vijayakumar, K., Kadam, V. J., & Sharma, S. K. (2021). Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering*, *29*(3), 275–284. https://doi.org/10.1177/1063293X211025105

Vo, N. N. Y., Liu, S., Li, X., & Xu, G. (2021). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems*, *212*, 106586. https://doi.org/https://doi.org/10.1016/j.knosys.2020.106586

Wang Blippar, J., Xu Blippar, J., & Wang Blippar, X. (2018). *Combination of Hyperband and Bayesian Optimization for Hyperparameter Optimization in Deep Learning*. https://arxiv.org/abs/1801.01596v1

Wang, L., Ye, W., Zhu, Y., Yang, F., & Zhou, Y. (2023). Optimal parameters selection of back propagation algorithm in the feedforward neural network. *Engineering Analysis with Boundary Elements*, *151*, 575–596. https://doi.org/https://doi.org/10.1016/j.enganabound.2023.03.033

Waters, P. (2022). *Lead generation in Business-to Business marketing and sales : how can Heimo Films improve their lead generation process?* http://www.theseus.fi/handle/10024/753271

Wu, M., Andreev, P., & Benyoucef, M. (2024). The state of lead scoring models and their impact on sales performance. *Information Technology and Management*, *25*(1), 69–98. https://doi.org/10.1007/S10799-023-00388-W/TABLES/8

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

Yang, G. R., & Wang, X.-J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*, *107*(6), 1048–1070. https://doi.org/10.1016/j.neuron.2020.09.005

Yim, W. Y., Khaw, K. W., Lim, S. T., & Chew, X. (2024). Enhancing Conversions and Lead Scoring in Online Professional Education. *International Journal of Management, Finance and Accounting*, *5*(1), 15–63. https://doi.org/10.33093/IJOMFA.2024.5.1.2

Zabor, E. C., Reddy, C. A., Tendulkar, R. D., & Patil, S. (2022). Logistic Regression in Clinical Studies. *International Journal of Radiation Oncology*Biology*Physics*, *112*(2), 271–277. https://doi.org/10.1016/J.IJROBP.2021.08.007

Zelikovitz, S., & Hirsh, H. (2001). Using LSI for text classification in the presence of background text. *Proceedings of the Tenth International Conference on Information and Knowledge Management*, 113–118. https://doi.org/10.1145/502585.502605

Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable Cities and Society*, *39*, 508–518. https://doi.org/https://doi.org/10.1016/j.scs.2018.02.016

Zhang, G. (2023). Comparison of Machine Learning Models and Traditional Models for Forecasting in the Economy. *Computer Life*, *11*(3), 1–6. https://doi.org/10.54097/N11MBN72

Zhao, Y., Wang, T., Bove, R., Cree, B., Henry, R., Lokhande, H., Polgar-Turcsanyi, M., Anderson, M., Bakshi, R., Weiner, H. L., & Chitnis, T. (2020). Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. *Npj Digital Medicine 2020 3:1*, *3*(1), 1–8. https://doi.org/10.1038/s41746-020-00338-8

Zhu, F., Zhang, W., Chen, X., Gao, X., & Ye, N. (2023). Large margin distribution multi-class supervised novelty detection. *Expert Systems with Applications*, *224*, 119937. https://doi.org/10.1016/J.ESWA.2023.119937

Zhuang, H., Wang, X., Bendersky, M., & Najork, M. (2020). Feature Transformation for Neural Ranking Models. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1649–1652. https://doi.org/10.1145/3397271.3401333

DOI: https://doi.org/10.22452/mjs.vol45sp1.6
Malaysian Journal of Science 45 (Special Issue 1): 44-57 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024 Special issue

Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026) https://mjs.um.edu.my

Evaluating Sedimentation Effects on Mangrove Growth: Comparative Analysis of Coastal Zones in Batu Pahat, Malaysia

Muhammad Aliyuddin Mohd Kamal[1a], Nor Baizura Hamid[2b*], Mardiha Mokhtar[3b], Aslila Abd Kadir[4b], Mohd Effendi Daud[5a] and Mohammad Ashraf Abdul Rahman[6c]

Abstract: Mangrove ecosystems, often called the rainforests of the ocean, are essential coastal habitats that offer key environmental advantages. Nevertheless, they are confronted with an impending danger from sedimentation, a complicated mix of natural and human-caused elements. The research examines the various impacts of sedimentation on the well-being of mangroves, specifically looking at Pantai Punggor and Pantai Perpat in Batu Pahat, Johor, Malaysia. Avicennia marina, a type of mangrove, is impacted by sediment accumulation and is chosen for its importance in these areas. UAV photogrammetry was applied because of its cost-effectiveness and high spatial resolution. Data underwent processing with the aid of Global Mapper and Pix4D programs, in addition to meticulous mission preparation and camera calibration. Sediment properties were examined per ASTM guidelines, utilizing sieve analysis, specific gravity, and moisture content methods with extra determination of salinity and pH from the in-situ process. The assessment of mangrove characteristics involved measuring trunk diameter with field data and tree height with UAV-derived Digital Surface and Terrain Models. The findings show notable differences between the zones in Pantai Punggor and Pantai Perpat, highlighting the negative impacts of sedimentation on the development of mangroves. Zone 3C aligns most closely with the specified ideal characteristics for mangrove height, soil composition, and environmental conditions. The results reveal significant differences in mangrove growth across sedimentation zones, with Zone 3C emerging as a key area for conservation efforts. The findings provide actionable insights for improving sediment management strategies and ensuring the sustainability of mangrove ecosystems.

Keywords: Sedimentation, mangrove, Avicennia marina, sediment deposition

1. Introduction

The mangrove ecosystems in Malaysia, known as the sea forests, have great ecological importance and are found extensively along the country's coastlines. These special forests along the coast of Malaysia are crucial for the country's diverse wildlife, providing a home for a wide variety of plants and animals, such as various types of fish, birds, and crustaceans. Also, they act as vital breeding grounds for economically important fish species, sustaining the local fishing sector. Nevertheless, these precious ecosystems are at risk from multiple sources, with sedimentation becoming a major issue. The problem of sedimentation arises from a mix of natural factors like tidal action and river flow, as well as human activities such as deforestation and changes in land use, which together upset the sediment deposition balance and endanger these ecosystems.

The health of mangroves can be impacted by sediment transport in various ways. Excessive sediment deposition can alter the physical and chemical composition of the soil, leading to a reduction in nutrients and oxygen available to mangrove roots (Ndungu et al., 2019). The aftermath is that mangroves may grow more slowly, become more susceptible to pests and illnesses, and eventually die. (J. Wang et al., 2019).

Along with soil characteristics, sediment transport impacts the hydrology of mangrove habitats. According to Basyuni et al., (2020), sediment accumulation in waterways can block them and decrease the flow, causing higher salinity and waterlogging in some places and lower salinity and drying out in others. This may change the mangrove species distribution and negatively impact the environment's biodiversity. Many methods have been created to address the harmful effects of sediment transport on the well-being of mangroves. Sediment-capturing infrastructure such as sediment ponds and check dams are deployed to reduce silt migration downstream (Razi et al., 2019). Two additional initiatives to reduce sedimentation rates are restoring injured mangrove regions and implementing sustainable land management techniques. Basyuni et al., (2020).

To conclude, sediment transportation poses a significant threat to the health and sustainability of Malaysia's mangrove ecosystems. Mangrove ecosystems are indispensable for

Authors information:

[a]Department of Infrastructure & Geomatic Engineering. Faculty of Civil Engineering and Build Environment, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor. E-mail: aliyuddinkamal@gmail.com[1], effendi@uthm.edu.my[5],

[b]Department of Civil Engineering. Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia (Pagoh Campus), Pagoh Higher Education Hub, KM 1, Jalan Panchor, 84600 Panchor, Johor. E-mail: norbaizura@uthm.edu.my[2], mardiha@uthm.edu.my[3], aslila@uthm.edu.my[4],

[c]Department of Civil Engineering Technology. Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia (Pagoh Campus), Pagoh Higher Education Hub, KM 1, Jalan Panchor, 84600 Panchor, Johor. E-mail: ashrafr@uthm.edu.my[6]

*Corresponding Author: norbaizura@uthm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

maintaining coastal stability, acting as natural buffers against erosion (Chang & Mori, 2021), and supporting a wide variety of species. In Malaysia, these ecosystems face increasing threats from sedimentation driven by tidal actions, deforestation, and changes in land use. Excessive sediment deposition can alter soil properties, disrupt water flow, and hinder mangrove growth. Although global studies have investigated the impacts of sedimen tation, there is limited research on its specific effects in Batu Pahat, Malaysia. This study addresses this gap by analyzing how sediment accumulation impacts Avicennia marina growth, utilizing advanced UAV photogrammetry and sediment analysis techniques. The novelty of this research lies in its precise, technology-driven approach to understanding these challenges and offering practical solutions.

1.1 Mangrove Ecosystem Characteristic

Mangrove ecosystems are characterized by their remarkable adaptive abilities to thrive in challenging coastal environments,

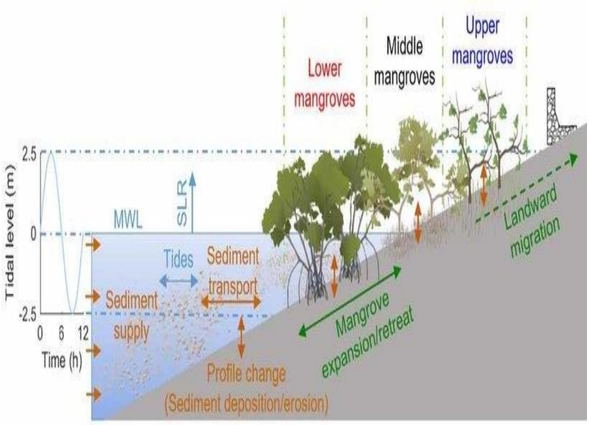


Figure 1. Schematic overview of the bio-morphodynamic modeling (Xie et al., 2022)

making them rich and unique habitats. These environments usually include six various types of mangroves and ten plant species associated with the habitat, amounting to a total of 16 species. The makeup of these species may differ across various places, with certain regions showing greater diversity. Avicennia marina, Suaeda maritima, and Sesuvium portulacastrum are some of the notable species that are common in the area. These mangroves play a crucial role in coastal ecosystems and offer necessary ecological services. Previous research offers significant insights into the mangrove ecosystem's species diversity and geographical spread (Singh, 2020). Figure 1 depicts the schematic overview of the bio-morphodynamic modeling.

One of the defining characteristics of mangrove habitats is their ability to thrive in challenging conditions. Mangroves have progressed to withstand high temperatures, salinity, low oxygen levels, and contaminated environments. (Mohanta et al., 2020). Their roots are specially adapted to handle waterlogged soils, which can vary widely in salinity, ranging from 2% to 90%.[5]. The ability of mangroves to adapt to shifts in sea level and harsh weather conditions allows them to be resilient. Their distinct physical makeup enables them to strain and capture sediments and nutrients, which are vital in maintaining water quality and stabilizing shorelines.

Mangrove forests can differ in their physical appearance, as mangrove trees can be small shrubs or tall giants reaching heights of 5 to 25 meters. The height and shape of the mangroves are influenced by different factors, such as the age of the forest and local environmental conditions. The wide variety of traits creates a broad spectrum of habitats for different wildlife species, turning mangrove ecosystems into a hotspot for biodiversity. Furthermore, mangrove environments are recognized for their importance in supporting a diverse range of aquatic and land-based creatures, demonstrating their critical significance for the overall well-being and ability to withstand the challenges of coastal ecosystems across the globe (Hai et al., 2020).

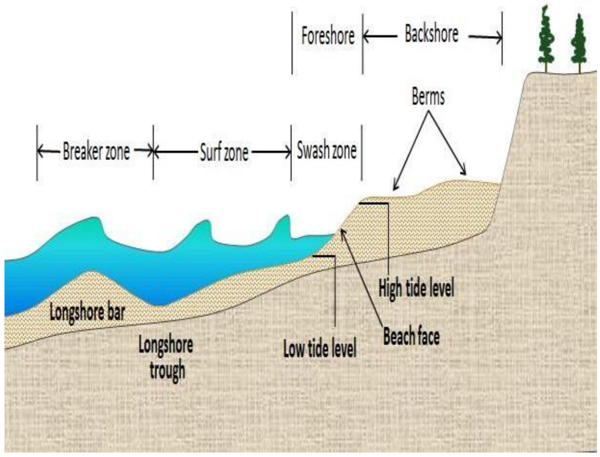


Figure 2. The components of a sandy marine beach (Earle et al., 2015).

1.2 Sediment Transport

Sediment transport is a dynamic and complex process primarily driven by fluid flow forces like water, wind, or human activities. In the case of water-driven sediment transport, rivers and streams are the main pathways for sediment movement (Dietrich et al., 2020). River flow creates shear stress on the riverbed, leading to sediment particles' dislodgement and downstream transport. (Parker et al., 2016). Fluvial transport is a process that is required to shape the river channels and build sediment along river networks. It plays a vital role in restructuring soil and sediments across landscapes. A component of the sandy marine beach is shown in Figure 2.

The zigzag pattern of particles on beaches is caused by the movement of coastal sediment due to waves and their circular motion. (Komar, 2017). Longshore sediment transport plays a role in shaping barrier islands and beaches and influences coastal ecosystems' dynamics. Moreover, sediments in estuaries and tidal flats are transported by tidal currents, moving them around within these coastal areas (McLachlan et al., 2021). Coastal sediment transport is essential for preserving the balance and structure of shoreline environments.

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

Sediment transport driven by wind is common in dry and coastal regions. Fine particles, such as sand and dust, are easily entrained and carried by wind breezes (Nickling et al., 2019). Aeolian processes, including saltation (hopping and bouncing of particles) and suspension (particles suspended in the air), can carry sediments over long distances (Pahtz et al., 2020). This wind-driven sediment transport can have a range of outcomes, from manipulating desert landscapes to impacting air quality and human health, particularly in areas liable to dust storms (Lancaster et al., 2019). Understanding these different modes of sediment transport is necessary for controlling and lessening their impacts on both natural and human-modified environments.

1.3 Sediment Deposition Affecting Mangrove

Mangroves play a crucial role in influencing sediment deposition and, in turn, are significantly impacted by sediment rates. These ecosystems are nature's way of regulating sediment transport and maintaining the elevation of coastal areas. They effectively reduce sediment movement by slowing down water flow and providing a habitat where suspended sediments can settle. The intricate root networks and prop roots of mangroves act as natural barriers, trapping sediments and promoting

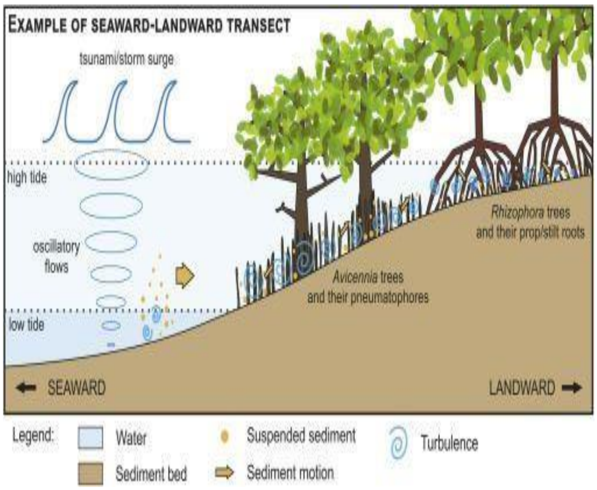


Figure 3. The process of sedimentation at mangrove area (Minor et al., 2021).

sediment deposition within their vicinity, as observed in the study by Besset et al. (2019). This process not only helps mangrove ecosystems grow taller but also leads to the formation of mudflats and marshy regions in coastal areas.

Sediment accumulation rates in mangrove environments are influenced by various factors, including tidal patterns, vegetation density, and hydrodynamics, as demonstrated by research by Asp et al. (2018). These variables interact to determine the extent and rate of sediment buildup within the mangrove ecosystem. Such sediment deposition is vital for mangroves' overall health and development, allowing them to thrive in their unique coastal habitats and providing essential ecological services.

Figure 3 shows the process of how the sediment carried into the mangrove area. Furthermore, mangrove ecosystems are known for enriching sediment with organic matter and nutrients. Mangroves absorb and retain these nutrients, resulting in a nutrient-rich environment that supports diverse organisms, as evidenced by studies like Salmo et al., (2029). The complex nutrient cycling processes in mangrove sediments play a vital role in maintaining the ecosystem's health, benefiting both the flora and fauna that call these habitats home. In addition to their ecological significance, mangroves serve as natural barriers against coastal erosion by reducing wave energy and preventing shoreline erosion through sediment accumulation, making them indispensable in safeguarding coastal communities and infrastructure, as highlighted in research by Saintilan et al. (2020). The intricate relationship between mangrove ecosystems and sediment rates underscores their importance in coastal management, conservation efforts, and the broader environmental context.

1.4 Existing Research at Pantai Punggor and Pantai Perpat

The study by Siddiq et al. (2023) emphasizes the important influence of sediment accumulation on the ecological well-being of mangroves within the research site. Accretion and erosion are critical in shaping their state due to dynamic processes.

The research conducted by Kaamin et al. (2018) expands on this understanding and discusses real-world findings from research carried out in the identical region. The study records noticeable alterations in the shape of the shoreline over time, indicating that the relationship between sediment movement and coastal characteristics is continually changing.

Additionally, utilizing UAV photogrammetry is revolutionizing how sediment volumes are measured, and shoreline changes are monitored, providing unmatched precision and efficiency. (Mokhtar et al., 2023). The potential of this technology extends further to assess vertical vegetation structure, like determining tree heights through creating and analysing Digital Terrain Models (DTM) and Digital Surface Models (DSM). The versatility and practicality of UAV-based techniques in improving our understanding of coastal ecosystems and their complex dynamics are highlighted by their multi-dimensional functionality.

2. Materials and method

Figure 4. Study area, including the zones.

2.1 Description of the Site Location

The study initially focused on Pantai Punggor and Pantai Perpat as the chosen research areas. One of the areas in Batu Pahat that experience notable erosion and accretion phenomena

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

is Pantai Punggor and Pantai Perpat. The coastal area in Batu Pahat shows considerable diversity (Kaamin et al., 2018). Sabri et al,. (2023) research on shoreline changes reveals how the condition of mangroves on both beaches can be impacted within a year due to variations in the shoreline (Mokhtar et al., 2020).

Figure 4. shows the study area and the zones. This research was conducted at Pantai Punggor and Pantai Perpat, two coastal zones in Batu Pahat, Johor, Malaysia. These areas were divided into three zones based on sedimentation levels: Zone 1 (minimal sedimentation), Zone 2 (moderate sedimentation), and Zone 3 (high sedimentation). Avicennia marina was chosen as the focal species due to its prevalence in these environments and its sensitivity to sediment changes (Hassan et al., 2016). The area that already has buildings (Zone 2) was neglected because this study focuses on finding the effect of sediment deposition on mangrove growth.

2.2 UAV Photogrammetry

Unmanned Aerial Vehicles (UAVs) were used for their cost-efficiency and ability to capture high-resolution spatial data (Sabri et al., 2022). Key steps included careful mission planning, camera calibration, and the integration of Ground Control Points (GCPs) to ensure accuracy (Siddiq et al., 2021). Data from the UAV flights were processed using Pix4D and Global Mapper software to create Digital Surface Models (DSM) and Digital Terrain Models (DTM). Tree heights were derived by subtracting the DTM from the DSM, and accuracy was validated using root mean squared error (RMSE) analysis.

2.3 Characteristics of Site Location

2.3.1 Mangrove Characteristic

The data for the mangrove sample was collected along 100 meters of coastline for both locations. Mangrove trunks were assessed according to their diameter and height. Field data collection at Pantai Punggor and Pantai Perpat calculated the average diameter of mangrove trees to determine the trunk diameter. The tree's height was calculated using UAV photogrammetry to generate DSM and DTM. Subtracting the DTM from the DSM determines the tree's height as outlined in (Ruzgiene et al., 2015). The model's accuracy was assessed based on the maps' root mean squared error (RMSE).

2.3.2 Sediment Characteristics

The sediment sample for this investigation was obtained from the highest tide location of each zone in the Pantai Punggor and Pantai Perpat using a hand auger. Soil samples were collected in accordance with ASTM International's Standard Practice for Field Collection of Soil Samples for Subsequent Lead Determination (ASTM E 1727) (Pazi et al., 2021) to ensure consistency and reliability. The particle size distribution of the soil samples was determined through sieve analysis, specific gravity, and moisture content techniques following the Standard Test Method for Particle-Size Analysis of Soils (ASTM D 422) (Beretta et al., 2018) to understand sediment density and water retention capabilities. Additionally, tests for salinity and pH of the sediment were analyzed to evaluate soil conditions for mangrove health. These parameters were selected to provide a comprehensive understanding of the sediment's impact on mangrove health.

2.4 Analytical Framework

To enable effective comparison, the data were normalized using the min-max scaling technique. This method adjusted the varying measurement scales to a common range, ensuring a robust and accurate statistical analysis of both sediment characteristics and mangrove growth metrics typically between 0 and 1 (Cotton-Barratt et al., 2020).

The most widely used method of normalization is min-max normalization, which involves transforming the values to a scale between 0 and 1 using the following formula (Perminov et al., 2017):

$$X' = \frac{X - Xmin}{Xmax - Xmin} \tag{1}$$

Table 1: Mangrove Characteristics Based on Zones

Size	Zone 1			Zone 2			Zone 3		
	A	B	C	A	B	C	A	B	C
Avg Height (m)	6.08	5.91	7.28	3.79	7.29	5.14	2.46	2.70	7.09
Avg Diameter (m)	0.21	0.22	0.24	0.17	0.26	0.18	0.11	0.17	0.24
Density (Tree/m^2)	0.38	0.44	0.31	0.56	0.31	0.5	0.63	0.5	0.31

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

3. Results

3.1 Mangrove Characteristics

Table 1 provides a comparative analysis of tree characteristics across three different zones, each divided into three subcategories: A, B, and C.

In Zone 1, the average tree height ranges from 5.91 meters in subcategory B to 7.28 meters in subcategory C, with an average diameter varying between 0.21 meters (subcategory A) and 0.24 meters (subcategory C). The tree density in Zone 1 is highest in Subcategory B at 0.44 trees per square meter and lowest in Subcategory C at 0.31 trees per square meter.

In Zone 2, the average tree height spans from 3.79 meters in subcategory A to 7.29 meters in subcategory B. The average diameter in this zone is slightly larger than in Zone 1, with subcategory B having the largest diameter at 0.26 meters. The density varies significantly, with subcategory A showing the highest density at 0.56 trees per square meter and subcategory B the lowest at 0.31 trees per square meter.

Zone 3 exhibits the most variation in tree heights, ranging from 2.46 meters in subcategory A to 7.09 meters in subcategory C. The average diameter follows a similar trend, with the smallest trees in subcategory A (0.11 meters) and the largest in subcategory C (0.24 meters). Interestingly, Zone 3 has the highest tree density overall, with subcategory A reaching 0.63 trees per square meter, while subcategories B and C both have lower densities at 0.5 and 0.31 trees per square meter, respectively.

The study found notable differences in mangrove growth across the three zones. Zone 1 exhibited the tallest trees and largest trunk diameters, while Zone 3, heavily affected by sediment accumulation, showed stunted growth. For instance, trees in Zone 1C had an average height of 7.28 meters, compared to just 2.46 meters in Zone 3A. The findings highlight the detrimental impact of excessive sediment accumulation, which restricts nutrient and oxygen availability to mangrove roots.

3.2 Sediment Characteristics

Table 2 provides detailed soil characteristics for three zones with three subcategories (A, B, and C). The characteristics analyzed include Specific Gravity (Gs), Particle Distribution (Silt Clay) percentage, Moisture Content percentage, pH Level, and Salinity in parts per thousand (ppt).

In Zone 1, the Specific Gravity ranges from 2.63 (A) to 2.77 (C), indicating slight variations in soil density. The Particle Distribution (Silt Clay) is relatively high, with values between 75% (B) and 88% (C), suggesting a clay-rich soil. Moisture Content is highest in subcategory C at 35% and lowest in A at 23%. The pH levels slightly vary from neutral to slightly acidic, ranging from 6.49 (B) to 7.14 (A). Salinity levels are consistent, ranging between 24.5 and 25.8 ppt.

Table 2: Sediment Characteristics Based on Zones

Characteristic	Zone 1			Zone 2			Zone 3		
	A	B	C	A	B	C	A	B	C
Specific Gravity (Gs)	2.63	2.73	2.77	2.66	2.75	2.67	2.62	2.65	2.77
Particle Distribution (Silt Clay) (%)	82	75	88	17.71	85.86	25.72	10.34	14.27	84.89
Moisture Content (%)	23	25	35	16	31	17	6	8	33
pH Level	7.14	6.49	6.77	7.02	6.68	7.31	6.49	6.07	6.84
Salinity (ppt)	24.5	24.6	25.8	18.7	23.4	24.1	21.5	20.4	25.8

Zone 2 exhibits similar patterns in Specific Gravity, with values ranging from 2.66 (A) to 2.75 (B). However, there is significant variability in Particle Distribution, with subcategory A having a notably lower percentage (17.71%) than B (85.86%). Moisture Content varies from 16% (A) to 31% (B). The pH levels are fairly consistent and slightly acidic, ranging from 6.68 (B) to 7.02 (A). Salinity in Zone 2 is generally lower than in Zone 1, with values between 18.7 (A) and 24.1 (C) ppt.

In Zone 3, the Specific Gravity is similar across subcategories, ranging from 2.62 (A) to 2.77 (C). The Particle Distribution varies widely, from as low as 10.34% (A) to 84.89% (C), indicating a

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

diverse soil composition. Moisture Content is lowest in subcategory A at 6% and highest in C at 33%. The pH levels indicate more acidic conditions in Zone 3, ranging from 6.07 (C) to 7.31 (A). Salinity also varies, with subcategory C showing the highest value at 25.8 ppt, similar to Zone 1, while subcategory A has the lowest at 21.5 ppt.

The sediment analysis revealed that Zone 3 had the highest silt-clay content, reaching up to 84.89% in subzone 3C. This high concentration of fine particles correlates with poor drainage and reduced aeration, which are detrimental to mangrove growth (Gao et al., 2023). Zone 1, on the other hand, had a more balanced particle distribution and neutral pH levels, making it more conducive to healthy mangrove development. Moisture content was also notably higher in Zone 3, creating waterlogged conditions unfavorable for mangroves. These findings demonstrate the critical relationship between sediment composition and mangrove health.

3.3 Normalised Result

In Table 3, the detailed analysis reveals the varying conditions of the mangrove area and sediment properties in the study area. Zone 1 stands out due to the robust tree growth and higher salinity in subcategory C, while Zone 2 displays strong growth characteristics in subcategory B. On the contrary, Zone 3's subcategory A indicates the least favorable conditions, with growth-related characteristics near zero, signifying an unsuitable environment for tree growth.

Table 3: Normalised Data of Each Zone

Zone		Normalised Characteristics				
		Height of Tree (m)	Trunk Diameter (m)	Particle Distribution %	pH Level	Salinity (ppt)
1	A	0.48	0.83	1.09	0.91	0.70
	B	0.46	0.92	0.98	0.36	0.71
	C	0.64	1.08	1.18	0.59	0.86
2	A	0.18	0.50	0.11	0.81	0.00
	B	0.64	1.25	1.15	0.52	0.57
	C	0.36	0.58	0.23	1.05	0.65
3	A	0.00	0.00	0.00	0.36	0.34
	B	0.03	0.50	0.06	0.00	0.20
	C	0.61	1.08	1.14	0.65	0.86

This research builds on existing knowledge by providing a detailed, site-specific analysis of sedimentation impacts on mangroves. Studies like Besset et al. (2019) have reported similar challenges, where excessive sedimentation restricts root oxygenation and nutrient absorption. However, this study's use of UAV photogrammetry adds a new layer of precision, offering highly accurate spatial and elevation data that enhance our understanding of these dynamics. For instance, the ability to generate Digital Terrain Models has provided clearer insights into the micro-topography influencing sediment deposition patterns.

The findings underscore the urgent need for interventions to mitigate sedimentation in mangrove ecosystems. Practical measures such as constructing sediment traps and restoring degraded areas can help manage sediment levels effectively (Horstman et al., 2014). Moreover, policymakers should focus on sustainable land-use practices to reduce sediment influx into coastal areas. By integrating advanced monitoring tools like UAV photogrammetry into conservation strategies, authorities can enhance the resilience of mangrove ecosystems and promote long-term ecological sustainability.

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

4. Conclusion

This study highlights how sedimentation impacts mangrove growth by altering soil and water conditions. Through the use of UAV photogrammetry and standardized sediment analysis, the research provides detailed insights into these effects. The findings support the development of practical sediment management strategies and stress the importance of conserving mangrove ecosystems. The research provides an in-depth analysis of how sediment transport impacts mangrove forests, revealing that excessive sediment buildup can alter soil properties, limiting nutrient and oxygen availability to mangrove roots. This, in turn, may impede growth and increase susceptibility to pests and diseases. These findings focus on the role of local authorities such as the Forestry Department and Department of Irrigation and Drainage on the necessity of effective management strategies, urging conservation efforts to prioritize reducing sediment pollution and restoring water flow patterns within mangrove ecosystems. The study emphasizes the importance of mangroves in coastal resilience and ecological balance, calling for continued research by other universities and dedicated conservation efforts to protect these vulnerable coastal habitats in Malaysia and elsewhere.

Challenges in collecting field data emerge from the dense structure of mangrove habitats, which makes it difficult to access due to thick vegetation. Furthermore, despite its potential for data collection, there are lingering doubts about the reliability of UAV monitoring for data collection. Additionally, the sampling procedure is limited by the need to carry out collections when tides are at their lowest, increasing the research project's logistical challenges. To ensure that the sample is collected in good condition, it is important to create a systematic and appropriate schedule for data collection in future studies. Furthermore, an appropriate timetable can help prevent situations like blurry or cached images. This is due to the significance of the images in producing a high-quality ortho mosaic, which is crucial for generating precise data for DSM and DTM.

5. Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia through Tier 1 (Vot Q970) and GPPS (Vot Q303).

6. References

Adrian L., Allan M. (2000). On the significance of the H/E ratio in wear control: a nanocomposite coating approach to optimised tribological behavior Wear 246:1-11.

Asp, N., Gomes, V., Schettini, C., Souza-Filho, P., Siegle, E., Ogston, A., Nittrouer, C., Silva, J., Nascimento, W., Souza, S., Pereira, L., & Queiroz, M. (2018). Sediment dynamics of a tropical tide-dominated estuary: Turbidity maximum, mangroves and the role of the Amazon River sediment load. Estuarine Coastal and Shelf Science, 214, 10–24. https://doi.org/10.1016/j.ecss.2018.09.004

ASTM D 422. (2014). Standard Test Method for Particle-Size Analysis of Soils. American Society of Testing and Materials.

ASTM E 1727. (2020). Standard Practice for Field Collection of Soil Samples for Subsequent Lead Determination. American Society of Testing and Materials.

Basyuni, M., Ramli, M. F., Ariffin, J., & Abdul Aziz, H. (2020). Impacts of sedimentation on mangrove ecosystems: A review. Sains Malaysiana, 49(3), 587-599.

Beretta, F., Shibata, H., Cordova, R., De Lemos Peroni, R., Azambuja, J., & Costa, J. F. C. L. (2018). Topographic modelling using UAVs compared with traditional survey methods in mining. REM - International Engineering Journal, 71(3), 463–470. https://doi.org/10.1590/0370-44672017710074

Besset, M., Gratiot, N., Anthony, E. J., Bouchette, F., Goichot, M., & Marchesiello, P. (2019). Mangroves and shoreline erosion in the Mekong River delta, Viet Nam. Estuarine Coastal and Shelf Science, 226, 106263. https://doi.org/10.1016/j.ecss.2019.106263

Chang, C. W., & Mori, N. (2021). Green infrastructure for the reduction of coastal disasters: a review of the protective role of coastal forests against tsunami, storm surge, and wind waves. Coastal Engineering Journal, 63(3), 370–385. https://doi.org/10.1080/21664250.2021.1929742

Cotton-Barratt, O., MacAskill, W., & Ord, T. (2020). Statistical normalization methods in interpersonal and intertheoretic comparisons. The Journal of Philosophy, 117(2), 61–95. https://doi.org/10.5840/jphil202011725

Dietrich, W. E., Day, G. A., Parker, G., & Hobbs, B. F. (2020). The physics and evolution of subaqueous sand dunes. Annual Review of Fluid Mechanics, 52, 29-59.

Earle, S., & Earle, S. (2015, September 1). 17.3 Landforms of Coastal Deposition. Pressbooks. https://opentextbc.ca/geology/chapter/17-3-landforms-of-coastal-deposition/.

Gao, J., Chen, L., Lu, H., Meng, X., Cao, X., Wang, D., Xie, L., Kwan, K., & Hu, B. (2023). The dynamic mechanism of rapid sediment deposition in the mangrove region of the Guangxi Beibu Gulf. Ocean Dynamics, 73(12), 743–760. https://doi.org/10.1007/s10236-023-01579-3

Hai, N., Dell, B., Phuong, V., & Harper, R. (2020). Towards a more robust approach for the restoration of mangroves in Vietnam. Annals of Forest Science, 77(1). https://doi.org/10.1007/s13595-020-0921-0

Hassan, M. I., & Rahmat, N. H. (2016). THE EFFECT OF COASTLINE CHANGES TO LOCAL COMMUNITY'S SOCIAL-ECONOMIC. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W1, 25–36. https://doi.org/10.5194/isprs-archives-xlii 4-w1-25-2016

Horstman, E., Dohmen-Janssen, C., Bouma, T., & Hulscher, S. (2014). Tidal-scale flow routing and sedimentation in mangrove forests: Combining field data and numerical modelling. Geomorphology, 228, 244–262. https://doi.org/10.1016/j.geomorph.2014.08.011

Kaamin, M., Daud, M. E., Sanik, M. E., Ahmad, N. F. A., Mokhtar, M., Ngadiman, N., & Yahya, F. R. (2018). Mapping shoreline position using unmanned aerial vehicle. AIP Conference Proceedings. https://doi.org/10.1063/1.5055465

Komar, P. D. (2017). Beach Processes and Sedimentation. 3rd ed. Routledge.

Lancaster, N. (2019). Aeolian Geomorphology: An Introduction [Wiley].

McLachlan, Andrew, Alan C. Brown, and Thomas A. Schlacher. (2021). The Biology of Beaches and Coasts. 3rd ed. Oxford University Press.

Minor, M. L., Zimmer, M., Helfer, V., Gillis, L. G., & Huhn, K. (2021). Flow and sediment dynamics around structures in mangrove ecosystems—a modeling perspective. In Elsevier eBooks (pp. 83–120). https://doi.org/10.1016/b978-0-12-816437-2.00012-4

Mohanta, Pradhan, B., & Sahu, S. (2020). Assessment of species diversity and physicochemical characteristics of mangrove vegetation in Odisha, India. In Elsevier eBooks (pp. 135–151). https://doi.org/10.1016/b978-0-12-819532-1.00006-8

Mohd Razi, N. A. A., Mohd Razif, N. H., Ahmad, A., & A. Aziz, N. (2019). Evaluating the effectiveness of sedimentation pond and sediment trap for river management: A review. Malaysian Journal of Civil Engineering, 31(3), 284-297.

Mokhtar, M., Daud, M. E., Kaamin, M., Sabri, M. S., Hayazi, F., Hisham, M. H., Siddiq, M. I., Azmi, M. a. M., Hamid, N. B., & Noor, H. M. (2023). Beach profile and shoreline sediment properties at eroded area in Batu Pahat. Construction Technologies and Architecture. https://doi.org/10.4028/p-ny3700

Mokhtar, M., Hayazi, F., Hamzah, A. S. A., Isnin, M. S. S., Kaamin, M., Azmi, M. A. M., & Hamid, N. B. (2020). Condition Survey of Eroded Area in Pantai Punggur, Malaysia Shoreline Using UAV Photogrammetry. Solid State Technology, 63(3), 826-833.

Ndungu, L. K., Steele, J. H., Hancock, T. L., Bartleson, R. D., Milbrandt, E. C., Parsons, M. L., & Urakawa, H. (2019). Hydrogen peroxide measurements in subtropical aquatic systems and their implications for cyanobacterial blooms. Ecological Engineering, 138, 444–453. https://doi.org/10.1016/j.ecoleng.2019.07.011

Nickling, W. G. (2019). "Aeolian Sediments and Processes." In Coastal Sediments, edited by S. L. Namikas and J. D. Rosati, 161-183. Springer.

Pähtz, T., & Durán, O. (2020). Unification of Aeolian and Fluvial Sediment Transport Rate from Granular Physics. Physical Review Letters, 124(16). https://doi.org/10.1103/physrevlett.124.168001

Parker, G. (2016). Physics of sediment transport by wind and water. Annual Review of Fluid Mechanics, 48, The :711-738.

Pazi, A. M. M., Khan, W. R., Nuruddin, A. A., Adam, M. B., & Gandaseca, S. (2021). Development of Mangrove sediment Quality Index in Matang Mangrove Forest Reserve, Malaysia: A synergetic approach. Forests, 12(9), 1279. https://doi.org/10.3390/f12091279

Perminov, Nikolay & Smirnov, Maksim & Nigmatullin, Raoul & Talipov, Anvar & Moiseev, Sergey. (2017). SRA vs histograms in the noise analysis of SPAD.

Ruzgienė, B., Berteška, T., Gečyte, S., Jakubauskienė, E., & Aksamitauskas, V. Č. (2015). The surface modelling based on UAV Photogrammetry and qualitative estimation. Measurement, 73, 619–627. https://doi.org/10.1016/j.measurement.2015.04.018

Sabri, M. S., Daud, M. E., & Mokhtar, M. (2022). Effect of soil shear strength on shoreline changes at Batu Pahat coastal area. Journal of Sustainable Underground Exploration, 2(1). doi: 10.30880/jsue.2022.02.01.002.

Saintilan, N., Khan, N. S., Ashe, E., Kelleway, J. J., Rogers, K., Woodroffe, C. D., & Horton, B. P. (2020). Thresholds of mangrove survival under rapid sea level rise. Science, 368(6495), 1118–1121. https://doi.org/10.1126/science.aba2656

Salmo, S. G., Malapit, V., Garcia, M. C. A., & Pagkalinawan, H. M. (2019). Establishing rates of carbon sequestration in mangroves from an earthquake uplift event. Biology Letters, 15(3), 20180799. https://doi.org/10.1098/rsbl.2018.0799

Siddiq, M. I., Daud, M. E., Kaamin, M., Mokhtar, M., Omar, A. S., & Duong, N. A. (2022). Investigation of Pantai Punggur Coastal Erosion by using UAV Photogrammetry. International Journal of Nanoelectronics & Materials, 15, 61-69.

Siddiq. "THE INVESTIGATION OF SHORELINE CHANGES BY USING UAV PHOTOGRAMMETRY AT PANTAI PUNGGUR ." Faculty of Civil and Built Environment Universiti Tun Hussein Onn Malaysia , July 2023.

Singh, J. K. (2020). Structural characteristics of mangrove forest in different coastal habitats of Gulf of Khambhat arid region of Gujarat, west coast of India. Heliyon, 6(8), e04685. https://doi.org/10.1016/j.heliyon.2020.e04685

Wang, J., Sun, J., Yu, Z., Li, Y., Tian, D., Wang, B., Li, Z., & Niu, S. (2019). Vegetation type controls root turnover in global grasslands. Global Ecology and Biogeography, 28(4), 442–455. https://doi.org/10.1111/geb.12866

Xie, D., Schwarz, C., Kleinhans, M. G., Zhou, Z., & Van Maanen, B. (2022). Implications of coastal conditions and Sea-Level rise on mangrove vulnerability: A Bio-Morphodynamic Modeling study. Journal of Geophysical Research Earth Surface, 127(3). https://doi.org/10.1029/2021jf006301

DOI: https://doi.org/10.22452/mjs.vol45sp1.7
Malaysian Journal of Science 45 (Special Issue 1): 58-65 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

https://mjs.um.edu.my

In Silico Characterization of Selected Carbohydrate Active Enzymes (CAZymes) from Malaysia's Unique Genome Assemblies Through Genomic Data Mining

Alvin Owen Miharil@Jukiri[1a], Aisyah Mohamed Rehan[2a*], Azzmer Azzar Abdul Hamid[3b] and Kamarul Rahim Kamarudin[4c]

Abstract: In silico characterization of protein is usually performed to elucidate the structure and function of protein targets prior to experimental approach, which helps to narrow down target candidates, and reduce overall time, cost and effort. Carbohydrate-active enzymes (CAZymes) are enzymes involved in conversion of biomass to industrially valuable products. More research is needed to characterise CAZymes from Malaysia. This study aims to elucidate CAZymes from selected genome assemblies isolated in Malaysia by in silico genomic data mining. The National Center for Biotechnology Information database is utilized to investigate Malaysia's genome assemblies, discovering 1386 complete genomes. From 2019 to 2024, there are 34 bacterial genomes isolated by Malaysian researchers. CAZymes from six bacterial genomes were categorised. Three CAZymes from *Thermobifida fusca* were chosen for subsequent protein modelling due to their acceptable homology range to available protein structures. To characterise the structural and functional properties of these CAZymes, online protein modelling tools were used (dbCAN3 meta server, I-TASSER and ExPASy SWISS-MODEL), as well as to evaluate the quality of the protein models (Ramachandran plot assessment, Verify3D and QMEAN4 score). Then, the ligand-binding sites inside each CAZyme sequence were predicted using I-TASSER's COACH algorithm. Three target CAZymes annotated as endo-1,4-beta-xylanase, amylo-alpha-1,6-glucosidase and endoglucanase respectively have been modelled, their protein model assessed for its quality, and their active sites and potential ligand predicted. Through in-silico characterization, the proteins' structure and function have been elucidated. Future studies, employing refined tertiary models, protein-ligand docking, and molecular dynamics simulations, aim to identify novel inhibitors or activators. This will facilitate experimental validation and unlock diverse industrial biotechnology applications. Potential benefits include enhanced enzyme catalysis for biofuel production, development of novel biopharmaceuticals, and optimization of metabolic pathways for sustainable chemical synthesis. These findings could significantly impact industrial processes and contribute to bio-based economies.

Keywords: CAZymes, In-silico Protein Modeling, Genomic Data Mining

1. Introduction

Carbohydrates are a major class of large biopolymers found in all cells, along with nucleic acids, proteins, and lipids. Carbohydrate-active enzymes (CAZymes) are enzymes that can degrade, modify, or create glycosidic bonds. These enzymes are essential catalysts in numerous biological processes, driving the breakdown and synthesis of complex carbohydrates. CAZymes are divided into several classes and are secreted by microorganisms, allowing the microbial conversion of plant biomass into industrially valuable products like bioenergy and biofuel production (Wardman et al., 2022). This microbial activity complements the plant's own carbohydrate processing pathways. Plants use photosynthesis to convert carbon dioxide and water into sugars, which are further converted into carbohydrates like starches and cellulose (Sharkey, 2024). CAZyme genes are abundant in plant and plant-degrading microbes' genomes. Precursors produced by these enzymes can be used to generate bio-based products like food, paper, textiles, animal feed, and biofuels (Bains et al., 2024).

CAZymes are crucial in bioenergy, agriculture, human nutrition, and disease prevention. They break down complex carbohydrates in the gut microbes, allowing them to be absorbed by the intestinal epithelium (Lee et al., 2022). CAZymes are also involved in biological processes underlying diseases like cancer, diabetes, Alzheimer's, and AIDS (Ye et al., 2022; Nin-Hill et al., 2023). This broad range of functional roles underscores the importance of exploring CAZymes from diverse sources, with significant scientific and industrial interest.

Due to Malaysia's biodiversity, bioprospecting and discovering new products are extensively performed by scientists and commercial companies. Bioinformatic tools are utilized to enable these discoveries. The CAZy database, dbCAN3 meta server (Zheng et al., 2023), and PlantCAZyme database have

Authors information:
[a]Department of Chemical Engineering Technology, Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub, KM 1, Jalan Panchor, 86400, Panchor, Johor, MALAYSIA. E-mail: an200042@student.uthm.edu.my[1], aisyahr@uthm.edu.my[2]
[b]Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia, Kuantan Campus, Bandar Indera Mahkota, 25200 Kuantan, Pahang, MALAYSIA. E-mail: azzmer@iium.edu.my[3]
[c]Department of Technology and Natural Resources, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub, KM 1, Jalan Panchor, 84600 Panchor, Johor, MALAYSIA. E-mail: kamarulr@uthm.edu.my[4]

**Corresponding Author: aisyahr@uthm.edu.my*

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

assisted in the genome mining of CAZymes from bacteria, fungi, viruses, and plants (Zhang et al., 2018; Yang et al., 2024). Malaysia's unique genome assembly has led to the discovery of a novel beta-glucosidase from *Jeotgalibacillus Malaysiensis*, a marine bacterium (Liew et al., 2018; Godse et al., 2021). This is similar to a genomic mining study on *Glaciozyma antarctica*, a psychrophilic yeast from Antarctica (Nooraisyah et al., 2015). Over the years, advances have been made to unearth the potential of psychrophilic enzymes studied from this yeast and its cold adaptation strategies (Yusof et al., 2021).

While CAZymes are implicated in diverse fields, including bioenergy and disease, their potential within Malaysia's unique biodiversity remains largely unexplored. Existing databases like CAZy and dbCAN3 facilitate genome mining, yet a significant gap exists in the in-silico elucidation of novel CAZymes from Malaysian genome assemblies. This research aims to provide in-silico characterization of CAZymes from selected Malaysian genome assemblies, offering new insights for industrial biotechnology and biomedical applications, leveraging Malaysia's untapped genomic resources.

2. Materials and methods

2.1 Genome Sequence Assembly

In this research, the main approach used was reference-based genome assembly. The complete genome sequence assembly for bacteria, fungi, virus, and plant division submitted by Malaysian depositors in the NCBI database and NCBI GenBank between 2019 and until the end of 2024 were used (Sayers et al., 2022). Filters were activated to exclude anomalous and partial assembly. Animal genome assemblies were also excluded. Protein FASTA sequences for each assembly were downloaded.

2.2 CAZymes annotation

Protein FASTA sequences from each assembly were then submitted to the dbCAN3 meta server to annotate the proteins using HMMER and DIAMOND via CAZy, dbCAN, and PPR tools, respectively (Zhang et al., 2018; Sayers et al., 2022). The results from the annotation tools on the dbCAN3 meta server were retrieved from the dbCAN3 results page (Zheng et al., 2023).

2.3 Selection of target Novel CAZymes

Genome assemblies for potential novel CAZymes, which are crucial for understanding glucose metabolism, carbohydrate breakdown, and organisms' interactions with carbohydrates, were selected.

2.4 Protein modelling of target Novel CAZymes

3D protein modeling will be used to determine the structure of CAZymes that lack existing structural data. BLASTp will be used to check the availability of structural homologs in Protein Data Bank (PDB). The protein tertiary structure will be predicted using three different server which were I-TASSER (Zhou et al., 2022) and ExPASy SWISS-MODEL (Waterhouse et al., 2024). All three predicted structures will then be validated using Ramachandran plot assessment, Verify3D and QMEAN4 score. From the validation, the best predicted structure will be selected for structural refinement. Model structure refinement will be performed using ModRefiner algorithm tool from I-TASSER database (Zhou et al., 2022).

2.5 Ligand prediction & docking experiment

Ligand prediction in CAZymes involves predicting potential substrates with which enzymes may bind or act. Machine-learning-based ligand predictors use techniques like support vector machines (SVM) and decision trees to identify critical residues, geometries, and conformations. Molecular docking is used to visualize binding modes and determine binding affinities. Popular docking applications like AutoDock, GOLD, and Glide help researchers understand enzyme-ligand pairings, leading to the design or optimization of new CAZymes for specific applications. This research will use COACH algorithm from I-TASSER server suite that generates ligand binding-site predictions by matching the target models with proteins in the BioLiP database (Zhou et al., 2022).

3. Results and discussions

3.1 Genome assemblies deposited by Malaysian researcher

The NCBI database is a valuable tool for acquiring genome sequence assemblies focusing on Malaysia's unique genetic composition. Researchers can access a unique reservoir of genetic data, providing insights into the region's biodiversity. A search using the NCBI database server showed 1386 complete genome assemblies by Malaysian scientists, with filters used to narrow the search and a refined selection process focusing on bacterial genomes. The search resulted in 34 complete bacterial genome assemblies, one from viruses, and zero from plants (Table 1, Sayers et al., 2022).

3.2 Annotation of CAZymes families

The study focused on bacterial genome assemblies due to time constraints and their potential as future CAZyme producers. Six randomly chosen bacterial genomes were analyzed using the NCBI dataset to identify potential advantages (Table 2). Direct NCBI annotation was used to retrieve and calculate CAZyme numbers within the genome assemblies. The dbCAN3 database was used to confirm NCBI annotations (Zheng et al., 2023). The NCBI genome assembly search was filtered for complete genome assemblies, resulting in complete annotations for each genome sequence. However, compiling annotations was challenging due to different specific CAZyme names and families (Sayers et al., 2022).

Only two genomes, *Thermobifida fusca* and *Parageobacillus caldoxylosilyticus*, were chosen for structure and function prediction after analysis and risk assessment of their CAZyme annotations. The exclusion criteria were to remove bacteria that is pathogenic to humans or can cause infection to human. This is critical for handling researchers' safety as well as for assessing research feasibility with low risk. These chosen genomes, recognised for their efficient breakdown abilities and resistance to high temperatures, respectively, provide potential avenues for industrial and environmental biotechnology advancements.

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

Table 1. List of 34 genome sequence assemblies with complete genome

Genome Assembly ID:		Species name:
1.	**ASM1503458v1**	*Thermobifida fusca* (high G+C Gram-positive bacteria*)*
2.	**ASM1927293v1**	*Parageobacillus caldoxylosilyticus* (firmicutes)
3.	**ASM2520084v1**	*Pseudomonas aeruginosa* (g-proteobacteria)
4.	**ASM159764v2**	*Mycobacterium tuberculosis* (high G+C Gram-positive bacteria)
5.	**ASM1907668v1**	*Acinetobacter baumannii* (g-proteobacteria)
6.	**ASM199636v2**	*Vibrio parahaemolyticus* (g-proteobacteria)
7.	**ASM3070428v1**	*Vibrio parahaemolyticus* (g-proteobacteria)
8.	**ASM2542621v1**	*Stenotrophomonas maltophilia* (g-proteobacteria
9.	**ASM967662v1**	*Limosilactobacillus fermentum* (firmicutes)
10.	**ASM1735231v1**	*Priestia megaterium* (firmicutes)
11.	**ASM2954238v1**	*Streptococcus suis* (firmicutes)
12.	**UKMPMC2000**	*Burkholderia pseudomallei* (b-proteobacteria)
13.	**UKMD286**	*Burkholderia pseudomallei* (b-proteobacteria)
14.	**ASM1421793v1**	*Shigella sonnei* (enterobacteria)
15.	**ASM2429712v1**	*Streptomyces cavourensis* (high G+C Gram-positive bacteria)
16.	**ASM2458281v1**	*Acinetobacter colistiniresistens* (g-proteobacteria)
17.	**ASM3207508v1**	*Komagataeibacter nataicola* (a-proteobacteria)
18.	**ASM970730v1**	*Rhodococcus sp. AQ5-07* (high G+C Gram-positive bacteria)
19.	**ASM2580919v1**	*Paenibacillus sp. PSB04* (firmicutes)
20.	**ASM3132634v1**	*Aureispira sp. CCB-E* (CFB group bacteria)
21.	**ASM2241457v2**	*Streptomyces sp. MUM 178J* (high G+C Gram-positive bacteria)
22.	**ASM3277389v1**	*Teredinibacter sp. KSP-S5-2* (g-proteobacteria)
23.	**ASM2054666v1**	*Marinobacter sp. CA1* (g-proteobacteria)
24.	**ASM432894v1**	*Aquitalea sp. USM4* (b-proteobacteria)
25.	**ASM2615116v2**	*Cryobacterium sp. SO2* (high G+C Gram-positive bacteria)
26.	**ASM421021v2**	*Cryobacterium sp. SO1* (high G+C Gram-positive bacteria)
27.	**ASM970728v1**	*Arthrobacter sp. AQ5-05* (high G+C Gram-positive bacteria)
28.	**ASM2996405v1**	*Arthrobacter sp. EM1* (high G+C Gram-positive bacteria)
29.	**ASM2315155v2**	*Photobacterium sp. CCB-ST2H9* (g-proteobacteria)
30.	**ASM1340350v1**	*Acinetobacter baumannii* (g-proteobacteria)
31.	**UKMR15**	*Burkholderia pseudomallei* (b-proteobacteria)
32.	**ASM2995351v1**	*Bacillus altitudinis* (firmicutes)
33.	**ASM2124932v1**	*Mycolicibacterium fortuitum* (high G+C Gram-positive bacteria)
34.	**ASM452194v2**	*Streptomyces sp. S6* (high G+C Gram-positive bacteria)

3.3 CAZymes from Malaysia's unique genome assemblies

This study analyzed target genome assemblies for potential novel CAZymes, which are crucial for understanding glucose metabolism, carbohydrate breakdown, and organisms' interactions with carbohydrates. Protein modelling of these CAZymes was performed using in silico tools like I-TASSER (Zhou et al., 2022) and ExPASy SWISS-MODEL (Waterhouse et al., 2024) for protein structure prediction and analysis. The Ramachandran plot evaluation, Verify3D, and QMEAN4 score were used for quality validation of the structural models, ensuring a reliable protein model for active sites and ligand binding site prediction. This research could aid in developing new diagnostic tools for infectious illnesses and microbial-related ailments.

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

Table 2. List of six genome sequence assemblies annotated based on their CAZymes

ID no.	Species Name	CAZymes families* (number of gene)					Benefits
		GHs	GTs	PLs	CEs	AAs	
ASM1503458v1	*Thermobifida fusca*	14	30	0	0	1	• Decomposition of Organic Matter • Production of Cellulases • Biotechnological Applications
ASM1927293v1	*Parageobacillus caldoxylosilyticus*	13	17	0	0	0	• High-Temperature Applications • Enzyme Production Xylanolytic Activity
ASM2520084v1	*Pseudomonas aeruginosa*	4	24	0	0	0	• Biodegradation • Antimicrobial Properties • Research Value
ASM159764v2	*Mycobacterium tuberculosis*	6	22	2	4	0	• Improved Diagnosis and Treatment • Public Health Benefits
ASM1907668v1	*Acinetobacter baumannii*	1	14	6	2	0	• Understanding Antibiotic Resistance • Importance in Infection Control
ASM3070428v1	*Vibrio parahaemolyticus*	4	23	15	1	0	• Improved Food Safety • Importance of Public Health Surveillance

*CAZymes families abbreviations: Glycoside Hydrolases= GHs, GlycosylTransferases = GTs, Polysaccharide Lyases = PLs, Carbohydrate esterases = Ces, Auxiliary Activities = Aas

3.4 Selection of target novel CAZymes

Research on novel Carbohydrate-Active Enzymes (CAZymes) for organisms like *Thermobifida fusca* and *Parageobacillus caldoxylosilyticus* is crucial for industrial biotechnology. These enzymes exhibit specific capabilities for breaking down complex polysaccharides, leading to advancements in biofuel production, waste management, and environmentally friendly solutions, leveraging microorganisms' natural proficiency in biomass degradation.

Thermophilic cellulolytic bacteria, like *Thermobifida fusca*, have potential industrial and biotechnological uses. Malaysian scientists analyzed the species' genome sequence, revealing 112 carbohydrate-active enzyme genes, including 40 glycoside hydrolases (GHs), 36 glycosyltransferases (GTs), 22 carbohydrate-binding modules (CBMs), nine carbohydrate esterases (CEs), three polysaccharide lyases (PLs) and two auxiliary activities (AAs) (Yildiz et al., 2022). Two genes for lytic polysaccharide monooxygenase (LPMO) were also discovered. The ninth genome was used to investigate potential industrial and agricultural uses. Three novel CAZymes were selected from 109: xylanase, glucosidase, and endoglucanase. These enzymes are essential in lignocellulose-degrading microorganisms, breaking down complex plant polysaccharides into simpler sugars. Their synergistic activity is crucial for effective biomass deconstruction, a crucial step in bioeconomy's efforts to exploit renewable resources (Nargotra et al., 2022).

Parageobacillus caldoxylosilyticus is a unique species from the thermophilic genus Parageobacillus (Wang et al., 2022). Because of its high adaptability towards temperature, this bacterium is a promising candidate to produce industrially applicable thermostable enzymes, such as xylanase and maltogenic amylase (Wang et al., 2022). This makes it an interesting subject for researchers studying bioconversion processes and biofuel generation. It also has potential biotechnological uses beyond biofuel production, such as bioethanol and biochemical synthesis. However, due to a lack of genomic knowledge, the extent of its powers and underlying genetic pathways is still in its infancy. *Parageobacillus thermoglucosidasius*, with high sequence similarities to *P. caldoxylosilyticus*, possesses more research findings on its genome assembly, offering potential research and application opportunities (Wang et al., 2022).

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

Table 3. The Carbohydrate hydrolysis enzymes (GH) of T. fusca

GH family	Product size (No of amino acid)	Protein/enzymes	Accession number
GH10	399	Endo-1,4-beta-xylanase	QOS59282.1
GH176	712	Amylo-alpha-1,6-glucosidase	QOS59445.1
GH6, CBM2	441	Endoglucanase	QOS57731.1

3.5 Protein modelling target novel CAZymes

Protein modelling is being used to study novel Carbohydrate-Active enzymes (CAZymes), providing a new approach to understanding and using these biological catalysts for biofuel production and medicine creation. Scientists use advanced bioinformatics methods to make 3D modelling predictions, using platforms like I-TASSER and ExPASy SWISS-MODEL. The models' correctness and dependability are assessed using computational tests like Ramachandran plot analysis, Verify3D, and QMEAN4 score (Zhou et al., 2022; Waterhouse et al., 2024).

3.6 Identification of target CAZymes genes on Thermobifida fusca

Researchers have identified Carbohydrate-Active Enzymes (CAZymes) in *Thermobifida fusca*, a bacterium known for its cellulolytic capabilities. Three CAZymes, xylanase, glucosidase, and endoglucanase, were chosen based on their high homologous fraction to well-characterized enzymes (Table 3). This discovery opens up new possibilities for biofuel generation and plant biomass biodegradation, enhancing environmental sustainability and renewable energy sources.

3.7 Determination of protein model quality

Protein models created using computational techniques are crucial for predicting active sites and simulating ligand binding. Verify3D, a tool on ExPASy SWISS-MODEL, helps scientists examine the structural integrity and functional usefulness of modelled CAZymes by comparing their 3D atomic models' suitability and correctness to their amino acid sequences. The correctness is determined by the environments of amino acid residues, such as the area of residue buried within the model, the fraction of side chain area covered by polar atoms, and the local secondary structure. This method allows scientists to verify structural predictions and ensure enzymes function like biological counterparts. By confirming the structural validity of these models, strengthens the basis for subsequent analyses, such as ligand docking and molecular dynamics simulations, ultimately facilitating the development of biotechnological and industrial applications. Figure 1 and Figure 2 shows the 3D modelling output for each side of the CAZymes.

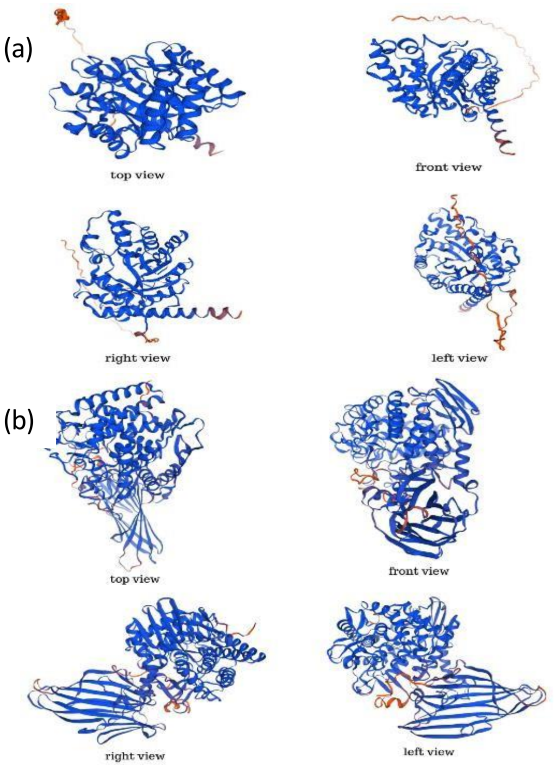


Figure 1. 3D modelling of endo-1,4-beta-xylanase (a) and 3D modelling of amylo-alpha-1,6-glucosidase (b)

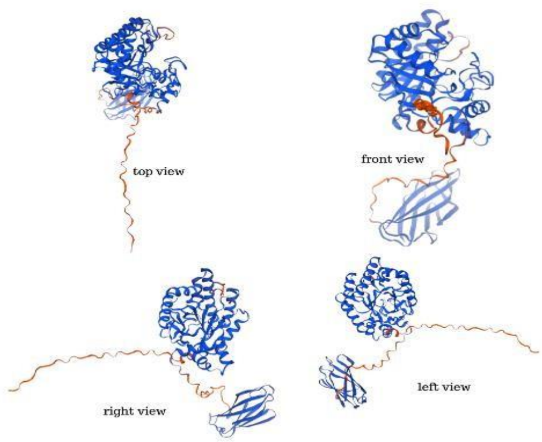


Figure 2. 3D modelling of Endoglucanase

3.8 Ramanchandran plot assessment

The Ramachandran plot is crucial for understanding the conformational dynamics of carbohydrate-active enzymes (CAZymes), which perform various biological tasks. By studying phi and psi torsion angles within CAZymes, researchers can determine structural flexibility and rigidity for substrate identification and catalytic activity. This analytical technique has enabled the discovery of structure-function correlations within

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

CAZyme families, thereby optimizing their applications in industrial and biotechnological processes. Figure 3 presents the Ramachandran plots for each modeled CAZyme, with detailed statistical summaries provided in Table 4. Notably, the xylanase model, depicted in Figure 3(a), exhibits the most favorable Ramachandran plot characteristics, demonstrating a lower number of outlier residues compared to the other CAZymes.

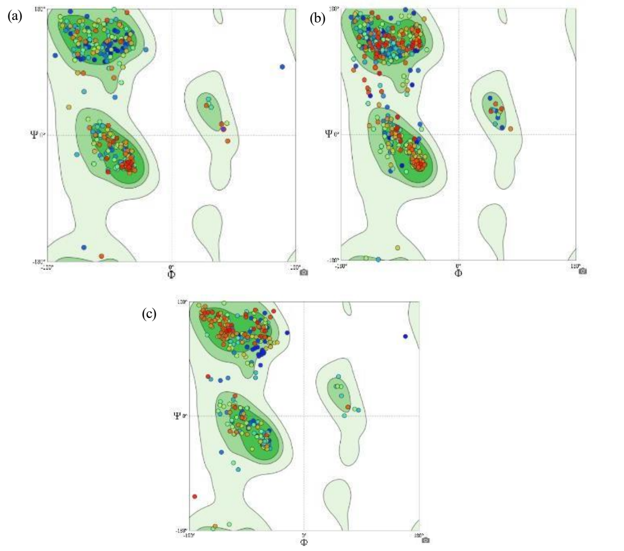


Figure 3. Ramachandran plot for Xylanase (endo-1,4-beta-xylanase) (a), Glucosidase (amylo-alpha-1,6-glucosidase) (b), and Endoglucanase (c)

This observation indicates a more energetically favorable arrangement of amino acid residues in the xylanase structure, suggesting higher model quality and potentially greater reliability in predicting its functional behavior. The reduced number of outliers signifies a well-defined secondary structure, which is crucial for accurate ligand binding and catalytic activity predictions. This superior conformation in xylanase reinforces its potential for targeted industrial applications, where precise structural understanding is paramount for enzyme engineering and optimization.

3.9 QMEAN4 score

The QMEAN4 score is a composite measure of protein model quality and structural reliability, indicating their accuracy. It is crucial in understanding the molecular mechanism of carbohydrate-activated enzymes (CAZymes), which are essential for carbohydrate breakdown, biosynthesis, and modification. An accurate QMEAN4 score indicates a well-structured and reliable model, impacting bioengineering experiments, biotechnological applications, and drug discovery. Figure 4 shows the output of QMEAN4 score for each of the CAZymes. Endoglucanase and xylanase show that output of -043 and -0.92 almost reached the ideal score of QMEAN4 score which is 1 and 0. Therefore, the structure of endoglucanase and xylanase are of relatively high structural accuracy, albeit below the ideal score of 1. These scores suggest that the generated models, particularly for endoglucanase and xylanase, provide a reliable foundation for subsequent analyses. The relatively close scores to the ideal indicate that the models are well constructed and likely to represent the true structures.

The implications of these QMEAN4 scores are significant. Firstly, the confidence in the model accuracy strengthens the validity of our ligand prediction findings. Secondly, these reliable models can be used for virtual screening of potential inhibitors or activators, which could lead to novel drug candidates or improved industrial enzyme performance. Furthermore, these models can guide site-directed mutagenesis experiments, enabling the fine-tuning of enzyme activity for specific biotechnological applications. While the QMEAN4 scores indicate good model quality, it's essential to acknowledge that they are in-silico predictions. Therefore, experimental validation, such as X-ray crystallography or NMR spectroscopy, is necessary to confirm the structural accuracy and refine our understanding of these CAZymes. This combined approach will maximize the potential of these enzymes for industrial and biomedical applications.

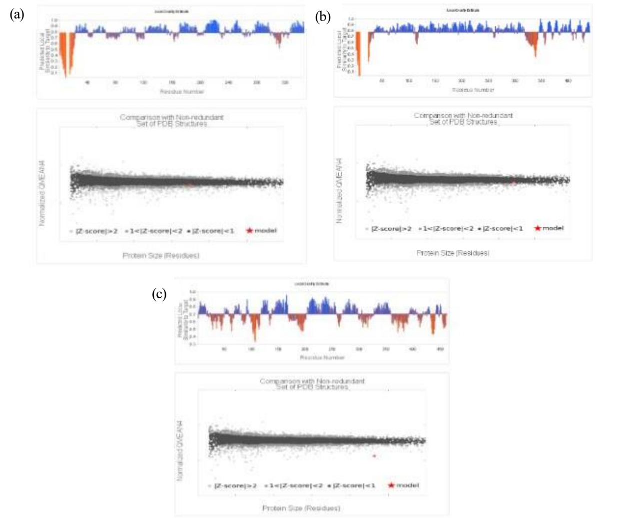


Figure 4. QMEAN4 score for Xylanase (endo-1,4-beta-xylanase) (a), Glucosidase (amylo-alpha-1,6-glucosidase) (b), and Endoglucanase (c)

3.10 Ligand site prediction

I-TASSER is a technique that predicts ligand site interactions with carbohydrate-active enzymes (CAZymes), which play a crucial role in carbohydrate breakdown, biosynthesis, and modification. By searching a large library of protein structures and improving models with ab initio folding simulations, I-TASSER helps understand how CAZymes interact with their substrates, revealing information about their catalytic processes (Zhou et al., 2022). This knowledge is essential for creating efficient enzymes for industrial uses like biofuel production and producing inhibitors for medicinal applications.

Figure 5 shows the ligand prediction for each of the CAZymes. The I-TASSER's COACH algorithm output on Xylanase protein model predicted a 137-cluster ligand (Figure 5(a)), designated 4-O-beta-D-xylopyranosyl-beta-D-xylopranose (BXP), with critical residues i.e. E48, N49, K52, H85, W89, Q92, N133, E134, Q209,

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

Table 4. Ramachandran plot detail for Xylanase (endo-1,4-beta-xylanase) (a), Glucosidase (amylo-alpha-1,6-glucosidase) (b), and Endoglucanase (c)

Positions	A257	A323
Torsion angles phi (φ) and psi (ψ)	-105.76 and 124.56	-97.85 and 0.56
Confidence	0.99	0.98
MolProbity score	1.11	
Clash score	0.82	
Ramachandran-favoured regions	94.83%	
Ramachandran outliers	0.52%	
Rotamer outliers	0.92%	

(a)

Positions	A428	A630
Torsion angles phi (φ) and psi (ψ)	-115.07 and 127.86	-95.84 and -4.04
Confidence	0.28	0.71
MolProbity score	2.34	
Clash score	3.21	
Ramachandran-favoured regions	86.55%	
Ramachandran outliers	6.29%	
Rotamer outliers	6.15%	

(b)

Positions	A402	A314
Torsion angles phi (φ) and psi (ψ)	-118.65 and 137.63	-65.91 and -40.07
Confidence	0.99	0.97
MolProbity score	1.11	
Clash score	0.32	
Ramachandran-favoured regions	91.80%	
Ramachandran outliers	1.82%	
Rotamer outliers	0.60%	

(c)

E240, W281, and W288, indicating their potential interactions and relevance in ligand binding.

Meanwhile, for the glucosidase, the predicted ligand site with a C-score of 0.10, indicated a 14-cluster ligand as alpha-D-glucopyranose (GCL) (Figure 5(b)). The binding region revealed residues i.e. R288, D289, W419, D421, H559, and E604, suggesting potential interactions with the glucoside enzyme. For endoglucanase ligand site prediction, the identified ligand is beta-D-glucopyranose (BGC), with a C-score of 0.28 and a 30-cluster size (Figure 5(c). The binding region revealed residues i.e. W72, Y104, P147, D148, S220, W262, K290, and D296 to be potentially involved in binding interactions.

These findings are significant for several reasons. Firstly, understanding the precise ligand-binding interactions of these CAZymes can facilitate the design of more efficient enzymes for industrial applications. For instance, optimized xylanases, glucosidases, and endoglucanases can enhance the breakdown of plant biomass in biofuel production, contributing to sustainable energy solutions. Secondly, the identified binding residues serve as potential targets for inhibitor development. This is particularly relevant in medicinal applications, where CAZyme inhibitors can be used to control carbohydrate metabolism in diseases like diabetes. Furthermore, the varying C-scores obtained for ligand predictions highlight the complexity of these interactions and the need for further experimental validation. The identified ligands and interacting residues provide a framework for future

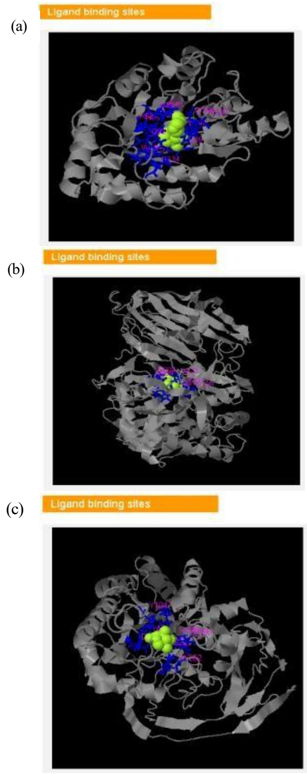


Figure 5. Ligand site prediction for Xylanase (endo-1,4-beta-xylanase) (a), Glucosidase (amylo-alpha-1,6-glucosidase) (b), and Endoglucanase (c)

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

molecular dynamics simulations and in vitro assays, which will refine our understanding of these enzymes and their potential applications. The detailed ligand predictions from this study underscore the power of in-silico methods in elucidating enzyme function and paving the way for targeted biotechnological interventions.

4. Conclusion

This study successfully mined the NCBI database for genome sequence assemblies, prioritizing bacterial candidates and ultimately focusing on *Thermobifida fusca* and *Parageobacillus caldoxylosilyticus* due to their demonstrated industrial relevance and biosafety profiles. The study identified three new carbohydrate-active enzymes from *Thermobifida fusca*, including endo-1,4-beta-xylanase, amylo-alpha-1,6-glucosidase, and endoglucanase. In-silico methodologies, including protein modeling and active site prediction, were employed to elucidate the structural and functional properties of these enzymes, with quality assessment confirming model reliability. Future research should prioritize refined tertiary structure modeling, detailed protein-ligand docking experiments, and comprehensive molecular dynamics simulations. These investigations will facilitate the identification of specific inhibitors or activators, crucial for optimizing enzyme activity. This refined understanding will directly support experimental validation and pave the way for targeted industrial applications, particularly in biomass conversion and bioprocessing.

5. Acknowledgement

The authors would like to thank the technical staff from the Chemical Engineering Technology Laboratory Cluster, Pagoh Campus Laboratory Management Office for their help and assistance. This research was supported by Universiti Tun Hussein Onn Malaysia through Tier 1 (Vote No.: H851).

6. References

Bains, R. K., Nasseri, S. A., Wardman, J. F., & Withers, S. G. (2024). Advances in the understanding and exploitation of carbohydrate-active enzymes, Current Opinion in Chemical Biology, 80, 102457.

Godse, R., Bawane, H., Tripathi, J., & Kulkarni, R. (2021). Unconventional β-glucosidases: a promising biocatalyst for industrial biotechnology, Applied Biochemistry and Biotechnology, 193, 2993-3016.

Lee, J. Y., Tsolis, R. M., & Bäumler, A. J. (2022). The microbiome and gut homeostasis. Science, 377(6601), eabp9960.

Liew, K. J., Lim, L., Woo, H. Y., Chan, K. G., Shamsir, M. S., & Goh, K. M. (2018). Purification and characterization of a novel GH1 beta-glucosidase from Jeotgalibacillus malaysiensis. International Journal of Biological Macromolecules, 115, 1094-1102.

Nargotra, P., Sharma, V., Lee, Y. C., Tsai, Y. H., Liu, Y. C., Shieh, C. J., ... & Kuo, C. H. (2022). Microbial lignocellulolytic enzymes for the effective valorization of lignocellulosic biomass: a review, Catalysts, 13(1), 83.

Nooraisyah, M. N., Siti Nur Hasanah, M. Y., Mahadi, N. M., Abu Bakar, F. D., & Murad, A. M. A. (2015). Genome Mining For Glycoside Hydrolases From The Psychrophilic Yeast Glaciozyma Antarctica PI12, Malaysian Applied Biology, 44(1).

Nin-Hill, A., Piniello, B., & Rovira, C. (2023). In silico modelling of the function of disease-related CAZymes, Essays in Biochemistry, 67(3), 355-372.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information, Nucleic acids research, 50(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Sharkey, T. D. (2024). The end game (s) of photosynthetic carbon metabolism, Plant Physiology, 195(1), 67-78.

Wang, Y., Wang, C., Chen, Y., Cui, M., Wang, Q., & Guo, P. (2022). Heterologous expression of a thermostable α-galactosidase from Parageobacillus thermoglucosidasius isolated from the lignocellulolytic microbial consortium TMC7, Journal of Microbiology and Biotechnology, 32(6), 749.

Waterhouse, A. M., Studer, G., Robin, X., Bienert, S., Tauriello, G., & Schwede, T. (2024). The structure assessment web server: for proteins, complexes and more, Nucleic Acids Research, gkae270.

Wardman, J. F., Bains, R. K., Rahfeld, P., & Withers, S. G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome, Nature Reviews Microbiology, 20(9), 542-556.

Yang, Q., Chang, S., Zhang, X., Luo, F., Li, W., & Ren, J. (2024). The fate of dietary polysaccharides in the digestive tract, Trends in Food Science & Technology, 104606.

Ye, S., Shah, B. R., Li, J., Liang, H., Zhan, F., Geng, F., & Li, B. (2022). A critical review on interplay between dietary fibers and gut microbiota, Trends in Food Science & Technology, 124, 237-249.

Yildiz, S. Y., Finore, I., Leone, L., Romano, I., Lama, L., Kasavi, C., Nicolaus, B., Oner, E. T., & Poli, A. (2022). Genomic Analysis Provides New Insights Into Biotechnological and Industrial Potential of Parageobacillus thermantarcticus M1, Frontiers in Microbiology, 13. https://doi.org/10.3389/fmicb.2022.923038

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Peter Kamp Busk, Xu, Y., & Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Research, 46(W1), W95–W101.

Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., & Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation, Nucleic Acids Research, 51(W1), W115-W121.

Zhou, X., Zheng, W., Li, Y., Pearce, R., Zhang, C., Bell, E. W., ... & Zhang, Y. (2022). I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction, Nature Protocols, 17(10), 2326-2353.

DOI: https://doi.org/10.22452/mjs.vol45sp1.8
Malaysian Journal of Science 45 (Special Issue 1): 66-74 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

https://mjs.um.edu.my

Lateral offset Displacement Fiber Optic Sensor for Detecting Adulteration of Pure Olive Oil with Palm Oil

Nur Ain Amila Binti Rizal[1a], Nurul Nadia Adnan[2a,b*], and Farah Diana Muhammad[3c]

Abstract: An optical fiber sensor was developed to detect the adulteration of pure olive oil with palm oil using a lateral offset displacement technique. Pure olive oil is a high-quality cooking oil often used in cooking, but its quality degrades due to adulteration. This study employed a single-mode fiber (SMF) displacement sensor, employing lateral offset distances of 1 µm, 2 µm, and 3 µm, to analyze the sensor's response to varying concentrations of palm oil in olive oil, ranging from 10% to 50%. The refractive indices of pure olive and palm oil were initially measured at 1.4698 and 1.4431, respectively, with a proportional decrease in refractive index observed as the concentration of palm oil increased. The maximum sensitivity of 0.888 dBm/% was achieved with a 3 µm lateral offset distance, highlighting the correlation between offset distance and the sensor sensitivity. The power loss increased consistently with higher concentrations of palm oil and greater offset distances, with the highest power loss of -5.2 dBm recorded at 50% adulteration and 3 µm offset. These results underscore the sensor's accuracy and sensitivity in detecting and quantifying adulteration, making it a valuable tool for quality control and ensuring the authenticity of high-value edible oils in the food industry.

Keywords: Fiber optic sensor, olive oil, adulteration, lateral offset, single-mode fiber

1. Introduction

Over the past decades, advancements in optical fiber technology have been remarkable, significantly impacting various fields, including sensing and measurement. The foundational discovery by Kao and Hockham in 1965, which demonstrated that impurity removal in glass could reduce light attenuation below 20 dB/km, laid the groundwork for developing low-loss optical fibers (Xie, 2022). Since then, continuous research and innovation have propelled the evolution of fiber optic sensors, which are now extensively used in detecting adulteration in various oils (Ansari et al., 2024; Srinivasulu & Rao, 2019).

Fiber optic sensors, known for their high sensitivity and accuracy, have been widely studied for identifying adulterants in oils such as lard in olive oil (Puspita et al., 2021), kerosene in petrol and diesel (Chauhan & Singh, 2022), and kerosene in petrol (Ferdous et al., 2022). These sensors typically rely on changes in refractive index or spectral shifts to detect adulteration levels. For example, a heptagonal photonic crystal fiber-based sensor using terahertz signals has demonstrated high sensitivity and low loss in detecting kerosene adulteration in petrol (Chauhan & Singh, 2022). Similarly, a multimode fiber optic sensor has effectively detected and quantified traces of palm oil adulterant in coconut oil by measuring wavelength shifts due to refractive index changes (Jadhav et al., 2022). These advancements underscore the potential of fiber optic sensors in ensuring the quality and authenticity of edible oils.

In this paper, we propose a novel lateral offset displacement approach in fiber optic sensors for detecting adulteration in pure olive oil (POO). Lateral offset in fiber optic cables involves deliberately misaligning or displacing optical fibers sideways, enhancing the sensor's capabilities by improving sensitivity and accuracy. This technique has been successfully applied to various sensing applications, including detecting hydrogen concentration (Zhang et al., 2022), chemical concentrations (Tian et al., 2022), curvature, and temperature changes (Liu et al., 2018).

Olive oil, derived from olives, is highly valued for its health benefits, including reducing the risk of cardiovascular diseases and inflammation due to its rich content of monounsaturated fats, phenolic compounds, and vitamin E (Guasch-Ferré et al., 2014; Nocella et al., 2017; Aparicio & Harwood, 2013). Its widespread use extends beyond culinary applications (Boskou, 2009) to cosmetics (Rodrigues et al., 2015), medicines (Caramia et al., 2012), and soaps (Girgis, 2003), making it a versatile product (Stanciu, 2023). However, olive oil is often adulterated with lower-quality oils such as palm (Rohman & Man, 2010), paraffin

Authors information:
[a]Department of Physics and Chemistry, Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Educational Hub, 84600, Pagoh, Johor, Malaysia.
E-mail: hw230011@student.uthm.edu.my[1],
nadia@uthm.edu.my[2]
[b]Photonics Devices and Sensor Research (PDSR), Department of Physics and Chemistry, Faculty of Applied Physics and Chemistry, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, Educational Hub, 84600, Pagoh, Johor, Malaysia. E-mail: nadia@uthm.edu.my[2]
[c]Department of Physics, Faculty of Science, University Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia.
E-mail: farahdiana@upm.edu.my[3]

*Corresponding Author: nadia@ uthm.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.9
Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

(Bunaciu et al., 2022), and used cooking oil (Li et al., 2018), driven by economic incentives to increase profits.

Detecting adulterants like palm oil in olive oil is crucial for maintaining pure olive oil's quality, authenticity, and health benefits. Palm olein oil, a cost-effective liquid fraction derived from palm oil (Elaeis guineensis), is commonly used in the food industry due to its high yield and low cost (Padial-Jaudenes et al., 2020; Deffense, 1985; Mamat et al., 2005). However, concerns regarding its high saturated fat content (Imoisi et al., 2015), environmental impact (Beyer et al., 2020), and associated social issues highlight the importance of reliable detection methods.

Previous studies have shown the potential of fiber optic sensors for detecting adulteration in edible oils. However, many of these methods rely on complex setups, focus on specific types of adulterants, or struggle to detect a broad range of concentrations. They often lack the sensitivity needed for identifying low levels of adulteration and face challenges regarding real-time applications.

As mentioned earlier, the novelty of this study lies in its lateral offset displacement technique, which enhances sensor sensitivity and precision. This method offers a linear, repeatable response to adulteration levels with minimal environmental interference and without complex sample preparation. Additionally, the sensor's reliance on straightforward refractive index measurements and power loss analysis makes it practical for industrial applications. This approach not only demonstrates superior sensitivity compared to earlier multimode fiber systems but also enables simpler implementation with minimal sample preparation and lower operational complexity.

This study aims to establish a robust, efficient tool for quality control, addressing the unmet need for reliable, cost-effective, and real-time detection of edible oil adulteration within the global food industry.

2. Methodology

2.1 Sample Preparation

The olive and palm oil needed for the experiment were bought from a nearby supermarket. To preserve the integrity of the samples, they were stored in a dark location before the experiment to prevent exposure to conditions that might alter their composition. Adulterated olive oil samples were prepared by mixing pure olive oil with palm oil in varying concentrations of 10%, 20%, 30%, 40%, and 50%. Table 1 summarizes the sample volumes used for each concentration.

Table 1: Volume of The Sample.

Concentration (%)	Volume of Pure Olive Oil (ml)	Volume of Palm Oil (ml)
0	50	0
10	45	5
20	40	10
30	35	15
40	30	20
50	25	25
100	0	50

2.2 Instrumentation

The Sumitomo Electric Type-82C HD Core Aligning Fusion Splicer was employed to splice the optical fibers with high precision, ensuring minimal signal loss and optimal alignment of the fiber cores. This splicer is renowned for its robust core alignment technology, which is critical for maintaining the integrity of the optical signals. The EXFO's ELS-50 was used as the light source. This device, known for its high stability and accuracy, provides a controlled light beam through optical fibers. It operated at a wavelength of 1550 nm, which was selected due to its low attenuation in SMF and its suitability for refractive index-based sensing applications. The EXFO FPM-22A Optical Power Meter was utilized to measure power loss, offering high sensitivity and precision in detecting changes in power levels. This meter's capability to accurately measure optical power loss is crucial for determining the sensor's response to different sample concentrations.

2.3 Determination of Refractive Indices

The refractive index of a material depends on various factors, including its chemical composition, density, and temperature. Different materials have different refractive indexes, and this property is critical in numerous optical applications, such as lenses, prisms, fiber optics, and waveguides. A refractometer will be used to test the sensor's refractive index. The refractive indices of pure olive oil and palm oil were measured, with variations recorded for each concentration of adulterated oil. Refractive indices were determined using the refractometer Atago PAL–BX/RI Digital Hand–Held Pocket Dual Scale Refractometer. The changes in refractive index with varying adulteration levels provided additional confirmation of sensor performance.

2.4 Adulteration Detection

The fiber is stripped, cleaned, and cleaved for a standard splice. It is then placed into the HD Core Aligning Fusion Splicer Type-82C from Sumitomo Electric, which typically aligns the fibers automatically. Next, the fibers are laterally adjusted using the splicer's fine adjustment controls or an external micro-positioner to achieve the desired offset. The fusion splice is performed, ensuring the offset remains intact. Finally, the splice is visually inspected, and the optical performance is tested to confirm the desired effect. The splicer's built-in imaging system was used to capture each splice, and the lateral displacement was measured from these images by determining the distance between the fiber core centers in the captured frame. The splicing procedure was repeated on separate samples for each offset value to confirm reproducibility.

Figure 1 depicts the experimental setup used in this study, consisting of a single-mode fiber (SMF), an optical light source, and an optical power meter. The sensor in this research is the laterally offset SMF. Light from the optical source travels through

DOI: https://doi.org/10.22452/mjs.vol45sp1.9
Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

the first fiber, leaking into the cladding and exciting high-order cladding modes. The second fiber then couples the light back into the core, creating an interference pattern. The fiber's cross-section is intentionally arranged so that the cores of the fibers do not overlap, resulting in power loss during light transmission.

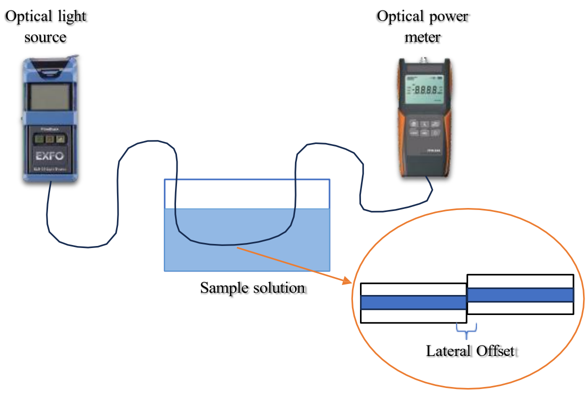


Figure 1. Experimental Setup for Adulteration Sensing

Detection is established as the cladding modes interact, with light dispersing across both the core and cladding. The lateral offset distance is varied at 0 µm, 1 µm, 2 µm, and 3 µm to analyze their effect on power loss and sensitivity. Finally, an Optical Power Meter measures the sensor's response in terms of optical signal loss.

The samples were prepared and tested under identical conditions, strictly following measurement protocols. Each measurement was repeated to verify consistency, and outliers were identified and excluded. The power loss data were then analyzed to quantify the relationship between adulteration levels, offset distance, and sensor performance. All measurements were performed at room temperature (25 ± 2°C) to minimize temperature effects, which are expected to be negligible within this range. Future work will focus on evaluating and compensating for potential temperature-induced variations to ensure consistent sensor performance under different environmental conditions.

3. Results and discussion

In this research, the refractive indices of the test samples were measured first. Table 2 displays the correlation between the percentage concentration of adulterants and the refractive index of pure olive oil compared to adulterated olive oil with palm oil. The refractive index of the olive oil analyzed in this study was 1.4698, which is consistent with the reported range of 1.4677 to 1.4705 for olive oil as specified by Codex Stan 33-1981 (Food and Agriculture Organization of the United Nations [FAO] & World Health Organization [WHO], 1981).

Table 2: Refractive Index Value of Olive Oil Adulteration.

Concentrations	Refractive Index
POO	1.4698
10%	1.4701
20%	1.4698
30%	1.4693
40%	1.4691
50%	1.4688
PO	1.4431

On the other hand, the refractive index value for palm oil obtained in this study was 1.4431, which falls outside the range specified by Codex Stan 210–1999, which is 1.459 to 1.460 (Food and Agriculture Organization of the United Nations [FAO] & World Health Organization [WHO], 1999). This deviation may be attributed to the degradation of palm oil. Over time, oils can degrade, leading to changes in their chemical composition due to oxidation and hydrolysis, which can alter the fatty acid profile and, consequently, the refractive index (Andrew et al., 2012).

The following table presents the findings from blending olive and palm oil at various concentrations. When combined with 50% adulterant, the refractive index for olive oil adulteration was ≤1.4688. At 40%, the refractive index was ≤1.4691. A 30% concentration resulted in a refractive index of ≤1.4693. For the 20% combination, the refractive index was ≤1.4698. Finally, with a concentration of 10%, the obtained value was ≤1.4701. The refractive index showed a small increase at 10% palm oil concentration because, at low concentrations, the saturated fatty acids in palm oil can occupy small gaps between the unsaturated molecules of olive oil, making the mixture a bit denser and increasing the refractive index. Such non-ideal mixing behavior and molecular packing effects have been reported in vegetable oil systems, where differences in fatty acid composition influence both density and optical properties (Amos-Tautua & Onigbinde, 2013). As the palm oil content increases further, the mixture becomes increasingly dominated by palm oil, resulting in a gradual decrease in the overall refractive index. Thus, it suggests that the refractive index can be a valuable indicator of the quality of edible oils.

Figure 2 illustrates the refractive index (RI) values for various olive oil-palm oil mixture concentrations, depending on the volume of adulterants added to the olive oil. The RI decreased proportionally with the increasing amount of palm oil adulterant. Pure olive oil has a refractive index of 1.4698. The addition of

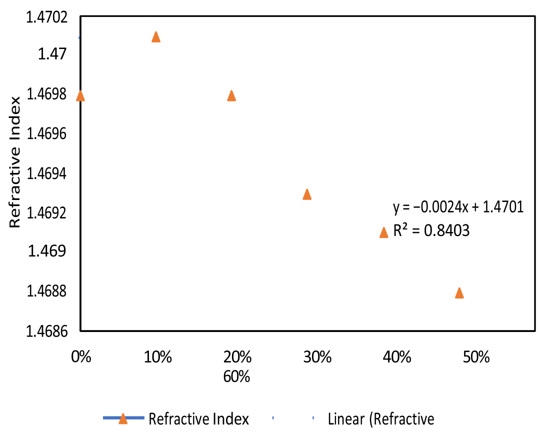


Figure 2. Refractive Index (RIU) Vs Concentration of Palm Oil in Olive Oil.

DOI: https://doi.org/10.22452/mjs.vol45sp1.9
Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

palm oil altered the RI value of the mixture. The slope of the line in the graph represents the fiber's sensitivity to different concentrations of adulteration. A steeper slope indicates greater sensitivity of the fiber sensor (Tian et al., 2022). According to the graph, the sensitivity of the fiber sensor to palm oil is 0.8403 RIU/%.

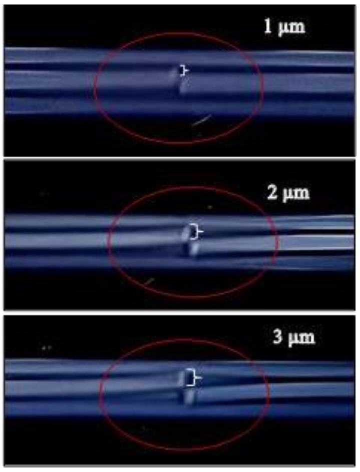


Figure 3. Captured image from the fusion splicer for 1, 2, and 3 µm offset.

Figure 3 shows the images of the spliced offset fiber from the splicing machine, as recorded from the fusion splicer.

Table 3. Power Loss for Olive Oil Adulteration with Palm Oil

Concentration (%)	Power Loss (dBm)			
	0 µm	1 µm	2 µm	3 µm
0	-3.69	-4.1	-4.56	-5.13
10	-3.73	-4.08	-4.58	-5.14
20	-3.75	-4.09	-4.59	-5.16
30	-3.76	-4.1	-4.63	-5.17
40	-3.77	-4.12	-4.67	-5.19
50	-3.79	-4.14	-4.63	-5.2

The relationship between the concentration of palm oil and the power loss was measured at different offset distances. The data shows that as the offset distance increases, the power loss also increases for all concentrations, indicating a greater degree of signal attenuation with larger offset gaps. For example, at 50% concentration, the power loss values are -3.79 dBm at 0 µm, -4.14 dBm at 1 µm, -4.63 dBm at 2 µm, and -5.2 dBm at 3 µm. This trend is consistent across all concentrations, suggesting that even a small physical gap can cause additional signal attenuation. Furthermore, there is a noticeable increase in power loss with higher concentrations of palm oil at each offset distance. At 3 µm offset, the power loss progresses from -5.13 dBm for 0% concentration to -5.2 dBm for 50% concentration, demonstrating that adulteration affects the optical properties of the oil mixture, likely due to changes in refractive index or other optical characteristics (Bodurov et al., 2013).

However, there is a slight inconsistency in the power loss trend at offsets of 1 µm and 2 µm. This can be attributed to most of the light remaining confined within the fiber core, resulting in weaker mode coupling and lower sensitivity to refractive index changes. Small differences in alignment or the sample condition may also slightly affect the measurements, causing the power loss to appear inconsistent. In contrast, at a 3 µm offset, stronger coupling between the core and cladding modes enhances light interaction with the surrounding oil, giving a more stable and linear response as the palm oil concentration increases.

To ensure the consistency and reliability of the sensor's performance, statistical process control (SPC) was applied to the measured power loss data across varying levels of adulteration (0%, 10%, 20%, 30%, 40%, 50%, and 100% palm oil). SPC was employed through the development of control charts to monitor the process behavior and identify any potential variations that might indicate inconsistencies in sensor output. Upper Control Limit (UCL) and Lower Control Limit (LCL) were calculated based on the data's mean and standard deviation, following standard SPC methods. The plotted data show that all power loss measurements fell within the UCL and LCL across the tested adulteration levels. This confirms that the sensor operated within statistical control limits, and no significant outliers or trends indicating process shifts or deviations were observed, reinforcing the sensor's reliability under experimental conditions.

The findings highlight that both the offset distance and the concentration of palm oil adulteration significantly impact power loss. Larger offsets lead to higher power loss, emphasizing the importance of maintaining minimal offsets in experimental setups to reduce signal degradation. Additionally, higher concentrations of palm oil in olive oil result in increased power loss, providing a potential method for detecting and quantifying adulteration in olive oil by measuring power loss. This can be attributed to the changes in the optical properties of the oil due to the presence of palm oil, as supported by the work of Bodurov et al., 2013 and Khosroshahi, 2018. These results are critical for ensuring the integrity and quality of olive oil, as they offer a quantitative approach to identifying adulteration.

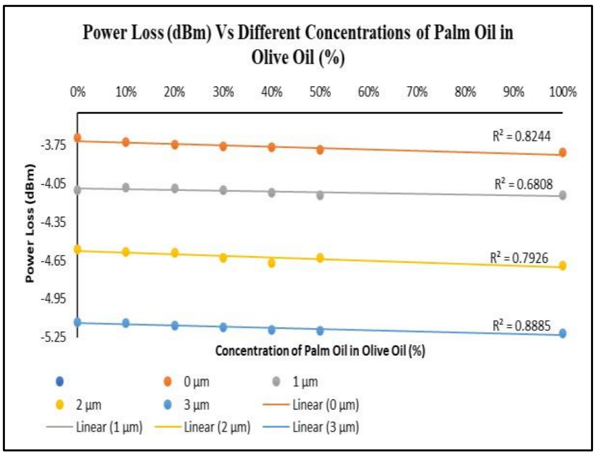


Figure 4. Power Loss (dBm) Versus Different Concentrations of Palm Oil in Olive Oil.

DOI: https://doi.org/10.22452/mjs.vol45sp1.9
Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

Figure 4 illustrates the relationship between power loss (dBm) and the concentration of palm oil in olive oil (%). The sensitivities observed for offsets of 0 µm, 1 µm, 2 µm, and 3 µm were 0.824 dBm/%, 0.680 dBm/%, 0.792 dBm/%, and 0.888 dBm/%, respectively. The data reveal a clear trend: as the lateral offset increases, the power loss increases for the same level of adulteration. This indicates that a larger offset results in greater sensitivity and more pronounced detection of adulterants. The sensitivity values at the 3 µm lateral offset distance showed a stable trend within control limits, confirming the repeatability and accuracy of the sensor's responses. The increase in sensitivity with greater offset distance suggests that the sensor's ability to detect changes in the refractive index due to palm oil adulteration improves as the offset widens. A greater lateral offset enhances coupling between the fiber core and cladding modes, allowing more light to leak into the cladding and interact with the surrounding oil medium. This enhanced light-sample interaction makes the sensor more responsive to changes in refractive index, resulting in higher sensitivity. Additionally, the observed linear decrease in output intensity with increasing offset highlights the impact of misalignment on sensor performance, emphasizing the importance of optimizing offset distances for effective adulteration detection.

Compared to previous studies about edible oil detection sensors that use fiber grating, U-bent fiber, or multimode interference, the lateral offset displacement sensor achieves a comparable or higher sensitivity using only standard SMFs and fusion splicing. This straightforward design removes the need for complex spectral analysis and enables direct, real-time power loss measurement with minimal calibration requirements.

4. Conclusion

This study successfully developed and demonstrated an optical fiber sensor based on the lateral offset displacement approach for detecting adulterated pure olive oil with palm oil. Utilizing the SMF-SMF displacement sensor, the research effectively analyzed the sensor's responses to varying concentrations of adulterants in olive oil. The experimental setup utilized sensing probes with offset distances at 1 µm, 2 µm, and 3 µm. The initial refractive indices for pure olive and palm oil were measured at 1.4698 and 1.4431, respectively.

Key findings revealed that the sensor achieved its highest sensitivity at a 3 µm lateral offset distance, with a sensitivity value of 0.888 dBm/%. This finding highlights the positive correlation between increased offset distance and enhanced sensor sensitivity. The sensor demonstrated high accuracy in detecting even small quantities of palm oil in olive oil, underscoring its potential as a reliable tool for quality control in the food industry. This study addresses several research gaps by optimizing the lateral offset displacement to improve sensitivity and precision while minimizing sample preparation and environmental interference.

Furthermore, the reliance on straightforward refractive index measurements and power loss analysis enhances the sensor's practicality for real-time applications. By integrating the benefits of high sensitivity, operational simplicity, and broad applicability, this study provides a robust solution to current challenges in edible oil adulteration detection. The lateral offset displacement approach demonstrates significant potential for ensuring product authenticity, enhancing quality control, and safeguarding consumer safety in the food industry. Future work could explore extending the technique to other high-value edible oils and further refining the sensor's performance for lower levels of adulteration.

5. Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia through Tier 1 Grant Research (Q407) and Postgraduates Research Grant (GPPS) (Vot Q680). Special thanks from the author to the Faculty of Applied Sciences and Technology (FAST) of Universiti Tun Hussein Onn Malaysia, giving full support in instruments.

6. References

Amos-Tautua, B. M. W., Onigbinde, A. O. (2013). Physicochemical properties and fatty acid profiles of crude oil extracts from three vegetable seeds. Pakistan Journal of Nutrition, 12(7): 647–650.

Andrew, C., Itodo, A. U., & Etim, E. E. (2012). Thermoxidative degradation of commonly used vegetable oils: a comparative study. Journal of Emerging Trends in Engineering and Applied Sciences, 3(6): 924–928.

Ansari, M. T. I., Raghuwanshi, S. K., Shadab, A., & Kumar, S. (2024). Detection of edible oil adulteration using fiber Bragg grating sensor: a fast and accurate approach. Optical Fibers and Sensors for Medical Diagnostics, Treatment, and Environmental Applications XXIV, 12835: 69–74.

Aparicio, R., Harwood, J. (2013).Handbook of olive oil: 431-478. Springer.

Beyer, R. M., Durán, A. P., Rademacher, T. T., Martin, P., Tayleur, C., Brooks, S. E., Sanderson, F. J. (2020). The environmental impacts of palm oil and its alternatives. Biorxiv.

Bodurov, I., Vlaeva, I., Marudova, M., Yovcheva, T., Nikolova, K., Eftimov, T., & Plachkova, V. (2013). Detection of adulteration in olive oils using optical and thermal methods. In Bulgarian Chemical Communications (Vol. 45).

Boskou, D. (2009). Culinary applications of olive oil—Minor constituents and cooking. Olive oil: Minor constituents and health: 1-4.

Bunaciu, A. A., Fleschin, S., & Aboul-Enein, H. Y. (2022). Determination of some edible oils adulteration with paraffin oil using infrared spectroscopy. Pharmacia, 69: 827-832.

Caramia, G., Gori, A., Valli, E., Cerretani, L. (2012). Virgin olive oil in preventive medicine: From legend to epigenetics. European Journal of Lipid Science and Technology, 114(4): 375-388.

Chauhan, M., Singh, V. K. (2022). ZnO coated evanescent field based fiber optic fuel adulteration sensor. 2022 Workshop on Recent Advances in Photonics (WRAP): 1–2.

Deffense, E. (1985). Fractionation of palm oil. Journal of the American Oil Chemists' Society, 62(2): 376–385.

Ferdous, A. H. M. I., Anower, M. S., Musha, A., Habib, M. A., Shobug, M. A. (2022). A heptagonal PCF-based oil sensor to detect fuel adulteration using terahertz spectrum. Sensing and Bio-Sensing Research, 36.

Food and Agriculture Organization of the United Nations, World Health Organization. (1981). Codex standard for olive oils and olive pomace oils (Codex Stan 33-1981). Codex Alimentarius.

Food and Agriculture Organization of the United Nations, World Health Organization. (1999). Codex standard for named vegetable oils (Codex Stan 210–1999). Codex Alimentarius.

Girgis, A. Y. (2003). Production of high quality castile soap from high rancid olive oil. Grasas y Aceites, 54(3): 226-233.

Guasch-Ferré, M., Hu, F. B., Martínez-González, M. A., Fitó, M., Bulló, M., Estruch, R., Ros, E., Corella, D., Recondo, J., Gómez-Gracia, E., Fiol, M., Lapetra, J., Serra-Majem, L., Muñoz, M. A., Pintó, X., Lamuela-Raventós, R. M., Basora, J., Buil-Cosiales, P., Sorlí, J. V, Salas-Salvadó, J. (2014). Olive oil intake and risk of cardiovascular disease and mortality in the PREDIMED Study.

Imoisi, O., Ilori, G., Agho, I., Ekhator, J. (2015). Palm oil, its nutritional and health implications (Review). Journal of Applied Sciences and Environmental Management, 19(1): 127.

Jadhav, M. S., Mastiholi, B., Kulkarni, V. K., Lalasangi, A. S., Raikar, U. S. (2022). DETECTION OF ADULTERANT TRACES IN EDIBLE OIL BASED ON FIBER OPTIC SENSOR TECHNOLOGY. In International Journal of Engineering Applied Sciences and Technology (Vol. 7).

Khosroshahi, M. E. (2018). Effect of Temperature on Optical Properties of Vegetable Oils Using UV-Vis and Laser Fluorescence Spectroscopy. Optics and Photonics Journal, 08(07): 247–263.

Liu, J., Zhang, X., Yang, J., Kang, J., Wang, X. (2018). Common difference temperature compensation based fiber refractive index sensor through asymmetrical core-offset splicing. Optics Communications, 427: 261–265.

Li, Y., Fang, T., Zhu, S., Huang, F., Chen, Z., & Wang, Y. (2018). Detection of olive oil adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and SiPLS. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 189: 37-43.

Mamat, H., Nor Aini, I., Said, M., Jamaludin, R. (2005). Physicochemical characteristics of palm oil and sunflower oil blends fractionated at different temperatures. Food Chemistry, 91(4): 731–736.

Nocella, C., Cammisotto, V., Fianchini, L., D'Amico, A., Novo, M., Castellani, V., Stefanini, L., Violi, F., Carnevale, R. (2017). Extra Virgin Olive Oil and Cardiovascular Diseases: Benefits for Human Health. Endocrine, Metabolic & Immune Disorders - Drug Targets, 18(1): 4–13.

Padial-Jaudenes, M., Castanys-Munoz, E., Ramirez, M., Lasekan, J. (2020). Physiological impact of palm olein or palm oil in infant formulas: A review of clinical evidence. In Nutrients (Vol. 12, Issue 12: 1–19. MDPI AG.

Puspita, I., Kurniawan, F., Muhamad Hatta, A., Koentjoro, S. (2021). U-bent plastic optical fiber for lard adulteration sensor in edible oil. In Halal Research (Vol. 1).

Rodrigues, F., Pimentel, F. B., Oliveira, M. B. P. (2015). Olive by-products: Challenge application in cosmetic industry. Industrial Crops and Products, 70: 116-124.

Rohman, A., Man, Y. C. (2010). Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. Food research international, 43(3): 886-892.

Srinivasulu, S., Rao, S. V. (2019). Multimode optical fiber extrinsic sensor—Determination of adulteration of certain edible oils. International Journal of Scientific Research and Review, 7(03): 311-318.

Stanciu, I. (2023). Study of the Rheological Behavior of Olive oil at Shear Rates Between 3.3 and 120s-1. Oriental Journal Of Chemistry, 39(1): 126–128.

Tian, R., Zhang, H., Feng, Y., Lu, Q., Chen, Z., Xue, L., Jiang, P., Shi, W., & Yao, J. (2022). Sensitivity-enhanced all-fiber displacement sensor based on Lyot filter. In Proceedings of SPIE: Advanced Sensor Systems and Applications XII (Vol. 12321): 97–102. SPIE.

Xie, C. (2022). Optical Interconnect Technologies for Datacenter Networks. In Datacenter Connectivity Technologies: 1–31. River Publishers.

Zhang, Y., Liu, Y., Shi, B., Han, B., Wang, M., Zhao, Y. (2022). Lateral offset single-mode fiber-based Fabry–Pérot interferometers with Vernier effect for hydrogen sensing. Analytical Chemistry, 95(2): 872–880.

DOI: https://doi.org/10.22452/mjs.vol45sp1.9
Malaysian Journal of Science 45 (Special Issue 1): 75-80 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

https://mjs.um.edu.my

Phytoremediation of Domestic Wastewater Using Aquatic Plants and Their Characterization: A Sustainable Approach for Wastewater Treatment

Yean Ling Pang[1a,b*],Sharon Chin Pea Ng[2a], Steven Lim[3a,b] and Jia Wei Tai[a]

Abstract: Water pollution caused by domestic wastewater discharge poses significant environmental challenges, necessitating sustainable and efficient treatment solutions. This study explores the potential of aquatic plants, specifically *Azolla microphylla* and *Lemna minor* for the phytoremediation of domestic wastewater. The removal efficiency of these plants was evaluated based on various factors including plant type, duration, weight and solution pH with a focus on the removal of both organic and inorganic compounds, such as ammoniacal nitrogen, total phosphate, color, and turbidity. The findings demonstrated the effectiveness of both species in removing these pollutants, with *Lemna minor* exhibiting superior performance. The highest removal efficiencies for ammoniacal nitrogen, color, and turbidity were 82.0%, 74.3%, and 77.4%, respectively, achieved with 7g of *Lemna minor* in 300 ml of domestic wastewater at natural solution pH. Pre- and post-phytoremediation characterization was conducted using X-ray diffraction (XRD), thermogravimetric analysis (TGA), and scanning electron microscopy (SEM) incorporated with energy-dispersive X-ray spectroscopy (EDX) to analyze the structural and compositional changes in the plant biomass. This study highlights the potential of *Azolla microphylla* and *Lemna minor* as natural phytoremediators for wastewater treatment, contributing to sustainable waste management.

Keywords: Azolla microphylla; Domestic wastewater; Lemna minor; Phytoremediation

1. Introduction

Over the past decade, the water crisis has emerged as a critical concern in Malaysia, particularly in the state of Selangor. It is approximately about 8.32 million of citizens staying in Selangor and Federal Territories of Kuala Lumpur, which accounting for about 26% population in Malaysia. The estimated gross domestic production per capita in Selangor is $43,625 (Taweesan et al., 2023). Approximately 10% of the urban population is served by the public sewerage system in Kuala Lumpur, while the majority relies on non-sewered sanitation systems. The effluents from these systems are directly discharged into open drains or nearby water bodies (Taweesan et al., 2023). Meanwhile, domestic wastewater from rural areas does not undergo any treatment before releasing back into the environment, resulting in significant environmental harm. The composition of domestic wastewater can be divided into three main categories: black water, yellow water, and grey water. The wastewater that contains pathogens can be categorized as black water, while urine and wastewater rich in nutrients can be categorized as yellow water. Meanwhile, the wastewater contains cleaning agents that can be categorized as grey water. In developing countries, critical ecological issues including compromised water quality and biodiversity loss were attributed to the untreated wastewater as the research found that 90% of domestic wastewater was discharged without a proper treatment (Werkneh & Gebru, 2023).

According to Indah Water Konsortium (IWK), majority of sewage treatment plants in the country utilize primary treatment systems such as septic tanks and Imhoff tanks followed by oxidation ponds with low-cost operating requirements in the secondary stage (Ariffin & Sulaiman, 2015). No tertiary systems are in place to remove excessive nutrients such as nitrogen and phosphate, thereby, the discharge sewage fail to meet regulatory effluent standards due to these limitations in treatment processes. This has created an urgent need to develop solutions for the safe management of domestic wastewater to achieve the targets outlined in Sustainable Development Goal No. 6 (SDG6), as highlighted in the WHO/UNICEF report (Unicef, 2023), which focuses on water, sanitation, and hygiene. The targets aim to improve the conditions of unsafely managed sanitation by 2030. If domestic wastewater is not properly treated before being discharged into natural water bodies, it can contribute to water eutrophication, exacerbating the problem (Al-dhawi et al., 2023). In many countries, including urbanized areas in Southeast Asia, only 20% wastewater treatment is considered safely managed. Alarmingly, approximately 2.8 billion population still rely on low-standard sanitation facilities, leading to significant negative impacts and the spread of fecal pathogen infections.

Authors information:
[a]Department of Chemical Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, 43000 Kajang, Selangor, MALAYSIA.
E-mail: pangyl@utar.edu.my[1], sharonncp@1utar.my[2], stevenlim@utar.edu.my[3]
[b]Centre of Advanced and Sustainable Materials Research (CASMR), Universiti Tunku Abdul Rahman, 43000 Kajang, Selangor, Malaysia.

*Corresponding Author: pangyl@utar.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.10
Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

Consequently, the integration of municip al domestic wastewater treatment with holistic resource utilization has emerged as an urgent concern that requires immediate attention (Wang et al., 2022).

Phytoremediation techniques have been employed in the treatment of weak wastewater to reduce the organic substances in the wastewater and slow down eutrophication. This emerging approach in wastewater treatment offers a promising and sustainable alternative by harnessing the natural capabilities of aquatic plants to remediate polluted water. Phytoremediation offers numerous benefits such as minimal costs and energy needs, straightforward operation and maintenance, ecological sustainability, and reliance on natural processes (Mustafa & Hayder, 2021; Tan et al., 2023). Recent research has revealed that specific species of aquatic macrophytes can tolerate stressful conditions and high concentrated pollutants in the water sources (Ansari et al., 2020). Floating aquatic plants, such as duckweeds species are commonly utilized in wastewater treatment since duckweeds are cost-effectiveness and ease to access. Duckweeds belong to a small family of aquatic floating plants or macrophytes, comprising 37 species distributed worldwide (Appenroth et al., 2018). Different species display varying patterns of growth and development even when grown in the similar medium. Duckweeds are known to be among the fastest-growing aquatic plants globally, with short doubling times (1.24 days) on nutrient-rich waters (Coughlan et al., 2022). Under favourable growth conditions, they can rapidly cover the entire water surface within a few days. Under suitable environmental conditions, the yearly production of duckweed has been documented to be 55 tons per hectare of dry biomass (Yang et al., 2021). Meanwhile, the *Azolla* plants have a rapid growth rate with the doubling time of 3 to 10 days (Thiruvenkatachari et al., 2021). This characteristic is attributed to its adaptability to different aquatic environments, making it another potential aquatic plant for phytoremediation. *Azolla microphylla* exhibits nitrogen-fixing ability due to symbiotic association with cyanobacteria through its endosymbiont, enabling growth with minimal nutrient requirements. It has been estimated that *Azolla* spp. can fix about 1.1 kg nitogen$\cdot$ha^{-1}day^{-1} and has been used as green manure as fertiliser to enhance rice production by about 20% (Prasad et al., 2016).

While there has been significant research conducted related to phytoremediation, limited attention has been given to the phytoremediation of domestic wastewater with the use of *Azolla microphylla* and duckweed species for reducing pollutants in domestic wastewater. Therefore, this study seeks to evaluate the effectiveness of both plant species in removing a wide range of contaminants from domestic wastewater, including chemical oxygen demand (COD), ammonia, nitrate, nitrite, and total phosphate. The effects of various operating parameters, such as plant type, treatment duration, plant weight, and solution pH, on phytoremediation efficiency were further discussed.

2. Methodology

2.1 Sampling of Wastewater

The raw domestic wastewater was collected from Kajang Utama Regional Sewage Treatment Plant HLT010 (2.972° N, 101.792° E). The wastewater samples were kept at 4 °C before being used in the experiment. Sodium hydroxide pellet (99.9%) and hydrochloric acid (37%) were purchased from Sigma-Aldrich and used as received without further purification.

2.2 Preparation of Accumulator Plants

The preparation of floating macrophytes followed the methodology described by Wibowo *et al.* (2023). The plants used in the experiment, including *Azolla microphylla* and *Lemna minor*, were purchased from the marketplace as free-floating specimens. Prior to experimentation, the roots of plants were cleaned under the running tap water to remove the attached soil and dirt. Then, the plants were placed in a basin of tap water and allowed to acclimatize for a week prior to the start of the sampling study. The tap water in the basin was changed weekly to maintain healthy growth of macrophytes and part of the biomass was removed to ensure enough space for further growth. Any dead leaves were removed prior to the experiment.

2.3 Characterization of Accumulator Plant

The plant morphology and elemental analysis for *Azolla microphylla* and *Lemna minor* were analysed by a Hitachi S-3400 scanning electron microscope with energy dispersive X-ray (EDX). The crystalline structure of the plants was determined via X-ray diffraction (XRD) using a Shimadzu XRD-6000. Additionally, thermogravimetric analysis (TGA) was conducted with a Perkin Elmer STA8000 to evaluate the thermal behavior of the plants.

2.4 Analysis of Wastewater

The physicochemical characteristics of domestic wastewater were determined at the beginning (raw) and monitored daily throughout the treatment period. During phytoremediation of domestic wastewater, various physicochemical conditions were observed including pH, turbidity, ammoniacal nitrogen, total phosphate, colour, and chemical oxygen demand (COD) using colorimetric methods. Prior to the liquid sample analysis, all wastewater samples before and after phytoremediation were filtered through a 0.45-micron filter to remove suspended solids, except for the turbidity analysis. The measurements were taken in triplicate, and the average removal efficiency was calculated using Equation (1), as described by Wibowo et al. (2023).

$$\text{Removal efficiency (\%)} = \frac{W_i - W_e}{W_i} \times 10 \tag{1}$$

where W_i = influent concentration and W_e = effluent concentration

The analytical methods for domestic wastewater was adopted and modified from Goren et al. (2021) and Worku et al. (2024). The solution pH was performed using a pH meter (Sartorius- PB-10 Basic Bench Top pH Meter). Turbidity, ammoniacal nitrogen, total phosphate, colour, and chemical oxygen demand (COD) were measured using colorimetric method by a potable DR 890 colorimeter (HACH, Loveland, Co., USA). The unit of turbidity was measured in Formazin Attenuation Units (FAU), where 1 FAU = 1 NTU = 1 FTU and conducted in the range of 21 to 1000 FAU under program number 95. The total phosphate of the wastewater samples was assessed by the method PhosVer® 3 with acid persulfate digestion method using program number 82. The

DOI: https://doi.org/10.22452/mjs.vol45sp1.10
Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

analysis was conducted for a detectable range of 0.06 to 3.50 mg/L PO_4^{3-} were used. The ammoniacal nitrogen (NH_3-N) analysis of the wastewater samples was performed by AmVer™ Salicylate Test 'N Tube™ Method and detected at high range of 0.4 to 50.0 mg/L NH_3-N under program number 67.

The COD analysis of the wastewater samples was performed using United States Environmental Protection Agency (USEPA) reactor digestion method (8000) under Program No. 16. The COD digestion reagent vials for detectable range of 3 to 150 mg/L low range COD were used and samples were preheated to 200 ℃ in a DRB 200 reactor. The true colour of filtered initial and final wastewater sample was performed according to platinum-cobalt standard method at 420 nm, under program number 19. The removal of suspended solids in the wastewater sample was conducted using 0.45 micron-filtration. The results followed the standard of 1 colour unit being equal to 1 mg/L platinum as chloroplatinate ion.

2.5 Parameter Studies during Phytoremediation

The phytoremediation study investigated the effects of different hyperaccumulator plants (*Azolla microphylla* and *Lemna minor*), remediation duration (1, 2, and 3 days), plant weights (5, 10, 15, 20, and 25 g), and solution pH (3, 5, 7, 9, and 11) on domestic wastewater. Each experiment used approximately 10 g (± 0.5 g) of plant in 300 ml of wastewater under ambient conditions and natural pH, with controls lacking plants. For pH adjustment, 0.1 M HCl and 0.1 M NaOH were used. Optimal parameters from each stage were carried forward to the rest of the parameter studies.

3. Results and discussion

3.1 Characterization of Accumulator Plant

Figure 1(a) shows the thermal behaviours of *Azolla microphylla* and *Lemna minor* over a broad temperature range (30 to 900 °C). The aquatic plants contain cell-wall components such as cellulose, hemicellulose, and lignin. It was observed that *Azolla microphylla* and *Lemna minor* underwent three phases of weight lost, below 200, 200-400 and above 400 °C. Osman et al. reported that the weight loss was attributed to the removal of moisture, decomposition of the hemicellulose followed by lignin pyrolysis decomposition above 600 °C. Miranda et al. (2016) reported that the weight lost happened at above 200 was denoted to the fractionation of carbohydrates, proteins and lipids. Meanwhile, the further weight loss was primarily due to the gasification of the carbonaceous residue, where highly non-volatile carbon compounds vaporized to produce carbon monoxide and carbon dioxide. Insignificant weight loss of *Azolla microphylla* and *Lemna minor* samples likely indicate the presence of minor quantities of lignin-like compounds.

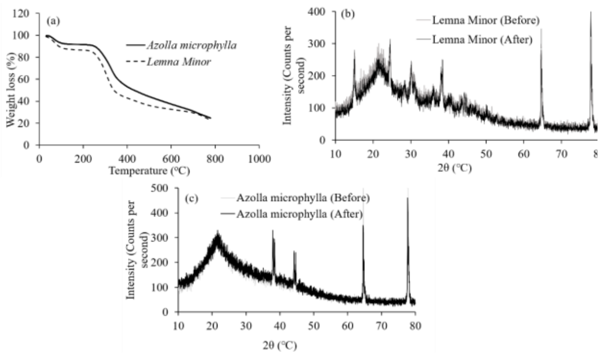


Figure 1. (a) Thermal analysis and XRD patterns of (b) Lemna minor (c) Azolla microphylla before and after phytoremediation.

The volatile content of the samples was accessed by calculating the percentage of weight lost between 120 and 650 °C (primary volatiles) and between 650 and 950 °C (secondary volatiles). The ash content (%) of the sample was determined by measuring the percentage of solids remaining after the combustion process. Based on the proximate analysis of the plant samples, it was found that the moisture loss from *Lemna minor* (11 %) was higher compared to *Azolla microphylla,* (7 %). The total volatiles for *Azolla microphylla* and *Lemna minor* were 96 % and 91 %, respectively. This indicated that *Azolla microphylla* was having higher secondary volatiles compared to *Lemna minor*. *Azolla* species have been reported to be rich in proteins, essential amino acids, vitamins A and B12, beta-carotene, growth-promoting intermediates, and minerals such as calcium, phosphorus, potassium, iron, copper, and magnesium (Muradov et al., 2014). The high nutrient content, substantial phosphorus (P) and nitrogen (N), coupled with its low lignin levels, renders Azolla an effective nutrient feed for livestock.

Figure 1(b) and Figure 1(c) present the XRD patterns of both *Lemna minor* and *Azolla microphylla* before and after phytoremediation. The patterns exhibit significant noise, underscoring the high degree of amorphousness characteristic of these aquatic plants. *Lemna minor* shows the XRD peaked at about 15.7 and 22.8°, corresponding to the (110) planes typical of cellulose and starch components. According to Azwar et al. (2022), duckweed species exhibits low crystallinity and drying could modifying the crystalline structure and resulting in a transformation from cellulose I to cellulose II. Meanwhile, *Azolla microphylla* demonstrated a broad peak observed between 2θ = 10–30° was corresponding to the (002) plane of amorphous carbon in lignocellulosic compounds (Puglia et al., 2024). This broad peak might be attributed to the lignocellulosic content, primarily composed of cellulose, hemicellulose, and lignin. In

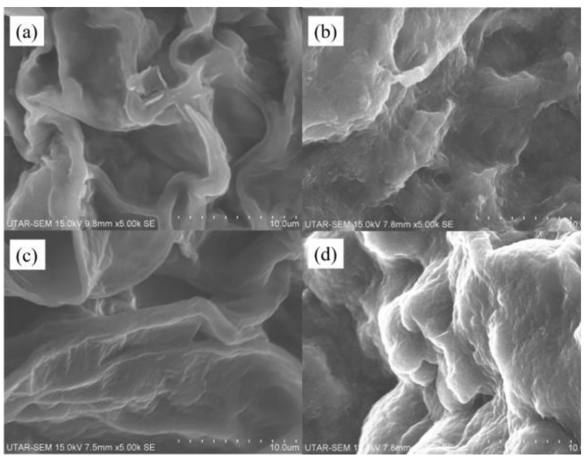


Figure 2. SEM Images of Lemna minor (a) before (b) after phytoremediation and Azolla microphylla (c) before (d) after phytoremediation.

DOI: https://doi.org/10.22452/mjs.vol45sp1.10
Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

natural lignocellulosic structures, cellulose exhibits both crystalline and amorphous regions, while hemicellulose and lignin are entirely amorphous. In is interesting to highlighted that both *Lemna minor* and *Azolla microphylla* samples showed no observable changes in terms of XRD peak intensities at 2θ values before and after phytoremediation.

Figure 2(a) and Figure 2(b) demonstrate the surface morphology of *Lemna minor* before and after phytoremediation of domestic wastewater. It could be observed the fresh leaf structure of *Lemna minor* had deep pores and free space, which might provide more surface area and facilitate the biosorption process. The surface of *Azolla microphylla* also demonstrated a similar structure as *Lemna minor* as shown in Figure 2(c), exhibited free spaces and a large surface area, enhancing its capacity for nutrient biosorption in domestic wastewater. In addition, *Azolla microphylla* was also known for the small, hair-like structures on its leaves called trichomes which could provide a large surface area for absorption of nutrients and pollutants in water (Akhtar et al., 2021). However, shrinking on leaf structures could be observed on the plant after phytoremediation due to the drying process of leaf prior to SEM analysis as shown in Figure 2(b) and Figure 2(d).

Elemental analysis for the fresh plant samples and after phytoremediation were also conducted to identify the elemental compositions. Based on Table 1, the results showed that the increment of phosphorus (1 to 2 wt%) and nitrogen (2 to 3 wt%) by *Azolla microphylla* after phytoremediation of domestic wastewater, indicating nutrient uptake had occurred. A similar observation was observed with *Lemna minor.* The result showed increment of phosphorus (3 to 4 wt%) and nitrogen (2 to 5 wt%). The findings from this study indicated a higher composition of element nitrogen and phosphorus in the *Lemna minor* compared to *Azolla microphylla*. Previous research highlighted the high tolerance of *Lemna minor* to high levels of ammonium ions in the wastewater (Britto & Kronzucker, 2002). Furthermore, *Lemna minor* was reported to have the ability to transport nitrogen and effectively to accumulate large amount of nitrogen within its biomass (Guo et al., 2020). These reports correlated with the findings of this study where *Lemna minor* showed higher removal efficiency than *Azolla microphylla* in terms of ammoniacal nitrogen and total phosphorus.

Table 1: EDX analysis of Azolla microphylla and Lemna minor before and after phytoremediation

	Weight percentage (%)			
	C	O	N	P
Lemna minor				
Before	57.4	37.1	2.0	3.5
After	51.2	38.8	5.4	4.6
Azolla microphylla				
Before	69.7	29.1	2.5	2.7
After	53.5	39.0	3.6	3.9

3.2 Parameter Studies

Figure 3(a) shows the phytoremediation ability of domestic wastewater by using *Azolla microphylla* and *Lemna minor*. It was observed that *Lemna minor* performed better than *Azolla microphylla* in terms of removal efficiency of the domestic wastewater. The removal efficiency of domestic wastewater

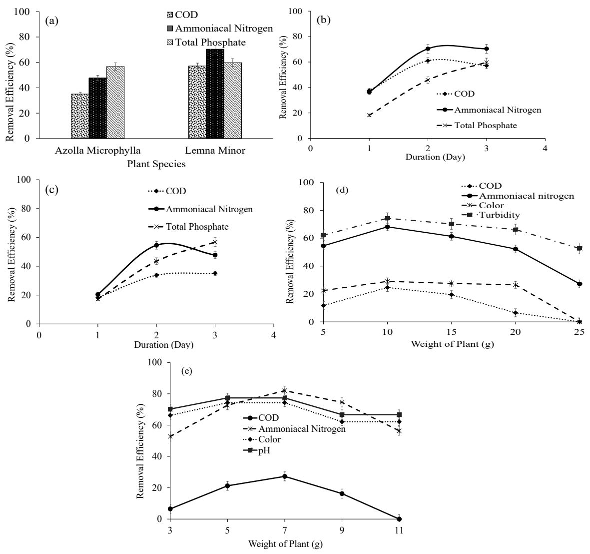


Figure 3. Effect of (a) plant species, duration on (b) Lemna minor (c) Azolla microphylla (d) plant weight of Lemna minor (e) solution pH during phytoremediation (Domestic wastewater volume = 300 ml, plant weight = 10 g (± 0.5 g), ambient conditions, natural solution pH, treatment duration = 3 days).

using *Lemna minor* on the harvested day in terms of ammoniacal nitrogen, total phosphate and COD, were 70.5 %, 59.8 % and 57.1 %, respectively. The findings also indicated that *Lemna minor* possessed deeper spaces and larger surface area as compared to *Azolla microphylla,* which aids in the better absorption of nutrients (i.e. ammoniacal nitrogen and total phosphate) from the domestic wastewater.

The results were compared with findings from the literature review on the improvement of water quality. The findings were consistent with historical data, demonstrating the efficient absorption of nitrogen and phosphorus components in swine wastewater by Azolla microphylla and Lemna minor (Muradov et al., 2014). Another study compared the effectiveness of both plants in treating heavy metal wastewater. Polechońska et al. (2022) reported that Lemna species could uptake more K, Mg and Na than Azolla species, but the content of Cd, Cr, Cu, Fe, Ni, and Pb in Azolla species was higher than in Lemna species. The same authors proved that the metal accumulation index for Lemna species was higher than Azolla species. Lemna species, characterized by submerged reduced roots and fronds that are approximately 50% submerged, were advantageous for absorption, whereas only the bottom layer of Azolla species had direct contact with the contaminated water, which impacted its performance.

Figure 3 (b) and (c) show that phytoremediation treatment duration had a positive effect on the removal efficiency of

DOI: https://doi.org/10.22452/mjs.vol45sp1.10
Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

ammoniacal nitrogen, total phosphate, and COD for both Lemna minor and Azolla microphylla, respectively. At the beginning of the experiment, both plants could achieve similar removal efficiency. However, the findings revealed that Lemna minor outperformed Azolla microphylla in achieving a higher removal efficiency from day 2 onwards. On the second day of the experiment, Lemna minor exhibited a higher removal efficiency of domestic wastewater in terms of total phosphate, ammoniacal nitrogen, and COD, that were 45.9%, 70.5%, and 61%, respectively. On day 3, the COD removal efficiency by Lemna minor was reduced from 61% to 57.1%.

The fluctuation results on the COD removal result was aligned with the current observations. Nesan et al. (2020) reported that the increase in COD values may be due to several factors: the plant's adaptation ability under saline conditions, the induction of organic cell death and lysis by high salinity, resulting in the release of organic substances into the medium, or the excretion of organic substances by the plantlets in response to biotic stress caused by difference in salinity levels between the plants and surroundings. Furthermore, the changes in leaves colour from greenish to pale yellow indicated plant cell death due to high exposure of nitrogen ammonia in the wastewater which had a chlorination effect on the plants. Kumar & Deswal (2020) found that duckweed species showed better phosphorus removal efficiency (70.21% on day 13) and COD removal (62.19%) for rice mill wastewater as compared to Salviniaceae species (61.38% and 60%). They claimed that the negative growth rate of Salvinia starting on day 10 was caused by the presence of periphyton in the salvinia tub.

Lemna minor was chosen to investigate the optimal weight during phytoremediation of domestic wastewater and results are shown in Figure 3 (d). It was found that the optimal weight for the specific conditions was 10 g of Lemna minor. In this study, the highest removal efficiency of domestic wastewater using 10 g of Lemna minor in terms of ammoniacal nitrogen, COD, colour and turbidity were 68.2 %, 24.7 %, 29.1 % and 74.3 %, respectively. The increase in Lemna minor biomass from 5 g to 10 g resulted in enhanced removal efficiencies, attributed to the substantial increase in nutrient uptake capacity of the macrophyte. However, further increases in biomass beyond 20 g were observed could reduce phytoremediation efficiency. Several factors contributed to this decline, including nutrient depletion, plant overcrowding, and increased oxygen demand, all of which negatively impacted the phytoremediation capability of Lemna minor. Consequently, an optimal biomass of 10 g was selected for subsequent parameter studies.

The effect of solution pH of 3, 5, 7, 9 and 11 during phytoremediation of domestic wastewater using Lemna minor was evaluated, and the results are demonstrated in Fig. 3 (e). Consideration must be given to the effect of solution pH on plant growth during extended restoration processes to ensure stable development. It was observed the Lemna minor could bioaccumulate higher removal efficiency when the solution pH ranged from 5 to 7. The duckweed species requires a pH of 5–9, 60 mg/L of nitrogen and at least 1 mg/L of phosphorus for optimal growth (Ekperusi et al., 2019). Ekperusi et al. (2019) and Heitzman et al. (2024) reported that the optimal duckweeds growth rates in laboratory happened in the range of 6.5–7.5. Meanwhile, the same authors pointed out that an adequate supply of nitogen, phosphors, potassium and other essential nutrients is required for optimal duckweed growth. Thus, most of the available form of nitrogen and phosphorous would be readily bioaccumulated by Lemna minor at this optimal condition.

Jones et al. (2023) reported that the ammonia (NH_3) proportion increases with rising water pH. NH_3 is more toxic to duckweed due to its lipid-solubility, which allows it to penetrate cell membranes, while the ionized form is less harmful. Thus, elvated water pH leads to the formation of ammonia in the solution, which can be toxic and prone to volatilization (Ekperusi et al., 2019). Acidifying the media to shift the NH_4^+: NH_3 equilibrium allows duckweed to grow in higher concentrations of ammoniacal nitrogen. However, it was found that extremely acidic condition could lead to stress on the plant, which negatively impact plant's cell membrane permeability, reducing the photosynthesis of the plant and its growth rate. Furthermore, the solubility of phosphorus and nitrogen decreased at highly acidic conditions which reduces the uptake of these nutrients by the plant.

5. Conclusion

This study investigated the phytoremediation of domestic wastewater using *Lemna minor* and *Azolla microphylla*. The SEM images showed deep pores in both plants, indicating significant surface area for nutrient bioaccumulation. The EDX analysis confirmed higher nitrogen and phosphorus levels in the plants post-phytoremediation, with *Lemna minor* exhibiting greater nutrient uptake than *Azolla microphylla*. Several parameters were assessed, including plant type, duration, weight, and solution pH. *Lemna minor* outperformed *Azolla microphylla* in phytoremediation efficiency, achieving 45.9% total phosphate, 70.5% ammoniacal nitrogen, and 61.0% COD removal after three days. Factors such as nutrient depletion, plant overcrowding, and increased oxygen demand affected *Lemna minor*'s performance. Optimal conditions for *Lemna minor* were found at a pH of 5 to 7, with ammoniacal nitrogen removal efficiency ranging from 72.7% to 82.0% and COD removal from 21.3% to 27.3%. Extreme solution pH conditions (below pH 3 or above pH 9) could reduce removal efficiency due to plant stress and free ammonia toxicity. The best results were achieved at phytoremediation of domestic wastewater using 10 g of *Lemna minor* for about two days at a solution pH of 7. This research demonstrates the effectiveness of *Lemna minor* and *Azolla microphylla* in phytoremediation which contributing to sustainable wastewater treatment practice.

6. Acknowledgement

The authors would like to express their gratitude to the Ministry of Higher Education (MoHE) Malaysia for funding through the Fundamental Research Grant Scheme (FRGS/1/2022/TK05/UTAR/02/34), and the Universiti Tunku Abdul Rahman (UTAR) for the support provided by the UTAR Research Fund (IPSR/RMC/UTARRF/2020-C2/P01). This research was also partially supported by the Kurita Overseas Research

Grant (21Pmy076) from the Kurita Water and Environment Foundation.

7. References

Akhtar, M., Sarwar, N., Ashraf, A., Ejaz, A., Ali, S., & Rizwan, M. (2021). Beneficial role of Azolla sp. in paddy soils and their use as bioremediators in polluted aqueous environments: implications and future perspectives. Archives of Agronomy and Soil Science, 67(9), 1242–1255. https://doi.org/10.1080/03650340.2020.1786885

Al-dhawi, B. N. S., Kutty, S. R. M., Ghaleb, A. A. S., Almahbashi, N. M. Y., Saeed, A. A. H., Al-Mekhlafi, A. B. A., Alsaeedi, Y. A. A., & Jagaba, A. H. (2023). Pretreated palm oil clinker as an attached growth media for organic matter removal from synthetic domestic wastewater in a sequencing batch reactor. Case Studies in Chemical and Environmental Engineering, 7, 100294. https://doi.org/10.1016/j.cscee.2022.100294

Ansari, A. A., Naeem, M., Gill, S. S., & AlZuaibr, F. M. (2020). Phytoremediation of contaminated waters: An eco-friendly technology based on aquatic macrophytes application. Egyptian Journal of Aquatic Research, 46(4), 371–376. https://doi.org/10.1016/j.ejar.2020.03.002

Appenroth, K. J., Sowjanya Sree, K., Bog, M., Ecker, J., Seeliger, C., Böhm, V., Lorkowski, S., Sommer, K., Vetter, W., Tolzin-Banasch, K., Kirmse, R., Leiterer, M., Dawczynski, C., Liebisch, G., & Jahreis, G. (2018). Nutritional value of the duckweed species of the Genus Wolffia (Lemnaceae) as human food. Frontiers in Chemistry, 6, 1–13. https://doi.org/10.3389/fchem.2018.00483

Ariffin, M., & Sulaiman, S. N. M. (2015). Regulating Sewage Pollution of Malaysian Rivers and its Challenges. Procedia Environmental Sciences, 30, 168–173. https://doi.org/10.1016/j.proenv.2015.10.030

Azwar, E., Chan, D. J. C., Kasan, N. A., Rastegari, H., Yang, Y., Sonne, C., Tabatabaei, M., Aghbashlo, M., & Lam, S. S. (2022). A comparative study on physicochemical properties, pyrolytic behaviour and kinetic parameters of environmentally harmful aquatic weeds for sustainable shellfish aquaculture. Journal of Hazardous Materials, 424, 127329. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.127329

Britto, D. T., & Kronzucker, H. J. (2002). NH4+ toxicity in higher plants: A critical review. Journal of Plant Physiology, 159(6), 567–584. https://doi.org/https://doi.org/10.1078/0176-1617-0774

Coughlan, N. E., Walsh, É., Bolger, P., Burnell, G., O'Leary, N., O'Mahoney, M., Paolacci, S., Wall, D., & Jansen, M. A. K. (2022). Duckweed bioreactors: challenges and opportunities for large-scale indoor cultivation of Lemnaceae. Journal of Cleaner Production, 336, 130285. https://doi.org/10.1016/j.jclepro.2021.130285

Ekperusi, A. O., Sikoki, F. D., & Nwachukwu, E. O. (2019). Application of common duckweed (Lemna minor) in phytoremediation of chemicals in the environment: State and future perspective. Chemosphere, 223, 285–309. https://doi.org/https://doi.org/10.1016/j.chemosphere.2019.02.025

Goren, A. Y., Yucel, A., Sofuoglu, S. C., & Sofuoglu, A. (2021). Phytoremediation of olive mill wastewater with Vetiveria zizanioides (L.) Nash and Cyperus alternifolius L. Environmental Technology and Innovation, 24, 102071. https://doi.org/10.1016/j.eti.2021.102071

Guo, L., Jin, Y., Xiao, Y., Tan, L., Tian, X., Ding, Y., He, K., Du, A., Li, J., Yi, Z., Wang, S., Fang, Y., & Zhao, H. (2020). Energy-efficient and environmentally friendly production of starch-rich duckweed biomass using nitrogen-limited cultivation. Journal of Cleaner Production, 251, 119726. https://doi.org/10.1016/j.jclepro.2019.119726

Heitzman, B. S., Bueno, G. W., Camargo, T. R., Proença, D. C., Yaekashi, C. T. O., da Silva, R. M. G., & Machado, L. P. (2024). Duckweed application in nature-based system for water phytoremediation and high-value coproducts at family agrisystem from a circular economy perspective. Science of The Total Environment, 919, 170714. https://doi.org/https://doi.org/10.1016/j.scitotenv.2024.170714

Jones, G., Scullion, J., Dalesman, S., Robson, P., & Gwynn-jones, D. (2023). Lowering pH enables duckweed (Lemna minor L.) growth on toxic concentrations of high-nutrient agricultural wastewater. Journal of Cleaner Production, 395, 136392. https://doi.org/10.1016/j.jclepro.2023.136392

Kumar, S., & Deswal, S. (2020). Phytoremediation capabilities of Salvinia molesta, water hyacinth, water lettuce, and duckweed to reduce phosphorus in rice mill wastewater. International Journal of Phytoremediation, 22(11), 1097–1109. https://doi.org/10.1080/15226514.2020.1731729

Liu, Z., & Tran, K. (2021). A review on disposal and utilization of phytoremediation plants containing heavy metals. Ecotoxicology and Environmental Safety, 226, 112821. https://doi.org/10.1016/j.ecoenv.2021.112821

Miranda, A. F., Biswas, B., Ramkumar, N., Singh, R., Kumar, J., James, A., Roddick, F., Lal, B., Subudhi, S., Bhaskar, T., & Mouradov, A. (2016). Aquatic plant Azolla as the universal feedstock for biofuel production. Biotechnology for Biofuels, 9(1), 1–17. https://doi.org/10.1186/s13068-016-0628-5

Muradov, N., Taha, M., Miranda, A. F., Kadali, K., Gujar, A., Rochfort, S., Stevenson, T., Ball, A. S., & Mouradov, A. (2014). Dual application of duckweed and Azolla plants for wastewater treatment and renewable fuels and petrochemicals production. Biotechnology for Biofuels, 7(1), 1–17. https://doi.org/10.1186/1754-6834-7-30

Mustafa, H. M., & Hayder, G. (2021). Cultivation of S. molesta plants for phytoremediation of secondary treated domestic wastewater. Ain Shams Engineering Journal, 12(3), 2585–2592. https://doi.org/10.1016/j.asej.2020.11.028

Nesan, D., Selvabala, K., & Chan Juinn Chieh, D. (2020). Nutrient uptakes and biochemical composition of Lemna minor in brackish water. Aquaculture Research, 51(9), 3563–3570. https://doi.org/https://doi.org/10.1111/are.14693

Polechońska, L., Szczęśniak, E., & Klink, A. (2022). Comparative analysis of trace and macro-element bioaccumulation in four free-floating macrophytes in area contaminated by copper smelter. International Journal of Phytoremediation, 24(3), 324–333. https://doi.org/10.1080/15226514.2021.1937932

Prasad, S. M., Kumar, S., Parihar, P., Singh, A., & Singh, R. (2016). Evaluating the combined effects of pretilachlor and UV-B on two Azolla species. Pesticide Biochemistry and Physiology, 128, 45–56. https://doi.org/https://doi.org/10.1016/j.pestbp.2015.10.006

Puglia, D., Luzi, F., Tolisano, C., Rallini, M., Priolo, D., Brienza, M., Costantino, F., Torre, L., & Del Buono, D. (2024). Cellulose nanocrystals and lignin nanoparticles extraction from Lemna minor L.: Acid hydrolysis of bleached and ionic liquid-treated biomass. Polymers, 16(10). https://doi.org/10.3390/polym16101395

Tan, H. W., Pang, Y. L., Lim, S., & Chong, W. C. (2023). A state-of-the-art of phytoremediation approach for sustainable management of heavy metals recovery. Environmental Technology and Innovation, 30, 103043. https://doi.org/10.1016/j.eti.2023.103043

Taweesan, A., Kanabkaew, T., Surinkul, N., & Polprasert, C. (2023). Convenient solutions to inconvenient truth: Domestic wastewater management-based approaches to sustainable development goal no. 6. Environmental and Sustainability Indicators, 18, 100255. https://doi.org/10.1016/j.indic.2023.100255

Thiruvenkatachari, S., Saravanan, C. G., Edwin Geo, V., Vikneswaran, M., Udayakumar, R., & Aloui, F. (2021). Experimental investigations on the production and testing of azolla methyl esters from Azolla microphylla in a compression ignition engine. Fuel, 287, 119448. https://doi.org/https://doi.org/10.1016/j.fuel.2020.119448

Unicef. (2023). Progress on household drinking water, sanitation and hygiene 2000-2022: Special focus on gender. https://data.unicef.org/resources/jmp-report-2023/

Wang, Q., Wang, X., Hong, Y., Liu, X., Zhao, G., Zhang, H., & Zhai, Q. (2022). Microalgae cultivation in domestic wastewater for wastewater treatment and high value-added production: Species selection and comparison. Biochemical Engineering Journal, 185, 108493. https://doi.org/https://doi.org/10.1016/j.bej.2022.108493

Werkneh, A. A., & Gebru, S. B. (2023). Development of ecological sanitation approaches for integrated recovery of biogas, nutrients and clean water from domestic wastewater. Resources, Environment and Sustainability, 11, 100095. https://doi.org/10.1016/j.resenv.2022.100095

Wibowo, Y. G., Syahnur, M. T., Al-Azizah, P. S., Arantha Gintha, D., Lululangi, B. R. G., & Sudibyo. (2023). Phytoremediation of high concentration of ionic dyes using aquatic plant (Lemna minor): a potential eco-friendly solution for wastewater treatment. Environmental Nanotechnology, Monitoring and Management, 20, 100849. https://doi.org/10.1016/j.enmm.2023.100849

Worku, A., Tibebu, S., & Kassahun, E. (2024). Phytoremediation potential of Duranta erecta in cascade hydroponics for effective domestic wastewater treatment. Results in Engineering, 23(July), 102837. https://doi.org/10.1016/j.rineng.2024.102837

Yang, G.-L., Yang, M.-X., Lv, S.-M., & Tan, A.-J. (2021). The effect of chelating agents on iron plaques and arsenic accumulation in duckweed (Lemna minor). Journal of Hazardous Materials, 419, 126410. https://doi.org/https://doi.org/10.1016/j.jhazmat.2021.126410

DOI: https://doi.org/10.22452/mjs.vol45sp1.10
Malaysian Journal of Science 45 (Special Issue 1): 81-87 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024 Special issue

Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026) https://mjs.um.edu.my

Utilising Electrical Resistivity Tomography for Peat Soil Stratigraphy Analysis at Parit Nipah

Muhammad Arif Azhar Zaihan[1a], Mohd Khaidir Abu Talib[2a,b*], Siti Nor Hidayah Arifin[3c],Kasbi Basri[4c], Muhammad Nur Hidayat Zahri[5d], and Muhammad Faris Hadi Muhamad Shahrulzaman[6a]

Abstract: This research examined Electrical Resistivity Tomography (ERT) for analysing peat soil stratigraphy at Parit Nipah in Johor, Malaysia. Analyses were carried out with the Wenner and Schlumberger arrays along with two lines to assess their effectiveness for peat profiling. The electrode spacing was set at 2 metres for the longer configuration and 1 metre for the shorter one for Line 1, and 1 metre for the longer configuration and 0.5 metres for the shorter one for Line 2. This allowed for a thorough investigation of both lateral and vertical variations in resistivity. The maximum depth reached during the investigation was 16 metres on Line 1 and 7 metres on Line 2 by utilising an ABEM Terrameter SAS 4000, steel electrodes, and Lund imaging cables. The ZondRes2d software was used for data processing and inversion, creating two-dimensional resistivity models. The Schlumberger array emerged as the preferred method for peat profiling, showing minimal differences compared to results from traditional peat sampling methods. While the Wenner protocol was also effective, it exhibited a slightly greater margin of error. This study highlights the ability of ERT to analyse peat stratigraphy efficiently and non-invasively, underscoring its advantages for peatland conservation and sustainable management practices.

Keywords: : Electrical Resistivity Tomography, Schlumberger Array, Wenner Array, Peat Profiling, Non-invasive Method.

1. Introduction

The identification of soil profiles is crucial in geotechnical investigation. To develop geotechnical designs for construction projects, it is essential to gather sufficient data regarding the soil profile. Conventional methods, like drilling, have been effective for data collection. However, they are often expensive, time-consuming, and invasive. Large infrastructure projects, such as railway and road construction, often require numerous boreholes to obtain adequate data for comprehensive analysis, which subsequently increases project duration and costs.

Geophysical techniques, like Electrical Resistivity Tomography (ERT), are becoming increasingly popular as non-invasive methods that provide broader exploration capabilities and efficient data analysis.

Authors information:
[a]Faculty of Civil Engineering and Built Environment, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, MALAYSIA. E-mail: arifazharzaihan243@gmail.com[1], khaidir@uthm.edu.my[2], hf230031@student.uthm.edu.my[6]
[b]Research Center for Soft Soil (RECESS), Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, MALAYSIA. E-mail: khaidir@uthm.edu.my[2]
[c]Faculty of Civil Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, MALAYSIA. E-mail: siti.nor.hidayah@utm.my[3], kasbibinbasri@gmail.com[4]
[d]Azurite Engineering PLT, Batu Pahat 86400, MALAYSIA. Email: hidayat.zahari@hotmail.com[5]

*Corresponding Author: khaidir@uthm.edu.my

Advances in these geophysical methods have allowed for the mapping of the spatial distribution in electrical resistivity and induced polarisation within subsurface layers, thereby facilitating the assessment of subsurface variations (Slob, 2004).

The ERT technique measures the electrical resistivity of the ground surface, which is influenced by the resistivity variations of different material layers. This method involves injecting current into the ground using two current electrodes and measuring the resulting potential difference. A core aspect of Electrical Resistivity Tomography (ERT) is the computation of the resistivity pseudo-section, which entails comparing the observed apparent resistivity pseudo-section with a calculated pseudo-section (Everett, 2013).

ERT is considered superior to other electrical methods because it can yield quantitative results using a precisely controlled source of specific dimensions. The electrical properties measured by ERT are significantly affected by geological features such as mineral composition, fluid content, porosity, and water saturation levels. By mapping the distribution of electrical resistivity across the soil, it becomes possible to differentiate between various soil layers (Herman, 2001; Loke, 1999). As a result, these non-invasive geophysical techniques provide significant advantages across different fields, such as geology, geotechnical engineering, and environmental research (Miele et al., 1996).

Despite its many benefits, ERT has limitations, particularly in areas with complex subsurface characteristics, like peatlands, which exhibit heterogeneous properties and significant

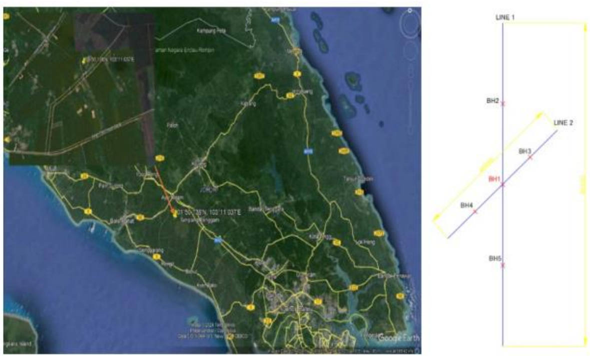
Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

fluctuations in water tables. In regions with highly variable peat properties, resistivity variability can lead to ambiguous interpretations, as the method may not accurately distinguish between different peat layers or the presence of decomposed organic material. Furthermore, significant changes in the water table can complicate resistivity measurements by introducing temporal variations that impact data consistency (Binley et al., 2002). High moisture content, characteristic of peatlands, can further obscure the delineation of subsurface features due to water's dominant influence on electrical resistivity (Reynolds, 2011). Addressing these limitations typically requires integrating ERT data with supplementary methods, such as drilling or other geophysical techniques, to enhance accuracy and reliability in complex environments.

2. Site Description

The research site was located in Parit Nipah, Johor, Malaysia. The area lies within a Quaternary region, characterised by the presence of marine and continental deposits such as peat, sand, silt, clay, and small amounts of gravel, as depicted in Figure 1. This region is a sparsely populated area and primarily used for palm tree cultivation. The analysis conducted with the peat sampler indicated that the uppermost 4 metres of soil at the research region consisted mainly of peat. Hence, the location selection was selected due to its substantial peat depth, which is among the deepest recorded in the area. The location also offered convenient access, making it suitable for conducting research activities.

3. Field Investigation



(a) (b)

Figure 1. (a) Study location with coordinate. (b) Illustration of 2 ERT line conducted at the field.

The field study entails geo-electrical studies using Electrical Resistivity Tomography (ERT) and a peat sampler for profiling. The study's objective was to ascertain peat depth by employing Non-Destructive Testing (NDT) and validating the findings through conventional invasive methods.

3.1 Peat Sampler

The peat sampler was used to collect the partially disturbed samples. The approaches conform to the requirements specified by Eijkelkamp Agrisearch Equipment (E.A., 2014). The peat sample tests were conducted to verify the accuracy of the soil stratigraphy collected from the ERT method. The peat sampler was used as a more affordable and straightforward alternative to the use of an expensive and complicated borehole. Five inquiry locations were designated, three of them located along each line of the ERT array, and one situated at the midpoint between Line 1 and Line 2. Profiling was made at regular intervals of 0.5 metres, reaching a maximum depth of 5 metres.

3.2 Electrical Resistivity Tomography

The Electrical Resistivity Tomography (ERT) was conducted using the Wenner and Schlumberger arrays for both Line 1 and Line 2. The experiments were conducted on a rainy day in May. The protocols selected for comparison aimed to determine the most appropriate protocol array configuration for peat profiling. The Schlumberger array is specifically designed to detect variations in lateral resistivity, while the Wenner array is used to recognise changes in vertical resistivity (Loke, 2004). Additionally, the gradient array in Electrical Resistivity Tomography (ERT) maps subsurface resistivity distributions to determine peat deposits' thickness and lateral changes (Yusa et al., 2019).

The equipment used in the test consists of the ABEM Terrameter SAS 4000, steel electrodes, electrode selector, jumper, Lund imaging cable, and steel electrodes. A total of 61 steel electrodes were used, which is the maximum number of electrodes that can be used for two Lund imaging cables. There are two sets of electrodes, one using 2-metre electrodes for the long setup and another using 1-metre electrodes for the short setup. The spacing for the long setup was 1 metre, while the spacing for the short setup was 0.5 metres. Reducing the distance between measurements can enhance the number of measurements collected, resulting in more detailed outcomes and decreasing the likelihood of noise that may introduce ambiguity into the results (Baines et al., 2002). The distance between electrodes directly affects the number of measurements, the scope of research, the vulnerability to interference, and the accuracy in both horizontal and vertical dimensions. The ZondRes2d software was used for data processing and inversion. The ZondRes2d inversion code facilitated the development of a two-dimensional inverse model of resistivity measurements in the subsurface using the measured apparent resistivity (Griffiths & Barker, 1993).

3.3 Comparison between ERT and Peat Sampler

Compared with conventional peat samplers, Electrical Resistivity Tomography (ERT) offers several practical benefits. ERT is a non-invasive method that reduces environmental disruption and preserves the peatland's integrity, which is essential for conservation efforts and in areas where protecting natural ecosystems is crucial. Additionally, ERT is more efficient for large-scale studies, as it can gather extensive subsurface data across wide areas in a shorter time. In contrast, although reliable for acquiring direct physical samples, traditional peat samplers require significant labour, are time-consuming, and involve intrusive methods that may disturb the peatland.

From a financial standpoint, while ERT may require a larger initial investment in equipment than conventional peat samplers, its capability to survey extensive areas and deliver continuous

DOI: https://doi.org/10.22452/mjs.vol45sp1.11
Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026)

data makes it more cost-effective over time. Moreover, ERT offers better spatial resolution, allowing for detailed insights into lateral and vertical variations in subsurface resistivity, which is vital for comprehending peat distribution and characteristics. On the other hand, conventional peat samplers yield localised data points that may not represent the broader variability of the peatland accurately.

However, ERT has its drawbacks, such as sensitivity to noise from environmental factors and the need for specialised software and expertise for data analysis. In contrast, conventional peat samplers provide direct physical samples for laboratory examinations, offering complementary insights, such as organic content, humification levels, and cation exchange capacity, which can enhance the interpretation of ERT results.

In conclusion, ERT offers notable advantages regarding non-invasiveness, efficiency, and extensive coverage, making it a valuable tool for studying peatlands. Nevertheless, traditional peat samplers are essential for validating and supplementing ERT data, ensuring a comprehensive understanding of peatland characteristics.

4. Results and discussion

Peat properties tests were also conducted to obtain information about the type of peatlands and the specific conditions under which the study was conducted. Table 1 summarises the peat properties in the study area.

Peat soil retains a significant amount of moisture due to its organic materials and the remains of plant cells, which have the ability to store water. A soil moisture test conducted on peat samples revealed changes in measurement due to the elevated water levels and the soil's high water-holding capacity. Its loose and porous texture allows the soil fibres to retain substantial water owing to their inherent high retention abilities (Huat et al., 2011). The results indicate a considerable amount of organic matter typical of peat soils, along with inorganic substances derived from various environmental and geological processes (Mesri & Ajlouni, 2007).

The peat soil in the examined area displays a high liquid limit of 178.22%, attributed to its large proportion of organic matter and moisture, significant humification, high fibre content, and the distinctive structural traits associated with peat (Huat et al., 2011). The classification of Parit Nipah peat soil as haemic peat, along with specific plant and organic materials, environmental conditions that preserve fibrous elements, and the soil's geotechnical features, contributes to the observed 58.68% fibre content. The elevated fibre content in the samples can be linked to these factors. Adnan (2008) identified that the specific gravity of peat ranges from 1.1 to 1.9, based on the BS 1377: Part 2: 1990, Clause 8.4 protocol. The high organic content, minimal mineral matter, and increased porosity contribute to a lower density when compared with mineral soils (Adnan, 2008). The acidity of peat soil results from the high production of organic acids during the decomposition of plant matter, the absence of neutralising minerals, anaerobic conditions during peat formation, and pyrite oxidation, with all these factors contributing to increased acidity in peat soils (Hikmatullah & Sukarman, 2014).

4.1 Soil Profile from Peat Sampler

The peat sampler was utilised at five specific coordinates to gather samples: (N 01°50.138’, E 103° 11.037’) for borehole 1, (N 01°50.150’, E 103°11.034’) for borehole 2, (N 01°50.144’, E103°11.040’) for borehole 3, (N 01°50.133’, E 103°11.034’) for borehole 4, and (N 01°50.129’, E 103°11.042’) for borehole 5.

The soil profiles obtained are presented in Figure 2 through Figure 6. The findings indicated that peats were identified at depths of 4 metres in boreholes 2, 3, and 5 and at a depth of 4.5 metres in boreholes 1 and 4. Subsequently, there was a transitional stratum comprising a mixture of peat and soft clay, followed by a soft clay layer. The table shows identifiable fibres, such as roots, which serve as indicators of various kinds of peat. The acquired sample exhibits subtle variations in peat colours, transitioning from a dark reddish-brown to a reddish-brown tone. This implies that the extent of peat decomposition varies depending on the depth. The extent of decomposition is directly influenced by factors such as moisture, organic composition, and fibre characteristics.

After analysing the tables, boreholes 1, 2, 3, and 5 exhibit an absence of peat soil in the peat sampler when extracted at depths ranging from 0.5m to 1.0m. This is attributed to the fibrous nature of the peat. The soft roots could not be included within the peat sampler, resulting in the disappearance of a small amount of peat soil and water that was removed due to the fibrous nature of the peat.

Table 1: Peat Properties

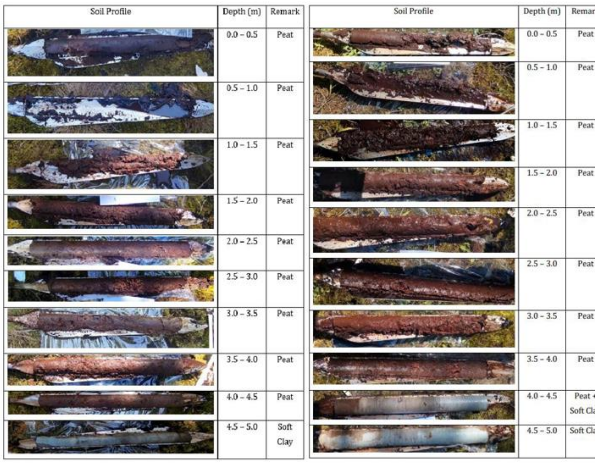
Peat properties	Value	Previous study
Moisture content	689.84 %	250 – 985.40% (N. Wahab et al, 2019)
Ash content	16.47 %	0.8% to 32.8% (A.A.M. Addly et al., 2022)
Organic content	83.53 %	95.88-98.48% (Wahab A. et al., 2022)
Fiber content	58.68 %	40 – 63% (Zainorabidin A. et al., 2014)
Liquid limit	178.22 %	220 – 417 % (Prabir K. et al., 2017)
Specific gravity	1.28	1.1 – 1.9 (N. Wahab et al, 2019)
pH	3.88	3 – 4 (Razali et al., 2013)

DOI: https://doi.org/10.22452/mjs.vol45sp1.11
Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026)

4.2 Soil Resistivity Profile

Geoelectrical technology, particularly Electrical Resistivity Tomography (ERT), was utilised to ascertain the soil strata in Parit Nipah. The primary objective of these experiments was to examine geophysical methodologies for determining the depth of peat in the designated study areas. Two surveys of Electrical Resistivity Tomography (ERT) were carried out in Parit Nipah. The resulting 2-D stratigraphic profiles are illustrated in Figure 7 and Figure 8. The obtained images display clear differentiation between soil layers, facilitating precise evaluation of the depth of peat. The initial line, with a total of 80 metres, consisted of a brief setup of 1 metre and an extended setup of 2 metre. The second line, with a total length of 40 metres, consisted of a 0.5 metre short configuration and a 1-metre-long setup.

5. Correlation Between Ert Method And Peat Sampler



(a) (b)

Figure 2. (a) Soil Profile obtained from peat sampler at borehole 1. (b) Soil Profile obtained from peat sampler at borehole 2.

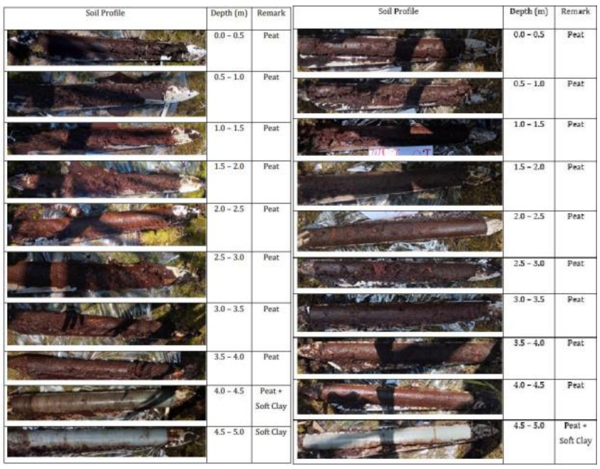


Figure 3. (a) Soil Profile obtained from peat sampler at borehole 3. (b) Soil Profile obtained from peat sampler at borehole 4.

The peat soil profile acquired from the peat sampler was overlaid for comparison purposes. A comparative analysis was carried out between the soil stratigraphy attained using the electrical resistivity approach and that obtained using the peat sampler. The percentage error was calculated using the conventional formula [(observed value – exact value)/exact value x 100]. Tables 2 and 3 summarise the comparison between the soil profiles obtained using the Electrical Resistivity Tomography (ERT) technology and the peat sampler. The layer was constructed with data obtained from a prior study in Parit Nipah. The study determined that the resistivity of peat in Parit Nipah ranges from 100.8 to 139.5 ohm.m, whereas clay resistivity ranges from 5.4 to 22.3 ohm.m (1).

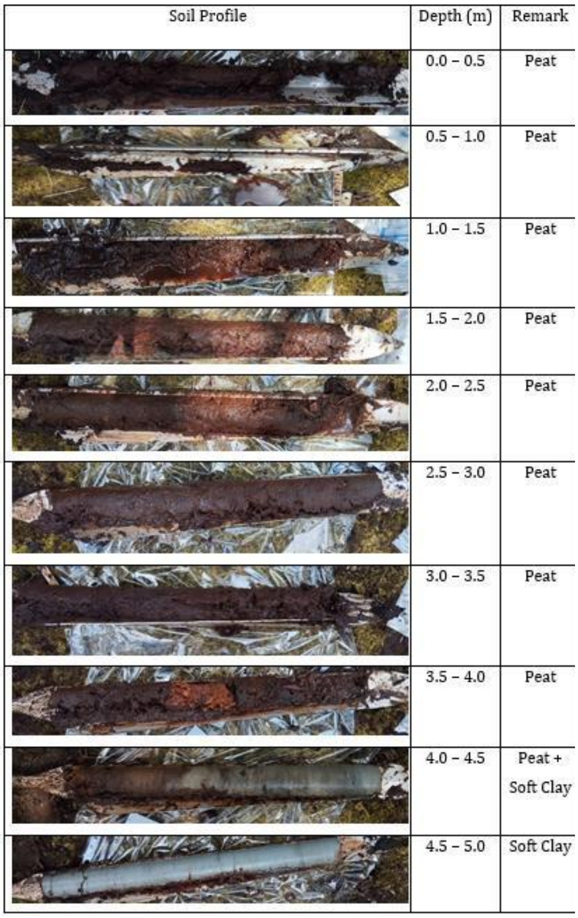


Figure 4. Soil Profile obtained from peat sampler at borehole 5.

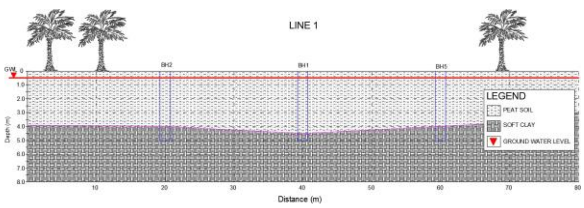


Figure 5. Soil stratigraphy at Parit Nipah, Johor (Line 1)

DOI: https://doi.org/10.22452/mjs.vol45sp1.11
Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026)

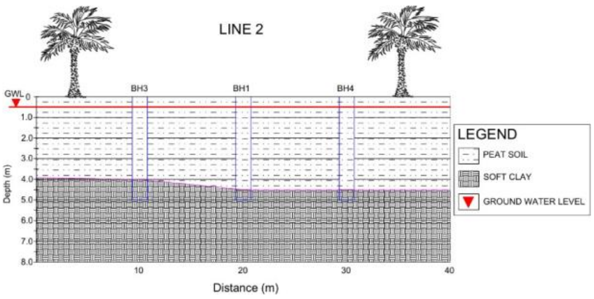


Figure 6. Soil stratigraphy at Parit Nipah, Johor (Line 2)

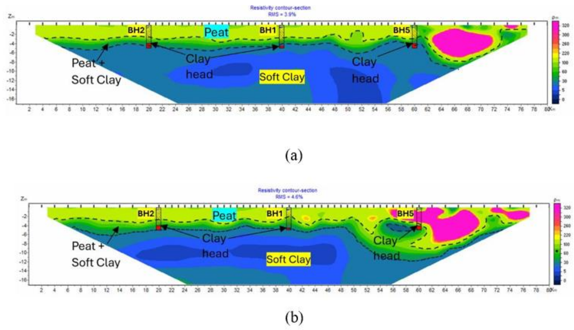


Figure 7. (a) Schlumberger array for line 1 (80m). (b) Wenner array for line 1 (80m).

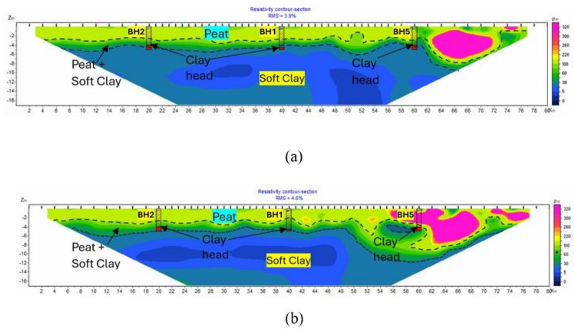


Figure 8. (a) Schlumberger array for line 1 (80m). (b) Wenner array for line 1 (80m).

The error percentages for the peat layer on Line 1, as determined using the Schlumberger array and Wenner array, were 23% and 30%, respectively, compared to the data acquired from the peat sampler. In Line 1, the transition layer between peat and soft clay showed percentage inaccuracy of 6% and 7% for the Schlumberger array and Wenner array, respectively. The summarised data is presented in Table 2.

The error percentages for the peat layer on Line 2, as determined by the Schlumberger array and Wenner array, were 21% and 25%, respectively, compared to the data obtained from the peat sampler. The Schlumberger array and Wenner array methods had percentage errors of 5% and 17%, respectively, in the transition layer between the peat and soft clay layer at Line 2. The summarised data is displayed in Table 3.

To summarise, the results indicate that decreasing the space between electrodes in the Schlumberger array results in enhanced accuracy and heightened sensitivity, in contrast with using wider electrode spacing. Okpoli (2013) noted that reducing the distance between electrodes results in a shorter penetration depth, hence enhancing sensitivity to changes in the electrical resistivity of the subsurface. However, using the Wenner array in the peat region, Line 2 had a smaller margin of error than Line 1, supported by Okpoli's findings. However, in the transition zone, Line 2 showed a larger disparity than Line 1, mainly due to discrepancies between the extrapolated software data and the actual data collected in the field. This may clarify why the accuracy differences are more noticeable for electrodes with shorter intervals on Line 2 than on Line 1. The results also showed that the electrical resistivity method effectively identifies differences in soil layers, and the most favourable technique for peat profiling is the Schlumberger array, as it demonstrates little inaccuracy.

6. Conclusion

Electrical Resistivity Tomography (ERT) technology for evaluating soil profiles has shown promising results. The soil profile obtained through the ERT method closely aligns with data gathered from the peat sampler, indicating that ERT effectively and efficiently characterises soil profiles, making it a suitable option for peatland research.

Among the various methods tested, the Schlumberger technique was particularly effective for profiling peat, as it had minimal discrepancies compared with the peat sampler data. Additionally, the ERT method improves investigative capabilities and avoids invasive techniques, making it an appealing choice for non-invasive subsurface exploration in peatlands.

To further enhance the efficacy of ERT in analysing peatlands, the study recommends focusing on essential peat characteristics, including organic content, moisture content, and the degree of humification. Incorporating these factors into the analysis will improve the understanding of the diverse peat resistivity values, thereby improving the accuracy and reliability of ERT findings.

In summary, this study underscores the promise of ERT as a valuable non-invasive tool for characterising peatlands. The strong correlation between ERT-derived soil profiles and peat sampler data highlights its relevance for geotechnical and environmental research. Its cost efficiency and ability to reduce intrusive methods further reinforce its importance for extensive peatland studies. The broader implications suggest that ERT could significantly assist land managers and conservationists in the sustainable management of peatlands. Practical applications include mapping peat thickness for carbon stock evaluations, monitoring subsurface water flow, and pinpointing areas at risk of degradation. Future developments in ERT techniques may further transform peatland research and conservation efforts by addressing the challenges associated with complex peat characteristics.

DOI: https://doi.org/10.22452/mjs.vol45sp1.11
Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026)

Table 2: Summarized data for line 1

	Line 1					
Protocols	**Schlumberger**			**Wenner**		
Peat Area (PS) (m)	0 – 4.0	0 – 4.5	0 – 4.0	0 – 4.0	0 – 4.5	0 – 4.0
Peat Area (ERT) (m)	0 – 3.0	0 – 3.0	0 – 3.8	0 – 2.5	0 – 2.1	0 – 4.0
Error (%)	25	33	5	37	53	0
Average Error (%)		23			30	
Transition Area (PS) (m)	4.0 – 4.5	4.5 – 5.0	4.0 – 4.5	4.0 – 4.5	4.5 – 5.0	4.0 – 4.5
Transition Area (ERT) (m)	3.0 – 4.3	3.0 – 4.5	3.0 – 4.3	2.5 – 5.0	2.1 – 4.9	4.0 – 4.9
Error (%)	4	10	4	11	2	8
Average Error (%)		6			7	

Table 3: Summarized data for line 2.

	Line 2					
Protocols	**Schlumberger**			**Wenner**		
Peat Area (PS) (m)	0 – 4.0	0 – 4.5	0 – 4.5	0 – 4.0	0 – 4.5	0 – 4.5
Peat Area (ERT) (m)	0 – 3.0	0 – 3.2	0 – 3.0	0 – 2.5	0 – 3.0	0 – 4.2
Error (%)	25	28	33	37	33	6
Average Error (%)		21			25	
Transition Area (PS) (m)	4.0 – 4.5	4.5 – 5.0	4.5 – 5.0	4.0 – 4.5	4.5 – 5.0	4.5 – 5.0
Transition Area (ERT) (m)	3.0 – 5.0	3.2 – 5.0	3.0 – 5.2	2.5 – 5.2	3.0 – 4.6	4.2 – 6.5
Error (%)	11	0	4	15	8	30
Average Error (%)		5			17	

7. Acknowledgement

This research is supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Geran TIER1 (Q859), Geran Penyelidikan Pascasiswazah GPPS (Q791) and GPPS (Q666). Special gratitude to the Research Management Centre (RMC), Faculty of Civil Engineering and Built Environment (FKAAB) and Research Centre for Soft Soil (RECESS) of UTHM for the guidance, encouragement, and valuable support.

8. References

A.A.M. Addly et al. IS THE RESIDUAL ASH METHOD APPLICABLE TO TROPICAL PEATLANDS? Mires and Peat, Volume 28 (2022), Article 08, 16 pp., http://www.mires- and-peat.net/, ISSN 1819-754X International Mire Conservation Group and International Peatland Society, DOI: 10.19189/MaP.2020.GDC.StA.2147

Adnan, A. (2008). Geotechnical Characteristics of Peat. ResearchGate.

Basri, K., Wahab, N., Talib, M. K. A., & Zainorabidin, A. (2019). Sub-surface profiling using electrical resistivity tomography (ERT) with complement from peat sampler. Civil Engineering and Architecture, 7(6).

Baines, D., et al., Electrical resistivity ground imaging (ERGI): a new tool for mapping the lithology and geometry of channel-belts and valley-fills. Sedimentology, 2002. 49(3): p. 441-449.

Binley, A., Hubbard, S. S., Huisman, J. A., & Revil, A. (2002). Geophysical characterization of subsurface properties. Springer.

Everett, M.E., Near-surface applied geophysics. 2013: Cambridge University Press.

Equipment, E.A., Peat Sampler Operating Instruction. Netherlands: Eijkelkamp Agrisearch Equipment, 2014.

Griffiths, D. and R. Barker, Two-dimensional resistivity imaging and modelling in areas of complex geology. Journal of applied Geophysics, 1993. 29(3-4): p. 211-226.

HabibM.Mohamad, B.Kasbi. (2021) Inventigating Peat Soil Stratigraphy and Marine Clay Formation Using the Geophysical Method in Padas Valley, Northen Borneo.

Herman, R., An introduction to electrical resistivity in geophysics. American Journal of Physics, 2001. 69(9): p. 943-952.

Hikmatullah and Sukarman (2014): Cultivated Peat Soils in Trial of ICCTF.

Huat, B. B., Kazemian, S., Prasad, A., & Barghchi, M. (2011). State of an art review of peat: General perspective. International Journal of the Physical Sciences, 6(8), 1988-1996.

Loke, M., Electrical imaging surveys for environmental and engineering studies. A practical guide to, 1999. 2.

Loke, M., Tutorial: 2-D and 3-D electrical imaging surveys. 2004.

Mesri, G., & Ajlouni, M. (2007). Engineering properties of fibrous peats. Journal of Geotechnical and Geoenvironmental Engineering, 133(7), 850-866.

Miele, M., et al. Rectangular Schlumberger resistivity arrays for delineating vadose zone clay- lined fractures in shallow tuff. in Symposium on the Application of Geophysics to Engineering and Environmental Problems 1996. 1996. Society of Exploration Geophysicists.

Mohd et al., Study of Various Soil Stabilization Techniques That Suitable for Parit Nipah Peat Soil. Recent Trends in Civil Engineering and Built Environment Vol. 3 No. 1 (2022) p. 1884-1892

Muhammad Yusa, Ari Sandyavitri, Sigit Sutikno. Application of electrical resistivity test to estimate carbon storage of tropical peat deposit (Case study of Bengkalis island) MATEC Web Conf. 276 05004 (2019) DOI: 10.1051/matecconf/201927605004

N. Wahab, M. K. A. T, K. Basri, and M. M. Rohani, “Segregation peat fiber and pre-consolidation pressure effect on the physical properties of reconstituted peat soil,” Interna tional Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 6S3, pp. 640–647, 2019.

Norhaliza Wahab, Mohd Khaidir Abu Talib, Ahmad Khairul Abd Malik, Aziman Madun (2023) Effect Of Cement Stabilized Peat On Strength, Microstructure, And Chemical Analysis,Physics And Chemistry Of The Earth,ELSEVIER,129,1,ELSEVIER,1-8,ISSN:14747065

Norhaliza Wahab, Mohd Khaidir Abu Talib, Nurlatifah Jumien, Aziman Madun, Faizal Pakir (2023) Comparative Study Of Peat Properties In Johore Malaysia,Asia-Pacific Journal Of Science And Technology,KHON KAEN UNIVERSITY, THAILAND,28,4,KHON KAEN UNIVERSITY, THAILAND,1-8,ISSN:25396293

Okpoli, C.C., Sensitivity and resolution capacity of electrode configurations. Geophysics, 2013.

Prabir K. Kolay and Siti Noor Linda Taib (2017). Physical and Geotechnical Properties of Tropical Peat and Its Stabilization. DOI: 10.5772/intechopen.

Razali, Siti & Hj. Bakar, Ismail & Zainorabidin, Adnan. (2013). Behaviour of Peat Soil in Instrumented Physical Model Studies. Procedia Engineering. 53. 145–155. 10.1016/j.proeng.2013.02.020.

Reynolds, J. M. (2011). An introduction to applied and environmental geophysics. John Wiley & Sons.

Slob, E., Optimal acquisition and synthetic electrode arrays, in SEG Technical Program Expanded Abstracts 2004. 2004, Society of Exploration Geophysicists. p. 1389-1392.

Wahab, A., Hasan, M., Mohd Kusin, F., Embong, Z., Zaman, Q. U., Babar, Z. U., & Imran, M.

S. (2022). Physical Properties of Undisturbed Tropical Peat Soil at Pekan District, Pahang, West Malaysia. International Journal of Integrated Engineering, 14(4), 403 414. https://penerbit.uthm.edu.my/ojs/index.php/ijie/article/view/8667

Zainorabidin, A., & Zolkefle, S. N. A. (2014). Dynamic behavior of Western Johore Peat, Malaysia. Published by Asia Pacific Chemical Biological and Environmental Engineering Society (APCBEES)

DOI: https://doi.org/10.22452/mjs.vol45sp1.11
Malaysian Journal of Science 45 (Special Issue 1): 88-94 (June 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

FAST-UTHM Global Nexus Conference 2024

Special issue

Malaysian Journal of Science 45 (Special Issue 1): 95 -103 (June 2026)

https://mjs.um.edu.my

Investigation Public Holiday is Factor Affecting Dengue Cases by Using Arima and Arimax Model in Selangor and Sarawak from 2015 to 2019

Nur Dayana Abdul Rahman[1a], Loshini Thiruchelvam[2a*], Nur Balqishanis Zainal Abidin[3b,f], Sarat Chandra Dass[4c], Balvinder Singh Gill[5d], Nur Amalina Mat Jan[6a] and Vijanth Sagayan Asirvadam[7e]

Abstract: This study investigated the relationship between dengue cases and human movement, as represented by public holidays in Malaysia for two states, Selangor and Sarawak, using a weekly dataset from 2015 to 2019. This study developed dengue prediction models by employing Autoregressive Integrated Moving Average (ARIMA) and ARIMA with public holiday as the exogenous variable (ARIMAX) time series models. The Box Jenkins approach for model development with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) selection criteria are applied to select the optimum model by using the training dataset. This study, however, found the positive effect of human movement during public holidays is only apparent in three study areas; Klang and Sabak Bernam districts in Selangor state and Sibu division in Sarawak. The AIC and BIC values for Klang, Sabak Bernam, and Sibu were (-573.51, -556.89), (-162.93, -146.32), and (-446.64, -420.06), respectively. Despite this, the difference in the AIC and BIC values between ARIMA and ARIMAX models in other study areas is not large, and the study further found that the ARIMAX model of each study area did not fail to perform comparably well, with some study areas showing better accuracy scores when compared to their counterpart ARIMA model in the testing dataset. This concludes that public holidays, specifically the effect of human movement during that period, are significant factors for predicting dengue cases.

Keywords: Box Jenkins, dengue, Malaysia, prediction models, ARIMAX

1. Introduction

Dengue is a viral disease, transmitted by Aedes mosquito vectors and occurs in the tropical and subtropical regions of the world, including our country, Malaysia. In Malaysia, Selangor has been recognized as the most urbanized state, with several well-known districts such as Petaling and Klang, known to have the highest cases of dengue in previous years. Locations that can hold still water, such as containers, water tanks, or pots, are the types of places where mosquitoes can breed in large numbers. There are six stages in the process of dengue infection, as illustrated in Figure 1 below (Thiruchelvam et al., 2017). The metamorphosis of mosquitoes' life cycles consists of eggs that develop into larvae, then pupae, and finally into adult mosquitoes (Rueda, 2008). This process occurs over one to three weeks during the first stage of the dengue infection.

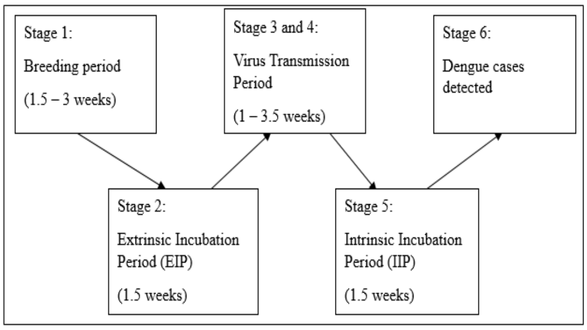


Figure 1. Stages of dengue infection process.

Dengue virus can be transmitted only by female Aedes mosquitoes after they bite infectious humans who have the dengue virus (Pliego Pliego et al., 2017), and it takes on average around two weeks for those female Aedes mosquitoes to activate the dengue virus inside their bodies (Chan & Johansson, 2012; Thiruchelvam et al., 2018). This process occurs in stage two and is known as the Extrinsic Incubation Period (EIP). Stages three and four are the transmission periods of the dengue virus that are transmitted by mosquitoes and the infected human, respectively, and usually occur around one to four weeks. Here, an infected female Aedes mosquito will bite an uninfected human, causing that human to acquire the dengue virus. Meanwhile, in stage four, the infected human in stage three will transmit the dengue virus

Authors information:
[a]Department of Physical and Mathematical Science, Faculty of Science, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, MALAYSIA. E-mail: dynraz@1utar.my[1], loshini@utar.edu.my[2], amalinaj@utar.edu.my[6],
[b]Faculty of Computing & Informatics (FCI), Multimedia Universiti, Persiaran Multimedia, 63100 Cyberjaya, Selangor. E-mail: nurbalqishanis@mmu.edu.my[3]
[c]School of Mathematical and Computer Sciences, Heriot–Watt University Malaysia, Putrajaya. E-mail: s.dass@hw.ac.uk[4]
[d]National Public Health Laboratory (MKAK), Sungai Buloh. Ministry of Health Malaysia. E-mail: drbsgill@moh.gov.my[5]
[e]Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 31750 Seri Iskandar, Perak, Malaysia. E-mail: vijanth_sagayan@utp.edu.my[7]
[f]Centre for Advanced Analysics, CoE Artificial Intelligence, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia

*Corresponding Author: loshini@utar.edu.my

Received: May 25, 2025
Accepted: July 30, 2025
Published: June 30, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

to a new uninfected female mosquito which bites them during a blood meal (Cheong et al., 2013). Thus, there is a greater chance for other healthy humans to be infected too in the future. As for stage five, it is the Intrinsic Incubation Period (IIP), which is the time taken for the infected human to have the dengue virus activated in their body and takes about two weeks to occur (Chan & Johansson, 2012; Thiruchelvam et al., 2018).

In the last stage, the infected human will have onset of symptoms of dengue, such as fever, rashes, and vomiting (Thiruchelvam et al., 2017), after which he or she will seek a doctor's consultation. About a week is needed for the clinical test results to confirm if the person has dengue fever. Thus, on average, it takes about one to two weeks from when the infected person starts having dengue symptoms to clinical confirmation of dengue cases.

Considering the time period between the occurrences of virus transmission and dengue cases being detected clinically to be on average two to three weeks, this study seeks to investigate if there is a possibility of higher dengue cases being detected when there is a public holiday during the two to three weeks before the virus transmission period. Here, public holidays are used as a proxy for human movement, with the hypothesis that when there is a public holiday, more people might travel into the specified district, causing more human-mosquito contact and thus leading to more spread of infectious Aedes mosquito bites.

Statistical approaches using Time Series Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) models are among the most commonly used methods in epidemiological research for predicting infectious diseases, such as dengue (Mahmud et al., 2021). Moreover, ARIMAX is a significant and recognized statistical model that allows exogenous variables to improve the model's forecasting accuracy (Islam et al., 2023). Thus, with the above rising issues, this study aims to determine if travel during public holidays increases the spread of dengue cases within districts in Selangor by using ARIMAX models. Additionally, several divisions from Sarawak state will also be investigated since these two states differ widely in terms of ethnicity.

2. Materials and methods

2.1 Study region

This study focuses on two states in Malaysia: Selangor in Peninsular Malaysia, located in the west of Malaysia, and Sarawak on Borneo Island, located in the east of Malaysia. Based on the Disease Control Division under the Ministry of Health Malaysia (MOH), Malaysian Space Agency (MYSA), and Ministry of Science, Technology and Innovation Malaysia (MOSTI) (2023), Selangor recorded the highest dengue cases with a total of 47,953 cases from December 2019 until October 2023. Another state, Sarawak, is taken as a comparison because it may have differences in culture and travel behaviour compared to Selangor due to significant differences in ethnicity between these states. Selangor consists of Malay and Bumiputera as the largest population, followed by Chinese, Indians, and some minority groups (Selangor Kawasanku, 2023), whereas Sarawak consists of Iban as the largest population, followed by Chinese, Malay, Bidayuh, Melanau, and a number of other local ethnic groups too (Badaruddin et al., 2024).

Figure 2 presents nine districts in Selangor: Gombak, Hulu Langat, Hulu Selangor, Klang, Kuala Langat, Kuala Selangor, Petaling, Sabak Bernam, and Sepang. According to the Selangor Travel (2024) website, the highest population district in Selangor is Petaling, followed by Hulu Langat and Klang. Petaling has the largest population in this state because of its urban centre, modern facilities, healthcare institutions, international colleges, and its public transportation and infrastructure are well developed, making it easy to access any place. Petaling is the most progressive district due to its rapid urbanisation. Hulu Langat's main economic activity is agriculture, and it is popular for its natural surroundings. As for Klang, it was the capital city of Selangor from 1875 until 1880.

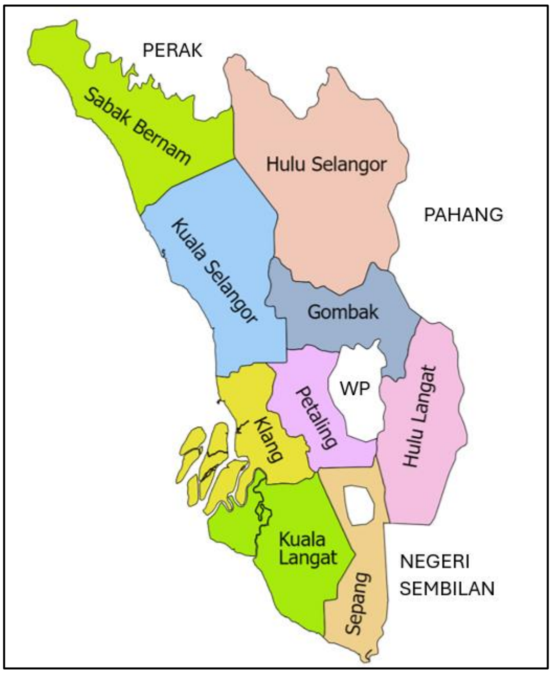


Figure 2. Districts in Selangor.

Table 1 presents the estimated area, population, and population density for nine districts in Selangor. As shown in Table 1, the most densely populated district is Petaling, followed by Klang and Hulu Langat, whereas the least dense district is Hulu Selangor, followed by Kuala Selangor and Kuala Langat.

Meanwhile, Sarawak consists of more than ten divisions, as illustrated in Figure 3. In this study, division is used as the unit of analysis, with only three divisions being examined, which are Kuching, Miri, and Sibu. These three divisions are selected due to the rapid expansion of developing cities along with the increasing population, where most of the forested area is used for urban development (Harvie et al., 2020). Based on the Sarawak Government (2024) website, Kuching division comprises three districts: Kuching, Lundu, and Bau districts, while Miri division includes Miri and Marudi districts. There are three districts in Sibu division: Sibu, Kanowit, and Selangau districts. Table 2 provides the values of estimated areas, population, and population density

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

for these divisions. As shown in Table 2, Kuching is the most densely populated division, whereas Miri and Sibu have relatively lower population densities compared to Kuching.

Table 1: Estimated Area and Population of Selangor District.

District	Estimated Area (km²)	Population	Population Density (people/km²)
Gombak	650	942,400	1,449.85
Hulu Langat	840	1,400,461	1,667.21
Hulu Selangor	1,756	243,029	138.40
Klang	626	1,088,942	1,739.52
Kuala Langat	858	307,449	383.33
Kuala Selangor	1,195	281,711	235.74
Petaling	484	2,298,130	4,748.20
Sabak Bernam	997	107,057	107.38
Sepang	600	325,244	542.07

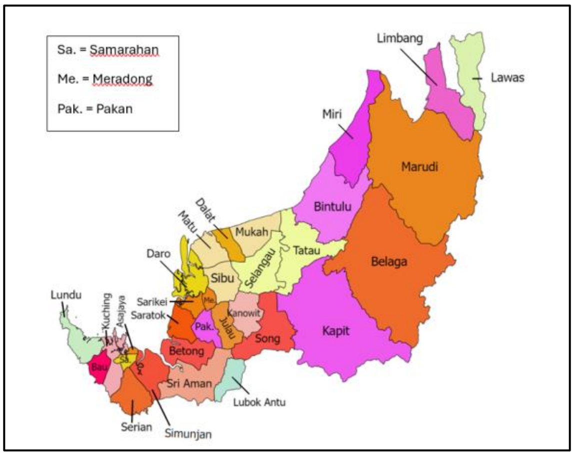


Figure 3. Districts in Sarawak.

Table 2: Estimated Area and Population of Sarawak District.

District	Estimated Area (km²)	Population	Population Density (people/km²)
Kuching	4,560	812,900	178.27
Miri	26,777	433,800	16.20
Sibu	8,278	349,700	42.24

2.2 Public Holidays

Table 3 shows a list of public holidays during the study period from 2015 to 2019 for the Selangor and Sarawak states. There are two types of holidays: national and state holidays. The national holidays include Chinese New Year, school holidays, Eid Fitri, and Eid Adha, while the state holidays are Gawai (orange-colored) and Deepavali (green-colored), celebrated by Sarawak and Selangor states, respectively.

Before including public holiday data in ARIMAX models, the data are transformed into binary variables. For example, the weeks with any public holidays listed in Table 3 are considered as '1' while the weeks that have no holidays are listed as '0'.

2.3 Data Description

Data on dengue cases were obtained from the Ministry of Health (MoH) Malaysia and represent a weekly dataset from 2015 until 2019, spanning a total of 260 weeks. The range and average of dengue cases for each study region are presented in Table 4. The data were then cleansed and normalized before being imported into the R programming language to develop Autoregressive Integrated Moving Average (ARIMA) and ARIMAX models. As this study uses a two-week lag of dengue cases, the data for 2015 started at week three, while other years began at week one. The data were then divided into two parts for training and testing the models.

3. Methodology

$$Y_t' = \phi_1 Y_{t-1}' + \phi_2 Y_{t-2}' \dots + \phi_p Y_{t-p}' + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

3.1 ARIMA and ARIMAX models for dengue prediction

$$Y_t' = \sum_{t=1}^{5} \gamma_i A_i + \sum_{l=1}^{n_b} \sum_{i=a}^{m_l} \beta_i X_{t-i}'^{(l)} + \sum_{i=a}^{m_l} \beta_i W_{t-i}^{(l)} + \phi_1 Y_{t-1}' + \phi_2 Y_{t-2}' + \cdots + \phi_p Y_{t-p}' + \varepsilon_t - \theta_1 \varepsilon_{t-1} + -\theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

Both ARIMA and ARIMAX models can be represented as the following equations:
and

where Y'$_t$ is the stationarized dengue observation at time *t*. A_i is the population density at time *i* with $i = 1,2, \dots 5$ since the study uses a five-year timeline. Coefficients of γ_i are the respective values of population density to be estimated. $X_{t-i}'^{(l)}$ refers to the climate parameters to be included. The n_b in the climate parameter refers to the number of climate parameters being included. Variable $W_{t-i}^{(l)}$ is the public holiday binary information, denoted as 1 for presence and 0 for absence of public holidays during time $(t - \mathrm{i})$ and ε_t is the error value of the model at time t. The l and m_l in both climate and public holiday variables refer to lag order in the ARIMAX model and the maximum lag order considered, respectively, with a being the initial incorporated lagged-time.

The vectors $\beta_i = a, \dots, m_l$, $\underline{\phi} = (\phi_1, \phi_2, \dots, \phi_\mathrm{p})$ and $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_\mathrm{q})$ are the unknown parameters to be estimated, with each $\phi_\mathrm{i} \in \Re$ in (1) and (2) representing the correlation for previous dengue cases and $\theta_\mathrm{i} \in \Re$ in (1) and (2) representing the correlation for moving average of previous error terms to be included. The parameters are estimated using the Maximum Likelihood (ML) approach, and residual diagnostics are carried out on the obtained models to ensure the errors are white-noise.

There are two phases in this section, where Phase I involves the development of the ARIMA model and Phase II involves the inclusion of public holiday information, in the form of binary data, into the finalized ARIMA model of Phase I, forming the ARIMAX model. Both phases consist of three stages: model identification, model estimation, and diagnostic check. The detailed steps of these three stages are illustrated in Figures 4 and 5.

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

Table 3: List of Public Holidays in Selangor and Sarawak Based on Year (2015-2019).

List of public holidays in Selangor and Sarawak based on year				
2015	2016	2017	2018	2019
Week 07 Chinese New Year	Week 6 Chinese New Year	Week 4 Chinese New Year	Week 7 Chinese New Year	Week 6 Chinese New Year
Week 11 School holidays (1^{st} term)	Week 11 School holidays (1^{st} term)	Week 12 School holidays (1^{st} term)	Week 12 School holidays (1^{st} term)	Week 13 School holidays (1^{st} term)
Week 22 Mid-year school holidays and Gawai	Week 22 Mid-year school holidays and Gawai	Week 22 Mid-year school holidays and Gawai	Week 22 Gawai	Week 22 Mid-year school holiday, Gawai and Eidul Fitri
Week 23 Mid-year school holidays	Week 23 Mid-year school holidays	Week 23 Mid-year school holidays	Week 24 Mid-year school holidays and Eidul Fitri	Week 23 Mid-year school holidays
Week 28 Eidul Fitri	Week 27 Eidul Fitri	Week 26 Eidul Fitri	Week 25 Mid-year school holidays	Week 33 School holidays (2^{nd} term) and Eid Adha
Week 38 School holidays (2^{nd} term) and Eid Adha	Week 37 School holidays (2^{nd} Term) and Eid Adha	Week 35 School holidays (2^{nd} Term) and Eid Adha	Week 34 School holidays (2^{nd} Term) and Eid Adha	Week 44 Deepavali
Week 45 Deepavali	Week 44 Deepavali	Week 42 Deepavali	Week 45 Deepavali	Week 48-52 School holidays (last term)
Week 47-52 School holidays (last term)	Week 48-52 School holidays (last term)	Week 48-52 School holidays (last term)	Week 48-52 School holidays (last term)	

Table 4: Range of Dengue Cases for Selangor Districts and Sarawak Division.

District/Division	Range of dengue cases (2015-2019)	Average dengue cases for five years (2015-2019)
Gombak	[34 – 349]	135
Hulu Langat	[80 – 527]	243
Hulu Selangor	[03 – 77]	22
Klang	[70 – 615]	188
Kuala Langat	[04 – 67]	25
Kuala Selangor	[01 – 54]	16
Petaling	[116 – 1025]	391
Sabak Bernam	[01 – 19]	5
Sepang	[10 – 123]	41
Kuching	[01 – 19]	6
Miri	[10 – 123]	3
Sibu	[01 – 40]	11

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

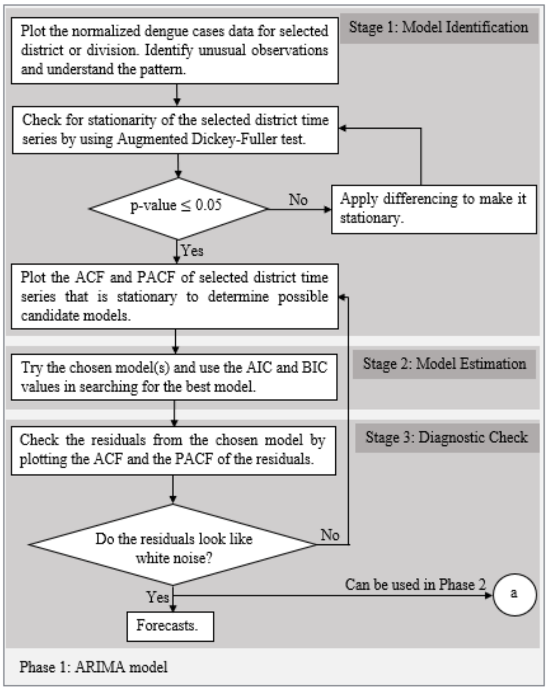


Figure 4. Flowchart of ARIMA model

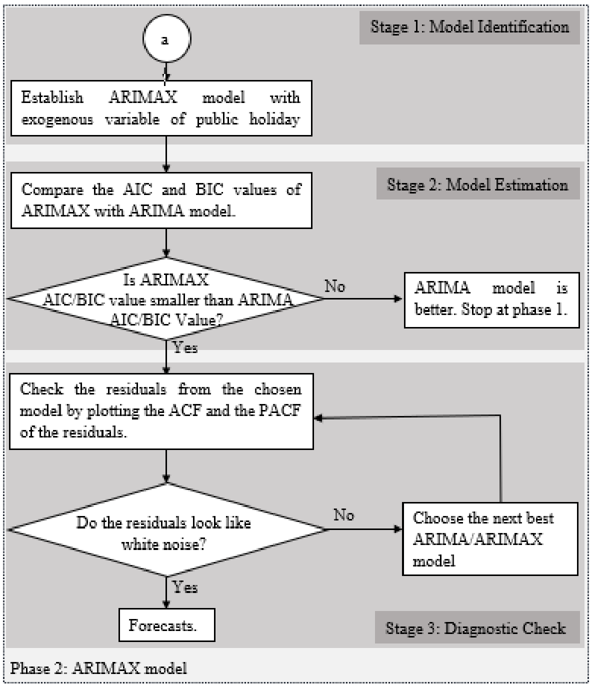


Figure 5. Flowchart of ARIMAX model.

3.2 Measurement Tools

This study employed three measurement tools to select the optimal model: the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Root Mean Square Error (RMSE).

The RMSE formulation with n is as the sample size is formulated as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(Predicted_i - Actual_i)^2}{n}} \quad (3)$$

4. Results

4.1 ARIMA Model Development

Sabak Bernam district is used here as an example to illustrate the development of a dengue prediction model, as can be seen in Figures 6-8. Figure 6 shows a time series of the weekly dengue cases for Sabak Bernam from the year 2015 until 2018. Although there are some high case numbers, it is not seen as observable peaks considering that the time series were within the 0-1 range of normalization. To confirm further, the study looks at the plots of Autocorrelation (ACF) and partial autocorrelation (PACF). Since the spikes managed to decay and fall within the confidence interval bands, it can be concluded that the time series is stationary. This is further confirmed through the Augmented Dickey-Fuller test, which found a p-value of 0.01, rejecting the null hypothesis and confirming that the time series is stationary. Anything within or close to the blue line is statistically close to zero and anything outside the blue line is statistically non-zero (Yasmin & Moniruzzaman, 2024).

Based on Figure 6, there are two significant spikes (at lag one and two) observed in the PACF plot. Thus, the candidate models would be ARIMA (1,0,0) and ARIMA (2,0,0). In choosing the best model fit under this condition, the pair of AIC and BIC ensures avoiding overfitting and underfitting of the model simultaneously. Next, RMSE is used to select the model that gives the best prediction accuracy for the testing data. Thus, applying AIC, BIC, and RMSE ensures selection of a model that has a good fit together with accurate prediction.

Table 5 shows the candidate ARIMA models for Sabak Bernam. Based on Table 5, ARIMA (2,0,0) is selected as the model with the best fit and also having the highest prediction accuracy as the pair of AIC and BIC while the RMSE value of training data is among the lowest. Figure 8 shows that the lags of ACF and PACF for the selected best ARIMA model are within the threshold line, which indicates the residuals are white noise. Hence, the ARIMA(2,0,0) model is the model set for predicting dengue cases in Sabak Bernam district. Figure 9 shows the plotted Sabak Bernam data with forecasted ARIMA models. In choosing the best model fit, two aspects are given importance: first, to ensure the selected model is not too complex, that is, accepting all the proposed variables/parameters into the model and causing an overfitting issue, in which the model tends to learn the current dataset too much. This will cause the selected model not to perform well in a new dataset. This can be addressed through BIC, which tends to select simpler models and avoid overfitting. On the other hand, problems with selecting a too simple model without sufficient information provided through the inclusion of parameters mean that the model cannot display the actual or existing relationship, and thus is unable to perform well with the available dataset itself. This situation can be overcome by introducing the AIC, which safeguards against underfitting error.

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

In summary, applying AIC and BIC model selection criteria simultaneously will ensure selecting the best model fit, with both overfitting and underfitting issues being taken care of. Next, RMSE is used on the testing dataset (for data of the year 2019) to select the model which gives the best prediction accuracy. Thus, applying AIC, BIC, and RMSE ensures selection of a model that has good fit together with accurate prediction.

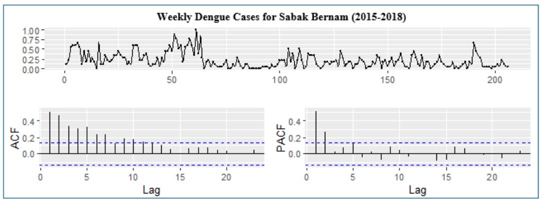


Figure 6. Time series of Sabak Bernam district from the year 2015 until 2018.

```
        Augmented Dickey-Fuller Test

data:  trainsb
Dickey-Fuller = -4.0569, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Figure 7. Result of Augmented Dickey-Fuller test for Sabak Bernam.

Table 5: Parameters for Candidate ARIMA Models in Sabak Bernam.

Candidate Model	AIC value	BIC Value	RMSE
ARIMA(1,0,0)	-141.0769	-131.0933	0.1674164
ARIMA(2,0,0)	-154.76	-141.4485	0.1696358

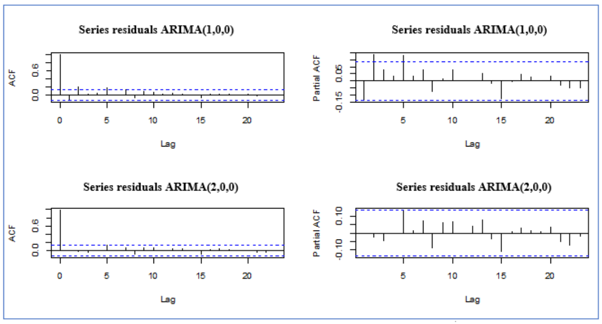


Figure 8. The residuals for each ARIMA model.

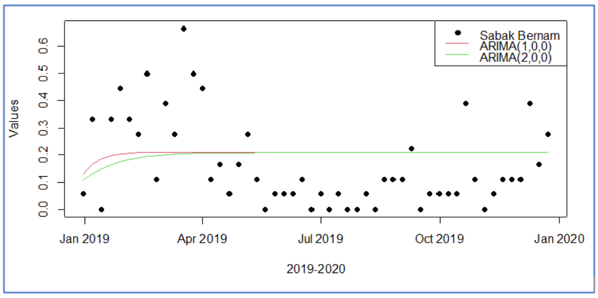


Figure 9. Plotted Sabak Bernam data with forecasted ARIMA models.

4.2 ARIMAX Model Development

The ARIMAX model is an extension of the ARIMA model where X represents the exogenous variable, which in this study is the public holidays, measured on a weekly basis. PH2, PH3, and PH2,3 represent 2-week lagged, 3-week lagged, and 2- and 3-week lagged variables, respectively. Sabak Bernam district is used again to illustrate the ARIMAX model development.

Based on Table 6, it can be seen that ARIMAX(2,0,0,PH2) is the best model as it has the smallest AIC and BIC values compared to other models. After selecting the most fitting and accurate model, a diagnostic check is carried out on the selected model, as shown in Figure 10. The residuals of ARIMAX(2,0,0,PH_2) are found to be white noise. Thus, the model can be finalized.

Figure 11 shows the forecasted values and range of the ARIMAX(2,0,0,PH_2) model for 52 weeks. The model is seen to forecast dengue cases to some acceptable extent. Prior Table 7 highlighted in orange in Table 7, and the bold models show the best model with the lowest AIC and BIC values for each district and division's training dataset. Although only three study areas showed positive results with the inclusion of public holiday variables in the training dataset, the ARIMAX models of each study area did not fail to perform comparably well, with some study areas shown to have better accuracy compared to their counterpart ARIMA model in the testing dataset. This indicates that public holiday, specifically the effect of human movement during that period, is a meaningful factor for predicting dengue cases.

Table 6. Parameters for ARIMAX models in Sabak Bernam.

Candidate Model	AIC value	BIC Value	RMSE
ARIMAX(1,0,0,PH_2)	-152.28	-138.97	0.1806147
ARIMAX(2,0,0,PH_2)	-163.10	-146.46	0.1894346
ARIMAX(1,0,0,PH_3)	-140.94	-127.63	0.1618569
ARIMAX(2,0,0,PH_3)	-153.12	-136.48	0.1674543
ARIMAX(1,0,0,$PH_{2,3}$)	-141.86	-128.54	0.1711214
ARIMAX(2,0,0,$PH_{2,3}$)	-155.51	-138.87	0.1782661

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

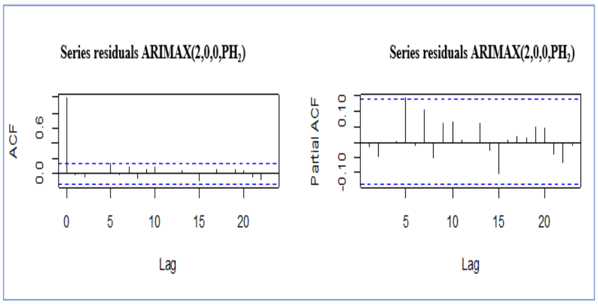


Figure 10. Residuals for Sabak Bernam ARIMAX(2,0,0,PH$_2$) model.

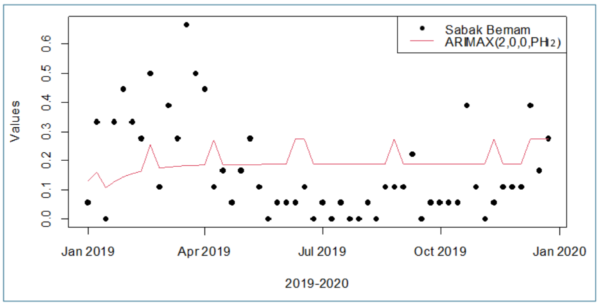


Figure 11. Plotted Sabak Bernam with forecasts of ARIMAX(2,0,0,PH$_2$) model

5. Discussion

Klang and Sabak Bernam districts in Selangor and Sibu division in Sarawak demonstrate better prediction as ARIMAX models in their training dataset, concluding that incorporating public holiday information provides additional data and is not considered redundant in dengue prediction for those study areas. Besides, all the ARIMAX models performed comparably well, with some study areas showing better accuracy, measured through the RMSE values, compared to their counterpart ARIMA models in testing data, thus indicating that ARIMAX models can make accurate predictions as well. The obtained results explain the significant impact of human movements and contacts during public holidays as a major contributing factor to the transmission and spread of dengue in both these study areas. The surplus of human movement, especially during public holidays, was observed in those study areas.

These findings align with prior studies (Thiruchelvam et al., 2021; Ramadona et al., 2023; Syukri & Wardiah, 2023), which show the pattern of dengue cases in a study area is also influenced by the pattern of dengue in its neighboring regions. This is believed to occur due to human movement involving dengue-infected people who might or might not have symptoms yet (Shahid et al., 2020; Hossain et al., 2020). Since once a mosquito bites the infected person, there is always a chance for the virus to be transmitted to other people, spreading the virus to the community through several cycles of mosquito bites (Hossain, 2022). Thus, population movement is considered a possible factor in dengue transmission. Additionally, a study by Siriyasatien et al.

Table 7. AIC, BIC and RMSE Values for All Districts

Division	Model	AIC value	BIC Value	RMSE
Gombak	**ARIMA(1,0,2)**	**-427.22**	**-410.58**	0.3254
	ARIMAX(1,0,2,PH$_3$)	-426.04	-406.08	**0.2716**
Hulu Langat	**ARIMA(2,0,2)**	**-342.84**	**-322.87**	0.2176
	ARIMAX(2,0,2,PH$_2$)	-341.17	-317.87	**0.1828**
Hulu Selangor	**ARIMA(4,0,2)**	**-337.46**	**-310.84**	**0.2537**
	ARIMAX(4,0,2,PH$_2$)	-336.13	-306.18	0.2983
Klang	ARIMA(2,1,2)	-573.73	-557.12	0.3477
	ARIMAX(2,1,2,PH$_2$)	**-574.07**	**-554.13**	**0.3271**
Kuala Langat	**ARIMA(2,1,0)**	**-255.18**	**-245.21**	**0.1707**
	ARIMAX(2,1,0,PH$_2$)	-253.62	-240.33	0.1780
Kuala Selangor	**ARIMA(4,1,2)**	**-319.6**	**-296.34**	**0.1959**
	ARIMAX(4,1,2,PH$_3$)	-319.16	-292.58	0.2039
Petaling	**ARIMA(2,1,3)**	**-532.94**	**-512.99**	**0.1374**
	ARIMAX(2,1,3,PH$_2$)	-531.83	-508.57	0.1452
Sabak Bernam	ARIMA(2,0,0)	-154.76	-141.45	**0.1696**
	ARIMAX(2,0,0,PH$_2$)	**-163.10**	**-146.46**	0.1894
Sepang	**ARIMA(0,1,1)**	**-248.89**	**-242.24**	**0.1732**
	ARIMAX(0,1,1,PH$_{2,3}$)	-246.89	-236.92	0.1734
Kuching	**ARIMA(0,1,1)**	**-542.26**	**-535.62**	**0.3475**
	ARIMAX(0,1,1,PH$_3$)	-540.48	-530.51	0.3491
Miri	**ARIMA(2,0,1)**	**-842.74**	**-826.1**	**0.3786**
	ARIMAX(2,0,1,PH$_3$)	-840.84	-820.88	0.3860
Sibu	ARIMA(5,1,2)	-445.37	-418.78	**0.0737**
	ARIMAX(4,1,2,PH$_2$)	**-446.64**	**-420.06**	0.0771

(2018) mentioned tourism as one of the reasons for people to travel, besides traveling for work or trade.

Our findings, especially for the rural district of Sabak Bernam, align with several previous studies conducted in countries such as Bangladesh, Singapore, and China (Aguiar et al., 2014; Hossain, 2022; Lin et al., 2016; Seah et al., 2021; Shahid et al., 2020; Sippy et al., 2019; Wu et al., 2018). For example, a study by Hossain (2022) concluded that the movement of migrants residing in the

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

capital city, Dhaka, to their hometowns during Eid al-Adha caused peak infections and new outbreaks in those cities. This conclusion was supported by a study by Hossain et al. (2020), which found that after a few weeks of the Eid holiday, most admitted dengue patients were reported from other cities, rather than the capital, Dhaka, which is the common case. This finding also aligns with another study by Rafi et al. (2020), which found that 40% of dengue patients in the northwestern city had a recent history of traveling to Dhaka. A study (Hay, 2013) highlighted that dengue cases increase during any holiday season or carnival, as the city's population is denser at that time, thereby increasing the chances of dengue infection.

6. Conclusion

This study was conducted to investigate the best models for predicting dengue cases in the districts and divisions of Selangor and Sarawak, in Malaysia, respectively, using ARIMA and ARIMAX models. This study has identified the ARIMAX model as a better model in two study districts, Klang and Sabak Bernam, in Selangor state, and also in Sibu division, in Sarawak, with the AIC and BIC values of (-573.51, -556.89), (-162.93, -146.32) and (-446.64, -420.06), respectively, and with the rest of the districts still having comparably good ARIMAX models too. Nonetheless, the difference between AIC and BIC values of ARIMA and ARIMAX models in other study areas is not that large. Furthermore, this study showed that the ARIMAX model of each study area performs comparably well, with some study areas showing better accuracy results when compared to their corresponding ARIMA model in the testing dataset. It can be hypothesized that dengue cases are influenced by human movement from the capital city to semi-urban and rural areas during public holidays in Malaysia.

Additionally, forecasting is a crucial early warning mechanism that can facilitate proactive planning and response for clinical and public health services. Based on these findings, recommendations can be addressed to the government, especially the Ministry of Health (MoH), so that they can take timely action in preventing and spreading awareness about dengue through media such as Twitter, radio, and television, especially whenever cases increase, for example reminding the public of steps such as wearing long clothes, applying mosquito repellents, and avoiding being outdoors early in the morning and evening before dusk, which are known as the peak biting periods of Aedes mosquitoes. Moreover, health services, especially government health services, can prepare themselves to combat dengue.

7. References

Aguiar, M., Rocha, F., Pessanha, J. E. M., Mateus, L., & Stollenwerk, N. (2014). Carnival or football, is there a real risk for acquiring dengue fever in Brazil during holidays seasons? Scientific Reports, 5. https://doi.org/10.1038/srep08462

Chan, M., & Johansson, M. A. (2012). The Incubation Periods of Dengue Viruses. PLoS ONE, 7(11). https://doi.org/10.1371/journal.pone.0050972

Cheong, Y. L., Burkart, K., Leitão, P. J., & Lakes, T. (2013). Assessing weather effects on dengue disease in Malaysia. International Journal of Environmental Research and Public Health, 10(12), 6319–6334. https://doi.org/10.3390/ijerph10126319

Harvie, S., Nor Aliza, A. R., Lela, S., & Razitasham, S. (2020). Detection of dengue virus serotype 2 (DENV-2) in population of Aedes mosquitoes from Sibu and Miri divisions of Sarawak using reverse transcription polymerase chain reaction (RT-PCR) and semi-nested PCR. In Tropical Biomedicine (Vol. 37, Issue 2).

Hossain, M. S. (2022). Megacity-centric mass mobility during Eid holidays: a unique concern for infectious disease transmission in Bangladesh. In Tropical Medicine and Health (Vol. 50, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s41182-022-00417-4

Hossain, M. S., Siddiqee, M. H., Siddiqi, U. R., Raheem, E., Akter, R., & Hu, W. (2020). Dengue in a crowded megacity: Lessons learnt from 2019 outbreak in Dhaka, Bangladesh. PLoS Neglected Tropical Diseases, 14(8), 1–5. https://doi.org/10.1371/journal.pntd.0008349

Islam, M. A., Hasan, M. N., Tiwari, A., Raju, M. A. W., Jannat, F., Sangkham, S., Shammas, M. I., Sharma, P., Bhattacharya, P., & Kumar, M. (2023). Correlation of Dengue and Meteorological Factors in Bangladesh: A Public Health Concern. International Journal of Environmental Research and Public Health, 20(6). https://doi.org/10.3390/ijerph20065152

Lin, H., Liu, T., Song, T., Lin, L., Xiao, J., Lin, J., He, J., Zhong, H., Hu, W., Deng, A., Peng, Z., Ma, W., & Zhang, Y. (2016). Community Involvement in Dengue Outbreak Control: An Integrated Rigorous Intervention Strategy. PLoS Neglected Tropical Diseases, 10(8). https://doi.org/10.1371/journal.pntd.0004919

Mahmud, N., Syuhada Muhammat Pazil, N., Jamaluddin, H., Aqilah Ali, N., Teknologi MARA, U., & Perlis Kampus Arau, C. (2021). Prediction of Dengue Outbreak: A Comparison Between ARIMA and Holt-Winters Methods ARTICLE HISTORY ABSTRACT. ESTEEM Academic Journal, 17, 101–111.

MOH, D. C. D. M. of H., MYSA, M. S. A., & MOSTI, M. of S. T. and I. (2023). Dengue Announcement by State in Malaysia for the year 2023. https://idengue.mysa.gov.my/ide_v3/index.php

Pliego Pliego, E., Velázquez-Castro, J., & Fraguela Collar, A. (2017). Seasonality on the life cycle of Aedes aegypti mosquito and its statistical relation with dengue outbreaks. Applied Mathematical Modelling, 50, 484–496. https://doi.org/10.1016/j.apm.2017.06.003

Rafi, A., Mousumi, A. N., Ahmed, R., Chowdhury, R. H., Wadood, A., & Hossain, G. (2020). Dengue epidemic in a non-endemic zone of Bangladesh: Clinical and laboratory profiles of patients. PLoS Neglected Tropical Diseases, 14(10), 1–14. https://doi.org/10.1371/journal.pntd.0008567

Rueda, L. M. (2008). Global diversity of mosquitoes (Insecta: Diptera: Culicidae) in freshwater. In Hydrobiologia (Vol. 595, Issue 1, pp. 477–487). https://doi.org/10.1007/s10750-007-9037-x

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)

Sarawak Government. (n.d.). Retrieved February 15, 2024, from https://sarawak.gov.my/web/home/article_view/240/175/?id=158#175

Seah, A., Aik, J., Ng, L. C., & Tam, C. C. (2021). The effects of maximum ambient temperature and heatwaves on dengue infections in the tropical city-state of Singapore – A time series analysis. Science of the Total Environment, 775. https://doi.org/10.1016/j.scitotenv.2021.145117

Selangor.travel | Your Smart Online Travel Guide In Selangor. (n.d.). Retrieved January 24, 2024, from https://selangor.travel/

Shahid, F., Ony, S. H., Albi, T. R., Chellappan, S., Vashistha, A., & Alim Al Islam, A. B. M. (2020). Learning from Tweets: Opportunities and Challenges to Inform Policy Making during Dengue Epidemic. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1). https://doi.org/10.1145/3392875

Sippy, R., Herrera, D., Gaus, D., Gangnon, R. E., Patz, J. A., & Osorio, J. E. (2019). Seasonal patterns of dengue fever in rural Ecuador: 2009-2016. PLoS Neglected Tropical Diseases, 13(5). https://doi.org/10.1371/journal.pntd.0007360

Siriyasatien, P., Chadsuthi, S., Jampachaisri, K., & Kesorn, K. (2018). Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. In IEEE Access (Vol. 6, pp. 53757–53795). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2018.2871241

Thiruchelvam, L., Asirvadam, V. S., Dass, S. C., Daud, H., & Gill, B. S. (2017). K-Step Ahead Prediction Models for Dengue Occurrences.

Thiruchelvam, L., Dass, S. C., Asirvadam, V. S., Daud, H., & Gill, B. S. (2021). Determine neighboring region spatial effect on dengue cases using ensemble ARIMA models. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-84176-y

Thiruchelvam, L., Dass, S. C., Zaki, R., Yahya, A., & Asirvadam, V. S. (2018). Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia. Geospatial Health, 13(1), 102–109. https://doi.org/10.4081/gh.2018.613

Wu, X., Lang, L., Ma, W., Song, T., Kang, M., He, J., Zhang, Y., Lu, L., Lin, H., & Ling, L. (2018). Non-linear effects of mean temperature and relative humidity on dengue incidence in Guangzhou, China. Science of the Total Environment, 628–629, 766–771. https://doi.org/10.1016/j.scitotenv.2018.02.136

Yasmin, S., & Moniruzzaman, Md. (2024). Forecasting of area, production, and yield of jute in Bangladesh using Box-Jenkins ARIMA model. Journal of Agriculture and Food Research, 16, 101203. https://doi.org/10.1016/j.jafr.2024.101203

DOI: https://doi.org/10.22452/mjs.vol45sp1.12
Malaysian Journal of Science 45 (Special Issue 1): 95-103 (June 2026)