Vol. 45 | Special Issue 2 | November 2026

MALAYSIAN JOURNAL OF SCIENCE

Special Issue on 5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

MJS Guest Editors
Dr. Ruhaila Md Kasmani, Universiti Malaya

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

Guest Editors
Dr. Nur Haizum binti Abd Rahman, Universiti Malaysia Pahang Al-Sultan Abdullah
Dr. Nadirah binti Mohd Nasir, Universiti Malaysia Pahang Al-Sultan Abdullah
Dr. Norhafizah binti Md Sarif, Universiti Malaysia Pahang Al-Sultan Abdullah
Dr. Ku Muhammad Na'im bin Ku Khalif, Universiti Malaysia Pahang Al-Sultan Abdullah
Dr. Adam Shariff bin Adli Aminuddin, Universiti Malaysia Pahang Al-Sultan Abdullah

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Published: 15-11-2026

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).

The 5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

The 5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025) is an academic conference organized by the Centre for Mathematical Sciences at Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). Taking place on 23–24 May 2025, this event marks the fifth edition of the series, continuing the legacy of previous gatherings held in 2017, 2019, 2022, and 2023. Convening under the theme "Synergies in Mathematics, Statistics, and Data Science: Bridging Insights for a Dynamic World," the conference gathers leading academics, researchers, educators, and industry practitioners worldwide. It serves as a premier interdisciplinary forum for sharing research discoveries, addressing emerging trends, and formulating solutions to practical challenges encountered in the mathematical sciences.

The conference is guided by four core objectives designed to elevate mathematical disciplines and foster global scholarly exchange. Primarily, it seeks to connect renowned, established authorities with emerging researchers across various mathematical, statistical, and data analytics domains. It simultaneously establishes a collaborative and welcoming environment for scholars to present original ideas, share technical methodologies, and contribute to scientific progress. Furthermore, ICoAIMS 2025 promotes mathematics, statistics, and data analytics as vital professions to the wider community, underscoring their critical role in engineering, commercial industries, and organizational decision-making. Lastly, the event cultivates international dialogue, cross-border cooperation, and mutual understanding among mathematicians and data experts globally.

The scope of ICoAIMS 2025 spans eight foundational and applied tracks across the mathematical sciences. These encompass Pure and Applied Mathematics, covering areas such as algebra, topology, dynamical systems, biomechanics, computational methods, fluid mechanics, and wave propagation, as well as Computational Mathematics, which examines numerical modeling, cryptography, and algorithms. The program also places substantial emphasis on modern data and operational fields, including Statistics, Operational Research, and Data Science, spanning topics like machine learning, deep learning, big data analytics, forecasting, and optimization. In addition, interdisciplinary tracks spotlight Engineering & Industrial Applications, including green energy, nanotechnology, robotics, and logistics, alongside developments in Mathematics Education and educational statistics.

The organizers gratefully acknowledge the support of Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), as well as the valued collaboration of Persatuan Sains dan Matematik Malaysia (PERSAMA), Suruhanjaya Syarikat Malaysia (SSM), Mie University (Japan), and Thaksin University (Thailand). Sincere appreciation is extended to the *Malaysian Journal of Science* (MJS) for partnering with the conference to offer publication opportunities through a dedicated Special Issue. Furthermore, heartfelt gratitude is owed to the guest editorial committee, Dr. Nur Haizum binti Abd Rahman, Dr. Nadirah binti Mohd Nasir, Dr. Norhafizah binti Md Sarif, Dr. Ku Muhammad Na'im bin Ku Khalif, and Dr. Adam Shariff bin Adli Aminuddin, as well as the appointed peer reviewers, international contributors, and participating authors whose collective scholarly dedication ensures the success of ICoAIMS 2025.

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026) https://mjs.um.edu.my

Forecasting Electricity Demand using Seasonal ARIMA and Asymmetric GARCH Models: A Case Study for Malaysia

Syarranur Zaim[1a], Siti Roslindar Yaziz[2a]*, Roslinazairimah Zakaria[3a], Noor Fadhilah Ahmad Radi[4b] and Muhammad Hisyam Lee[5c]

Abstract: Accurate energy load forecasting is important for enhancing functionality and reliability of strategy development within the power generation industry. Electricity systems exhibit distinct reactions to positive and negative demand shocks, making asymmetric models particularly useful for capturing these variations. Incorporating asymmetric Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models into electricity demand forecasting can enhance forecasting accuracy by capturing the responses to demand shocks. It has been demonstrated that non-seasonal Autoregressive Integrated Moving Average (ARIMA) with asymmetric GARCH-type model is promising for forecasting highly volatile data. However, Malaysia's electricity demand time series data exhibits strong seasonal fluctuations, necessitating the use of seasonal models. From this perspective, an evaluation is carried out in this study to determine the effectiveness of a seasonal ARIMA model combined with an asymmetric GARCH-type model for forecasting electricity demand in Malaysia. Additionally, it proposes a new procedure for forecasting electricity demand. This study utilized weekly peak electricity usage over the timeframe of 2005 to 2016. The results suggest the optimal combination of seasonal ARIMA with asymmetric GARCH-type model for electricity demand prediction is $\mathrm{SARIMA}(1,1,0)(0,1,0)_{52}$ integrated with $\mathrm{EGARCH}(1,2)$ structure using Student's *t*-distribution. This model was selected due to its simplicity and strong performance, reflected in its low error metrics: root mean squared error (RMSE) of 1021.6340, mean absolute error (MAE) of 867.6399, and mean absolute percentage error (MAPE) of 5.14%. The study demonstrates the model's superiority in capturing volatility and seasonality compared to traditional approaches. Given its enhanced predictive performance, the proposed seasonal ARIMA with asymmetric GARCH-type model provides critical insights for policymakers and energy sector planners in forecasting electricity demand in Malaysia.

Keywords: Forecasting Electricity Demand, ARIMA, GARCH, Asymmetric.

1. Introduction

Electricity is essential due to its feasibility, convertibility, and mobility, but since it cannot be stored economically, continuous supply and demand prediction is necessary (Keyno *et al.,* 2009). Electricity demand in a power system is largely affected by factors such as weather, socio-economic conditions, and demographic variables (Al-Alawi & Islam, 1996). In Malaysia, electricity usage is significantly affected by climate variations and economic dynamics. According to the Department of Statistics Malaysia (2018), the nation's highest recorded electricity usage, reaching 17,788 MW as observed on 19 April 2016, underscores the necessity of precise forecasting, particularly during extreme weather events. The hot and dry conditions during that period led consumers to use electricity at its highest level. Consequently, the evaluation of power system efficiency emphasizes peak load rather than lower load levels.

Furthermore, electricity is a fundamental energy source that underpins modern society, driving socio-economic development, and contributing significantly to various facets of modern infrastructure, including transportation networks, industrial production, resource extraction, and digital connectivity. It also serves as a key tool in shaping strategic planning and future energy policies. With electricity consumption steadily increasing, its diverse applications have driven human civilization to remarkable advancements. As a result, electricity demand is deeply interconnected with all aspects of development (Yasmeen & Sharif, 2014).

In competitive electricity industry, precise prediction of load requirements and electricity pricing is essential for efficient energy system management and operation, as electricity must be delivered in real time due to limited storage capability, making any imbalance between supply and demand a significant challenge for all stakeholders (Lisi & Shah, 2020). Accurate electric load forecasting and modeling are essential for efficient energy use, cost management, and economic stability, helping policymakers and organizations prevent power shortages, reduce waste, and avoid high costs (Qiu, 2013; Kavaklioglu *et al.,* 2009).

Authors information:
[a]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia, E-mail: PSS21001@student.umpsa.edu.my[1], roslinazairimah@umpsa.edu.my[3]
[b]School of Mathematical Sciences, College of Computing Informatics and Media, UiTM Shah Alam, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia, E-mail: noorfadhilah@uitm.edu.my[4]
[c]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia, E-mail: mhl@utm.my[5]

*Corresponding Author: roslindar@umpsa.edu.my[2]

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

To ensure accurate forecasting for effective power system management, various modeling techniques have been proposed including traditional statistical models such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) (Zaim *et al.,* 2025; Duong *et al.,* 2023; Andoh *et al.,* 2021; Goswami & Kandali, 2020) as well as exponential smoothing methods like Simple, Holt's, and Brown's models (Ishak *et al.,* 2022). In addition, advanced machine learning techniques have been investigated (Lee *et al.,* 2022). Hybrid approaches, such as the SARIMA-GARCH model, have also been applied to capture both linear and volatility dynamics in electricity demand (Zaim *et al.,* 2023; Makobea, 2023).

However, predicting electricity demand and prices remains a challenging task because of the irregular and unpredictable patterns in the data (Shah *et al.,* 2022). Moreover, electricity demand often exhibits volatility due to several factors, including unexpected weather changes, economic disruptions, or policy shifts. These shocks may influence demand differently depending on whether they are positive (e.g., an economic boom increasing demand) or negative (e.g., a natural disaster reducing demand).

Electricity systems exhibit distinct reactions to positive and negative demand shocks. Positive shocks, such as higher-than-expected demand, often require rapid adjustments in generation, increasing system instability and operational strain. In contrast, negative shocks, such as lower-than-expected demand, may have minimal impact if the excess capacity can be effectively stored or absorbed, as seen with hydropower reservoirs or battery storage systems. Asymmetric models are particularly useful in quantifying these differing reactions, enabling more accurate forecasting and improved grid management strategies.

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is widely recognized as an effective tool for modeling and forecasting the volatility of time series data (Rakshit *et al.*, 2023). Despite its popularity, the GARCH model has notable limitations, particularly its assumption of symmetry, which overlooks the direction of shocks and fails to reflect how volatility may respond differently to positive and negative shocks of equal magnitude (Rakshit *et al.*, 2023).

The standard GARCH model assumes that the volatility responds equally to shocks, suggesting that shocks of equal magnitude, whether positive or negative, have an identical effect on future volatility. Positive shocks occur when higher-than-expected demand leads to increased production, potentially causing grid instability or requiring operational adjustments. These responses can, in turn, amplify volatility in the electricity system. Negative shocks, such as lower-than-expected demand, might have a smaller or delayed impact due to the presence of demand-response mechanisms or surplus generation capacity. These factors help mitigate the immediate effects of reduced demand on the electricity system. Thus, the inability of the GARCH model to distinguish between the effects of positive and negative shocks limits its applicability in accurately modeling electricity demand volatility. Incorporating asymmetric GARCH models into electricity demand forecasting can enhance forecasting accuracy by capturing the impact of demand shocks. Additionally, they support policy planning by offering insights into how certain events, such as holidays or outages, may asymmetrically affect volatility. By addressing the limitations of symmetric GARCH models and utilizing asymmetric alternatives, forecasts for electricity demand become more robust, particularly in managing volatile and shock-sensitive periods.

Several studies have demonstrated the importance of using asymmetric GARCH models in overcoming the shortcomings of symmetric GARCH models in effectively capturing the effects of positive and negative shocks on volatility. Ndege *et al.* (2023) examined the application of asymmetric GARCH-type models to address the limitations of symmetric GARCH models in forecasting volatility in the Kenyan exchange rate market. The study compared various asymmetric Conditional Heteroscedasticity models, including Exponential GARCH (EGARCH), Threshold GARCH (TGARCH), and Asymmetric Power Autoregressive Conditional Heteroscedastic (APARCH). Using secondary data on exchange rates obtained from the Central Bank of Kenya, spanning from January 1993 to June 2021, the authors identified the most representative model according to the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Log-Likelihood criterion, and error minimization metrics such as mean error (ME) and root mean absolute error (RMAE). Their findings revealed that the ARMA (3,0) - APARCH (1,1) model with a skewed normal distribution is the best fit for the exchange rate data (AIC = -4.6871, BIC = -4.5860). While Rakshit *et al.* (2023) investigated price volatility in the potato markets of major Indian cities, emphasizing the role of seasonal production, weather abnormalities, and perishability in creating volatility. They highlighted that GARCH models cannot capture asymmetric price volatility, where positive and negative shocks of similar magnitude have different effects. To address this, EGARCH, APARCH, and Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) models were employed to analyze weekly wholesale potato prices in Agra, Ahmedabad, Bengaluru, Delhi, Kolkata, and Mumbai. The study confirmed the presence of asymmetry using News Impact Curves (NICs), which quantified the degree of volatility asymmetry across different markets.

Chen *et al.* (2019) employed various GARCH-type models, including GJR-GARCH, APARCH, and EGARCH, in wind power forecasting to account for asymmetry in volatility, observing that negative shocks exert a milder effect compared to positive ones. While, Lim and Sek (2013) examined volatility in the Malaysian stock market by applying both symmetric and asymmetric GARCH models, assessing their performance through error metrics such as mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) on data from January 1990 to December 2010, divided into pre-crisis, crisis, and post-crisis periods. Their findings showed that symmetric GARCH models performed better during stable periods (pre- and post-crisis), while asymmetric GARCH models were more effective during volatile periods like the 1997 financial crisis. This research highlights the importance of using different GARCH models depending on market conditions, offering valuable insights for improving volatility modeling in various applications, such as electricity demand forecasting.

These studies underscore the effectiveness of asymmetric GARCH models in capturing complex volatility dynamics, highlighting their importance in capturing the volatility dynamics

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

of time series data. By addressing the asymmetry in volatility caused by positive and negative shocks, such models demonstrate improved forecasting accuracy compared to their symmetric counterparts. These findings are particularly relevant to electricity demand forecasting, where volatility and shock asymmetry are critical factors, emphasizing the need for similar modeling techniques. By capturing complex volatility dynamics, these asymmetric GARCH-type models provide strong foundation for enhancing electricity demand forecasting.

Through a systematic review, Chalissery *et al.* (2022) found that most highly cited studies in the field used the EGARCH model to capture asymmetric volatility, and many authors considered EGARCH to be the best-performing among asymmetric GARCH models. They also noted a significant contribution in the literature from researchers comparing GARCH model variants and advancing their theoretical foundations. However, Tehrani *et al.* (2022) showed that in the case of European electricity markets, asymmetric GARCH models like SGARCH, GJR-GARCH, and APARCH offered only minor improvements over the simpler symmetric GARCH (SGARCH) model. Their findings suggest that including an asymmetric parameter is not always necessary, as the SGARCH model can effectively capture volatility in practice.

While these studies highlight the significance of asymmetric GARCH models across various domains, there remains a gap in applying these advanced models to electricity demand forecasting, particularly in Malaysia, where electricity demand exhibits unique characteristics such as strong seasonality, volatility, and asymmetric shock responses. Additionally, there is lack of a detailed framework designed for one-step-ahead electricity demand forecasting that accounts for the distinct characteristics of Malaysia's electricity load trends. In line with previous studies on time series data with significant volatility, the present study will utilize asymmetric variants of the GARCH model, specifically EGARCH, APARCH, and TGARCH (also known as GJR-GARCH) in order to explore asymmetries in electricity demand volatility.

Furthermore, previous studies have not provided a thorough approach designed for one-step-ahead electricity demand forecasting. This study addresses these gaps by not only applying asymmetric GARCH models to forecast weekly maximum electricity demand but also developing a tailored framework for one-step ahead forecasting. The framework is designed to account for the distinct patterns of Malaysia's electricity demand while serving as a potential template for analyzing regions with similar electricity load patterns. Therefore, this study aims to bridge these gaps by applying asymmetric GARCH models to weekly maximum electricity demand data, identifying the best-fitting model, and using it to improve forecasting accuracy.

The study is guided by the following specific objectives:

1. To utilize SARIMA with asymmetric GARCH-type models to fit the weekly maximum electricity demand data.
2. To determine the optimal SARIMA with an asymmetric GARCH-type model that best fits the weekly maximum electricity demand data.
3. To forecast weekly maximum electricity demand using the best-performing SARIMA with asymmetric GARCH-type model.

By achieving these objectives, the study not only aims to enhance the understanding of electricity demand volatility but also introduces a robust framework for modeling and forecasting electricity demand. This framework accommodates the particular challenges of Malaysia's power load and offers a practical tool for energy management and planning in regions with similar load patterns.

2. Modelling Framework and Procedures

2.1 Model Specification and Descriptions

2.1.1 SARIMA

Analysis of electrical requirement data highlights periodic trends resulting from seasonal influences. To model and forecast these recurring seasonal effects, the Box-Jenkins SARIMA model, is widely used. A SARIMA model is developed through the inclusion of seasonal elements in the ARIMA model. It can be represented as $SARIMA(p,d,q)(P,D,Q)_s$, where *p* stands for the non-seasonal AR order, *d* is the degree of non-seasonal differencing, and *q* refers to the non-seasonal MA order. The parameters *P, D,* and *Q* represent the seasonal AR order, seasonal differencing, and seasonal MA order, while *s* indicates the length of the seasonal period. The formulation for $SARIMA(p,d,q)(P,D,Q)_s$ is expressed mathematically, as shown in Equation (1),

$$\Phi_P(B^s)\varphi_p(B)(1-B)^d(1-B^s)^D\dot{y}_t = \Theta_Q(B^s)\theta_q(B)a_t \quad (1)$$

where

$$\dot{y}_t = \begin{cases} y_t - \mu, & if\ d = D = 0 \\ y_{t,} & otherwise \end{cases}$$

and y_t is the observed time series data at time *t*, the operator of $\varphi_p(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$ and $\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in terms of B of degree *p* and *q*, while the operator of $\Phi_P(B^s) = 1 - \sum_{I=1}^{P} \Phi_I (B^s)^I$ and $\Theta_Q(B^s) = 1 - \sum_{J=1}^{Q} \Theta_J (B^s)^J$ are polynomials in terms of B^s of order *P* and *Q*, $\nabla_D^s = (1-B^s)^D$, B is the backward shift operator, and the random errors a_t are assumed to be independently and identically distributed (IID) with mean zero and constant variance σ^2.

2.1.2 EGARCH

To address certain limitations of the GARCH model in modeling financial time series, Nelson (1991) introduced the $EGARCH(r,s)$ model. Specifically, to capture the asymmetric impact of positive and negative asset returns, he incorporated a weighted innovation term. In this context, let s_t denotes the stationary time series data, a_t represents the random error at

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

time t, μ_t is the conditional mean of s_t, and σ_t^2 is the conditional variance of the error term at time t,

$$s_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{r} \alpha_i [|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}] + \sum_{j=1}^{s} \beta_j \, ln(\sigma_{t-j}^2) \quad (2)$$

where, a positive a_{t-i} contributes $\alpha_i(1+\gamma_i)\,|\varepsilon_{t-i}|$ to the log volatility, whereas a negative a_{t-i} gives $\alpha_i(1-\gamma_i)\,|\varepsilon_{t-i}|$, where $\varepsilon_{t-i} = \frac{a_{t-i}}{\sigma_{t-i}}$. The parameter γ_i thus signifies the leverage effect, controlling the asymmetric impact of positive and negative shocks on volatility, enabling the model to capture this effect, with γ_i usually negative in financial data (Tsay, 2013).

2.1.3 APARCH

The APARCH model, proposed by Ding *et al.* (1993), is expressed as follows:

$$s_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^{\delta} = \alpha_0 + \sum_{i=1}^{r} \alpha_i (|a_{t-i}| + \gamma_i a_{t-i})^{\delta} + \sum_{j=1}^{s} \beta_j \, \sigma_{t-j}^{\delta} \quad (3)$$

where s_t and a_t represent the stationary component of the series and the random error at time t, μ_t is the conditional mean of s_t, σ_t^2 is the conditional variance of the error term, and ε_t is independently and identically distributed (IID) with zero mean and unit variance. The positive real number δ, together with coefficients $\alpha_0 > 0, \alpha_i \geq 0, |\gamma_i| < 1$, and $\beta_j \geq 0$, satisfy regularity conditions to ensure positive volatility.

The parameter γ_i controls the leverage effect, where the asymmetric response of volatility to positive and negative shocks. Three special cases of the APARCH model are of interest, for $\delta = 1$ and γ_i = 0, the model reduces to the standard GARCH model, while if $\delta = 2$, it becomes the GJR-GARCH or TGARCH model. The $APARCH\ (1,1)$ model is often used in practice (Tsay, 2013).

2.1.4 TGARCH

Another volatility model commonly used to handle leverage effects is the TGARCH model (Glosten *et al.*, 1993; Zakoian, 1994), also known as GJR-GARCH model. The $TGARCH(r,s)$ model is defined as,

$$s_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} (\alpha_i + \gamma_i N_{t-i}) a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \, \sigma_{t-j}^2 \quad (4)$$

where $N_{t-i} = 1$ if $a_{t-i} < 0$, and 0 otherwise. Here, s_t is the stationary series, μ_t is the conditional mean, a_t is the innovation, and ε_t is independently and identically distributed (IID) with zero mean and unit variance. The parameters $\alpha_0 > 0, \alpha_i \geq 0, \gamma_i \geq 0$, and $\beta_j \geq 0$ ensure positive variance. The parameter γ_i captures the leverage effect, as negative shocks, $a_{t-i} < 0$ contribute $(\alpha_i + \gamma_i) a_{t-i}^2$ to volatility, while positive shocks, $a_{t-i} \geq 0$ contribute only $\alpha_i a_{t-i}^2$.

2.2 Proposed Framework for Forecasting Electricity Demand using SARIMA with Asymmetric GARCH-Type Models

Figure 1 presents the proposed research framework incorporating SARIMA with asymmetric GARCH-type models, detailing the systematic steps used for forecasting electricity demand. The framework adopts the four fundamental stages of the Box-Jenkins methodology which follows three primary steps: identifying a suitable model, estimating its parameters, conducting diagnostic evaluations, and forecasting to evaluate forecast accuracy. While the original framework proposed by Yaziz (2019) focused on non-seasonal, highly volatile time series data, this study extends its application to encompass seasonal characteristics within similarly volatile series.

2.2.1 Stage I: Identification

To determine the suitable parameters for the SARIMA and asymmetric GARCH-type models, various statistical tests can be performed. The process begins with assessing the stationarity of the time series using the Augmented Dickey-Fuller (ADF) test. If the series is found to be non-stationary, apply differencing, seasonal differencing, or transformations to achieve stationarity. Once stationarity is achieved, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the non-seasonal component are analyzed to identify the optimal orders for the autoregressive (AR), integrated (I), and moving average (MA) terms. Additionally, the presence of seasonal patterns in the data is examined to determine the appropriate seasonal cycle length. Following this, ACF and PACF plots of the seasonal component are used to estimate the seasonal AR (SAR) and seasonal MA (SMA) parameters. These steps form the foundation for developing a well-specified SARIMA model before incorporating the volatility structure through an asymmetric GARCH-type model.

The process of specifying the asymmetric GARCH-type model begins with constructing a SARIMA model for the stationary time series data to remove any serial correlation. Once the SARIMA model is fitted, the residuals are examined for the presence of conditional heteroscedasticity using the Ljung-Box Q (LBQ) test, which helps identify ARCH effects in the residual series. If such effects are detected, the next step involves analyzing the ACF and PACF plots of the squared residuals. These plots are used to determine the appropriate lag orders, denoted as r and s, for configuring the asymmetric GARCH-type model.

In general, identifying the optimal parameters for a SARIMA model combined with an asymmetric GARCH-type structure is an iterative process, often requiring the evaluation of multiple candidate models to determine the best fit.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

2.2.2 Stage II: Estimation Phase

The process of statistical estimation in a SARIMA model integrated with an asymmetric GARCH-type component requires estimating the parameters for both components separately. The first step entails examining the stationarity and seasonal characteristics of the time series, typically through tests such as the ADF test and seasonal decomposition methods like STL. Once these properties are identified, the SARIMA parameters are estimated using Maximum Likelihood Estimation (MLE), focusing on selecting the most suitable orders for the AR, MA, SAR, and SMA components. Subsequently, parameters for the GARCH structure are also estimated via MLE, which includes determining the optimal lag orders for the ARCH and GARCH terms. To finalize the model, selection criteria such as the AIC and BIC, introduced by Akaike (1974) and Schwarz (1978), respectively, are employed to identify the best-fitting SARIMA-GARCH configuration based on the lowest information criterion values.

2.2.3 Stage III: Model Validation

Diagnostic evaluation is a vital part of model development, as it assesses the adequacy of the model and helps identify any potential limitations. For SARIMA models combined with asymmetric GARCH-type models, diagnostic checking is especially important to adequately capture both the time series dynamics and the volatility aspects. A well-fitted model should produce residuals that exhibit white noise behavior (Tsay, 2013). During the diagnostic checking stage, tests for serial correlation, heteroscedasticity, and normality are conducted to ensure that the errors exhibit white noise behavior.

To detect correlation across residual lags, one approach is to assess serial dependence through residual correlation plots and confirm findings with an LBQ test. If autocorrelation is detected, model adjustments may be necessary. Heteroscedasticity, on the other hand, is assessed by subjecting the squared residuals to the LBQ test. Ideally, the model's residuals should yield a normal distribution and free from serial dependence or ARCH/GARCH effects. The Jarque-Bera (JB) test is employed to evaluate the normality of the residual distribution. However, if the residuals deviate from normality, alternative distributional assumptions, such as Student's *t* or Generalized Error Distribution (GED), should be considered. Student's t distribution is selected for its robustness in handling heavy-tailed data.

2.2.4 Stage IV: Forecasting

Forecast accuracy measures the degree to which a model's predictions match the actual results of future events. It serves as a key indicator of how well the model performs in forecasting, typically represented as a percentage or numerical value. Additionally, it is important to continually evaluate and refine forecasting models to ensure they remain accurate and relevant over time. The forecasting accuracy is evaluated using MAE, RMSE, and MAPE scores, as shown in Equations (5), (6), and (7), respectively,

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}{n}} \quad (6)$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \quad (7)$$

where y_t and $\hat{y}_t$ represent the actual and predicted values at time t, respectively, and n indicates the number of out-of-sample observations. The best forecasting model is identified as the one with the smallest prediction error. Nevertheless, when the evaluation outcomes differ depending on the metric used, MAPE is preferred, as it generally offers greater stability than other performance measures (Wang *et al.,* 2012). According to Girish

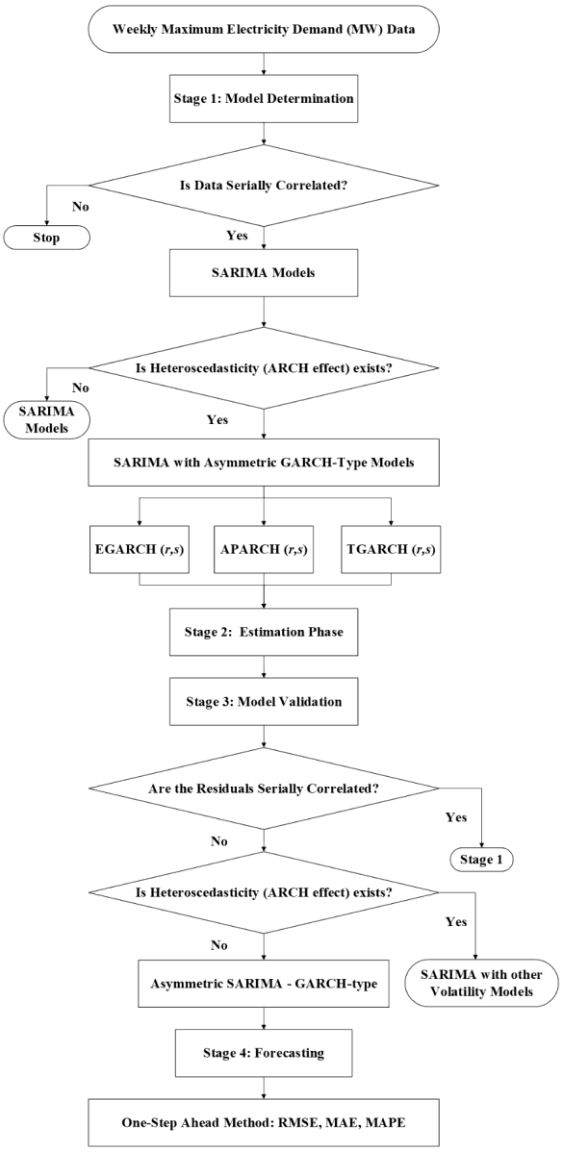


Figure 1. Proposed framework for forecasting electricity demand using SARIMA with Asymmetric GARCH-Type models.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

(2016), a forecasting model is generally regarded as fairly accurate when its MAPE value is around 5%.

3. Findings and Analysis

3.1 Data Overview

The study retrieved weekly peak electricity usage data from Single Buyer (SB) official website (https://www.singlebuyer.com.my/) covering the period from 2005 to 2016 and is classified as secondary data. SB operates under the authority of the Energy Commission and is responsible for electricity supply management and operational functions. A summary of the weekly demand figures is shown in Table 1. The dataset is divided into training and testing subsets based on the conventional 90:10 ratio, which is commonly applied for developing forecasting models (Yaziz, 2019; Chatfield, 2001). However, since the dataset follows weekly and seasonal characteristics, the dataset is divided into 92:8 ratio for the estimate to forecast, ensuring a better cycle. This results in 624 observations used for training and 52 observations for testing, corresponding to a full year of weekly data.

Although monthly or quarterly seasonality is commonly used in time series analysis, these broader intervals may fail to capture subtle fluctuations influenced by weather changes and environmental factors. Weekly forecasting offers higher temporal resolution, enabling a more precise understanding of short-term demand variations, which is valuable for effective electricity supply planning (Taylor, 2010). According to Hor *et al.* (2005), the monthly demand shows a strong cyclic pattern due to yearly temperature changes. However, the authors noted that these variations are clearer in daily demand. This suggests that weekly forecasts may be more effective than monthly or quarterly ones for managing electricity supply.

Table 1. Weekly Maximum Electricity Demand Dataset Overview.

Time Period	Total Data Points	Training Set (In-Sample)	Testing Set (Out-of-Sample)
2005 - 2016	676	Weeks 1 to 624	Weeks 625 to 676

3.2 Time Series Modeling using SARIMA and Asymmetric GARCH-Type Models

Forecasting was done using a hybrid of SARIMA and asymmetric GARCH-type models (Figure 1). Meanwhile, Figure 2 presents the pattern of in-sample weekly peak electricity load, revealing a clear upward trend over the 2005–2016 period. Despite the overall growth, short-term fluctuations are present, which likely result from seasonal factors, economic conditions, and other external influences.

The data displays both a strong seasonal structure and a consistent upward trajectory. Consequently, weekly peak electricity demand from 2005 to 2016 is used as the basis for projecting electricity usage for the subsequent year.

Outcome of the decomposition process for weekly peak electricity load data is illustrated in Figure 3. The data has been broken down into four key elements which are raw data, extracted periodic component, trend line with smoothing, and residuals. The consistency of seasonal fluctuations justifies the additive decomposition model. The seasonal component reveals a distinct and regular pattern, repeating in a predictable cycle. These components, when summed, reconstruct the original series shown in the top panel. Residuals, depicted in the bottom panel, quantify the unexplained variation post-detrending and deseasonalizing.

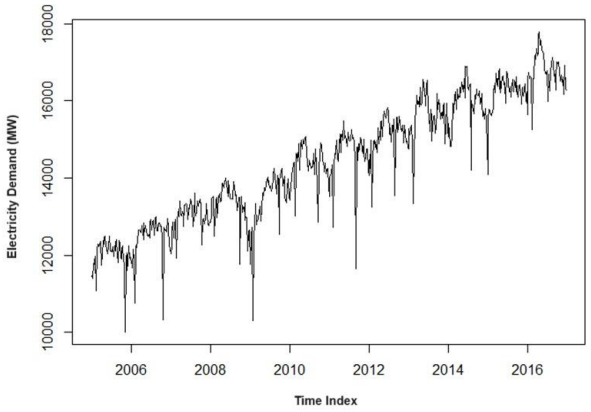


Figure 2. Weekly peak electricity load observations.

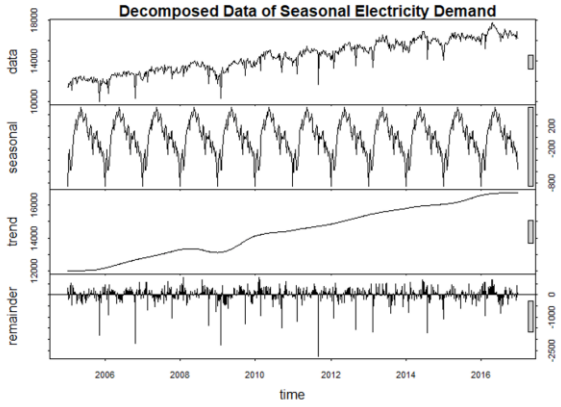


Figure 3. Time series decomposition of weekly maximum electricity load.

Table 2 presents the numerical results, which demonstrate seasonal effects that recur annually over 52 weeks. The analysis estimates the impact for each week using weekly data. Employing R statistical software, seasonality was evaluated in the time series, revealing a consistent pattern that repeats every 52 weeks.

As illustrated in Figure 4, after regular differencing, the series still exhibits clear seasonal patterns and nonstationarity, indicating that regular differencing alone is insufficient to achieve

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Table 2. Periodic Variation Effects on Weekly Electricity Demand Peaks.

Period (Week)	Estimated Seasonal Coefficient
1	-746.1064
2	-453.5700
3	-421.1485
4	-182.8890
5	-195.2180
6	-458.8856
7	-165.1864
8	-39.0005
9	-49.5013
10	103.7071
11	176.2628
12	133.8702
13	318.8347
14	83.2800
15	277.8137
16	218.9154
17	342.9762
18	334.2281
19	327.7232
20	452.6493
21	411.6273
22	386.4160
23	220.0440
24	199.5053
25	208.0522
26	330.1781
27	229.6411
28	94.9300
29	75.9146
30	-4.3645
31	40.9505
32	108.1245
33	151.9653
34	123.2089
35	-24.9987
36	-122.3736
37	42.1005
38	47.4898
39	49.4952
40	17.9409
41	24.3953
42	136.8664
43	-71.7872
44	-168.3808
45	-182.1498
46	-139.3503
47	-295.6563
48	-264.4669
49	-414.0220
50	-242.9042
51	-424.7041
52	-602.4426

stationarity. To address the remaining seasonality, a seasonal differencing operation is performed, as illustrated in Figure 5. After both regular and seasonal differencing, the series in Figure 5 appears stationary, with the trend and seasonality largely removed. This transformation makes the series suitable for further modeling, such as with the SARIMA framework.

To ensure the appropriateness of SARIMA modeling, LBQ test is conducted on the first-differenced series to check for serial correlation. With *p*-value of 2.2×10^{-16}, which is less than 0.05, the null hypothesis can be rejected with confidence, indicating the presence of serial correlation at the 5% significance level which supports the application of the Box-Jenkins methodology.

From ten SARIMA models initially considered, seven models are found to be statistically significant, as shown in Table 3. Three models, specifically, one with a single non-seasonal and seasonal MA term, another with two non-seasonal and single seasonal MA terms, and the third model with two non-seasonal MA terms. Although these models are theoretically parsimonious, they did not appear among the significant models and thus are excluded based on model diagnostics.

Additionally, heteroscedasticity checks are performed on the error terms of the significant Seasonal Box-Jenkins models through the application of LBQ test to their squared values. Only three models exhibit signs of ARCH effects, indicating the presence of volatility and making them suitable for extension with GARCH models.

Based on the diagnostic checking results, the $\mathrm{SARIMA}(1,1,0)(0,1,0)_{52}$ model is statistically significant, with no evidence of serial correlation in the residuals, but it exhibits heteroscedasticity and non-normal residuals. These characteristics indicate that this model is the most suitable candidate for integration with a GARCH-type model to better capture volatility and improve the distributional assumptions of the residuals.

To ensure the appropriateness of SARIMA modeling, LBQ test is conducted on the first-differenced series to check for serial correlation. With p-value of 2.2 × 10-16, which is less than 0.05, the null hypothesis can be rejected with confidence, indicating the presence of serial correlation at the 5% significance level which supports the application of the Box-Jenkins methodology.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

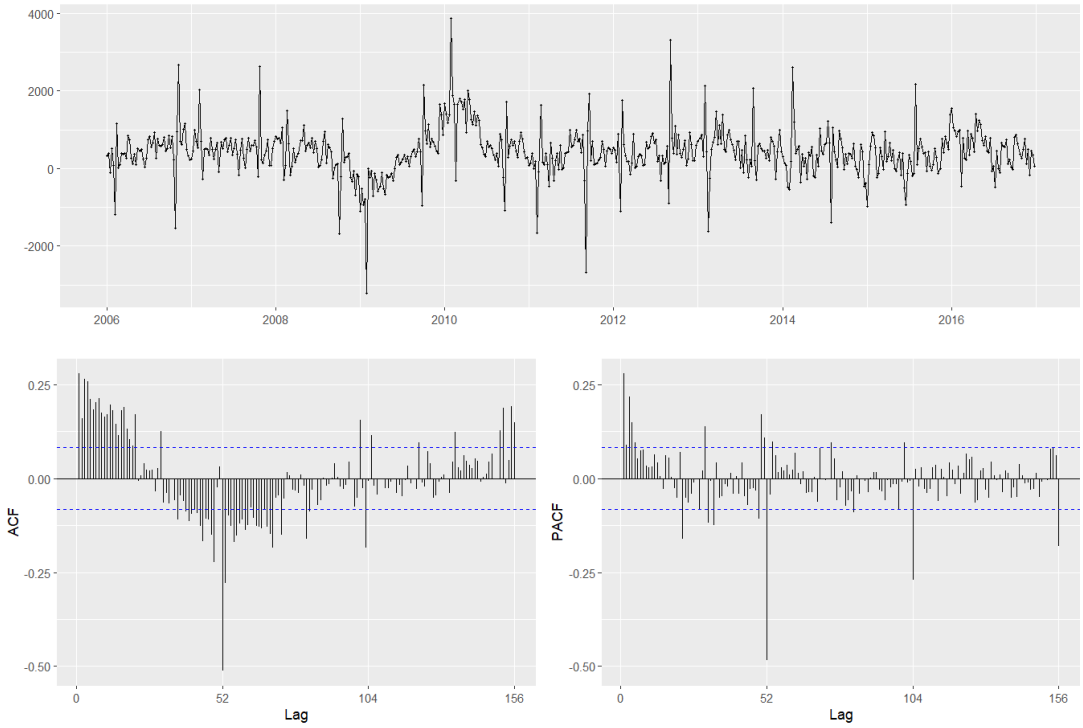


Figure 4. Regular differenced time series of weekly maximum electricity load.

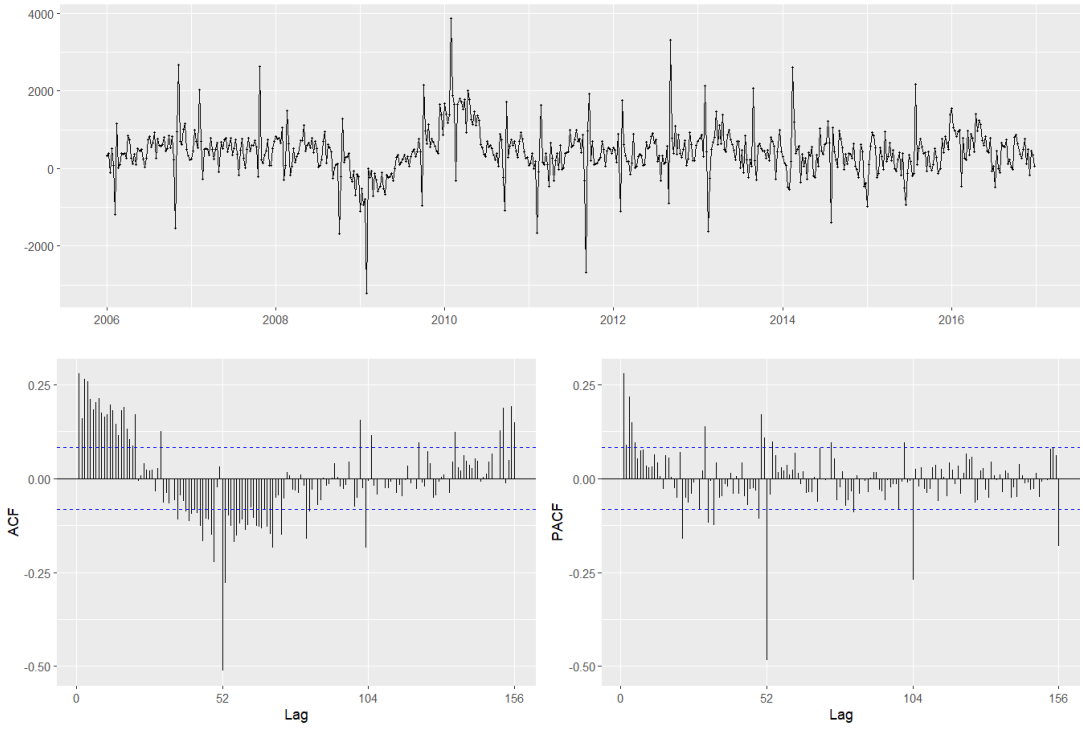


Figure 5. Seasonally differenced time series of weekly maximum electricity load.

From ten SARIMA models initially considered, seven models are found to be statistically significant, as shown in Table 3. Three models, specifically, one with a single non-seasonal and seasonal MA term, another with two non-seasonal and single seasonal MA terms, and the third model with two non-seasonal MA terms. Although these models are theoretically parsimonious, they did not appear among the significant models and thus are excluded based on model diagnostics.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Table 3. Diagnostic Checking on Selected SARIMA Models.

SARIMA Model	Significance Check	Autocorrelation Check	Heteroscedasticity Status	Normality Check
$\text{SARIMA}(1,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 1	Present	Not normal JB: 715.1444 *p*-value: 0.0000
$\text{SARIMA}(2,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 2	Present	Not normal JB: 963.0602 *p*-value: 0.0000
$\text{SARIMA}(3,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 3	Present	Not normal JB: 966.8753 *p*-value: 0.0000
$\text{SARIMA}(4,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 4	Absent (Lag 4-42, Lag 45-49)	Not normal JB: 1022.2090 *p*-value: 0.0000
$\text{SARIMA}(5,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 5	Absent (Lag 3-49)	Not normal JB: 1105.5060 *p*-value: 0.0000
$\text{SARIMA}(6,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 6	Absent (Lag 1-49)	Not normal JB: 1110.8090 *p*-value: 0.000
$\text{SARIMA}(7,1,0)(0,1,0)_{52}$	Significant	No correlation up to lag 7	Absent (Lag 1-49)	Not normal JB: 1143.2470 *p*-value: 0.0000

Additionally, heteroscedasticity checks are performed on the error terms of the significant Seasonal Box-Jenkins models through the application of LBQ test to their squared values. Only three models exhibit signs of ARCH effects, indicating the presence of volatility and making them suitable for extension with GARCH models.

Based on the diagnostic checking results, the $\text{SARIMA}(1,1,0)(0,1,0)_{52}$ model is statistically significant, with no evidence of serial correlation in the residuals, but it exhibits heteroscedasticity and non-normal residuals. These characteristics indicate that this model is the most suitable candidate for integration with a GARCH-type model to better capture volatility and improve the distributional assumptions of the residuals.

Among these, a SARIMA model with a single non-seasonal AR term, is selected not only for its parsimonious structure but also due to its superior forecasting performance. As presented in Table 4, it achieves the lowest MAPE of 2.88%, along with the lowest RMSE and MAE among the significant models. Thus, the selection is based on a combination of statistical significance, presence of heteroscedasticity for GARCH modeling, and better forecasting accuracy with fewer parameters, aligning with the principle of parsimony.

Table 5 shows the diagnostic results for the selected SARIMA-GARCH-type models. All models are statistically significant and show no serial correlation in their residuals, which means the models have captured the time series patterns well. When looking at volatility (heteroscedasticity), the EGARCH and TGARCH models perform better. They successfully remove the volatility effects in the data up to lag 51. In contrast, both APARCH models still show signs of remaining heteroscedasticity, meaning they may not handle changing variance as effectively. For normality, all models fail the test where JB values are very high with *p*-values of 0.0000. This shows that the residuals are not normally distributed, likely because of heavy tails or skewed data. Because of this, using distributions like Student's *t* or GED is important when estimating the models. In summary, the EGARCH and TGARCH models are better at handling volatility, but none of the models meet the normality assumption which is expected for real-world electricity demand data.

The proposed procedure utilizes the widely adopted AIC and BIC criteria to determine the most suitable SARIMA model combined with a GARCH-type specification. These criteria impose penalties for model complexity, favoring models that balance goodness of fit with simplicity. Accordingly, the preferred model is the one that achieves the lowest AIC or BIC score while accurately representing both the underlying time series pattern and the volatility behavior. When comparing models with differing numbers of parameters, the parsimony principle promotes choosing the less complex model if it provides comparable performance. According to McLeod (1993), the principle of parsimony suggests selecting the simplest model available.

Table 6 and Table 7 present the best performing model, which is the $\text{SARIMA}(1,1,0)(0,1,0)_{52} - \text{EGARCH}(1,2)$ with Student's *t* distribution, as it minimizes both AIC and BIC while achieving the highest log-likelihood, yet adheres to the parsimony principle. This model effectively captures heteroscedasticity and adapts well to heavy-tailed data, making it a robust choice. Close second is the TGARCH(1,2) model with Student's *t* distribution,

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Table 4. Evaluation of Forecasting Accuracy for Selected SARIMA Models.

SARIMA Model	Forecast Performance on Test Data		
	RMSE	MAE	MAPE (%)
$SARIMA(1,1,0)(0,1,0)_{52}$	606.4686	477.2099	2.88
$SARIMA(2,1,0)(0,1,0)_{52}$	627.2976	492.4028	2.97
$SARIMA(3,1,0)(0,1,0)_{52}$	622.4103	486.3534	2.94
$SARIMA(4,1,0)(0,1,0)_{52}$	623.4419	488.5009	2.95
$SARIMA(5,1,0)(0,1,0)_{52}$	624.5307	489.3780	2.95
$SARIMA(6,1,0)(0,1,0)_{52}$	632.1847	493.3851	2.98
$SARIMA(7,1,0)(0,1,0)_{52}$	637.3993	496.5803	3.00

Table 5. Diagnostic Tests for Selected SARIMA-GARCH-type Models.

SARIMA-GARCH-type Model	Significance Check	Autocorrelation Check	Heteroscedasticity Status	Normality Check
$SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,1)$	Significant	No correlation up to lag 1	Absent (Lag 1-51)	Not normal JB: 731.7946 *p*-value: 0.0000
$SARIMA(1,1,0)(0,1,0)_{52} - APARCH(1,1)$	Significant	No correlation up to lag 1	Present	Not normal JB: 940.6238 *p*-value: 0.0000
$SARIMA(1,1,0)(0,1,0)_{52} - TGARCH(1,1)$	Significant	No correlation up to lag 1	Absent (Lag 1-51)	Not normal JB: 737.9878 *p*-value: 0.0000
$SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$	Significant	No correlation up to lag 1	Absent (Lag 1-51)	Not normal JB: 835.9982 *p*-value: 0.0000
$SARIMA(1,1,0)(0,1,0)_{52} - APARCH(1,2)$	Significant	No correlation up to lag 1	Present	Not normal JB: 467.8978 *p*-value: 0.0000
$SARIMA(1,1,0)(0,1,0)_{52} - TGARCH(1,2)$	Significant	No correlation up to lag 1	Absent (Lag 1-51)	Not normal JB: 738.6357 *p*-value: 0.000

which performs nearly as well as EGARCH(1,2) but is slightly less favorable due to marginally higher values of model selection criteria (AIC and BIC). PARCH models, on the other hand, are the least effective, likely due to parameter instability and higher AIC and BIC values.

Table 8 presents the comparison of different SARIMA-GARCH-type models in capturing the volatility dynamics of the dataset. Among the models evaluated, $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$ demonstrates the best fit and forecasting performance, making it the preferred selection as it effectively addresses the heteroscedasticity detected in the SARIMA residuals. Diagnostic checks confirm that the residuals of the fitted model no longer exhibit heteroscedasticity, indicating the EGARCH model's strength in capturing the volatility pattern. It also exhibits the lowest AIC and BIC values, indicating superior model fit. Additionally, it uses Student's *t* distribution, which effectively captures the heavy-tailed nature of the data. These findings confirm that the $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$ model provides the best fit for the dataset, making it the most reliable choice for volatility forecasting.

Table 9 compares forecasting accuracy of two models: $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,1)$ and $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$. The results show that the $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$ model performs slightly better, with lower values for RMSE (1021.6340), MAE (867.6399), and MAPE (5.14%) highlights the model's utility for operational forecasting compared to $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,1)$ model. This indicates that $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$ model provides more accurate forecasts for electricity demand in Malaysia.

Figure 7 shows the $SARIMA(1,1,0)(0,1,0)_{52} - EGARCH(1,2)$ model (based on Student's *t* assumptions) one-step-ahead predictions for weekly peak electricity load for the next 52 weeks. The plot shows the forecasted values as a solid blue line, while the solid green line represents the recorded electricity usage. The values obtained from the forecast, shown with a margin of ±2 standard errors by the dotted lines, fluctuate between 17,788 MW and 10,005 MW throughout the 52-week validation period, closely following actual data and indicating the model effectively captures the underlying trend. In this study, the 52 weeks represent a full year of testing data, which is sufficient to capture a comprehensive trend, as it encompasses an entire year's worth of information.

4. Conclusion

In conclusion, the proposed SARIMA model with the asymmetric EGARCH model shows promising performance for forecasting electricity demand based on Malaysian data. Utilizing Student's *t* distribution, the $SARIMA(1,1,0)(0,1,0)_{52} -$

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Table 6. Estimation Results for SARIMA-GARCH-type Models.

SARIMA-GARCH-type Model	Parameter Estimation	Innovations of Distributions		
		Normal	**Student's *t***	**GED**
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,1)$	C	-39.3657 (0.0046)	-26.1627 (0.0442)	-23.0548 (0.0594)
	φ_1	-0.3361 (0.0000)	-0.3463 (0.0000)	-0.3105 (0.0000)
	α_0	5.5135 (0.0000)	4.8090 (0.0001)	5.2714 (0.0002)
	α_1	0.5917 (0.0000)	0.7686 (0.0000)	0.6900 (0.0000)
	γ_1	-0.3399 (0.0000)	-0.2775 (0.0045)	-0.2692 (0.0035)
	β_1	0.5369 (0.0000)	0.5872 (0.0000)	0.5470 (0.0000)
	T-DIST. DOF	-	3.2125 (0.0000)	-
	GED PARAMETER	-	-	1.0333 (0.0000)
	AIC	15.6360	**15.4284**	15.4600
	BIC	15.6818	**15.4818**	15.5134
	Log-l	-4450.262	-4390.095	-4399.095
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{APARCH}(1,1)$	C	-2088.547 (0.0000)	-21.1864 (0.0055)	-4889.562 (0.0087)
	φ_1	-1.0748 (0.0000)	-2.6982 (0.0000)	-16.4900 (0.3330)
	α_0	10898.72 (0.0000)	489238.3 (0.4069)	449543.5 (0.4935)
	α_1	-1.8605 (0.0000)	-2.6402 (0.8720)	-2.5944 (0.7999)
	γ_1	-0.1793 (0.0003)	-0.3892 (0.9541)	-0.4335 (0.9271)
	β_1	0.8217 (0.0000)	0.4670 (0.4689)	0.4253 (0.4296)
	T-DIST. DOF	-	2.0000 (0.0000)	-
	GED PARAMETER	-	-	0.9288 (0.5948)
	AIC	21.7735	17.8080	26.4637
	BIC	21.8193	17.8613	26.5170
	Log-l	-6199.460	-5068.271	-7535.141
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{TGARCH}(1,1)$	C	-36.4173 (0.0406)	-26.6810 (0.0442)	-22.7238 (0.0731)
	φ_1	-0.3087 (0.0000)	-0.3464 (0.0000)	-0.3099 (0.0000)
	α_0	199967.6 (0.0000)	183751.8 (0.0001)	172128.2 (0.0000)
	α_1	0.1197 (0.0079)	0.2181 (0.1338)	0.1741 (0.1285)
	γ_1	0.8071 (0.0000)	0.9425 (0.0157)	0.7677 (0.0100)
	β_1	0.1358 (0.0290)	0.1753 (0.1150)	0.1446 (0.2127)
	T-DIST. DOF	-	3.2270 (0.0000)	-
	GED PARAMETER	-	-	1.0384 (0.0000)
	AIC	15.6297	15.4215	15.4539
	BIC	15.6754	15.4749	15.5073
	Log-l	-4448.458	-4388.135	-4397.357

* Values in parentheses represent *p*-values, and Log-l refers to log-likelihood.

$\mathrm{EGARCH}(1,2)$ model demonstrates superior effectiveness in forecasting electricity demand in the Malaysian context with its low MAPE value of 5.14%.

However, this study is limited by the lack of updated electricity demand data, utilizing data from 2017 to 2023. A more robust demonstration of SARIMA with asymmetric GARCH-type model for forecasting electricity demand in Malaysia would require more recent data updates. Additionally, the study employs weekly maximum electricity demand, which may not capture more granular variations, such as daily or intraday fluctuations, that are critical for real-time energy management. Future research in this area should focus on applying multistep forecasting while incorporating updated data.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

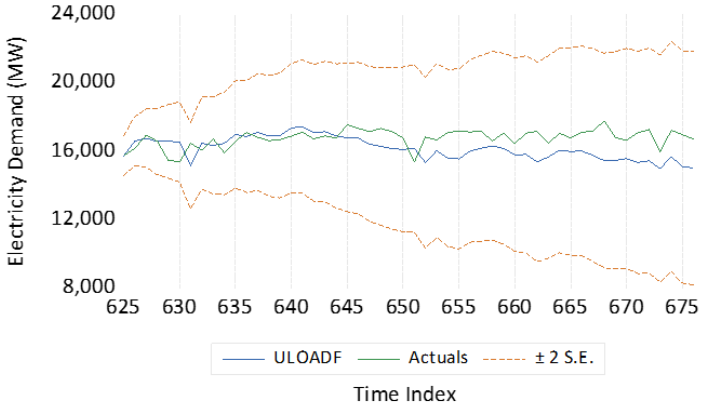


Figure 6. Weekly electricity demand forecast using $\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,1)$.

Table 7. Estimation Results for SARIMA-GARCH-type Models.

SARIMA-GARCH-type Model	Parameter Estimation	Innovations of Distributions		
		Normal	**Student's *t***	**GED**
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,2)$	C	-37.1886 (0.0094)	-25.5287 (0.0371)	-24.7191 (0.0392)
	φ_1	-0.3608 (0.0000)	-0.3913 (0.0000)	-0.3504 (0.0000)
	α_0	6.2604 (0.0000)	5.2072 (0.0000)	5.6656 (0.0000)
	α_1	0.4786 (0.0000)	0.6321 (0.0000)	0.5546 (0.0000)
	γ_1	-0.2652 (0.0000)	-0.2129 (0.0057)	-0.2216 (0.0031)
	β_1	0.9667 (0.0000)	1.0187 (0.0000)	0.9750 (0.0000)
	β_2	-0.4831 (0.0000)	-0.4571 (0.0000)	-0.4523 (0.0005)
	T-DIST. DOF	-	3.2852 (0.0000)	-
	GED PARAMETER	-	-	1.0595 (0.0000)
	AIC	15.6223	**15.4143**	15.4528
	BIC	15.6757	**15.4753**	15.5138
	Log-l	-4445.349	-4385.068	-4396.036
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{APARCH}(1,2)$	C	-28.5307 (0.2417)	-16.6803 (0.0333)	-4704.133 (0.0080)
	φ_1	-0.3850 (0.0000)	-2.0846 (0.0000)	-18.0671 (0.3474)
	α_0	1231.358 (0.0000)	491143.4 (0.8507)	489108.2 (0.5315)
	α_1	-0.0059 (0.5705)	-2.3778 (0.9069)	-2.3331 (0.8302)
	γ_1	-1.0000 (0.5941)	-0.5770 (0.9483)	-0.5944 (0.9234)
	β_1	0.1693 (0.0000)	0.4128 (0.9307)	0.4461 (0.3411)
	β_2	-0.9883 (0.0000)	-0.1068 (0.9500)	-0.0748 (0.8931)
	T-DIST. DOF	-	2.0000 (0.0000)	-
	GED PARAMETER	-	-	0.9304 (0.5927)
	AIC	15.8846	17.2301	26.5867
	BIC	15.9380	17.2911	26.6477
	Log-l	-4520.114	-4902.580	-7569.215

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Table 7. Continued.

SARIMA-GARCH-type Model	Parameter Estimation	Innovations of Distributions		
		Normal	Student's *t*	GED
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{TGARCH}(1,2)$	C	-37.8297 (0.0385)	-26.4966 (0.0446)	-23.3647 (0.0663)
	φ_1	-0.2955 (0.0000)	-0.3524 (0.0000)	-0.3099 (0.0000)
	α_0	222391.5 (0.0000)	187752.7 (0.0001)	184096.4 (0.0000)
	α_1	0.1523 (0.0015)	0.2400 (0.1184)	0.1847 (0.1092)
	γ_1	0.7531 (0.0000)	0.8227 (0.0204)	0.7334 (0.0095)
	β_1	0.1534 (0.0250)	0.2613 (0.0781)	0.1863 (0.1728)
	β_2	-0.0795 (0.0081)	-0.0788 (0.1668)	-0.0675 (0.1579)
	T-DIST. DOF	-	3.2342 (0.0000)	-
	GED PARAMETER	-	-	1.0453 (0.0000)
	AIC	15.6222	15.4196	15.4526
	BIC	15.6756	15.4805	15.5136
	Log-l	-4445.332	-4386.571	-4395.999

* Values in parentheses represent *p*-values, and Log-l refers to log-likelihood.

Table 8. Selected SARIMA with GARCH-type Models.

SARIMA-GARCH-type Model	Distribution Type	AIC	BIC
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,1)$	Student's *t*	15.4284	15.4818
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{APARCH}(1,1)$	Student's *t*	17.8080	17.8613
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{TGARCH}(1,1)$	Student's *t*	15.4215	15.4749
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,2)$	Student's *t*	**15.4143**	**15.4753**
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{APARCH}(1,2)$	Normal	15.8846	15.9380
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{TGARCH}(1,2)$	Student's *t*	15.4196	15.4805

Table 9. Evaluation of Forecasting Performance of SARIMA-EGARCH Models.

SARIMA-EGARCH Model	Forecast Performance on Test Data		
	RMSE	MAE	MAPE (%)
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,1)$	1041.1700	882.5311	5.23
$\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,2)$	**1021.6340**	**867.6399**	**5.14**

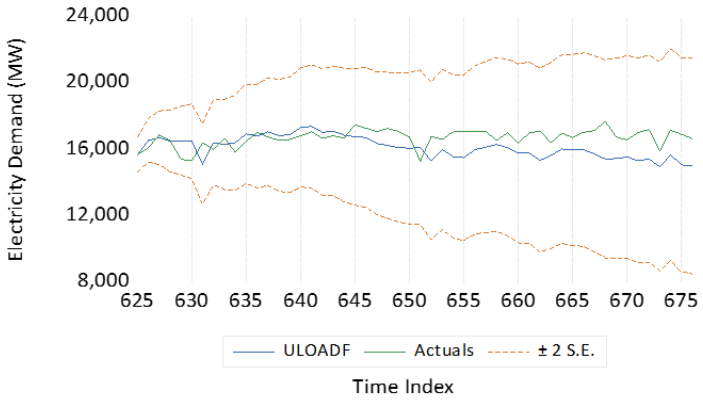


Figure 7. Weekly electricity demand forecast using $\mathrm{SARIMA}(1,1,0)(0,1,0)_{52} - \mathrm{EGARCH}(1,2)$.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

5. Acknowledgement

The authors gratefully acknowledge the funding provided by UMPSA via research grant entitled Forecasting Electricity Demand for Domestic Usage in Malaysia using Seasonal Box-Jenkins Model with GARCH-Type (Research Grant RDU240306).

6. References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.

Al-Alawi, S. M., & Islam, S. M. (1996). Principles of electricity demand forecasting. I. Methodologies. Power Engineering Journal, 10(3), 139-143.

Andoh, P. Y. A., Sekyere, C. K. K., Mensah, L. D., & Dzebre, D. E. K. (2021). Forecasting electricity demand in Ghana with the SARIMA model. Journal of Applied Engineering and Technological Science, 3(1), 1-9.

Chalissery, N., Anagreh, S., Nishad T, M., & Tabash, M. I. (2022). Mapping the trend, application and forecasting performance of asymmetric GARCH models: a review based on bibliometric analysis. Journal of Risk and Financial Management, 15(9), 406.

Chatfield, C. (2001). Time-series forecasting. Retrieved from http://www.crcnetbase.com/isbn/9781584880639.

Chen, H., Zhang, J., Tao, Y., & Tan, F. (2019). Asymmetric GARCH type models for asymmetric volatility characteristics analysis and wind power forecasting. Protection and Control of Modern Power Systems, 4(4), 1-11.

Department of Statistics Malaysia. (2018). Malaysia's economic performance report (DOSM/BPPIB/1.2018/Series 10 (Electricity Sector in Malaysia)). Retrieved from https://www.dosm.gov.my/v1/uploads/files/6_Newsletter/newsletter%202018/Series%2010_Electricity%20Sector.pdf.

Duong, T. K., Phan, D. H., & Nguyen, D. M. (2023). Application of Sarima Model in Load Forecasting in Hanoi City. International Journal of Energy Economics and Policy, 13(3), 164-170.

Girish, G. P. (2016). Spot electricity price forecasting in Indian electricity market using autoregressive - GARCH models. Energy Strategy Reviews, 11(12), 52-57.

Goswami, K., & Kandali, A. B. (2020). Electricity demand prediction using data driven forecasting scheme: ARIMA and SARIMA for real-time load data of Assam. In 2020 International Conference on Computational Performance Evaluation (ComPE), IEEE, 570-574.

Hor, C. L., Watson, S. J., & Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. IEEE transactions on power systems, 20(4), 2078-2085.

Ishak, I., Othman, N. S., & Harun, N. H. (2022). Forecasting electricity consumption of Malaysia's residential sector: Evidence from an exponential smoothing model. F1000Research, 11(54), 54.

Kavaklioglu, K., Ceylan, H., Ozturk, H. K., & Canyurt, O. E. (2009). Modeling and prediction of Turkey's electricity consumption using artificial neural networks. Energy Conversion and Management, 50(11), 2719-2727.

Keyno, H. S., Ghaderi, F., Azade, A., & Razmi, J. (2009). Forecasting electricity consumption by clustering data in order to decline the periodic variable's affects and simplification the pattern. Energy Conversion and Management, 50(3), 829-836.

Lee, M. H. L., Ser, Y. C., Selvachandran, G., Thong, P. H., Cuong, L., Son, L. H., ... & Gerogiannis, V. C. (2022). A comparative study of forecasting electricity consumption using machine learning models. Mathematics, 10(8), 1329.

Lim, C. M., & Sek, S. K. (2013). Comparing the performances of GARCH-type models in capturing the stock market volatility in Malaysia. Procedia Economics and Finance, 5, 478-487.

Lisi, F., & Shah, I. (2020). Forecasting next-day electricity demand and prices based on functional models. Energy Systems, 11(4), 947-979.

Makobea, O. S. (2023). Analysing electricity consumption in South Africa using volatility forecasting models (Doctoral dissertation, North-West University (South Africa)).

McLeod, A. I. (1993). Parsimony, model adequacy and periodic correlation in time series forecasting. International Statistical Review/Revue Internationale de Statistique, 387-393.

Ndege, E. M., Muriithi, D. K., & Wagala, A. (2023). Application of Asymmetric-GARCH type models to the Kenyan exchange rates. European Journal of Mathematics and Statistics, 4(4), 78-92.

Qiu, Z. (2013). Electricity consumption prediction based on data mining techniques with particle swarm optimization. International Journal of Database Theory and Application, 6(5), 153-164.

Rakshit, D., Paul, R. K., Yeasin, M., Emam, W., Tashkandy, Y., & Chesneau, C. (2023). Modeling asymmetric volatility: A news impact curve approach. Mathematics, 11(13), 2793.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461-464.

Shah, I., Iftikhar, H., & Ali, S. (2022). Modeling and forecasting electricity demand and prices: A comparison of alternative approaches. Journal of Mathematics, 2022.

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. European Journal of Operational Research, 204(1), 139-152.

Tehrani, S., Juan, J., & Caro, E. (2022). Electricity spot price modeling and forecasting in European markets. Energies, 15(16), 5980.

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

Tsay, R. S. (2013). An Introduction to analysis of financial data with R. Hoboken, N.J: Wiley.

Wang, B., Huang, H., & Wang, X. (2012). A novel text mining approach to financial time series forecasting. Neurocomputing, 83, 136-145.

Yasmeen, F., & Sharif, M. (2014). Forecasting electricity consumption for Pakistan. International Journal of Emerging Technology and Advanced Engineering, 4(4), 496-503.

Yaziz, S. R. (2019). Modified Box-Jenkins and GARCH for Forecasting Highly Volatile Time Series Data, PhD Thesis, Universiti Malaysia Pahang.

Zaim, S, Yusoff, W. N. S. W., Mohamad, N. N., Radi, N. F. A., and Yaziz, S. R. (2023). Forecasting of electricity demand with seasonal highly volatile characteristics using Box-Jenkins - GARCH model", MATEMATIKA, 39(3), 293-313.

Zaim, S., Yaziz, S. R., Zakaria, R., Mohamad, N. N., & Radi, N. F. A. (2025). Electricity demand forecasting in malaysia using seasonal Box-Jenkins model. Jurnal Teknologi (Sciences & Engineering), 87(2).

DOI: https://doi.org/10.22452/mjs.vol45sp2.1
Malaysian Journal of Science 45 (Special Issue 2): 1-15 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026) https://mjs.um.edu.my

Using Time-Series Anomaly Detection Approach for Identifying Arbitrage Opportunities

Jose Abad[a], Darren Pabiloña[a], Stephen Ty[a], and Jeric Briones[a,b*]

Abstract: Since market anomalies are tedious and difficult to manually identify, the use of anomaly detection techniques to find statistical arbitrage opportunities is becoming increasingly popular. As such, this study compares the performance of different pairs trading strategies on the top 100 Philippine stocks listed on the Philippine Stock Exchange to assess the feasibility of using machine learning-based anomaly detection models. Specifically, time series anomaly detection through generative adversarial networks and auto-encoder with regression were employed to detect statistical arbitrage opportunities. The comparison across the strategies was limited to the top five cointegrated pairs (of order one) with the smallest sum of squared residuals. Results show that, even after accounting for transaction costs, AER outperformed the benchmarks (distance and cointegration approaches) by a wide margin, while TadGAN achieved a slightly better performance than the benchmarks. These findings suggest that anomaly detection models have the potential to yield higher profits than conventional approaches.

Keywords: anomaly detection, machine learning, pairs trading, Philippine stocks, statistical arbitrage

1. Introduction

As most transactions in financial markets carry some level of risk, strategies have been formulated to minimize said risk. One commonly used strategy involves taking advantage of arbitrage. Arbitrage refers to the practice of exploiting price differences in markets to buy and sell assets for a profit. For example, one type of arbitrage locks in a *riskless* profit by (simultaneously) entering into transactions in two or more markets. To illustrate, consider a stock that is listed in both the New York Stock Exchange (NYSE) and the London Stock Exchange (LSE). Suppose the stock price is $12 in NYSE and £10 in LSE, and that the prevailing exchange rate is $1.25 per £1. To take advantage of this arbitrage opportunity, a trader can buy shares (for example, 1,000 shares, costing $12,000) in NYSE, sell the shares in LSE (earning £10,000), and convert the proceeds of the sale from British pounds to U.S. dollars (equivalent to $12,500). This transaction would net $500 in profit for the trader.

Another type of arbitrage is statistical arbitrage, which uses price differences indicated by statistical testing. Due to the nature of this strategy, it carries minimal risk. An example of this arbitrage strategy is *pairs trading*. Pairs trading is a form of statistical arbitrage that hedges risk by trading securities that show a strong historical correlation of moving together. It relies on the assumption of mean reversion between the prices of such pairs (Carrasco Blázquez et al., 2018). For example, pairs trading in the stock market involves three steps: identifying a pair of historically highly correlated stocks, monitoring the price spread between the stocks, and trading according to the magnitude of their price spread. In literature, two established methods for performing these steps are the *distance* and the *cointegration* approaches. The distance approach conducts pairs trading by choosing pairs with the smallest sum of squared deviations between their cumulative return indices, which indicates that these prices "move together" (Gatev et al., 2006). The cointegration approach trades pairs that exhibit cointegration, which is the property that a linear combination of two $I(1)$ stock prices produces a stationary time series (Krauss, 2017). Here, $I(1)$ refers to a time series process whose order of integration is one.

While arbitrage strategies can be used to minimize or mitigate risks in transactions, their effectiveness relies on the ability of traders to detect opportunities to employ them. These opportunities are often present whenever market anomalies exist. Market anomalies refer to patterns in a market, such as the stock market, that contradict the efficient market hypothesis (Canady, 2022). Although they are difficult to identify manually, recent studies have shown that machine learning and artificial intelligence can be potentially used to detect them. Machine learning methods have exhibited viability in solving complex problems in other disciplines due to their ability to detect patterns even in complex data. This is especially evident with issues that might involve features that are tedious or difficult to manually solve or identify. For instance, in network intrusion detection systems, researchers have found that machine learning models and deep neural networks can significantly improve detection rates and reduce false positives compared to conventional approaches (Ullah et al., 2024). Of interest to this work are machine learning models used for anomaly detection. For example, deep learning models, such as the Transformer model, have been considered for time series anomaly detection

Authors information:
[a]Department of Mathematics, Ateneo de Manila University, Quezon City, PHILIPPINES
[b]Department of Finance and Accounting, Ateneo de Manila University, Quezon City, PHILIPPINES

*Corresponding Author: jbriones@ateneo.edu

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

(Nakashima & Yairi, 2023). However, some of these models, including the Transformer model, have not performed as expected. TadGAN, built on generative adversarial networks (GANs), and auto-encoder with regression (AER), however, have exhibited improvements in anomaly detection for time series data (Geiger et al., 2020; Wong et al., 2022). As such, the focus of this work is on anomaly detection models suited to time series data.

Aside from the effectiveness of the methods used, market data are also assumed to carry information that these methods can make use of. It is then worth noting that the Philippine stock market, an emerging market, has been shown to be not weak-form informationally efficient (Aquino, 2002). That is, historical prices carry some information that could be used to detect the above-mentioned market anomalies. It then becomes natural to explore the applicability of machine learning methods for statistical arbitrage detection. While previous studies have shown their effectiveness in detecting anomalies in developed markets (Tallboys et al., 2022), their applicability in emerging markets is yet to be explored. This is where this work comes in: It uses data from the Philippine stock market to determine whether time series anomaly detection approaches are applicable. In doing so, the study could provide investors with tools they can use as they participate in emerging markets.

1.1 Objectives of the Study

Exploring the efficacy of machine learning–based anomaly detection models as statistical arbitrage detection models in the Philippine Stock Exchange (PSE) is the primary motivation for this study. That is, the study aims to determine if anomaly detection can identify statistical arbitrage opportunities. Specifically, the objectives of the study are as follows:

- to detect statistical arbitrage opportunities for selected stock pairs using TadGAN and AER; and
- to compare the performance of the time series anomaly detection approach with benchmark pairs trading strategies (namely distance and cointegration approaches).

1.2 Scope and Limitations

In assessing the applicability of the two aforementioned models, several assumptions for the market are considered. First, this study assumes that fractional shares *can* be traded and that stocks are fully liquid and tradeable daily. Moreover, only the effects of transaction costs are considered. Other costs such as from slippage, interest, and dividends are not considered in the evaluation of the models' performance. Furthermore, only statistical arbitrage is considered, and arbitrage opportunities are detected only for identified pairs of stocks that meet specified conditions. Finally, different market conditions are not noted, as this study focuses on the applicability and efficacy of the models.

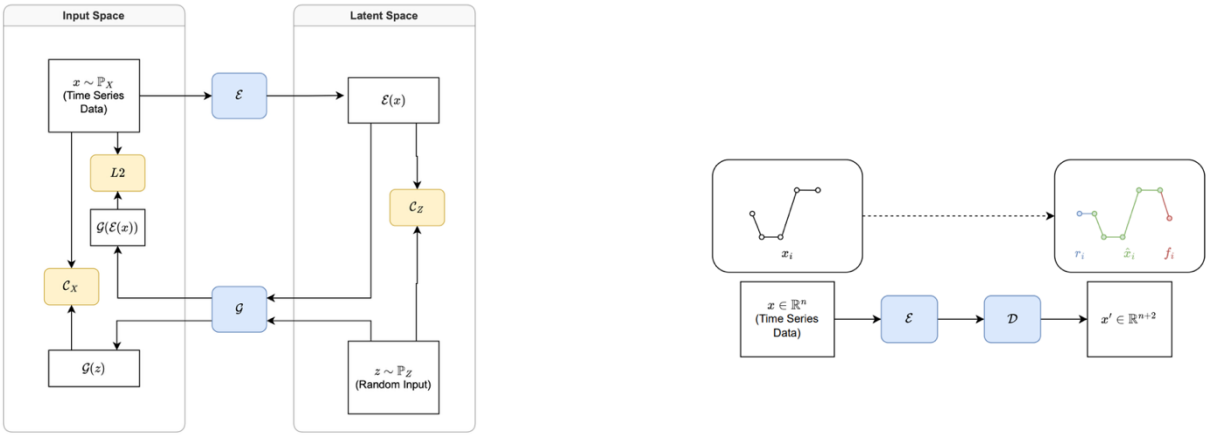
Aside from market assumptions, a few model assumptions are also made. First, the study does not do hyperparameter tuning and optimization. Instead, the models use values based on established practices found in existing related literature. Moreover, in the comparison, the study focusses on simulating and applying the identified pairs trading strategies. Finally, other machine learning models are not considered, as the study is focussed on time series anomaly detection models.

2. Pairs Trading and Anomaly Detection

As mentioned, a commonly used strategy to take advantage of statistical arbitrage involves pairs trading. This strategy relies on monitoring the spread of highly correlated pairs of stocks to generate statistical returns based on a set of systematic rules. In general, pairs trading involves three phases: formation of the pairs, construction of the residual series, and formulating a trading rule. When forming the stock pairs, traders often use statistical properties of stocks as a basis for pairing them. For example, the distance approach utilizes the Euclidean distance of historical prices to identify highly correlated pairs. This distance is monitored, and, when the spread deviates from the historical mean by one standard deviation (Fernández-Pérez et al., 2020), a position is opened. Under the assumption of mean reversion, the spread is expected to return to the historical mean, allowing the position to close at a profit. Although this approach works, using the sum of squared distances is not optimal. Instead, cointegrated pairs of time series are preferred, using a similar method of thresholding to open and close positions. This represents the cointegration approach. Instead of using highly correlated pairs based on their Euclidean distance, the approach uses cointegrated pairs of order 1. These pairs are $I(1)$ processes that have a stationary linear combination, and, by the error correction model, the combination has mean-reverting properties (Rad et al., 2016). However, the problem lies on how long the relationship will continue. Uncertainties on when to close the position may arise when the series does not seem to revert to its historical mean.

Aside from using the spread as a trigger to open positions, positions can also be opened when market anomalies are detected. This is because arbitrage opportunities may be associated with market anomalies. These anomalies are defined as breaks in patterns or trends that are assumed to have been caused by external factors outside of normal market conditions. A common form of anomaly detection is through prediction based on historical values. From this, a confidence interval is found, and when the actual observed value is outside of it, it is labelled an anomaly. An alternative method is through reconstruction. Latent features from the pair's correlation are captured to reconstruct the data instead. When the observed value varies greatly from the reconstructed value, it is labelled an anomaly. New, sophisticated forms of time series reconstruction have been used by incorporating different forms of machine learning methods to improve reconstruction. Latent features from the correlated pairs may be better represented through deep learning models than other models (such as stochastic models). Examples of deep learning models used for anomaly detection are TadGAN (which uses GANs to form time series reconstructions) and AER (which combines anomaly detection from predictive and reconstructive approaches to form a more robust anomaly detection model). These two models are illustrated in Figure 1.

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)



(a) TadGAN (b) AER

Figure 1. Illustration of the architectures of the time series anomaly detection models considered in this paper.

2.1 Time Series Anomaly Detection Using Generative Adversarial Networks (TadGAN)

TadGAN (Geiger et al., 2020) makes use of GANs to reconstruct time series data. It consists of two parts. First is the encoding of the data into a latent space of lower dimension (represented by the mapping function $\mathcal{E}: X \to Z$). This is followed by a generator that decodes information from the latent space reconstructing the time series data (represented by the mapping function $\mathcal{G}: Z \to X$). For each part, a critic is present to act as an adversary to the encoder and generator. In order to avoid the mode collapse problem common to GANs, TadGAN employs Wasserstein loss to train the critics. Furthermore, a cycle consistency loss is used to ensure better reconstruction through the process of encoding and decoding (i.e., $x_i \to \mathcal{E}(x_i) \to \mathcal{G}(\mathcal{E}(x_i))$, where $\mathcal{G}(\mathcal{E}(x_i)) \approx \hat{x}_i$). This gives the following objective function:

$$\min_{\{\mathcal{E},\mathcal{G}\}} \max_{\{\mathcal{C}_x \in \mathbf{C_x}, \mathcal{C}_z \in \mathbf{C_z}\}} V_X(\mathcal{C}_x, \mathcal{G}) + V_Z(\mathcal{C}_z, \mathcal{E}) + V_{L2}(\mathcal{E}, \mathcal{G}). \quad (1)$$

The first two terms are the Wasserstein loss for the critic models given by

$$V_X(\mathcal{C}_x, \mathcal{G}) = \mathbb{E}_{x\sim\mathbb{P}_X}[\mathcal{C}_x(x)] - \mathbb{E}_{z\sim\mathbb{P}_Z}[\mathcal{C}_x(\mathcal{G}(z))], \quad (2)$$
$$V_Z(\mathcal{C}_z, \mathcal{E}) = \mathbb{E}_{z\sim\mathbb{P}_Z}[\mathcal{C}_z(z)] - \mathbb{E}_{x\sim\mathbb{P}_X}[\mathcal{C}_z(\mathcal{E}(x))], \quad (3)$$

where $\mathcal{C}_x(x)$ is the critic score of the original time series x tested against itself, $\mathcal{C}_x(\mathcal{G}(z))$ is the critic score of the series $\mathcal{G}(z)$ generated from random noise tested against the original time series x, $\mathcal{C}_z(z)$ is the critic score of a random latent space sample z tested against itself, and $\mathcal{C}_z(\mathcal{E}(x))$is the critic score of an encoded time series $\mathcal{E}(x)$ tested against a random latent space sample z. Furthermore, note that $\mathbb{P}_X$ and $\mathbb{P}_Z$ are the probability distributions over X and Z, respectively. Thus, $x \sim \mathbb{P}_X$ and $z \sim \mathbb{P}_Z$ indicate that time series x and z are $\mathbb{P}_X$- and $\mathbb{P}_Z$-distributed, respectively. As such, $\mathbb{E}_{x\sim\mathbb{P}_X}[\cdot]$ and $\mathbb{E}_{z\sim\mathbb{P}_Z}[\cdot]$ are the Lebesgue integrals over probability space $\mathbb{P}_X$ and $\mathbb{P}_Z$, respectively. However, these cannot be calculated explicitly and are instead estimated using the arithmetic mean. Finally, the cycle consistency lost based on the L2-norm is given by

$$V_{L2}(\mathcal{E}, \mathcal{G}) = \mathbb{E}_{x\sim\mathbb{P}_X}[x - \mathcal{G}(\mathcal{E}(x))], \quad (4)$$

where x refers to the original time series and $\mathcal{G}(\mathcal{E}(x))$ refers to the reconstructed time series. For the model architecture, the encoder and generator are both built on long short-term memory (LSTM) networks. These models are known for being able to generate outputs based on previously stored information. In contrast, the two critics utilize dense layers with a leaky rectified linear unit (Leaky ReLU) activation function to distinguish between real and generated data.

2.2 Time Series Anomaly Detection Using Auto-Encoder with Regression (AER)

AER, developed by Wong et al. (2022), is another anomaly detection model; it uses both prediction-based and reconstruction-based anomaly detection. These two approaches were considered as they have their respective strengths and weaknesses. It uses an LSTM auto-encoder to reconstruct the time series and an LSTM regressor to predict the subsequent values of the time series. To address the drawbacks of prediction-based anomaly detection, AER's prediction model is bidirectional; it also predicts the previous values of the time series. Consideration of these three parts of the model yields the following objective function:

$$\text{Loss} = \frac{\gamma}{2} V_{\text{pred}}(t_{i-1}, r_{i-1}) + \frac{\gamma}{2} V_{\text{pred}}(t_{i+n}, f_{i+n}) + (1 - \gamma)\, V_{\text{pred}}(t_{i:i+n-1}, y_{i:i+n-1}), \quad (5)$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

where n is the number of observations, γ is the weight of the regression losses on the objective function, $V_{\text{pred}}(a, b)$ is the prediction loss given by the mean squared error between the pairs of values a and b, $V_{\text{rec}}(a, b)$ is the reconstruction error given by the mean squared error between the time series sequences a and b, and t_i, r_i, f_i, and y_i are the actual, reverse predicted, forward predicted, and reconstructed values, respectively. Both the encoder and the decoder utilize bidirectional LSTMs to capture information in both forward and backward directions. The encoder maps the data to a latent space, while the decoder reconstructs the series with two additional data points: one at the beginning for one-step reverse prediction, and one at the end for one-step forward prediction. The rest corresponds to the reconstructed data.

3. Methods and Implementation

To determine the efficacy of TadGAN and AER, post-pandemic daily closing prices from 2021 to 2023 in the Philippine stock market were used. To be included in the dataset, the stock must be among the 100 stocks with the largest market capitalization (valued as of December 2023). They must also be listed in the PSE from 2021 to 2023. In forming the pairs, prices from 2021 to 2022 are taken into account, with the top five pairs being chosen based on the combined criteria of the cointegration and distance methods. The trading period used for simulation entails prices from 2023. Lastly, the pipelines are retrained every 20 trading days. Results from the two anomaly detection models (TadGAN, AER) are compared with benchmark methods (distance, cointegration). A summary of the pairs trading pipeline can be seen in Figure 2. Details of the network architecture for TadGAN and AER used in this paper can be seen in Figure 3.

3.1 Formation of Pairs and Residual Series

The five pairs are chosen using a combination of the distance and cointegration method criteria (Rad et al., 2016). The combined selection criteria have two parts. First, each pair is tested for cointegration of order 1. In this step, the augmented Dickey-Fuller test (Ajewole et al., 2020) is first used to identify which of the stock price time series are $I(1)$ processes. Those that are not $I(1)$ processes are removed from consideration. Afterwards, each pair of $I(1)$ stock prices is tested for cointegration of order 1 using the two-step Engel-Granger method (Moura, 2010). The second part involves computing the sum of squared distances (SSD) for each cointegrated pair found. In this step, the SSD of the stock pair *X* and *Y* is given by $\text{SSD} = \sum_t \left[S_t^{\text{d}}\right]^2$, where

$$S_t^{\text{d}} = \frac{X_t}{X_0} - \frac{Y_t}{Y_0}, \tag{6}$$

and X_t and Y_t are the historical stock prices. The five pairs with the smallest SSD are then chosen. After the pairs are formed, the residual series is constructed. Series S_t^{d} is the residual series used in the distance method; the residual series used in the cointegration method is the stationary linear combination of the cointegrated pair X and Y given by

$$S_t^{\text{c}} = X_t - \gamma Y_t, \tag{7}$$

where γ is the cointegration coefficient obtained from the two-step Engel-Granger method.

3.2 Formation of Trading Rules

The trading rules for the pairs trading strategies considered are as follows. For the distance and cointegration methods, positions are opened when the residual series is far from the "mean" (Fernández-Pérez, 2020). The position is closed upon mean reversion. To be specific, the trading rules for opening a position are the following:

- if the residual series is more than one historical standard deviation *above* the mean, sell X short and buy Y; and
- if the residual series is more than one historical standard deviation *below* the mean, buy X and sell Y short.

While the two methods have similar trading rules, the number of shares involved and the mean used differ. For the distance method, the number of shares involved is one currency unit worth of X and Y, with the mean set to 0. For the cointegration method, however, one share of X and γ shares of Y are used, with the mean set to the historical mean.

For TadGAN and AER, positions are opened when statistical arbitrage opportunities are found. The existence of these opportunities is determined by detecting anomalies in the cointegration residual series. Upon the end of the anomalous sequence, the positions are closed. For these methods, the trading rules are inherited from the cointegration method when the cointegration residuals are used. That is,

- if S_t^{c} is an anomalous point and is *above* the reconstructed series, one share of X will be sold short while γ shares of Y will be bought; and
- if S_t^{c} is an anomalous point and is *below* the reconstructed series, γ shares of Y will be sold short while one share of X will be bought.

3.3 Performance Evaluation

The mean daily per cent profit and loss (%pnl) is used to evaluate the profitability of the aforementioned methods. The formula employed is a generalized version of the one used by Gatev et al. (2006), modified to accommodate the trading rules of the cointegration and machine learning approaches. To be specific, %pnl at time t, denoted by r_t, is

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

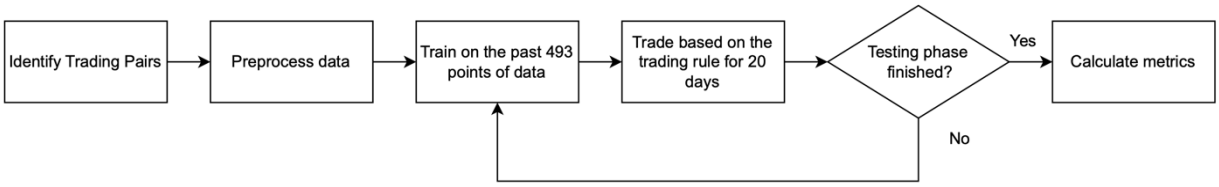


Figure 2. Experiment design flowchart. The pairs trading pipeline includes forming the stock pairs, training the anomaly detection models, and evaluating their profitability.

TadGAN Layer	Output Shape	Activation Function Update	Recurrent
Encoder			
Input	10×1		
LSTM	10×100	Tanh	Sigmoid
Flatten	1000		
Dense	5	Linear	
Critic Z			
Flatten	5		
Dense	100	Linear	
Leaky ReLU	100	Leaky ReLU ($\alpha = 0.2$)	
Dropout	100		
Dense	100	Linear	
Leaky ReLU	100	Leaky ReLU ($\alpha = 0.2$)	
Dropout	100		
Dense	1	Linear	
Generator			
Flatten	5		
Dense	5	Linear	
Reshape	5		
Bidirectional LSTM	5×128	Tanh	Sigmoid
UpSampling1D	10×128		
Bidirectional LSTM	10×128	Tanh	Sigmoid
TimeDistributed Dense	10×1	Linear	
Activation	10×1	Tanh	
Critic X			
Conv1D	8×64		
Leaky ReLU	8×64	Leaky ReLU ($\alpha = 0.2$)	
Dropout	8×64		
Conv1D	6×64		
Leaky ReLU	6×64	Leaky ReLU ($\alpha = 0.2$)	
Dropout	6×64		
Conv1D	4×64		
Leaky ReLU	4×64	Leaky ReLU ($\alpha = 0.2$)	
Dropout	4×64		
Conv1D	2×64		
Leaky ReLU	2×64	Leaky ReLU ($\alpha = 0.2$)	
Dropout	2×64		
Flatten	128		
Dense	1	Linear	

AER Layer	Output Shape	Activation Function Update	Recurrent
Encoder			
Input	10×1		
Bidirectional LSTM	20	Tanh	Sigmoid
Decoder			
Repeat Vector	12×20		
Bidirectional LSTM	12×20	Tanh	Sigmoid
TimeDistributed Dense	12×1		Linear

Figure 3. Breakdown of the layers present in the TadGAN (left) and AER (right) architectures.

$$r_t = \frac{w_{l,t} r_{l,t} - w_{s,t} - r_{s,t}}{C}, \quad (8)$$

where $r_{l,t}$ and $r_{s,t}$ are the percentage increases in the prices of the long and short stocks, respectively. The weights $w_{l,t}$ and $w_{s,t}$ correspond to the values of the respective long and short positions at time t. These weights are indexed by t in order to mimic the evolution of the profit and loss. Finally, C is the value of the long position at the start of the transaction.

To account for transaction costs, these values are adjusted by subtracting a transaction fee as a percentage of the trade. Note that this adjustment is done every time a position is opened or closed. That is, if a transaction occurs at time t, the computed %pnl is adjusted to include a transaction fee. Thus, the adjusted %pnl at time t, denoted by r_t^*, is

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

$$r_t^* = \begin{cases} r_t - r_{\mathrm{TC}} & \text{if a position is opened/closed at } t \\ r_t & \text{otherwise} \end{cases}, \quad (9)$$

where r_{TC} is the transaction fee expressed as a percentage of the trade. In this work, the transaction fee is assumed to be $r_{\mathrm{TC}} = 0.5\%$.

4. Results and Discussion

4.1 Profit and Loss Analysis

Table 1 shows the unadjusted mean daily %pnl when the following pairs trading strategies are employed for the five stock pairs: TadGAN, AER, the distance approach, and the cointegration approach. With one-sample *t*-test, it was observed that the mean %pnl for the anomaly detection methods is greater than zero for three out of five pairs with at least 90% confidence. Only one pair (PAL–RFM) obtained a positive %pnl with 90% confidence for the cointegration method. Therefore, using TadGAN and AER for pairs trading can yield (more) significant returns than the benchmark methods. This showcases the efficacy of using a time series anomaly detection approach. To be specific, this illustrates the profitability of using machine learning–based anomaly detection models such as TadGAN and AER.

While TadGAN and AER resulted in more profitable trades, some drawbacks worth mentioning were observed. Table 2 shows the number of times a position was opened or closed during the trading period. The anomaly detection methods had more trades than the benchmark methods. This implies that these methods would incur higher transaction costs due to the more frequent buying and selling of stocks. While lower transaction costs for the benchmark methods may appear as an advantage, their infrequent trades (such as the number of trades for the PGOLD–FPH pair using the distance method) might signal another, more pressing drawback of the method, such as the possibility that the trade will never close.

With a hypothetical transaction cost of 0.5% incurred per trade, Table 3 shows that the machine learning methods lost a greater percentage of their profits due to transaction costs. Worth noting is that AER was the most significantly affected method. Despite the loss of profit to transaction costs, AER still outperformed the other three strategies. This can also be observed in Fig. 4 where AER achieves a profit of more than 80% of the total capital employed to all pairs at the beginning of the trading period. However, these results do not account for market liquidity and slippage. Given AER's active trading strategy, trading delays caused by limited trading volume may affect the price trades are executed at, which may reduce the profit generated by the strategy.

4.2 Residual Analysis

Aside from mean daily %pnl and trade counts, residual plots can provide insights into the efficacy of the anomaly detection methods. First, note that for the residual plot of one pair (see Fig. 5a), the position never closed due to the series not crossing 0. That is, mean reversion never happened. This has the potential to incur huge losses due to keeping a short position open. However, looking at the residual plot of another pair (see Fig. 5b), anomaly detection actually improved this by setting a dynamic "normal," as opposed to the static "normal" of the benchmark methods. While this would incur losses, since we expect the movement to be from the residual series rather than the reconstructed series to earn profits, this removes the uncertainty (regarding when to close) that the benchmark methods experience.

While the results presented above indicate the potential of anomaly detection models, it is also important to contextualize them vis-à-vis the nature of emerging markets. For example, emerging markets may harbour characteristics that could affect the profitability of a trading strategy. Compared with developed markets, emerging markets are more volatile. Within emerging markets, it is further observed to be correlated with trading volume (Girard & Biswas, 2007). As such, while increased volatility could provide more pairs trading opportunities, its effects on trading volume may affect its profitability, especially for the anomaly detection strategies.

On the other hand, higher volatility could suggest a more probable occurrence of breakouts. A breakout or breakdown may cause a shift in the trend of a stock's price. This, in turn, may affect the possibility of mean reversion. These phenomena may occur more frequently in emerging markets, making them a concern for the benchmark approaches – these approaches use mean reversion as a condition to close positions. However, the dynamic mean of the anomaly detection approaches addresses this issue. That is, since the anomaly detection approaches do not rely on a static mean, there is no concern that the position will never close.

5. Conclusion

In this paper, we evaluated the efficacy and profitability of using anomaly detection models to identify arbitrage opportunities in an emerging market, namely the Philippine stock market. Different pairs trading strategies – TadGAN, AER, and distance and cointegration approaches – were employed and simulated using historical daily closing stock prices. Results showed that the time series anomaly detection approaches yielded more significant returns as compared to the benchmark approaches. This, however, was at the expense of executing more trades. Nonetheless, the adjusted mean daily %pnl showed that AER still outperformed the other strategies. Furthermore, it should be noted that both TadGAN and AER avoided the problem of positions not being closed due to mean reversion not happening.

It can thus be concluded that TadGAN and AER can be used to detect statistical arbitrage opportunities using the cointegration residuals. That is, anomaly detection models, such as TadGAN and AER, may be applied in pairs trading; of the four strategies tested,

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

Table 1. Resulting mean daily %pnl for the pairs trading strategies considered. The markers for mean return are interpreted as follows: *** significant at 99% level, ** significant at 95% level, * significant at 90% level.

Pair	TadGAN	AER	Distance	Cointegration
TFHI–STR	0.0062*	0.0090***	0.0033	0.0030
PGOLD–FPH	0.0006	0.0023**	0.0003	0.0001
ALI–MAC	0.0006	0.0016	0.0000	−0.0001
COSCO–FLI	−0.0004	0.0015	−0.0013	−0.0013
PAL–RFM	0.0020**	0.0115***	0.0003	0.0029*

Table 2. Trade counts for each pairs trading strategies considered. It can be seen that strategies employing an anomaly detection approach had more trade counts than the benchmark strategies.

Pair	TadGAN	AER	Distance	Cointegration
TFHI–STR	19	40	7	6
PGOLD–FPH	6	48	3	2
ALI–MAC	11	33	2	2
COSCO–FLI	8	41	1	1
PAL–RFM	8	54	2	8

Table 3. Summary of how aggregate %pnl are affected by transaction costs per trading strategy. A 0.5% fee is subtracted from the %pnl whenever a position is opened/closed.

Strategy	Mean Daily %pnl	Adjusted Mean Daily %pnl	Retained Profit
TadGAN	0.0090	0.0069	76.97%
AER	0.0260	0.0171	66.01%
Distance	0.0026	0.0021	80.19%
Cointegration	0.0046	0.0039	85.14%

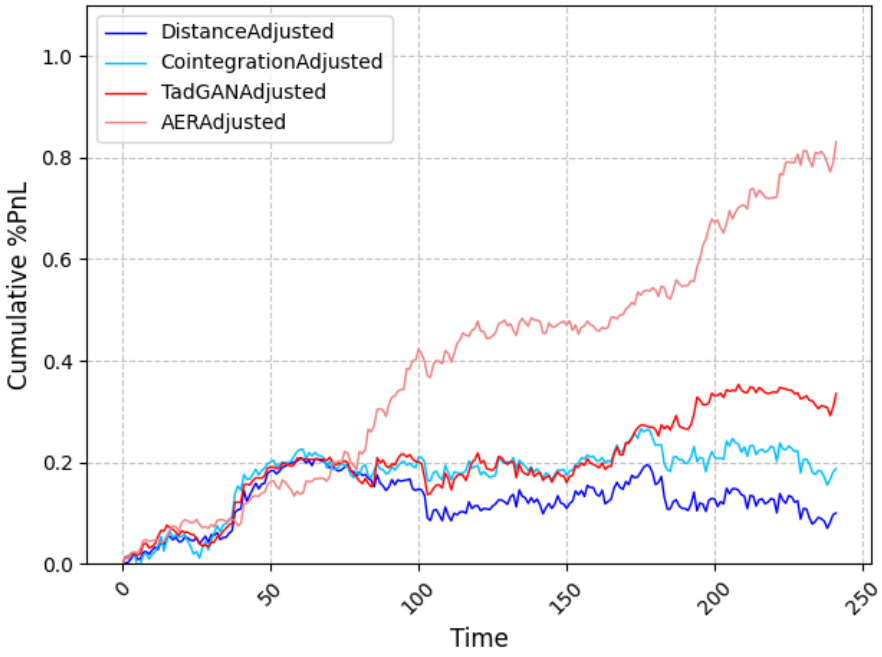


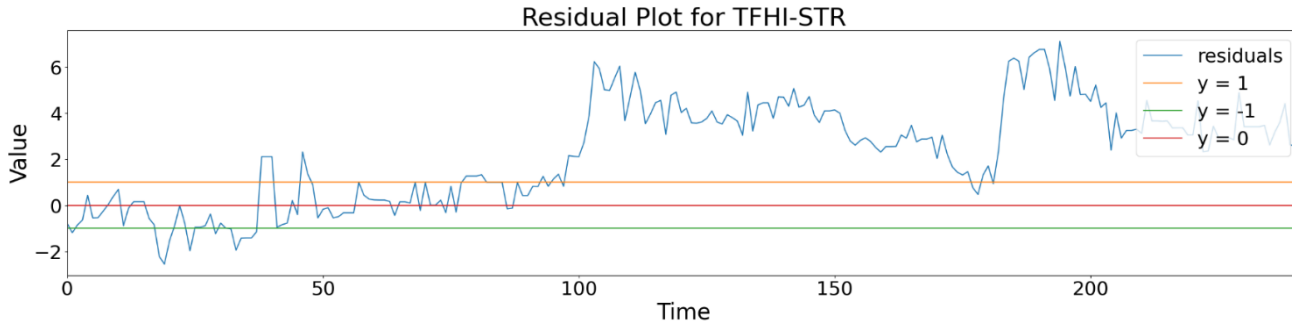
Figure 4. Cumulative %pnl over the trading period for aggregated pairs trading strategies, considering a 0.5% trading cost.

AER had the best performance. The distance and cointegration approaches encountered the problem of positions not closing at mean reversion. This is because at some point, the residual series did not exhibit mean-reverting properties. Contrarily, while the anomaly detection models executed more trades than the benchmark methods (and may close on a loss), overall, they addressed the problem of trades not closing and obtained higher returns than the benchmarks.

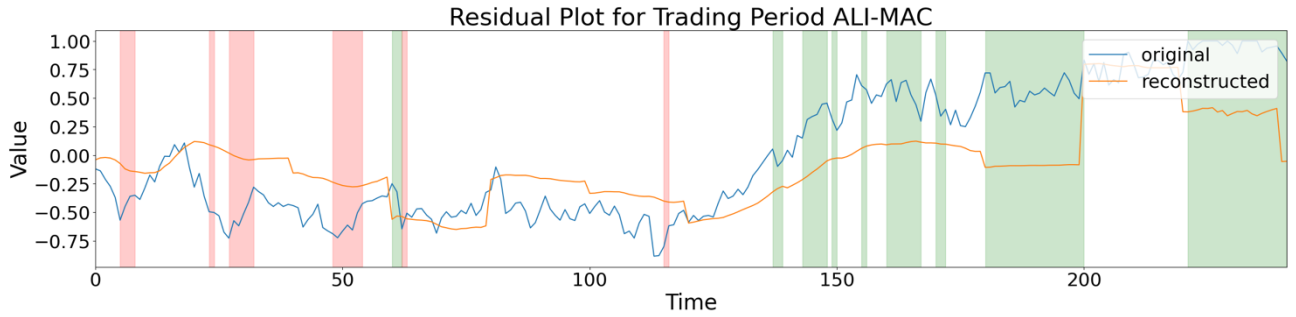
5.1 Future Considerations

While this paper was able to establish the applicability and efficacy of TadGAN and AER in pairs trading, their viability must be further tested. Considering other factors, such as transaction costs, interest expenses, or dividends, and incorporating further data such as bid–ask spreads may provide a more realistic

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)



(a) Cointegration residuals of one stock pair



(b) Original and reconstructed cointegration residuals of another stock pair

Figure 5. Cointegration residuals plots of two stock pairs. The green region indicates when the first trading rule (short the first stock, long the second) is triggered, while the red region is for the second rule (long the first stock, short the second). **(a)** illustrates the basis for how trading is triggered for the benchmark methods. **(b)** illustrates the original and reconstructed residuals, which are used to identify when trading is triggered for the anomaly detection methods.

measure of AER's performance. As the feasibility of the anomaly detection models has been established in this work, their practicability can be considered in future work.

Furthermore, a return calculation tailored to the cointegration approach could be formalized, as this study provided its own performance metric (%pnl). Note that this metric is not exactly similar to returns. Formulating this could more accurately represent the performance of these strategies. Lastly, improvements to the two models can still be explored to enhance their profitability. For example, utilizing both models' residuals could potentially address their limitations and help improve how anomalies are detected. Hyperparameter tuning and optimization could also be considered to optimize the resulting profit and loss metrics.

6. Acknowledgement

The authors would like to thank the Ateneo de Manila University Department of Mathematics and Kristian Mamforte, Christian Paul Chan Shio, and Juan Carlo Mallari for their constructive feedback on this work. A conference presentation of this work was funded by the Ateneo John Gokongwei School of Management Dean's Office.

7. References

Ajewole, K., Adejuwon, S., & Jemilohun, V. (2020). Test for stationarity on inflation rates in Nigeria using augmented Dickey Fuller test and Phillips-Persons test. *IOSR Journal of Mathematics*, *16*(3), 11–14.

Aquino, R. (2002). *Informational efficiency characteristics of the Philippine stock market* (Technical Report). UP College of Business Administration.

Canady, T. S. (2022). *Making sense of market anomalies*.

Carrasco Blázquez, M., De la Orden De la Cruz, C., & Prado Rom´an, C. (2018). Pairs trading techniques: An empirical contrast. *European Research on Management and Business Economics*, *24*(3), 160–167.

Fernández-Pérez, A., de las Nieves López-García, M., & Requena, J. P. R. (2020). On the sensibility of the pairs trading strategy: The case of the FTS Stock Market Index. *Studies of Applied Economics*, *38*(3).

Gatev, E., Goetzmann, W., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, *19*(3), 797–827.

Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A., & Veeramachaneni, K. (2020). *TadGAN: Time series anomaly detection using generative adversarial networks*.

Girard, E., & Biswas, R. (2007). Trading volume and market volatility: Developed versus emerging stock markets. *Financial Review*, *42*(3), 429–459.

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, *31*(2), 513–545.

Moura, M. (2010). Testing the Taylor model predictability for exchange rates in Latin America. *Open Economies Review*, *21*, 547–564.

Nakashima, T., & Yairi, T. (2023). Assessing the performance of transformer for time series anomaly detection. In *PHM Society Asia-Pacific Conference* (Vol. 4).

Rad, H., Low, R. K. Y., & Faff, R. (2016). The profitability of pairs trading strategies: Distance, cointegration and copula methods. *Quantitative Finance*, *16*(10), 1541–1558.

Tallboys, J., Zhu, Y., & Rajasegarar, S. (2022). Identification of stock market manipulation with deep learning. In *Advanced Data Mining and Applications* (pp. 408–420). Springer International Publishing.

Ullah, F., Ullah, S., Srivastava, G., & Lin, J. C.-W. (2024). IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic. *Digital Communications and Networks*, *10*(1), 190–204.

Wong, L., Liu, D., Berti-Equille, L., Alnegheimish, S., & Veeramachaneni, K. (2022). *AER: Auto-encoder with regression for time series anomaly detection*.

DOI: https://doi.org/10.22452/mjs.vol45sp2.2
Malaysian Journal of Science 45 (Special Issue 2): 16-24 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026) https://mjs.um.edu.my

Evaluation of a Harmful-Animal Detection Method Combining Techniques for Object Detection and Skeleton Estimation

Shinichi Bota[1a]* and Naoyuki Morimoto[2a]

Abstract: This paper presents a new harmful-animal detection method that combines conventional algorithms for image recognition with skeleton estimation techniques. In many places around the world, wild animals encroach on human territory, causing personal injury and damage to crops. One way to address this social problem is to implement a harmful-animal detection system using cameras and object detection algorithms; however, current methods face the challenge of low detection accuracy, limiting their practical application. This study aims to address this challenge through a novel approach that combines conventional algorithms for image recognition with skeleton estimation techniques. The proposed method uses transfer learning on two pre-trained models: YOLOv8—known for its high accuracy and real-time performance—and HRNet, a widely used model for animal skeleton estimation. For training, a dataset extracted from AP-10K was used. In the proposed method, animals in the input image are detected by combining the inference results from two image recognition algorithms. The initial result is obtained from the input image itself, while the subsequent result is derived from a skeleton image created by plotting estimated keypoints and connecting lines. The final inference result is weighted and combined to improve the accuracy of animal detection in the image. The proposed method was evaluated by comparing its detection results with those of a conventional detection method while varying the minimum confidence threshold (*conf*). The evaluation demonstrated that the proposed method performs as well as or better than the conventional detection method. Furthermore, the proposed method exhibited greater overall stability, which simplified the process of *conf* parameter adjustment, making it more user-friendly. These findings suggest that the proposed method can contribute to the development of a more practical and effective harmful-animal detection system.

Keywords: image recognition, object detection, skeleton estimation, YOLOv8, AP-10K

1. Introduction

In many parts of the world, wild animals encroach on human territory, causing personal injury and significant damage to crops—one of the major social issues humanity currently faces. This phenomenon, commonly referred to as human–wildlife conflict, is believed to be driven by competition between humans and wildlife for the same resources, whether food or space (König et al., 2021).

For example, in Malaysia, data up to 2019 show an increasing trend in human–wildlife conflicts since 2012 (Xin et al., 2024), highlighting the persistent nature of this problem. Moreover, wild animals entering human territory often need to be culled, further complicating the situation. As land development and environmental changes caused by global warming continue, these conflicts are expected to persist. While discussions about coexistence between humans and wildlife have become more common in recent years, it is crucial to address this issue promptly to establish a society in which humans and animals can coexist.

In response to this problem, harmful-animal detection systems using cameras and object detection algorithms have been developed in the field of information technology; however, effective harmful-animal detection systems capable of significantly reducing wildlife damage have yet to be realized. One of the primary challenges lies in improving detection accuracy, which is hindered by the limited availability of data. The scarcity of animal data can be attributed to several factors: animals do not cooperate in data collection, the wide variety of animal species, and the low populations of some species, making it difficult to gather large datasets (Jiang et al., 2022). Previous studies have explored methods addressing data scarcity, such as data augmentation (Kumar et al., 2024) and synthetic datasets combining generated images (Mu et al., 2020); however, synthetic data, while promising, introduces biases that are difficult to eliminate. Therefore, this study aims to develop a harmful-animal detection system using an alternative approach to address the data scarcity issue.

This paper proposes a new method for animal detection that combines conventional algorithms for image recognition with skeleton estimation techniques. While prior studies have improved detection accuracy using keypoints or applied detected keypoints for behavior classification and analysis (Chen et al., 2023), this study focuses instead on inferring animal types directly from estimated keypoint information. To the best of our knowledge, no previous study has explored this approach. The method proposed in this study integrates the results of conventional detection algorithms with those obtained from

Authors information:
[a] 1577 Kurimamachiya-cho Tsu city, Graduate School of Engineering, Mie University, Mie, JAPAN. E-mail: 425m527@m.mie-u.ac.jp[1], morimoto@info.mie-u.ac.jp[2]

*Corresponding Author: 425m527@m.mie-u.ac.jp

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

images annotated with skeleton-estimated information. We hypothesize that performance could be improved primarily by utilizing inference results from conventional detection algorithms and supplementing them with inference results derived from skeleton estimation techniques.

Skeleton estimation techniques are used to predict the positions of keypoints from images or videos. During animal detection using YOLO, bounding boxes often accurately detect the size of the detected object; however, they are prone to mistakenly including background elements or incorrectly inferring the animal class. To address this, we propose combining skeleton estimation techniques to increase the information available for inference. Incorporating skeleton information can mitigate the issue of data scarcity and enhance detection robustness. In addition, conventional detection methods often require the optimization of multiple parameters for prediction, such as the minimum confidence threshold parameter. We aim to address this challenge as well.

In this study, we focus on detecting the following classes: cow, horse, and bear. Bears were chosen because of the high risk they pose to humans. If a human accidentally encounters a bear, it is difficult to escape or defend oneself. For example, in Japan and Malaysia, bears are responsible for numerous incidents of crop damage and personal injuries annually (Xin et al., 2024). Cows and horses were selected because their skeletal structures are similar to those of bears and because they inhabit similar regions. Since our method incorporates skeleton information, demonstrating its effectiveness in animals with similar skeletal structures could indicate its broader applicability.

The proposed method uses models that are transfer-learned from the pre-trained models of YOLOv8 (Jocher et al., 2023) and HRNet-w48. YOLOv8 is known as an object detection algorithm with excellent accuracy and real-time performance. HRNet-w48 is a high-accuracy model included in MMPose (MMPose Contributors, 2020), a skeleton estimation algorithm library. The dataset used for training is AP-10K (Yu et al., 2021), which includes 23 animal families and 54 species with high-quality keypoint annotations.

To evaluate the proposed method, we compare its performance with a conventional detection method by varying *conf*, a parameter that determines the minimum confidence threshold. Performance is measured using standard metrics, such as precision, recall, and F1 score. The ANIMAL-POSE DATASET (Cao et al., 2019) is used for performance evaluation.

The results demonstrated that the proposed method achieved higher and more stable evaluation metric values than the conventional detection method. Although the proposed method also required the optimization of several parameters, the changes in performance caused by varying these parameters were smaller than those observed with the conventional detection method. This indicated that the proposed method ensured a certain level of accuracy without extensive parameter optimization, making it more user-friendly. Furthermore, when comparing the average performance of both methods, the conventional detection method exhibited a slightly higher recall. However, the proposed method achieved higher precision, resulting in a higher F1 score. Therefore, the proposed method can be considered slightly superior to the conventional detection method.

2. Related Research

Dave et al. (2023) used YOLOv8 to improve the detection accuracy of lion, tiger, bear, and leopard classes using YOLOv8 models, including YOLOv8m, YOLOv8l, and YOLOv8x. YOLO was chosen because of its proven high accuracy in the field of real-time object detection. The authors noted that model performance could be enhanced through data augmentation techniques. Inspired by this research, we decided to adopt YOLOv8 in our study as well; however, unlike Dave et al. (2023), we chose to use YOLOv8n. We also aimed to improve accuracy by using a distinct approach from data augmentation techniques.

Furthermore, using skeleton estimation techniques for detection, Chen et al. (2023) proposed an animal motion prediction model combining object detection and pose estimation to enable autonomous driving systems to assess animal postures more accurately. Initially, they used a convolutional neural network–based object detection method to detect animals in images. They then employed an animal pose estimation method to extract keypoints from the detected animal images. Finally, the images were classified into three categories of motion: running, walking, and standing. After extracting the keypoints, the authors used ResNet and MLP models to classify the pose. The MLP model classifies actions based on the angles of the keypoints. While their study focused on classifying behaviors from keypoint information, our study aims to infer the type of animal from the obtained keypoint data.

3. Proposed Method

3.1 Overview

For the proposed method, a system was constructed that primarily utilizes inference results from conventional detection algorithms and supplements them with inference results derived from skeleton estimation techniques. The overall structure of the proposed method is shown in Figure 1.

This method consists of a combination of three main algorithms: an image recognition algorithm, a skeleton estimation algorithm, and a skeleton image recognition algorithm.

The image recognition algorithm detects animals in the input image and outputs the predicted class, bounding box size, and confidence score. This process is identical to that of standard object-detection algorithms. Additionally, for each detected bounding box, the input image is cropped, and the cropped image is stored. The skeleton estimation algorithm estimates the skeleton from the input image. Similar to the image recognition algorithm, it detects animals in the input image and outputs the predicted class, bounding box size, and confidence score; however, unlike the image recognition algorithm, the input image for this algorithm is a skeleton image generated by overlaying the keypoints predicted by the skeleton estimation algorithm and the

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

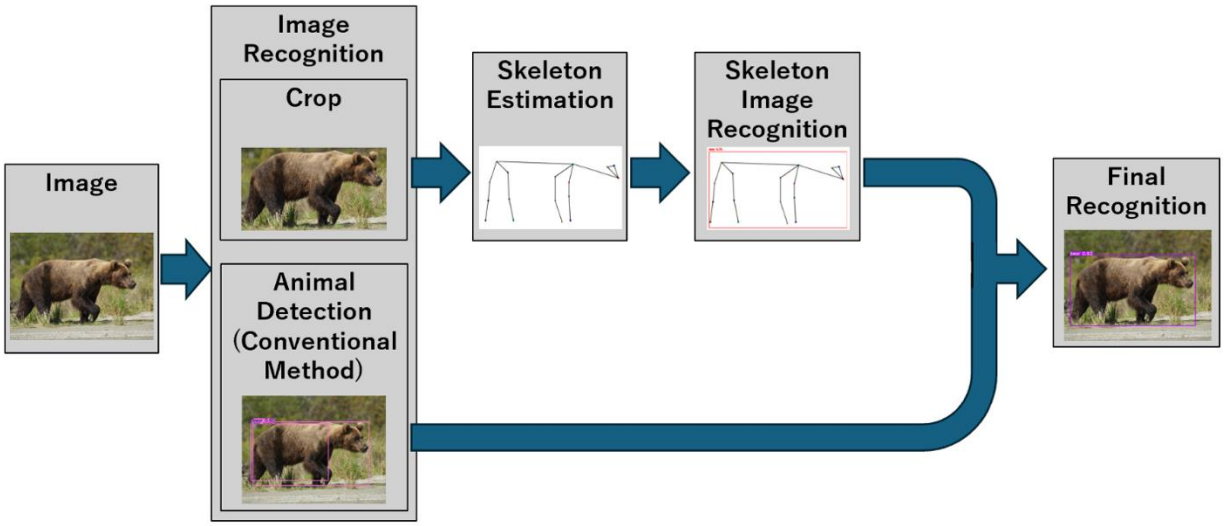


Figure 1. Overview of the proposed method

lines connecting these keypoints onto a white background. In this paper, such images are referred to as skeleton images.

Finally, the system outputs a weighted combination of the inference results from the image recognition algorithm and the skeleton image recognition algorithm.

3.2 Pre-trained Models for Each Algorithm

3.2.1 Dataset

This section describes the pre-trained models used by the three algorithms. The dataset used to create each pre-trained model was a modified version of the AP-10K dataset, containing only data for cows, horses, black bears, and brown bears, which had been extracted and processed accordingly. The authors of the AP-10K dataset explained in their paper that they had recruited 13 well-trained annotators and conducted three rounds of cross-checks and corrections both manually and automatically (Yu et al., 2021). Consequently, this dataset was deemed to possess high annotation quality and was selected for use in this study. While the proposed method uses skeletal information for classification, it was determined that there is no distinction between the skeletons of black bears and brown bears. Therefore, the data for these two species were combined into a new class labeled “bear.” The data corresponding to the three labels were selected from AP-10K and randomly divided into training, validation, and test sets. These data were split into three sets: 60% for training, 20% for validation, and 20% for testing, a partitioning scheme commonly used in machine learning research.

The models utilized for the image recognition and skeleton estimation algorithms were trained using a dataset derived from the AP-10K dataset. In this process, the original images and their corresponding label information from AP-10K were converted into YOLO format. For the skeleton image recognition algorithm, the model was trained on a dataset composed of skeleton images generated from keypoint information extracted from AP-10K, along with label information converted into the YOLO format. In the skeleton images, each keypoint was assigned a distinct color, and the keypoints were connected based on the label information, as illustrated in Figure 2. It was assumed that there might be patterns in the distances between and positions of the keypoints; therefore, the keypoints were color coded to facilitate learning.

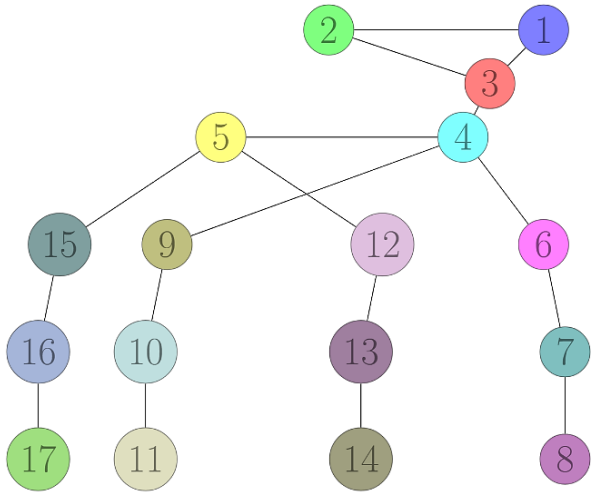


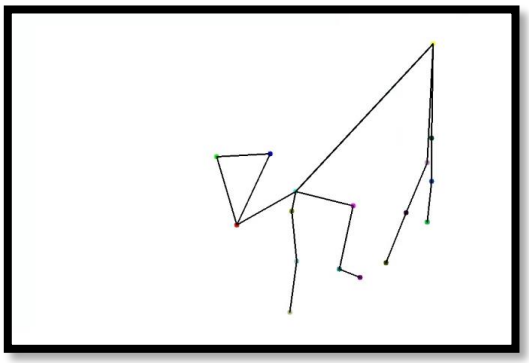
Figure 2. Keypoint connections

To provide a concrete example, the dataset includes images such as the one shown in Figure 3a. Based on the keypoint information extracted from the labels for this image, a corresponding skeleton image was generated, as shown in Figure 3b. The dataset comprising images such as those in Figure 3a was used to train the models for the image recognition algorithm and the skeleton estimation algorithm, whereas the dataset comprising skeleton images, represented in Figure 3b, was employed to train the model for the skeleton image recognition algorithm.

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)



(a) Base image



(b) Skeleton image

Figure 3. Examples of images used for training each model

3.2.2 Model Training

The image recognition algorithm model was created by applying transfer learning to YOLOv8n, which is known for its high accuracy and real-time performance. This algorithm is the same as the one used for the conventional detection method. The details of its use as a conventional detection method are explained later. To ensure a fair comparison between the proposed method and the conventional detection method, both methods were trained using the same parameters. Transfer learning was conducted using YOLO's default settings, except for the number of epochs, to achieve general animal detection.

Regarding the skeleton estimation algorithm, although several such models exist—including pre-trained models trained on the AP-10K dataset within MMPose—we chose not to use them. Instead, we developed our own pre-trained model. This decision was made because we believe that the accuracy of skeleton estimation would have a significant impact on the results of the proposed method. To improve the accuracy of skeleton estimation, we reduced the number of animal species in the transfer learning dataset from 54 to three and developed a new model.

The model used in the skeleton estimation algorithm was created by applying transfer learning to the pre-trained model hrnet_w48-8ef0771d.pth, available from the MMPose library, which specializes in skeleton estimation algorithms. The code for transfer learning using the HRNet-w48 pre-trained model and the AP-10K dataset is provided in MMPose's GitHub repository (MMPose Contributors, 2020). Using this code as a reference, we implemented the skeleton estimation algorithm.

Similar to the image recognition algorithm, the skeleton image recognition algorithm model was created by applying transfer learning to YOLOv8n. The parameters for training were the same, except for the number of epochs, which were adjusted. As previously mentioned, the dataset used for transfer learning consisted of skeleton images. The representative parameters used for training the models are presented in Table 1.

Table 1. Parameters

Argument	Image recognition algorithm	Skeleton estimation algorithm	Skeleton image recognition algorithm
Epochs	1000	840	1000
Batch	16	32	16
Image size	640×640	256×256	640×640
Optimizer	Auto	Adam	Auto

3.3 Novel Mechanism of the Proposed Method

3.3.1 Skeleton Estimation Method

The most novel aspect of the proposed method is the integration of skeleton estimation techniques. This section explains the skeleton estimation method, followed by a discussion in the next section on how the inference results from the original and skeleton images are integrated. Finally, the execution flow of the proposed method is described.

The pre-trained model employed in the skeleton estimation algorithm follows a top-down approach. A key assumption of top-down skeleton estimation methods is that they expect a single instance to be present in the input. Therefore, when using such approaches, it is common to initially detect objects using an object detection algorithm, followed by keypoint detection for each object within its bounding box.

In the proposed method, the image recognition algorithm is executed prior to the skeleton estimation algorithm. During this process, the images are cropped according to the inferred bounding boxes, and the resulting cropped images are stored. These cropped images are then used as inputs to the skeleton estimation algorithm. In other words, the input images are the cropped outputs of the image recognition algorithm, with each one expected to contain only a single animal.

The skeleton estimation algorithm assumes that the input image contains only a single animal and estimates its skeleton. For keypoints in the estimation results with a confidence score greater than 0.5, the following processing steps are performed:

- Keypoints are displayed on a white background image the same size as the cropped image, with each keypoint assigned a different color.
- When keypoints are adjacent, they are connected according to Figure 2.

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

In MMPose, the *visualize* function, which performs visual prediction, has a default keypoint visualization threshold of 0.3. For the proposed method, this threshold was set to a higher value of 0.5. The rationale for this decision is that skeleton estimation results are used for subsequent inferences. Using data with low confidence scores would imply relying on data that were likely to have been misdetected during skeleton estimation. If such misdetected data were treated as correct data in later steps, they might negatively impact the accuracy of subsequent inference results. Therefore, the threshold was set higher to minimize this risk.

3.3.2 Integration Method

The inference results from the image recognition and skeleton image recognition algorithms are integrated under specific conditions. To begin with, when a bounding box is detected from the skeleton image estimated by the skeleton estimation algorithm, the results are discarded if the area of the inferred bounding box is less than half the area of that detected by the image recognition algorithm from the input image. This condition was established because when cropped images were created using the image recognition algorithm, there were cases in which the cropped images contained multiple animals or none at all.

The size of the bounding box to be output as the final inference result was set to the size inferred during the execution of the image recognition algorithm. The primary reason for this decision was the difficulty of determining the correct bounding box size from the skeleton image. Skeleton images generated by the skeleton estimation algorithm sometimes contain only a few keypoints, depending on the image. This occurs due to the high threshold setting or the presence of challenging images for estimation. In such cases, inferring the bounding box from the skeleton image becomes problematic. Furthermore, when inferring the bounding box using the skeleton image recognition algorithm, the bounding box is detected within the cropped image. Therefore, the inferred bounding box is, at most, the size of the bounding box detected by the image recognition algorithm, but is usually smaller. As a result, the size of the bounding box to be output as the final inference result was set to the size inferred by the image recognition algorithm.

When the results are integrated, a weighted sum is calculated. The confidence score inferred from the input image is multiplied by 0.7, while the confidence score inferred from the skeleton image is multiplied by 0.3, and the results are summed. The reason for assigning more weight to the confidence score inferred from the input image is that the method's purpose is to rely primarily on the results of conventional detection algorithms, with the results obtained from skeleton estimation techniques serving as supplementary information to improve performance. Under these conditions, the optimal ratio was determined using the evaluation method described in this paper. Consequently, a ratio of 0.7 for the confidence score inferred from the input image and 0.3 for that inferred from the skeleton image was adopted, as it yields the most balanced overall performance. The evaluation results are provided in Appendix Table A. The mathematical formulation of the integration method is presented in Equation (1).

$$\text{score}(R) = \begin{cases} \text{score}(R_1) \times 0.7 + \text{score}(R_2) \times 0.3, \text{if bbox}(R_2) \geq \frac{1}{2} \times \text{bbox}(R_1) \\ \text{score}(R_1) \times 0.7, \text{if bbox}(R_2) < \frac{1}{2} \times \text{bbox}(R_1) \end{cases} \quad (1)$$

In this formulation, R is the final detection result, $\text{score}(R)$ is the final confidence score, R_1 is the detection result from the original image, $\text{score}(R_1)$ is the confidence score of R_1, and $\text{bbox}(\text{R}_1)$ indicates the area of the bounding box derived from the original image. Likewise, R_2 is the detection result from the skeleton image, $\text{score}(\text{R}_2)$ is the confidence score of R_2, and $\text{bbox}(\text{R}_2)$ indicates the area of the bounding box derived from the skeleton image.

Since the final inference results of the proposed method aim to be in the same output format as the YOLO inference results, the summed confidence score is also expressed between 0 and 1, mirroring the YOLO confidence score range. However, it should be noted that because the inference method of YOLO and the calculation method of the proposed method differ, the confidence score output of the proposed method cannot be directly compared to that of YOLO.

3.3.3 Execution Flow

The proposed method assumes 2D images as inputs. First, the image recognition algorithm is executed. Using the pre-trained model, the algorithm then detects animals within each input image. The output includes the class of the detected animal, the bounding box size, and the confidence score for each detection. The confidence score is compared to a predefined threshold. If the score exceeds this threshold, the label information and the cropped image, obtained by trimming the region corresponding to the inferred bounding box, are stored. These cropped images are then used by the skeleton estimation algorithm. The inference results from this step are combined with those from the skeleton image recognition algorithm to detect the animal present in the image.

Next, the skeleton estimation algorithm is executed. In this algorithm, a pre-trained model is employed to estimate the skeleton for each cropped image stored by the image recognition algorithm. The output consists of the positions of keypoints and their respective confidence scores. Keypoints with confidence scores exceeding a threshold of 0.5 are plotted on a white background image the same size as the cropped image. Additionally, the keypoints are connected by lines based on their numbering. The resulting image, which includes both keypoints and connecting lines, is stored as the output of this algorithm. This output image is a skeleton image in which the keypoints and the lines connecting them are drawn on a white background matching the size of the input image.

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

Finally, the skeleton image recognition algorithm is executed. This algorithm uses a pre-trained model to infer the animal in each stored skeleton image. The output consists of the animal class, the bounding box size, and the confidence score, similar to the output of the image recognition algorithm. If the confidence score exceeds the predefined threshold, the inference results are stored.

The inference results from the image recognition and skeleton image recognition algorithms are integrated. For each cropped image, the bounding boxes detected by both algorithms are compared. If the overlap between the bounding boxes from the image recognition and skeleton image recognition algorithms exceeds a predefined threshold of 0.5, the confidence scores from both algorithms are weighted and summed. The confidence score from the image recognition algorithm is assigned a weight of 0.7, while that from the skeleton image recognition algorithm is assigned one of 0.3. After calculating the weighted sum, the animal class with the highest confidence score is selected. A text file is generated containing three pieces of information: the detected animal class, the bounding box size detected by the image recognition algorithm, and the highest confidence score.

The confidence score recorded in the text file is compared to the predefined threshold. If the score falls below this threshold, the cropped image is classified as containing no animal, and the corresponding information in the text file is deleted. If the score exceeds the threshold, the information in the text file is retained as the label for the cropped image. The labels are then organized according to the inference results from the image recognition algorithm, and a label is generated for each input image. An image is generated displaying labels following the output format of YOLO. Each label contains the following information: class information, x center, y center, width, height, and confidence score. The execution flow of the proposed model is presented as a pseudocode in Algorithm 1.

4. Performance Evaluation and Comparison

4.1 Methods for the Evaluation Experiment

To conduct a performance evaluation, a labeled dataset is essential, as it serves as the ground truth for assessing the performance of predicted labels. In this study, the evaluation dataset was derived from the ANIMAL-POSE DATASET, from which only data for three animal classes—cows, horses, and bears—were extracted for the evaluation.

Three widely used metrics in machine learning—precision, recall, and F1 score—were employed for performance evaluation. During the evaluation, the intersection-over-union threshold was set to 0.5. The predicted and ground truth classes were compared to assess performance only if the bounding box overlap exceeded 50%. In YOLOv8, one of the inference arguments was *conf*, which sets the minimum confidence threshold for detection. The default value of this parameter was set to 0.25. In our study, this value was varied for evaluation purposes. For each minimum confidence threshold, evaluation results were calculated for each class as well as for the overall average.

A detailed explanation of the performance evaluation follows. By varying the minimum confidence threshold, the conventional detection method and the proposed method were evaluated using three metrics. The results were compared to assess the effect of combining skeleton estimation techniques. For each method, the minimum confidence threshold was set to [0.2, 0.25, 0.3, 0.4, 0.5, 0.6], including a default value of 0.25. The conventional detection method has only one place for setting the *conf* parameter; hence, it calculates six evaluation results. In contrast, the proposed method has three locations for setting it; hence, 216 evaluation results are computed. Specifically, the *conf* parameter is set at three distinct points: the image recognition algorithm, the skeleton image recognition algorithm, and the weighted integration of the results from each algorithm.

As a method of performance comparison, this study employed box plots focusing on the overall mean evaluation results—specifically, macro precision, macro recall, and macro F1 scores. Initially, a comprehensive performance comparison was conducted between the conventional detection method and the proposed method. Three tables are presented, corresponding to macro precision, macro recall, and macro F1 scores, with each table containing two box plots. The first box plot represents the total results for the conventional detection method, where six evaluation results were obtained by varying the minimum confidence threshold and are grouped into a single box plot. The second box plot represents the total results for the proposed method, where 216 evaluation results were obtained by varying the minimum confidence threshold and are grouped into a single box plot.

Subsequently, the performance variations were examined using box plots when changing the *conf* parameter at the final integration stage of the proposed method, where the inference results from each algorithm are weighted and integrated. Three tables are presented, corresponding to macro precision, macro recall, and macro F1 scores. The proposed method was evaluated by varying the *conf* parameter in six ways, with 36 evaluation results grouped into a single box plot for each minimum confidence threshold. In general, the object detection algorithms showed that increasing the minimum confidence threshold above YOLO's default value improved performance. This was because precision increased as only highly confident detections were retained, whereas recall decreased because of a reduction in the number of detections. As a result, the F1 score, functioning as the harmonic mean of the two, improved, leading to a more balanced detector. The proposed method was examined to determine whether changes in the *conf* parameter led to variations similar to those observed in conventional detection methods.

Furthermore, the average evaluation results for each class and the overall average evaluation results of the conventional detection method and the proposed method are presented, allowing for a comparison of which method demonstrated superior performance on average.

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

Algorithm 1 Proposed Model

```
Input: image - Input image
Output: R_list - Final confidence scores

function Main(image)
    (cropped_images, labels) ← ExecuteImageRecognition(image)
    R_list ← []
    for i = 1 to length(cropped_images) do
        crop ← cropped_images[i]
        label ← labels[i]
        (confidence_original, size_original) ← ComputeConfidenceAndSize(label)
        keypoint_label ← ExecuteSkeletonEstimation(crop)
        filtered_keypoints ← FilterKeypointsByConfidence(keypoint_label, 0.5)
        skeleton_image ← OutputSkeleton(filtered_keypoints)
        label_skeleton ← ExecuteSkeletonImageRecognition(skeleton_image)
        (confidence_skeleton, size_skeleton) ← ComputeConfidenceAndSize(label_skeleton)
        if size_skeleton ≥ 0.5 × size_original then
            R ← 0.7 × confidence_original + 0.3 × confidence_skeleton
        else
            R ← 0.7 × confidence_original
        end if
        append(R, R_list)
    end for
    return R_list
end function

function ExecuteImageRecognition(image)
//Infer the animal from an image (conventional detection method)
     return (cropped_images, labels) // Multiple cropped images and their labels
end function

function ExecuteSkeletonEstimation(image)
//Estimate the skeleton from an image
    return keypoint_label // Contains 17 keypoints (x, y, confidence)
end function

function ExecuteSkeletonImageRecognition(image)
// Infer the animal from a skeleton image
    return label
end function

function ComputeConfidenceAndSize(label)
// Compute the confidence score and bounding box size from the label
    return (label.confidence, label.width × label.height)
end function

function FilterKeypointsByConfidence(keypoint_label, threshold)
// Find keypoints exceeding the threshold
    return {kp ∈ keypoint_label | kp.confidence > threshold}
end function

function OutputSkeleton(keypoints)
// Draw keypoints and the lines connecting them on a white background
    return skeleton_image
end function
```

4.2 Results and comparison

As mentioned earlier, the conventional detection method described here used the same pre-trained model as the image recognition algorithm used by the proposed method. Therefore, a model trained through transfer learning using YOLOv8n on the dataset extracted from AP-10K was used.

The performance results for each method across the three metrics are represented as box plots in Figure 4. In addition, the performance of the proposed method across the three metrics,

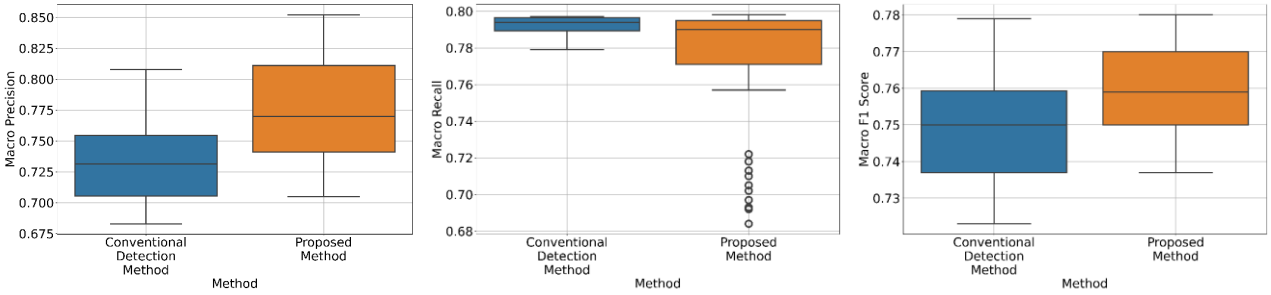
DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

when the *conf* parameter was varied at the final integration stage, is represented as box plots in Figure 5. Table 2 shows the average performance results for each method across the three metrics.

According to Figure 4, the proposed method demonstrated superior overall performance and stability. This is shown by the smaller spread of the whiskers in the box plots and the numerical values, indicating that the proposed method achieved more stable and overall better performance. In Figure 4c, the maximum values for both the conventional and proposed methods are almost identical; however, the range of the box for the conventional method is wider, highlighting the increased stability of the proposed method. From Figure 4a and b, it can be observed that the proposed method achieved a higher precision than the conventional detection method, although the recall decreased slightly. Comparing these two figures, the box plots of precision and recall for the proposed method are both centered on 0.78, indicating that the proposed method maintained a well-balanced performance. Therefore, from Figure 4, it can be concluded that the proposed method exhibits both good balance and high stability.

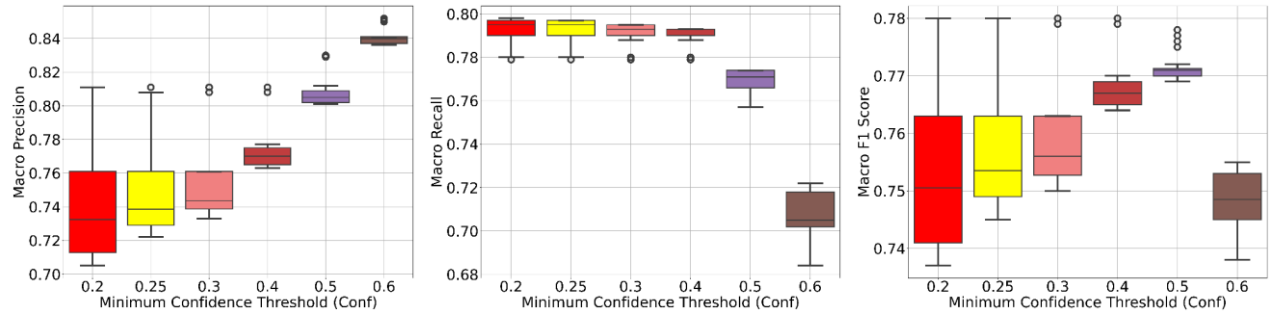
Figure 5 displays the performance variations of the proposed method by adjusting the *conf* parameter at the final integration stage, where the inference results from each algorithm are weighted and integrated. The data confirmed that the highest performance stability was achieved when the *conf* parameter was set to 0.5. Additionally, for scenarios in which recall performance was prioritized while maintaining stable detection, a *conf* parameter of 0.4 was an effective option.

Here, we discuss not only the overall evaluation results but also those for each animal class. Prior to this research, YOLO generally detected bounding box sizes with high accuracy, whereas the inference for animals was occasionally inaccurate. The data presented in Table 2 reveal that the conventional detection method showed higher performance for recall than for



(a) Macro precision (b) Macro recall (c) Macro F1 score

Figure 4. Performance comparison of the two different methods



(a) Macro precision (b) Macro recall (c) Macro F1 score

Figure 5. Performance of the proposed method for different *conf* parameters

Table 2. Mean Performance Difference Across Methods

Method	Class	Precision	Recall	F1 score	Macro precision	Macro recall	Macro F1 score
Conventional detection method	Cow	0.736	0.600	0.660	0.736	0.792	0.749
	Horse	0.586	0.898	0.709			
	Bear	0.885	0.876	0.880			
Proposed method	Cow	0.796	0.580	0.668	0.779	0.774	0.760
	Horse	0.619	0.895	0.731			
	Bear	0.921	0.845	0.880			

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

Table 3. Mean Performance Using Only Keypoint Information

Method	Class	Precision	Recall	F1 score	Macro precision	Macro recall	Macro F1 score
Skeleton	Cow	0.570	0.349	0.416			
Recognition	Horse	0.571	0.783	0.656	0.625	0.515	0.529
Method	Bear	0.734	0.414	0.514			

precision. Generally, precision and recall are regarded as being in a trade-off relationship, requiring a balance between these metrics to improve model performance. From Table 2, it is evident that the proposed method achieved a balanced trade-off between macro precision and macro recall. In addition, the proposed method outperformed the conventional detection method in terms of the F1 score for all classes, demonstrating stable and high performance across all classes. This indicates that the proposed method, by incorporating skeleton estimation techniques, successfully addresses the problem of detecting background objects or making incorrect inferences for animals, which is one of the key objectives of this study.

The results of inferring an animal solely from the estimated keypoint information are presented in Table 3.

As shown in Table 3, using the same evaluation method as that in Table 2, recall was low. These results suggest that inferring an animal solely from keypoint information is difficult. Thus, in animal detection, keypoint information should be utilized as supplementary data for inference, as employed in this study.

To assess the feasibility of the proposed method for harmful-animal detection, the execution time per image was measured. The experimental environment used for the evaluation is summarized in Table 4.

Table 4. Experimental Environment

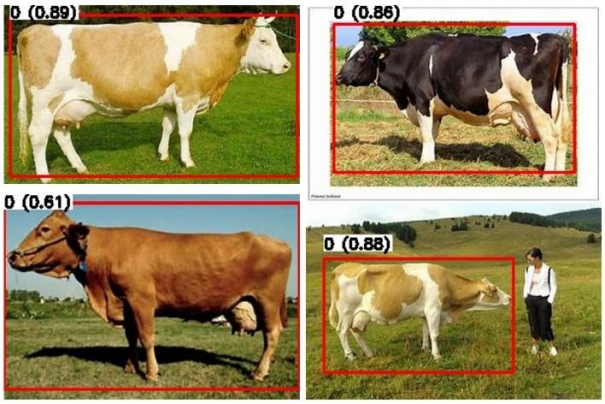
Environment	
CPU	Intel Core i5-12400 2.50 GHz
GPU	NVIDIA GeForce RTX 3080 10 GB VRAM
Memory	16 GB
Architecture	x86_64
Distribution	WSL-Ubuntu 22.04

The experimental measurements indicate that the proposed method required approximately 5.66 seconds per image. In contrast, using only the object detection model resulted in an execution time of approximately 2.33 seconds, representing a 2.4-fold increase. While it can be observed that the proposed method took longer to execute, the comparison results demonstrated that it achieved higher stability without requiring complicated optimization. This enhanced stability improves the robustness of the detection system.

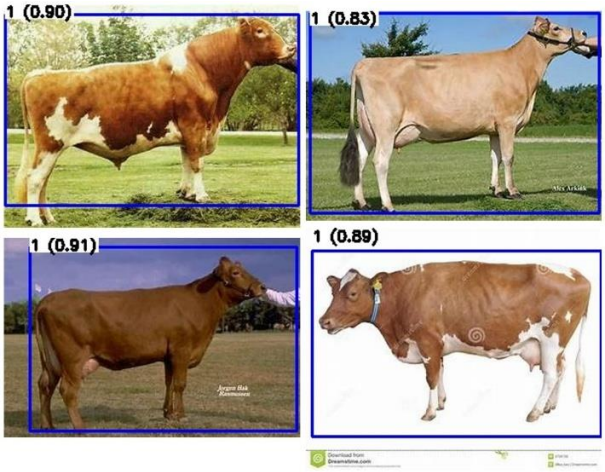
As a result, it is shown that stability can be increased by using skeletal estimation techniques, which we believe may be one way to address the problem of a lack of data. Therefore, it can be concluded that such techniques are highly effective for animal detection tasks.

In a qualitative analysis of the proposed model's performance, among the three categories, bears were classified with high accuracy, as indicated in Table 2. Since the ultimate goal of this study is to develop a method for detecting harmful animals, this result can be considered favorable. However, a notable concern is the low evaluation scores for cows. An analysis of the detection output revealed that cows were frequently misclassified as horses. Furthermore, both the conventional method and the skeleton estimation-based approach exhibited a tendency to do this. This outcome can be attributed to the close similarity in the appearance and skeletal structures of cows and horses.

An analysis was conducted to identify the characteristics of the images that were prone to misclassification. As shown in Figure 6a, images featuring a sagging neck region and a lateral extension of the neck were often classified correctly. In contrast, images without a sagging neck, as observed in Figure 6b, were frequently misclassified. This tendency can be attributed to the fact that horses generally exhibit less neck sagging than cows.



(a) When the detection is correct



(b) When the detection is incorrect

Figure 6. Qualitative analysis of detection

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

5. Conclusion

In this paper, we proposed a new method for animal detection that combines conventional algorithms for image recognition with skeleton estimation techniques. The objective of the proposed method is to improve detection accuracy by leveraging inference results from conventional detection algorithms and supplementing them with results derived from skeleton estimation techniques.

For this study, we focused on detecting three animal classes: cow, horse, and bear. Among the datasets containing these species, we utilized a high-quality annotated dataset extracted from AP-10K for training. To evaluate the model's performance, we employed a dataset extracted from the ANIMAL-POSE DATASET. The proposed method consists of a combination of three algorithms: an image recognition algorithm, a skeleton estimation algorithm, and a skeleton image recognition algorithm. The models used in the proposed method include transfer learning applied to pre-trained YOLOv8 and HRNet-w48 models. In generating skeleton images, we devised specific procedures for skeleton estimates outputs and threshold settings. Furthermore, during the integration of inference results from input images and skeleton images, we adjusted the weighting, threshold settings, and output methods throughout the integration process.

The proposed method and the conventional detection method were compared by varying the minimum confidence threshold to assess the performance differences. For the conventional detection method, a model trained via transfer learning on a pre-trained YOLO model using a dataset extracted from AP-10K was employed. The comparison results revealed that while the conventional detection method exhibited a significant performance gap between the nonoptimized and optimized parameter settings, the proposed method demonstrated a smaller performance difference between these two settings, indicating that it maintained relatively high performance, even without parameter optimization. In other words, the proposed method outperformed the conventional detection method in terms of performance stability. As a result, the proposed method can be regarded as a more user-friendly model, as it requires less parameter tuning than the conventional detection method.

The proposed method utilizes skeletal information for animal detection. Accordingly, experiments were conducted on animals exhibiting similar skeletal structures. As a result, the effectiveness of the proposed method was demonstrated for animals with similar skeletal features. This finding indicates that the method is likely to be applicable to animals with dissimilar skeletal structures, although such animals were not included in the dataset used in this study. Additionally, although the dataset primarily contained images of animals in a lateral standing position, it also included images of animals lying down, sitting, and running. Since the method successfully classifies these postures, it is anticipated that this approach can adapt to various animal postures.

Finally, we address the challenges and future perspectives. This study aimed to evaluate whether a combination of object detection and skeleton estimation techniques improves the performance of harmful-animal detection. To achieve this, YOLO, a widely recognized object detection model, and HRNet, a widely used skeleton estimation model, were employed. Future research should investigate which combination of object detection and skeleton estimation models achieve optimal performance for practical implementation. In particular, it is essential to evaluate the performance of transformer-based object detection models when combined with skeleton estimation models. Furthermore, while the proposed method integrates an object detection model with a skeleton estimation model, comparing its performance with methods that combine two object detection models could be an interesting direction for future research.

Ensuring the practical applicability of the proposed method requires addressing several challenges. One key consideration is the impact of low-quality images. In this study, training was conducted using high-quality images from AP-10K, which may have introduced bias. Consequently, when applied to real-world scenarios, preprocessing techniques such as super-resolution may be required to enhance image quality.

In addition, while this study focused solely on image-based systems, real-world applications of animal detection algorithms require compatibility with video processing and real-time inference. Although this study did not focus on real-time performance, there are situations in which real-time capabilities are necessary for practical application in the real world. Therefore, as a potential future direction, pursuing real-time performance should be considered.

Furthermore, although this research utilized 2D images for training, we believe that incorporating 3D images could further enhance the performance of skeleton estimation. Future research should also aim to broaden the scope of the animal species included in the dataset. Ultimately, the goal is to advance this system into a more practical and deployable solution for real-world applications.

5. Author Contributions

S. B. designed the study, developed the methodology, implemented the system, analyzed the data, evaluated the results, and wrote the original draft. N. M. supervised the research and managed the project. All authors revised the manuscript, approved the final version for publication, and agree to be accountable for all aspects of the work. They will ensure that any questions related to its accuracy or integrity are appropriately investigated and resolved.

6. References

Cao, J., Tang, H., Fang, H. S., Shen, X., Lu, C., & Tai, Y. W. (2019). Cross-domain adaptation for animal pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9498–9507).

Chen, K., Yu, Y., & Zhou, H. (2023). Animal movement prediction based on the corroboration of object detection and pose estimation. *Highlights in Science, Engineering and*

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

*Technol*ogy, 39, 997–1004. https://doi.org/10.54097/hset.v39i.6692

Dave, B., Mori, M., Bathani, A., & Goel, P. (2023). Wild animal detection using YOLOv8. *Procedia Computer Science*, *230*, 100–111.

Jiang, L., Lee, C., Teotia, D., & Ostadabbas, S. (2022). Animal pose estimation: A closer look at the state-of-the-art, existing gaps and opportunities. *Computer Vision and Image Understanding*, *222*, 103483.

Jocher, G., Qiu, J., & Chaurasia, A. (2023). *Ultralytics YOLO* (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

König, H. J., Ceaușu, S., Reed, M., Kendall, H., Hemminger, K., Reinke, H., Ostermann-Miyashita, E.-F., Wenz, E., Eufemia, L., Hermanns, T., Klose, M., Spyra, M., Kuemmerle, T., & Ford, A. T. (2021). Integrated framework for stakeholder participation: Methods and tools for identifying and addressing human–wildlife conflicts. *Conservation Science and Practice*, *3*(3), e399.

Kumar, T., Brennan, R., Mileo, A., & Bendechache, M. (2024). Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access, 12,* 187536–187571.

MMPose Contributors. (2020). *OpenMMLab pose estimation toolbox and benchmark*. https://github.com/open-mmlab/mmpose

Mu, J., Qiu, W., Hager, G. D., & Yuille, A. L. (2020). Learning from synthetic animals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12386–12395).

Xin, L. Q., Mat Ghani, S. N. A., Rasudin, N. S., Abdul Ghafar, N., Abd Rashid, N. H., Ten, D. C. Y., Saaban, S., Edinur, H. A., & Abdullah, M. T. (2024). Characterisation of human–wildlife conflict and casualties caused by wildlife attacks in Peninsular Malaysia. *IIUM Medical Journal Malaysia*, *23*(01), 83–90.

Yu, H., Xu, Y., Zhang, J., Zhao, W., Guan, Z., & Tao, D. (2021). AP-10k: A benchmark for animal pose estimation in the wild. *arXiv preprint arXiv:2108.12617*.

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

7. Appendix

Table A. Evaluation Results for Determining the Weighting Ratio During Integration

Ratio	Class	Precision	Recall	F1 score	Macro precision	Macro recall	Macro F1 score
0.5_0.5	Cow	0.71	0.74	0.72	0.76	0.80	0.78
	Horse	0.76	0.83	0.79			
	Bear	0.82	0.83	0.82			
0.6_0.4	Cow	0.68	0.82	0.75	0.79	0.83	0.80
	Horse	0.76	0.81	0.78			
	Bear	0.92	0.84	0.88			
0.7_0.3	Cow	0.71	0.82	0.76	0.79	0.83	0.80
	Horse	0.75	0.81	0.78			
	Bear	0.90	0.84	0.87			
0.8_0.2	Cow	0.70	0.86	0.77	0.78	0.84	0.80
	Horse	0.76	0.81	0.78			
	Bear	0.87	0.84	0.86			
0.9_0.1	Cow	0.68	0.86	0.76	0.77	0.84	0.80
	Horse	0.75	0.81	0.78			
	Bear	0.87	0.84	0.86			

DOI: https://doi.org/10.22452/mjs.vol45sp2.3
Malaysian Journal of Science 45 (Special Issue 2): 25-36 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026) https://mjs.um.edu.my

Deconvolving Neuronal Ensembles from Macroscopic Neural Activities in Voltage-Sensitive Dye Image Streams on Mouse Cerebrum Slices

Karin Horiuchi[1a], Santa Fukuda[2b], Masahiro Ohtani[3a] and Yuki Hayashida[4b]*

Abstract: Previous clinical trials have demonstrated that electrical microstimulation delivered within the primary visual cortex (V1) in human subjects with acquired blindness can evoke focal light sensations in their visual fields, supporting the feasibility of intracortical visual prostheses. Our previous animal studies using voltage-sensitive dye (VSD) imaging examined the spatiotemporal patterns of cortical neural activity in response to electrical current-pulse stimuli, varying stimulus parameters to optimize stimulation design. These studies revealed specific relationships between stimulus parameters and neural responses at millisecond temporal resolution and submillimeter spatial scales. However, as VSD images typically represent voltage changes across all cytoplasmic membranes of cells responding to stimuli, activities of individual neuron populations and their functional connections are obscured by the macroscopic responses. This study aims to investigate stimulus-induced neuronal activities at a finer spatial resolution by employing image processing techniques. Specifically, spatial and temporal cross-correlation analyses were applied to VSD image streams to identify spatially distinguishable patterns of pixel clusters exhibiting temporally correlated VSD signals. In certain cases, these neuronal ensembles, each smaller than 100 micrometers in diameter, displayed time-fixed correlation in membrane voltage change, suggesting functional connectivity between them. The image processing method demonstrates the potential of separating distinct activities of cortical neuronal ensembles from macroscopic VSD imaging data, offering a valuable tool for detailed studies of stimulus-induced network dynamics and, in turn, optimizing stimuli for visual prostheses.

Keywords: Visual prosthesis, Neuronal ensemble, Voltage-sensitive dye imaging, Spatio-temporal cross-correlation analyses

1. Introduction

Previous clinical trials have demonstrated that electrical stimulation of the primary visual cortex (V1) using tissue-penetrating microelectrodes evokes light perception, termed phosphenes, in the visual field of blind patients (Schmidt et al., 1996; Fernández et al., 2021). This phenomenon highlights the potential of cortical visual prostheses, which allow blind individuals to perceive spots of light. Building on these findings, recent years have witnessed progress in both fundamental research and prosthetic devices (Chen et al., 2020; Hayashida et al., 2022). Since the functionality of intracortical prostheses depends on precisely designed and controlled stimulation, a comprehensive understanding of the physiological basis of neuronal activity responding to these stimuli is essential.

Physiological experiments in rodents, both in vivo and with brain slice preparations, have offered valuable insights into the spatiotemporal characteristics of neural responses to microstimulation in V1. Using the voltage-sensitive dye (VSD) imaging technique, it has been demonstrated that microstimulation pulses can elicit action potentials in neural populations within milliseconds in both sighted and blind animals (Fehérvári et al., 2015; Miyamoto et al., 2017). These population action potentials subsequently trigger focal activation of cortical neural circuits, involving competitive dynamic interactions between excitatory and inhibitory signaling (Tanaka et al., 2019). Moreover, these studies have revealed specific relationships between stimulus parameters and the resulting neural responses (Miyamoto et al., 2017; Tanaka et al., 2019; Hayashida et al., 2022), and they contribute to explaining, for example, the suppressive effects observed in clinical trials when high-frequency pulse-train stimulation was applied (Hayashida et al., 2014; Nomoto et al., 2017; Fukuda et al., 2022).

While VSD imaging in the slice preparation has provided critical insights into neural response dynamics with millisecond temporal resolution and submillimeter spatial scales (Tanaka et al., 2019), it reflects voltage changes across all cytoplasmic membranes of cells responding to stimuli. Consequently, it obscures the activities of individual neuron populations or neuronal elements, leaving many details to be clarified at finer spatial resolutions. In this study, we aim to investigate neuronal responses at microscopic resolution, on the scale of tens of micrometers, by applying image analysis techniques. This approach aims to enhance our understanding of the functional connectivity between neuronal elements in response to microstimulation in V1.

Authors information:
[a]Dept. of Electrical Engineering, National Institute of Technology Nara College, Nara, Japan, E-mail: ae1175@nara.kosen-ac.jp[1]
E-mail: ohtani@elec.nara-k.ac.jp[3]
[b]Dept. Information Engineering, Grad. Engineering, Mie University, Mie, Japan,
E-mail: 423d054@m.mie-u.ac.jp[2]

*Corresponding Author: hayashida@info.mie-u.ac.jp[4]

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

2. Methods

2.1 Physiological Experiment with VSD Imaging

The Animal Experiment Committee of Mie University approved animal care and experimental procedures in this study. The method of VSD imaging was similar to that used in our previous studies (Tanaka et al., 2019; Fukuda et al., 2021). In brief, cerebral tissue slices (~0.3-mm thick, containing V1) were prepared from a 4-week-old C57BL/6J mouse and stained with VSD, NK3630 (Nippon Kankoh-Shikiso Kenkyusho). A glass micropipette electrode (~6-µm tip diameter, ~1-MΩ tip resistance) was inserted in layer II/III of V1, and a return electrode (Ag/AgCl-pellet) was placed in the recording chamber. A train of cathodic-first biphasic current pulses (20 µA/phase, 0.2 msec/phase, repetition rate of 50 or 200 Hz) was delivered to layer II/III via the electrode. The transmitted light images of a rectangular region along the mediolateral axis within layer II/III (154 × 820 µm on the slice) were captured at 1000 frames per second with 12 × 64 binned pixels using an Electron Multiplying Charge-Coupled Device camera (iXon 3 DU-897, Andor Technology).

2.2 VSD Imaging Data and Image Processing

Data analysis and image processing were carried out using MATLAB (ver. R2024a, MathWorks). The pre-processing of VSD imaging data followed procedures similar to those in our previous study (Tanaka et al., 2019). After averaging across multiple trials of the recording, each pixel value in the image was converted to the relative change in the transmission light intensity, denoted as $\Delta T(t_n)/T_0$ at time t_n . A decrease and an increase in $\Delta T(t_n)/T_0$ correspond to the membrane hyperpolarization and depolarization, respectively. To decrease noise, the converted images were subjected to spatial filtering (Gaussian, 3×3-pixel mask).

Figure 1 presents the workflow for the image processing steps described in this study. The initial step involves subtracting the accumulating component of neural response to the pulse-train stimulus (Fukuda et al., 2021). This step is crucial for isolating local neuronal activities that would otherwise be masked by the global response to the stimulus pulse, and for facilitating the deconvolution of underlying networks in the following analyses. The time course of $\Delta T(t_n)/T_0$ for each image pixel is processed using a 6th-order Bessel low-pass temporal filter with a cutoff frequency of 20 Hz. After the group delay introduced by the filtering is compensated to ensure accurate temporal alignment, the filtered time course is subtracted from the original $\Delta T(t_n)/T_0$ time course. This provides a processed time series of $\Delta T(t_n)/T_0$ images that preserves local neuronal responses at a finer spatial scale.

The second step involved manually selecting spatially clustered pixels that are visually distinguishable from the surrounding pixels in the subtracted time-series images generated by the initial step. For each of the pixel clusters, the processed $\Delta T(t_n)/T_0$ time courses within a 3×3-pixel area were averaged to obtain the corresponding local response time course.

In the third step, the time points during which the local response was either spatially distinct or reached local extrema in time were identified for each pixel cluster. The image frames at these time points were then averaged, producing an image map that highlights specific pixel patterns whose activity correlates with that of the given pixel cluster. This resulting image is referred to as the "spatial correlation map."

In the fourth step, the temporal cross-correlation function between the selected pixel cluster and specific pixels in the spatial correlation map is computed using the local response time courses at these two locations (denoted as $x(t)$ and $y(t)$)). The cross-correlation function $R(\Delta t)$ is defined as:

$$R(\Delta t) = \int_{T_{min}}^{T_{max}} x(t) \cdot y(+\Delta t)\ dt \qquad (1)$$

where Δt represents the time lag between the signals in two locations, T_{min} and T_{max} denote the start and end time points, respectively, of the duration over which the cross-correlation is calculated. The cross-correlation function was then filtered using a 6th-order Bessel low-pass filter with a cutoff frequency of 400 Hz, and the group delay introduced by the filtering was compensated. As an alternative to manually selecting a specific pixel cluster in the second step, if the signal-to-noise ratio in the time course of each pixel is sufficiently high, then all pixels can be included in the subsequent analyses. In this case, temporal cross-correlation functions were first computed between a given reference pixel and every other pixel. Subsequently, the peak amplitudes of these functions were mapped onto a two-

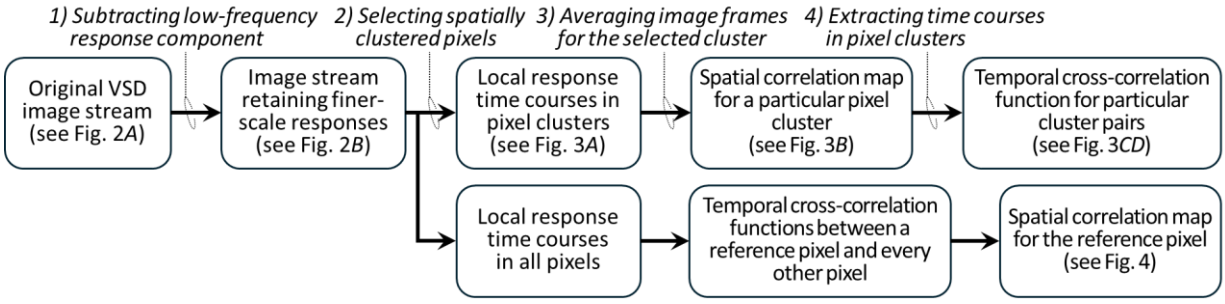


Figure 1. Image processing workflows. The diagram illustrates the four sequential steps.

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

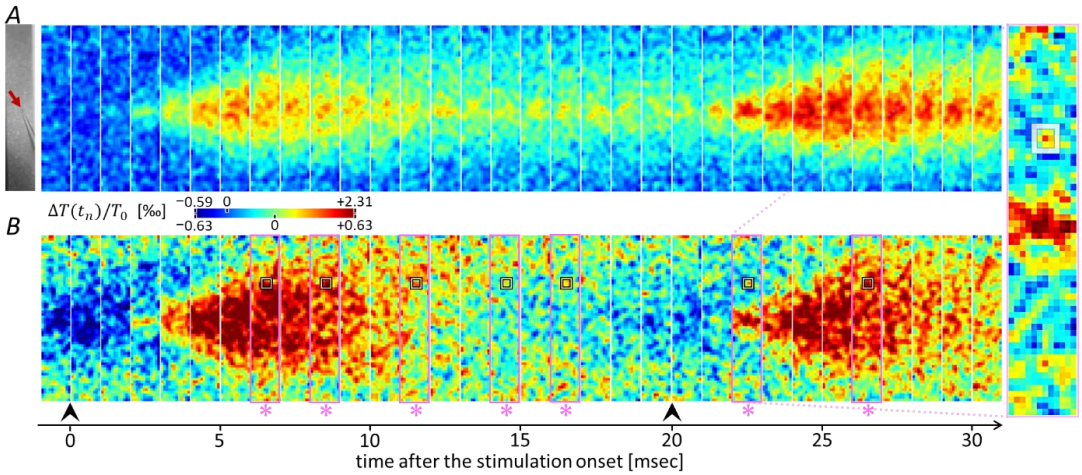


Figure 2. Unmasking local neuronal activities in VSD imaging. *A*) A near-infrared microphotograph of the tissue in the imaging field (left-most) and time-series VSD images (32 image frames) recorded from the tissue in response to stimulus current pulses delivered from the electrode tip inserted in the tissue (indicated by an arrow) at time 0 and 20 msec after the stimulation onset (indicated by arrowheads). *B*) Time-series images generated by subtraction of the low-frequency component from the original images shown in *A*. Asterisks indicate the image frames used for the averaging process. An enlarged view of one of those frames is shown on the right-most. W*hite* squares indicate an example of selected pixel clusters. The color codes differ between *A* and *B*.

dimensional spatial layout corresponding to the pixel locations, resulting in spatial correlation maps for the given pixel (see *lower* workflow in Figure 1).

3. Results

3.1 Local VSD Responses Unmasked from Global Response

Figure 2*A* shows an example of VSD images of the tissue responding to a train of microstimulation pulses (50-Hz repetition rate and 20-μA/phase current amplitude). The leftmost image displays a transmitted image of a portion of the layer II/III of V1, referred to as the imaging field, with the stimulating electrode tip inserted into the tissue at the center of the field (indicated by an arrow). The remaining frames show time-series VSD images in the imaging field using a color-coded scale. Consistent with our previous studies (Tanaka et al., 2019; Fukuda et al., 2021), an increase in $\Delta T(t_n)/T_0$ was observed near the electrode tip within one to two milliseconds after the onset of the stimulus pulse (indicated by an arrowhead at the 0-msec time point), representing action potentials elicited in a neural population (Tanaka et al., 2019). Subsequently, the excitation became more pronounced in amplitude and spatially propagated toward the periphery over the next several milliseconds. This represents synaptic propagations of neural excitation (Tanaka et al., 2019; Fukuda et al., 2021). The response then gradually decayed in both amplitude and spatial extent, reflecting synaptic inhibition of neural activity [6].

In response to the second stimulus pulse (indicated by an arrowhead at the 20-msec time point), $\Delta T(t_n)/T_0$ increased to higher amplitudes than those elicited by the first pulse due to the accumulation of synaptic and non-synaptic responses (Fukuda et al., 2021). These global response characteristics were consistently observed across various tissue slice samples we used in the present and previous studies. On the other hand, the finer-scale spatial patterns of the responses appeared to vary among samples. To visually enhance such local response patterns, a low-frequency response component was subtracted from the original image sequence, resulting in time-series images shown in Figure 2*B*. Compared to the original images (Figure 2*A*), each of these image frames highlights patchy response patterns, with some spatially clustered pixels becoming discernible from the surrounding areas (indicated by white squares as examples). Since their locations were similar across multiple image frames (indicated by asterisks), each pixel cluster was presumed to represent distinct neuronal elements rather than speckle noise.

3.2 Presumed Neuronal Elements Deconvolved in Correlation Map

Figure 3*A-a* shows the "local response time course" (see Figure 1) for the example pixel cluster highlighted in Figure 2*B*. As shown in Figure 3*B-a*, this pixel cluster (indicated by a *green* square) was located about 150 μm from the stimulating electrode tip. Figure 3*B-b* displays the "spatial correlation map" for this pixel cluster (see Figure 1), highlighting a specific pattern of pixels considered to be functionally correlated with the selected pixel cluster. Figure 3*B-c* offers a three-dimensional view of this map, showing another peak with a pixel value comparable to that of the original cluster (indicated by a *grey* arrow, corresponding to the *grey* square in sub-panel *b*). The location of this secondary peak is near the stimulating electrode tip, as shown in subpanel *a* (*grey* square). Figure 3*A-b* represents the local response time course at the location of this secondary peak. The inset depicts a superimposition of the two response time courses at an expanded time scale, confirming repetitive membrane excitations in response to the current pulses at both locations. The stimulus pulses were delivered for a 400-msec duration (indicated by

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

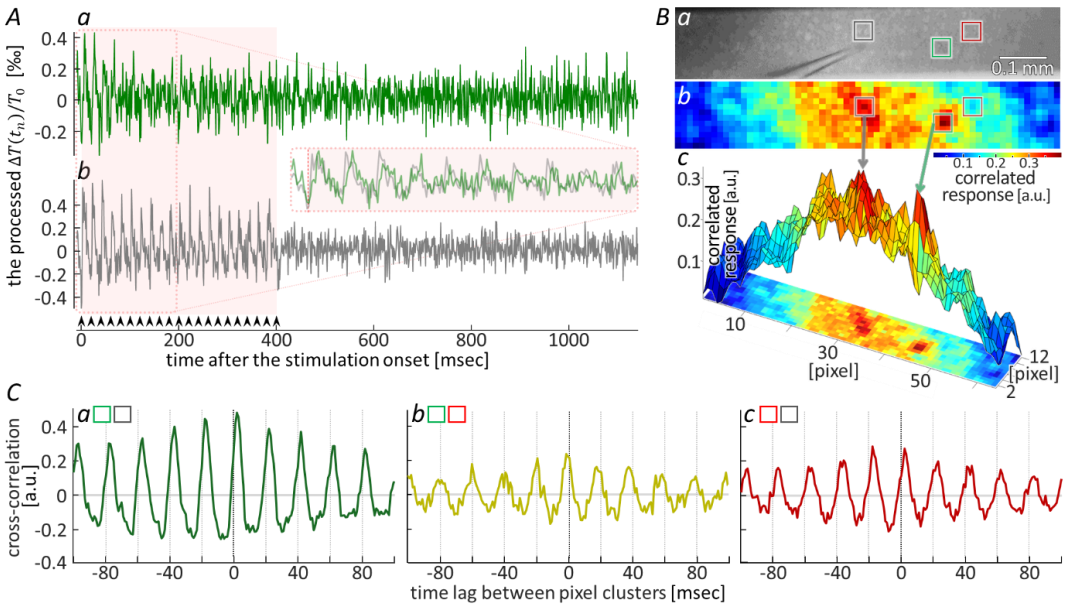


Figure 3. Deconvolved activities of neuronal elements. *A*) Local response time courses for the pixel clusters indicated by the *green* and *black* squares in *B-ab*. Inset: Superimposition of time-course segments encircled by the *dotted* rectangle. *B*) Spatial correlation map for the particular pixel cluster indicated by the *green* square in *B-ab*. *C*) Temporal cross-correlation functions for three different pairs of 3×3-pixel areas: pair of *green* and *grey* squares in *B-ab* (*a*), pair of *green* and *red* squares (*b*), and pair of *green* and *grey* squares (*c*).

arrowheads in panel *A*), and the subsequent analyses were conducted over this time period.

Figure 3*C-a* shows the temporal cross-correlation function between the two response time courses, demonstrating temporally correlated 50-Hz activity with a 2-msec time lag between the responses. For comparison, a temporal cross-correlation function was also calculated between the selected pixel cluster (*green* square in panel *B-ab*) and a nearby pixel region (*red* square in panel *B-ab*) that showed relatively low correlation with the selected cluster, despite its spatial proximity (Figure 3*C-b*). A cross-correlation function was calculated between this example pixel region and the pixel cluster near the stimulating electrode tip (*black* square in panel *B-ab*). In both cases, the cross-correlation amplitude was significantly lower than that observed for the pair of pixel clusters exhibiting prominent peaks on the spatial correlation map (i.e., the clusters indicated by *green* and *black* squares). This result supports the interpretation that the specific pattern observed in the spatial correlation map reflects the functional coupling of neuronal elements associated with the selected pixel cluster.

In the results discussed above, a specific pixel cluster was manually selected to ensure, through visual inspection, the presence of distinct VSD signals clearly contrasted with the surrounding areas (e.g., *white* squares in Figure 2). This step may be required in certain situations, such as when the signal-to-noise ratio is insufficient and/or spatially inhomogeneous due to uneven staining of the tissue with the dye. However, such manual selection could introduce subjective bias into the subsequent analyses. Therefore, we additionally explored an alternative image processing workflow, as illustrated in Figure 1. In this workflow, spatial cross-correlation maps were computed for every pixel in the image frame, excluding those located along the image boundaries, resulting in a total of 620 maps (corresponding to 10 × 62 pixels). Among these maps, specific patterns of correspondence were observed across multiple maps generated from different reference pixels.

Figure 4 depicts two representative examples of such spatial cross-correlation maps, with the reference pixels indicated by *white* asterisks. Both maps display patchy clusters of pixels exhibiting relatively high correlation values, including areas surrounding the reference pixel. Although the majority of these clusters are located near the stimulation site in each map, one cluster appears at a distance from the reference pixel, indicated by an arrow. Notably, the position of this distant cluster relative to the reference pixel in one map corresponds to the position of the reference pixel relative to the distant cluster in the other map. Such a correspondence across different maps may reflect shared or overlapping circuit dynamics between spatially separated regions.

4. Discussion

Following the proposed method (Figure 1), local neuronal activities evoked by the stimulus, previously obscured by the global neural response, were visually highlighted in the time-series images (Figure 2). This enabled the extraction of both spatial correlation maps and temporal cross-correlation functions from selected pixel regions (Figure 3). Additionally, an alternative

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

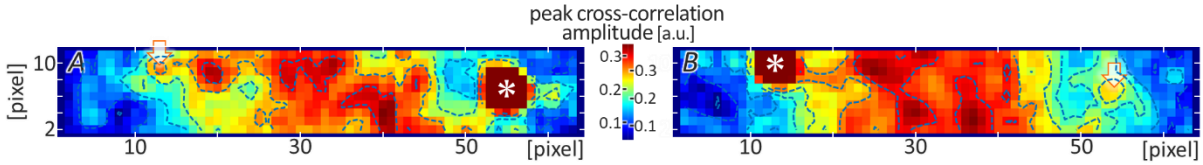


Figure 4. Spatial cross-correlation maps computed using different reference pixels. Each *white* asterisk in panels *A* and *B* indicates the reference pixel used to compute the corresponding map. Contour lines are overlaid on the color maps to highlight spatial patterns. Color bars indicate the peak amplitude of the temporal cross-correlation. Both the horizontal and vertical axes represent pixel indices.

image processing workflow was introduced to generate spatial cross-correlation maps, using any pixel in the imaging field as a reference (Figure 4). In each map, spatially discernible patchy structures, each less than 100 micrometers in diameter, were displayed.

We focused exclusively on image frames that captured neural activity during pulse-train stimulation. Baseline and post-stimulus activities, as well as responses to other types of stimulation, were not examined. Consequently, the present approach does not account for intrinsic functional coupling across different neural conditions. Moreover, the current method depends on manual comparisons of the generated spatial correlation maps (Figure 4). Therefore, further development is necessary to establish an efficient strategy for grouping similar correlation maps, such as through statistical segmentation or other suitable techniques, and for subsequently performing automatic deconvolution of neuronal components.

The validation of the proposed method is primarily qualitative. Incorporating comparisons with ground-truth neuronal recordings would increase the credibility of our approach. However, since VSD imaging captures voltage changes across all cytoplasmic membranes within the three-dimensional volume of the imaging field, obtaining direct neuronal recordings would require the placement of electrodes throughout the entire field of view, which is not practically feasible. Recent studies have introduced deconvolution methods for VSD imaging data (e.g., Benisty et al., 2022). However, these approaches have mainly focused on extracting action potentials from individual neurons rather than on revealing functional connectivity among neuronal elements. Furthermore, some of these methods assume that individual somata are already visible in the imaging data (e.g., Xie et al., 2021), which restricts their applicability to our experimental conditions. An alternative approach that may be feasible in our case is the use of confocal imaging with Ca^{2+}-sensitive dyes (cf. Tanaka et al., 2019), which can provide three-dimensional localization of neuronal somata. Further investigation is justified in this direction.

5. Conclusion

To the best of our knowledge, this study is the first to attempt the deconvolution of functional connectivity between neuronal ensembles responding to microstimulation, based on macroscopic VSD imaging data, through spatio-temporal cross-correlation analyses. The present findings suggest the feasibility of the proposed method as a tool for dissecting cortical network dynamics in response to microstimulation and, ultimately, for optimizing stimulation protocols for visual prostheses.

6. Acknowledgements

This research was supported in part by Grant-in-Aid for Scientific Research from MEXT, Japan (22K12781, 25K15892 to Y.H.), and JST SPRING, Japan (JPMJSP2137 to Mie Univ./S.F.).

7. References

Benisty H., Song A., Mishne G., Charles A. S. (2022) Review of data processing of functional optical microscopy for neuroscience, *Neuro-photonics*, 9 (4): 041402 (32 pp.).

Chen X., Wang F., Fernández E., Roelfsema P. R. (2020) Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex, *Science* 370 (6521): 1191-1196.

Fehérvári T. D., Okazaki Y., Sawai H., Yagi T. (2015) In vivo voltage-sensitive dye study of lateral spreading of cortical activity in mouse primary visual cortex induced by a current impulse, *PLoS One* 10 (7): e0133853 (24 pp.).

Fernández E., Alfaro A., Soto-Sánchez C., Gonzalez-Lopez P., Lozano A. M., Peña S., Grima M.D., Rodil A., Gómez B., Chen X., Roelfsema P. R., Rolston J. D., Davis T. S., Normann R. A. (2021) Visual percepts evoked with an intracortical 96-channel microelec-trode array inserted in human occipital cortex, *J. Clin. Invest.* 131 (23): e151331 (15 pp.).

Fukuda S., Nomoto T., Yoshida K., Hayashida Y., Yagi T. (2021) Accumulation of non-synaptic and synaptic excitations induced by microstimulation pulses in the mouse visual cortex layer II/III, *2021 43rd Ann. Int'l Conf. I.E.E.E. Med. Biol. Soc.*: 5152.

Fukuda S., Yoshida K., Nomoto T., Yagi T., Hayashida Y. (2022) Use-dependent inactivation of the mouse visual cortex in response to microstimulation train, *2022 E-Health Bioeng. Conf.*: 1-4.

Hayashida Y., Takeuchi K., Ishikawa N., Okazaki Y., Fehérvári T. D., Tanaka H., Yagi T. (2014) Voltage-sensitive dye imaging of the visual cortices responding to electrical pulses at different intervals in mice in vivo, 2014 36th *Ann. Int'l Conf. I.E.E.E. Med. Biol. Soc.*: 402-405.

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

Hayashida Y., Kameda S., Umehira Y., Ishikawa S., Yagi T. (2022) Multichannel stimulation module as a tool for animal studies on cortical neural pros-theses, *Front. Med. Technol.* 4: 927851 (21 pp.).

Miyamoto S., Suematsu N., Umehira Y., Hayashida Y., Yagi T. (2017) Age-related changes in the spatiotemporal responses to electrical stimulation in the visual cortex of rats with progressive vision loss, *Sci. Rep.* 7(1): 14165 (13 pp.).

Nomoto T., Tanaka Y., Hayashida Y., Yagi T. (2017) Imaging of population spikes induced by repetitive stimulus pulses in mouse cerebral slices in vitro, *2017 39th Ann. Int'l Conf. I.E.E.E. Med. Biol. Soc.*: 1109.

Schmidt E. M., Bak M. J., F T Hambrecht, C V Kufta, D K O'Rourke, P Vallabhanath (1996) Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex, *Brain* 119 (2): 507-522.

Tanaka Y., Nomoto T., Shiki T., Sakata Y., Shimada Y., Hayashida Y., Yagi T. (2019) Focal activation of neuronal circuits induced by microstimulation in the visual cortex, *J. Neural Eng.* 16 (3): 036007 (17 pp.).

Xie M. E., Adam Y., Fan L. Z., Böhm U. L., Kinsella I., Zhou D., Rozsa M., Singh A., Svoboda K., Paninski L., Cohen A. E. (2021) High-fidelity estimates of spikes and subthreshold waveforms from 1-photon voltage imaging in vivo, *Cell Reports* 35(1): 108954 (15 pp.).

DOI: https://doi.org/10.22452/mjs.vol45sp2.4
Malaysian Journal of Science 45 (Special Issue 2): 37-42 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special issue

Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

https://mjs.um.edu.my

Determining the Mixed Metric Dimension of a Graph with the Binary Gray Wolf Algorithm

Hendy[1ab], Mohammad Isa Irawan[2a*], Imam Mukhlash[3a], and Rinurwati[4a]

Abstract: Many computational techniques for extracting insights from data depend on embedding data within a Euclidean space. However, symbolic datasets, particularly biological sequence datasets resulting from high-throughput sequencing processes, often pose challenges for conventional embedding methods, as they tend to yield either excessively high-dimensional or overly simplified representations, thereby limiting their analytical applicability. To address these constraints, the concept of metric dimension (MD) in graph theory provides a rigorous and practical framework for representing points in the Euclidean plane based on their relative distances. In this study, we explore a hybrid concept that combines two distance parameters: MD and edge metric dimension (EMD). The mixed metric dimension (MMD) uniquely characterizes a graph's elements (edges and vertices) based on their distance vectors to a designated set of vertices. We developed a computational intelligence approach to determine the mixed metric dimension of graphs: the Binary Gray Wolf Optimizer (BGWO) algorithm. We integrate a distance-based matrix representation with population-based optimization methods. To improve solution performance, several modifications have been incorporated, including the validation of row uniqueness, the implementation of dynamic stagnation handling techniques, and the adoption of an elitism strategy. Our findings indicate that the BGWO achieves efficient completion times and delivers competitive performance for this problem.

Keywords: Binary gray wolf optimizer algorithm, edge metric dimension (EMD), mixed metric dimension (MMD)

1. Introduction

Unless otherwise stated, all graphs in this study are connected and simple. The distance between two vertices u and v in a graph G is represented as $d_G(u,v)$. A vertex $s \in V(G)$ is said to differentiate (or resolve) two vertices u and v if $d_G(u,s) \neq d_G(v,s)$. A subset $S \subseteq V(G)$, is referred to as a metric generator for G if each pair of vertices u and v in G is differentiated by at least one vertex in S. The smallest cardinality of such a subset is called the MD of G, denoted by $\dim(G)$. The concept originated independently through the contributions of Harrary and Melter (1976) and Slater (1975), who referred to it as resolving sets and locating sets, respectively.

The MD of graphs has a wide range of applications, including mathematical biology (Tillquist & Lladser, 2019). Collections of symbolic representations, especially sequences of biological macromolecules generated by advanced genomic sequencing technologies, often present challenges for conventional embedding methods. These techniques often result in excessively high-dimensional or overly simplistic representations, thereby reducing their usefulness for analytical purposes. To overcome these challenges, the graph-theoretical notion of the MD provides a rigorous and efficient framework for representing data points within the Euclidean plane. Furthermore, the concept of metric dimension is also used in various applications beyond data representation, including network topology analysis (Beerliova et al., 2006), sensor placement optimization (Chartrand et al., 2000), structural graph characterization and pharmaceutical chemistry (Imran et al., 2022), methods for processing patterns and images (Melter & Tomescu, 1984), autonomous robot movement (Khuller et al., 1996), and combinatorial optimization (Seb & Tannier, 2016).

1.1. Edge metric dimension

The metric dimension concept has recently evolved from focusing on vertex identification to edge identification in a graph (Kelenc, Tratnik, & Yero, 2018). For a graph $H = (V, E)$, a vertex $v \in V$ and an edge $e = uw \in E$, the distance between the vertex v and the edge e is defined as $d_G(e,v) = \min\{d_G(u,v), d_G(w,v)\}$. A vertex $x \in V$ distinguishes two edges $e_1, e_2 \in E$ if $d_G(x,e_1) \neq d_G(x,e_2)$. A set S of vertices is an edge metric generator if every pair of distinct edges in G is differentiated through at least one vertex in S. The minimum cardinality of an edge metric generator in G is known as the edge metric dimension, denoted by $dim_e(G)$. An edge metric basis of G is an edge metric generator with a cardinality of exactly $dim_e(G)$. For instance, consider a graph G, as illustrated in Figure 1. $Dim(H) = 2$ with the resolving set $S_1 = \{a_1, a_2\}$. The representations of all vertices: $r(a_1|S_1) = (0,1)$, $r(a_2|S_1) = (1,0)$, $r(a_3|S_1) = (2,1)$, $r(a_4|S_1) = (2,2)$, $r(a_5|S_1) = (1,2)$ are all different. $Dim_e(H) = 3$ with the edge metric generator of G is $S_2 = \{a_1, a_4, a_5\}$. The depictions of all edges: $r(a_1a_2|S_2) = (0,1,1)$, $r(a_1a_5|S_2) = (0,1,0)$, $r(a_2a_3|S_2) = (1,1,1)$, $r(a_2a_4|S_2) = (1,0,1)$, $r(a_2a_5|S_2) = (1,1,0)$, $r(a_3a_4|S_2) = (2,0,1)$, $r(a_4a_5|S_2) = (1,0,0)$ are all different.

Authors information:
[a]Dept. of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. E-mail: 7002211006@student.its.ac.id[1], mii@its.ac.id[2], imamm@matematika.its.ac.id[3], rinur@matematika.its.ac.id[4].
[b]Faculty of Engineering, Universitas Kadiri, Kediri, Indonesia. E-mail: 7002211006@student.its.ac.id[1]

*Corresponding Author: mii@its.ac.id

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Numerous studies were conducted on the EMD of various graph families and the outcomes of specific graph operations (Peterin & Yero, 2020; Zhang & Gao, 2020). Furthermore, research was also performed on the characterization of graphs with particular EMD (Zhu et al., 2019; Knor et al., 2021).

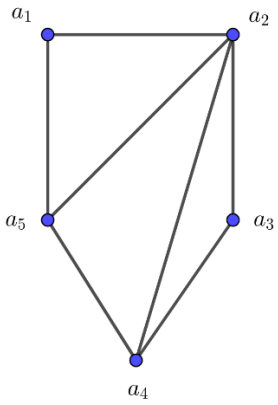


Figure 1. Graph H

1.2. Mixed metric dimension

This study focuses on a combined variation of the two previously mentioned parameters. Specifically, given a graph G, our objective is to distinguish elements of G (edges and vertices) through distance vectors to a predetermined set of vertices in G. A vertex a in a graph G is said to differentiate two elements u and v (which may be vertices or edges) if $d_G(u,a) \neq d_G(v,a)$. A set $S \subseteq G$ is defined as a mixed metric generator if any two distinct elements of G are differentiated by at least one vertex in S. The smallest size of such a set is the mixed metric dimension (MMD) of G, represented by $dim_m(G)$ (Kelenc et al. 2017).

For example, consider the graph H illustrated in Figure 1. It can be demonstrated that there is no mixed metric generator of H with cardinality < 4. The graph H has $dim_m(H) = 4$, where $S_3 = \{a_1, a_3, a_4, a_5\}$ as mixed metric generator. The representations of all elements of H: $r(a_1|S_3) = (0,2,2,1)$, $r(a_2|S_3) = (1,1,1,1)$, $r(a_3|S_3) = (2,0,1,2)$, $r(a_4|S_3) = (2,1,0,1)$, $r(a_5|S_3) = (1,2,1,0)$, $r(a_1a_2|S_3) = (0,1,1,1)$, $r(a_1a_5|S_3) = (0,2,1,0)$, $r(a_2a_3|S_3) = (1,0,1,1)$, $r(a_2a_4|S_3) = (1,1,0,1)$, $r(a_2a_5|S_3) = (1,1,1,0)$, $r(a_3a_4|S_3) = (2,0,0,1)$, $r(a_4a_5|S_3) = (1,1,0,0)$ are all different.

2. Literature review

The MMD possess several combinatorial properties. Every mixed-metric generator functions simultaneously as a metric generator and an edge metric generator. Consequently, the following relationship holds.

Remark 1. (Kelenc et al. 2017). For any graph G, $dim_m(G) \geq \max\{\dim(G), dim_e(G)\}$

Remark 2. (Kelenc et al. 2017). For any graph G of order n, $2 \leq dim_m(G) \leq n$

Several studies have investigated the MMD in different graph families and examined the effects of various graph operations (Koam et al., 2023; Danas, 2021; Abbas et al., 2024; Darmaji & Azahra, 2022; Sedlar & Škrekovski, 2021b). Furthermore, research was conducted on the characterization of graphs with a particular mixed-metric dimension (Sedlar & Škrekovski, 2021a).

Computing the MD of a graph can be modeled as an optimization problem. From a computational perspective, Diaz et al. (2017) and Hauptmann et al. (2012) demonstrated the complexity of the MDP, showing that it is NP-complete. Computational methods for the metric dimension problem have previously been explored by Kratica et al. (2009) on determining the strong MD of various graph classes using genetic algorithms. Kratica et al. (2012) formulated the problem of a strong MD as a linear integer programming model. Chen et al. (2016) designed a strongly polynomial-time exact algorithm to solve the minimum cost doubly resolving set problem for specific graph classes, while Hoffmann et al. (2016) implemented a linear-time algorithm for determining the MD of certain graph classes.

Further advancements occurred when Murdiansyah and Adiwijaya (2017) applied the particle swarm optimization algorithm in the same year to solve the MDP for hypercube graphs. Vietz et al. (2019) proposed a new approach by decomposing graphs and applying an algorithm to determine their metric dimension. Muhammad and Susilowati (2020) developed an algorithm to compute the metric dimension of a graph. Chelladurai (2020) designed an algorithm to determine the MD and basis of simple graphs. Mohamed et al. (2023) proposed an algorithm to address the dominant metric dimension problem by employing the binary Archimedes optimization algorithm.

This paper introduces a computational approach to a problem that extends the concept of metric dimension, known as the MMD. To date, no computational method has been developed to solve the MMDP. The approach adopted in this paper utilizes the binary gray wolf optimizer, a population-based metaheuristic algorithm inspired by the hunting behavior of gray wolves. Research findings show that the GWO is one of the most widely used algorithms for solving various optimization problems. A bibliometric analysis indicates that the gray wolf optimizer is the most frequently referenced metaheuristic in computer science (Hendy et al., 2023).

This paper adopts the binary variant of the Gray Wolf Optimizer, an optimization algorithm initially introduced by Mirjalili et al. (2014). This algorithm falls under the category of population-based swarm intelligence techniques. It is designed based on the hunting behavior and hierarchical structure of gray wolf (Canis lupus) packs. Gray wolves, members of the Canidae family, are classified as apex predators, indicating their dominance at the top of the food chain. These animals typically live in structured packs, with an average pack size of 5 to 12 members.

The pack's leader is called the alpha, who may be either a male or female wolf and holds the highest authority within the gray wolf hierarchy. The alpha is responsible for making key decisions. The next hierarchical level is the beta, a subordinate to the alpha who assists in decision making and performing

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

associated activities. At the lowest level is the omega, which acts as the group's scapegoat and is submissive to all dominant wolves. Wolves that do not belong to the alpha, beta, or omega groups but dominate the omega are classified as delta wolves. Although delta wolves are outranked by alpha and beta wolves, they exert authority over the omega.

The following principles are established in developing a mathematical model based on the social hierarchy of gray wolves: (1) The optimal solution, representing the optimal fitness value, is designated as α. (2) The next best solution is represented by β, while the third-highest ranked solution is denoted as δ. (3) All remaining wolves not belonging to these categories are classified as omega. (4) Within the GWO algorithm, the optimization process, analogous to the hunting process, is directed by the α, β, and δ wolves. In contrast, the ω wolves follow the guidance of these leading wolves.

A pack of gray wolves surrounds their prey during hunting. In mathematical modeling, this process is expressed by the following equation.

$$\vec{D} = \left|\vec{K}.\overrightarrow{Y_p}(i) - \vec{Y}(i)\right| \quad (1)$$

$$\vec{Y}(i+1) = \overrightarrow{Y_p}(t) - \vec{A}.\vec{D} \quad (2)$$

In this context, i represents the current iteration, $\vec{A}$ and $\vec{K}$ are coefficient vectors. $\overrightarrow{Y_p}$ denotes the coordinate vector representing the prey's position, and $\vec{Y}$ denotes the coordinate vector of the gray wolf. Vectors $\vec{A}$ and $\vec{K}$ are determined based on the following formulas:

$$\vec{A} = 2\vec{a}.\overrightarrow{n_1} - \vec{a} \quad (3)$$

$$\vec{K} = 2\,\overrightarrow{n_2} \quad (4)$$

Here, $\vec{a}$ linearly decays from an initial value of 2 to 0 across all iterations, where $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ represent randomly generated vectors within the range of [0,1].

In mathematically simulating gray wolf hunting behavior, the α (representing the most optimal potential solution), β, and δ possess better information regarding the prospective location of the prey. Therefore, the top three optimal solutions obtained are retained, while the remaining search wolves, including the omega members, update their positions with reference to where the leading wolves are located. This hunting process is modeled through equations (5) to (11) (Mirjalili et al., 2014).

$$\overrightarrow{D_\alpha} = \left|\overrightarrow{K_1}.\overrightarrow{Y_\alpha} - \vec{Y}\right| \quad (5)$$

$$\overrightarrow{D_\beta} = \left|\overrightarrow{K_2}.\overrightarrow{Y_\beta} - \vec{Y}\right| \quad (6)$$

$$\overrightarrow{D_\delta} = \left|\overrightarrow{K_3}.\overrightarrow{Y_\delta} - \vec{Y}\right| \quad (7)$$

$$\overrightarrow{Y_1} = \overrightarrow{Y_\alpha} - \overrightarrow{A_1}.\left(\overrightarrow{D_\alpha}\right) \quad (8)$$

$$\overrightarrow{Y_2} = \overrightarrow{Y_\beta} - \overrightarrow{A_2}.\left(\overrightarrow{D_\beta}\right) \quad (9)$$

$$\overrightarrow{Y_3} = \vec{Y} - \overrightarrow{A_3}.\left(\overrightarrow{D_\delta}\right) \quad (10)$$

$$\overrightarrow{Y_{(\iota+1)}} = \frac{\overrightarrow{Y_1}+\overrightarrow{Y_2}+\overrightarrow{Y_3}}{3} \quad (11)$$

3. Proposed BGWO algorithm for the mixed metric dimension problem

Modifications and refinements of the BGWO algorithm are crucial for its implementation in solving the MMDP. This necessity arises from the inherent characteristics of the solutions to the mixed metric dimension problem, which are represented as binary ordered, with not all randomly generated outputs producing feasible candidate solutions. This section offers a detailed explanation of the BGWO algorithm, adapted from its application to the feature selection problem, as described in prior work (Emary et al., 2016). This section also discusses various modification processes designed to enhance the algorithm's performance in solving the mixed metric dimension problem. The primary updating equation is presented as shown in equation (12). Refer to Algorithm 1 for further details.

$$X_i^{t+1} = Crossover(x_1, x_2, x_3) \quad (12)$$

where $Crossover(x_1, x_2, x_3)$ signifies an appropriate crossover operation involving solutions x, y, z, and x_1, x_2, x_3 are binary vectors denoting the impact of a wolf's movement towards the α, β, and δ gray wolves sequentially. The x_1, x_2, x_3 are determined using equations (13),(16), and (19) correspondingly.

$$x_1^d = \begin{cases} 1, \text{if } (x_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0, \quad \text{otherwise} \end{cases} \quad (13)$$

x_1^d represents the position vector of the alpha wolf in dimension d, while $bstep_\alpha^d$ signifies a binary step in dimension d, which can be computed as described in equation (14).

$$bstep_\alpha^d = \begin{cases} 1, \text{if } cstep_\alpha^d \geq rand \\ 0, \quad \text{otherwise} \end{cases} \quad (14)$$

$rand$ is a randomly generated number from a uniform distribution within the [0, 1] range. $cstep_\alpha^d$ reflects the continuous-valued step size for dimension d and can be computed using a sigmoidal function as outlined in equation (15).

$$cstep_\alpha^d = \frac{1}{1+e^{-10(A_1^d D_\alpha^d - 0{,}5)}} \quad (15)$$

where A_1^d and D_α^d are determined by equations (3), and (9) in the dimension d.

$$x_2^d = \begin{cases} 1, \text{if } (x_\beta^d + bstep_\beta^d) \geq 1 \\ 0, \quad \text{otherwise} \end{cases} \quad (16)$$

The vector x_β^d represents the position vector of the beta wolf in dimension d, and $bstep_\beta^d$ represents a binary step in dimension d that can be determined as in equation (17).

$$bstep_\beta^d = \begin{cases} 1, \text{if } cstep_\beta^d \geq rand \\ 0, \quad \text{otherwise} \end{cases} \quad (17)$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

where the $rand$ denotes a randomly generated number from a uniform distribution within the range ∈ [0, 1], and $cstep_\beta^d$ represents the continuous-valued step size for dimension d which can be computed using a sigmoidal function as outlined in equation (18).

$$cstep_\beta^d = \frac{1}{1+e^{-10(A_1^d D_\beta^d - 0{,}5)}} \quad (18)$$

where A_1^d and D_β^d are computed using equations (3) and (5) in the dimension d.

$$x_3^d = \begin{cases} 1, \text{if } (x_\delta^d + bstep_\delta^d) \geq 1 \\ 0, \quad \text{otherwise} \end{cases} \quad (19)$$

x_δ^d represents the position vector of the delta wolf in dimension d, while $bstep_\delta^d$ denotes a binary step in dimension d, calculated according to equation (20).

$$bstep_\delta^d = \begin{cases} 1, \text{if } cstep_\delta^d \geq rand \\ 0, \quad \text{otherwise} \end{cases} \quad (20)$$

where $rand$ is a random number drawn from the uniform distribution ∈ [0, 1], and $cstep_\delta^d$ is the continuous-valued step size for dimension d and can be calculated using a sigmoidal function as in equation (21).

$$cstep_\delta^d = \frac{1}{1+e^{-10(A_1^d D_\delta^d - 0{,}5)}} \quad (21)$$

As equation (22) demonstrates, a stochastic crossover strategy is implemented per dimension to perform the crossover operation among the solutions a, b, and c.

$$x_d = \begin{cases} a_d, & \text{if } rand < \frac{1}{3} \\ b_d, & \text{if } \frac{1}{3} \leq rand < \frac{2}{3} \\ c_d, & \text{otherwise} \end{cases} \quad (22)$$

where a_d, b_d, c_d represent the binary values for the first, second, and third parameters in dimension d, x_d denotes the crossover output at dimension d, and rand is a randomly generated number from a uniform distribution within the range [0,1].

Algorithm 1. Modified BGWO Algorithm for mixed metric dimension problem

Input: n number of vertices, Matrix C, popsize, MaxIt
Output: x_α optimal gray wolf, mixed metric dimension

1. Initalize a population of a popsize wolves' positions at random ∈ $[0,1]$
2. Find the representation of all vertices and edges with respect to wolves
3. Row uniqueness validation
4. Replace the population based on objective value
5. Find the fitness value
6. Find the α, β, and δ based on fitness while stopping criteria not met do

For each $wolf_i \epsilon\ pack$ do
Calculate x_1, x_2, x_3 using equations (13),(16), and (19)
$x_i^{t+1} \leftarrow$ crossover among x_1, x_2, x_3 using equation (22)
end (for)

I. Update a, A, C
II. Evaluate the positions of individual wolves
III. Update α, β, and δ
IV. Mutation
V. Validation
VI. Stagnation handling and elitism end (while)

The proposed method combines distance-based matrix representations and population-based optimization techniques modeled on the behavior of gray wolves in nature. This section describes the construction of the problem, initialization, iterative process, and enhancements incorporated into the BGWO approach.

I. Graph Representation and Matrix Construction The input graph is represented by an adjacency matrix A, which stores the weights between nodes. The diagonal entries of A are set to zero, indicating that the distance between any node and itself is zero. The distance is set to infinity (inf) if no direct edge exists between two nodes. To evaluate the distances between nodes and edges, the following matrices are constructed:

1. Distance Matrix D, contains the minimum distances along the shortest paths connecting every pair of vertices, computed using the Floyd-Warshall algorithm.
2. Edge-to-Node Distance Matrix B: For each edge in the graph, the minimum distance to all nodes is calculated and stored in Matrix B.
3. Combined Matrix C: The Matrix C merges D and B to provide a complete distance representation of both nodes and edges.

The Matrix C serves as the primary input to the optimization algorithm.

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

II. Binary Gray Wolf Optimization Framework

The steps of the BGWO algorithm are outlined as follows:

Step 1: Population Initialization

A population of popsize binary wolves is arbitrarily generated. Each wolf represents a candidate solution, where a one at position d indicates that vertex d is selected for the solution. If no valid solution is found in the initial population (i.e., all nodes are set to 0), the algorithm regenerates new wolves to ensure at least one vertex is selected.

Step 2: Fitness Evaluation

For each wolf in the population, the following steps are performed:

1. Matrix Representation: The binary vector of a wolf selects specific columns from Matrix C, generating a submatrix.
2. Lexicographic Sorting: The submatrix is sorted lexicographically to facilitate comparison between rows.
3. Row Uniqueness Validation: If any two rows in the sorted Matrix are identical, the wolf is deemed invalid. The wolf is updated by randomly flipping one 0 to a 1, and the validation is repeated until all rows are a unique representation of both vertices and edges.
4. Objective Value Calculation: The algorithm computes the maximum positional difference (j(i)) between consecutive rows of the sorted matrix. This value measures the quality of the solution.
5. Fitness Assignment: The fitness value is defined as 1/(1 + j(i)), where a lower j(i) yields a higher fitness score. This formulation encourages solutions with minimal vertex selections while maintaining row uniqueness.

Step 3: Selection of α, β, and δ Wolves

Based on the fitness values, the wolves are ranked, and the three best wolves are identified as α, β, and δ. These wolves guide the population's exploration and exploitation during the optimization process.

Step 4: Iterative Optimization

The algorithm proceeds iteratively for a maximum number of iterations, applying the following operations at each iteration:

1. Wolf Update: Each wolf updates its position based on the influence of alpha, beta, and delta. The update involves generating random coefficients (A and C), calculating new candidate positions (X1, X2, X3), and performing crossover to determine the final position.
2. Mutation: To maintain population diversity and avoid premature convergence, a mutation step randomly flips bits in the wolf vector with a certain probability.
3. Validation: After updating positions, the wolves undergo validation to ensure that all matrix representations maintain unique rows. Wolves failing this validation are repaired by modifying their binary vectors.
4. Objective Reevaluation: The algorithm recomputes the updated population's objective values and fitness scores.
5. Update α, β, and δ wolves: The α, β, and δ wolves are updated based on the new fitness rankings.

Step 5: Stagnation Handling

The algorithm tracks changes in the best objective value across iterations to address stagnation. If no improvement is observed over multiple iterations, the algorithm performs a partial restart by reinitializing a portion of the population. This strategy helps escape local optima and enhances the solution space exploration.

III. Termination and Result Visualization

The algorithm terminates when it reaches either the maximum iteration count or a convergence criterion. The best solution (alpha) is then extracted, representing the optimal selection of vertices for the MMDP. The script stores the MMD at each iteration. A plot is generated to visualize the progression of the metric dimension over time, demonstrating the algorithm's performance and how it approaches convergence.

4. Results

In this section, the experiments were conducted using MATLAB R2023b, installed on a Windows 10 Pro system equipped with a Core i5 processor and 16 GB of RAM. The results of the BGWO algorithm for the MMD of the cycle graphs, wheel graphs, path graphs, grid graphs, and complete bipartite graphs are presented in Tables 1, 2, 3, 4, and 5, respectively. Also, the upper bounds for the EMD and MMD of the generalized antiprism graph are provided. These theoretical results can then be compared with the outcomes obtained from the BGWO algorithm simulations.

An elitism process is required for path graphs P_n with $n \geq 12$, for ladder graphs L_n with $n \geq 9$, and for grid graphs $P_4 \times P_5$. Elitism ensures that the best solution found so far is never lost during iterations. After evaluating all individuals in the population at each iteration, the algorithm identifies the best-performing solution (alpha) based on the objective function. If this new best solution is better than the previously recorded best solution, it updates the record. Otherwise, the previous best solution is reinserted into the population, replacing the current worst individual. This process guarantees that the best solution persists through generations, preventing degradation due to random mutations or exploration steps. Consequently, elitism helps stabilize the algorithm, reduces fluctuations, and ensures convergence towards an optimal solution over time (see Fig. 2). This simulation shows that the BGWO algorithm produces the same results as those reported in (Kelenc, 2016; Sedlar & Škrekovski, 2021a; Sedlar & Škrekovski, 2021b).

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Table 1. Results of BGWO Algorithm for MMD on Cycle graphs.

Graph	n	m	popsize	It	dim_m	Mixed metric generator
Cycle graphs	4	4	20	1	3	$\{a_1, a_3, a_4\}$
	5	5	20	1	3	$\{a_1, a, a_4\}$
	6	6	20	1	3	$\{a_1, a_3, a_4\}$
	7	7	20	1	3	$\{a_1, a_4, a_6\}$
	8	8	20	1	3	$\{a_2, a_4, a_8\}$
	9	9	20	1	3	$\{a_1, a_5, a_9\}$
	10	10	20	1	3	$\{a_3, a_5, a_8\}$
	11	11	20	1	3	$\{a_3, a_5, a_9\}$
	12	12	20	1	3	$\{a_2, a_4, a_8\}$
	13	13	20	1	3	$\{a_2, a_8, a_{11}\}$
	14	14	20	1	3	$\{a_6, a_{13}, a_{14}\}$
	15	15	20	1	3	$\{a_5, a_{11}, a_{15}\}$

Table 2. Results of BGWO Algorithm for MMD on Wheel graphs.

Graph	n	m	popsize	It	dim_m	Mixed metric generator
Wheel graph	4	6	20	1	4	$\{w_1, w_2, w_3, w_4\}$
	5	8	20	1	4	$\{w_2, w_3, w_4, w_5\}$
	6	10	20	1	5	$\{w_2, w_3, w_4, w_5, w_6\}$
	7	12	20	1	6	$\{w_2, w_3, w_4, w_5, w_6, w_7\}$
	8	14	20	1	7	$\{w_2, w_3, w_4, w_5, w_6, w_7, w_8\}$
	9	16	20	1	8	$\{w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$
	10	18	20	1	9	$\{w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}\}$

Table 3. Results of BGWO Algorithm for MMD on path graphs.

Graph	n	m	popsize	It	dim_m	Mixed metric generator
Path graph	2	1	20	1	2	$\{v_1, v_2\}$
	3	2	20	1	2	$\{v_1, v_3\}$
	4	3	20	1	2	$\{v_1, v_4\}$
	5	4	20	1	2	$\{v_1, v_5\}$
	6	5	20	1	2	$\{v_1, v_6\}$
	7	6	20	1	2	$\{v_1, v_7\}$
	8	7	20	1	2	$\{v_1, v_8\}$
	9	8	20	1	2	$\{v_1, v_{10}\}$
	10	9	20	1	2	$\{v_1, v_{10}\}$
	11	10	20	1	2	$\{v_1, v_{11}\}$
	12	11	20	1	2	$\{v_1, v_{12}\}$
	13	12	20	1	2	$\{v_1, v_{13}\}$
	14	13	20	1	2	$\{v_1, v_{14}\}$
	15	14	20	1	2	$\{v_1, v_{15}\}$

Table 4. Results of BGWO Algorithm for MMD on Ladder and Grid graphs.

Graph	n	m	popsize	It	dim_m	Mixed metric generator
L_3	6	7	20	1	3	$\{l_1, l_3, l_6\}$
L_4	8	10	20	1	3	$\{l_3, l_5, l_8\}$
L_5	10	13	20	1	3	$\{l_1, l_5, l_8\}$
L_6	12	16	20	1	3	$\{l_2, l_7, l_{12}\}$
L_7	14	19	20	1	3	$\{l_2, l_8, l_{14}\}$
L_8	16	22	20	1	3	$\{l_1, l_8, l_{13}\}$
L_9	18	25	20	3	3	$\{l_5, l_{10}, l_{18}\}$
L_{10}	20	28	20	2	3	$\{l_1, l_{10}, l_{14}\}$
$P_3 \times P_3$	9	12	20	1	3	$\{p_1, p_3, p_7\}$
$P_3 \times P_4$	12	17	20	1	3	$\{p_1, p_3, p_{10}\}$
$P_4 \times P_5$	20	31	20	3	3	$\{p_1, p_4, p_{19}\}$
			20			

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Table 5. Results of BGWO Algorithm for MMD on Complete Bipartite graphs.

Graph	n	m	popsize	It	dim_m	Mixed metric generator
L_3	6	7	20	1	3	$\{l_1, l_3, l_6\}$
L_4	8	10	20	1	3	$\{l_3, l_5, l_8\}$
L_5	10	13	20	1	3	$\{l_1, l_5, l_8\}$
L_6	12	16	20	1	3	$\{l_2, l_7, l_{12}\}$
L_7	14	19	20	1	3	$\{l_2, l_8, l_{14}\}$
L_8	16	22	20	1	3	$\{l_1, l_8, l_{13}\}$
L_9	18	25	20	3	3	$\{l_5, l_{10}, l_{18}\}$
L_{10}	20	28	20	2	3	$\{l_1, l_{10}, l_{14}\}$
$P_3 \times P_3$	9	12	20	1	3	$\{p_1, p_3, p_7\}$
$P_3 \times P_4$	12	17	20	1	3	$\{p_1, p_3, p_{10}\}$
$P_4 \times P_5$	20	31	20	3	3	$\{p_1, p_4, p_{19}\}$
			20			

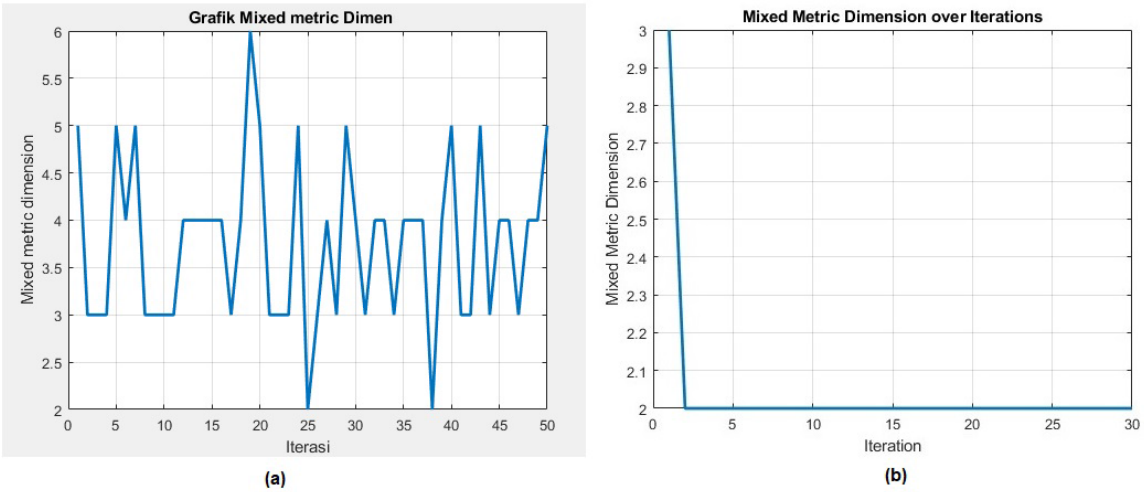


Figure 2. The MMD of P_{12} without (a) and with (b) elitism process

Now, we give the upper bound on the EMD and the upper bound on the MMD of the generalized antiprism graph.

Theorem 1. Edge metric dimension of the generalized antiprism graph A_3^n, $dim(A_3^n) \leq 2n$.

Proof. Suppose $V(A_3^n) = \{a_1^1, a_1^2, a_1^3, a_2^1, a_2^2, a_2^3, a_3^1, a_3^2, a_3^3, \dots, a_n^1, a_n^2, a_n^3\}$ and $S^n = \{a_1^1, a_1^2, a_2^2, a_2^3, a_3^1, a_3^3, \dots, a_n^x, a_n^y\} \subseteq V(A_3^n)$ with $(x, y) = \begin{cases} (1,2) \text{ if } n = 3t-2, \ t \in \{1,2,3,\dots\} \\ (2,3) \text{if } n = 3t-1, \ t \in \{1,2,3,\dots\} \text{ is the set of vertices located} \\ (1,3) \quad \text{otherwise} \end{cases}$
on paths that are opposite to each other, as illustrated in Figure 3. The representation of all edges in A_3^n is as follows:

$$r(a_1^1 a_1^2 | S^n) = (0,0,1,1,2,2, \dots, n-1, n-1)$$
$$r(a_1^2 a_1^3 | S^n) = (1,0,1,1,2,2, \dots, n-1, n-1)$$
$$r(a_1^1 a_1^3 | S^n) = (0,1,1,1,2,2, \dots, n-1, n-1)$$
$$r(a_2^1 a_2^2 | S^n) = (1,1,0,1,1,1, \dots, n-2, n-2)$$
$$r(a_2^2 a_2^3 | S^n) = (1,1,0,0,1,1, \dots, n-2, n-2)$$
$$r(a_2^1 a_2^3 | S^n) = (1,1,1,0,1,1, \dots, n-2, n-2)$$
$$r(a_3^1 a_3^2 | S^n) = (2,2,1,1,0,1, \dots, n-3, n-3)$$
$$r(a_3^2 a_3^3 | S^n) = (2,2,1,1,1,0, \dots, n-3, n-3)$$
$$r(a_3^1 a_3^3 | S^n) = (2,2,1,1,0,0, \dots, n-3, n-3)$$
$$\vdots$$
$$r(a_n^1 a_n^2 | S^n) = (n-1, n-1, n-2, n-2, n-3, n-3, \dots, x_1, y_1)$$
$$r(a_n^2 a_n^3 | S^n) = (n-1, n-1, n-2, n-2, n-3, n-3, \dots, x_2, y_2)$$
$$r(a_n^1 a_n^3 | S^n) = (n-1, n-1, n-2, n-2, n-3, n-3, \dots, x_3, y_3)$$

where:

$$(x_1, y_1) = \begin{cases} (0,0) \text{ if } n = 3t-2, \ t \in N \\ (0,1) \quad \text{otherwise,} \end{cases} \quad (x_2, y_2) = \begin{cases} (1,0) \text{ if } n = 3t-2 \text{ or } n = 3t, \ t \in N \\ (0,0) \quad \text{otherwise} \end{cases}$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

$$(x_3, y_3) = \begin{cases} (0,1) \text{ if } n = 3t-2 \text{ or } n = 3t, t \in N \\ (1,0) \text{ if } n = 3t-1, t \in N \\ (0,0) \text{ otherwise} \end{cases}$$

Next:

$$r(a_1^1 a_2^1 | S^n) = (0,1,1,1,1,1,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_1^2 a_2^1 | S^n) = (1,0,1,1,1,1,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_1^2 a_2^2 | S^n) = (1,0,0,1,1,2,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_1^3 a_2^2 | S^n) = (1,1,0,1,1,2,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_1^3 a_2^3 | S^n) = (1,1,1,0,2,1,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_1^1 a_2^3 | S^n) = (0,1,1,0,2,1,2,2,3,3, \dots, n-2, n-2)$$
$$r(a_2^1 a_3^1 | S^n) = (1,1,1,1,0,1,1,2,2,2, \dots, n-3, n-3)$$
$$r(a_2^2 a_3^1 | S^n) = (2,1,0,1,0,1,1,2,2,2, \dots, n-3, n-3)$$
$$r(a_2^2 a_3^2 | S^n) = (2,1,0,1,1,1,1,1,2,2, \dots, n-3, n-3)$$
$$r(a_2^3 a_3^2 | S^n) = (1,2,1,0,1,1,1,1,2,2, \dots, n-3, n-3)$$
$$r(a_2^3 a_3^3 | S^n) = (1,2,1,0,1,0,2,1,2,2, \dots, n-3, n-3)$$
$$r(a_2^1 a_3^3 | S^n) = (1,1,1,1,1,0,2,1,2,2, \dots, n-3, n-3)$$
$$r(a_3^1 a_4^1 | S^n) = (2,2,1,2,0,1,0,1,2,1, \dots, n-4, n-4)$$
$$r(a_3^2 a_4^1 | S^n) = (2,2,1,1,1,1,0,1,2,1, \dots, n-4, n-4)$$
$$r(a_3^2 a_4^2 | S^n) = (2,2,1,1,1,1,1,0,1,2, \dots, n-4, n-4)$$
$$r(a_3^3 a_4^2 | S^n) = (2,2,2,1,1,0,1,0,1,2, \dots, n-4, n-4)$$
$$r(a_3^3 a_4^3 | S^n) = (2,2,2,1,1,0,1,1,1,1, \dots, n-4, n-4)$$
$$r(a_3^1 a_4^3 | S^n) = (2,2,1,2,0,1,1,1,1,1, \dots, n-4, n-4)$$
$$r(a_4^1 a_5^1 | S^n) = (3,3,2,2,1,2,0,1,1,1, \dots, n-5, n-5)$$
$$r(a_4^2 a_5^1 | S^n) = (3,3,2,2,2,1,1,0,1,1, \dots, n-5, n-5)$$
$$r(a_4^2 a_5^2 | S^n) = (3,3,2,2,2,1,1,0,0,1, \dots, n-5, n-5)$$
$$r(a_4^3 a_5^2 | S^n) = (3,3,2,2,1,1,1,1,0,1, \dots, n-5, n-5)$$
$$r(a_4^3 a_5^3 | S^n) = (3,3,2,2,1,1,1,1,1,0, \dots, n-5, n-5)$$
$$r(a_4^1 a_5^3 | S^n) = (3,3,2,2,1,2,0,1,1,0, \dots, n-5, n-5)$$

$$\vdots$$

$$r(a_{(n-2)}^1 a_{(n-1)}^1 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_1, q_1)$$
$$r(a_{(n-2)}^2 a_{(n-1)}^1 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_2, q_2)$$
$$r(a_{(n-2)}^2 a_{(n-1)}^2 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_3, q_3)$$
$$r(a_{(n-2)}^3 a_{(n-1)}^2 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_4, q_4)$$
$$r(a_{(n-2)}^3 a_{(n-1)}^3 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_5, q_5)$$
$$r(a_{(n-2)}^1 a_{(n-1)}^3 | S^n) = (n-3, n-3, n-4, n-4, \dots, p_6, q_6)$$

$$r(a_{(n-1)}^1 a_n^1 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_1, s_1, t_1, u_1, v_1, w_1)$$

$$r(a_{(n-1)}^2 a_n^1 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_2, s_2, t_2, u_2, v_2, w_2)$$

$$r(a_{(n-1)}^2 a_n^2 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_3, s_3, t_3, u_3, v_3, w_3)$$

$$r(a_{(n-1)}^3 a_n^2 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_4, s_4, t_4, u_4, v_4, w_4)$$

$$r(a_{(n-1)}^3 a_n^3 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_5, s_5, t_5, u_5, v_5, w_5)$$

$$r(a_{(n-1)}^1 a_n^3 | S^n) = (n-2, n-2, n-3, n-3, \dots, 2,2, r_6, s_6, t_6, u_6, v_6, w_6)$$

where:

$$(p_1, q_1) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (1,2) \text{ if } n = 3t-1, t \in N \\ (2,1) \text{ otherwise} \end{cases}, \quad (p_2, q_2) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (1,2) \text{ if } n = 3t-1, t \in N \\ (2,1) \text{ otherwise} \end{cases}$$

$$(p_3, q_3) = \begin{cases} (1,2) \text{ if } n = 3t-2 \text{ or } n = 3t, t \in N \\ (1,1) \text{ otherwise} \end{cases}, \quad (p_4, q_4) = \begin{cases} (1,2) \text{ if } n = 3t-2 \text{ or } n = 3t, t \in N \\ (1,1) \text{ otherwise} \end{cases}$$

$$(p_5, q_5) = \begin{cases} (1,1) \text{ if } n = 3t, t \in N \\ (2,1) \text{ otherwise} \end{cases}, \quad (p_6, q_6) = \begin{cases} (1,1) \text{ if } n = 3t, t \in N \\ (2,1) \text{ otherwise} \end{cases}$$

$$(r_1, s_1) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (1,2) \text{ otherwise} \end{cases}, \quad (r_2, s_2) = \begin{cases} (1,1) \text{ if } n = 3t-1, t \in N \\ (2,1) \text{ otherwise} \end{cases}$$

$$(r_3, s_3) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (2,1) \text{ otherwise} \end{cases}, \quad (r_4, s_4) = \begin{cases} (1,2) \text{ if } n = 3t-2, t \in N \\ (2,1) \text{ if } n = 3t-1, t \in N \\ (1,1) \text{ otherwise} \end{cases}$$

$$(r_5, s_5) = \begin{cases} (1,2) \text{ if } n = 3t-2, t \in N \\ (2,1) \text{ if } n = 3t-1, t \in N \\ (1,1) \text{ otherwise} \end{cases}, \quad (r_6, s_6) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (1,2) \text{ otherwise} \end{cases}$$

$$(t_1, u_1) = \begin{cases} (1,1) \text{ if } n = 3t-2, t \in N \\ (0,1) \text{ otherwise} \end{cases}, \quad (t_2, u_2) = \begin{cases} (0,1) \text{ if } n = 3t-2, t \in N \\ (1,1) \text{ if } n = 3t-1, t \in N \\ (1,0) \text{ otherwise} \end{cases}$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

$$(t_3,u_3)=\begin{cases}(0,1)\ \text{if } n=3t-2, t\in N\\(1,1)\ \text{if } n=3t-1, t\in N\\(1,0)\ \text{otherwise}\end{cases},\ (t_4,u_4)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(1,0)\ \text{otherwise}\end{cases}$$

$$(t_5,u_5)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(1,0)\ \text{otherwise}\end{cases},\ (t_6,u_6)=\begin{cases}(1,1)\ \text{if } n=3t-2, t\in N\\(0,1)\ \text{otherwise}\end{cases}$$

$$(v_1,w_1)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(0,1)\ \text{otherwise}\end{cases},\ (v_2,w_2)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(0,1)\ \text{otherwise}\end{cases}$$

$$(v_3,w_3)=\begin{cases}(1,1)\ \text{if } n=3t-2, t\in N\\(1,0)\ \text{if } n=3t-1, t\in N\\(0,1)\ \text{otherwise}\end{cases},\ (v_4,w_4)=\begin{cases}(1,1)\ \text{if } n=3t-2, t\in N\\(1,0)\ \text{if } n=3t-1, t\in N\\(0,1)\ \text{otherwise}\end{cases}$$

$$(v_5,w_5)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(1,0)\ \text{otherwise}\end{cases},\ (v_6,w_6)=\begin{cases}(1,1)\ \text{if } n=3t-1, t\in N\\(0,1)\ \text{otherwise}\end{cases}$$

Because of all the edge representation are distinct, thus S^n is basis of A_3^n and $dim_e(A_3^n)\le 2n$. ■

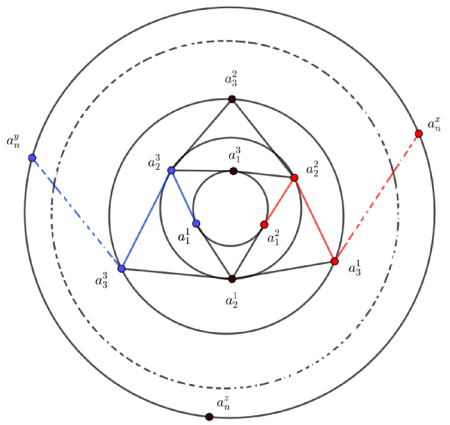


Figure 3. A basis of the generalized antiprism graph A_3^n

Theorem 2. MMD of the generalized antiprism graph A_3^n, $dim_m(A_3^n)\le 2n$.

Proof. Choose the set S^n as described in Theorem. 1. Consequently, the representation of the edges is presented as in the proof of Theorem 1. Furthermore, each vertex's representation in A_3^n can be stated as:

$$r(a_1^1|S^n)=(0,1,2,1,2,2,\dots,n-1,n-1)$$
$$r(a_1^2|S^n)=(1,0,1,2,2,2,\dots,n-1,n-1)$$
$$r(a_1^3|S^n)=(1,1,1,1,2,2,\dots,n-1,n-1)$$
$$r(a_2^1|S^n)=(1,1,1,1,1,1,\dots,n-2,n-2)$$
$$r(a_2^2|S^n)=(2,1,0,1,1,2,\dots,n-2,n-2)$$
$$r(a_2^3|S^n)=(1,2,1,0,2,1,\dots,n-2,n-2)$$
$$r(a_3^1|S^n)=(2,2,1,2,0,1,\dots,n-3,n-3)$$
$$r(a_3^2|S^n)=(2,2,1,1,1,1,\dots,n-3,n-3)$$
$$r(a_3^3|S^n)=(2,2,2,1,1,0,\dots,n-3,n-3)$$
$$\vdots$$
$$r(a_{(n-1)}^1|S^n)=(n-2,n-2,n-3,n-3,\dots,b_1,c_1)$$
$$r(a_{(n-1)}^2|S^n)=(n-2,n-2,n-3,n-3,\dots,b_2,c_2)$$
$$r(a_{(n-1)}^3|S^n)=(n-2,n-2,n-3,n-3,\dots,b_3,c_3)$$
$$r(a_n^1|S^n)=(n-1,n-1,n-2,n-2,,\dots,2,2,d_1,e_1,f_1,g_1)$$
$$r(a_n^2|S^n)=(n-1,n-1,n-2,n-2,,\dots,2,2,d_2,e_2,f_2,g_2)$$
$$r(a_n^3|S^n)=(n-1,n-1,n-2,n-2,,\dots,2,2,d_3,e_3,f_3,g_3)$$

where:

$$(b_1,c_1)=\begin{cases}(1,2)\ \text{if } n=3t-2, t\in N\\(2,1)\ \text{if } n=3t-1, t\in N\\(1,1)\ \text{otherwise}\end{cases},(b_2,c_2)=\begin{cases}(1,1)\ \text{if } n=3t-2, t\in N\\(1,2)\ \text{otherwise}\end{cases}$$

$$(b_3,c_3)=\begin{cases}(1,1)\ \text{if } n=3t-1, t\in N\\(2,1)\ \text{otherwise}\end{cases},\ (d_1,e_1)=\begin{cases}(1,1)\ \text{if } n=3t-1, t\in N\\(1,2)\ \text{otherwise}\end{cases}$$

$$(d_2,e_2)=\begin{cases}(1,1)\ \text{if } n=3t, t\in N\\(2,1)\ \text{otherwise}\end{cases},\ (d_3,e_3)=\begin{cases}(1,1)\ \text{if } n=3t-2, t\in N\\(1,2)\ \text{if } n=3t-1, t\in N\\(2,1)\ \text{otherwise}\end{cases}$$

$$(f_1,g_1)=\begin{cases}(1,1)\ \text{if } n=3t-1, t\in N\\(0,1)\ \text{otherwise}\end{cases},\ (f_2,g_2)=\begin{cases}(1,0)\ \text{if } n=3t-2, t\in N\\(0,1)\ \text{if } n=3t-1, t\in N\\(1,1)\ \text{otherwise}\end{cases}$$

$$(f_3,g_3)=\begin{cases}(2,1)\ \text{if } n=3t-2, t\in N\\(1,0)\ \text{otherwise}\end{cases}$$

Because of all the edge and vertex representations being different, S^n is a mixed generator of A_3^n and $dim_m(A_3^n)\le 2n$. ■

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Table 6. Results of BGWO Algorithm for EMD on generalized antiprism graphs A_3^n

Graph	n	m	popsize	It	dim_e	edge metric generator
A_3^3	9	21	20	1	6	$\{a_2, a_3, a_5, a_6, a_7, a_8\}$
A_3^4	12	30	20	1	8	$\{a_1, a_3, a_4, a_6, a_8, a_9, a_{11}, a_{12}\}$
A_3^5	15	39	20	2	10	$\{a_1, a_2, a_4, a_5, a_8, a_9, a_{10}, a_{12}, a_{13}, a_{14}\}$
A_3^6	18	48	20	1	12	$\{a_1, a_3, a_4, a_5, a_8, a_9, a_{10}, a_{11}, a_{13}, a_{14}, a_{16}, a_{18}\}$
A_3^7	21	57	20	1	14	$\{a_1, a_2, a_4, a_5, a_8, a_9, a_{10}, a_{12}, a_{14}, a_{15}, a_{16}, a_{18}, a_{19}, a_{21}\}$
A_3^8	24	66	20	2	16	$\{a_2, a_3, a_4, a_5, a_8, a_9, a_{11}, a_{12}, a_{14}, a_{15}, a_{17}, a_{18}, a_{19}, a_{21}, a_{22}, a_{23}\}$
A_3^9	27	75	20	2	18	$\{a_2, a_3, a_5, a_6, a_7, a_8, a_{10}, a_{11}, a_{13}, a_{15}, a_{17}, a_{18}, a_{19}, a_{21}, a_{22}, a_{23}, a_{26}, a_{27}\}$
A_3^{10}	30	84	20	2	20	$\{a_1, a_2, a_4, a_6, a_7, a_9, a_{10}, a_{12}, a_{13}, a_{15}, a_{16}, a_{18}, a_{20}, a_{21}, a_{23}, a_{24}, a_{25}, a_{27}, a_{28}, a_{30}\}$
A_3^{11}	33	93	20	2	22	$\{a_2, a_3, a_5, a_6, a_7, a_9, a_{11}, a_{12}, a_{14}, a_{15}, a_{16}, a_{18}, a_{19}, a_{20}, a_{23}, a_{24}, a_{26}, a_{27}, a_{28}, a_{30}, a_{32}, a_{33}\}$
A_3^{12}	36	102	20	2	24	$\{a_1, a_2, a_5, a_6, a_7, a_9, a_{10}, a_{11}, a_{14}, a_{15}, a_{17}, a_{18}, a_{20}, a_{21}, a_{22}, a_{23}, a_{25}, a_{27}, a_{29}, a_{30}, a_{31}, a_{32}, a_{34}, a_{35}\}$

Table 7. Results of BGWO Algorithm for MMD on generalized antiprism graphs A_3^n

Graph	n	m	popsize	It	dim_e	mixed metric generator
A_3^3	9	21	20	1	6	$\{a_1, a_2, a_5, a_6, a_7, a_9\}$
A_3^4	12	30	20	4	8	$\{a_2, a_3, a_4, a_6, a_7, a_8, a_{11}, a_{12}\}$
A_3^5	15	39	20	1	10	$\{a_1, a_2, a_5, a_6, a_7, a_9, a_{10}, a_{11}, a_{14}, a_{15}\}$
A_3^6	18	48	20	1	12	$\{a_2, a_3, a_4, a_6, a_7, a_8, a_{11}, a_{12}, a_{13}, a_{15}, a_{16}, a_{17}\}$
A_3^7	21	57	20	7	14	$\{a_1, a_3, a_4, a_5, a_8, a_9, a_{10}, a_{12}, a_{14}, a_{15}, a_{16}, a_{18}, a_{19}, a_{20}\}$
A_3^8	24	66	20	5	16	$\{a_2, a_3, a_4, a_6, a_7, a_8, a_{10}, a_{12}, a_{14}, a_{15}, a_{16}, a_{17}, a_{20}, a_{21}, a_{22}, a_{24}\}$
A_3^9	27	75	20	4	18	$\{a_2, a_3, a_4, a_6, a_7, a_8, a_{11}, a_{12}, a_{14}, a_{15}, a_{16}, a_{18}, a_{19}, a_{20}, a_{23}, a_{24}, a_{25}, a_{27}\}$
A_3^{10}	30	84	20	3	20	$\{a_1, a_2, a_5, a_6, a_7, a_9, a_{10}, a_{12}, a_{13}, a_{15}, a_{16}, a_{18}, a_{19}, a_{21}, a_{22}, a_{24}, a_{25}, a_{26}, a_{29}, a_{30}\}$
A_3^{11}	33	93	20	8	22	$\{a_1, a_2, a_5, a_6, a_7, a_9, a_{10}, a_{11}, a_{13}, a_{15}, a_{16}, a_{17}, a_{20}, a_{21}, a_{22}, a_{24}, a_{25}, a_{26}, a_{29}, a_{30}, a_{31}, a_{33}\}$
A_3^{12}	36	102	20	13	24	$\{a_1, a_2, a_5, a_6, a_7, a_9, a_{10}, a_{12}, a_{13}, a_{14}, a_{16}, a_{18}, a_{19}, a_{20}, a_{22}, a_{23}, a_{26}, a_{27}, a_{29}, a_{30}, a_{31}, a_{33}, a_{34}, a_{35}\}$

The following presents the simulation results of the BGWO algorithm for determining the EMD and MMD of the generalized antiprism, as shown in Tables 6 and 7. The results demonstrate that the BGWO algorithm yields values consistent with those established in Theorems 1 and 2, with efficient completion times indicated by the required number of iterations.

5. Conclusion

This paper presents a modified binary Gray Wolf Optimizer (BGWO) algorithm to address the mixed-metric dimension problem. The algorithm is designed to reduce the number of vertices required to uniquely identify all elements of a graph based on distance constraints. The proposed BGWO algorithm incorporates several enhancements aimed at improving performance in the MMDP. First, validation of the uniqueness of the row ensures that all solutions maintain valid matrix representations. Second, dynamic stagnation handling detects and resolves stagnation through partial population restarts. Third, mutation promotes diversity and prevents premature convergence. Lastly, iterative crossover balances exploration and exploitation by dynamically updating the positions of the wolves. In some graphs with many vertices or edges, an elitism process is necessary. Elitism ensures that the best solution found so far is never lost during iterations. The results demonstrate that the modified BGWO algorithm achieves efficient completion times and delivers competitive outcomes for this problem.

Moreover, the BGWO algorithm shows practical utility in the following areas: (1) conducting pattern recognition to identify resolving set patterns within specific graph families. The algorithm's outputs provide various minimum resolving sets, allowing the selection of resolving sets for pattern identification and analysis; (2) developing hypotheses about the mixed metric dimension values for specific types or families of graphs; (3) generating mixed metric dimension values that may coincide with or fall below the upper bounds obtained from theoretical results. This phenomenon occurs because theoretical results rely on fixed patterns or regularities, while the BGWO algorithm operates through random calculations using an iterative refinement mechanism.

6. Acknowledgments

The author sincerely acknowledges the support provided by various divisions within the Ministry of Higher Education, Science, and Technology through the BPI doctoral research scholarship and Universitas Kadiri, which facilitated the completion of this study.

7. References

Abbas, Waseem et al. 2024. "Investigating Metric Dimension and Edge Metric Dimension of Hexagonal Boron Nitride and Carbon Nanotubes." *European Journal of Pure and Applied Mathematics* 17(3): 2055–72.

Beerliova, Zuzana et al. 2006. "Network Discovery and Verification." *IEEE Journal on Selected Areas in Communications* 24(12): 2168–81.

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Chartrand, Gary, Linda Eroh, Mark A. Johnson, and Ortrud R. Oellermann. 2000. "Resolvability in Graphs and the Metric Dimension of a Graph." *Discrete Applied Mathematics* 105(1–3): 99–113.

Chelladurai, Xavier. 2020. "Algorithms for the Metric Dimension of a Simple Graph." *Data Science and Security, Lectures Notes in Network and Systems* (132): 91–105.

Chen, Xujin, Xiaodong Hu, and Changjun Wang. 2016. "Approximation for the Minimum Cost Doubly Resolving Set Problem." *Theoretical Computer Science* 609: 526–43. http://dx.doi.org/10.1016/j.tcs.2015.03.048.

Danas, Milica Milivojević. 2021. "The Mixed Metric Dimension of Flower Snarks and Wheels." *Open Mathematics* 19(1): 629–40.

Darmaji, and N. Azahra. 2022. "The Mixed Metric Dimension of Wheel-like Graphs." *Journal of Physics: Conference Series* 2157(1).

Diaz, Josep, Olli Pottonen, Maria Serna, and Erik Jan van Leeuwen. 2017. "Complexity of Metric Dimension on Planar Graphs." *Journal of Computer and System Sciences* 83(1): 132–58. http://dx.doi.org/10.1016/j.jcss.2016.06.006.

Emary, E., Hossam M. Zawbaa, and Aboul Ella Hassanien. 2016. "Binary Grey Wolf Optimization Approaches for Feature Selection." *Neurocomputing* 172: 371–81. http://dx.doi.org/10.1016/j.neucom.2015.06.083.

Harrary, F, and R.A Melter. 1976. "On the Metric Dimension of a Graph." *Ars Combin* 2: 191–95.

Hauptmann, Mathias, Richard Schmied, and Claus Viehmann. 2012. "Approximation Complexity of Metric Dimension Problem." *Journal of Discrete Algorithms* 14: 214–22. http://dx.doi.org/10.1016/j.jda.2011.12.010.

Hendy, Hendy, Mohammad Isa, Imam Mukhlash, and Samsul Setumin. 2023. "A Bibliometric Analysis of Metaheuristic Research and Its Applications." *Register: Jurnal Ilmiah Teknologi Sistem Informasi* 9(January): 1–17. http://www.journal.unipdu.ac.id/index.php/register/article/view/2675.

Hertz, Alain. 2020. "An IP-Based Swapping Algorithm for the Metric Dimension and Minimal Doubly Resolving Set Problems in Hypercubes." *Optimization Letters* 14(2): 355–67.

Hoffmann, Stefan, Alina Elterman, and Egon Wanke. 2016. "A Linear Time Algorithm for Metric Dimension of Cactus Block Graphs." *Theoretical Computer Science* 630: 43–62. http://dx.doi.org/10.1016/j.tcs.2016.03.024.

Imran, Muhammad, Ali Ahmad, Muhammad Azeem, and Kashif Elahi. 2022. "Metric-Based Resolvability of Quartz Structure." *Computers, Materials and Continua* 71(1): 2053–71.

Kelenc, Aleksander, Dorota Kuziak, Andrej Taranenko, and Ismael G. Yero. 2017. "Mixed Metric Dimension of Graphs." *Applied Mathematics and Computation* 314: 429–38. http://dx.doi.org/10.1016/j.amc.2017.07.027.

Kelenc, Aleksander, Niko Tratnik, and Ismael G. Yero. 2018. "Uniquely Identifying the Edges of a Graph: The Edge Metric Dimension." *Discrete Applied Mathematics* 251(January 2016): 204–20.

Khuller, Samir, Balaji Raghavachari, and Azriel Rosenfeld. 1996. "Landmarks in Graphs." *Discrete Applied Mathematics* 70(3): 217–29.

Knor, Martin et al. 2021. "Graphs with the Edge Metric Dimension Smaller than the Metric Dimension." *Applied Mathematics and Computation* 401.

Koam, Ali N.A., Ali Ahmad, Sadia Husain, and Muhammad Azeem. 2023. "Mixed Metric Dimension of Hollow Coronoid Structure." *Ain Shams Engineering Journal* 14(7): 102000. https://doi.org/10.1016/j.asej.2022.102000.

Kratica, Jozef, Vera Kovačević-Vujčić, and Mirjana Čangalović. 2009. "Computing the Metric Dimension of Graphs by Genetic Algorithms." *Computational Optimization and Applications* 44(2): 343–61.

Kratica, Jozef, Vera Kovačević-Vujčić, Mirjana Čangalović, and Milica Stojanović. 2012. "Minimal Doubly Resolving Sets and the Strong Metric Dimension of Some Convex Polytopes." *Applied Mathematics and Computation* 218(19): 9790–9801.

Melter, Robert A., and Ioan Tomescu. 1984. "Metric Bases in Digital Geometry." *Computer Vision, Graphics and Image Processing* 25(1): 113–21.

Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Andrew Lewis. 2014. "Grey Wolf Optimizer." *Advances in Engineering Software* 69: 46–61. http://dx.doi.org/10.1016/j.advengsoft.2013.12.007.

Mladenović, Nenad, Jozef Kratica, Vera Kovačević-Vujčić, and Mirjana Čangalović. 2012. "Variable Neighborhood Search for Metric Dimension and Minimal Doubly Resolving Set Problems." *European Journal of Operational Research* 220(2): 328–37.

Mohame, Basma, Mohammed Ami, and Linda Mohaise. 2023. "Binary Archimedes Optimization Algorithm for Computing Dominant Metric Dimension Problem." *Intelligent Automation and Soft Computing* 38(1): 19–34.

Muhammad, F., and L. Susilowati. 2020. "Algorithm and Computer Program to Determine Metric Dimension of Graph." *Journal of Physics: Conference Series* 1494(1): 0–5.

Murdiansyah, D.T, and Adiwijaya. 2017. "Computing the Metric Dimension of Hypercube Graphs by Particle Swarm Optimization Algorithm." *Recent Advances on Soft Computing and Data Mining, Advances in Intelligent Systems and Computing* 549.

Peterin, Iztok, and Ismael G. Yero. 2020. "Edge Metric Dimension of Some Graph Operations." *Bulletin of the Malaysian*

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

Mathematical Sciences Society 43(3): 2465–77. https://doi.org/10.1007/s40840-019-00816-7.

Seb, Andris, and Eric Tannier. 2016. "On Metric Generators of Graphs Author (s): András Sebő and Eric Tannier Published by : INFORMS Stable URL : Http://Www.Jstor.Org/Stable/30035689 On Metric Generators of Graphs." 29(2): 383–93.

Sedlar, Jelena, and Riste Škrekovski. 2021a. "Extremal Mixed Metric Dimension with Respect to the Cyclomatic Number." *Applied Mathematics and Computation* 404: 126238. https://doi.org/10.1016/j.amc.2021.126238.

Sedlar, Jelena, and Riste Škrekovski.. 2021b. "Mixed Metric Dimension of Graphs with Edge Disjoint Cycles." *Discrete Applied Mathematics* 300: 1–8. https://doi.org/10.1016/j.dam.2021.05.004.

Slater, P.J. 1975. "Leaves of Trees." *Congr. Numer* 14: 549–59.

Tillquist, Richard C., and Manuel E. Lladser. 2019. "Low-Dimensional Representation of Genomic Sequences." *Journal of Mathematical Biology*. https://doi.org/10.1007/s00285-019-01348-1.

Vietz, Duygu, Stefan Hoffmann, and Egon Wanke. 2019. 11355 LNCS Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) *Computing the Metric Dimension by Decomposing Graphs into Extended Biconnected Components: (Extended Abstract)*. Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-10564-8_14.

Zhang, Yuezhong, and Suogang Gao. 2020. "On the Edge Metric Dimension of Convex Polytopes and Its Related Graphs." *Journal of Combinatorial Optimization* 39(2): 334–50. https://doi.org/10.1007/s10878-019-00472-4.

Zhu, Enqiang, Andrej Taranenko, Zehui Shao, and Jin Xu. 2019. "On Graphs with the Maximum Edge Metric Dimension." *Discrete Applied Mathematics* 257: 317–24. https://doi.org/10.1016/j.dam.2018.08.031.

DOI: https://doi.org/10.22452/mjs.vol45sp2.5
Malaysian Journal of Science 45 (Special Issue 2): 43-54 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special issue

Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

https://mjs.um.edu.my

Comparative Analysis of Time Series Models: LSTM, Arima, FB Prophet for Predicting Rainfall and Streamflow

Nur-Adib Maspo[1a*], Aizul Nahar Harun[2b], Amelia Ritahani Ismail[3c] and Mimi Liza Abdul Majid[4d]

Abstract: Accurate forecasting of rainfall and streamflow is essential for water resource management and climate resilience in Malaysia, particularly under monsoon-driven variability. This study evaluates Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), and Facebook Prophet (FB Prophet) for predicting these hydrological variables using data from the Department of Irrigation and Drainage, Malaysia (2009–2019). LSTM, with its 50-memory-unit architecture, achieved the lowest Root Mean Square Error (RMSE) of 15.328631 for rainfall, excelling in capturing nonlinear, long-term patterns. ARIMA outperformed with an RMSE of 1.448643 for streamflow, using linear autocorrelations. FB Prophet showed higher errors: RMSE 20.869064 for rainfall, 2.319169 for streamflow, indicating limited suitability despite its seasonality modeling. The findings underscore the importance of tailoring model selection to hydrological variables: LSTM for complex rainfall, ARIMA for linear streamflow, while advocating hybrid approaches to improve accuracy. This research offers practical insights into the strengths of models, contributing to sustainable water management and addressing global water issues.

Keywords: LSTM, ARIMA, Facebook Prophet, time series analysis, rainfall prediction.

1. Introduction

Accurate and reliable prediction of rainfall and streamflow is crucial for effective water resource management, flood prevention, and climate adaptation, particularly in regions like Malaysia, where monsoon-driven variability poses significant challenges to water security and infrastructure resilience (Kumar et al., 2024). These hydrological phenomena are influenced by complex climatic patterns, such as the northeast and southwest monsoons, which bring heavy rainfall and variable streamflow to areas like Kelantan (Abdullah et al., 2024). This necessitates the use of advanced forecasting tools to support decision-making and reduce water-related risks. Time series models, which analyse historical data to forecast future trends, have become vital in hydrology, enabling stakeholders to address challenges such as flooding, drought, and water scarcity amid increasing climate variability and resource constraints.

To address this issue, this study conducts a comparative analysis of three prominent time series forecasting models: Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), and Facebook Prophet (FB Prophet). The purpose is to evaluate their performance in forecasting rainfall and streamflow using a dataset (2009–2019) obtained from the Department of Irrigation and Drainage (DID) Malaysia. This dataset comprises univariate time series from 24 rainfall stations, six streamflow stations, and six water level stations within the Sg. Kelantan drainage area. This offers a rich, monsoon-influenced context for assessing model performance, building on prior hydrological experiments conducted globally and regionally. For example, related studies, such as Adnan et al. (2017) on ARIMA for streamflow in Pakistan and Xu et al. (2020) on LSTM for Australian rivers, have demonstrated the potential of these models. However, only a limited number of researchers have examined the application of deep learning approaches, specifically LSTM, and their comparison with traditional statistical models such as ARIMA, within the context of Malaysia's distinctive monsoon-driven hydrological patterns, particularly using data from the Sg. Kelantan river basin, which experiences some of the most frequent and severe annual flooding events (Khairudin et al., 2020; Latif et al., 2021; Sherstinsky, 2020).

Authors information:
[a]Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Selangor, MALAYSIA. E-mail: nuradibmaspo@iium.edu.my[1]
[b]Department of Management of Technology, Malaysia-Japan International Institute of Technology, University of Technology Malaysia, 54100 Kuala Lumpur, MALAYSIA. E-mail: aizulnahar.kl@utm.my[2]
[c]Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Selangor, MALAYSIA. E-mail: amelia@iium.edu.my[3]
[d]Department of Information System, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Selangor, MALAYSIA. E-mail: mimiliza@iium.edu.my[4]

*Corresponding Author: nuradibmaspo@iium.edu.my

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

This gap highlights the need to assess model suitability under the specific climatic and hydrological conditions of Malaysia. Accordingly, this study aims to fill that gap by conducting a thorough evaluation of LSTM, ARIMA, and FB Prophet models. The main objectives are threefold:

1. To conduct a detailed review of the practical applicability of LSTM, ARIMA, and FB Prophet in hydrological forecasting.
2. To empirically analyse and compare their predictive accuracy using rainfall and streamflow data collected from 2009 to 2019 by the Department of Irrigation and Drainage Malaysia.
3. To provide evidence-based recommendations for model selection, optimization, and potential integration to improve forecasting accuracy.

LSTM is a sophisticated deep learning model within the recurrent neural network (RNN) family, created to capture long-term dependencies and complex temporal patterns in sequential data (Shumway et al., 2017; Sherstinsky, 2020). This makes it notably suitable for modelling the non-linear and seasonally varying rainfall patterns influenced by Malaysia's monsoons (Ramlan et al., 2021; Ibrahim et al., 2024). ARIMA, a traditional statistical model by Sivaramakrishnan et al. (2022), integrates autoregressive, differencing, and moving average components and is effective in modelling linear, autocorrelated time series, provided the data are stationary. FB Prophet, developed by Meta, is an automated forecasting tool that incorporates trend, seasonality, and holiday effects using a Bayesian framework (Xu et al., 2022). However, its use in hydrological contexts remains relatively limited.

Each model has distinct advantages and limitations: LSTM's ability to capture non-linearity comes with high computational demands and the need for hyperparameter tuning; ARIMA is restricted to linear trends but is computationally efficient; and FB Prophet provides automation and interpretability but lacks extensive use in hydrology. These differences require a thorough comparative analysis to assess their suitability for forecasting tasks within Malaysia's monsoon-influenced hydrological environment.

The importance of this research lies in its potential to enhance forecasting accuracy for rainfall and streamflow, thus supporting effective water resource management, flood control, and climate adaptation in Malaysia and similar regions. The paper is organized as follows: Section 2 reviews related studies, Section 3 describes the methodology, including data preprocessing and model development, Section 4 presents results and discussion, and Section 5 concludes with key findings and future research suggestions.

2. Related works

Time series forecasting in hydrology has evolved significantly, with several models applied to predict rainfall, streamflow, and water levels. This section reviews key studies on LSTM, ARIMA, and FB Prophet, identifying their strengths, limitations, and relevance to this study's focus on Malaysia's hydrological data.

2.1 ARIMA in Hydrology

ARIMA models have been extensively used for their ability to address linear, seasonal patterns in hydrological time series. Adnan et al. (2017) compared ARMA and ARIMA for monthly streamflow prediction from 1974 to 2010, reporting better ARIMA performance with lower RMSE and Mean Absolute Percentage Error (MAPE) values. Agaj et al. (2024) applied ARIMA to streamflow forecasting on the Yangtze River, obtaining satisfactory results in terms of RMSE and MAPE, although uncertainties arose due to sudden flow changes. Kobiela et al. (2022) highlighted ARIMA's effectiveness in linear scenarios but noted its difficulties with non-linear patterns, restricting its applicability to complex hydrological data. These findings indicate ARIMA's reliability for stationary, linear streamflow, but its limitations in non-linear contexts encourage the exploration of advanced models, such as LSTM.

2.2 LSTM Advances

Deep learning approaches, particularly LSTM, have gained prominence in capturing non-linear and long-term dependencies in time series. Ghimire et al. (2021) integrated convolutional neural networks with LSTM for hourly streamflow prediction at the Brisbane River, achieving high accuracy when modelling complex dynamics, with Nash-Sutcliffe Efficiency (NSE) values exceeding 0.9. Xu et al. (2020) compared LSTM with the SWAT model for river flow in the Upper Yangtze, noting LSTM's superiority in capturing peak flows, which are critical for flood forecasting, but at the expense of requiring significant computational resources. Sahoo et al. (2024) reported NSE values up to 0.971 for daily flow prediction in the Brahmani-Baitarani basin, while Ostadkalayeh et al. (2023) enhanced LSTM performance for streamflow using Kalman filtering, addressing tuning challenges. These studies underscore LSTM's potential for non-linear hydrological data but emphasize the need for careful hyperparameter optimization, a focus of this study.

2.3 FB Prophet Applications

FB Prophet, designed for automated forecasting with seasonality and trends, is less explored in hydrology but demonstrates promise. Aguilera et al. (2019) applied it to groundwater level prediction, indicating its utility in water management, though with moderate accuracy. Sulasikin et al. (2021) used FB Prophet for monthly rainfall prediction in Central Jakarta, reporting low Mean Squared Error (MSE) and RMSE, while Hossain et al. (2022) combined observed and CMIP5 data, enhancing predictions for dry periods. However, its performance lags in complex hydrological contexts, indicating a need for further validation, as addressed in this study.

Direct comparisons of LSTM, ARIMA, and FB Prophet in hydrology are limited. However, interdisciplinary studies offer valuable insights. Ibrahim et al. (2024) and Sivaramakrishnan et al. (2022) compared these models for monthly river flow forecasting, finding LSTM better suited for non-linear patterns, ARIMA for linear trends, and FB Prophet for seasonal data, although with higher errors in real-time applications. Recent work by Xu et al. (2022) explores hybrid LSTM-ARIMA models, indicating improved accuracy for drought forecasting, with an

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

accuracy higher than that of the three individual models, such as ARIMA, SVM, and LSTM.

3. Methodology

This research employed a machine learning pipeline approach, as visually represented in Figure 1, which outlines the sequential process from data collection to model evaluation. This figure highlights the integration of data processing, exploratory data analysis, data transformation, model development, forecasting, and performance evaluation. The specific techniques used in the study will be elaborated upon further in the following sections.

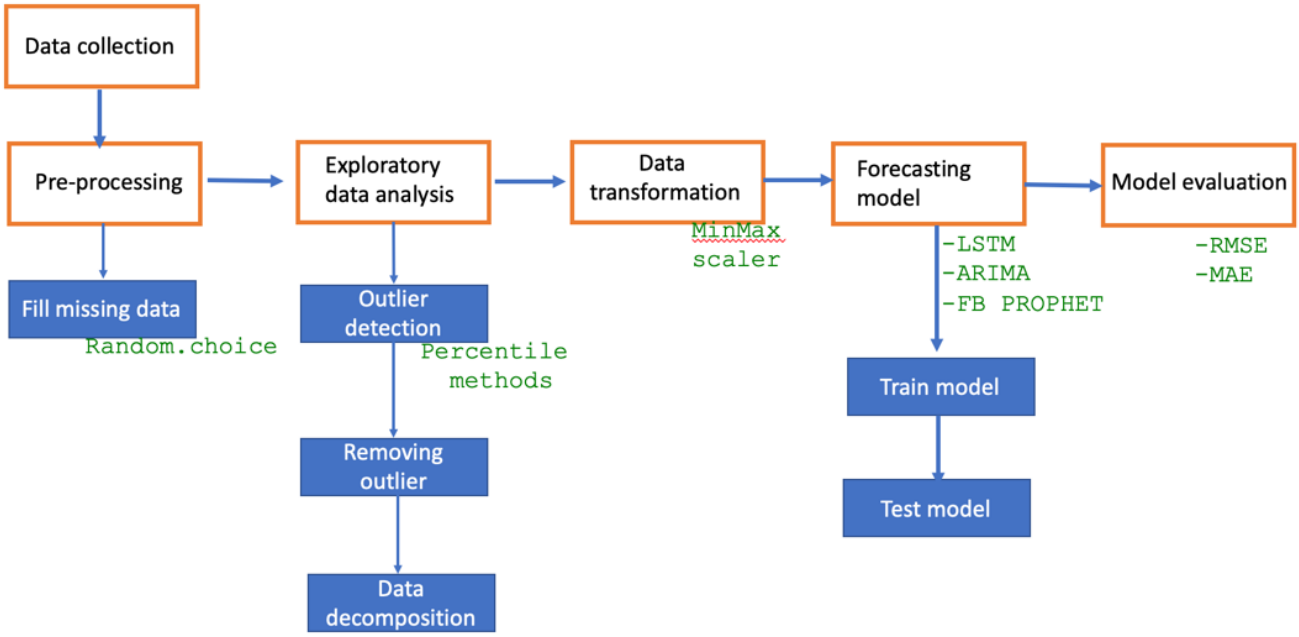


Figure 1. Model pipeline for flood forecasting using time series data of rainfall and streamflow.

3.1 Data Collection and Pre-Processing

The dataset for this study was obtained from the DID, Malaysia, comprising univariate time series data for rainfall (24 stations), streamflow (6 stations), and water level (6 stations) spanning the period from 2009 to 2019. Rainfall data from the 24 observation stations were compiled into a single file (rainfallmerger.csv), while streamflow and water level data were combined into streamflowmerger.csv and waterlevelmerger.csv, respectively. This aggregation enabled a unified analysis of hydrological variables.

Pre-processing was essential to ensure data quality for modelling. Missing values, identified by symbols such as "#" (e.g., 22#, 300#), were common in the raw data. These incomplete entries were replaced using the Random Choice method, which selects probable values from correlated stations based on a probability approach. This method outperformed a neural network approach, as it provided more accurate estimations of missing values by exploiting spatial correlations between stations. The pre-processing pipeline, as illustrated in Figure 1, begins with data collection and includes steps such as filling missing data, which is a key step before further analysis.

Exploratory Data Analysis (EDA) was conducted to visualize patterns and understand data attributes. Figure 2 depicted rainfall intensity across three stations, showing peaks during the northeast monsoon season (October to January) in Kelantan. Similarly, Figure 3 illustrates streamflow patterns, identifying a significant peak from late 2014 to early 2015. Pearson's correlation coefficient was used to assess relationships between stations, with Figure 4 demonstrating strong correlations (e.g., between stations 561051 and 5621052) due to their proximity within the Kelantan River network.

Outlier detection is another key component of EDA and was performed using percentile methods as presented in Figure 5. Outliers, such as those at station 4614001, were identified as extreme rainfall events but were addressed to prevent skewing model performance. These outliers were either removed or adjusted to minimize their impact on error metrics, such as RMSE, thereby ensuring accurate model fitting.

Feature extraction involved decomposing monthly time series data into trend, seasonality, and residuals. This decomposition provided deeper insights into rainfall patterns, revealing seasonal fluctuations without any clear trend and random residuals. Data transformation was then performed, employing normalization to scale variables to a [0, 1] range using the MinMax scaler. This scaling ensured that all features contributed equally during model training, thereby reducing bias in algorithms that are sensitive to variable scales.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

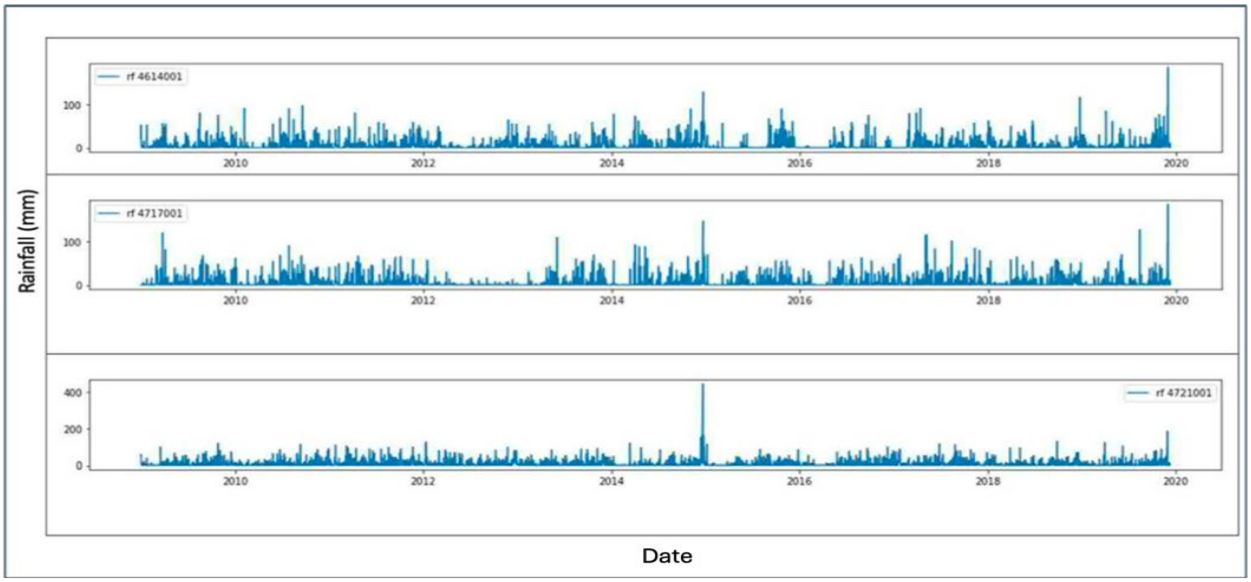


Figure 2. Rainfall intensity across three stations from 2009 to 2019.

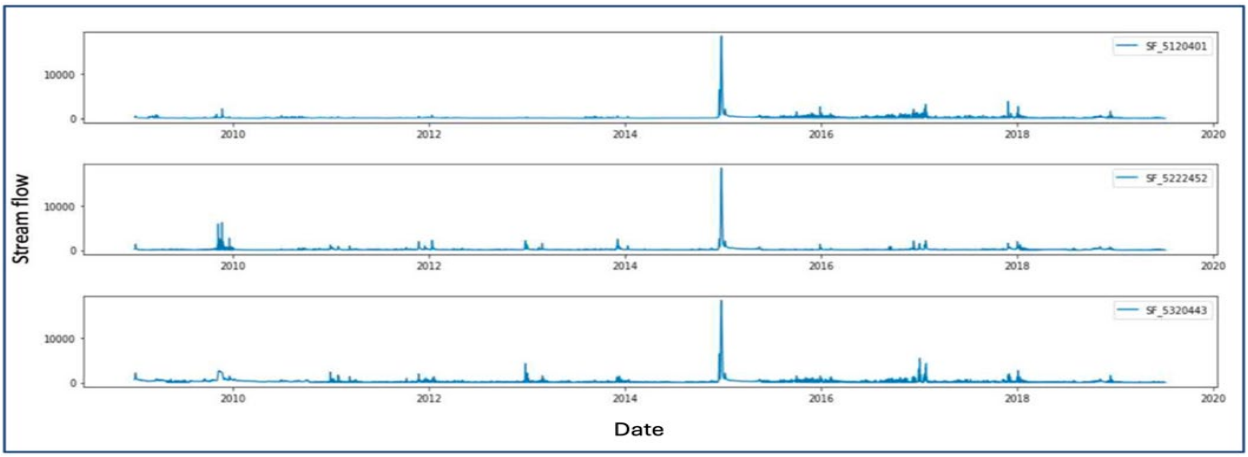
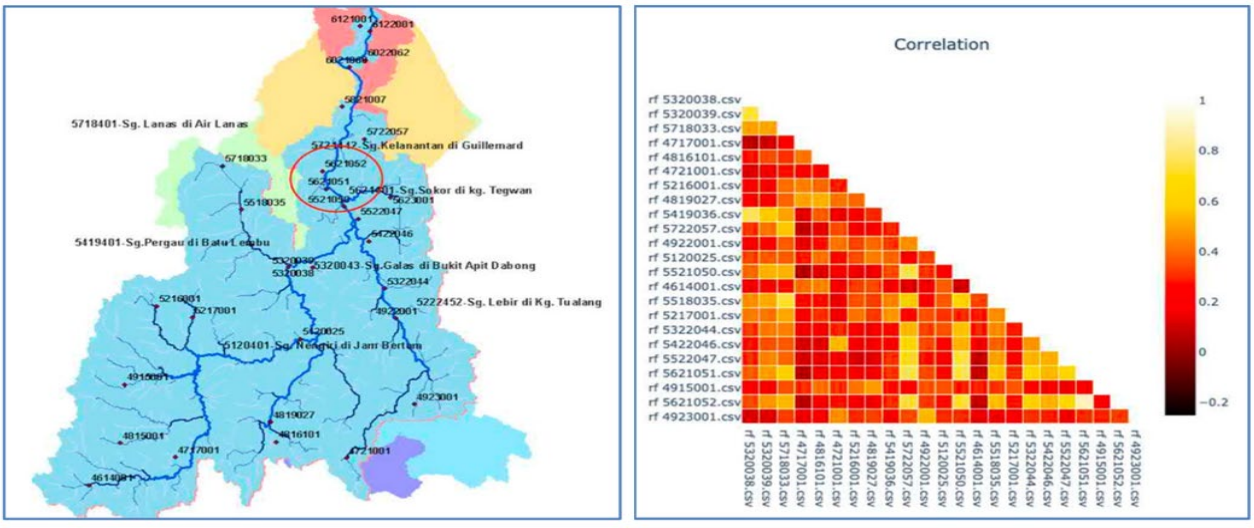


Figure 3. Stream flow pattern for 10 years of observation at 3 measuring stations from 2009 to 2019.



(a) Kelantan River network and location of rainfall observation

(b) Pearson's correlation of rainfall stations

Figure 4. Spatial distance and correlation of rainfall stations in the Kelantan River network. (a) Distance between observation stations. (b) Correlation of rainfall data between stations, with an example of strong correlation between station 561051 and its nearby station 5621052.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

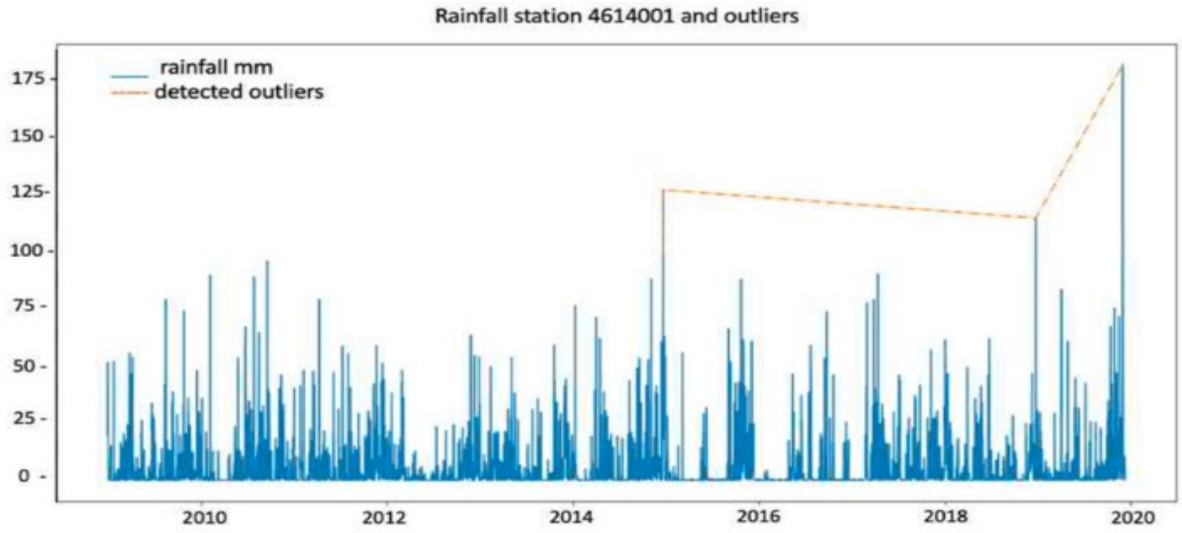


Figure 5. Graph representation of outlier detection.

4. Model development

The pre-processed dataset was divided into a 70% training set (covering January 1, 2009, to May 10, 2016) and a 30% testing set (from May 11, 2016, to July 8, 2019) to train and validate the models. The following distinct models were created and evaluated for forecasting rainfall and streamflow.

4.1 LSTM Model Architecture and Hyperparameter Configuration

In this study, a stacked Long Short-Term Memory (LSTM) neural network was utilized to forecast univariate time series data for rainfall and streamflow prediction. The model was developed using Keras' Sequential () API, enabling the linear stacking of layers. The architecture consists of three LSTM layers followed by a fully connected (Dense) output layer. To assess the impact of network capacity on prediction accuracy, experiments were conducted using three different LSTM configurations, varying the number of memory units (hidden units) in each layer: 30, 50, and 90 units, respectively. These units determine the dimensionality of the output space of each LSTM layer, which consequently influences the model's ability to capture complex temporal patterns in the data. The key architectural and hyperparameter components are outlined as follows:

1. Stacked LSTM Layers: Each configuration involves three LSTM layers stacked to capture hierarchical temporal features. The parameter return sequences=True is used in the first two layers to output the complete sequence of hidden states, which is necessary for stacking multiple LSTM layers. The final LSTM layer outputs only the last hidden state (return sequences = False by default), which serves as input to the Dense output layer.
2. Memory Units (30, 50, 90):
    a) The 30-unit configuration is designed to evaluate a lighter model with reduced computational demand, appropriate for scenarios with limited data or low complexity.
    b) The 50-unit configuration represents a balanced architecture, offering sufficient capacity for capturing temporal dependencies while reducing the risk of overfitting.
    c) The 90-unit configuration investigates a higher-capacity model, intended to capture more intricate patterns, although it may incur higher computational cost and potential overfitting if not adequately regularized.
4. Input Shape: The input shape = (1,1) signifies that the model receives sequences of length 1 with one feature per time step, suitable for univariate time series input. Although uncommon for LSTM networks, this setting was consistent with the structure of the experimental dataset and time-step configuration employed during training.
5. Dense Output Layer: A single-unit Dense layer is incorporated to produce the final output, corresponding to the predicted rainfall or streamflow value at the subsequent time step.
6. Optimizer: The Adam optimizer is chosen due to its adaptive learning rate and efficiency in training deep learning models.

This experimental setup allowed the study to evaluate the effect of different memory unit sizes on prediction accuracy. The findings contributed to understanding the trade-offs among model complexity, training duration, and forecasting accuracy in the context of monsoon-influenced hydrological data.

Hyperparameter tuning was performed to optimize performance, testing configurations with 30, 50, and 90 memory units over 81 epochs, as shown in Figures 6 to 8. It is shown that the 50-unit configuration produced the best fit for rainfall prediction, with minimal deviation between actual and predicted values. The LSTM model's architecture and tuning procedure are key components of the forecasting model stage, as shown in Figure 1, where it is positioned alongside ARIMA and FB Prophet for comparative evaluation.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1, 50)	10400
lstm_1 (LSTM)	(None, 1, 50)	20200
lstm_2 (LSTM)	(None, 50)	20200
dense (Dense)	(None, 1)	51

Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0

Figure 6. LSTM parameter with 50 memory units.

Layer (type)	Output Shape	Param #
lstm_12 (LSTM)	(None, 1, 30)	3840
lstm_13 (LSTM)	(None, 1, 30)	7320
lstm_14 (LSTM)	(None, 30)	7320
dense_4 (Dense)	(None, 1)	31

Total params: 18,511
Trainable params: 18,511
Non-trainable params: 0

Figure 7. LSTM parameter with 30 memory units.

Layer (type)	Output Shape	Param #
lstm_24 (LSTM)	(None, 1, 90)	33120
lstm_25 (LSTM)	(None, 1, 90)	65160
lstm_26 (LSTM)	(None, 90)	65160
dense_8 (Dense)	(None, 1)	91

Total params: 163,531
Trainable params: 163,531
Non-trainable params: 0

Figure 8. LSTM parameter with 90 memory units.

4.2 ARIMA

As a traditional statistical model, ARIMA requires the time series to be stationary, necessitating the removal of trends and seasonality. This is achieved by first differencing, transforming non-stationary data into a stationary format as shown in Figure 9. This transformation is a critical prerequisite for applying the ARIMA model, which assumes that the input time series exhibits stationarity so that statistical properties such as mean, variance, and autocorrelation structure remain constant over time.

Figure 9(a) shows the original time series data, displaying clear trends and strong seasonal variations, which indicate non-stationarity. These characteristics violate the fundamental assumptions of the ARIMA model, potentially resulting in inaccurate parameter estimation and unreliable forecasting performance. To overcome this issue, transformations such as differencing and seasonal decomposition are applied to eliminate trends and seasonal effects. The result of this transformation is illustrated in Figure 9 (b), where the adjusted series exhibits consistent variance and lacks long-term trends, thereby satisfying the conditions of stationarity. This stationarity allows the ARIMA model to effectively represent the underlying stochastic process and improve prediction accuracy.

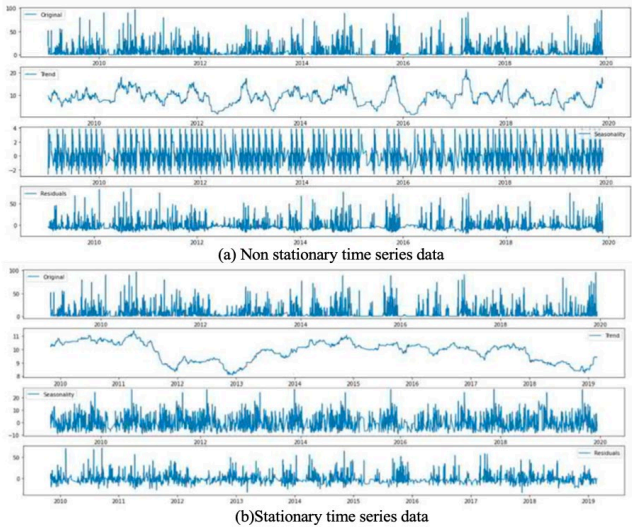


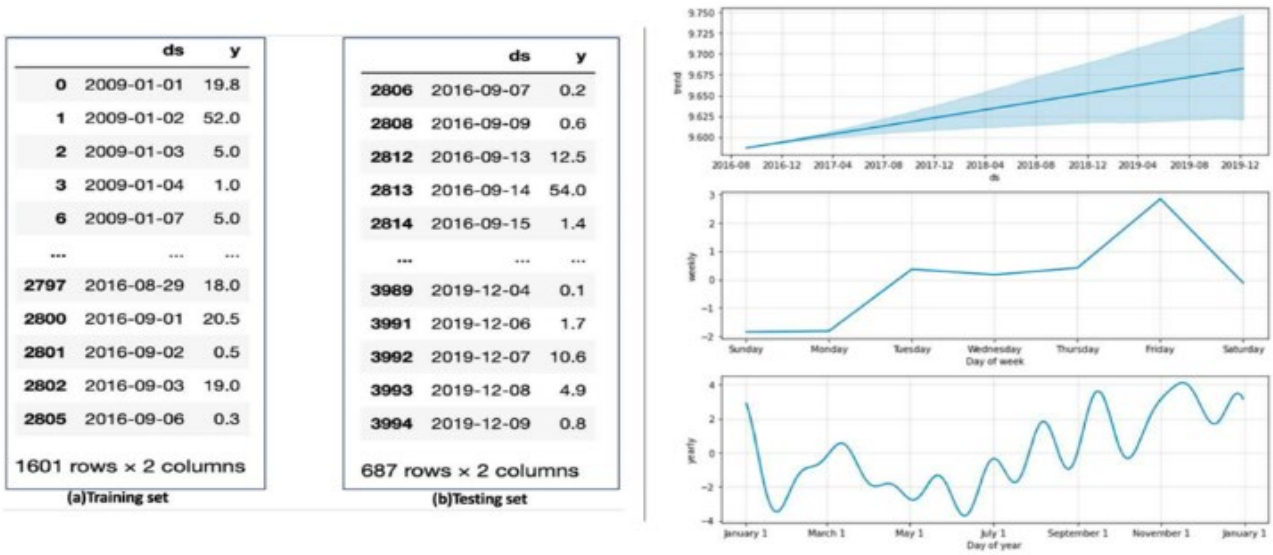
Figure 9. Transformation of (a)non-stationary time series data to (b)stationary time series.

In the context of hydrological time series forecasting, such as rainfall and streamflow, the transformation to stationarity is particularly important because the data are often influenced by seasonal climate patterns and long-term climatic changes. By ensuring the stationarity of the input series, the ARIMA model can be reliably used to capture temporal dependencies and enhance forecasting accuracy. The ARIMA model is then fitted to the training dataset, using 70% of the data for training and 30% for testing. The model's predictions are subsequently validated against actual observations. Due to its simplicity and effectiveness in handling linear and seasonal patterns, ARIMA remains a fundamental part of the forecasting framework and serves as a benchmark for comparison with other predictive models in this study.

4.3 FB Prophet

This automated forecasting tool was applied to the same dataset, with date and rainfall columns renamed to 'ds' and 'y', respectively (Figure 10). FB Prophet modelled trends, seasonality, and uncertainty, generating forecasts with upper and lower bounds. Figure 10 further demonstrated its ability to capture yearly trends and seasonal peaks, aligning with the monsoon season in Kelantan. The model's user-friendly nature and emphasis on seasonality are highlighted, where it is positioned as part of the forecasting model suite. The FB Prophet model reveals significant information that rainfall intensity increases yearly. Additionally, it shows that the peak value starts from September to January, which corresponds to the monsoon season in Kelantan state.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)



	ds	y
0	2009-01-01	19.8
1	2009-01-02	52.0
2	2009-01-03	5.0
3	2009-01-04	1.0
6	2009-01-07	5.0
...	...	...
2797	2016-08-29	18.0
2800	2016-09-01	20.5
2801	2016-09-02	0.5
2802	2016-09-03	19.0
2805	2016-09-06	0.3

1601 rows × 2 columns

(a)Training set

	ds	y
2806	2016-09-07	0.2
2808	2016-09-09	0.6
2812	2016-09-13	12.5
2813	2016-09-14	54.0
2814	2016-09-15	1.4
...	...	...
3989	2019-12-04	0.1
3991	2019-12-06	1.7
3992	2019-12-07	10.6
3993	2019-12-08	4.9
3994	2019-12-09	0.8

687 rows × 2 columns

(b)Testing set

Figure 10. Dataset (a) training set, (b) testing set and FB Prophet model prediction of trend, weakly and yearly.

4.4 Evaluation Metrics

Model performance was evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), comparing predictions against actual values across all stations. The RMSE is a standard statistical metric used to measure model performance and evaluate the accuracy of regression results, as defined by Equation (1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (1)$$

where:

- n denotes the total number of observations in the time series,
- y_i represents the actual observed value (rainfall or streamflow) at time step i,
- $\hat{y}_i$ denotes the predicted value at time step i generated by the forecasting model.

This formulation calculates the square root of the mean of the squared differences between predicted and actual values. A lower RMSE indicates a model with greater predictive accuracy. RMSE is especially effective in assessing the performance of forecasting models across multiple stations. The MAE is the average of the absolute value of all individual samples' deviations from the arithmetic mean. The MAE equation is shown in Equation (2).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (2)$$

where:

- n is the total number of observations in the time series,
- y_i is the actual observed value (e.g., rainfall or streamflow) at time step i,
- $\hat{y}_i$ is the corresponding predicted value generated by the forecasting model,
- $|\hat{y}_i - y_i|$ denotes the absolute error at each time step.

These metrics compared predicted values against actual observations across the selected rainfall and streamflow stations, as detailed in the Results and Discussion section. The evaluation procedure, depicted in Table 1 under "Model evaluation," utilized RMSE and MAE to determine the most accurate model for deployment, ensuring a reliable assessment of forecasting capabilities.

5. Results and discussion

5.1 Rainfall Prediction

The performance of the Long Short-Term Memory (LSTM) model for rainfall prediction was evaluated across five stations, as summarized in Table 1 below. The LSTM configuration with 90 memory units achieved the lowest Root Mean Square Error (RMSE) of 9.99 and Mean Absolute Error (MAE) of 9.88 at station 4614001, outperforming configurations with 50 memory units (RMSE 22.47 at 5722057) and 30 memory units (RMSE 12.05 at 4922001). Figure 11 illustrates the actual versus predicted rainfall values, with the 50-unit configuration demonstrating the best overall fit due to minimal deviation between observed and predicted data. This superior performance corresponds with

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

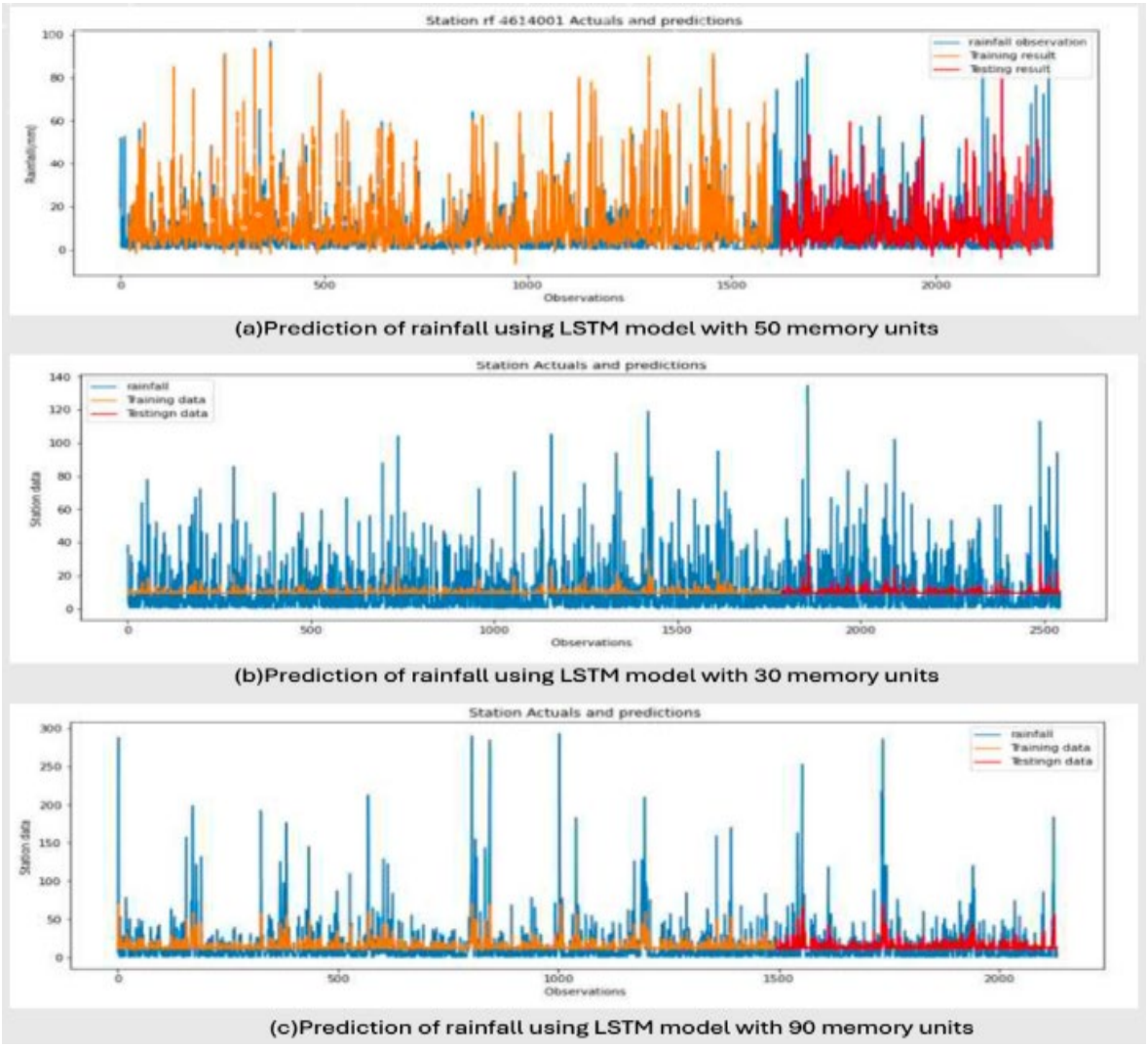


Figure 11. Prediction of rainfall using LSTM model with (a) 50 memory units, (b) 30 memory units, and (c) 90 memory units.

LSTM's capacity to capture non-linear, long-term dependencies in rainfall time series, as demonstrated by its structural flexibility.

The graph in Figure 11 illustrates the actual rainfall observations (blue bars), training predictions (orange bars), and testing predictions (red bars) against a time series index on the x-axis and rainfall in millimetres on the y-axis. The blue bars exhibit significant variability, with periodic peaks representative of Malaysia's monsoon-driven rainfall, particularly during the northeast monsoon (October–January) in Kelantan, reflecting the non-linear and seasonal characteristics of the data. The orange bars, covering the training period (approximately observations 1–1,400, or 70% of the data from 2009 to mid-2016), closely align with the blue bars, demonstrating that the LSTM model, likely configured with 30 memory units, 50 memory units, and 90 memory units as detailed in Table 1 and the model parameters representation in Figures 6 to 8, effectively captured the long-term dependencies and non-linear patterns during training. This alignment supports the model's low RMSE of 10.66 for rainfall prediction, as reported in Table 1. The model demonstrates strong predictive performance, highlighting its suitability for complex hydrological time series.

In the testing phase (red bars, from approximately observation 1,400 to 2,000, or mid-2016 to 2019), the predictions show noticeable deviations from the actual observations, particularly during peak rainfall events, where the red bars either overestimate or underestimate the blue bars, indicating challenges in generalizing to unseen data. These discrepancies highlight the limitations of LSTM in handling abrupt changes or extreme monsoon events, potentially due to the risks of overfitting noted in higher-memory configurations (e.g., 90 units) or the need for further hyperparameter tuning, as discussed in the methodology. Compared to ARIMA (RMSE 17.00) and FB Prophet (RMSE 17.01), which showed higher errors for rainfall at observation station 4614001 in Table 1, the LSTM predictions remain closer to actual values, reinforcing its effectiveness for non-linear rainfall patterns. However, the testing results suggest opportunities for improving model reliability, possibly through hybrid approaches combining LSTM's deep learning capabilities

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

Table 1. Model performance evaluation for rainfall data.

RF stations	Models	RMSE	MAE	Models	RMSE	MAE	Models	RMSE	MAE
Rf 4614001	LSTM (50)	10.763046	8.767914	LSTM (30)	11.977072	11.837367	LSTM (90)	9.991485	9.886242
	FB PROPHET	14.230670	9.83779	FB PROPHET	14.230670	9.83779	FB PROPHET	14.230670	9.83779
	ARIMA	14.157034	9.761772	ARIMA	14.157034	9.761772	ARIMA	14.157034	9.761772
Rf 4922001	LSTM (50)	12.38371	11.765164	LSTM (30)	12.055913	12.01011	LSTM (90)	12.847241	12.804568
	FB PROPHET	16.228314	10.909548	FB PROPHET	16.228314	10.909548	FB PROPHET	16.228314	10.909548
	ARIMA	16.26	11.474213	ARIMA	16.26	11.474213	ARIMA	16.26	11.474213
Rf 5120025	LSTM (50)	10.660235	10.313289	LSTM (30)	12.150119	11.709419	LSTM (90)	10.915791	10.896921
	FB PROPHET	17.014614	12.292773	FB PROPHET	17.014614	12.292773	FB PROPHET	17.014614	12.292773
	ARIMA	17.002148	12.16735	ARIMA	17.002148	12.16735	ARIMA	17.002148	12.16735
Rf 5216001	LSTM (50)	12.15431	11.680964	LSTM (30)	12.181483	12.125944	LSTM (90)	9.884912	9.87605
	FB PROPHET	16.849905	12.35696	FB PROPHET	16.849905	12.35696	FB PROPHET	16.849905	12.35696
	ARIMA	16.486853	11.465422	ARIMA	16.486853	11.465422	ARIMA	16.486853	11.465422
Rf 5722057	LSTM (50)	22.476722	15.601898	LSTM (30)	22.685858	20.437183	LSTM (90)	26.708829	26.22012
	FB PROPHET	27.811985	16.369283	FB PROPHET	27.811985	16.369283	FB PROPHET	27.811985	16.369283
	ARIMA	27.974476	16.445802	ARIMA	27.974476	16.445802	ARIMA	27.974476	16.445802

Table 2. Model performance of stream-flow prediction.

SF stations	Models	RMSE	MAE	Models	RMSE	MAE	Model	RMSE	MAE
5120401	LSTM (50)	248.238136	129.99157	LSTM (90)	283.860012	100	LSTM (30)	325.5082	187.3909
	FB PROPHET	444.404245	341.227676	FB PROPHET	444.404245	341.227676	FB PROPHET	444.404245	341.227676
	ARIMA	289.280855	159.50526	ARIMA	289.280855	159.50526	ARIMA	289.280855	159.50526
5320443	LSTM (50)	399.342057	316.356019	LSTM (90)	364.84575	251.740427	LSTM (30)	378.7617	190.1653
	FB PROPHET	480.669985	356.04208	FB PROPHET	480.669985	356.04208	FB PROPHET	480.669985	356.04208
	ARIMA	362.212159	230.521021	ARIMA	362.212159	230.521021	ARIMA	362.212159	230.521021
5222452	LSTM (50)	223.994856	171.491614	LSTM (90)	269.623867	147.057549	LSTM (30)	194.2497	121.3121
	FB PROPHET	672.831293	601.126042	FB PROPHET	672.831293	601.126042	FB PROPHET	672.831293	601.126042
	ARIMA	213.642699	118.80583	ARIMA	213.642699	118.80583	ARIMA	213.642699	118.80583
5419401	LSTM (50)	353.443639	184.719494	LSTM (90)	279.058315	126.267982	LSTM (30)	312.3706	145.2426
	FB PROPHET	557.17164	403.714354	FB PROPHET	557.17164	403.714354	FB PROPHET	557.17164	403.714354
	ARIMA	426.60987	235.975076	ARIMA	426.60987	235.975076	ARIMA	426.60987	235.975076
5621401	LSTM (50)	41.892632	16.621564	LSTM (90)	111.743602	103.804901	LSTM (30)	65.52391	41.23494
	FB PROPHET	740.342201	645.729629	FB PROPHET	740.342201	645.729629	FB PROPHET	740.342201	645.729629
	ARIMA	60.034331	17.058709	ARIMA	60.034331	17.058709	ARIMA	60.034331	17.058709
5721442	LSTM (50)	745.914703	394.267734	LSTM (90)	713.494153	407.799559	LSTM (30)	608.2709	365.5414
	FB PROPHET	592.60287	403.356575	FB PROPHET	592.60287	403.356575	FB PROPHET	592.60287	403.356575
	ARIMA	571.961758	311.628273	ARIMA	571.961758	311.628273	ARIMA	571.961758	311.628273

with ARIMA's linear focus or FB Prophet's seasonality handling, as proposed in the discussion, to enhance forecasting accuracy for Malaysia's monsoon-influenced hydrological challenges.

Further discussion on the LSTM architectures tested in this study is detailed in Figures 6-8, providing insight into the model's structure and parameter optimization.

1. For the 30-memory-unit configuration shown in Figure 7, the architecture consisted of three LSTM layers followed by a dense layer, with a total of 18,511 trainable parameters. Specifically, the layers included: an initial LSTM layer with an output shape of (None, 1, 30) and 3,840 parameters, a second LSTM layer with the same output shape and 7,320 parameters, a third LSTM layer with an output shape of (None, 30) and 7,320 parameters, and a dense layer with an output shape of (None, 1) and 31 parameters. This configuration, while computationally lighter, showed higher errors in rainfall prediction, indicating its limitations in capturing complex patterns compared to higher-memory configurations.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

2. For the 50-memory-unit configuration presented in Figure 6, the architecture consisted of three LSTM layers and a dense layer, totalling 50,851 trainable parameters. The layers were arranged as follows: an initial LSTM layer with an output shape of (None, 1, 50) and 10,400 parameters, a second LSTM layer with an output shape of (None, 1, 50) and 20,200 parameters, a third LSTM layer with an output shape of (None, 50) and 20,200 parameters, and a dense layer with an output shape of (None, 1) and 51 parameters. This setup provided the best fit for rainfall prediction, with an RMSE of 10.76, demonstrating an optimal balance of model complexity and predictive accuracy for nonlinear rainfall data.

3. The 90-memory-unit configuration, illustrated in Figure 8, featured three LSTM layers and a dense layer, totalling 163,531 trainable parameters. The layers included: an initial LSTM layer with an output shape of (None, 1, 90) and 33,120 parameters, a second LSTM layer with the same output shape and 65,160 parameters, a third LSTM layer with an output shape of (None, 90) and 65,160 parameters, and a dense layer with an output shape of (None, 1) and 91 parameters. This configuration, while computationally demanding, achieved the lowest error metrics at certain stations, such as 4614001, but displayed higher errors elsewhere, suggesting possible overfitting or limited generalizability for some datasets.

These three LSTM architecture configurations, optimized using the Adam algorithm, highlight LSTM's adaptability in hydrological time series forecasting. Overall, LSTM consistently outperformed ARIMA and FB Prophet in rainfall prediction, as indicated in Table 1. These results verify the effectiveness of LSTM in modelling the non-linear, seasonal patterns of rainfall, consistent with findings from Ghimire et al. (2021) and Xu et al. (2020).

5.2 Stream-flow Prediction

For streamflow prediction, the results across six stations indicated that LSTM with 50 memory units performed best at stations 5120401 (RMSE = 248.23), 5419401 (RMSE = 353.44), and 5621401 (RMSE = 41.89), surpassing other configurations.

However, ARIMA achieved lower RMSE values at stations 5310443 (362.21), 5222452 (213.64), and 5721442 (571.96), demonstrating its strength in handling linear and autocorrelated data. FB Prophet showed higher errors across all stations, indicating limited suitability for streamflow forecasting.

The LSTM architectures described earlier were also evaluated for streamflow, but their performance varied due to the linear characteristics of streamflow data. The 50-unit configuration, with its balanced parameter count and design, provided competitive results, while the 90-unit setup occasionally underperformed, possibly due to overfitting. ARIMA's superior performance in streamflow prediction reflects its effectiveness with stationary, linear time series, consistent with Adnan et al. (2017).

5.3 Model Comparison

The comparative analysis presented in Tables 1 and 2 highlights the distinct strengths and limitations of each forecasting model. LSTM demonstrated strong performance in capturing non-linear rainfall patterns, particularly with the 50-memory-unit configuration (50,851 parameters), which achieved optimal accuracy across several rainfall stations. In contrast, ARIMA performed best for streamflow forecasting, effectively modelling linear and seasonal patterns with lower RMSE values across key stations. FB Prophet consistently underperformed for both rainfall and streamflow, showing limited suitability for hydrological time series due to its generalized structure, though its ease of automation and seasonality modelling may benefit specific, less complex applications, as mentioned by Sulasikin et al. (2021).

An investigation into the LSTM configurations revealed that model performance is highly sensitive to memory unit size. The 90-unit setup, while yielding improved accuracy at some stations, showed high computational costs (163,531 parameters) and signs of overfitting. The 30-unit model (18,511 parameters), on the other hand, lacked sufficient capacity to capture complex temporal dependencies, resulting in underfitting. These findings highlighted the importance of architectural flexibility and hyperparameter tuning in LSTM modelling. The trade-offs between complexity, accuracy, and computational efficiency suggest the need for hybrid modelling approaches that combine the deep learning strengths of LSTM with the statistical robustness of ARIMA for improved forecasting in hydrological applications.

6. Conclusion and recommendation

This study compared LSTM, ARIMA, and FB Prophet for time series forecasting of rainfall and streamflow using DID Malaysia data (2009–2019). LSTM performed best for rainfall prediction, achieving an RMSE of 10.76 and an MAE of 8.76 at station 4614001, using a 50-memory-unit architecture (50,851 parameters) that balanced complexity and accuracy. ARIMA outperformed others for streamflow, achieving lower RMSE values at stations 5310443 (362.21), 5222452 (213.64), and 5721442 (571.96) by effectively modeling linear trends. FB Prophet demonstrated limited suitability for both variables due to its generalized structure. These results highlight the importance of selecting forecasting models based on data characteristics to support informed water resource management. Future work should refine FB Prophet by adjusting changepoint sensitivity to better capture abrupt hydrological shifts and explore hybrid models combining its interpretability with the predictive capacity of deep learning or statistical methods to improve accuracy and climate resilience.

7. Acknowledgement

The authors would like to thank the Department of Irrigation and Drainage Malaysia (DID) for providing the data to conduct this study. We would also like to thank IIUM and MJIIT-UTM for the sponsorship during this study.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

8. References

Abdullah, M. S., Selamat, A., Do, N. Q., Abu, M. A., Lim, K. C., & Krejcar, O. (2024, December). Rainfall Prediction based on Time-series Data and Temporal Convolutional Network for Flood Forecasting in Malaysia. In 2024 IEEE International Conference on Computing (ICOCO) (pp. 404-409). IEEE.

Adnan, R. M., Yuan, X., Kisi, O., & Curtef, V. (2017). Application of time series models for streamflow forecasting. Civil and Environmental Research, 9(3), 56-63.

Agaj, T., Budka, A., Janicka, E., & Bytyqi, V. (2024). Using ARIMA and ETS models for forecasting water level changes for sustainable environmental management. Scientific Reports, 14(1), 22444.

Aguilera, H., Guardiola-Albert, C., Naranjo-Fernández, N., & Kohfahl, C. (2019). Towards flexible groundwater-level prediction for adaptive water management: using Facebook's Prophet forecasting approach. Hydrological Sciences Journal, 64(12), 1504-1518.

Bakhshi Ostadkalayeh, F., Moradi, S., Asadi, A., Moghaddam Nia, A., & Taheri, S. (2023). Performance improvement of LSTM-based deep learning model for streamflow forecasting using Kalman filtering. Water Resources Management, 37(8), 3111-3127.

Ghimire, S., Yaseen, Z. M., Farooque, A. A., Deo, R. C., Zhang, J., & Tao, X. (2021). Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Scientific Reports, 11(1), 17497.

Hossain, M. M., Anwar, A. F., Garg, N., Prakash, M., & Bari, M. (2022). Monthly rainfall prediction at catchment level with the Facebook Prophet model using observed and CMIP5 decadal data. Hydrology, 9(6), 111.

Ibrahim, N., Ahmad, N., Jan, N.A.M., Zainudin, Z., Jamil, N.S. and Azlan, A., 2024, September. Comparative Analysis of ARIMA and LSTM Approaches for Monthly River Flow Forecasting in Terengganu. In 2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 1-6). IEEE.

Khairudin, N. B. M., Mustapha, N. B., Aris, T. N. B. M., & Zolkepli, M. B. (2020, September). Comparison of machine learning models for rainfall forecasting. In 2020 international conference on computer science and its application in agriculture (ICOSICA) (pp. 1-5). IEEE.

Kobiela, D., Krefta, D., Król, W., & Weichbroth, P. (2022). ARIMA vs LSTM on NASDAQ stock exchange data. Procedia Computer Science, 207, 3836-3845.

Kumar, V., Kedam, N., Kişi, Ö., Alsulamy, S., Khedher, K., & Salem, M. A. (2024). A Comparative Study of Machine Learning Models for Daily and Weekly Rainfall Forecasting. Water Resources Management. https://doi.org/10.1007/s11269-024-03969-8Latif, S. D., Ahmed, A. N., Sathiamurthy, E., Huang, Y. F., & El-Shafie, A. (2021). Evaluation of deep learning algorithm for inflow forecasting: a case study of Durian Tunggal Reservoir, Peninsular Malaysia. Natural Hazards, 109(1), 351-369.

Latif, S. D., Ahmed, A. N., Sathiamurthy, E., Huang, Y. F., & El-Shafie, A. (2021). Evaluation of deep learning algorithm for inflow forecasting: a case study of Durian Tunggal Reservoir, Peninsular Malaysia. Natural Hazards, 109(1), 351-369.

Ramlan, S. Z., & Mohd Deni, S. (2021). Rainfall prediction in flood prone area using deep learning approach. In Soft Computing in Data Science: 6th International Conference, SCDS 2021, Virtual Event, November 2–3, 2021, Proceedings 6 (pp. 71-88). Springer Singapore.

Sahoo, A., Parida, S. S., Samantaray, S., & Satapathy, D. P. (2024). Daily flow discharge prediction using integrated methodology based on LSTM models: Case study in Brahmani-Baitarani basin. HydroResearch, 7, 272-284.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. Time series analysis and its applications: with R examples, 75-163.

Sulasikin, A., Nugraha, Y., Kanggrawan, J. I., & Suherman, A. L. (2021, August). Monthly rainfall prediction using the Facebook Prophet model for flood mitigation in Central Jakarta. In 2021 International Conference on ICT for Smart Society (ICISS) (pp. 1-5). IEEE.

Sivaramakrishnan, S., Fernandez, T. F., Babukarthik, R. G., & Premalatha, S. (2022). Forecasting time series data using arima and facebook prophet models. In Big data management in Sensing (pp. 47-59). River Publishers.

Xu, Wei, Yanan Jiang, Xiaoli Zhang, Yi Li, Run Zhang, and Guangtao Fu. "Using long short-term memory networks for river flow prediction." Hydrology Research 51, no. 6 (2020): 1358-1376.

Xu, D., Zhang, Q., Ding, Y., & Zhang, D. (2022). Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environmental Science and Pollution Research, 29(3), 4128-4144.

DOI: https://doi.org/10.22452/mjs.vol45sp2.6
Malaysian Journal of Science 45 (Special Issue 2): 55-65 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special issue

Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

https://mjs.um.edu.my

Deterministic and Stochastic Gompertz Prey-Predator Model for the Effect of Thymoquinone on Oral Squamous Carcinoma Cell Lines (HSC-4)

Shabana Tabassum[1a], Norhayati Rosli[2b*], Wastuti Hidayati Suriyah[3c], and Solachuddin Jauhari Ichwan Arief[4d]

Abstract: Oral squamous cell carcinoma (OSCC) is one of the most common malignancies affecting the head and neck region. Thymoquinone (TQ), the primary bioactive compound of black seed (*Nigella sativa*), has shown antineoplastic properties in various tumor cell types. However, research on TQ effects in tumor cell models, particularly OSCC, remains at an early stage. TQ, a phytochemical with confirmed anticancer potential, has been shown to inhibit OSCC progression mainly through induction of apoptosis. In this study, a prey–predator stochastic Gompertz model is used to characterize the dynamics of the HSC-4 OSCC cell line in response to TQ treatment. The model is first formulated as an ordinary differential equation (ODE), then extended to a stochastic differential equation (SDE) to capture biological noise arising from intrinsic and extrinsic sources within the cellular environment. The Markov Chain Monte Carlo (MCMC) method is utilized to estimate the model's kinetic parameters, while fourth-order stochastic Runge-Kutta (SRK4) algorithms are applied for numerical simulation. Model outputs are validated against in vitro experimental data from TQ-treated HSC-4 cells. The deterministic framework is analyzed in terms of the positivity of solutions and equilibrium states, including the disease-free equilibrium (DFE), the treatment-failure equilibrium (TFE), and the successful treatment equilibrium (SFE). The root mean-square error (RMSE) is computed to assess the consistency between simulated outcomes and experimental observations. The findings demonstrate the model's capability in capturing and explaining the dynamic behavior of HSC-4 cells under TQ treatment.

Keywords: Deterministic model, stochastic model, thymoquinone, oral cancer.

1. Introduction

The five-year survival rate for oral cancer patients was 69%. The incidence rate of new oral cavity and pharynx cancer was 11.5 per 100,000 men and women annually. The mortality rate stood at 2.6 per 100,000 men and women per year. These figures are age-adjusted, derived from cases reported between 2017 and 2021 and deaths recorded from 2018 to 2022 (Global Cancer Observatory, 2025). The sources and methods used in compiling the Global Cancer Observatory: Cancer Today GLOBOCAN estimates have been documented and described online at the Global Cancer Observatory. According to the National Institutes of Health (NIH), the estimated new cases of oral cavity cancer were 59,660 from 2021 to April 2025, and the estimated deaths were 12,770 for the same period (National Cancer Institute, 2025). Cancer spreads by avoiding immune cell destruction, evading growth suppressors, and dysregulating proliferative signaling. Adjuvant chemotherapy and radiation therapy are administered as needed after surgical removal, which is the cornerstone of the current treatment strategy. Many treatment plans have also been developed using targeted approaches to specific molecular markers, such as breast cancers linked to breast cancer tissues. According to Jallah et al. (2023), treatments that target multiple cancer progression pathways may be a useful adjunct to cancer care. Cancer presents itself in a range of ways, including through irregular proliferative signaling, evasion of growth suppressors, and avoidance of immune cell destruction. Surgical excision, adjuvant chemotherapy, and radiation therapy, where necessary, are the mainstays of the current therapeutic approach. Targeted methods to specific molecular targets, like breast malignancies linked to the Breast Cancer gene (BRCA), have also demonstrated considerable therapeutic success. Multiple oncogenic pathway-targeting treatments could be a valuable addition to cancer therapy. The International Classification of Diseases for Oncology, Third Edition, was used to classify all cancer cases; however, the International Classification of Childhood Cancer was used for carcinomas that occurred in children and young adults (Steliarova-Foucher et al., 2005; Steliarova-Foucher et al., 2017).

Authors information:
[a]Mathematics and Statistics Missouri University of Science and Technology. 202 Rolla Building, 400 W. 12th St., Rolla, MO 65409-0020, USA. E-mail: stabassum@mst.edu[1]
[b]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, MALAYSIA, E-mail: norhayati@umpsa.edu.my[2]
[c]Faculty of Dentistry, Universitas YARSI, Jl. Let. Jend. Suprapto Kav. 13. Jakarta 10510, INDONESIA. E-mail: wastuti.hidayati@yarsi.ac.id[3]
[d]PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, BRUNEI. E-mail: solachuddin.ichwan@ubd.edu.bn[4]

*Corresponding Author: norhayati@umpsa.edu.my

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

Tumors of the lip, oral cavity, and throat are all included in the phrase "oral cancer." The survival rate for oral cancer is about 80% if the disease is detected early (Steliarova-Foucher et al., 2005). Treatment options for oral cancer include surgery, chemoradiotherapy, and targeted therapy. However, combined therapy strategies have unsatisfactory outcomes in advanced illness because chemoradiotherapy can cause harmful effects on healthy cells, and cancer can recur. Recently, cancer treatment has evolved into targeted therapies that inhibit the apoptotic activity of cancer cells. The use of chemotherapy may be encouraged by an alternative strategy for reducing the growth of cancer. Using plant products, commonly referred to as nutritional supplements, is one strategy for advanced oral cancer care. These naturally occurring compounds are physiologically accessible, have several molecular targets for cancer chemoprevention, are cost-effective, and produce non-toxic effects (Steliarova-Foucher et al., 2005; Steliarova-Foucher et al., 2017). In the context of head and neck cancer, oral squamous cell carcinoma remains a highly challenging disease due to its high recurrence rate and relatively simple metastasis (González-Moles et al., 2022; Cadoni et al., 2017; Kasteb et al., 1990). Determining an effective tumor diagnostic and molecular therapeutic target is therefore essential. Through the suppression of serine/threonine kinase 1 (human) AKT and extracellular signal-regulated kinase signaling pathways, TQ reduces tumor angiogenesis and proliferation (Gali-Muhtasib et al., 2006).

2. Thymoquinone and cancer

For a while, TQ has been verified scientifically through the degradation of thymol with hydrogen peroxide. According to recent studies, black seed (Nigella sativa), also known as "Habbtul-barakah" in Arabic, is the most beneficial plant in Mediterranean and West Asian traditional medicine, and it can be used to target the progression of cancer. According to the Prophet's (PBUH) historical account, this plant has therapeutic potential (Gali-Muhtasib et al., 2006). The primary active compound, TQ, which is a strong anti-carcinogenic antioxidant and anti-mutagenic agent, is associated with the biological activities of Nigella sativa seeds (Tabassum et al., 2022). Due to its invasive properties, the oral cancer cell line Human Tongue Squamous Carcinoma Cell Line HSC-4 has been widely used to study metastatic squamous cell carcinoma (Steliarova-Foucher et al., 2005). For example, using oral cancer HSC-4 cell lines for personalized medical care, a recent study by Tabassum et al. (2024) examined the cytotoxic effects of TQ combined with cisplatin (CDDP). Even at lower doses (IC16.8 μM), they found that enhancing TQ with CDDP improves its carcinogenic cytotoxicity against HSC cell lines while reducing its adverse effects on healthy cells (Tabassum et al., 2024). Additionally, significant side effects of current treatment methods, such as surgery and chemoradiotherapy, remain major causes of treatment dissatisfaction. Local relapse and regional or systemic lesions are often the cause of death due to oral cancer (Alaufi et al., 2017).

The flowering plant black cumin (Nigella sativa) contains a phytochemical called TQ. Its anti-inflammatory, anti-tumor, anti-cancer, anti-neoplastic, anti-bacterial, anti-diabetic, antihypertensive, anti-parasitic, and antioxidant properties have been demonstrated in both in vitro and in vivo studies (Chen et al., 2011). Research on TQ's toxicity to normal cells or animal models remains limited. Despite the scarcity of studies on TQ's use in oral cancer, it has shown promising anticancer effects (Hannan et al., 2021). Using a mathematical model to depict the underlying mechanism can improve both quantitative and qualitative understanding of how HSC-4 responds to TQ. Both quantitative and qualitative knowledge can be applied in biomedical research. Ordinary differential equations are often used to develop mathematical models describing the proliferation of cancer cells under treatment (AbuKhader, 2013). However, cancer cell behavior is influenced by many uncontrollable factors, including stress, patient lifestyle, environmental exposures (such as pollution and radiation), genetic predisposition, immune system variability, and pre-existing chronic conditions. Deterministic models based on ordinary differential equations (ODEs) cannot capture the inherent variability and randomness related to these factors.

To address this, ODEs can be extended to stochastic differential equations (SDEs), which include random fluctuations and noise, providing a more realistic framework for modeling the complex and uncertain dynamics of oral cancer development and progression. To illustrate the behavior of oral cancer cell lines in the presence of TQ, we present a stochastic model in this study. We employ a predator-prey paradigm to illustrate the interaction between HSC-4 cell lines and TQ. Assuming that the HSC-4 growth rate follows the Gompertz law, the predator-prey model is modified by incorporating the Gompertz rule. This results in the development of the deterministic Gompertz predator-prey model. The SDE model is created by perturbing the deterministic model with uncontrollable factors, which are modeled by the Wiener process. Mathematical models of the HSC-4 cancer cell lines take the form of ODEs and SDEs in response to TQ. We then present the dynamics of the system regarding positive solutions and the existence of an equilibrium. The four-stage Stochastic Runge-Kutta method (SRK4) is applied for simulation, and parameter estimation is performed using the Markov Chain Monte Carlo (MCMC) method.

3. Mathematical model

Figure 1 depicts the schematic representation of the relationship between the cancer cell and TQ. Based on Figure 1, when TQ reacts with cancer cells, some cells are destroyed, while others continue to proliferate and form multiple cells. However,

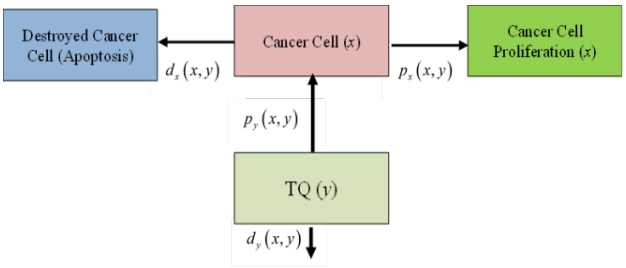


Figure 1. Schematic representation of TQ's relationship with cancerous cells.

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

the presence of TQ can slow down the proliferation of cancer cells by interfering with the growth factors.

The cancer cell regression and proliferation in the presence of TQ, shown in Figure 1, can be translated into the predator-prey type model of

$$\frac{dx}{dt} = p_x(x,y) - d_x(x,y)$$
$$\frac{dy}{dt} = p_y(x,y) - d_y(x,y) \quad (1)$$

where $p_x(x,y)$ is the cancer cell growth term, and $d_x(x,y)$ is referred to as the destruction of cancer cells. $p_y(x,y)$ is referred to as the proliferation of immune cells induced by TQ and $d_y(x,y)$ represents the inhibition of the TQ. Both $p_x(x,y)$ and $p_y(x,y)$ depend on the interaction between the cancer cell and TQ. In numerous problems of non-exponential growth of cancer cells, the Gompertz model is described as the kinetics cell proliferation (Emanuel & Emmanuel, 2000; Mazlan et al., 2017). For the specific functions $d_x(x,y)$, $p_y(x,y)$ and $d_y(x,y)$, the generic model (1) reduces to specific models derived by Kuznetsov et al. (1994) as follows

$$p_x(x,y) = ax - bx\ln x$$
$$d_x(x,y) = \alpha xy$$
$$p_y(x,y) = \frac{\rho xy}{g+x} \quad (1)$$
$$d_y(x,y) = mxy.$$

Substituting Eq. (2) in Eq. (1) leads to a system of ODEs

$$\frac{dx}{dt} = (ax - bx\ln x) - \alpha xy$$
$$\frac{dy}{dt} = \frac{\rho xy}{g+x} - mxy \quad (3)$$

In this model, parameter a is the intrinsic growth rate of the cancer cell and b reflects the rate of antiangiogenetic response. The attraction of TQ to cancer cells is captured through a Michaelis-Menten term, $\frac{\rho xy}{g+x}$, which introduces a saturation effect observed in biological uptake processes. Additionally, the term involving ρ represents the proliferation rate of TQ and g is a positive constant. The competitive interaction between cancer cells and TQ is expressed through mortality and inhibition terms, represented mathematically by the terms $-\alpha xy$ and $-mxy$, respectively. Numerous mathematical equations have been developed to predict the dynamics of cancer cell growth, with the assumption that proliferation constants remain time-invariant and are not significantly influenced by temporal changes (Norton, 1988; Hu et al., 2024).

The Gompertz model with stochastic parameters has been extensively researched because it can accurately represent the cell decelerations that occur during the growth of tumor cells (AbuKhader, 2013). A clinical data-stimulated stochastic model of Breast cancer, using the Gompertz equation, is employed for tumor development in Kazerouni et al. (2020), where tumor growth initially shows a similar pattern. The Gompertzian properties are modified before applying a random time-dependent process that causes it to become kinetically variable (Rana, 2020). Since the outcomes of the noisy phenomena affect biological structures, such as varying brain control over personal traits or muscular activity, like anxiety, which causes irregular variations, equation (1) should be extended to the stochastic form of SDEs (Rana, 2020; Ly et al., 2022). Consequently, the Wiener process must disturb the intrinsic growth rate parameters. Assuming that the intrinsic growth rate parameter varies according to noisy behavior, and randomness is incorporated into this, for example

$$a \to a + \sigma\frac{dW}{dt}$$

$$\rho \to \rho + \sigma\frac{dW}{dt}$$

here $\sigma > 0$ represents the rate of absorption for $t > 0$ is the Gaussian procedure with an increase in time h with mean zero and W_t is the is the Wiener process. The stochastic model of Gompertz is formulated as

$$dx = (ax - bx\ln x - \alpha xy)\,dt + \sigma x dW_t$$
$$dy = \left(\frac{\rho xy}{g+x} - mxy\right)dt + \sigma\left(\frac{xy}{g+x}\right)dW_t \quad (4)$$

4. Experimentation

HSC-4 cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 1% Penicillin-Streptomycin and 10% fetal bovine serum (Gibco, US). Cell proliferation was monitored over five days in serum-containing medium. Cells were seeded in a labeled 6-well plate at a density of 4 × 10^4 cells/mL.

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

Cells in plate one were collected after 24 hours on day 1. Before counting, cells were rinsed with PBS, trypsinized, and resuspended in an equal volume of medium. This procedure was repeated daily for continued analysis. To prepare stock solutions, TQ was dissolved in molecular-grade DMSO (100 mg/mL, Gibco, US). HSC-4 cells were then treated with various TQ concentrations for 0, 12, 24, 36, and 48 hours to assess its in vitro toxicity on oral cancer cells. All experiments were conducted in triplicate for reliability. Results demonstrated that TQ significantly reduced the viability of oral cancer cells while having a minimal effect on normal cells. TQ exhibited toxicity in a cell-type-dependent manner, with its effects on HSC-4 cells increasing with concentration and exposure time. The IC50 values at 0, 12, 24, 36, and 48 hours remained constant at 0.0012 mg/mL. Figure 2 illustrates HSC-4 cells under control conditions, before treatment initiation. In Figure 3, the effects of TQ treatment on the cancer cell line are clearly apparent. After 48 hours of TQ exposure, a significant reduction in cancer cell count was observed, confirming the anticancer potential of TQ.

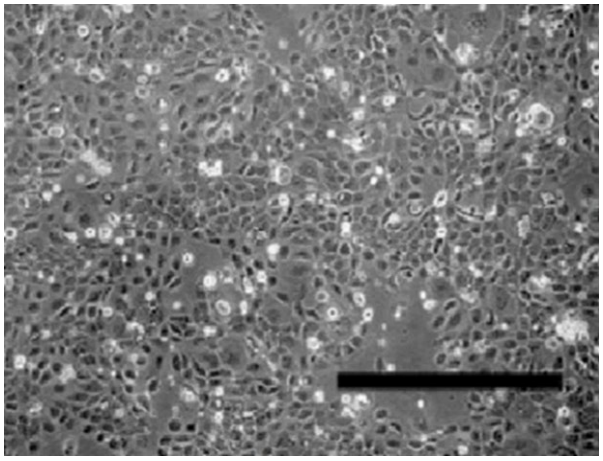


Figure 1. HSC-4 cell line (control).

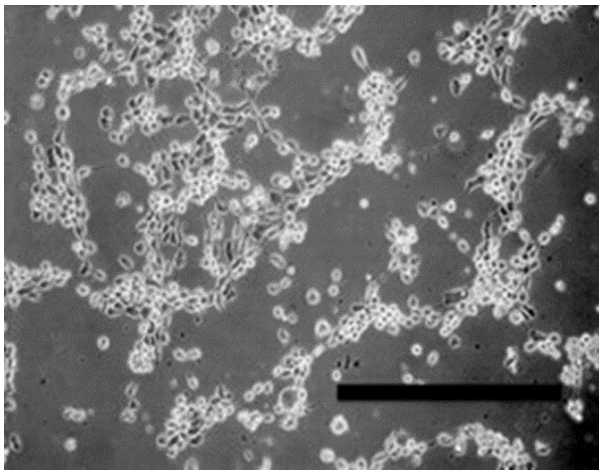


Figure 2. HSC-4 cell line after 24 hours of treatment.

5. Model analysis

We have established a mathematical model and in vitro experimentation that capture the TQ interaction with cancer cells. In this section, the characteristics of the deterministic model (3) in terms of equilibrium stability and positive solutions are investigated. We show that all solutions of system (3) with the initial conditions remain positive.

Theorem 1: Let $x(t)$ and $y(t)$ be the solutions to the system (3) with initial conditions $x(0)>0,\ y(0)>0.$ Then for all $t>0,$ the solutions of (3) remain positive, i.e., $x(t)>0,\ y(t)>0.$

Proof: System of ODEs (3) can be rewritten in a separable form of

$$\begin{aligned}\frac{dx}{x}&=(a-b\ln x-\alpha y)\,dt\\ \frac{dy}{y}&=\left(\frac{\rho x}{g+x}-mx\right)dy\end{aligned}\qquad(5)$$

By integrating both sides of system (5) from 0 to *t*, yield

$$x(t)=x(0)\exp\left[\int_0^t\left(a-b\ln x(s)-\alpha y(s)\right)ds\right]$$

$$y(t)=y(0)\exp\left[\int_0^t\left(\frac{\rho x(s)}{g+x(s)}-mx(s)\right)ds\right]$$

Given that $x(0)>0,\ y(0)>0$ and the exponential function is always positive, then it follows $x(t)>0,\ y(t)>0,$ for all $t>0.$ Thus, both $x(t)$ and $y(t)$ remain positive for all $t>0,$ which completes the proof.

Next, we identify the disease-free equilibrium (DFE), treatment failure equilibrium (TFE), and successful treatment equilibrium (SFE) points. After a period of treatment, it is assumed that the cancer cells may not increase but remain stable at the equilibrium points. To determine the equilibrium points, the right-hand side of the system (3) is set to zero such that

$$\begin{cases}(ax-bx\ln x)-\alpha xy=0\\ \dfrac{\rho xy}{g+x}-mxy-ey=0\end{cases}\qquad(2)$$

For disease-free equilibrium points, i.e., in the absence of tumor, it is clear that the cancer cells and the TQ, $y=0.$ Hence, the trivial is $E_0^*=(0,0).$ All biologically feasible equilibria admitted by system (3) and study the dynamics of the system

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

around the treatment failure equilibrium (TFE) and successful treatment equilibrium (SFE) points are investigated. The treatment failure equilibrium occurs when the treatment level $y=0,$ implying the treatment does not affect the dynamics of the system. Using equation (6) yields either $x=0$ or $\ln(x)=\frac{a}{b},$ which implies $x=e^{\frac{a}{b}}$. Therefore, the treatment failure equilibrium is $E_1^*=\left(e^{\frac{a}{b}},0\right)$. The equilibrium for success treatment occurs when the treatment level of TQ, $y>0.$ From equation (6), we have $y=\frac{a-b\ln(x)}{\alpha}.$ Substituting y into the second equilibrium in equation (6) yields $y\left(\frac{\rho x}{g+x}-mx\right)=0.$ Since $y\neq 0,$ the term $\frac{\rho x}{g+x}-mx=0,$ i.e., $x=\frac{\rho}{m}-g>0,$ i.e. $\rho>gm.$ Hence, the success treatment equilibrium is

$$E_2^*=\left(\frac{\rho}{m}-g,\frac{a-b\ln\left(\frac{\rho}{m}-g\right)}{\alpha}\right),\quad \rho>gm.$$

Let $x(t)$ and $y(t)$ are solutions for system (3), and the system admits:

(i) Disease-free equilibrium, $E_0^*=(0,0).$

(ii) Treatment-failure equilibrium, $E_1^*=\left(e^{\frac{a}{b}},0\right).$

(iii) Successful treatment equilibrium,

$$E_2^*=\left(\frac{\rho}{m}-g,\frac{a-b\ln\left(\frac{\rho}{m}-g\right)}{\alpha}\right),\quad \frac{\rho}{m}-g>0.$$

6. Results and discussion

Figures 4 (a) – (c) show the dynamics in time and phase space plot of the Gompertz prey-predator model (3) for cancer cells (x) in the presence of TQ (y). The cancer cells tend to approach the equilibrium point of 0.19, as indicated by the red line graph in Figure 4(a). Meanwhile, the concentration of TQ tends toward 0.040, as indicated by the blue line in Figures 4(a) and 4(c).

Table 1 presents the estimated kinetic parameter values for the ODE (3) and SDE (4) models, obtained using the MCMC approach.

Table 1. The Kinetic Parameters of ODE and SDE

ODE (1)	SDE (2)
$g=0.1$	$g=0.2$
$b=0.95$	$b=0.95$
$m=0.71$	$m=0.71$
$\rho=1.1$	$\rho=1.1$
$\alpha=-0.003$	$\alpha=-0.003$
$a=0.6$	$a=0.4$
	$\sigma=0.01$

Figure 5 (a) and (b) present the experimental data and the simulated results for ODE (3) and SDE (4), respectively. In Figure 5(a), 500 simulations were conducted for SDE (4). Then, the average of the simulated results is computed. Both ODE and SDE demonstrate a decreasing trend of HSC-4 cell lines in the presence of TQ as an anticancer agent. The simulated results align with the experimental data of HSC-4 in the presence of TQ.

The root mean square error (RMSE) values are presented in Table 2, providing a quantitative measure of the models' adequacy in fitting the experimental data. Low values of RMSE signify a good fit, indicating that the model predictions closely correspond with the observed data. Since ODE and SDE Gompertz models yield RMSE values close to zero, they demonstrate sufficient ability to capture the underlying patterns in the experimental data. This result suggests that both models effectively represent the system's behavior.

Table 1. RMSE for HSC-4 Gompertz Stochastic and Deterministic Models

Model	RMSE
Deterministic (ODE)	0.015052227
Stochastic (SDE)	0.015043384

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

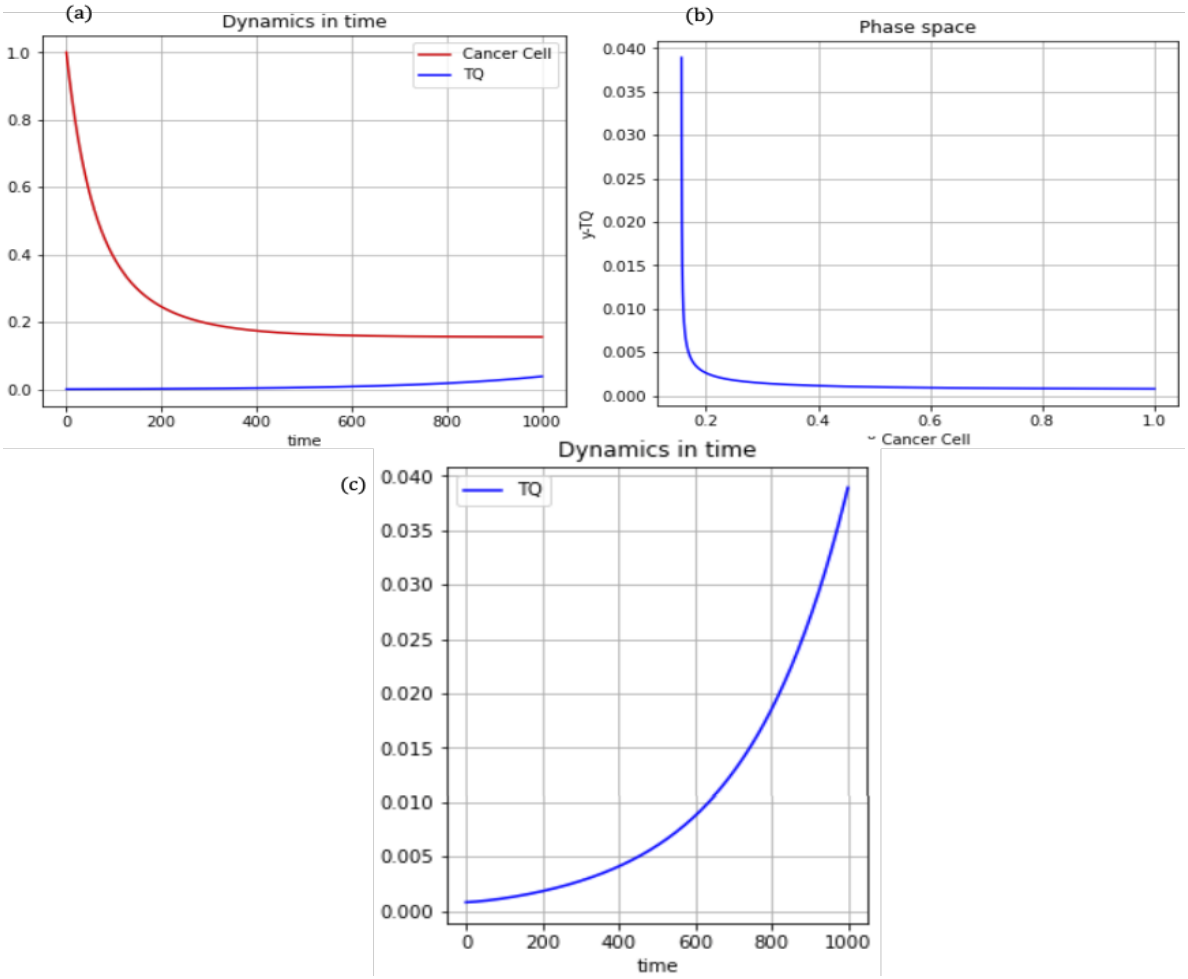


Figure 4. The phase plane plot of the Gompertz model (3) for $a = -0.5,\ b = 0.4,\ \alpha = 4.0,\ \rho = 1.1,\ g = 0.2$ and $m = 2.82.$

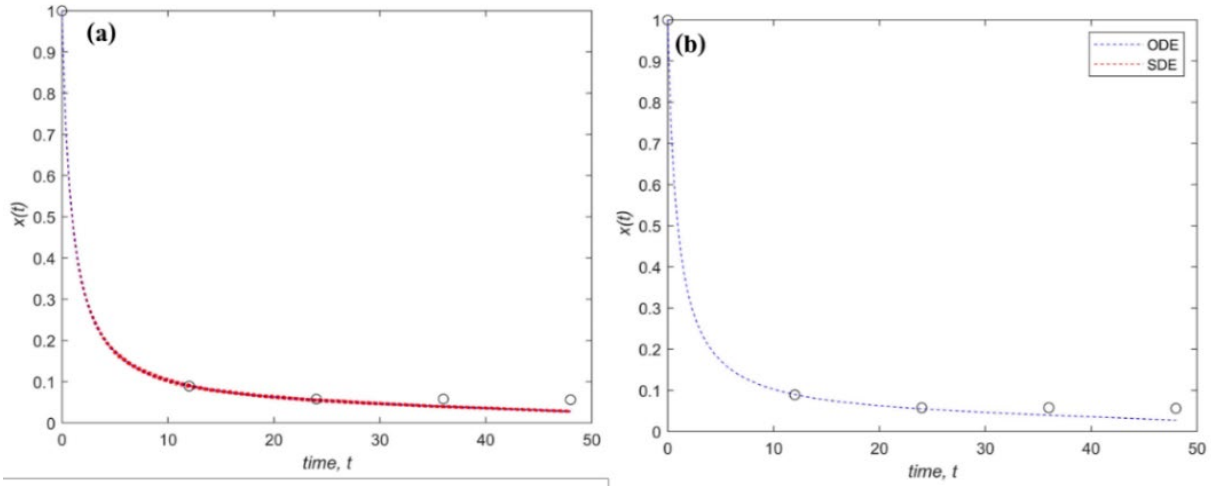


Figure 5. The Numerical Simulation for 500 simulations of the Gompertz model for the ODE (3), SDE (4), and the Experimental Data of HSC-4 Cell Lines.

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

7. Conclusion

This study presents both deterministic and stochastic formulations of the Gompertz model to characterize the growth behavior of HSC-4 oral cancer cells in response to thymoquinone (TQ), a promising anticancer agent. The deterministic model assumes a predictable and smooth tumor growth trajectory, governed by fixed parameters and initial conditions. It is effective for representing the average behavior of the system. However, it fails to capture random biological fluctuations and uncertainties that arise from patient-specific variability, environmental factors, or measurement noise. To address this limitation, a stochastic version of the Gompertz model is introduced by incorporating noise terms into the differential equations. This extension permits modeling the probabilistic dynamics of tumor growth, offering a more realistic representation of the variability in cell response to TQ treatment.

The model demonstrates that TQ exhibited cytotoxicity in a cell-type-dependent manner, with its effects on HSC-4 cells increasing with concentration and exposure time. The observed decline in HSC-4 cell populations over time indicates a positive therapeutic response, supporting TQ's potential as a targeted oral cancer therapy. Furthermore, the study identifies equilibrium points, including the disease-free equilibrium (DFE), tumor-free equilibrium (TFE), and steady-state free equilibrium (SFE). These equilibrium states require further examination through both local and global stability analyses to understand the system's long-term behavior under various treatment conditions. Sensitivity analysis of the model parameters is also essential to assessing how changes in biological or pharmacological inputs affect model outcomes. This will enhance the reliability and predictive capability of the model, enabling more accurate and personalized treatment strategies. These aspects are intended to be explored in future research.

8. Acknowledgement

The authors acknowledge Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) through the Distinguished Research Grant (RDU223008), the Research & Innovation Department under Tabung Persidangan Dalam Negara (TPDN), and the Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS), grant number FRGS/1/2019/STG06/UMP/02/2, for the financial support.

8. References

Abu Khader, M. M. (2013). Thymoquinone in the clinical treatment of cancer: Fact or fiction? Pharmacognosy Reviews, 7(14), 117.

Alaufi, O. M., Noorwali, A., Zahran, F., Al-Abd, A. M., Al-Attas, S. (2017). Cytotoxicity of thymoquinone alone or in combination with cisplatin (CDDP) against oral squamous cell carcinoma in vitro. Scientific Reports, 7: 13131.

Cadoni, G., Giraldi, L., Petrelli, L., Pandolfini, M., Giuliani, M., Paludetti, G., ..., Boccia, S. (2017). Prognostic factors in head and neck cancer: A 10-year retrospective analysis in a single institution in Italy. Acta Otorhinolaryngologica Italica, 37(6): 458–466.

Chen, P. N., Chu, S. C., Kuo, W. H., Chou, M. Y., Lin, J. K., & Hsieh, Y. S. (2011). Epigallocatechin-3 gallate inhibits invasion, epithelial–mesenchymal transition, and tumor growth in oral cancer cells. Journal of Agricultural and Food Chemistry, 59 (8):3836–3844.

Emanuel, E. J., Emanuel, L. L. (1992). Four models of the physician-patient relationship. JAMA, 267(16): 2221–2226.

Gali-Muhtasib, H., Roessner, A., Schneider-Stock, R. (2006). Thymoquinone: A promising anti-cancer drug from natural sources. The International Journal of Biochemistry & Cell Biology, 38(8): 1249–1253.

Global Cancer Observatory. (2025). International Agency for Research on Cancer. https://gco.iarc.who.int (accessed May 16, 2025)

González-Moles, M. Á., Aguilar-Ruiz, M., & Ramos-García, P. (2022). Challenges in the early diagnosis of oral cancer, evidence gaps and strategies for improvement: A scoping review of systematic reviews. Cancers, 14(19): 4967.

Hannan, M. A., Rahman, M. A., Sohag, A. A. M., Uddin, M. J., Dash, R., Sikder, M. H., ... Kim, B. (2021). Black cumin (Nigella sativa L.): A comprehensive review on phytochemistry, health benefits, molecular pharmacology, and safety. Nutrients, 13(6): 1784.

Hu, G., Li, B., Geng, Z. (2024). Analysis of a stochastic hybrid Gompertz tumor growth model driven by Lévy noise. Stochastic Models, 1–22.

Jallah, J. K., Dweh, T. J., Anjankar, A., & Palma, O. (2023). A review of the advancements in targeted therapies for breast cancer. Cureus, 15(10).

Kasteb, F. H., Saileau, K., & Meffert, R. M. (1990). Quantitative evaluation of human gingival epithelial cell attachment to implant surfaces in vitro. The International Journal of Periodontics & Restorative Dentistry, 10(1): 68–79.

Kazerouni, A. S., Gadde, M., Gardner, A., Hormuth, D. A., Jarrett, A. M., Johnson, K. E., ... & Yankeelov, T. E. (2020). Integrating quantitative assays with biologically based mathematical modeling for predictive oncology. iScience, 23(12).

Kuznetsov, V. A., Makalkin, I. A., Taylor, M. A., & Perelson, A. S. (1994). Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. Bulletin of Mathematical Biology, 56(2): 295–321.

Lv, H., He, G., Cheng, H., & Peng, Y. (2022). Stochastic behaviors of an improved Gompertz tumor growth model with coupled two types of noise. Heliyon, 8(11).

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

Mazlan, M. S. A., Rosli, N., Ichwan, S. J. A., & Azmi, N. S. (2017). Modelling the cancer growth process by stochastic differential equations with the effect of chondroitin sulfate (CS) as anticancer therapeutics. Journal of Physics: Conference Series, 890(1): 012085.

National Cancer Institute. (2025). SEER Cancer Stat Facts: Oral Cavity and Pharynx Cancer. https://seer.cancer.gov/statfacts/html/oralcav.html (accessed, May 16, 2025)

Norton, L. (1988). A Gompertzian model of human breast cancer growth. Cancer Research, 48(24): 7067–7071.

Rana, S. M. S. (2020). Dynamic complexity in a discrete-time predator-prey system with Michaelis-Menten functional response: Gompertz growth of prey. Computational Ecology and Software, 10(3): 117.

Steliarova-Foucher, E., Colombet, M., Ries, L. A. G., et al. (2017). International incidence of childhood cancer, 2001–10: A population-based registry study. The Lancet Oncology, 18(6): 719–731.

Steliarova-Foucher, E., Stiller, C., Lacour, B., Kaatsch, P. (2005). International Classification of Childhood Cancer, third edition. Cancer, 103(7): 1457–1467.

Tabassum, S., Rosli, N., & Arief, S. J. I. (2024). Stochastic Gompertz model for the oral cancer cell line HSC-3 response to an anticancer thymoquinone. AIP Conference Proceedings 2895(1).

Tabassum, S., Thakur, V., Rosli, N., Ichwan, S. J. A., Mishra, P., & Suriyah, W. H. (2022). Therapeutic implications of thymoquinone and its molecular and functional mechanisms against oral and lung cancer. Gene Reports, 27, 101600.

DOI: https://doi.org/10.22452/mjs.vol45sp2.7
Malaysian Journal of Science 45 (Special Issue 2): 66-73 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026) https://mjs.um.edu.my

Numerical Computation of Second Rank Polarization Tensor with Speeding Up Techniques

Suzarina Ahmed Sukri[1a*], Yeak Su Hoe[2a], Taufiq Khairi Ahmad Khairuddin[3a] and Nurfarrisha Asnida Khairul Akmar[4a]

Abstract: This research focused on the techniques used to accelerate the numerical computation of the second rank polarization tensor (PT), which plays a crucial part in many real-life applications such as mathematics and engineering. In order to achieve an accurate numerical solution of PT with reduced computation time, this study employed a different yet simple technique for handling the higher singularity in the mathematical computation of PT itself. By leveraging numerical integration techniques together with the adaptive mesh refinement technique, it shows that the computation of the numerical solution can be achieved within a shorter time without sacrificing accuracy. From the results, a lower computational time is achieved with a higher precision numerical solution, making these techniques practical in complex systems, especially in engineering applications. This research not only improves computational speed but also enhances efficiency and accuracy by implementing adaptive geometrical modifications with the aids of meshing tool which is Netgen Mesh Generator.

Keywords: Polarization tensor, numerical integration, adaptive mesh refinement, computational efficiency

1. Introduction

The polarization tensor (PT) has attracted considerable attention from researchers, particularly in engineering and science. In this context, the term "tensor" originates from the field of mathematics, where it is used to describe the characteristics of objects that may vary according to their directions. This phenomenon appears in several real-life problems, for example, in the properties of crystalline substances, where these properties differ but correspond to their directions. In this case, we describe it as anisotropic. A researcher named Polya represented PT as an induced mass of an object for which objects with no conductivity move in a fluid (Polya, 1947). The second rank PT contains both physical and geometrical properties of an object. In subsequent years, PT has been further developed, where, instead of PT, researchers derived the mathematical formulation of PT and called it the Generalized Polarization Tensor (GPT). GPT is expressed as an asymptotic formula, where the dominant terms in that formula are referred to as the second rank PT (SRPT) or the first order PT (FPT). Since PT contains information regarding object inclusion, these characteristics make it useful for distinguishing between utilized and non-utilized objects. Overall, the concept of GPT, introduced by Polya and Szego, has significantly expanded the original version of PT pioneered by Polya. This concept has been widely employed by other researchers in the same field, as seen in Ammari and Kang (2007), Friedman and Vogelius (1989), de Jong et al. (2000), and Polya (1997). The application of PT can be seen in many areas, including electrosensing, medical imaging, electrical imaging, metal detection, and many other uses (Ammari & Kang, 2003, 2006, 2007; Bardet & Vasseur, 2005; Bannister & Gupta, 2009; Roux & Gloaguen, 2016; Alkhalifah, 2014). For instance, PT was used in a model of electrical impedance tomography (EIT), where the system is employed by an engineer to measure voltages on the surface of a body, which can determine the image of objects that might appear inside human bodies (Ammari & Kang, 2007). Meanwhile, breast anomalies detected by T-scan detection are demonstrated by Ammari and Kang (2003), where PT was employed in mathematical modeling of the anomaly's image, as illustrated in Figure 1(a), where the breast abnormality is reconstructed using the model of PT.

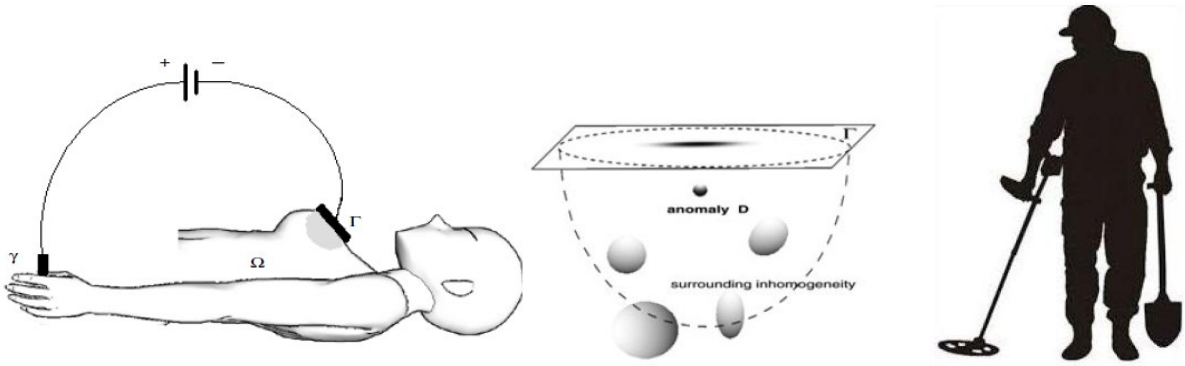
The computation of the second rank PT is crucial in various applications where it requires accuracy to ensure no false results are obtained. For example, in metal detection, we can differentiate between metallic and non-metallic objects, which necessitates the use of highly accurate methods. The inaccuracy of these numerical results can lead to unreliable numerical techniques, which may cause high costs in real-life problems. Previously, several numerical methods were developed to compute SRPT for various geometries, such as the Boundary Element Method (BEM). Among all numerical methods used, BEM seems to gain the most attention from researchers because of its ability to handle complex mathematical formulations, which in this case is the integral of PT itself. For example, Jin et al. (2021) estimate a PT for a cylindrical shape with high conductivity. In their research, BEM was used, and their results show a highly accurate outcome when compared to the analytical solution derived. Other methods used by previous researchers include semi algebraic methods, linear and quadratic element numerical

Authors information:
[a]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310 Johor, MALAYSIA. E-mail: suzarina@utm.my[1], s.h.yeak@utm.my[2], taufiq@utm.my[3], nurfarrishaasnida@graduate.utm.my[4]

*Corresponding Author: suzarina@utm.my

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)



(a) (b)

Figure 1. (a) T-scan configuration for detection of breast anomaly. (b) Detection of a metal in landmine using metal detector (Ammari and Kang, 2003).

integration, coupling techniques using interpolation and minimization, among others, as seen in works by Khairuddin and Lionheart (2013, 2016), Capdeboscq et al. (2012), Lu et al. (2013), and Sukri et al. (2020). The precision of numerical solutions can vary based on the geometrical shapes used. Various geometries were employed in PT computation, such as benchmarking geometries, spheres and ellipsoids, and other real-life shapes like coins and guns. By applying these methods, researchers aim to understand and improve the effectiveness of PT calculations, especially in real-life applications such as metal detection or similar fields, which can ensure reliable yet efficient outcomes in critical engineering problems.

However, achieving these so-called accurate methods can sometimes incur higher time demands for computation. Since high accuracy is crucial to ensuring that the solution is reliable, the process of obtaining results could make it less practical for real-life applications where rapid and accurate results are critical. The singularity of the PT integral itself makes it challenging to solve analytically, and numerically, it consumes considerable time. To address this issue, this study aims to provide a speed-up technique by implementing geometric modifications as well as exploiting the adaptive properties of PT to ensure that computations are completed in a shorter time while still achieving accurate numerical solutions. By implementing a speeding technique, the computational cost can be optimized in terms of time, enabling computations to be completed faster without sacrificing solution accuracy. These strategies include geometric modifications and adjustments used in PT calculations, which enhance computational efficiency. This study aims to provide a practical solution balancing efficiency and precision for real-life problems where speed and reliability are considered key factors in the methods.

In this paper, we employed a quadratic element integration from a study by Sukri et al. [18], where the speeding up technique is used in the singular integral operator of K_B^* . We perform a comparison of the computation for K_B^* without speeding technique with the new computed singular integral operator K_{BN}^* in terms of its error as well as its execution time. In the next section, we introduce the mathematical background of PT, which is crucial for better insight and understanding of PT, especially first order PT.

2. The second rank polarization tensor

The transmission problem which has been defined by Ammari and Kang (2007), is the origination of the terminology of PT for which a Lipschitz bounded domain called the domain of B is considered where the origin of the domain, O is stated inside B. The conductivity of object is denoted as k which in this case cannot be equivalent to 1 (which is considered to be the same conductivity as the surrounding). The range of the conductivity is from non-negative numbers to infinite numbers $(0 < k \neq 1 \leq +\infty)$. The solution of this transmission problem in equation (1) is denoted as u where H is the harmonic function in $\mathbb{R}^3$.

$$u(x) - H(x) = \sum_{|i|=|j|=1}^{+\infty} \frac{(-1)^{|i|}\partial_x \Gamma(x) M_Q(x_i, y_j)\partial^j H(0)}{i!\, j!}, as\ |x| \to +\infty. \quad (1)$$

In this case, $|i|$ and $|j|$ are multi-indices,$\Gamma(x)$ is the fundamental solution of the Laplacian while $M_Q(x_i, y_j)$ is the first order PT. The definition of the first order PT is represented in terms of an integral equation containing a system of linear equations over the boundary of the domain B which is stated in equation (2).

$$M_Q(x_i, y_j) = \int_{\partial B} y_j \psi_i(y) d\Omega(y)\,, x, y \in \partial B. \quad (2)$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

The notation for the coordinates for element y is denoted as y_j, while $\psi_i(x)$ is defined as a system of equations as in equation (3).

$$\psi_i(y) = (\lambda I - K_B^*)(N_{V_x}). \quad (3)$$

where $\lambda = 1/2 + 1/(k+1)$ while N_{V_x} is the normal vector of element x with formula $N_{V_x} = V(x).\nabla x_i$. In order to obtain the system of equations, $\psi_i(y)$, we need to solve the singular integral operator K_B^*, where this integral is defined as Cauchy Principle Value integral stated in equation (4).

$$K_B^*\psi(x_i) = \frac{1}{4\pi} P.V. \int_{\partial B} \frac{\langle x_i - y_j, V_{x_i}\rangle}{|x_i - y_j|^3} \psi(y_i) d\Omega(y). \quad (4)$$

where $\psi(x_i) \in \partial B$ such that $L^2(\partial B)$ is a space of square integrable functions. In our study, we aim to focus on the computation of K_B^* whereby the speeding-up technique is applied to this integral. The reason the integral is chosen is that when we referred to the study by Sukri et al. (2020), it is observed that the matrix that is formed as Gaussian quadrature is implemented using to a symmetric matrix. Hence, instead of calculating the whole matrix, we only calculate the upper or lower part of the matrix and multiply it by two. It seems that the process is simple. However, in order to define the geometry in Netgen Mesh Generator, which is the meshing software used to discretize the geometry, we need to define a hollow object, where, in this case, we only consider the boundary of the element. After we implemented the speeding process, we compared the numerical solutions in terms of reduction time percentage as well as its relative error. As mentioned in the study by Ammari and Kang (2007), we refer to the following theorem which can validate the result and the numerical solutions that have been computed.

2.1 Theorem and Proving

Theorem 1 The matrix 3×3 of the first order polarization tensor is symmetric and positive definite if the conductivity of the object inclusion is more than 1, while negative definite if the conductivity is between 0 and 1.

Theorem 1 has been proven by Khairuddin et al. (2013) where the researchers separate the proof involving ellipsoid object inclusion $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$ into two parts:

[1] Matrix of the first order PT, $M(k, B)$ is positive definite matrix if and only if $k > 1$;

[2] Matrix of the first order PT, $M(k, B)$ is positive definite matrix if and only if $k < 1$.

Proof For part [1], the researchers assume $k > 1$ and $M(k, B)$ is the first-order PT for the ellipsoid. Then, the matrix of PT is represented as

$$M(k,B) = (k-1)|B| \begin{bmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{bmatrix}.$$

Let $\alpha_i = (k-1)|B|M_i$ for $i = 1,2$ and 3. The conductivity $k > 1$ since $k - 1 > 0$ and clearly $|B| > 0$ as well as $M_i > 0$ which satisfies Proposition 1 with $0 < k \neq 1 < +\infty$.

Proposition 1 The depolarization factors, D_i for $i = 1,2,3$ satisfies $0 < D_1 < 1$

$\alpha_i > 0$, which is the diagonal of matrix $M(k, B)$ and $M(k, B)$, is a positive definite matrix. Next, assume $M(k, B)$ is a positive definite matrix. We need to show $k > 1$.

$$M(k,B) = (k-1)|B| \begin{bmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{bmatrix}.$$

Let $\alpha_i = (k-1)|B|M_i$ for $i = 1,2$ and 3. Since $M(k, B)$ is a positive definite matrix, then the eigenvalues, $\alpha_i > 0$. Here, $|B| > 0$. Since $0 < k \neq 1 < +\infty$ and $0 < D_1 < 1$ based on Proposition 1, $M_i > 0$. For $\alpha_i > 0$, $(k-1) > 0$ and we can conclude that $k > 1$.

For part [2], assume $k < 1$. We need to show the $M(k, B)$ is negative definite matrix. Using the same approach, we let $M(k, B)$ be the first order PT for ellipsoid. Then, the matrix of PT is represented as

$$M(k,B) = (k-1)|B| \begin{bmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{bmatrix}.$$

Let $\alpha_i = (k-1)|B|M_i$ for $i = 1,2$ and 3. The conductivity $k < 1$, then $k - 1 < 0$, and clearly $|B| > 0$, as well as $M_i > 0$, which satisfies Proposition 1 with $0 < k \neq 1 < +\infty$. $\alpha_i > 0$, which is the diagonal of matrix $M(k, B)$ and it conclude that $M(k, B)$ is a negative definite matrix. Next, assume $M(k, B)$ is a positive definite matrix. We need to show $k < 1$.

$$M(k,B) = (k-1)|B| \begin{bmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{bmatrix}.$$

Let $\alpha_i = (k-1)|B|M_i$ for $i = 1,2$ and 3. Since $M(k, B)$ is a positive definite matrix, then the eigenvalues, $\alpha_i > 0$. Here, $|B| > 0$. Since $0 < k \neq 1 < +\infty$ and $0 < D_1 < 1$ based on Proposition 1, $M_i > 0$. For $\alpha_i > 0$, $(k-1) < 0$, we can conclude that $k < 1$. In the next section, the analytical solution was derived by Ammari and Kang [2], and is represented where the analytical solution was used as a benchmark. Different notations of $M(k, B)$ were used in order to differentiate between the analytical solution M_A and the approximated solution M_E.

2.2 The Analytical Solutions of the Second Rank Polarization Tensor for Sphere and Ellipsoid

The derived analytical solution provided by Ammari and Kang (2007) is based on the ellipsoid with radius 1. First, the researchers consider an ellipsoid with the general equation $x^2/A + y^2/B + z^2/C = 1$. The condition for the semi principal axes A, B and C is $0 < C = B = A$. The analytical solution of the ellipsoid is represented by a 3×3 matrix for which the main

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

diagonal is in integral form containing the semi principal axes A, B and C as represented in the following equation.

$$M_A(x_i, y_j)|_{(k,Ellipsoid)} = \frac{4\pi\varphi(A \cdot B \cdot C)}{3} \begin{bmatrix} \frac{1}{1+D_1\varphi} & 0 & 0 \\ 0 & \frac{1}{1+D_2\varphi} & 0 \\ 0 & 0 & \frac{1}{1+D_3\varphi} \end{bmatrix}. \quad (5)$$

where $\varphi = k - 1$ while D_1, D_2 and D_3 are in the form of an integral containing the semi principle axes of A, B and C, given as

$$D_1 = \int_1^{+\infty} \frac{B \cdot C}{t^2\sqrt{(A^2(t^2-1)+B^2)}\sqrt{A^2(t^2-1)+C^2}} dt,$$
$$D_2 = \int_1^{+\infty} \frac{B \cdot C}{\sqrt{(A^2(t^2-1)+B^2)}\left(t^2-1+\left(\frac{B}{A}\right)^2\right)^2\sqrt{A^2(t^2-1)+C^2}} dt, \quad (6)$$
$$D_3 = \int_1^{+\infty} \frac{B \cdot C}{\sqrt{(A^2(t^2-1)+B^2)}\left(t^2-1+\left(\frac{C}{A}\right)^2\right)^2\sqrt{A^2(t^2-1)+C^2}} dt.$$

To obtain the analytical solution for a sphere, we ensure that the semi principal axes are equal to each other $A = B = C$. By setting semi-principal axes to become equal to each other, the integral for D_1, D_2 and D_3 yield to $1/3$. Hence, the resulting first order PT for a sphere can be computed using the following formula

$$M_A(x_i, y_j)|_{(k,Sphere)} = \frac{4\pi\varphi r^2}{3} \begin{bmatrix} \frac{1}{1+\frac{1}{3}\varphi} & 0 & 0 \\ 0 & \frac{1}{1+\frac{1}{3}\varphi} & 0 \\ 0 & 0 & \frac{1}{1+\frac{1}{3}\varphi} \end{bmatrix}. \quad (7)$$

From the integral of PT in (2), we implement the previous method by Sukri et al. (2020), together with the acceleration techniques which are applied to equation (4), and then compare them with the analytical solution in equations (5) and (7). In this study, we employ a symmetrical geometry with known analytical solutions specifically a sphere, to ensure that our technique are both reliable and effective in terms of accuracy.

3. Speeding up techniques using adaptive geometrical modification

3.1 Netgen Mesh Generator (NG Solve version 6.2)

We start the speeding technique by discretizing a hemisphere into triangular meshes in a meshing software called Netgen Mesh Generator. It is a user-friendly software developed by Joachim Schöberl (1997) that can generate both triangular and quadrilateral meshes for two-dimensional and three-dimensional objects. However, despite its advantages, there is a limitation to this software, as users cannot set the desired number of meshes. The flowchart shown in Figure 2 represents the process of discretizing an object using Netgen Mesh Generator. The flowchart illustrates the steps involved, starting from the initialization of geometry and ending with the final generation of meshes for the geometry.

However, not every geometric shape is provided by the software developer within the application. Other shapes, such as a hemisphere, are not defined in Netgen Mesh Generator. Therefore, if we want a specific shape, we need to define it first using another programming language, which is Python. Due to its compatibility with Netgen Mesh Generator software, we can directly import the predefined shape into Netgen Mesh Generator for mesh generation purposes. In this case, we create a hemisphere with an open bottom, as illustrated in Figure 3. The reason a hemisphere was chosen is that we aim to duplicate it to become a sphere and then calculate its PT while comparing it with the given analytical solution. These steps allow users to customize their own geometry, which can be complicated or complex geometries.

Based on Figure 3, as we duplicate the upper part of the hemisphere, it is observed that there is redundancy in the number of nodes at the axis of symmetry where in this case our axis of symmetry is at $z = 0$. The redundancy of the nodes can make the mesh structure more complex and increase the potential for inaccurate computation and analysis. To address this, several techniques must be employed to remove the redundant nodes. This technique involves removing, redefining and renumbering the nodes accordingly. After successfully defining the new nodes, surface elements and meshes, the computation of the singular integral operator of PT, K_{BN}^* can be carried out. From the properties of matrix of K_B^*, we categorize the computation into three cases.

4. Cauchy principle valiue integral of polarization tensor

The computation of the PT is based on these cases. In order to avoid repetitive computation of the normal vector, V_{x_i}, the computation of a singular integral operator is calculated separately from the original equation of K_B^* and is stored in pre-allocated memory in Matlab. This is to ensure that, minimal memory is used to compute the CPV integral. The integral of the CPV now is expressed as

$$K_B^*\psi(x_i) = \frac{1}{4\pi} P.V. \int_{\partial B} \frac{\langle x_i - y_j \rangle}{|x_i - y_j|^3} \psi(y_i) d\Omega(y). \quad (8)$$

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

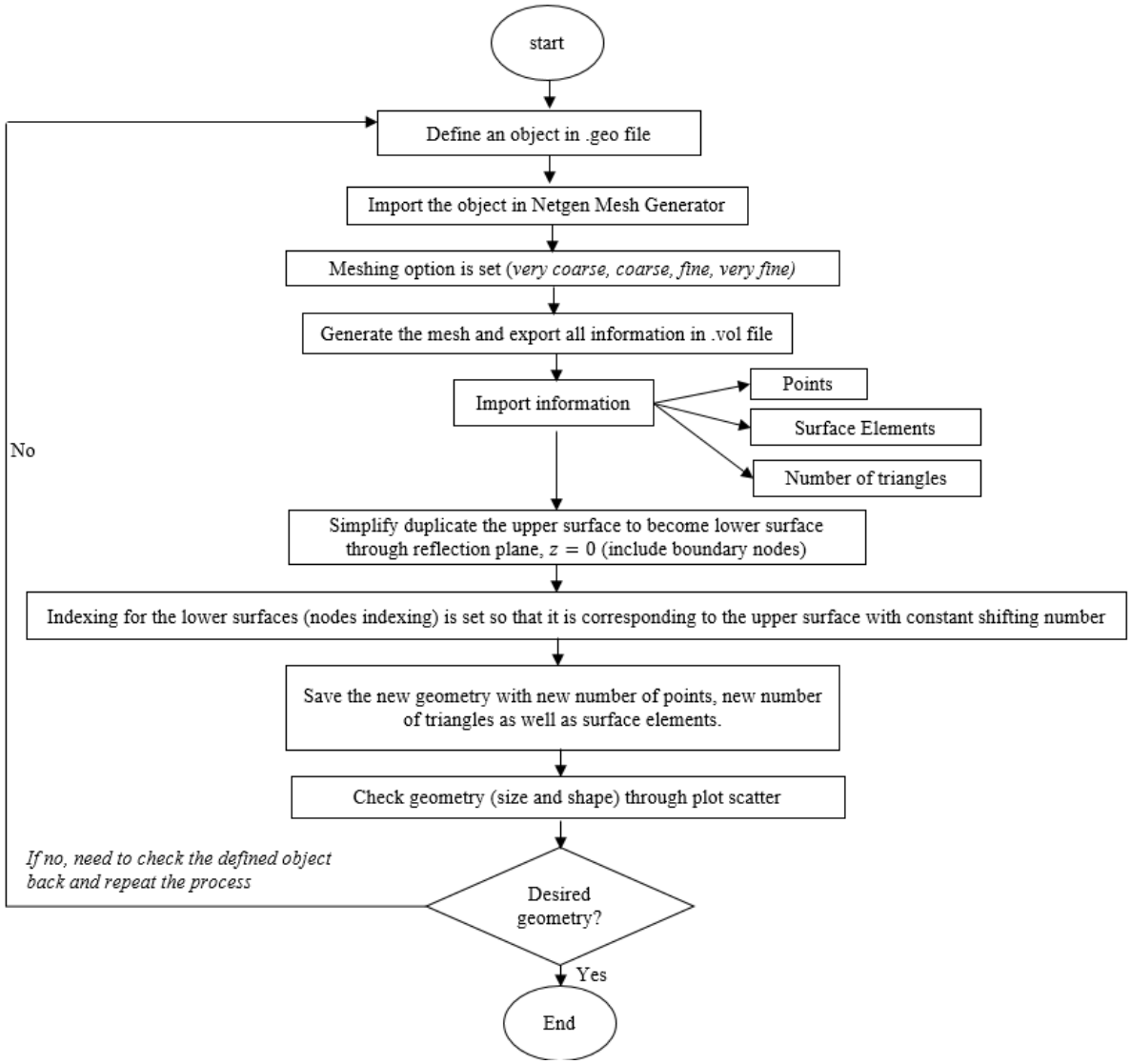
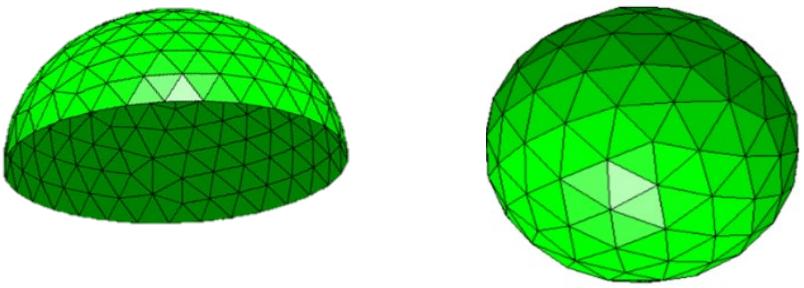


Figure 2. Flow chart illustrating the process to create a sphere from a hemisphere.



(a) (b)

Figure 3. (a) A hemisphere with open bottom is discretized into 359 meshes with radius 3. (b) Duplication of a hemisphere with open bottom to become a sphere with 718 meshes with radius 3.

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

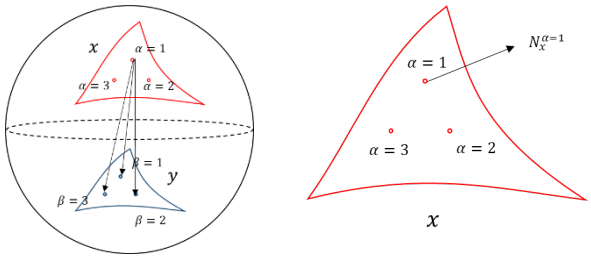
In this case, $x - y$ is the difference between elements within the geometry which is expressed in terms of the distance between the Gaussian points in each element as

$$x_{N_x}^{\alpha=1} - y_{N_y}^{\alpha=1,2,3} = \begin{bmatrix} x_{N_x}^{\alpha=1} - y_{N_y}^{\alpha=1} \\ x_{N_x}^{\alpha=1} - y_{N_y}^{\alpha=2} \\ x_{N_x}^{\alpha=1} - y_{N_y}^{\alpha=,3} \end{bmatrix},$$
$$x_{N_x}^{\alpha=2} - y_{N_y}^{\alpha=1,2,3} = \begin{bmatrix} x_{N_x}^{\alpha=2} - y_{N_y}^{\alpha=1} \\ x_{N_x}^{\alpha=2} - y_{N_y}^{\alpha=2} \\ x_{N_x}^{\alpha=2} - y_{N_y}^{\alpha=,3} \end{bmatrix}, \quad (9)$$
$$x_{N_x}^{\alpha=3} - y_{N_y}^{\alpha=1,2,3} = \begin{bmatrix} x_{N_x}^{\alpha=3} - y_{N_y}^{\alpha=1} \\ x_{N_x}^{\alpha=3} - y_{N_y}^{\alpha=2} \\ x_{N_x}^{\alpha=3} - y_{N_y}^{\alpha=,3} \end{bmatrix}.$$

where N_x and N_y are the number of surface elements. The normal vector for the element x is expressed as

$$V_{x_{N_x}}^{\alpha=1,2,3} = \begin{bmatrix} V_{x_{N_x=1}}^{\alpha=1,2,3} \\ \vdots \\ V_{x_{N_x=N}}^{\alpha=1,2,3} \end{bmatrix}, \quad (10)$$

In this paper, we focus on a sphere with radius 3, as illustrated in Figure 4 a), while the normal vector for element x, V_x for three Gaussian points, $\alpha = 1,2$ and 3, is illustrated in Figure 4 b).



(a) (b)

Figure 4. a) The distance between element x and element y for a sphere with radius 3 b) Normal vector for element x, V_x.

The integral of CPV yields a matrix of $3N_x \times 3yN_x$. It is observed that the upper triangular matrix generates the same values for the lower triangular matrix of K_{BN}^*. Therefore, instead of computing all elements of the matrix, we categorize the matrix according to six cases in total. In this case, we can minimize the time taken for the computation of K_{BN}^* without neglecting its accuracy.

$$\begin{aligned}
Case\ (i) \rightarrow K_{BN}^*([i,gp],[i+N_{el},gp]) &= \langle -1,-1,-1\rangle \\ &\cdot K_{BN}^*([i+N_{el},gp],[i,gp]), \\
Case\ (iia) \rightarrow K_{BN}^*([i,gp],[j+N_{el},gp]) &= \langle 1,1,-1\rangle \\ &\cdot K_{BN}^*([i+N_{el},gp],[j,gp]), \\
Case\ (iib) \rightarrow K_{BN}^*([j+N_{el},gp],[i,gp]) &= \langle 1,1,-1\rangle \\ &\cdot K_{BN}^*([j,gp],[i+N_{el},gp]), \\
Case\ (iiia) \rightarrow K_{BN}^*([i,gp],[j,gp]) &= \langle -1,-1,-1\rangle \\ &\cdot K_{BN}^*([i+N_{el},gp],[j \\ &+N_{el},gp]), \\
Case\ (iiib) \rightarrow K_{BN}^*([j,gp],[i,gp]) &= \langle -1,-1,-1\rangle \\ &\cdot K_{BN}^*([j+N_{el},gp],[i \\ &+N_{el},gp]),
\end{aligned} \quad (11)$$

From equation (11), i and j are the indexing of the elements x and y, are the number of surface elements for the upper part of the sphere. gp is the Gaussian point of three point Gaussian quadrature, which is expressed in shape functions, $N_k(\xi,\eta)$ in terms of ξ and η where $k = 1,2,3,4,5$ and 6, as

$$\begin{aligned}
N_1 = 2\xi^2 - \xi, N_2 = 2\eta^2 - \eta, N_3 &= 2\xi^2 + 2\eta^2 + 4\xi\eta - \xi - \eta \\ &+ 1, \\
N_4 = 4\xi\eta, N_5 = -4\eta^2 - 4\xi\eta + 4\eta, N_6 &= -4\xi^2 - 4\xi\eta + 4\eta.
\end{aligned} \quad (12)$$

Instead of computing the matrix of K_{BN}^* as a whole, we compute the matrix by cases. Then time reduction is observed. Although it seems that, many modifications are needed in order to complete the matrix computation, it is observed that the time reduction is significant. The percentage of time reduction is calculated using the formula in equation (13).

$$Time\ reduction = \frac{Time\ \ to\ execute\ (Duplicated\ N)}{Time\ to\ execute\ (N\ without\ duplication)} \times 100\% \quad (13)$$

Based on the modification of the geometry, it is observed that the total time reduction percentage can be as low as 12%. However, some external factors that we cannot control, such as the processor running in the background of MATLAB and memory handling, can cause the reduction time percentage to vary for each duplication of elements. After completing the speeding technique, we compare the relative error in equation (14) for the second-rank PT, where we must ensure that the technique does not affect the computation of the PT itself.

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

$$Relative\ Error = \frac{|M_E - M_A|}{|M_E|}. \quad (14)$$

M_E is the exact solution for second rank PT while M_A is the computed second rank PT based on the speeding technique and without the speeding technique. In the next section, numerical examples based on a sphere with radius 3 and conductivity, is $k = 1.5$ are computed and the time reduction is recorded.

5. Numerical examples of second rank polarization tensor

5.1 Comparative Analysis of the Execution Time for K_B^* and K_{BN}^*

This section presents an analysis of the execution time to compute the CPV integral with and without speeding technique. The performance of the numerical computation is observed based on the time taken for the numerical computation to finish. However, we must take account other factors that may influence the time of execution. Computing the CPV integral can be computationally intensive due to the singularity of the integral itself. A comparative study is conducted, where we compare the time execution , K_{BN}^* with speeding technique, K_B^* for different numbers of surface elements. In order to perform the comparative analysis, the mesh size or number of surface elements for a sphere is generated using Netgen Mesh Generator. However, the mesh size cannot be set according to the user's preference. Therefore, for the numerical computation for K_B^*, the number of surface elements that is obtained from the software is $N = 28, 42, 130, 278, 354, 714$ and 2842 meshes, while the time to compute it is recorded in Table 1.

Table 1. Time to execute the numerical computation for K_B^* of a sphere with radius 3 and conductivity, $k = 1.5$.

Mesh size, N	Time to execute K_B^* (second)
28	0.812500
42	1.718750
130	15.46875
278	68.8750
354	113.03125
718	571.96875
2842	7390.203125

Table 2. Time to execute the numerical computation for K_{BN}^* of a sphere with radius 3 and conductivity, $k = 1.5$.

Mesh size, N	Time to execute K_{BN}^* (second)
28	0.578125
42	1.406250
116	12.593750
280	58.312500
464	94.406250
652	396.125000
2608	6474.515625

Meanwhile, for the computation of the new K_{BN}^*, instead of generating a full sphere, as described in the previous section, we generate a hemisphere with the same radius as the sphere generated in Table 1. As a result, the mesh size is effectively half of the resulting mesh described in Table 2. We generate multiple mesh sizes and then duplicate the hemisphere to form a complete sphere (while ensuring that the boundaries at $z = 0$ do not overlap). This allows us to match the mesh size in Table 1, ensuring fair comparison between K_{BN}^* and K_B^* .

Table 1 and Table 2 depicts the time for execution for the CPV integral both with and without geometry modification. By comparing the results from the Table, the numerical computation for the new K_{BN}^* is lower than K_B^* as the number of mesh increases. This suggests that the computation using the geometry modification is more optimized and less complex than the original computation of the CPV integral.

Table 3 shows the percentage of time reduction calculated using equation (13). A reduction in computation time for the CPV integral with a radius of 3 and conductivity $k = 1.5$, across tested mesh sizes is depicted as a range of approximately 12% to 30% time reduction. This indicates that the geometry modification leads to a reduction in computation time. However, the time reduction depends on the mesh size, whereas, the highest performance of 30.74% occurs at mesh size 718 (after duplication) while the lowest performance can be seen at mesh size 2842, with performance 13.29. External factors such as background processes, memory availability as well as operating system overhead may influence the time reduction percentage. It is observed that, as the number of surface elements increases, the time to execute is increases. This is due to the size of the matrix computation in the CPV integral itself.

Table 3. Percentage of time reduction to execute the numerical computation for the new sphere with radius 3 and conductivity, $k = 1.5$.

Mesh Size, N (before duplication)	Mesh Size, N (after duplication)	Percentage of time reduction $(\%)$ for K_{BN}^*
14	28	28.85
21	42	18.18
65	130	18.59
139	278	15.34
177	354	16.48
359	718	30.74
1421	2842	12.39

Figure 5 depicts the difference between the execution time for the computation of K_{BN}^* and K_B^*. From the figure, we observe that as the number of elements increases, the computation increases. This is due to the increasing size of the matrix of the CPV integral. As the matrix size increases, computational resources such as memory and processing power also increase. This leads to a longer computation time for the singular integral of the CPV integral as reflected in Figure 5, where both execution time for K_{BN}^* and K_B^* is increase as the number of surface elements increases. Larger matrix sizes can pose challenge in

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

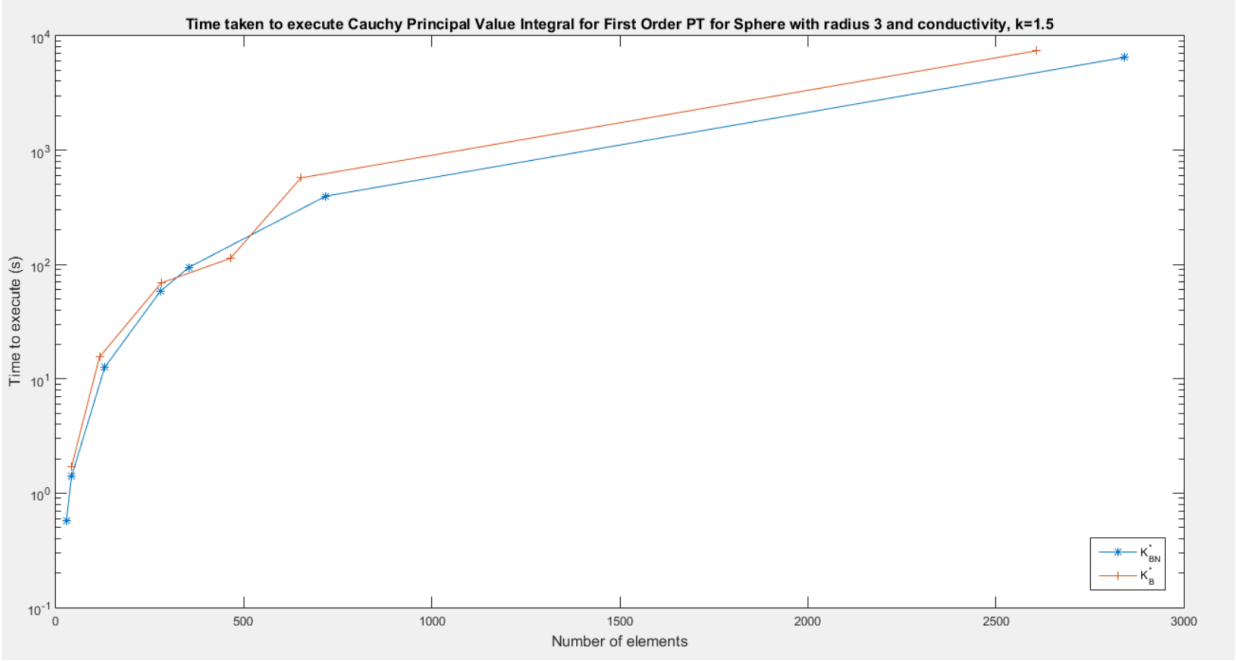


Figure 5. Execution time for K^*_{BN}(with speeding) and K^*_B(without speeding) for the first order PT calculation on a sphere with radius 3, conductivity, $k = 1.5$ and the number of surface elements for $N^* = 28, 48, 130, 278, 354, 718,$ and 2842 while $N = 42, 116, 280, 464, 652$ and 2608.

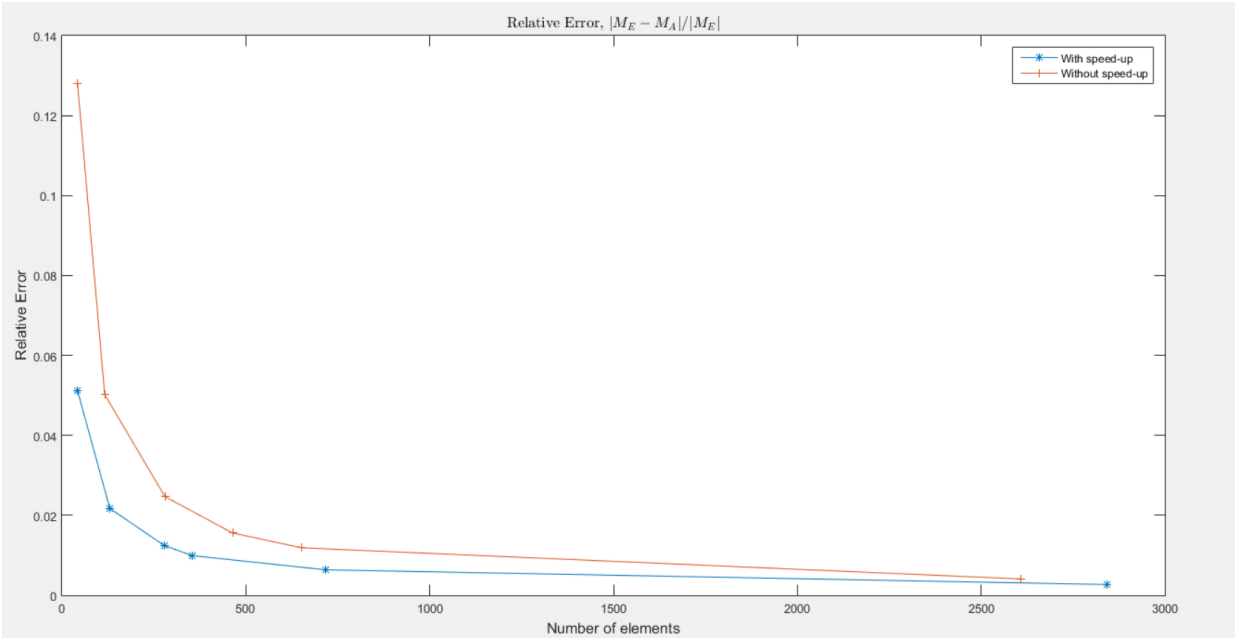


Figure 6. Convergence of the relative error using for the first order PT calculation on a sphere with radius 3, conductivity, $k = 1.5$ and the number of surface elements for $N^* = 28, 48, 130, 278, 354, 718,$ and 2842 while $N = 42, 116, 280, 464, 652$ and 2608.

numerical stability. The stability of a larger matrix size is critical and often has a higher sensitivity to perturbations where, any changes or perturbations can make the matrix ill-conditioned, thus rendering it unstable. Hence, computation by cases, as mentioned in equation (11), is required, which can lead to a lower relative error for the entire process of obtaining the second rank PT, as observed in Figure 6. From Figure 6, the computed relative error using equation (14) shows that the relative error for the second rank PT with the speeding technique applied is lower than the relative error without the speeding technique.

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

6. Conclusion

In conclusion, this paper has demonstrated the effectiveness of the speeding technique for tackling high singularity integrals, as shown in the integral of PT. The techniques proposed are easy to implement and can be applied to symmetrical object inclusion, where this technique not only reduces computation time but also has proven effective in real-life problems where faster computation is essential. By addressing the challenges associated with higher singularities in the mathematical computation of PT, we have shown that it is possible to achieve accurate numerical solutions while significantly reducing computation time. The integration of advanced numerical techniques, alongside adaptive mesh refinement, has been instrumental in enhancing both speed and precision. The results indicate that our approach meets the demands of complex systems and offers practical solutions for engineering applications where computational efficiency is important. The implementation of adaptive geometrical modifications, enabled by the Netgen Mesh Generator, further emphasizes the potential of these techniques in optimizing computational processes. Overall, this study contributes valuable insights into the intersection of numerical methods and engineering practices, paving the way for future research and applications in the field. The advancements outlined herein not only improve current methodologies but also lay a foundation for continued exploration of more efficient computational techniques.

7. Acknowledgement

The authors would like to acknowledge the financial support from Universiti Teknologi Malaysia (UTM) for the financial funding under the UTM Encouragement Research (UTMER) grant scheme with reference number PY/2024/00856 and cost center number Q.J130000.3854.42J43.

8. References

Alkhalifah, T. A. "Efficient computation of polarization tensors for the detection of buried landmines." Journal of Applied Geophysics, 107 (2014): 74-81.

Ammari, H., and Kang, H., “Properties of the generalized polarization tensors,” *Multiscale Model. Simul.* 1, 335–348 (2003).

Ammari, H., and Kang, H., *Polarization and moment tensors: with applications to inverse problems and effective medium theory*, Vol. 162 (Springer Science & Business Media, 2007).

Bannister, J. R., and Gupta, V. "Electromagnetic detection of buried objects using the polarization tensor." Journal of Geophysical Research, 114 (2009): B12202.

Bardet, J. P., and Vasseur, J. "Generalized polarization tensors in the study of magnetically induced currents." Journal of Applied Physics, 97(10) (2005): 10C102.

Capdeboscq, Y., Karrman, A. B., and Nedelec, J. C., “Numerical computation of approximate generalized polarization tensors,” *Appl. Anal.* 91, 1189-203 (2012).

de Jong, W., Cremer, F., Schutte, K., and Storm, J., “Usage of polarization features of land mines for improved automatic detection,” in *Detection and Remediation Technologies for Mines and Minelike Targets V*, Vol. 4038, 241-252 (SPIE, 2000).

Friedman, A., and Vogelius, M., “Identification of small inhomogeneities of extreme conductivity by boundary measurements: a theorem on continuous dependence,” *Arch. Rational Mech. Anal.* 105, 299-326 (1989).

Jin, Zhongwen, Ganghua Qin, Haidong Fan, Ruochen Huang, Ziqi Chen, Qian Zhao, Mingyang Lu, Zhijie Zhang, Anthony Peyton, and Wuliang Yin. "Boundary-element analysis of magnetic polarization tensor for metallic cylinder." IEEE Access 9 (2021): 63250-63256.

Lu, Zhijie Zhang, Anthony Peyton, and Wuliang Yin. "Boundary-element analysis of magnetic polarization tensor for metallic cylinder." IEEE Access 9 (2021): 63250-63256.

Kang, H., and Ammari, H. "Generalized polarization tensors and their applications." Inverse Problems, 22(1) (2006): 305-317.

Khairuddin, T. K. A., and Lionheart, W., “Some properties of the first order polarization tensor for 3-D domains,” *Matematika* 19, 1-18 (2013).

Khairuddin, T. K. A., and Lionheart, W. R., “Computing the first order polarization tensor: welcome BEM++!,” *Menemui Matematik* 35, 15-20 (2013).

Khairuddin, T. K. A., and Lionheart, W. R., “Characterization of objects by electrosensing fish based on the first order polarization tensor,” *Bioinspir. Biomim.* 11, 055004 (2016).

Lu, M., Zhao, Q., Hu, P., Yin, W., and Peyton, A. J., “Prediction of the asymptotical magnetic polarization tensors for cylindrical samples using the boundary element method,” in *2015 IEEE Sensors Applications Symposium (SAS)*, 1-4 (IEEE, 2015).

Pólya, G., “A minimum problem about the motion of a solid through a fluid,” *Proc. Natl. Acad. Sci.* 33, 218-221 (1947).

Roux, M., and Gloaguen, R. "Application of generalized polarization tensors to electromagnetic modeling of complex geological structures." Geophysical Journal International, 205(3) (2016): 1836-1855.

S. A. Sukri, Y. S. Hoe, and T. K. A. Khairuddin, “Quadratic element integration of approximated First Order Polarization Tensor for sphere,” Malaysian J. Fundam. Appl. Sci. 16(5), 560–565 (2020).

Schöberl, Joachim. "NETGEN An advancing front 2D/3D-mesh generator based on abstract rules." Computing and visualization in science 1, no. 1 (1997): 41-52.

DOI: https://doi.org/10.22452/mjs.vol45sp2.8
Malaysian Journal of Science 45 (Special Issue 2): 74-82 (November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026) https://mjs.um.edu.my

Simulation Study of Agglomerative Clustering Algorithms for Detecting Outliers in Univariate Circular Data

Nur Syahirah Zulkipli[1a], Siti Zanariah Satari[2a*] and Wan Nur Syahidah Wan Yusoff[3a]

Abstract: Outlier detection techniques for univariate circular data are distinct from those designed for bivariate or multivariate contexts due to differences in data structure and dimensionality. In earlier research, clustering-based procedures specifically those utilizing agglomerative algorithms have proven effective for detecting outliers in bivariate circular datasets. This study aims to evaluate the performance of agglomerative clustering algorithms such as single-linkage (SL), complete-linkage (CL) and average-linkage (AL) by conducting a simulation study for a clustering-based outlier detection (COD) procedure for univariate circular data. The performance of the clustering algorithms from COD procedure assessed across various simulated data conditions under a specific outlier scenario. In the simulation study, outliers are embedded within synthetic datasets using a defined outlier scenario, where the mean direction is shifted by varying degrees of contamination to distinguish outliers from inliers. The simulation study revealed that CL-Satari/Di algorithm has the best performance on detecting outliers correctly with minimal masking effect under specific data conditions. Meanwhile, the performance of SL-Satari/Di algorithm found to be the best algorithm with low swamping effect even though not performing well in detecting correct outliers. This result also revealed that the concentration parameter, percentage of outliers, and sample size affect the performance of the clustering algorithms. The findings demonstrate that the algorithms effectively detect outliers with low masking and swamping effects when the concentration parameter is large. The simulation results also found that the algorithms performed well in detecting outliers when percentage of outliers is small with low masking effect but producing high swamping effect. Moreover, large sample size was found to be less sensitive to masking effect, while a smaller sample size is less sensitive to swamping effect. This work extends existing research on identifying outliers within univariate circular datasets by employing a clustering-based procedure as its core framework.

Keywords: Circular data, Agglomerative clustering algorithms, Outlier detection, Outlier scenario, Simulation study.

1. Introduction

Numerous techniques have been established for identifying outliers in univariate linear data. However, approaches specific to univariate circular data have received comparatively less attention. This is primarily due to the fundamental differences in data structure, as the strategies used for linear data are not directly applicable to circular data owing to its periodic and bounded nature. Detecting outliers in circular data is crucial for accurate analysis with the suitable procedure for both univariate and multivariate datasets. Outliers can cause problems such as in modelling and forecasting because they affect how parameters are estimated (Ramlee et al., 2024; Sert & Kardiyen, 2023; Satari et al., 2021; Satari & Ku Khalif, 2020). In circular data, the concept of an outlier as an extreme value is not valid due to its closed and bounded range (Abuzaid, 2020). Since circular data has a different geometric structure from linear data, detecting outliers in it needs special methods (Sert & Kardiyen, 2023). In some cases, such as in medical research or fraud detection, outliers may represent important discoveries rather than errors. Therefore, Mahmood et al. (2017a) suggested using robust statistical methods or adjusting outliers using clean data points, instead of simply removing or ignoring them.

Several studies have explored the detection of outliers in univariate circular data. Researchers such as Mahmood et al. (2017b), El et al. (2019), Abuzaid (2020), and Rattanawong et al. (2023) have proposed various methods specifically designed for this purpose. Among these, the row deletion technique has been widely adopted and frequently cited in the literature (Collett, 1980; Abuzaid et al., 2009; Hung et al., 2012; Mahmood et al., 2017b). However, a key limitation of these approaches is their focus on detecting only one or a few outliers at a time (Abuzaid, 2020; Satari & Ku Khalif, 2020). Sebert et al. (1998) highlighted the usefulness of clustering in outlier detection, noting that clustering algorithms can group outlying observations into a distinct cluster. In particular, agglomerative clustering has been recognized for its sensitivity to outliers and commonly used for identifying the outliers.

Satari (2015) demonstrated that agglomerative clustering could effectively detect multiple outliers by introducing a clustering-based approach specifically for bivariate circular data. Later, Di et al. (2017a, 2017b) extended the method using single-linkage clustering for circular-circular regression models, incorporating novel circular similarity measures along with a

Authors information:
[a]Pusat Sains Matematik, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, MALAYSIA. E-mail: syahirahzulkipliwork@gmail.com[1], zanariah@umpsa.edu.my[2], wnsyahidah@umpsa.edu.my[3]

*Corresponding Author: zanariah@umpsa.edu.my

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

stopping rule proposed by Satari (2015). Furthermore, Satari et al. (2019) evaluated the applicability of the single-linkage algorithm for Down and Mardia circular-circular regression models under three different outlier scenarios. This line of research was further extended by Satari et al. (2021) to compare several agglomerative clustering algorithms for detecting multiple outliers in the Down and Mardia circular-circular regression models. This study motivated to compare single-linkage, complete-linkage and average-linkage algorithms using various circular distances, with performance evaluated through simulations across different outlier scenarios, concentration levels, sample sizes, and contamination degrees.

However, the existing clustering-based outlier detection procedures for bivariate circular data were not directly applicable to univariate circular data due to differences in dimensional structure and data generation processes. Recognizing this limitation, Satari and Ku Khalif (2020) highlighted the need to develop a new clustering-based method specifically tailored for univariate circular data. In response, Zulkipli et al. (2024a) proposed and successfully developed the clustering-based outlier detection (COD) procedure and applied to the real-life univariate circular biological data. Accordingly, this study aims to assess the performance of algorithms within the COD framework under a range of data conditions through simulation. To perform the simulation study, the location of outlier will be set using a simulated data generation procedure with a specific outlier scenario formulation. As defined by Sebert et al. (1998), an "outlier scenario" refers to how outlying data points are positioned relative to the core (or “clean”) observations. In the context of univariate circular data, outliers are typically introduced by altering the mean direction, which is a key parameter (Collett, 1980; Abuzaid et al., 2012; Rambli et al., 2012). This study adopts Outlier Scenario 1 (OS1), introduced by Zulkipli et al. (2022), to simulate data under various experimental conditions. At the end of this study, the most effective clustering algorithms and the optimal confidence interval for the stopping rule will be identified, along with the best performing data conditions such as sample size, concentration parameter, and percentage of outliers.

2. Methodology

2.1 Clustering-Based Outlier Detection (COD) Procedure for Univariate Circular Data

Several studies have utilized clustering-based procedure to identify outliers in bivariate circular data, as widely reported in the literature. In this study, the clustering-based procedure for univariate circular data namely COD procedure are outlined in three stages. As a first step, appropriate circular similarity measures were employed to compute the dissimilarity between observations for each clustering algorithm. The similarity distance is calculated by using Satari, Di and Chang distances. According to Zulkipli et al. (2022), the similarity measures produced by the Satari and Di distances are equivalent when dealing with univariate circular datasets. Hence, these circular distance measures are collectively denoted as the Satari/Di distance throughout the study. The formulas for the Satari/Di and Chang-chien similarity distances are given in Eq. 1 and Eq. 2.

Satari/Di distance (Zulkipli et al., 2022):

$$d_{ij(Satari)} = \sum_{k=1}^{p}\left(\pi - \left|\pi - \left|\theta_{ik} - \theta_{jk}\right|\right|\right) \quad (1)$$

Chang-chien distance (Hung et al., 2012):

$$d_{ij(Chang-Chien)} = \sqrt{\sum_{k=1}^{p}\left(1-\cos(\theta_{ik}-\theta_{jk})\right)^2}\,. \quad (2)$$

At the second stage, the agglomerative clustering algorithms are chosen which are single-linkage (SL), complete-linkage (CL), and average-linkage (AL) algorithms to detect the outliers in univariate circular data. The *N* observations are clustered by using SL, CL and AL clustering algorithms and the following are the steps for grouping the observations given by Johnson and Wichern (2014):

Step 1: It starts with *N* clusters and each cluster contain a single observation, $\theta_1,\theta_2,\ldots,\theta_n$.

Step 2: Calculate the distance matrix between all possible pairs of clusters using circular distance in Eq. 1 and Eq. 2, respectively according to circular similarity distance.

Step 3: Find the smallest value of any $d_{\theta_i\theta_j}$ (distance between θ_i and θ_j) in distance matrix to calculate the nearest members and merge the corresponding observations, let say θ_1 and θ_2 are the nearest members and now they formed a cluster. The distance between cluster (θ_1,θ_2) and the other cluster or single observation, $\theta_1,\theta_2,\ldots,\theta_n$ will be calculated. Let a single observation (θ_3) be a cluster and the distance between cluster (θ_1,θ_2) and (θ_3) can be computed as given in Eq. 3 to Eq. 5 for SL, CL and AL, respectively. The distance of $d_{\theta_1\theta_3}$ and $d_{\theta_2\theta_3}$ are referred from the distance matrix in Step 2.

Single-linkage (SL):

$$d_{(\theta_1,\theta_2)\theta_3} = \min\left\{d_{\theta_1\theta_3}, d_{\theta_2\theta_3}\right\}. \quad (3)$$

Complete-linkage (CL):

$$d_{(\theta_1,\theta_2)\theta_3} = \max\left\{d_{\theta_1\theta_3}, d_{\theta_2\theta_3}\right\}. \quad (4)$$

Average-linkage (AL):

$$d_{(\theta_1,\theta_2)\theta_3} = \frac{\sum_i\sum_k d_{ik}}{N_{(\theta_1\theta_2)}N_{\theta_3}}, \quad (5)$$

where $N_{(\theta_1\theta_2)}$ and N_{θ_3} are the number of observations in clusters $(\theta_1\theta_2)$ and θ_3.

Step 4: The rows and columns in the distance matrix responding to the merged cluster(s) are deleted. Now, a new distance matrix is updated by re-calculating the distance matrix after forming new cluster in Step 3.

Step 5: Repeat Step 3 and Step 4 if there are more than one cluster remain in the dendrogram until all the clusters are merged into a single cluster.

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

Next, for the third stage of COD procedure, dendrogram is generated as the output to visualize the clusters and stopping rule. The dendrogram serves to identify the optimal number of clusters, requiring a "cut" at a specific height to distinguish outlying observations from the main data cluster. The stopping rule can be calculated according to the different level of confidence interval as mention in Satari (2015). The specific multipliers λ_P are displayed in Satari (2015) for a selection of possibly *P*-values such as 0.90 and 0.95. Given the appropriate values of λ_P and their range of validity in terms of κ as in Eq. 6.

$$\begin{aligned} P=0.90,\ \lambda_P=1.69,\quad \kappa\geq 0.65\ \ (\rho\geq 0.31)\\ P=0.95,\ \lambda_P=2.06,\quad \kappa\geq 0.80\ \ (\rho\geq 0.37)\end{aligned} \tag{6}$$

Purposely, Satari (2015) choose a specific multiplier $\lambda_P=2.06$ and this study also utilizes 90% confidence interval by using multiplier $\lambda_P=1.69$. Thus, given the stopping rule used in this study as in Eq. 7 and Eq. 8.

$$\bar{h}+1.69s_h \tag{7}$$

$$\bar{h}+2.06s_h \tag{8}$$

The Eq. 7 and Eq. 8 show that $\bar{h}$ is the mean of the heights for all $N-1$ clusters, *N* is a total number of cluster which contain single observations and s_h is the circular standard deviation of the heights given by Eq. 9.

$$s_h=\sqrt{-2\log \bar{R}_h} \tag{9}$$

where $\bar{R}_h$ is the mean resultant length of the height for $N-1$ clusters. Thus, any cluster that exceeds the threshold defined by the stopping rules in Eq. 7 and Eq. 8 will be classified as a candidate outlier group. In summary, six clustering algorithms are employed in this study for identifying outliers in univariate circular dataset, as outlined in Table 1. Each COD algorithm is named based on the agglomerative hierarchical clustering and the circular similarity measures. For instance, CL-Satari/Di denotes the use of complete-linkage clustering combined with the Satari/Di distance.

Table 1. Clustering algorithms and stopping rule to identify outliers

Clustering method	Circular similarity measures	Algorithm name	Stopping rule
Single-linkage (SL)	Satari/Di distance	SL-Satari/Di	90% and 95%
	Chang-chien distance	SL-Chang	
Complete-linkage (CL)	Satari/Di distance	CL-Satari/Di	
	Chang-chien distance	CL-Chang	
Average-linkage (CL)	Satari/Di distance	AL-Satari/Di	
	Chang-chien distance	AL-Chang	

2.2 Outlier Scenario and Simulated Data Generation

In this study, simulated data for univariate circular data is generated with a specific outliers scenario. A simulated data is generated from all combinations of different sample size, n and concentration parameter, κ. To create the simulated data, the "clean" observations or called inliers were randomly generated from von Mises distribution, $VM(\mu,\kappa)$ with mean direction is fixed at 0. Seven values of concentration parameter, $\kappa=1.5,2,5,10,30,50,100$ with four sample size, $n=15,30,50,100$ are considered to generate the simulated univariate circular dataset. The outliers will be introduced in the dataset by 5%, 10%, 15% and 20%. In this study, the outliers will be placed by shifting the specific parameter which is mean direction, μ. The outliers were placed away from inlying observations at specified distance by shifting the mean direction with different degree of contamination, λ in the range of $0\leq\lambda\leq 1$ which defined by Eq. 10,

$$\bar{\theta}_{out}=\mu+\lambda\pi \tag{10}$$

where the mean direction, μ is fixed with 0. The formulation of outlier scenario in Eq. 10 namely as Outlier Scenario 1 (OS1) proposed by Zulkipli et al. (2022). Six contamination levels of $\lambda=0.0,0.2,0.4,0.6,0.8,1.0$ will be set to the simulated data with OS1. The procedure of generating the synthetic dataset with OS1 is briefly explained as follows:

i) Inliers are randomly generated from $VM(\mu,\kappa)$ with the mean direction, $\mu=0$ and the concentration parameter is set at $\kappa=1.5,2,5,10,30,50,100$.

ii) The outliers with 5%, 10%, 15% and 20% are randomly generated from $VM(\mu+\lambda\pi,\kappa)$, where the contamination value, $\lambda=0.0,0.2,0.4,0.6,0.8,1.0$. The mean direction, μ and the concentration parameter, κ are fixed according to condition used in (i).

iii) Inliers and outliers are combined together in one dataset called simulated dataset.

The simulated data is generated for large sample size up to $n=100$ and it is large enough for univariate circular data since this study are not focusing on big data. Moreover, the run time for simulating sample size more than 100 is time consuming. It takes a few days to run the simulation for sample size, $n>30$. The combination of data conditions to generate simulated dataset for simulation study are summarized in Table 2 and there will be 672 simulated datasets with OS1 are generated. In the Results and Discussion section, only selected data conditions will be presented to illustrate key findings and insights.

Table 2. Combination of data conditions and total synthetic datasets generated for simulation study

Sample size, n	Concentration parameter, κ	Percentage outliers	Contamination level, λ
15	1.5	5%	0.0
30	2	10%	0.2
50	5	15%	0.4
100	10	20%	0.6
	30		0.8
	50		1.0
	100		

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

2.3 Simulation Study Design and Performance Indicator

A simulation study will be conducted using Python to assess the performance of the COD procedure in detecting outliers within univariate circular data. The number of outliers denoted as *out* will be introduced into each simulated dataset is determined by

$$out = out\%(n) \tag{11}$$

where $out\%$ is the percentage of outliers and n is sample size. The COD procedure is simulated for 5000 times for each outlier scenario to investigate its performance. In each iteration for simulation study, the six algorithms stated in Table 1 will be performed to identify the outliers from generated synthetic dataset. Three performance indicators defined by Sebert et al. (1998) will be used in this study to evaluate the performance of six algorithms in Table 1 according to each pair of condition in simulation study. The formulas of "success" probability as well as masking and swamping effect as performance indicators are displayed in Table 3.

Note that *s* is the total number of simulations. The closer the *pout* value to 1, the more successful the proposed COD procedure is at identifying all outliers, indicating that no masking has occurred. To make sure that the clustering algorithms from COD procedure performs very well, the masking rate need to be low. Masking becomes a more critical issue than swamping when the primary goal of the analysis is to identify influential subsets of observations. Meanwhile, *pswamp* occurs when an inlier is classified as an outlier by the clustering algorithms from COD procedure. Hence, this will be noted as "*false alarm"*. Similar with *pmask*, low *pswamp* values is considered to be good performance for the clustering algorithms from COD procedure to detect outliers.

3. Result and Discussion

3.1 Simulation Results

The simulation results as measured by the three performance indicators; *pout, pmask* and *pswamp* are presented in Table 4 to Table 6, respectively. The "success" probability of detecting all outliers (*pout*) result as shown in Table 4 found that as contamination level, λ increases, the *pout* also increases. It indicates that the probability of the procedures successfully identified all outliers is high as λ is approaching to 1. The results in Table 4 also show that the *pout* values are decreasing for all algorithms when the percentage of outliers are increasing. Moreover, the *pout* values are increasing when κ getting larger and all algorithms produce similar pattern. However, the *pout* values slightly decrease as n increase for most data conditions. A similar pattern was observed across all algorithms.

The results of masking effect (*pmask*) for OS1 are presented in Table 5. The results of *pmask* found that as λ increase, the *pmask* values gradually decrease and approaching to zero which indicates the algorithm is well-performed since there are no outliers falsely detected as inliers. The *pmask* values also decreasing as the value of κ getting larger for most data conditions. Meanwhile, the *pmask* values slightly increase as the percentage of outliers increase especially when $\lambda \geq 0.2$ for most data conditions. The *pmask* values also gradually increase when n increase especially when $\lambda < 0.6$ for most data conditions. The result of swamping effect (*pswamp*) for OS1 are presented in Table 6. The results show that as λ increase, the *pswamp* values gradually decrease and approaching to zero which indicates the algorithms are not falsely detecting inliers as outliers. Similar with *pmask* result, the *pswamp* values also decreasing as the value of κ getting larger for most data conditions. Moreover, the *pswamp* values gradually decrease when n increase for most data conditions. However, the *pswamp* values slightly increase as the percentage of outliers increase especially when $\lambda > 0.2$ for most data conditions

Table 3. Performance indicator to evaluate the performance of the COD procedure

Performance indicator	Formula
"Success" probability (*pout*)	$pout = \frac{"success"}{s}$ (12) where *pout* is the probability that all the planted outliers are successfully detected and *"success"* is number of dataset that the method successfully identified all the planted outliers.
Masking effect (*pmask*)	$pmask = \frac{"failure"}{(out)(s)}$ (13) where *pmask is* the probability of outliers are falsely detected as inliers, *"failure*" is the number of outliers in all data set that detected as inliers, *out* is the planted outliers.
Swamping effect (*pswamp*)	$pswamp = \frac{"false"}{(n-out)(s)}$ (14) where *pswamp* is the probability of inliers are falsely detected as outliers, "*false*" is the number of inliers in all data set that detected as outliers, *n* is sample size, *out* is the planted outliers.

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

Table 4. "Success" probability (pout) as a performance measure for agglomerative clustering algorithms

Algorithm			SL-Satari/Di		SL-Chang		CL-Satari/Di		CL-Chang		AL-Satari/Di		AL-Chang	
Outlier %			5%	15%	5%	15%	5%	15%	5%	15%	5%	15%	5%	15%
n	κ	λ												
15	1.5	0.0	0.1368	0.0056	0.1424	0.0078	0.3538	0.0564	**0.3618**	**0.0600**	0.2130	0.0138	0.2136	0.0148
		0.2	0.1758	0.0164	0.1816	0.0204	0.3982	0.1038	**0.4064**	**0.1110**	0.2550	0.0440	0.2580	0.0434
		0.4	0.3122	0.0618	0.3192	0.0636	0.5488	0.2416	**0.5544**	**0.2502**	0.4094	0.1386	0.4144	0.1392
		0.6	0.4870	0.1500	0.4922	0.1564	0.6922	0.4340	**0.7006**	**0.4408**	0.5832	0.3200	0.5852	0.2996
		0.8	0.6354	0.2696	0.6424	0.2776	0.7898	0.5816	**0.7974**	**0.5924**	0.7350	0.4512	0.7362	0.4590
		1.0	0.7102	0.3294	0.7156	0.3442	0.8418	0.6362	**0.8480**	**0.6456**	0.7994	0.5106	0.8028	0.5200
	5	0.0	0.1544	0.0104	0.2316	0.0624	0.2828	0.0252	**0.2826**	**0.0252**	0.2192	0.0170	0.2240	0.0178
		0.2	0.3510	0.0722	0.4132	0.1074	0.4576	0.1998	**0.4570**	**0.1998**	0.4168	0.1430	0.4194	0.1468
		0.4	0.7158	0.3220	0.7146	0.3374	0.7158	0.6074	**0.7146**	**0.6066**	0.7614	0.5104	0.7600	0.5164
		0.6	0.9094	0.6738	0.9074	0.6800	0.9094	0.8734	**0.9074**	**0.8724**	0.9506	0.8364	0.9494	0.8360
		0.8	0.9926	0.8986	0.9926	0.8974	0.9870	0.9798	**0.9870**	**0.9802**	0.9954	0.9542	0.9950	0.9528
		1.0	0.9982	0.9624	0.9984	0.9630	0.9984	0.9956	**0.9986**	**0.9954**	0.9990	0.9840	0.9992	0.9840
50	1.5	0.0	0.0048	0.0008	0.0090	0.0060	**0.3448**	**0.1392**	0.3002	0.1252	0.1572	0.0614	0.1092	0.0440
		0.2	0.0092	0.0018	0.0130	0.0054	**0.3820**	**0.1722**	0.3452	0.1512	0.2228	0.0778	0.1630	0.0552
		0.4	0.0492	0.0088	0.0394	0.0096	**0.5336**	**0.2750**	0.5002	0.2488	0.3908	0.1662	0.3288	0.1284
		0.6	0.1788	0.0486	0.1434	0.0408	**0.7012**	**0.4370**	0.6796	0.4138	0.6242	0.3410	0.5632	0.2896
		0.8	0.3638	0.1486	0.3082	0.1310	**0.8014**	**0.5866**	0.7880	0.5634	0.7590	0.5090	0.7178	0.4594
		1.0	0.4696	0.2332	0.4100	0.2128	**0.8302**	**0.6508**	0.8200	0.6320	0.8302	0.5860	0.7984	0.5420
	5	0.0	0.0094	0.0026	0.2084	0.1040	**0.2510**	**0.1146**	0.1852	0.0862	0.1374	0.0568	0.0918	0.0388
		0.2	0.0568	0.0088	0.3004	0.1608	**0.4564**	**0.2068**	0.3758	0.1512	0.3346	0.1382	0.2472	0.0914
		0.4	0.3490	0.1294	0.3730	0.2522	**0.8144**	**0.5964**	0.7424	0.5252	0.7144	0.5258	0.5756	0.4056
		0.6	0.7530	0.5990	0.6536	0.5252	**0.9596**	**0.8956**	0.9314	0.8638	0.9316	0.8580	0.8500	0.7620
		0.8	0.9598	0.9254	0.9172	0.8876	**0.9958**	**0.9852**	0.9926	0.9810	0.9928	0.9786	0.9788	0.9496
		1.0	0.9922	0.9860	0.9784	0.9732	**0.9994**	**0.9980**	0.9990	0.9972	0.9996	0.9982	0.9972	0.9930

Table 5. Masking effect (pmask) as a performance measure for agglomerative clustering algorithms

Algorithm			SL-Satari/Di		SL-Chang		CL-Satari/Di		CL-Chang		AL-Satari/Di		AL-Chang	
Outlier %			5%	15%	5%	15%	5%	15%	5%	15%	5%	15%	5%	15%
n	κ	λ												
15	1.5	0.0	0.8632	0.8583	0.8576	0.8525	0.6462	0.6394	**0.6382**	**0.6345**	0.7870	0.7815	0.7864	0.7813
		0.2	0.8242	0.8261	0.8184	0.8201	0.5902	0.5920	**0.5836**	**0.5847**	0.7250	0.7366	0.7420	0.7460
		0.4	0.6878	0.7343	0.6808	0.7296	0.4512	0.4737	**0.4456**	**0.4642**	0.5906	0.6239	0.5856	0.6213
		0.6	0.5130	0.6000	0.5078	0.5911	0.3078	0.3213	**0.2894**	**0.3135**	0.4168	0.4410	0.4148	0.4371
		0.8	0.3646	0.4479	0.3576	0.4389	0.2002	0.2113	**0.1926**	**0.2043**	0.2650	0.2895	0.2638	0.2829
		1.0	0.2898	0.3768	0.2844	0.3681	0.1582	0.1737	**0.1520**	**0.1668**	0.2006	0.2306	0.1972	0.2243
	5	0.0	0.8456	0.8417	0.7684	0.7622	0.7172	0.6991	**0.7174**	**0.6993**	0.7808	0.7786	0.7760	0.7654
		0.2	0.6490	0.7187	0.5868	0.6752	0.5424	0.5796	**0.5430**	**0.5799**	0.5832	0.6341	0.5806	0.6292
		0.4	0.2666	0.4311	0.2436	0.4175	0.2842	0.2635	**0.2554**	**0.2641**	0.2386	0.2919	0.2400	0.2887
		0.6	0.0498	0.1815	0.0486	0.1773	0.0906	0.0679	**0.0526**	**0.0679**	0.0494	0.0761	0.0506	0.0751
		0.8	0.0074	0.0483	0.0074	0.0481	0.0130	0.0085	**0.0091**	**0.0095**	0.0046	0.0273	0.0050	0.0181
		1.0	0.0018	0.0151	0.0016	0.0163	0.0026	0.0023	**0.0014**	**0.0023**	0.0010	0.0054	0.0008	0.0055
50	1.5	0.0	0.8888	0.8869	0.9033	0.9015	**0.3406**	**0.3394**	0.3370	0.3790	0.5245	0.5268	0.5955	0.5961
		0.2	0.8513	0.8543	0.8687	0.8772	**0.3060**	**0.3085**	0.3002	0.3377	0.4593	0.4655	0.5342	0.5397
		0.4	0.7244	0.7589	0.7643	0.7956	**0.2175**	**0.2270**	0.2394	0.2480	0.3183	0.3284	0.3708	0.3755
		0.6	0.5178	0.5613	0.5765	0.6130	**0.1313**	**0.1512**	0.1418	0.1608	0.1683	0.1811	0.2052	0.2116
		0.8	0.3229	0.3514	0.3787	0.3913	**0.0849**	**0.1021**	0.0908	0.1085	0.0920	0.0985	0.1103	0.1147
		1.0	0.2458	0.2337	0.2920	0.2585	**0.0746**	**0.0829**	0.0786	0.0879	0.0616	0.0704	0.0739	0.0803
	5	0.0	0.8921	0.8895	0.6133	0.6125	**0.4335**	**0.4313**	0.5135	0.5154	0.5747	0.5730	0.6690	0.6679
		0.2	0.7015	0.7623	0.4927	0.5547	**0.2706**	**0.2993**	0.3403	0.3688	0.3691	0.4008	0.4595	0.4913
		0.4	0.3449	0.4587	0.3619	0.4650	**0.0866**	**0.1136**	0.1206	0.1427	0.1239	0.1290	0.1994	0.2096
		0.6	0.1037	0.1272	0.1619	0.2026	**0.0207**	**0.0320**	0.0313	0.0377	0.0204	0.0222	0.0565	0.0406
		0.8	0.0151	0.0145	0.0339	0.0304	**0.0026**	**0.0036**	0.0037	0.0047	0.0024	0.0028	0.0073	0.0066
		1.0	0.0027	0.0020	0.0076	0.0043	**0.0005**	**0.0013**	0.0006	0.0014	0.0001	0.0002	0.0009	0.0009

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

Table 6. Swamping effect (pswamp) as a performance measure for agglomerative clustering algorithms

Algorithm			SL-Satari/Di		SL-Chang		CL-Satari/Di		CL-Chang		AL-Satari/Di		AL-Chang	
Outlier %			5%	15%	5%	15%	5%	15%	5%	15%	5%	15%	5%	15%
n	κ	λ												
15	1.5	0.0	**0.1404**	**0.1409**	0.1452	0.1464	0.3567	0.3575	0.3629	0.3619	0.2136	0.2119	0.2151	0.2138
		0.2	**0.1381**	**0.1351**	0.1433	0.1403	0.3590	0.3638	0.3644	0.3707	0.2115	0.2113	0.2137	0.2132
		0.4	**0.1301**	**0.1305**	0.1361	0.1350	0.3599	0.3656	0.3655	0.3746	0.2031	0.2114	0.2061	0.2143
		0.6	**0.1166**	**0.1134**	0.1216	0.1174	0.3548	0.3623	0.3632	0.3725	0.1958	0.1902	0.1981	0.1916
		0.8	**0.1013**	**0.1017**	0.1047	0.1050	0.3555	0.3599	0.3629	0.3726	0.1770	0.1725	0.1781	0.1781
		1.0	**0.0964**	**0.1006**	0.0991	0.1048	0.3453	0.3583	0.3544	0.3710	0.1659	0.1646	0.1690	0.1706
	5	0.0	**0.1560**	**0.1558**	0.2334	0.2387	0.3001	0.2979	0.3000	0.2978	0.2230	0.2243	0.2296	0.2287
		0.2	**0.1344**	**0.1258**	0.1957	0.1701	0.2801	0.2663	0.2800	0.2662	0.2029	0.1900	0.2078	0.1949
		0.4	**0.0725**	**0.0584**	0.0960	0.0665	0.2108	0.1641	0.2102	0.1639	0.1243	0.0924	0.1278	0.0957
		0.6	**0.0204**	**0.0165**	0.0222	0.0176	0.1156	0.0729	0.1130	0.0729	0.0423	0.0274	0.0437	0.0276
		0.8	**0.0038**	**0.0045**	0.0038	0.0045	0.0618	0.0396	0.0567	0.0402	0.0107	0.0077	0.0110	0.0077
		1.0	**0.0048**	**0.0022**	0.0048	0.0021	0.0509	0.0319	0.0496	0.0337	0.0048	0.0031	0.0048	0.0031
50	1.5	0.0	**0.1114**	**0.1114**	0.0955	0.0966	0.6626	0.6607	0.6231	0.6212	0.4765	0.4766	0.4019	0.4060
		0.2	**0.1105**	**0.1117**	0.0965	0.0960	0.6603	0.6678	0.6255	0.6279	0.4820	0.4865	0.4041	0.4129
		0.4	**0.1095**	**0.1130**	0.0946	0.0981	0.6623	0.6684	0.6256	0.6359	0.4772	0.4916	0.4119	0.4256
		0.6	**0.1083**	**0.1176**	0.0920	0.1038	0.6590	0.6645	0.6224	0.6371	0.4681	0.4847	0.4009	0.4200
		0.8	**0.1068**	**0.1206**	0.0894	0.1089	0.6509	0.6520	0.6174	0.6267	0.4629	0.4709	0.3949	0.4110
		1.0	**0.1033**	**0.1238**	0.0867	0.1146	0.6556	0.6509	0.6215	0.6247	0.4599	0.4657	0.3961	0.4076
	5	0.0	**0.1093**	**0.1096**	0.3872	0.3852	0.5660	0.5706	0.4837	0.4865	0.4276	0.4284	0.3339	0.3314
		0.2	**0.0966**	**0.0885**	0.3521	0.3298	0.5592	0.5583	0.4755	0.4705	0.4144	0.4038	0.3168	0.3123
		0.4	**0.0519**	**0.0519**	0.1483	0.1772	0.5120	0.5233	0.4187	0.4354	0.3124	0.2987	0.1931	0.1923
		0.6	**0.0204**	**0.0290**	0.0189	0.0325	0.4858	0.4849	0.3875	0.3979	0.2035	0.1828	0.0737	0.0716
		0.8	**0.0092**	**0.0137**	0.0045	0.0083	0.4696	0.4583	0.3665	0.3792	0.1290	0.1057	0.0302	0.0292
		1.0	**0.0058**	**0.0077**	0.0023	0.0041	0.4682	0.4567	0.3649	0.3761	0.1105	0.0867	0.0218	0.0216

3.2 Performance Analysis for COD Procedure

The performance analysis particularly with OS1 were performed by comparing the performance of clustering algorithms. The performance of the COD procedure also analyzed and compare in various data conditions and confidence interval (C.I) of stopping rule. The results were summarized in table and the graphs are plotted in various data conditions as measured by *pout, pmask* and *pswamp* to visualized:

i. the best clustering algorithm
ii. the best concentration parameter
iii. the best percentage of outliers
iv. the best sample sizes
v. the best confidence interval of stopping rule

3.2.1 Comparison Analysis of Clustering Algorithms

The performance of six clustering algorithms are analyzed and compared based on *pout, pmask* and *pswamp*. Table 7 show the best algorithms on identifying multiple outliers in univariate circular datasets across varying conditions as measured by *pout, pmask* and *pswamp* for OS1.

The results show that the performance of CL-Satari/Di is the best by referring to the *pout* value especially when $\kappa \le 2$ and $n \ge 30$ for any percentage of outliers and both 90% and 95% C.I. of stopping rule. CL-Chang has the best performance to successfully detect all outliers especially when concentration parameter, $\kappa \le 2$ and small sample size, $n = 15$ for all percentage of outliers and both 90% and 95% C.I. of stopping rule. Moreover, the simulation study found that, AL-Chang has the best performance for *pout* especially when $\kappa = 100$ and $n = 100$. Meanwhile, SL-Chang is the best algorithms when the dataset has large concentration parameter and sample size $(\kappa = 100,\ n = 100)$ for all percentage of outliers and both 90% and 95% C.I. of stopping rule.

The summary result of *pmask* in Table 7 shows that CL-Satari/Di has the best performance on detecting outliers with low masking effect for most of data conditions. However, the result shows that CL-Chang is less sensitive to masking effect especially when the dendrogram is cut at 95% C.I. of stopping rule for small sample size, $n = 15$ when $\kappa < 10$. The simulation study found that AL-Chang has the best performance for *pmask* especially when $\kappa = 100$ and $n \ge 50$ for any percentage outliers and both 90% and 95% C.I. of stopping rule. This indicates that when the dataset has large concentration parameter and sample size, AL-Chang is the best to detect outliers in univariate circular data for OS1 without or less likely to falsely detect outliers as inliers. The result in Table 7 also show that the performance of SL-Satari/Di is the best for all data conditions by referring to the *pswamp* value, indicating that this algorithm is less sensitive to swamping effect. Thus, SL-Satari/Di is the best clustering algorithms on detecting outliers in univariate circular biological data without or less likely to falsely detect inliers as outliers for OS1 regardless of concentration parameter, sample size, percentage of outliers and C.I. of stopping rule. Figure 1 shows a few plots of all clustering algorithms performance for OS1 with combination of certain conditions such as $n = 15$, $\kappa = 1.5$ with 5% of outliers and $n = 50$, $\kappa = 5$ for 15% of outliers.

Generally, the curves of *pout, pmask* and *pswamp* show similar pattern for both 5% and 15 % of outliers. It can be seen

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

Table 7. The best algorithms on identifying multiple outliers in univariate circular data for highlighted conditions

95% C.I	***pout***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	CL-Chang	CL-Chang	SL-Chang	AL-Chang	CL-Chang	CL-Chang	SL-Chang	AL-Chang
$n=50$	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang
90% C.I	***pout***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang
$n=50$	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang
95% C.I	***pmask***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	CL-Chang	CL-Chang	SL-Chang	AL-Chang	CL-Chang	CL-Chang	SL-Chang	AL-Chang
$n=50$	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang
90% C.I	***pmask***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang	CL-Satari/Di	CL-Satari/Di	SL-Chang	AL-Chang
$n=50$	CL-Satari/Di	CL-Satari/Di	CL-Satari/Di	AL-Chang	CL-Satari/Di	CL-Satari/Di	CL-Satari/Di	AL-Chang
95% C.I	***pswamp***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di
$n=50$	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di
90% C.I	***pswamp***							
	5% Outliers				15% Outliers			
Sample size	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$	$\kappa=1.5$	$\kappa=5$	$\kappa=50$	$\kappa=100$
$n=15$	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di
$n=50$	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di	SL-Satari/Di

from Figure 1, the curves of *pout* gradually increase and approaching to 1 as λ increase. The larger the sample size, the faster the curves approach to 1. Besides, it can be seen that the *pmask* gradually decreases as the λ increase and as n increase, the faster the curves of *pmask* approaching to zero. Similarly happen to the curves of *pmask* when κ increase, the faster the curves approaching to zero. Meanwhile, the plots of *pswamp* display in Figure 1 show that the value of *pswamp* remain constant as the value of λ increase. However, as n and κ increase, the curves slightly approaching to zero as the value of λ increase especially for SL-Chang, AL-Satari/Di, and AL-Chang algorithms.

3.2.2 Comparison Analysis of Concentration Parameter

There are seven concentration parameters considered in this study, $\kappa=1.5,2,5,10,30,50,100$. In this study, $\kappa=1.5$ is considered small concentration parameter which means that the dataset has a very large spread, meanwhile $\kappa=100$ is considered large concentration parameter which means that the dataset is concentrated at mean angle. The simulation study found that $\kappa=100$ is the best concentration parameter value on detecting multiple outliers in univariate circular data with various conditions as measured by *pout*, *pmask* and *pswamp* for OS1. This result shows that when $\kappa=100$, the outliers are most likely to detect correctly without masking and swamping effects for OS1 using both 90% and 95% C.I of stopping rule. In other words, $\kappa=100$ is large enough to explain how concentrated the dataset towards the mean angle and make the outliers obviously separated from the others. Hence, the clustering algorithms are able to detect outliers correctly without masking and swamping effects regardless of sample size, percentage of outliers and confidence interval of cutting height.

Figure 2 show the plots of *pout*, *pmask* and *pswamp* versus λ for OS1 using SL-Satari/Di and CL-Satari/Di clustering algorithms by looking at different value of κ with selected conditions using 95% C.I of stopping rule. Generally, the performance of all clustering algorithms shows similar pattern for *pout*, *pmask* and *pswamp* for most data conditions. The value of *pout* gradually increases and approaching to 1 as the value of λ increases for all value of κ for any other data conditions. The pout values approaching to 1 faster even with small λ when $\kappa\geq 30$ for all clustering algorithms. Meanwhile, the plots of *pmask* in Figure 2 show that the *pmask* values decreasing and approaching to zero faster even with small λ especially when $\kappa\geq 30$. Similarly, the *pswamp* values also declining and gradually approaching to zero as the value of λ increases especially when $\kappa=100$. The plots of *pswamp* show that the values remain constant at specific value as

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

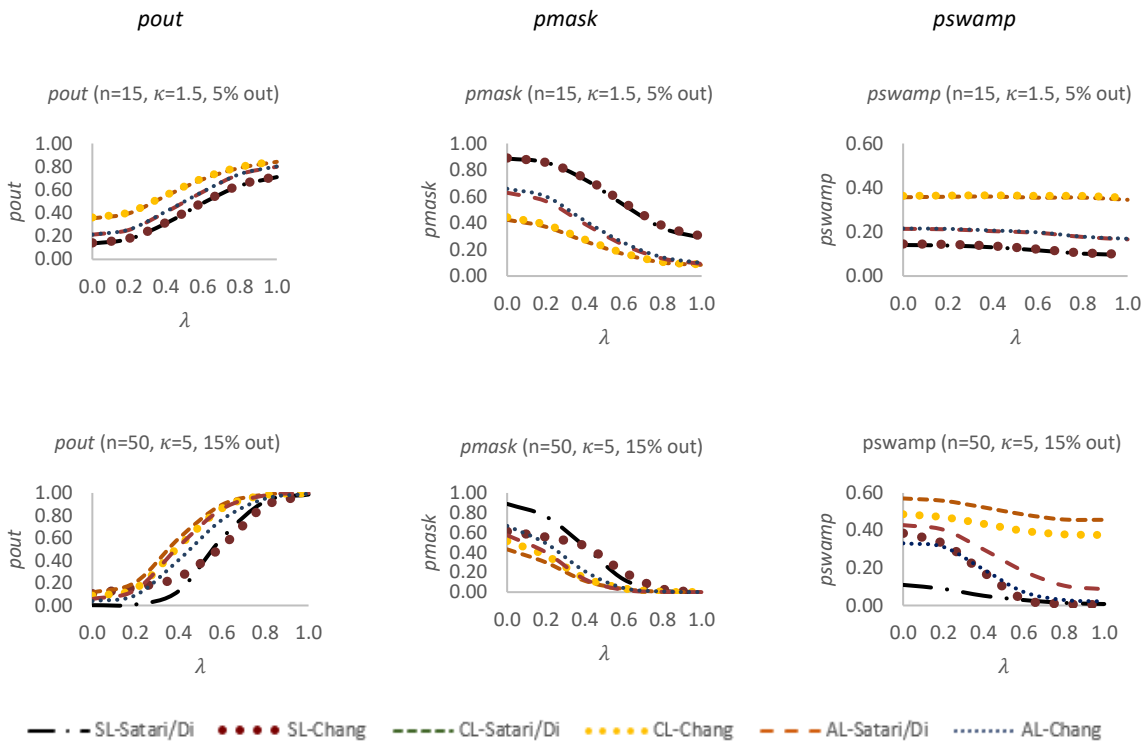


Figure 1. Plots *pout*, *pmask* and *pswamp* versus λ with selected condition ($n=15$, $\kappa=1.5$ with 5% outliers and $n=50$, $\kappa=5$ for 15% of outliers) using 95% C.I stopping rule

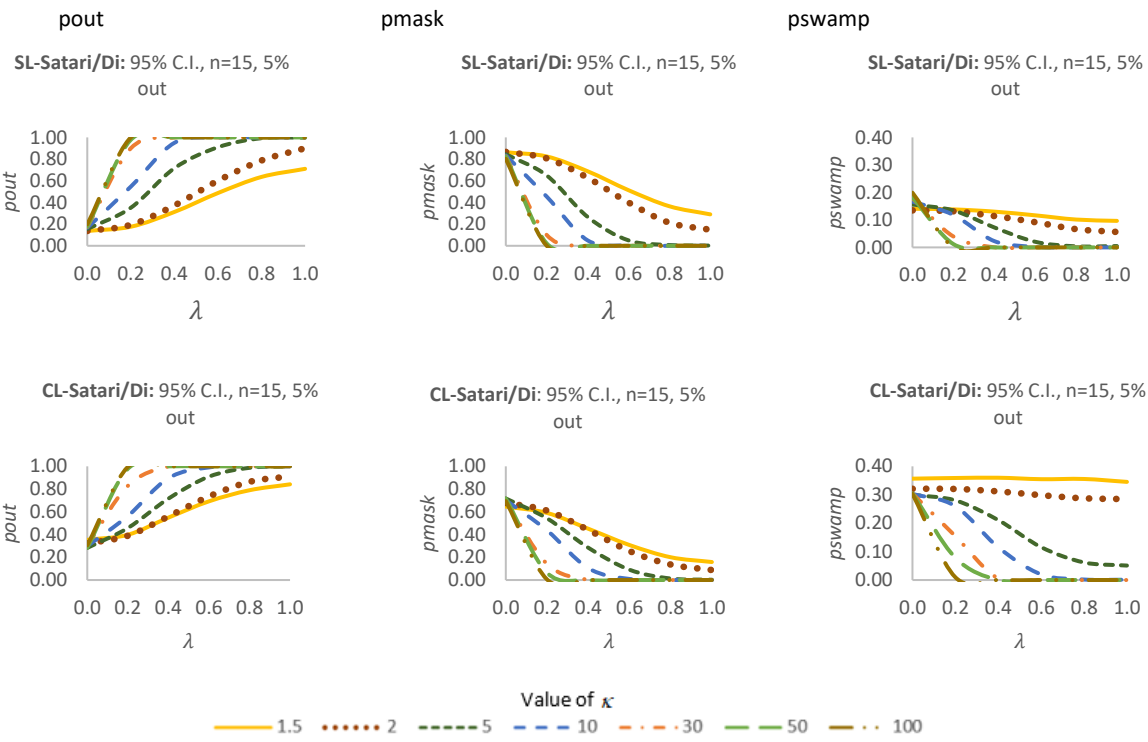


Figure 2. OS1: Plot of *pout*, *pmask* and *pswamp* versus λ for SL and CL algorithms with selected conditions ($n=15$ with 5% of outliers using 95% C.I of stopping rule

the value of λ increases particularly when $\kappa \leq 2$. Hence, the plots in Figure 2 show that the performance of *pout*, *pmask* and *pswamp* for OS1 are at the best even with small λ when $\kappa=100$ for all clustering algorithms. This study also found that the results for the best concentration parameter are similar when applying the 90% C.I of stopping rule.

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

3.2.3 Comparison Analysis of Percentage Outliers

In simulation study, the outliers are planted in simulated data generated from von Mises distribution with outliers strategically placed at a specified distance from the inlying observations using the formulation of OS1. Four percentage of outliers are chosen which are 5%, 10%, 15% and 20% from the sample size. Table 8 summarized the best percentage outlier for *pout*, *pmask* and *pswamp* by considering the highlighted data conditions for all clustering algorithms using both 90% and 95% C.I of stopping rule based on the indicators *pout*, *pmask* and *pswamp*, respectively.

found that 5% is the best percentage of outliers for all clustering algorithms by using 90% and 95% C.I of cutting height regardless any n and κ. It indicates that the probability for the clustering algorithms falsely detected outliers as inliers is lower when the outliers are not more than 5% from sample size. Table 8 also shows that SL-Satari/Di and SL-Chang might have low swamping effect (*pswamp*) when minimal outlier presence (5%), particularly when $n \geq 30$. In contrary, CL-Satari/Di, CL-Chang, AL-Satari/Di and AL-Chang might have small values of *pswamp* when the dataset consists of 20% of outliers across most data conditions. It indicates that when the percentage of outliers are high and the

Table 8. Summary of the best percentage outlier based on three performance indicators by considering the highlighted sample size for all and both 90% and 95% C.I of stopping rule

Clustering algorithms	*pout*		*pmask*		*pswamp*	
	$n=15$	$n \geq 30$	$n=15$	$n \geq 30$	$n=15$	$n \geq 30$
SL-Satari/Di	5%	5%	5%	5%	20%	5%
SL-Chang	5%	5%	5%	5%	5%	5%
CL-Satari/Di	5%	5%	5%	5%	20%	20%
CL-Chang	5%	5%	5%	5%	20%	20%
AL-Satari/Di	5%	5%	5%	5%	20%	20%
AL-Chang	5%	5%	5%	5%	20%	20%

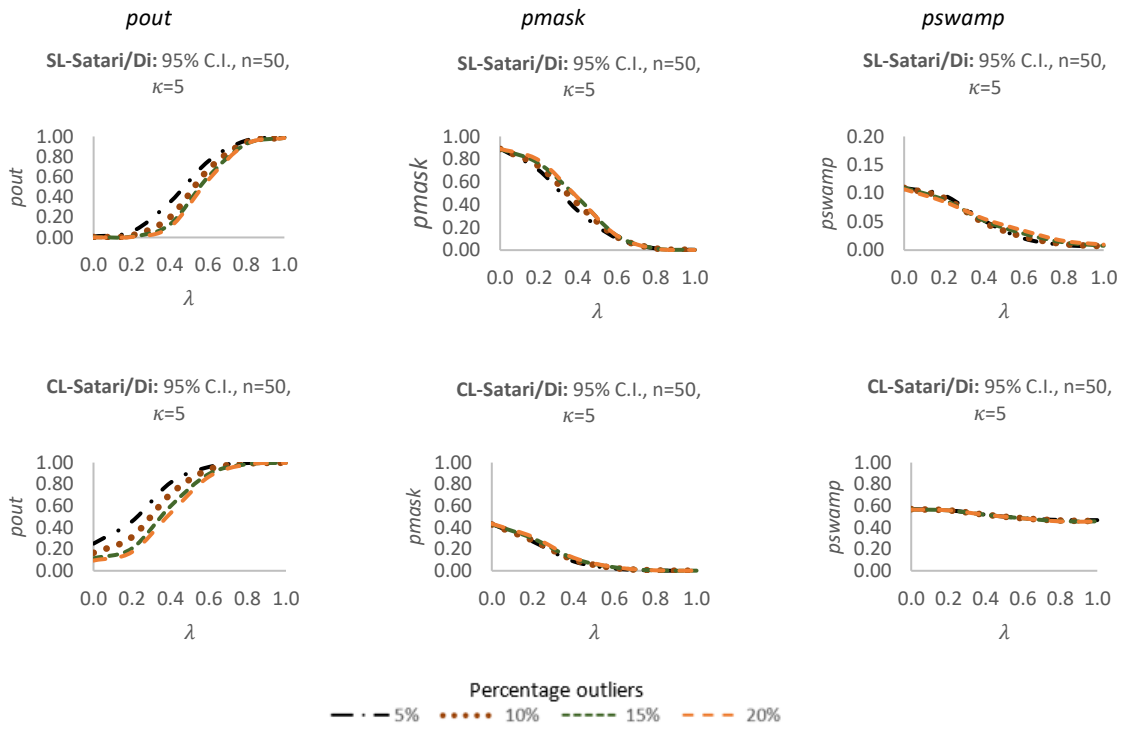


Figure 3. Plots *pout, pmask* and *pswamp* versus λ for different percentage of outliers with selected clustering algorithms for dataset $n = 50$ with $\kappa = 15$ using 95% C.I stopping rule

The simulation results show that 5% of outliers is the best condition for any clustering algorithms to successfully detect the correct outliers in univariate circular data in any data conditions using both 90% and 95% C.I of stopping rule . The performance of *pout* indicates that the outliers are more successfully detected when the dataset have small percentage of outliers for any clustering algorithms regardless any n and κ for both 90% and 95% C.I of stopping rule. Similarly, the results based on *pmask* probability of falsely detecting inliers as outliers is very low for CL-Satari/Di, CL-Chang, AL-Satari/Di and AL-Chang.

Figure 3 show the plot of *pout, pmask* and *pswamp* performance against level of contamination, λ for OS1 using SL-Satari/Di and CL-Satari/Di algorithms. The results are illustrated across varying percentages of outliers, based on selected data conditions and a 95% confidence interval applied to the stopping rule. The plots of *pout* for all clustering algorithms show that 5%

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

of outliers performed the best compared to others. It is also show that the *pmask* values decreasing as λ increasing and it can be seen that 5% of outliers producing the lowest value of *pmask* compare to others for all algorithms. The *pswamp* plots in Figure 3 also show various pattern and it is found that the *pswamp* values are almost the same for all percentage outliers. The *pswamp* plots show that the *pswamp* value remain constant at specific value as the value of λ increases especially when the data has small sample size and concentration parameter for most of clustering algorithms. This study also found that the plots presented in Figure 3 have similar pattern to the results obtained using the 90% C.I of stopping rule.

3.2.4 Comparison Analysis of Sample Sizes

Four sample sizes, $n=15,30,50,100$ are chosen in this study where $n=15$ is considered small sample size and $n=100$ is considered as large sample size. This study found that the best sample size using the 90% C.I of stopping rule is similar to that obtained with the 95% C.I of stopping rule. The results from simulation study found that the probability of algorithm to detect outliers correctly (*pout*) when $n=15$ using 90% and 95% C.I of stopping rule. Meanwhile, the algorithms are mostly to detect outliers with low masking effect (*pmask*) when $n=100$ regardless any conditions for both C.I of stopping rule. This indicates that when sample size is large, any clustering algorithms used in this study have less sensitive to masking effect which occurs when outliers falsely detected as inliers regardless any conditions. The simulation study also found that the best sample size for outlier detection with low swamping effect is when $n=15$ by using any clustering algorithms for most data conditions. This indicates that when sample size is small, the clustering algorithms use in this study are less sensitive to swamping effect which occurs when inliers falsely detected as outliers regardless any conditions. Generally, the performance of all clustering algorithms shows similar pattern for *pout, pmask* and *pswamp* for most data conditions using both 90% and 95% C.I of stopping rule. Figure 4 visualized the plots of *pout, pmask* and *pswamp* against varying λ for OS1 using SL-Satari/Di and AL-Chang algorithms. The results are shown for selected data conditions with a 95% confidence interval applied to the stopping rule. The plots of *pout* in Figure 4 show that SL-Satari/Di achieved superior performance when $n=15$ using 95% C.I of stopping rule. The plot of *pmask* for SL-Satari/Di and AL-Chang in Figure 4 show the *pmask* performance have slightly differences among all sample size. Moreover, the plots of *pswamp* in Figure 4 found that most of clustering algorithms value remain constant at certain value as the value of λ increases. In conclusion, the plots in Figure 4 indicate that most clustering algorithms perform the best in detecting outliers with minimal swamping effects under small sample conditions but might lead to high masking effect.

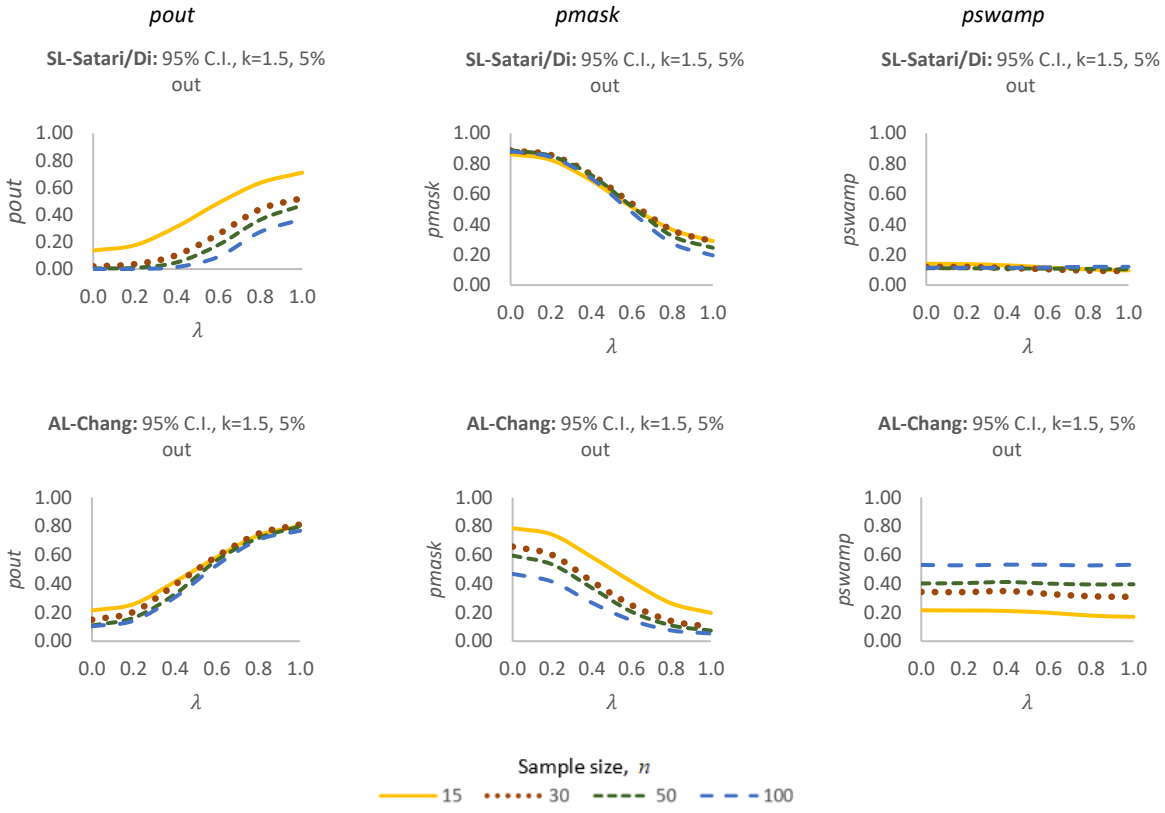


Figure 4. Plots *pout*, *pmask* and *pswamp* versus λ for selected clustering algorithms for selected dataset ($\kappa=1.5$ with 5% of outliers) using 95% C.I of stopping rule

3.2.5 Comparison Analysis of Confidence Interval of Stopping Rule

The results based on *pout* and *pmask* indicate that a 90% C.I for the stopping rule provides the most effective performance in identifying multiple outliers within univariate circular data across

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

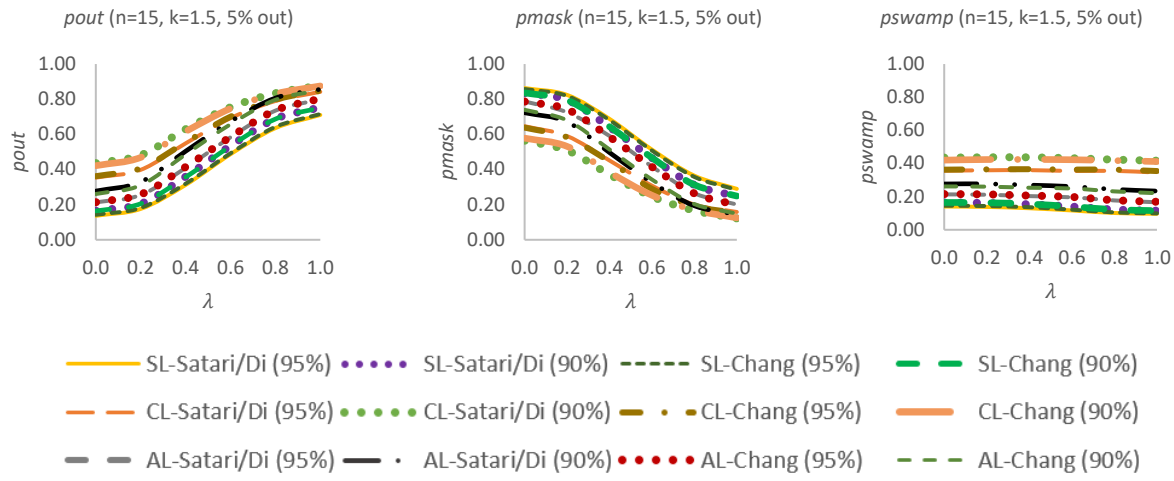


Figure 5. Plots *pout*, *pmask* and *pswamp* versus λ for all algorithms for $n=15$, $\kappa=1.5$ with 5% of outliers

most data conditions, regardless of the clustering algorithm applied. The plots of *pout* and *pmask* against value of λ in Figure 5 show that 90% C.I of stopping rule from all clustering algorithms has better performance even though the differences with 95% C.I are slightly noticeable. This indicates that by cutting the height of dendrogram at 90% C.I, any clustering algorithms used in this study are most likely to successfully detect outliers correctly without or with low masking effect. However, 95% C.I is the best for most data conditions by referring to the *pswamp* value as visualized in Figure 5 which indicates that 95% C.I of stopping rule is less sensitive to swamping effect by using any clustering algorithms.

4. Conclusions

This study aims to assess the effectiveness of the COD procedure in detecting outliers within univariate circular datasets. The simulation was carried out using synthetic datasets, where outliers were introduced using the formulation of OS1. The outliers from OS1 are introduced by shifting the mean direction of the inliers by a specific distance defined by the contamination level. The mean direction was initially set to zero and the outliers are placed at a specific distance from the inliers depending on the contamination level, λ. The value of λ defines how far the outliers are moved from the central of inlying data points or called inliers. There are six different λ used in the simulation study and this study shows that different magnitudes of contamination in OS1 affect the outlier detection process. The simulation results show that the *pout* values gradually increase and approaching to one as λ and κ increase. Meanwhile, the *pout* values gradually decreasing as the percentage of outliers and n increase. The *pmask* values gradually decreasing and approaching to zero as λ and κ increase. As the percentage of outliers and n increase, the *pmask* values are slightly increasing. As conclusion from *pswamp* results for OS1, the values of *pswamp* gradually decreasing and approaching to zero as the values of λ, κ and n increase. However, the *pswamp* values slightly increase as the percentage of outliers increase. Generally, the results of *pout*, *pmask* and *pswamp* performance revealed that all clustering algorithms have similar result pattern. Table 9 presents a summary of the best-performing results from the simulation study under Outlier Scenario 1 (OS1) using three performance measure which are *pout, pmask* and *pswamp*

The result from simulation study with OS1 found that CL-Satari/Di has the best performance on detecting outliers with low masking effect for particular data conditions. Meanwhile, the performance of SL-Satari/Di is not good compared to other algorithms for *pout* and *pmask* but this algorithm found to be the best algorithm based on *pswamp*. This indicates that SL-Satari/Di is less sensitive to swamping effect for OS1. This study found that SL-Satari/Di is convincing to identify outliers in univariate circular data without or less likely to misidentified inliers as outliers for OS1. The results show that the best sample size for detecting outliers varies depending on the clustering algorithm and data conditions. It is found that large sample size is less sensitive to masking effect, while a smaller sample size is less sensitive to swamping effect. This study also shows that the mean direction and contamination levels play a crucial role in determining how outliers are generated. Thus, the topic on outlier scenario is interesting to study deeper as it can be extended by adjusting the concentration parameter to define the new outliers scenario and compare with OS1 from this study.

In addition, this paper focuses on simulation-based validation to assess the performance of the COD procedure under various conditions. A companion studies from Zulkipli et al. (2024a, 2024b) present a real-life data application of the method to univariate circular biological data. Together, these studies provide both theoretical and practical validation of the COD procedure. Future studies should consider to extend the evaluation by comparing the best algorithm, CL-Satari/Di from COD procedure

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

against other existing outlier detection methods for univariate circular data.

5. Acknowledgments

The authors sincerely appreciate the thoughtful assessments and constructive suggestions provided by the editors and reviewers, which have played an important role in improving the overall quality of this manuscript. Financial support for this research was provided by Universiti Malaysia Pahang Al-Sultan Abdullah and the Ministry of Higher Education Malaysia through the UMPSA Internal Grant (PGRS210328), RDU1901168, and FRGS/1/2019/STG06/UMP/02/6.

6. References

Abuzaid, A. H. (2020). Detection of outliers in univariate circular data by means of the outlier local factor (LOF). *Statistics in Transition New Series, 21*(3), 39–51.

Abuzaid, A. H., Hussin, A. G., Rambli, A., & Mohamed, I. (2012). Statistics for a new test of discordance in circular data. *Communications in Statistics—Simulation and Computation*, *41*, 1882–1890.

Abuzaid, A. H., Mohamed, I. B., & Hussin, A. G. (2009). A new test of discordancy in circular data. *Communications in Statistics - Simulation and Computation*, *38*(4), 682–691.

Collett, D. (1980). Outliers in circular data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(1), 50–57.

Di, N. F. M., & Satari, S. Z. (2017). The effect of different distance measures in detecting outliers using clustering-based algorithm for circular regression model. *AIP Conference Proceedings*, *1842*, 030016. https://doi.org/10.1063/1.4982059

Di, N. F. M., Satari, S. Z., & Zakaria, R. (2017). Detection of different outlier scenarios in circular regression model using single-linkage method. *Journal of Physics: Conference Series*, *890*(1). https://doi.org/10.1088/1742-6596/890/1/012088.

El, H.I., Ahmed, S., Abuzaid, A.H., & Awar, I.I. (2019). Detection of Outliers in Circular Data using Kernel Density Function. *Life Sciences: An International Journal (LSIJ)*, *1*(1), 1–11.

Hung, W. L., Chang-Chien, S. J., & Yang, M. S. (2012). Self-updating clustering algorithm for estimating the parameters in mixtures of von Mises distributions. *Journal of Applied Statistics*, *39*(10), 2259–2274.

Johnson, R., & Wichern, D. (2014). *Applied multivariate statistical analysis* (6th ed.). Pearson.

Mahmood, E. A., Rana, S., Hussin, A. G., & Midi, H. (2017a). Adjusting outliers in univariate circular data. *Pertanika Journal of Science & Technology*, *25*(4), 1147–1158

Mahmood, E. A., Rana, S., Midi, H., & Hussin, A. G. (2017b). Detection of outliers in univariate circular data using robust circular distance. *Journal of Modern Applied Statistical Methods*, *16*(2), 418–438.

Rambli, A., Ibrahim, S., & Mohamed, I. (2012). On discordance test for the wrapped normal data. *Sains Malaysiana*, *41*(6), 769–778.

Ramlee, I. M., Ibrahim, S., Zhe, L. W., & Yusoff, M. I. (2024). Detection of Outliers in Circular Regression Model via DFBETAcIS Statistic. *Sains Malaysiana*, *53*(4), 935–951.

Rattanawong, B., Papukdee, N., & Srisodaphol, W. (2023). Detection of Outliers in Univariate Circular Data Using New Cut-off Points for the Circular Distance. *Philippine Journal of Science, 152*(5), 2019–2026.

Satari, S. Z. (2015). *Parameter estimation and outlier detection for some types of circular model* (Doctoral dissertation, University of Malaya).

Satari, S. Z., & Ku Khalif, K. M. N. (2020). Review on outliers identification methods for univariate circular biological data. *Advances in Science, Technology and Engineering Systems*, *5*(2), 95–103.

Satari, S. Z., Muhammad Di, N. F., Zubairi, Y. Z., & Hussin, A. G. (2021). Comparative study of clustering-based outliers detection methods in circular-circular regression model. *Sains Malaysiana*, *50*(6), 1787–1798.

Sebert, D. M., Montgomery, D. C., & Rollier, D. A. (1998). A clustering algorithm for identifying multiple outliers in linear regression. *Computational Statistics and Data Analysis*, *27*, 461–484.

Sert, S., & Kardiyen, F. (2023). A method for detecting outliers in linear-circular non-parametric regression. *PLoS ONE*, *18*(6 June), 1–17.

Zulkipli, N. S., Satari, S. Z., & Wan Yusoff, W. N. S. (2022). The effect of different similarity distance measures in detecting outliers using single-linkage clustering algorithm for univariate circular biological data. *Pakistan Journal of Statistics and Operation Research*, *18*(3), 561–573.

Zulkipli, N. S., Satari, S. Z., & Wan Yusoff, W. N. S. (2024a). The Multiple Outliers Detection for Circular Univariate Data Using Different Agglomerative Clustering Algorithms. *AIP Conference Proceedings,* 2895(1): p. 090007.

Zulkipli, N. S., Satari, S. Z., & Wan Yusoff, W. N. S. (2024b). Abnormalities detection in Apert syndrome using hierarchical clustering algorithms. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *44*(4), 135–174.

DOI: https://doi.org/10.22452/mjs.vol45sp2.9
Malaysian Journal of Science 45 (Special Issue 2): 83-94(November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026) https://mjs.um.edu.my

Forecasting Indonesia's Inflation-at-Risk Using Vector Autoregression with Exogenous Variables Model

Ni Kadek Juliarini[1a], Dedy Dwi Prastyo[2a*], Muhammad Sjahid Akbar[3a]

Abstract: Inflation is a critical economic indicator influencing stability and monetary policy, making accurate forecasting essential. This study employs the Vector Autoregression with Exogenous Variables (VARX) model to simultaneously forecast Indonesia's inflation across various sectors, including food, health, housing, clothing, and education, incorporating the Financial Condition Index (FCI) and Financial Stress Index (FSI). The FSI evaluates financial stress using data from banking, foreign exchange, stock markets, and government securities, while the FCI assesses economic conditions through Principal Component Analysis (PCA). VARX is rarely used in this context and differs from previous studies in that it focuses on the risk of extreme inflation. In addition, this study introduces the use of the Correlation Information Criterion (CIC) as an evaluation tool, which has not been widely applied in similar research. The fluctuation patterns of each sectoral inflation generally display distinct characteristics, enabling a comprehensive evaluation of the VARX model's ability to forecast data with varying fluctuation patterns. This approach differs from previous studies, which typically utilize data with uniform fluctuation patterns. Furthermore, the Seasonal Trend Decomposition Based on Loess (STL) method is employed to mitigate seasonal effects in the data and improve forecasting accuracy. This technique is rarely applied in this context. Analyzing monthly data from February 2010 to December 2024, the VARX(1) model effectively forecasts Indonesia's Inflation-at-Risk (IaR) with high accuracy, as measured using the Symmetric Mean Absolute Percentage Error (SMAPE) criterion. Seasonal inflation in the health sector is addressed using the Seasonal Trend Decomposition Based on Loess (STL) method, which further improves forecasting precision. However, the model struggles with extreme fluctuations, highlighting the need for future research into more advanced predictive models for volatile economic conditions.

Keywords: IaR, FCI, FSI, VARX, Forecasting.

1. Introduction

Economic growth is a key measure of societal welfare and is closely linked to inflation. High inflation often slows economic expansion, while low and stable inflation supports sustained growth. Inflation is defined as a continuous, widespread rise in the prices of goods and services, pushing overall price levels upward. Estimates of inflation-related information are essential for monitoring economic pressures that may impact the financial sector. Inflation-at-Risk (IaR) is a concept used to forecast inflation by considering risks arising from financial variables, specifically the Financial Stability Index (FSI) and the Financial Conditions Index (FCI). A widely used approach in analyzing economic dynamics is the Vector Autoregression (VAR) model, which was first introduced by Christopher Albert Sims in 1972, building upon ideas developed by Granger and Clive in 1969. The VAR model can be used for multivariate analysis, where all variables are considered to be endogenous. Each variable is explained by its past values as well as the past values of other endogenous variables in the system (Gujarati, 2003).

Authors information:
[a]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, INDONESIA. E-mail: juliarini.kdk@gmail.com[1], dd.prastyo@its.ac.id[2], m_sjahid_a@its.ac.id[3]

*Corresponding Author: dd.prastyo@its.ac.id

The VAR model is commonly utilized for multivariate time series analysis and has several extensions, including VARX for incorporating exogenous variables (Sesay et al., 2020; Farih & Prastyo, 2022), Generalized Space-Time Autoregressive (GSTAR), and GSTARMA for accommodating spatial effects (Akbar, et al., 2020), also with exogenous variables become GSTARX (Suhartono et al., 2018), and hybrid models combining VAR with Machine Learning techniques (Suhartono et al., 2018; Prastyo et al., 2018; Hikmawati et al., 2021; Novianto et al., 2018). In comparison to previous studies, Safuan et al. (2022) conducted a study to investigate the intermittent relationship between financial stress and key macroeconomic variables in Indonesia using the Financial Stress Index (FSI) model. The conclusion drawn from the study is that there is an episodic relationship between financial stress and Indonesia's key macroeconomic variables. Additionally, the FSI developed in the study can capture unusual financial stress events, such as the global financial crisis of 2008 and the early stages of the COVID-19 pandemic in Indonesia. Therefore, for future research related to macroeconomic variables, the FSI can be considered a key variable in the study.

Based on this discussion, given the importance of inflation forecasting, this research aims to forecast IaR using the VARX model, employing inflation data from categories such as food, health, housing, clothing, and education, which reflect the basic needs of the population. Furthermore, economic indicators like

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

the FSI and the FCI are also used. The FSI is developed to assess reactions to systemic vulnerabilities within the financial system (Park & Mercado, 2014). Meanwhile, FCI provides greater benefits for examining the relationship between macroeconomic variables (Juhro & Iyke, 2019).

2. Literature Review

This section discusses the literature review that underpins the research. The statistical theories that form the foundation of this study include Vector Autoregression with Exogenous Variables (VARX), granger causality, Seasonal Trend Decomposition Based On Loess (STL), Financial Stress Index (FSI), Financial Condition Index (FCI), Indonesia's Inflation, and Symmetric Mean Absolute Percentage Error (SMAPE).

2.1 Vector Autoregression with Exogenous Variables (VARX) Model

The VAR model expands upon the univariate Autoregression model and is used to analyze multivariate time series data (Zivot & Wang, 2003). The equation for the VAR(*p*) model is expressed in the following form:

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad p > 0. \tag{1}$$

where $\mathbf{y}_t$ represents the vector of endogenous variables, $\boldsymbol{\phi}_0$ represents the intercept, $\mathbf{\Phi}_1$ represents the parameter of the first-order autoregression (AR), and $\{\boldsymbol{\varepsilon}_t\}$ is assumed to be white noise, with a mean of zero and a variance of σ^2. Meanwhile, VARX is an extension of the VAR model that includes exogenous variables. According to Ocampo & Rodríguez (2012), the VARX(*p*) model is expressed in the following equation form:

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\gamma}_t \mathbf{y}_t + \boldsymbol{\varepsilon}_t; \ \boldsymbol{p > 0} \tag{2}$$

where $\mathbf{x}_t$ denotes a vector of exogenous variables and $\boldsymbol{\theta}_q$ represents the parameter of exogenous variables at lag *q*.

2.2 Granger Causality

The Granger causality test identifies causal relationships between variables in a model and is usually conducted after selecting the optimal lag for the VAR model. For example, in the context of a bivariate model where $y_t = (y_{1,t}, y_{2,t})'$ and $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, the VAR(1) model is represented by the following two equations (Tsay, 2002):

$$y_{1,t} = \phi_{1,0} + \Phi_{1,11} y_{1,t-1} + \Phi_{1,12} y_{2,t-1} + \varepsilon_{1,t} \tag{3}$$

$$y_{2,t} = \phi_{2,0} + \Phi_{1,21} y_{1,t-1} + \Phi_{1,22} y_{2,t-1} + \varepsilon_{2,t} \tag{4}$$

where $\Phi_{1,ij}$ represents the ij-th element of Φ_1 and $\phi_{i,0}$ denotes the i-th element of ϕ_0. Based on Equation (3), $\Phi_{1,12}$ indicates the linear dependence of $y_{1,t}$ on $y_{2,t-1}$ given $y_{1,t-1}$. Therefore, $\Phi_{1,12}$ represents the conditional effect of $y_{2,t-1}$ on $y_{1,t}$ condition on $y_{1,t-1}$. If $\Phi_{1,12} = 0$, then $y_{1,t}$ does not depend on $y_{2,t-1}$ and the model shows that $y_{1,t}$ only depends on its own past value. Similarly, if $\Phi_{1,21} = 0$, then Equation (4) shows that $y_{2,t}$ does not depend on $y_{1,t-1}$ when $y_{2,t-1}$ is known. Considering Equations (3) and (4) simultaneously, if $\Phi_{1,12} = 0$ and $\Phi_{1,21} \neq 0$ there is a one-way relationship from $y_{1,t}$ to $y_{2,t}$. If $\Phi_{1,12} = \Phi_{1,21} = 0$ then $y_{1,t}$ and $y_{2,t}$ are independent. However, if $\Phi_{1,12} \neq 0$ and $\Phi_{1,21} \neq 0$, there is a two-way relationship between the two variables (Tsay, 2002).

2.3 Seasonal Trend Decomposition Based on LOESS (STL)

Seasonal Trend Decomposition, using Locally Estimated Scatterplot Smoothing (LOESS), also known as STL (Cleveland et al., 1990), is an algorithm developed to break down time series data into three key components: trend, seasonal, and residual. The LOESS is a nonparametric regression method that offers high flexibility, as the curve estimates are automatically formed based on the data, unaffected by outliers. Mathematically, the STL model can be expressed as follows:

$$\mathbf{y}_t = \mathbf{S}_t + \mathbf{T}_t + \boldsymbol{\varepsilon}_t \tag{5}$$

where S_t represents the seasonal component at time *t*, T_t refers to the trend component at time *t*, and ε_t represents the residual at time *t*. When applied to the VAR model, for instance, VAR(1), the equation model becomes as follows:

$$\mathbf{D}_t = \mathbf{\Phi}_1 + \mathbf{D}_{t-1} + \boldsymbol{\varepsilon}_t \tag{6}$$

where $\mathbf{D}_t$ is the vector of seasonal exogenous variables at time *t*, consisting of $D_{1t}, D_{2t}, \ldots, D_{nt}$.

2.4 Financial Stress Index (FSI)

The FSI is an index designed to detect periods of significant stress within the financial system, although not every instance of high stress will result in a crisis (Safuan et al., 2022). FSI measures financial pressure using data from banking, foreign exchange markets, stock markets, and government securities. The FSI is represented by the following equation (Safuan et al., 2022):

$$\text{FSI}_t = \omega_1 \beta_t + \omega_2 \text{EMPI}_t + \omega_3 \vartheta_t^2 + \omega_4 \text{SRET}_t + \omega_5 \text{DMP}_t, \tag{7}$$

where $\omega_j; j = 1,2,3,4,5$ represents the weight of each component based on its standard deviation. The equation for each component is presented as follows:

$$\beta_t = \frac{cov(rf, rj)}{var\,(rj)}, \tag{8}$$

where β_t is the banking stress level at the time t, the rf is the return on IDX Finance, and rj is the return on JKSE. The second component is the foreign exchange market pressure ($EMPI_t$) which can be calculated with the following equation:

$$\text{EMPI}_t = \frac{(\Delta e_t - \mu_{\Delta e})}{\sigma_{\Delta e}} - \frac{(\Delta \text{RES}_t - \mu_{\Delta \text{RES}})}{\sigma_{\Delta \text{RES}}}, \tag{9}$$

where Δe and ΔRES denote the percentage change in the value of the Indonesian rupiah against the US dollar and Indonesia's foreign exchange reserves, respectively. Meanwhile, μ and σ represent the mean and standard deviation. The third component is stock market volatility ϑ_t^2 which is assumed to follow the Generalized Autoregressive Conditional Heteroskedasticity (GARCH(1,1)) model, expressed by the following equation:

$$\vartheta_t^2 = \omega + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \sigma_{t-1}^2 \tag{10}$$

The fourth component is the stock market return ($SRET_t$), which can be calculated using the following equation:

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

$$\mathrm{SRET_t} = log(P_t) - log(P_{t-1}) \quad (11)$$

where P_t is the closing stock price of the JKSE at period t. If the return on the JKSE is lower, it indicates that stress in the stock market is higher, and conversely, if the return is higher, it indicates lower stress. The final component is the government securities market pressure (DMP_t), which can be calculated with the following equation:

$$\mathrm{DMP}_t = \mathrm{YIDGB}_t - \mathrm{YUST}_t \quad (12)$$

where *YIDGB* is the annual yield on Indonesian government bonds with a five-year maturity, and *YUST* is the annual yield on US government bonds with a five-year maturity.

2.5 Financial Condition Index

The FCI is a variable used to examine the connection between economic and financial factors (Juhro & Iyke, 2019). The FCI variable is constructed based on the concept of Principal Component Analysis (PCA). PCA is a multivariate analysis technique that aims to reduce the dimensionality of data by transforming correlated variables into a set of new, uncorrelated variables, while retaining as much variance as possible (Ferrara et al., 2022). These new variables are referred to as Principal Components (PC). In linear algebra, a PC is a linear combination of n random variables (y_1,y_2,…,y_n). Geometrically, this linear combination represents a new coordinate system obtained by rotating the original system to serve as the new coordinate axes. These new axes explain the maximum variance, resulting in a simpler covariance structure. The condition for forming a PC, which is a linear combination of the vector $y = (y_1, y_2, \cdots, y_n)'$ in order to have the most significant variance, it is achieved by selecting eigenvalues $e_i = (e_1,e_2,\ldots,e_n)'$ such that $Var(e_i' y)$ is maximized, and $e_i'e_i = 1$ and $Cov(e_i'y, e_m'y) = 0$ for $m < i$.

The formation of PCs can be done using either the covariance matrix or the correlation matrix. The formation of PCs using the correlation matrix is done when the observed variables do not have the same unit of measurement, requiring standardization. As a result, the variance-covariance matrix of the standardized data will be the same as the correlation matrix of the original data, and the total variance of the principal components will be equal to the sum of the variances of the original variables. The random vector $y = (y_1, y_2, \cdots, y_n)'$ has a variance-covariance matrix Σ with eigenvalues in descending order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, The linear combination of the PCs is as follows.

$$\begin{aligned} \mathrm{PC}_1 &= e_1'\mathbf{y} = e_{11}y_1 + e_{21}y_2 + \cdots + e_{n1}y_n \\ \mathrm{PC}_2 &= e_2'\mathbf{y} = e_{12}y_1 + e_{22}y_2 + \cdots + e_{n2}y_n \\ &\vdots \\ \mathrm{PC}_n &= e_k'\mathbf{y} = e_{1n}y_1 + e_{2n}y_2 + \cdots + e_{nn}y_n \end{aligned} \quad (13)$$

where PC_1 is the first PC which has the largest variance, PC_2 is the second PC with the second-largest variance, PC_n is the *n*-th PC with the *n*-th largest variance, and e_n is the eigenvector of the *n*-th variable. The general model of the *j*-th PC can be written as follows:

$$\mathrm{PC}_j = e_j'\mathbf{y}, \quad j = 1,2,\ldots,n. \quad (14)$$

Thus, $Var\,(PC_j) = e_j'\mathbf{\Sigma}e_j = \lambda_i$, $j = 1,2,\ldots,n$ and $Cov\,(PC_j, PC_m) = e_j'\mathbf{\Sigma}e_j = 0$, $j \neq m$. PCs are independent of each other, and their variances correspond to the eigenvalues of the covariance matrix **Σ**, as follows:

$$\begin{aligned} \sigma_{11} + \sigma_{22} + \cdots + \sigma_{nn} &= \sum_{i=1}^{n} Var(y_i) \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_n \\ &= \sum_{j=1}^{n} Var(\mathrm{PC}_j). \end{aligned} \quad (15)$$

The percentage of total variance explained by PC_j is as follows:

$$\text{Proportion of variance for the } j\text{-th} = \frac{\lambda_j}{\sum_{i=1}^{n}\lambda_i}. \quad (16)$$

If *m* PCs are chosen where $m < n$, then:

$$\text{Proportion of variance for the } m\text{-thPC} = \frac{\lambda_1+\lambda_2+\cdots+\lambda_m}{\lambda_1+\lambda_2+\cdots+\lambda_n} \quad (17)$$

If PCA is linear, then the covariance matrix of the standardized data is used, as the main diagonal of the matrix contains values of one. Thus, the total population variance for the standardized variables is *n* which is the sum of the diagonal elements of the correlation matrix (ρ), so:

$$\text{Proportion of variance for the } j\text{-th} = \frac{\lambda_j}{n} \times 100\%. \quad (18)$$

The number of PCs selected should ideally explain a high cumulative variance, for example, between 70% and 80%.

2.6 Symmetric Mean Absolute Percentage Error (SMAPE)

According to (Chicco et al., 2021), the SMAPE is a forecasting accuracy measure used to address the shortcomings of the Mean Absolute Percentage Error (MAPE), where MAPE becomes undefined when the actual value y_t is zero. The equation for MAPE is as follows:

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\%, \quad (19)$$

where y_t is the actual data at the time *t* and $\hat{y}_t$ is the predicted data of y_t. This differs from SMAPE, as it can still provide results even when the actual value is zero. The equation for SMAPE is written as follows:

$$SMAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|\hat{y}_t - y_t|}{(|\hat{y}_t| + |y_t|)} \times 100\%. \quad (20)$$

SMAPE has a lower bound of 0% and an upper bound of 100%,

Table 1. Forecasting Accuracy Levels

SMAPE	Accuracy level
≤ 10%	Highly Precise Forecasting
10% < SMAPE ≤ 20%	Precise Forecasting
20% < SMAPE ≤ 50%	Moderately Precise Forecasting
> 50%	Imprecise Forecasting

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

where SMAPE will automatically reach the upper limit when the actual value is zero. The evaluation values produced by SMAPE have the following criteria as stated in Table 1.

3. Methodology

This study utilizes data associated with macroeconomics, finance, and policy matters in Indonesia. All data used are numerical with a monthly frequency from February 2010 to December 2024, totaling 179 data points for each variable. The variables and data structure used in the study are presented in Tables 2 and 3, respectively.

where T represents the total count of data points utilized. The detailed study analysis is depicted in the flowchart presented in Figure 1.

4. Results and Discussion

The inflation data used in this study include inflation for food, health, housing, clothing, and education. Other variables include FSI and FCI. FSI is calculated using Equation (7), and the results are shown in Figure 2.

The FSI graph illustrates the dynamics of financial stress that occur over a certain period. An increase in FSI reflects significant stress, which can be caused by various factors such as global or domestic economic crises, market uncertainty, tight monetary policies, or external shocks like commodity price fluctuations and natural disasters. This condition is often followed by a decline in economic activity, reduced investment, and diminished market confidence. Conversely, a weakening FSI indicates reduced

Table 2. Research Variables

Variable		Description	
Inflation (y_i); i = 1,…,5	$INFL_1$ (y_1)	Food Inflation	
	$INFL_2$ (y_2)	Health Inflation	
	$INFL_3$ (y_3)	Housing Inflation	
	$INFL_4$ (y_4)	Clothing Inflation	
	$INFL_5$ (y_5)	Education Inflation	
FSI (x_1)		IDX FIN ($x_{1,1}$)	Indonesia Stock Exchange Finance Index
		JKSE ($x_{1,2}$)	Jakarta Stock Exchange
		EXC ($x_{1,3}$)	Exchange Rate (USD IDR)
		CADEV ($x_{1,4}$)	Indonesia Foreign Exchange Reserves
		YIDGB ($x_{1,5}$)	Indonesia Bond Yield
		YUST ($x_{1,6}$)	US Bond Yield
$x_{1,i}$; i =1, …6 represents the components of FSI			
FCI (x_{2j}); $j = 1,2,\dots,s$		RATE ($x_{2j,1}$)	Interest Rate
		JIBOR ($x_{2j,2}$)	Jakarta Interbank Offered Rate
		JUB ($x_{2j,3}$)	Money Supply (M_2)
		CCI ($x_{2j,4}$)	Consumer Confidence Index
$x_{2\,j,i}$; i =1, …4 represents the components of FCI. s: the number of principal components (PCs) formed, such that the cumulative variance reaches between 70% and 80%.			

Table 3. Data Structure

		Endogenous Variables					Exogenous Variables			
t	Date	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$	$y_{4,t}$	$y_{5,t}$	$x_{1,t}$	$x_{21,t}$	⋯	$x_{2s,t}$
1	1/2/2010	$y_{1,1}$	$y_{2,1}$	$y_{3,1}$	$y_{4,1}$	$y_{5,1}$	$x_{1,1}$	$x_{21,1}$	⋯	$x_{2s,1}$
2	1/3/2010	$y_{1,2}$	$y_{2,2}$	$y_{3,2}$	$y_{4,2}$	$y_{5,2}$	$x_{1,2}$	$x_{21,2}$	⋯	$x_{2s,2}$
3	1/4/2010	$y_{1,3}$	$y_{2,3}$	$y_{3,3}$	$y_{4,3}$	$y_{5,3}$	$x_{1,3}$	$x_{21,3}$	⋯	$x_{2s,3}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
T	1/12/2024	$y_{1,T}$	$y_{2,T}$	$y_{3,T}$	$y_{4,T}$	$y_{5,T}$	$x_{1,T}$	$x_{21,T}$	⋯	$x_{2s,T}$

Table 4. Results of PCA Financial Condition Index (FCI)

	PC_1	PC_2	PC_3	PC_4
Standard Deviation	1.4726	0.9550	0.8947	0.34479
Proportion of Variance	0.5422	0.2280	0.2001	0.02972
Cumulative Proportion	0.5422	**0.7702**	0.9703	1

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

financial stress, economic recovery, supportive monetary policies, and increased market confidence. This situation typically fosters economic growth through increased investment and consumption. Based on Figure 2, the FSI rose sharply in 2020, reflecting the impact of the COVID-19 pandemic on financial markets, followed by a significant decline during the 2021-2023 period, indicating the success of the economic recovery. However, fluctuations at the end of the graph signal the continued risk of financial stress that should still be monitored. The overall FSI graph effectively reflects the pressure on financial stability in Indonesia. Meanwhile, the PCA process based on Equation (18) to form the FCI variable yields the following results.

The results of the Principal Component Analysis (PCA) show that there are four main components (PC_1 to PC_4) with different standard deviations and variance contributions. Based on the cumulative variance, the selection of principal components was made using the criterion that the component must explain at least 70% of the total variance of the data. According to Table 4, PC_1 contributes 54.22% to the total variance. When combined with PC_2, the cumulative variance explained increases to 77.02%, exceeding the 70% threshold. Therefore, PC_1 and PC_2 were selected as the main components because they represent most of

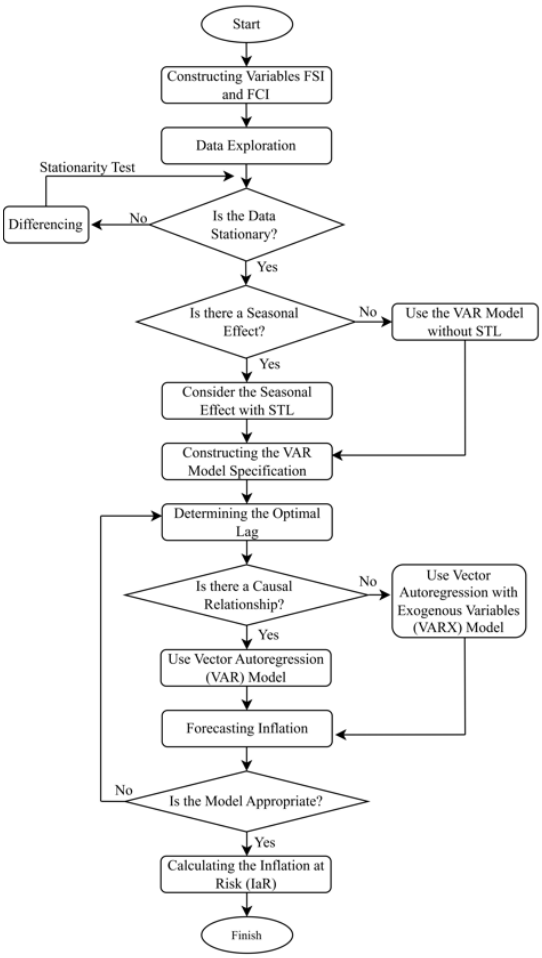


Figure 1. Step Analysis of Research

the information in the data. The combined visualization of the FSI and FCI is shown in Figure 3.

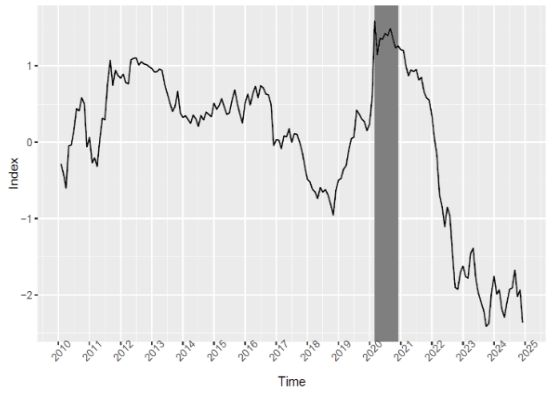


Figure 2. The Financial Stress Index (FSI) shows financial pressures in Indonesia.



Figure 3. Plot of FSI and FCI.

FCI is an index that reflects financial conditions. The visualization of FCI, represented by FCI_1 and FCI_2, illustrates the dynamics of financial conditions from February 2010 to December 2024. FSI and FCI are interrelated, where an increase in stress in financial markets tends to lead to a deterioration in financial conditions, and vice versa. Based on Figure 3, it is observed that during the 2022–2023 period, when the FSI declined in the aftermath of the COVID-19 pandemic, there was an increase in FCI as represented by FCI_2. Meanwhile, the visualization for inflation data is presented in Figure 4.

Based on the visualization of inflation data in Figure 4, it is observed that there is a tendency for seasonal patterns in the health inflation data shown in panel (b). Therefore, the Seasonal Trend Decomposition Based on LOESS (STL) method will be applied to the variable y_2. The next step is to perform a stationarity test on all variables using the Phillips-Perron test, with the following results as stated in Table 5.

With the alternative hypothesis that the data is stationary and a significance level (alpha) of 5%, it was concluded that all inflation variables are stationary, while FSI and FCI are not stationary. Therefore, differencing was performed on x_1, x_{21}, and x_{22}. After differencing, a stationarity test was conducted again, resulting in a p-value of 0.01 for all three variables, indicating that

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

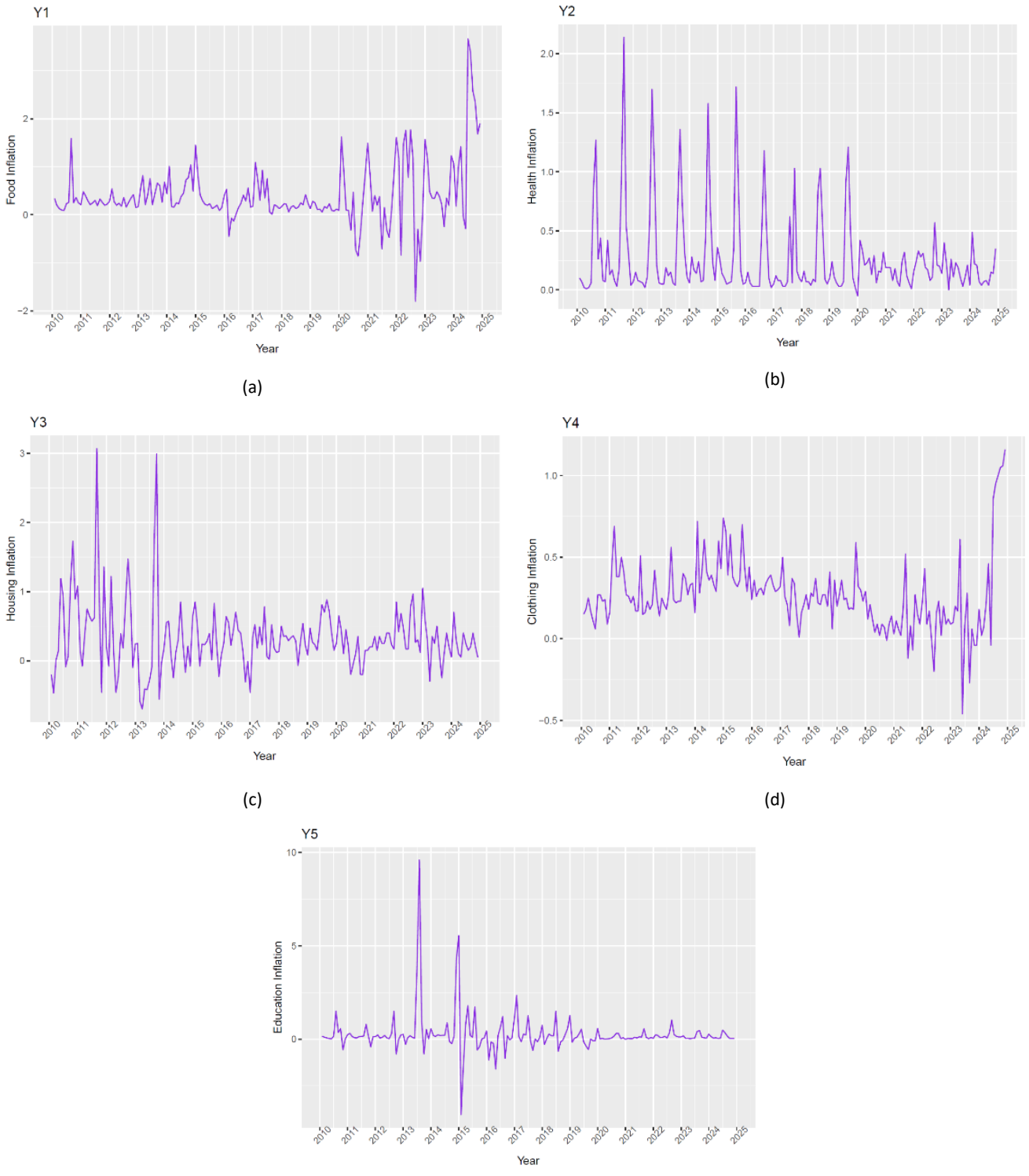


Figure 4. Plot Inflation Data

the data is now stationary. Subsequently, the data was divided into 20% out-of-sample and 80% in-sample portions. The VAR model to be used was then selected based on the lag with the smallest Correlation Information Criterion (CIC) value, as shown in Table 6.

Based on the CIC values in Table 6, it was determined that the optimal lag is in the VAR(1) model, as it has the smallest CIC value. The next step is to perform a Granger causality test to determine the relationships between variables, as presented in Table 7

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

Table 5. Stationarity Test

Variable	p-value
y_1	0.01
y_2	0.01
y_3	0.01
y_4	0.01
y_5	0.01
x_1	0.7789
x_{21}	0.7253
x_{22}	0.654

Table 6. The optimal lag of the VAR model is determined based on the CIC value.

Model	CIC
VAR(1)	**127.3302**
VAR(2)	176.9530
VAR(3)	208.3535
VAR(4)	241.7554

The Granger causality test in Table 7 indicates that the FSI and FCI variables do not have a significant causal relationship with other endogenous variables. Therefore, FSI and FCI can be used as exogenous variables. Additionally, Table 7 shows that several inflation variables do not have a causal relationship with other inflation variables. As a result, a restricted VAR approach is employed to simplify the model or eliminate insignificant variables. Subsequently, the VARX(1) model is estimated as follows:

$$y_1 = 0.421y_{1,t-1} + 0.375d_{2,t-1} + 0.073y_{5,t-1} + 0.078x_{1,t} + 0.028x_{21,t} + 0.125x_{22,t}$$

$$d_2 = 0.368d_{2,t-1} + 0.475y_{4,t-1} + 0.039y_{5,t-1} + 0.226x_{1,t} + 0.036x_{21,t} + 0.13x_{22,t}$$

$$y_3 = 0.483y_{3,t-1} + 0.149y_{5,t-1} + 0.015x_{22,t}$$

$$y_4 = 0.856y_{4,t-1} + 0.078x_{1,t} + 0.004x_{21,t} - 0.042x_{22,t}$$

$$y_5 = 0.561y_{4,t-1} + 0.226y_{5,t-1} + 0.124x_{1,t}$$

To forecast food inflation, there is a positive influence from its past data, as well as health inflation, education inflation, and external factors such as FSI and FCI. Health inflation is positively influenced by its past data, as well as inflation in clothing, education, and external factors such as FSI and FCI. Housing inflation is positively influenced by its past data, education inflation, and FCI. Clothing inflation is positively influenced by its past data, FSI, FCI sub 1, and FCI sub 2 have a negative influence. Education inflation is positively influenced by its past data, clothing inflation, and external factors such as FSI. The model is then used to forecast inflation. The predictions for the in-sample data are presented in a visualized format in Figure 5.

Table 7. Granger Causality Test

	F	df_1	df_2	*p*		*Chisq*	*df*	*p*	
y_1 and y_2	4.95	1	134	0.028	*	4.95	1	0.026	*
$y_1$and y_3	0.51	1	134	0.476		0.51	1	0.475	
$y_1$and y_4	1.83	1	134	0.179		1.83	1	0.177	
$y_1$and y_5	6.07	1	134	0.015	*	6.07	1	0.014	*
$y_1$and x_1	0.01	1	134	0.924		0.01	1	0.924	
$y_1$and x_{21}	0.30	1	134	0.588		0.30	1	0.587	
$y_1$and x_{22}	0.22	1	134	0.637		0.22	1	0.636	
y_1 ALL	3.40	7	134	0.002	**	23.79	7	0.001	**
y_2 and y_1	0.23	1	134	0.629		0.23	1	0.629	
$y_2$and y_3	4.16	1	134	0.043	*	4.16	1	0.041	*
$y_2$and y_4	21.87	1	134	<0.001	***	21.87	1	<0.001	***
$y_2$and y_5	6.69	1	134	0.011	*	6.69	1	0.010	*
$y_2$and x_1	0.82	1	134	0.367		0.82	1	0.365	
$y_2$and x_{21}	1.61	1	134	0.206		1.61	1	0.204	
$y_2$and x_{22}	2.84	1	134	0.094	.	2.84	1	0.092	.
y_2 ALL	6.66	7	134	<0.001	***	46.60	7	<0.001	***
y_3 and y_1	1.17	1	134	0.281		1.17	1	0.279	
$y_3$and y_2	3.58	1	134	0.061	.	3.58	1	0.058	.
$y_3$and y_4	5.18	1	134	0.024	*	5.18	1	0.023	*
$y_3$and y_5	10.36	1	134	0.002	**	10.36	1	0.001	**
$y_3$and x_1	2.17	1	134	0.143	.	2.17	1	0.141	.
$y_3$and x_{21}	1.86	1	134	0.175		1.86	1	0.173	
$y_3$and x_{22}	1.15	1	134	0.285		1.15	1	0.283	
y_3 ALL	6.74	7	134	<0.001	***	47.17	1	<0.001	***
y_4 and y_1	0.76	1	134	0.385		0.76	1	0.383	
$y_4$and y_2	1.32	1	134	0.253		1.32	1	0.251	
$y_4$and y_3	2.26	1	134	0.135		2.26	1	0.132	
$y_4$and y_5	3.48	1	134	0.064	.	3.48	1	0.062	.
$y_4$and x_1	0	1	134	0.953		0	1	0.953	
$y_4$and x_{21}	0.29	1	134	0.588		0.29	1	0.587	
$y_4$and x_{22}	0.79	1	134	0.376		0.79	1	0.374	
y_4 ALL	1.71	7	134	0.111		12	7	0.1	
y_5 and y_1	0.01	1	134	0.931		0.01	1	0.931	
$y_5$and y_2	0	1	134	0.963		0	1	0.963	
$y_5$and y_3	1.55	1	134	0.216		1.55	1	0.213	
$y_5$and y_4	2.84	1	134	0.094	.	2.84	1	0.092	.
$y_5$and x_1	0.65	1	134	0.422		0.65	1	0.420	
$y_5$and x_{21}	1.29	1	134	0.259		1.29	1	0.257	
$y_5$and x_{22}	0.24	1	134	0.622		0.24	1	0.622	
y_5 ALL	1.48	7	134	0.179		10.36	7	0.169	
x_1 and y_1	2.35	1	134	0.128		2.35	1	0.125	
$x_1$and y_2	0.82	1	134	0.367		0.82	1	0.366	
$x_1$and y_3	0.56	1	134	0.457		0.56	1	0.455	
$x_1$and y_4	0.15	1	134	0.701		0.15	1	0.701	
$x_1$and y_5	0.05	1	134	0.828		0.05	1	0.828	
$x_1$and x_{21}	2.80	1	134	0.096	.	2.80	1	0.094	.
$x_1$and x_{22}	3.20	1	134	0.076	.	3.20	1	0.074	.
x_1 ALL	1.02	7	134	0.417		7.17	7	0.411	
x_{21} and y_1	0.20	1	134	0.653		0.20	1	0.652	
$x_{21}$and y_2	0	1	134	0.997		0	1	0.997	
$x_{21}$and y_3	1.29	1	134	0.259		1.29	1	0.257	
$x_{21}$and y_4	0.12	1	134	0.730		0.12	1	0.730	
$x_{21}$and y_5	0.22	1	134	0.643		0.22	1	0.643	
$x_{21}$and x_1	2.97	1	134	0.087	.	2.97	1	0.085	.
$x_{21}$and x_{22}	2.45	1	134	0.120		2.45	1	0.117	
x_{21} ALL	1.48	7	134	0.179		10.38	1	0.168	
x_{22} and y_1	0.15	1	134	0.695		0.15	1	0.718	
$x_{22}$and y_2	2.26	1	134	0.135		2.26	1	0.182	
$x_{22}$and y_3	0.09	1	134	0.769		0.09	1	0.407	
$x_{22}$and y_4	0.88	1	134	0.349		0.88	1	0.339	
$x_{22}$and y_5	0.27	1	134	0.604		0.27	1	0.630	
$x_{22}$and x_1	0	1	134	0.982		0	1	0.960	
$x_{22}$and x_{21}	24.17	1	134	<0.001	***	24.17	1	<0.001	***
x_{22} ALL	4.09	1	134	<0.01	***	28.64	7	<0.001	***

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

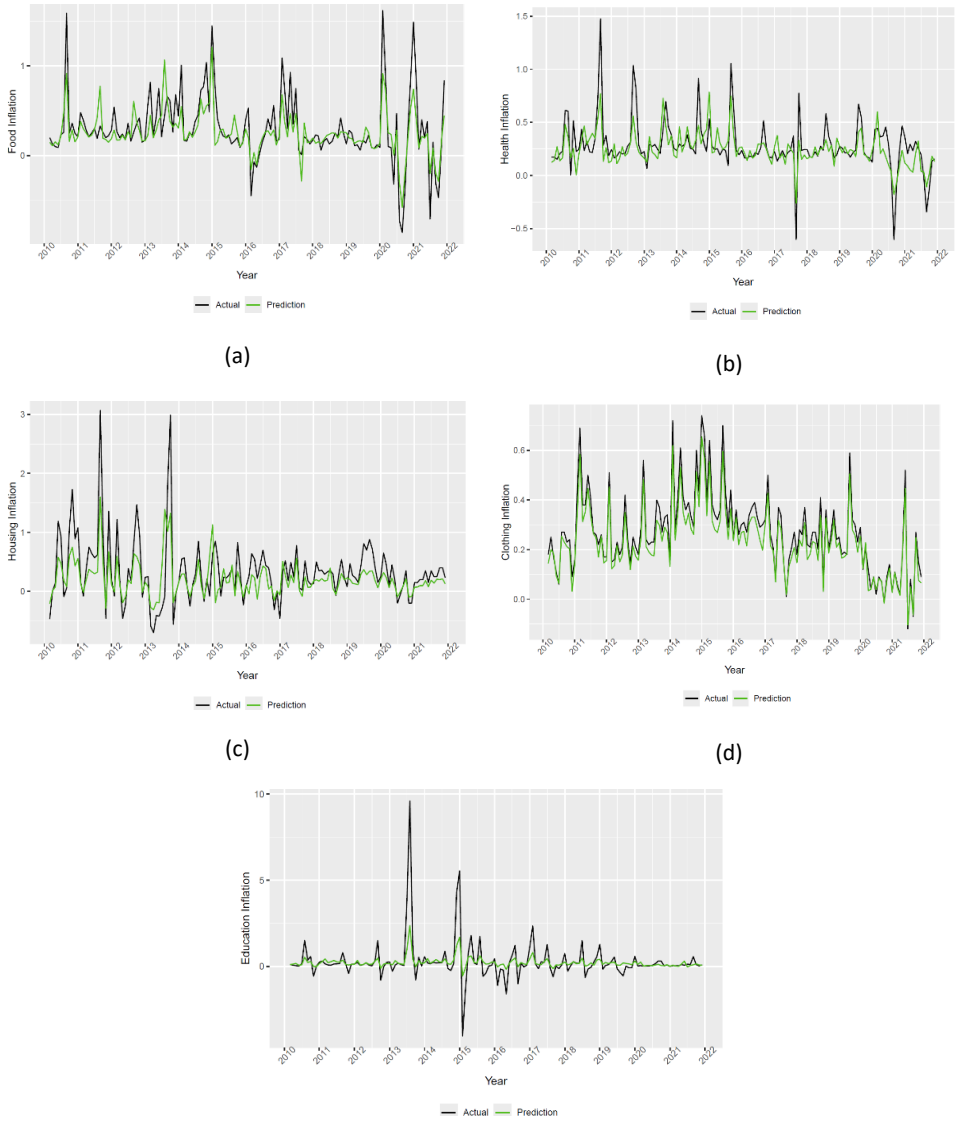


Figure 5. Forecasting In-Sample of Inflation

Figure 5 presents the forecasting results for inflation data during the in-sample period from February 2010 to December 2021, consisting of 143 data points or 80% of the initial dataset. The black line represents the actual inflation data, while the green line represents the model's predictions. Overall, the model accurately follows the actual data patterns. However, it struggles to capture extreme fluctuations. The in-sample forecasting accuracy results are shown in Table 8.

Table 8. Forecasting accuracy based on the SMAPE criterion.

Variable	SMAPE
y_1	21.5%
y_2	21.3%
y_3	38%
y_4	8.8%
y_5	49.7%

Based on the SMAPE criteria in Table 1, it is found that the forecast for clothing inflation is highly accurate, while forecasts for food inflation, health inflation, housing inflation, and education inflation are moderately accurate. Next, the forecast for the out-of-sample data from January 2022 to December 2024, consisting of 35 data points or 20% of the last data, is presented in Figure 6.

Table 9. Forecasting accuracy based on the SMAPE criterion.

Variable	SMAPE
y_1	40%
y_2	44.5%
y_3	27.7%
y_4	19%
y_5	49.6%

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

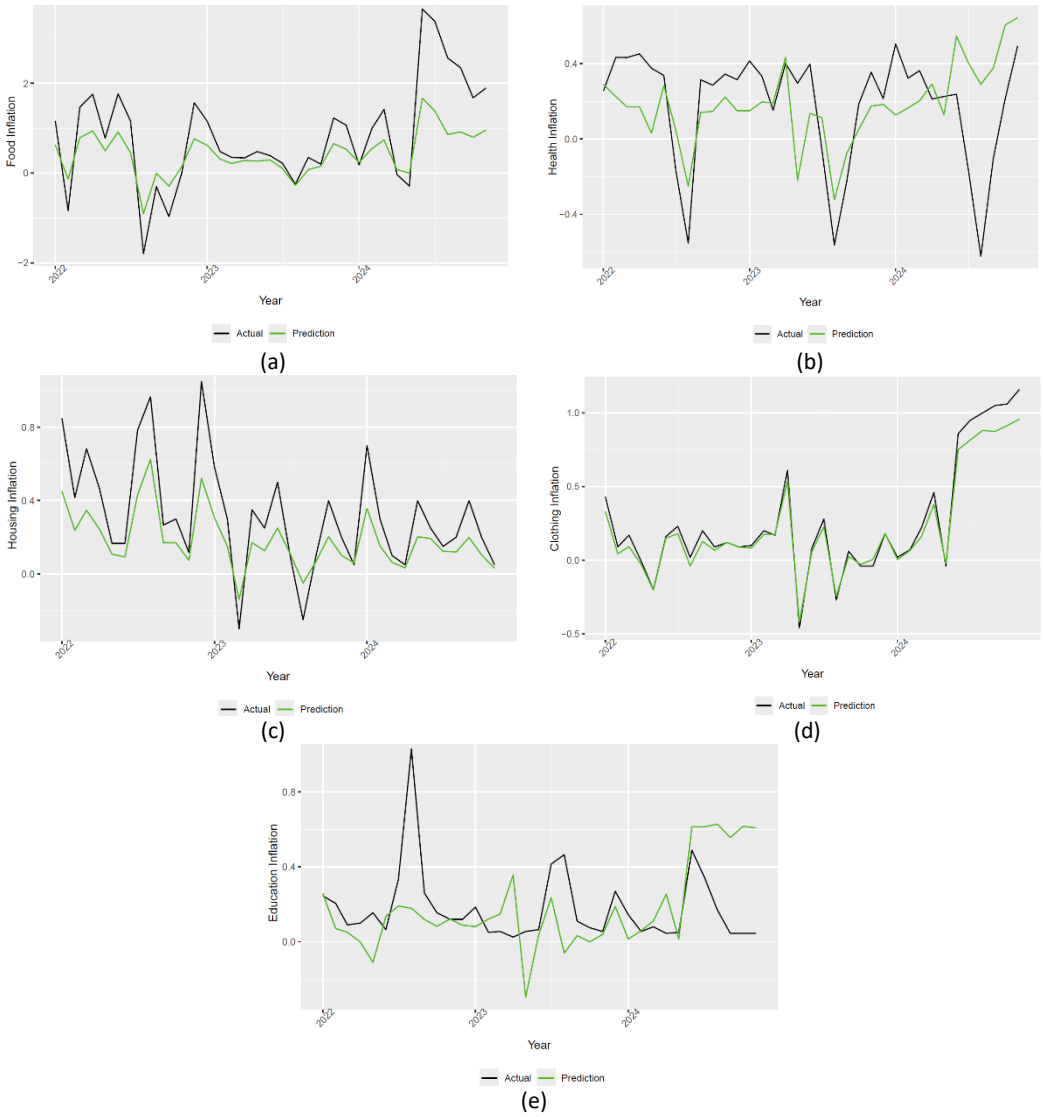


Figure 6. Forecasting the Out-of-Sample of Inflation

In general, the model is able to follow the actual data patterns well; however, it struggles to capture extreme fluctuations. The out-of-sample forecasting accuracy is presented in Table 9.

Based on the SMAPE forecasting criteria in Table 1, it is found that the forecast for clothing inflation is accurate, food inflation, health inflation, housing inflation, and education inflation are moderately accurate.

5. Conclusions

Based on the analysis results, the main finding of this study indicates that the VARX model is effective for predicting IaR in Indonesia. Generally, the forecast results follow the pattern of the actual data. The VARX model performs exceptionally well when the inflation data being analyzed exhibits relatively stable fluctuations, as seen in the clothing inflation component. In such cases, the model successfully captures the dynamic relationships between variables, as reflected in forecast visualizations that closely match the actual data and accurate SMAPE values. However, the model's effectiveness decreases when applied to data with extreme fluctuations, such as the more volatile education inflation component.

6. Acknowledgments

I would like to express my deepest appreciation to *Lembaga Pengelola Dana Pendidikan* (LPDP) for their scholarship support. I would also like to express my gratitude to the reviewers, particularly Wisnowan Hendy Saputra, for his support throughout this research.

7. References

Akbar, M. S., Setiawan, S., Suhartono, S., Ruchjana, B. N., Prastyo, D. D., Muhaimin, A., & Setyowati, E. (2020). A Generalized Space-Time Autoregressive Moving Average (GSTARMA) Model for Forecasting Air Pollutant in Surabaya. J. Phys.: Conf. Ser. 1490 (012022).

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 18.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics, 6(1), 3-73.

Farih, I., & Prastyo, D. D. (2022). Forecasting Electricity Consumption Based on Economics and Social Indicators Using VAR Model with Exogenous Variable: Evidence From East Java Province. in 022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Yogyakarta.

Ferrara, L., Mogliani, M., & Sahuc, J.-G. (2022). High-frequency monitoring of growth at risk. International Journal of Forecasting, 38, 582-595.

Gujarati, D. N. (2003). Basic Econometrics. New York: Gary Burke.

Hikmawati, F., Suhartono, S., & Prastyo, D. D. (2021). A Hybrid GSTARX-Jordan RNN Model for Forecasting Space-Time Data with Calendar Variation Effect. J. Phys.: Conf. Ser. 1752 (012013).

Juhro, S. M., & Iyke, B. N. (2019). Monetary Policy and Financial Condition in Indonesia. Bulletin of Monetary Economics and Banking, 21, 283-302.

Novianto, M. A., Suhartono, S., Prastyo, D. D., Suharsono, A., & Setiawan, S. (2018). GSTARIX Model for Forecasting Spatio-Temporal Data with Trend, Seasonal and Intervention. J. Phys.: Conf. Ser. 1097 (012076).

Ocampo, S., & Rodríguez, N. (2012). An Introductory Review of a Structural VAR-X Estimation and Applications. Revista Colombiana de Estadística, 35, 479-508.

Park, C.-Y., & Mercado, R. V. (2014). Determinants of financial stress in emerging market economies. Journal of Banking & Finance, 45, 199-224.

Prastyo, D. D., Nabila, F. S., Suhartono, S., Lee, M. H., & Suhermi, N. F.-F. (2018). VAR and GSTAR-Based Feature Selection in Support Vector Regression for Multivariate Spatio-Temporal Forecasting. In Soft Computing in Data Science (SCDS), Communications in Computer and Information Science, vol 937. Springer.

Safuan, S., Sugandi, E. A., Aziz, O. Q., & Triandhari, R. (2022). Indonesia's Financial Stress Events and Macroeconomic Dynamics. Bulletin of Monetary Economics and Banking, 25, 323-370.

Sesay, A., Suhartono, S., & Prastyo, D. D. (2020). Forecasting Exchange Rate Across Countries with Gold Price as Exogenous Variable Using Transfer Function and VARI-X Model. MATEMATIKA, 36(3), 181-196.

Suhartono, S., Gazali, M. M., & Prastyo, D. D. (2018). VARX and GSTARX Models for Forecasting Currency Inflow and Outflow with Multiple Calendar Variations Effect. MATEMATIKA, 34(3), 57–72.

Suhartono, S., Prastyo, D. D., Kuswanto, H., & Lee, M. H. (2018). Comparison between VAR, GSTAR, FFNN-VAR and FFNN-GSTAR Models for Forecasting Oil Production. MATEMATIKA, 34(1), 103–111.

Tsay, R. S. (2002). Analysis of Financial Time Series. Canada: John Wiley & Sons, Inc.

Zivot, E., & Wang, J. (2003). Modeling Financial Time Series with S-Plus. New York: Springer Science+Bu siness Media.

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.10
Malaysian Journal of Science 45 (Special Issue 2): 95-104(November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026) https://mjs.um.edu.my

Mixed-Data Sampling Quantile Regression (MIDAS-QR) for Nowcasting Indonesia's Economic Growth-at-Risk of Non-Oil and Gas Processing Industry Sector

Muhammad Rizal[1a], Dedy Dwi Prastyo[2a*], Muhammad Sjahid Akbar[3a]

Abstract: The non-oil and gas processing industry plays a vital role in economic stability, making timely Gross Domestic Product (GDP) nowcasting essential for managing economic fluctuations and shaping effective macroeconomic policies. This study applies the Mixed Data Sampling Quantile Regression (MIDAS-QR) model to nowcast GDP growth in the non-oil and gas processing industry, a key driver of economic stability. By integrating high-frequency financial indicators, such as the Financial Stress Index (FSI) and Financial Condition Index (FCI), derived through principal component analysis, the model effectively captures non-linear patterns using an Exponential Almon Lag Polynomial function. Covering data from January 2011 to December 2023, the MIDAS-QR model demonstrates strong predictive accuracy, particularly in the electronic goods and basic metals sectors. The best performing model for nowcasting is found in the median quantile ($\tau = 0.5$), which produces the smallest error in both the electronic goods and base metals sectors. This is because the median value tends to be closer to the actual value than the other quantiles. The symmetric Mean Absolute Percentage Error (sMAPE) value in the training data for the electronic goods sector is 7.32% and testing is 9.16%, while for the base metal sector, the sMAPE value in the training data is 5.21% and testing is 4.09%. These findings enhance Growth-at-Risk (GaR) assessments, providing policymakers and industry leaders with critical insights for navigating economic uncertainties and shaping effective macroeconomic policies.

Keywords: GaR, MIDAS-QR, FCI, GDP Growth, Nowcasting

1. Introduction

In recent decades, economic growth has often been accompanied by significant uncertainty, particularly in industries that are sensitive to global and domestic market dynamics. This uncertainty is illustrated by the concept of Growth-at-Risk (GaR), which considers the distribution of future economic growth in the face of various financial and economic risks. This approach enables the measurement of systemic risks to overall economic growth and provides an important surveillance tool for economic policymakers (Xu et al., 2023).

Economic growth is one of the crucial indicators that reflect a country's economic condition over a specific period. One way to measure this economic condition is through Gross Domestic Product (GDP), which represents the added value of goods and services produced in the economy. GDP can be measured based on current prices, which describe the value added of goods and services using prices from the current year, and based on constant prices, which use prices from a specific base year as a reference to measure real growth from year to year (Sujianto, 2024). GDP at current prices can be used to observe shifts in economic structure, while GDP at constant prices is used to track economic growth from year to year.

Authors information:
[a]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, INDONESIA. E-mail: muhdrhizal99@gmail.com[1], dd.prastyo@its.ac.id[2], m_sjahid_a@its.ac.id[3]

*Corresponding Author: dd.prastyo@its.ac.id

The most considerable contribution to Indonesia's GDP comes from the non-oil and gas processing industry sector, whose primary focus is the processing of raw materials into finished or semi-finished products that consumers or other industries can use (Fastasima et al., 2025). Data from the Ministry of Industry shows that this sector contributed the most to the national GDP, achieving 16.30% in the second quarter of 2023. The contribution of the non-oil and gas processing industry increased by 4.56% in the second quarter of 2023, surpassing the same period in 2022 by approximately 0.23 percentage points. The largest growth in the performance of the national manufacturing sector in the second quarter of 2023, specifically in the metal goods, computer, electronic goods, optical, and electrical equipment industries, increased by 17.32%, followed by the basic metal industry at 11.49%. The non-oil and gas processing industry plays a very significant role in driving the nation's economic growth because its contribution is able to provide a positive chain or multiplier effect for the economy, such as an increase in the added value of domestic raw materials, absorption of local labor, and high foreign exchange earnings from exports (Fastasima et al., 2025).

GDP data, which is a time series issued by the Central Statistics Agency (BPS), has a reasonably long release lag, occurring every three months. Hence, data publication occurs quarterly. This is due to the large scope and number of indicators that need to be calculated (Supriyatna et al., 2024). In addition, GDP data is delayed in its release by five weeks after the end of each quarter. This condition occurs at the regional and national

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

levels. Of course, this long delay limits the availability of up-to-date data. In the current economic climate, where a crisis can occur at any time, the lack of real-time data can lead to delays in decision-making. This can have profound implications, especially when governments and investors need accurate and fast information to deal with economic uncertainty (Ma'Arif, 2019). Therefore, the early and high-frequency release of economic indicator data is essential for a more real-time analysis of economic developments, thereby complementing GDP data that is generally released quarterly (Eurostat, 2016). For example, trends in inflation or unemployment data can serve as an early indicator of changes in economic growth, which can be used for forecasting or nowcasting future GDP.

A common solution to the problem of data with different frequencies is to transform the data so that the dependent and independent variables have the same frequency. However, this transformation can lead to the loss of important information, making it difficult to recognize the relationship between variables. In time series regression models, all variables used should have the same frequency; therefore, data with high frequencies need to be transformed into lower frequencies. To overcome this limitation, the Mixed Data Sampling (MIDAS) regression model, introduced by Ghysels et al. (2004), is employed. MIDAS can reduce high-frequency data to a lower frequency according to the target variable without losing information from the data (Xu M et al., 2025). The reduction process is performed by assigning weights that take into account the contribution of each period, resulting in more accurate results compared to using ordinary averages. Another advantage of the MIDAS regression model is that it can minimize the number of parameters that need to be estimated, making it a simpler regression model. This model is beneficial in economic contexts where various indicators are available at varying frequencies to estimate quarterly GDP, which is typically released at a lower frequency (Kuzin et al., 2011).

This study employed a different MIDAS-QR approach compared to previous studies. Research (Utari & Rosadi, 2018) used conventional MIDAS to forecast Indonesia's GDP with JCI as a predictor, but did not consider risk distribution (quantile). Meanwhile, Research (Saputra et al., 2024) proposed a hybrid QRNN-MIDAS model for GaR, which, although sophisticated, may reduce the interpretability of the results due to the complexity of the neural network. The advantage of our study lies in combining the flexibility of MIDAS with direct quantile analysis, allowing for more efficient and interpretable GaR estimation, especially for understanding the impact of high-frequency predictors on the tail of the GDP distribution.

This study aims to obtain nowcasting results for the GDP of the non-oil and gas processing industry subsector using the MIDAS-QR model, which facilitates model specification in the presence of mixed-frequency data. By utilizing quarterly GDP data, as well as various economic indicators available at a monthly frequency, this study aims to provide more accurate GDP predictions and a deeper understanding of GDP dynamics. The results of this study are expected to contribute both academically and practically in efforts to control the GDP of the non-oil and gas processing industry sub-sector in Indonesia.

2. Literature Review

2.1 Mixed Data Sampling (MIDAS) Regression

In time series regression analysis, the data used is generally available at the same frequency. However, in this study, the regression model is formed with predictors with different frequencies. To deal with time series data that has different frequencies, a Mixed Data Sampling model is used, hereafter referred to as MIDAS regression. Unlike the autoregressive (AR) model, MIDAS regression employs predictors with varying sampling frequencies. The AR model indirectly assumes that the data is taken at the same frequency as the previous period (Ghysels et al., 2004).

Suppose $y_t, t \in T$ is the univariate response variable of a low-frequency (quarterly) observation, where t represents the quarterly time period. Meanwhile $x_m, m \in M$ are the predictor variables of high-frequency (monthly) observations, where m represents the monthly time index. In general, the mixed frequency regression model with an autoregressive lag L can be expressed in the following equation (Ghysels et al., 2016):

$$y_t = \alpha_1 y_{t-L_1} + \cdots + \alpha_L y_{t-L_J} + \beta_0 x_{mt} + \beta_1 x_{mt-1} + \cdots + \beta_{m-1} x_{m(t-1)+1} + \varepsilon_t, \quad (1)$$

which is equivalent to

$$y_t = \sum_{j=1}^{J} \alpha_j y_{t-L_j} + \sum_{k=0}^{m-1} \beta_k x_{mt-k} + \varepsilon_t, \quad (2)$$

It is important to remember that the variable x must be represented according to time t at low periods (Saputra et al., 2024). Therefore, MIDAS regression involves linear projection using the stochastic process $\{x_k^{(t)}, k \in M\}$, which is obtained from the higher frequency (monthly). At each low frequency with period t, the process $x_k^{(t)}$ will be observed at a high frequency with a period $k = (t-1)m + 1, \ldots, tm$, so Equation (1) can be rewritten as follows.

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_L y_{t-L} + \beta_0^{(t)} x_m^{(t)} + \beta_1^{(t)} x_{m-1}^{(t)} + \cdots + \beta_{m-1}^{(t)} x_1^{(t)} + \varepsilon_t, \quad (3)$$

which is equivalent to

$$y_t = \sum_{j=1}^{J} \alpha_j y_{t-j} + \sum_{k=0}^{m-1} \beta_k^{(t)} x_{m-k}^{(t)} + \varepsilon_t \quad (4)$$

2.2 Determination of Parameters of the Weighting Function

One of the main aspects of the MIDAS regression model is the parameterization of the weight function $c(k; \boldsymbol{\delta})$. In addition, it is also important to ensure that the sum of the weight functions is one (Utari & Rosadi, 2018). Another goal of the weight function in MIDAS regression is to achieve flexibility while still maintaining the simplicity of the model.

The method of determining the parameters of the weight function used is called the Almon Exponential, as it is closely related to the Almon polynomial function. The Almon Exponential weight function is expressed as follows:

$$b(L_m, \boldsymbol{\delta}) = \sum_{k=0}^{K} c(k, \boldsymbol{\delta})\, L_m^k, c(k; \boldsymbol{\delta}) = \frac{\exp(\delta_1 k + \cdots + \delta_Q k^Q)}{\sum_{k=0}^{K} \exp(\delta_1 k + \ldots + \delta_Q k^Q)} \quad (5)$$

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

The specification is parsimonious, with only two parameters to estimate, $\boldsymbol{\delta} = (\delta_1, \delta_2)'$. This aims to simplify the model while still ensuring flexibility in specifying the polynomial form. Therefore, Equation (5) can be rewritten as follows (Ghysels et al., 2005):

$$c(k; \boldsymbol{\delta}) = \frac{\exp(\delta_1 k + \delta_2 k^2)}{\sum_{k=0}^{K} \exp(\delta_1 k + \delta_2 k^2)} \quad (6)$$

2.3 Determination of Optimum Lag

Determining the optimal lag is crucial in time series analysis. This is because the effects of an economic policy generally do not have an immediate impact on economic activity at the same time, but require time or inertia for the effects to be seen. The criterion used to determine the optimum lag length is the Akaike Information Criterion (AIC). The formula for this selection criterion is (Tsay, 2010):

$$AIC = n \ln(\hat{\sigma}_\varepsilon^2) + 2\, m \quad (7)$$

with $\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n}$, SSE is the Sum of Squared Error can be estimated from the sum of the squares of all error values, n is the number of observations, mIs the number of parameters estimated. The best model is the model with the smallest AIC value.

2.4 Quantile Regression (QR)

Quantile regression is a regression analysis approach introduced by Koenker and Basset in 1978 (Koenker & Bassett, 1978). A quantile regression model is proposed based on the reparameterization of the original distribution by incorporating quantile parameters (Saulo et al., 2023). Estimators in quantile regression are obtained by estimating the value of the quantile function of a distribution Y which is a function or a linear combination of X. Quantile regression can be used to explore the problem of heterogeneous data distribution (Koenker, 2017).

Suppose Y is a random variable with a cumulative distribution F_Y and τ is a natural number (probability) at a certain quantile index $0 < \tau < 1$ then the τth of F_Y can be expressed as (White & Kim, 2015):

$$Q_Y(\tau) = inf\ \{y: F(y) \geq \tau\}, \quad (8)$$

where $F_Y\ (y) = P(Y \leq y)$ is the cumulative distribution of Y. If $F(y)$ is a monotonically increasing function, then the τth quantile can be represented as:

$$Q_Y(\tau) = F^{-1}(\tau), \quad (9)$$

which can also be represented in the following way:

$$Q_Y(\tau) = \int_{-\infty}^{\infty} dF(y) = E[1\{Y \leq Q_Y(\tau)\}|\mathcal{F}_{t-1}] = \tau. \quad (10)$$

If the variable Y is affected by the covariate x, then the τth conditional quantile, which is then called quantile regression, is expressed as:

$$\hat{Q}(\tau|\mathbf{X}) = \mathbf{X}\hat{\boldsymbol{\beta}}(\tau), \quad (11)$$

So that the model for y at the τth quantile is:

$$\mathbf{y}(\tau) = \mathbf{X}\hat{\boldsymbol{\beta}}(\tau) + \varepsilon, \quad (12)$$

where $\boldsymbol{\beta}(\tau)$ are the regression model's parameters at the τth quantile and ε is the residual.

Based on the Equation (12), the vector $\boldsymbol{\beta}(\tau)$ can be estimated by $\hat{\boldsymbol{\beta}}(\tau)$ which is then called the model parameter estimate at the τth quantile. The Least Absolute Deviation (LAD) method is used to estimate the parameters in quantile regression. This method minimizes the absolute sum of the weighted residuals. The weight of the residual will be given a weight of τ if the residual ε_t is more than or equal to zero, while if the residual is less than zero, a weight of $1 - \tau$ (Koenker & Bassett, 1978). The parameter estimation can be obtained using the following equation:

$$\hat{\boldsymbol{\beta}}(\tau) = \min_{\beta} \left[\sum_{y \in \{t: y \geq \mathbf{X}\boldsymbol{\beta}\}} \tau |\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}| + \sum_{y \in \{t: y < \boldsymbol{X\beta}\}} (1-\tau)|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}|\right], \quad (13)$$

So, the estimated value of $\hat{\boldsymbol{\beta}}(\tau)$ is obtained for each $\tau - \text{th}$. If there is an autoregressive process in the response, then Equation (13) can be written as (Koenker, 2017):

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_\tau (y_t - \mathbf{Z}_{t-h}\hat{\boldsymbol{\beta}}(\tau)), \quad (14)$$

with $\rho_\tau = \begin{cases} \tau\varepsilon_t, \varepsilon_t \geq 0 \\ (\tau - 1)\varepsilon_t, \varepsilon_t < 0' \end{cases}$ where $\mathbf{Z}_{t-h} = \left(y_{t-h}, \mathbf{X}_m^{(t-h)}\right)_{T \times (L+p)}$ with h being the forecast horizon and ε_t being the residual of the parameter estimation result. At the end, the Equation (14) becomes :

$$\hat{\boldsymbol{\beta}}(\tau) = \min_{\boldsymbol{\beta}} \left[\frac{1}{t}\sum_{t=1}^{T} \rho_\tau (y_t - \hat{y}_t(\tau))\right], \quad (15)$$

where y_t denotes the response observation with the time index $t = 1,2, \ldots, T$ (Rantini et al., 2018).

2.5 Mixed Data Sampling Quantile Regression (MIDAS-QR)

The concept of MIDAS quantile regression (MIDAS-QR) is proposed as a solution to account for high-frequency financial data in real-time financial risk measurement. There is a problem in combining high-frequency financial indicators with lower-frequency GDP data. Higher-frequency data is thought to be able to predict lower-frequency data. However, the problem lies in aggregating the high-frequency data properly. The lag structure used previously does not allow the model to respond quickly to shocks, rendering it less useful for policy purposes. Therefore, to overcome this, Ferrara et al. (2020) proposed a combined method called MIDAS Quantile Regression (MIDAS-QR), which is then defined as:

$$y_t(\tau) = \sum_{j=1}^{L} a_j(\tau) y_{t-j} + \sum_{i=1}^{p} \beta_i(\tau) \sum_{k=0}^{m-1} w_i(\boldsymbol{\delta}(\tau), k) x_{i,m-k}^{(t)} + \varepsilon_t, \quad (16)$$

where $(\boldsymbol{\delta}(\tau), k)$ is a weighting function whose sum is close to one. It depends on the polynomial parameter vector $\boldsymbol{\delta}(\tau)$ and a sequence of lags k. Note that $x_m^{(t)}$ and y_t refers to monthly and quarterly data, respectively. Using the Almon lag polynomial, we can express the Equation (16) as follows:

$$y_t(\tau) = \sum_{j=1}^{L} a_j(\tau) y_{t-j} + \sum_{i=1}^{p} \beta_i(\tau) \sum_{k=0}^{m-1} w_i(\boldsymbol{\delta}, k) x_{i,m-k}^{(t)} + \varepsilon_t, \quad (17)$$

where $w_k(\boldsymbol{\delta}, k)$ is the Almon lag weight. One type of Almon lag is the exponential Almon lag polynomial. The advantage of the Almon lag polynomial is that the Equation (17) is linear and follows the principle of parsimony as it only depends on m parameters, which are weights for high-frequency variables.

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

2.6 Symmetric Mean Absolute Percentage Error (SMAPE)

The best forecasting is achieved by minimizing the error rate. The level of forecasting accuracy can be calculated using sMAPE (symmetric Mean Absolute Percentage Error). This forecasting evaluation aims to see the performance of the testing data in forecasting the time series model. sMAPE calculates a measure of error presentation with the following formula (Makridakis & Hibon, 2000):

$$\text{sMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)} \times 100\% \quad (18)$$

Table 1. Forecasting Accuracy Level Rating Scale

sMAPE	Forecasting Accuracy Level
$\leq 10\%$	Highly Accurate forecasting
$10\% < \text{sMAPE} \leq 20\%$	Good forecasting
$20\% < \text{sMAPE} \leq 50\%$	Reasonable forecasting
$> 50\%$	Inaccurate forecasting

Table 1 shows the interpretation of the results obtained with sMAPE. This table classifies the forecasting accuracy based on the sMAPE value. The lower the sMAPE value, the more accurate the forecasting, while higher values indicate less accurate forecasting (Hu et al., 2022).

3. Methodology

3.1 Research Data Source

The data used in this study were obtained from various sources as listed in Table 2. The period of data collection spans from 2011 to 2023. The data used in this study have different frequencies. The response variable is available at low frequency, which is quarterly, while the predictors are available at high frequency, which is monthly. This frequency difference reflects the varying intervals of measurement between the response variable and the predictors, which must be considered when selecting and applying the appropriate analytical methods to ensure accurate results.

3.2 Model Specifications

The model specification of the MIDAS-QR method is as follows. The MIDAS-QR model for $y_{1,t}$ (GDP Growth of Electronic Goods Industry Sector) is

$$y_{1,t}(\tau) = \sum_{j=1}^{L}\alpha_{i,j}(\tau)y_{1,t-j} + \beta_1(\tau)\sum_{k=0}^{m-1}w(\delta,k)FSI_{t-k} + \sum_{i=1}^{s}\gamma_{1,i}(\tau)\sum_{k=0}^{m-1}w(\delta^{'},k)G_{i,t-k} + \varepsilon_{1,t}(\tau).$$

The MIDAS-QR model for $y_{2,t}$ (GDP Growth of Basic Metals Sector) is

$$y_{2,t}(\tau) = \sum_{j=1}^{L}\alpha_{i,j}(\tau)y_{2,t-j} + \beta_2(\tau)\sum_{k=0}^{m-1}w(\delta,k)FSI_{t-k} + \sum_{i=1}^{s}\gamma_{2,i}(\tau)\sum_{k=0}^{m-1}w(\delta^{'},k)G_{i,t-k} + \varepsilon_{2,t}(\tau).$$

Table 2. Research Data Source

Data	Frequency	Source
GDP of the non-oil and gas processing industry subsector	Quarterly	Central Statistics Agency
FSI Global	Monthly	Office of Financial Research
EXC	Monthly	Yahoo Finance
JCI	Monthly	Yahoo Finance
CCI	Monthly	Organisation for Economic Co-operation and Development
INFL	Monthly	Central Statistics Agency
RATE	Monthly	Central Statistics Agency

3.3 Research Variables and Steps of Research

The variables used in this study comprise data consisting of 8 variables, as explained in Table 3. Figure 1 illustrates the research steps, specifically the process of obtaining the best model for nowcasting using the MIDAS-QR method, based on the data used.

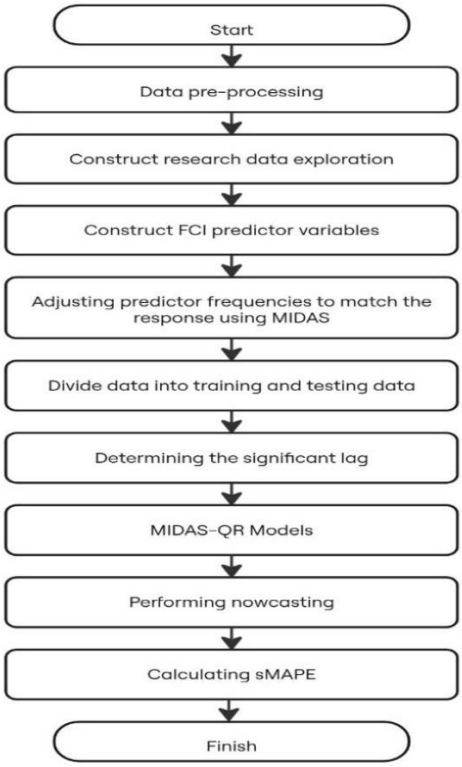


Figure 1. Step Analysis of Research

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

Table 3. Research Variables

	Variables		Description
Response	$(y_{1,t})$		GDP Growth of Electronic Goods Industry Sector, etc.
	$(y_{2,t})$		GDP Growth of the Basic Metals Sector
Predictor	FSI_t $(x_{1,t})$		*Financial Stress Index*
	FCI_t (G_i) *i*=1,...,*s* *s* is the number of PCs used	EXC $(x_{21,t})$	*Exchange Rate* (USD IDR)
		JCI $(x_{22,t})$	*Jakarta Composite Index*
		CCI $(x_{23,t})$	*Consumer Confidence Index*
		INFL $(x_{24,t})$	*Inflation Rate*
		RATE $(x_{25,t})$	*Interest Rate*

4. Results and Discussion

GDP growth data are calculated based on the GDP growth rate at constant 2010 base prices. The data was collected from the first quarter of 2011 (2011Q1) to the fourth quarter of 2023 (2023Q4). This growth data is used as the response variable in MIDAS-QR modeling and is then referred to as the response.

Descriptive statistics of these response variables were calculated, with the results summarized in Table 4. These values indicate a minimum range for the variables GDP Growth of the Electronic Goods Industry Sector and GDP Growth of the Basic Metals Sector, ranging from -6.87 to 8.44 and from -13.2 to 11.55, respectively. The average growth rates were recorded at 1.15 and 2.18, while the variances were 12.12 and 22.25. These relatively small variances indicate that the data has low volatility.

Table 4. Descriptive Statistics of Response Variables

Variables	Mean	Min	Max	Varians
$y_{1,t}$	1.149038	-6.87	8.44	12.12488
$y_{2,t}$	2.181154	-13.2	11.55	22.24781

To provide further visualization, the response variable is visualized in the form of a time series plot, as shown in Figure 2.

The first exogenous predictor variable used in MIDAS-QR modeling in this study is the Financial Stress Index (FSI). Data for the FSI is obtained from the Office of Financial Research, which presents an aggregated index to measure the level of financial stress in an economy. The index encompasses four major financial markets: the banking market, the foreign exchange market, the equity market, and the debt market. The FSI data is collected at a monthly frequency to reflect current market dynamics. The movement pattern of the FSI value during the study period is visualized in Figure 3.

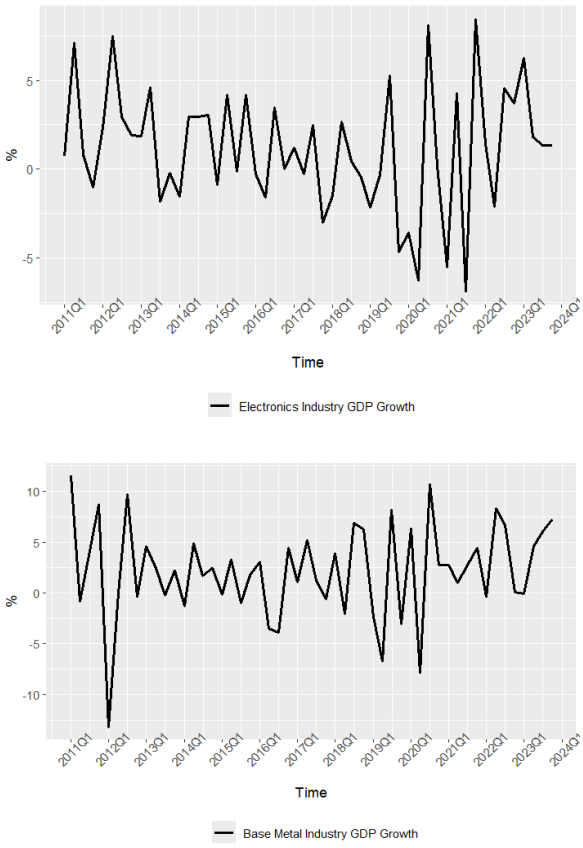


Figure 2. Indonesia GDP Growth Data

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

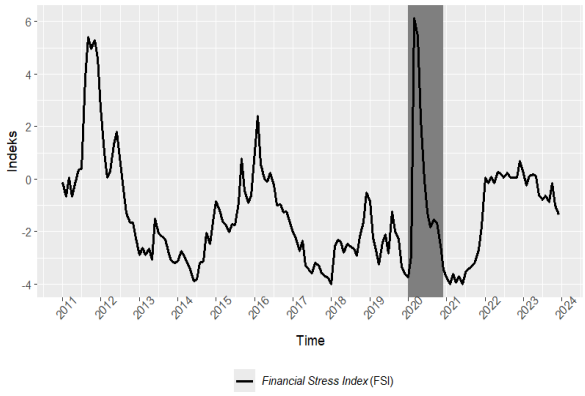


Figure 3. The FSI Time Series Plot

Based on Figure 3, it can be seen that the FSI index experienced significant fluctuations in specific periods, especially when economic conditions faced severe pressure. One example occurred from 2020 to 2021, which is characterized by the impact of the COVID-19 pandemic. This period was characterized by a sharp spike in FSI values, as shown in the black-marked area. The COVID-19 pandemic, which began in early 2020, placed tremendous pressure on the global economy, including that of Indonesia, by significantly impacting major financial markets. This pattern illustrates the high economic instability during the pandemic.

The abnormal pattern in FSI values during the COVID-19 pandemic period reflects the disruptions that have occurred in various economic sectors. Such instability is also evident in the graph, which shows a spike in the index value from 2020 to early 2021. The movement of the FSI during this period provides important insights to understand the impact of the pandemic on financial stress and overall economic conditions.

The second exogenous predictor variable, the Financial Conditions Index (FCI), is one of the primary indicators that reflect a country's financial condition. In this study, the FCI is calculated using the Principal Component Analysis (PCA) approach, which allows the extraction of information from various high-frequency financial variables, as depicted in Table 3, to form a composite index that represents overall financial conditions. The following is the process of forming FCI variables through the PCA process.

The PCA results indicate that there are five primary components (PC1, PC2, PC3, PC4, PC5) that account for the variability in the financial data used to construct the FCI. Based on the cumulative value of variance, the main component selection criteria must be able to explain at least 80% of the information in the data. In Table 5 above, three main components (PC1, PC2, and PC3) are identified in the formation of the composite index (FCI), as they account for 81.28% of the total relevant information.

The following is a visualization plot of the relationship between the FSI and FCI, which illustrates the relationship between the two indices more clearly and comprehensively. The relationship between FSI and FCI can be seen from the Figure 4. When financial market pressure, as represented by FSI, decreases, financial conditions, as measured by FCI, tend to improve; conversely, when financial market pressure increases,

Table 5. Results of PCA for FCI

	PC_1	PC_2	PC_3	PC_4	PC_5
Standard Deviation	1.4426	1.0416	0.9477	0.7188	0.64760
Proportion of Variance	0.4162	0.2170	0.1796	0.1033	0.08388
Cumulative Proportion	0.4162	0.6332	0.8128	0.9161	1.00000

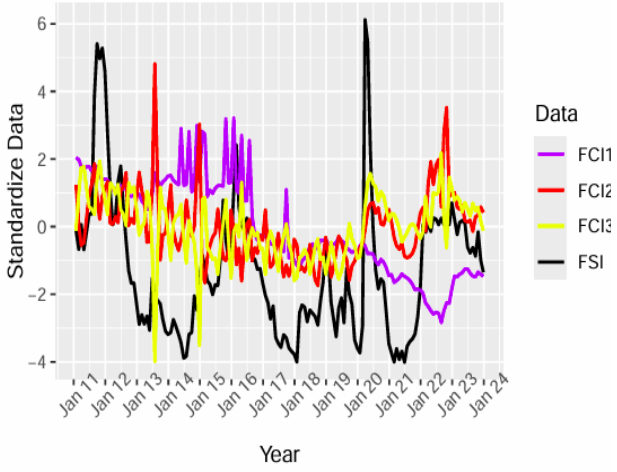


Figure 4. Illustration of the relationship between the Financial Stress Index and the Financial Condition Index

financial conditions also worsen. One of the significant events reflected in the graph is the impact of the COVID-19 pandemic, which occurred from late 2019 to 2020, as seen in the spike in FSI. The pandemic caused an increase in economic pressure, leading to a decline in financial conditions, which was reflected in the decrease in the FCI value. However, after the pandemic ended in 2021, there was a significant recovery as reflected in the increase in FCI1, FCI2, and FCI3, indicating an economic recovery followed by an improvement in financial conditions.

The choice of lag length used in this study refers to the lag interval ranging from lag 1 to lag 4 for each GDP sector, which corresponds to the data structure that divides 1 year into 4 quarters. Thus, the model accommodates the pattern of data changes between quarters in each year. The optimum lag test is shown in Table 6.

Based on Table 6, regarding the optimum lag test, the lag selection for each sector is determined by the smallest Akaike Information Criterion (AIC) value, indicating the most efficient model. In the Electronic Goods Industry sector and similar fields, the optimum lag is located at lag 1, with an AIC value of -3.097, which is the smallest value among lags 2 to 4. Meanwhile, for the Basic Metals sector, the optimum lag is at lag 2 with an AIC value of -25.119, which is lower than lags 1, 3, and 4. The selection of these lags is based on the principle that the smaller the AIC value, the better the model is at balancing accuracy and complexity. Thus, lag one is used for the Electronic Goods Industry sector and

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

lag 2 for the Basic Metals sector in the MIDAS-QR model analysis. Furthermore, the MIDAS-QR model is used to analyze the relationship between response and predictor variables at different quantiles. The model is applied to two main sectors: the GDP for the Electronic and Similar Goods Industry Sector and the GDP Growth of the Basic Metals Sector.

Table 6. Optimum Lag test

Variables	*Lag*	AIC
y_1	1	-3.097*
	2	-1.171
	3	0.811
	4	2.089
y_2	1	-24.937
	2	-25.119*
	3	-23.326
	4	-21.760

**lag* optimum

The MIDAS-QR model for the GDP of the Electronic and Similar Goods Industry Sector is presented in various quantiles as follows:

$$\hat{y}_{1,t}(0.05) = 0.225 - 0.193PDB_{L1,t-1} - 0.078.FSI_{t-1} + 0.601.FCI_t - 0.009.FCI2_t - 0.305FCI3_t$$
$$\hat{y}_{1,t}(0.50) = 0.331 - 0.232.PDB_{L1,t-1} - 0.143.FSI_{t-1} + 0.105.FCI_t + 0.374.FCI2_t + 0.228.FCI3_t$$
$$\hat{y}_{1,t}(0.95) = 0.761 - 0.512.PDB_{L1,t-1} + 0.368.FSI_{t-1} - 0.009.FCI_t + 0.030.FCI2_t + 0.249.FCI3_t$$

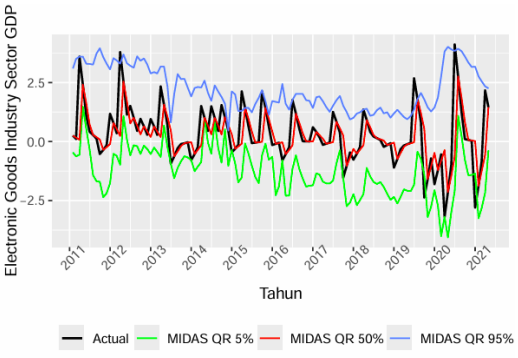
Furthermore, the MIDAS-QR model is also applied to the Basic Metals Sector GDP Growth sector at various quantiles as follows.

The MIDAS-QR model shows how various economic variables affect the growth of the electronic goods industry and the base metals sector at different levels of data distribution. The model offers a more comprehensive understanding of how macroeconomic factors impact these industry sectors under diverse economic conditions.

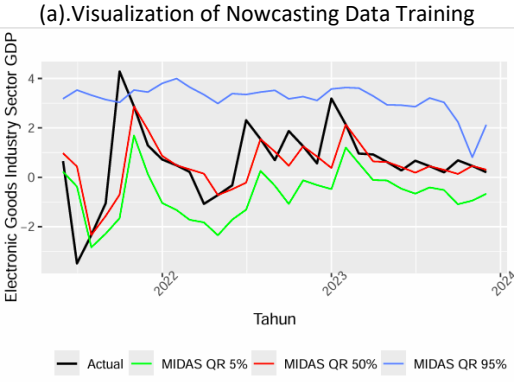
$$\hat{y}_{2,t}(0.05) = 0.716 - 0.674.PDB_{L1,t-1} - 0.409.PDB_{L2,t-2} - 0.670.FSI_{t-1} - 0.339.FCI_t + 0.120.FCI2_t + 1.031.FCI3_t$$
$$\hat{y}_{2,t}(0.50) = 1.066 - 0.656.PDB_{L1,t-1} - 0.236.PDB_{L2,t-2} - 0.094.FSI_{t-1} - 0.129.FCI_t + 0.260.FCI2_t + 0.194.FCI3_t$$
$$\hat{y}_{2,t}(0.95) = 1.447 - 0.523.PDB_{L1,t-1} - 0.359.PDB_{L2,t-2} + 0.138.FSI_{t-1} - 0.155.FCI_t + 0.090.FCI2_t - 0.171.FCI3_t$$

Table 7. Accuracy of Nowcasting Results on Training and Testing Data

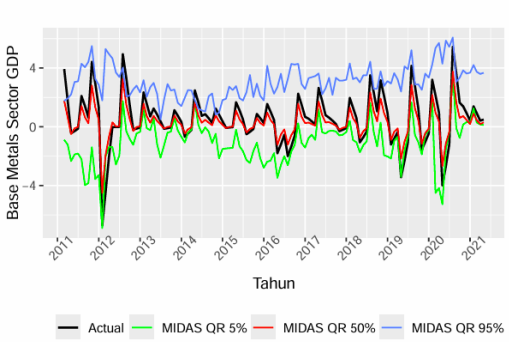
$\tau = 50\%$	
sMAPE *Training* (%)	sMAPE *Testing* (%)
7.32	9.61
5.21	4.09



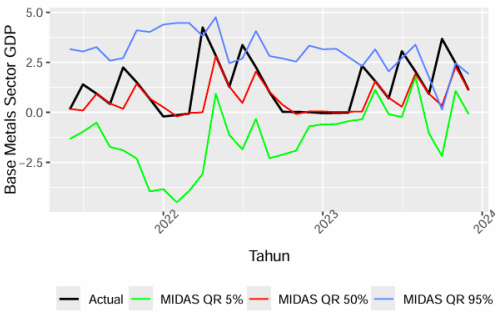
(a).Visualization of Nowcasting Data Training



(b). Visualization of Nowcasting Data Testing



(c). Visualization of Nowcasting Data Training



(d). Visualization of Nowcasting Data Testing

Figure 5. Visualization of Nowcasting Data for Training and Testing MIDAS-QR Model

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

Based on the calculation results, the following table and graph illustrate the performance of the MIDAS-QR model in predicting the GDP growth of the electronic goods industry sector and the basic metals sector, considering various quantiles.

Table 7 compares the sMAPE (Symmetric Mean Absolute Percentage Error) values on the training and testing data for each sector at $\tau = 50\%$. In the GDP of the electronic goods industry sector, the training sMAPE value is 7.32% and the testing sMAPE is 9.61%, indicating that this model has a relatively small prediction error when applied to both training and testing data. In the GDP of the basic metal sector, the sMAPE values for training and testing are 5.21% and 4.09%, respectively, indicating that this sector model has higher accuracy than the electronic goods industry sector. The accuracy of this model is improved, as reflected in the relatively lower sMAPE value. Based on Table 1, the assessment of forecasting accuracy level, as indicated by the sMAPE value, reveals that all sectors fall into the highly accurate category, as the sMAPE for each sector is below 10%.

Based on Figure 5 for the training period, it can be seen that the nowcasting results using MIDAS-QR have the advantage of capturing various levels of quantile distribution, especially for GDP growth, which has high volatility. It can be seen that the prediction for the quantile 0.05 (downside risk) yields a lower prediction, while the prediction for the quantile 0.95 (upside potential) yields a higher prediction. The median prediction of 0.5 is closer to the actual line, reflecting the median or more general expectation. In the electronic goods industry sector, the 0.95 quantile is able to capture extreme spikes more effectively than the 0.5 quantile, which is centered only on the middle value. In the base metals industry sector, the MIDAS-QR model, at a quantile of 0.05, provides some points where the downside prediction offers insight into how the sector could face a sharp decline under adverse conditions. However, these downside predictions remain relatively stable in some periods. At the 0.5 quantile, the predictions at this quantile give an idea of the expected normal conditions. The graph shows that the median prediction follows a good trend with the actual data; however, there are differences at some points with large fluctuations. Meanwhile, the 0.95 quantile indicates the upside potential, representing the most optimistic prediction with the highest value in the distribution. The large spikes in the electronic goods industry and basic metal industry seen in 2020 indicate an extraordinary event, such as a COVID-19 event.

Unlike traditional linear regression, which focuses only on the mean, quantile regression allows us to predict different points across the distribution of the target variable. This provides significant advantages in capturing more detailed information about the data distribution, especially in the face of extreme fluctuations and higher uncertainty, which traditional linear regression often struggles to capture, as it only examines the mean.

MIDAS-QR model on testing data for two sub-sectors of the non-oil and gas processing industry, namely electronic goods and basic metals. The model generally follows the seasonal patterns and trends of actual data quite well, although there are some deviations, especially in extreme fluctuations, as indicated by the striking difference between the actual line and the extreme quantile limits.

5. Conclusions

This study applies the MIDAS-QR model for nowcasting GDP growth in the non-oil and gas processing industry. By integrating high-frequency financial indicators such as FSI and FCI through PCA, the GaR model is evaluated. The model proved to be effective in capturing non-linear growth patterns by using the Almon Lag Exponential polynomial function. The results show that the model provides accurate predictions, especially in the electronic goods and base metals sectors, with the best performance achieved at the median quantile ($\tau = 0.5$), which results in the smallest error.

However, this study has several limitations, including data limitations and frequency differences between the response and predictor variables, which may impact the results. Therefore, further research can consider applying the MIDAS-QR model using weighting functions other than the Almon Exponential, such as the Beta weighting function, as well as increasing the frequency of variables in both the response and predictor to improve model accuracy.

The findings from this study provide policymakers and industry leaders with important insights for addressing economic uncertainty and developing effective macroeconomic policies. For future research, it is recommended to test this model on other sectors or by using a broader dataset to gain deeper insights. This research is also expected to provide policymakers with guidance on designing more effective policies to address economic uncertainty.

6. Acknowledgments

The first author gratefully acknowledges the Indonesian Endowment Fund for Education/ *Lembaga Pengelola Dana Pendidikan* (LPDP), Ministry of Finance Indonesia, for providing the financial support/studentship that enabled the completion of this research. Additionally, we would like to thank the reviewers for their valuable comments.

7. References

Eurostat. (2016). Overview of GDP Flash Estimation Methods. In Overview of GDP flash estimation methods. Luxembourg: Publications Office of the European Union.

Fastasima, B., Zaharani, H. H., Widodo, W. U., & Kiky Aprianto, N. E. (2025). Peran Kebijakan Pemerintah Dalam Mendorong Pertumbuhan Industri. Jurnal Riset dan Publikasi Ilmu Ekonomi, 3, 50-60.

Ferrara, L., Mogliani, M., & Sahuc, J. G. (2020). High-Frequency Monitoring of Growth-at-Risk. CAMA : Centre for Applied Macroeconomic Analysis.

Ghysels, E., Clara, P. S., & Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. Department of Economics University of North Carolina.

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

Ghysels, E., Clara, P. S., & Valkanov, R. (2005). There is a Risk-Return Trade-Off After All. Journal of Financial Economics, 76, 509–548.

Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed Frequency Data Sampling Regression Models: The R Package Midasr. Journal of Statistical Software, 72(4).

Hu, H., Zhao , C., Li, J., & Huang, Y. (2022). Stock Prediction Model Based on Mixed Fractional Brownian Motion and Improved Fractional-Order Particle Swarm Optimization Algorithm. Fractal and Fractional, 6, 1-20.

Koenker, R. (2017). Quantile Regression: 40 Years On. Centre for Microdata Methods and Practice (cemmap).

Koenker, R., & Bassett, G. (1978). Regression Quantiles. Econometrica, 46 No.1, 33-50.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area. International Institute of Forecasters, 27, 529-542.

Ma'Arif, S. (2019). Nowcasting Pertumbuhan Produk Domestik Bruto Indonesia Menggunakan Dynamic Factor Model. In Nowcasting Pertumbuhan Produk Domestik Bruto Indonesia Menggunakan Dynamic Factor Model. Surabaya: Tesis Institut Teknologi Sepulih Nopember.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results , Conclusions and Implications. International Journal of Forecasting, 16, 451-476.

Rantini, D., Octavanny, M. A., Kruba, R., Kuswanto, H., & Fithriasari, K. (2018). Quantile Regression Neural Network Model for Forecasting Consumer Price Index In Indonesia. Inferensi, 1(1), 39-43.

Saputra, W. H., Prastyo, D. D., & Kuswanto, H. (2024). Machine Learning Modeling on Mixed-frequency Data for Financial Growth at Risk. Procedia Computer Science, 234, pp. 397-403.

Saulo, H., Vila, R., & Borges, G. (2023). Modeling Income Data via New Parametric Quantile Regressions: Formulation, Computational Statistics, and Application. Mathematics, 11(2), 1-25.

Sujianto, A. E. (2024). Analisis Penentuan Produk Domestik Bruto Atau Gdp (2 Sektor). Neraca Manajemen Ekonomi, 5 No.8, 1-10.

Supriyatna, P. K., Prastyo, D. D., & Akbar, M. S. (2024). Application of the Dynamic Factor Model on Nowcasting Sectoral Economic Growth with High-Frequency Data. Media Statistika, 17(2), 128-139.

Tsay, R. S. (2010). Analysis of Financial Time Series. USA: Wiley.

Utari, D. T., & Rosadi, D. (2018). Forecasting Indonesian GDP Growth using Mixed Data Sampling (MIDAS) Regression. International Seminar on Innovation in Mathematics and Mathematics Education. Yogyakarta.

White, H., & Kim, T. -H. (2015). VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles. Journal of Econometrics, 187, 169-188.

Xu, M., Xu, Q., Jiang, C., & Zhuo, X. (2023). Mixed-frequency Growth-at-Risk with the MIDAS-QR Method : Evidence from China. Economic Systems, 47.

Xu, M., Xu, Q., Jiang, C., & Zhuo, X. (2025). High-Frequency Growth-at-Risk of China: the Role of Macro-financial Environment. Computational Economics.

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.11
Malaysian Journal of Science 45 (Special Issue 2): 105-113(November 2026)

MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394 - 3065

5th International Conference on Applied & Industrial Mathematics and Statistics 2025 (ICoAIMS 2025)

Special Issue

Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026) https://mjs.um.edu.my

Analysis of Relation Between Indicators in the Sustainable Development Goals in Indonesia Using Higher-Order PLS-PM

Fatimah Ihza Aulia[1a], Subchan Subchan[2b*]

Abstract: The world has been placing greater emphasis on the Sustainable Development Goals (SDGs) over the last decade, including in Indonesia. The SDGs represent a global commitment to enhancing community welfare and are structured around four dimensions: social, economic, environmental, and law and governance. The SDGs include 17 goals and some indicators for their measurement. Indonesia has achieved 62% of the 138 indicators targeted in 2022. The purpose of this research is to model the relationships among the dimensions, goals, and indicators of the SDGs, enabling an analysis of the main factors influencing sustainable development in Indonesia. Higher-Order Partial Least Square Path Modelling (PLS-PM) was used in this research to map the relationship between each dimension, goal, and indicator. The indicators were derived from the annual SDG report as the data from 34 provinces in 2023. The results of the 298 indicators are reduced to 75 indicators that exist its data on provincial level. Goal 3 "Good Health and Well-Being" is giving the most influence in social dimension of SDGs, while for the economic dimension and environment dimension are Goal 8 "Decent Work and Economic Growth" and Goal 13 "Climate Action". The validation of the model shows that its predictive power is 64.55%, indicating that the constructed model closely fits the data. The findings of this study may be the consideration for policymakers in planning sustainable development in Indonesia to achieve the SDGs by 2030.

Keywords: SDGs, Higher-Order PLS-PM, indicator analysis, disaggregated data

1. Introduction

The Sustainable Development Goals (SDGs), or the 2030 Agenda, were declared by the United Nations in 2015 and agreed upon by 193 countries as a global commitment to improving community welfare, replacing the Millennium Development Goals (MDGs) (Rajani & Boluk, 2022). SDGs comprise 17 goals that focus on issues within social, economic, environmental, and law and governance dimensions such as poverty, gender equality, clean energy, justice and others. Indonesia has consistently supported the achievement of the SDGs goals, shown its implementation aligned with national development planning at both the national and sub-national levels. According to the 2023 report on the implementation of SDGs, Indonesia has achieved 62% of the 138 indicators targeted in 2022 (Bappenas, 2023). Several indicators suggest that the achievement status of the SDGs in Indonesia is trending toward reaching targets. However, the majority remain in progress and require substantial efforts to meet the target in 2030. Notably, 55 indicators are at risk of not being achieved, demanding attention since it has been in progress for 10 years.

However, despite these efforts, one of the major challenges in achieving the SDGs lies in their broad and ambitious scope, which spans across 4 dimensions and involves 17 goals, 169 targets, and 241 indicators for measurement. The evaluation of the implementation of the MDGs revealed, among other things, a lack of synergy in efforts to achieve each of the goals (Alisjahbana & Murniningtyas, 2020). Therefore, it is important to understand the interconnections between goals and indicators in the implementation of the SDGs.

To solve this problem, we need a method to discover the relationship between goals and dimensions in SDGs. One of the methods is Partial Least Square Path Modeling (PLS-PM). This model has been done in several research field, such as education (Zafeer et al., 2024), social and management (Maia et al., 2024; Yang et al., 2024), science and technology (de Morais Fernandes et al., 2023; Sun et al., 2024), and other sectors. Likewise, for sustainable development, there have been many studies using PLS-PM to measure the SDGs all over the world. Some of them only discuss one Goal or one sector in a paper, for instance, Bihun et al. (2022) analyze sustainable tourism in Ukraine using PLS PM. This study discusses the most influencing factor in sustainable tourism from the perspective of social, economic, and environmental areas. In Pu & Cheng (2024), PLS-PM was used to see the water quality status in rivers and reservoirs, which is included in Goal 6, in terms of the spatial-temporal dynamics and influence of natural and socioeconomic factors. This paper used a case study in southwest China with annual and seasonal data. The development of studies in PLS-PM, Higher-Order PLS-PM, is also being done, in which the relationship between unobserved variables is carried out in tiers and consisting of lack of observed variables or indicators. Cataldo et al. analyze and address gender

Authors information:
[a]Department of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, INDONESIA. E-mail: fatimahihza1@gmail.com[1], subchan@its.ac.id[2]

*Corresponding Author: subchan@its.ac.id.

Received: October 28, 2025
Accepted: May 6, 2026
Published: November 15, 2026

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

equality in EU countries using Higher-Order PLS-PM to achieve SDG Goal 5 (Cataldo et al., 2024). In Cataldo et al. (2019), second-order model in Higher-Order PLS-PM were used to see the quality of life in the historic center of Naples in order to measure the well-being of the citizens (SDGs Goal 3). Meanwhile, there are only few of the studies discussing some of goals and the whole SDGs. Crocetta et al. proposed the model to see the relationship between all of the Goals in social area (Goal 1 to Goal 6) in third order model of Higher Order PLS-PM. Then, this research compares four approaches in higher-order system (Crocetta et al., 2021). The model of the whole SDGs, including all of the indicators related to Goals, Dimensions, and SDGs itself has been done in Cataldo et al. (2021). This paper using the case study of European countries.

Among the studies mentioned, PLS-PM has been widely used in various research, including those on SDGs. However, none of these studies specifically focus on the de tailed mathematical modeling of the entire SDG framework using the most recent data, particularly in the context of Indonesia. Therefore in this work, we propose mathematical modeling of sustainable development involves its dimensions, goals, and indicators. We show the mathematical model and analyze the relationship between each dimension, dimensions and correlated goals, each goal and goals and correlated indicators. The indicators were used as data derived from the annual SDG report by the Ministry of National Development Planning as data from 2019 to 2023. The propose of this research can be used as an evaluation and reference for policymakers in planning development in Indonesia to achieve the SDGs targets by 2030.

This paper is structured as follows. Section 2 describes the theory of the PLS-PM framework, while Section 3 describes Higher-Order PLS-PM. Then, the mathematical model is proposed in Section 4 along with the data of this paper. The interpretation and analysis results are discussed in Section 5. Finally, Section 6 concludes the paper.

2. Partial Least Square Path Modeling (PLS-PM)

Partial Least Square (PLS) is not a new concept in data analysis. PLS Structural Equation Models (SEM), which are more commonly known, are used as a tool for investigating complex models involving models that cannot be measured directly. PLS-PM tends to think of the cause-and-effect relationship of unobservable variables with their measurable indicators, putting PLS-PM in the PLS-SEM family (Cataldo et al., 2024).

PLS-PM is a statistical technique used for analyzing groups of observed variables, where each group is represented by a latent variable, and linear relationships are assumed to exist between these latent variables (Sanchez, 2013). Latent variable (LV) is an unobservable variable, on the other hand, an observed variable named as indicator or manifest variable (MV). We can indirectly measure LV by its consequences or effects indicators (MV). The primary distinction between formative and reflective approaches lies in their causal-effect relationships between the indicators and latent variables (LV). A model may contain a combination of both formative and reflective indicators, depending on the context of the problem.

PLS-PM comprised by two types of models, inner model and outer model. The inner model (or structural model) describes the relationship between each latent variable. Meanwhile, the outer model (or measurement model) has to do with the relationship between the latent variable and its block of indicators. Both an inner and an outer model are treated as linear relationships. Given a j number of LV and k number of MV, the mathematical equation for inner model, the relationship between LVs, expressed as:

$$LV_j = \sum_{i \to j} \beta_{ji} LV_i + error_j \quad (1)$$

Since the outer model describes the relationship between latent variable and its indicator, which can be interpreted in two types, formative and reflective, there are also two types of outer model, depends on its interconnection.

$$LV_j = \lambda_{jk} X_{jk} + error_j \ (Formative) \quad (2)$$

$$X_{jk} = \lambda_{jk} LV_j + error_{jk} \text{ (Reflective)} \quad (3)$$

Index k indicates the number of indicators in its LV. Parameter λ_{jk} are called loadings, as the relationship between LV and its MVs. The purpose of using PLS-PM is to get the estimates value of LVs and parameters, path coefficients and loadings.

3. Higher-Order PLS-PM

PLS-PM is a tool to measure an index of abstract concept named latent variable with respect to its relationship of other latent variables and its measurable indicators. But in fact, not all of the LV has indicator to be measured. Some of the LVs constructed as higher level of abstraction, which means, involve more than one dimension. To handle this case, researchers formed a new method for measure the LV without requiring all the LVs to have its associated MVs, later we call it as High-Order PLS-PM. Some approaches have been developed and proposed by researches to work with Higher Order PLS-PM: the Repeated Indicators Approach; the Hybrid Approach; and the Two Step Approach (Sanchez, 2013).

The approach used in this paper is the Mixed Two Step Approach. Based on Crocetta et al. (2021), it was found that the mixed two step approach showed the best results because it provided precise accuracy. The mixed two step approach started with the implementation of PLS-PM with re peated indicators. First, the indicators of lower-order LV are repeated for the higher-order LV. Then the algorithm of PLS-PM is run and we can obtain the scores of LV. After that, the scores of LVs are used to become the indicators for its higher-order LV. The algorithm of PLS-PM is run once again and we can obtain the finalize path coefficients and loadings for each LV and MV. This method use the components that represent best of each block and at the same time include the best prediction power for its higher-order LV, compared to repeated indicators approach. (Cataldo et al., 2021).

4. The Data and the Model

The concept of sustainable development goals, announced by the United Nations, is highly suitable with PLS-PM analysis that worked for complex and abstract concept. The SDGs aims to achieve all the 17 goals contained in four dimensions—social, economy, environment, and law and governance dimension. The

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

relationship between these three variables, SDGs, its dimensions and correlated Goals, can be represented in a form of path diagram. The variables are difficult to measure because of their complexity; thus, they require indicators for their measurement. For this reason, SDGs, the four dimensions, and the 17 goals are the latent variables (LV), whereas indicators in each goal are the manifest variables (MV).

not have indicators, it is addressed using the Higher-Order PLS-PM framework.

The first-order LVs, representing the Goals in the SDGs, are measured formatively. Similarly, the relationships between each order of latent variables are treated as formative modes, where the lower-order LVs form the higher-order LVs. The path diagram of the proposed SDGs model is presented in Figure. 1.

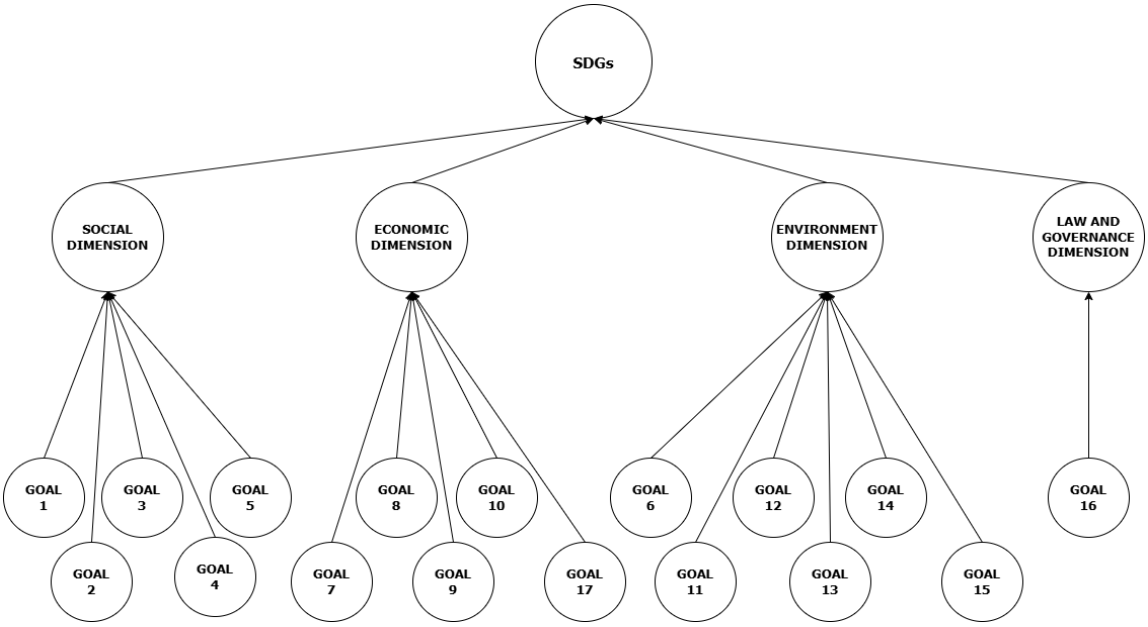


Figure 1. Path Diagram of SDGs.

UN provides indicators of every goal and able to be adapted depends on the country's condition. In 2023, Indonesia developed 298 indicators for measuring the achievement of 17 goals and the whole SDGs. These indicators are the data in this paper, derived from provincial-level data specifically addressing SDG achievements across 34 provinces in Indonesia. However, due to limited data on developing countries, there are no data for several indicators; thus, 298 indicators are being reduced to suit the data availability.

The construct comprises a third-order LV. The first-order LV represents the goals in SDGs, starting from Goal 1, Goal 2, Goal 3 to Goal 17. At this level, every LV has its own indicator or MVs for measurement. The data indicators were obtained from the annual SDGs report by the Ministry of National Development Planning (Bappenas, 2023).

The second-order LVs correspond to the four dimensions of the SDGs, which are formed by related goals. The social dimension encompasses Goals 1 to 5. The economic dimension is related to latent variables of clean energy, decent work, economic growth, industry, inequality, and partnership. The environmental dimension LV includes the most extensive range of latent variables. Meanwhile, only Goal 16 is associated with the dimension of law and governance.

These four second-order LVs are independent of MVs and collectively form a higher order LV, the SDGs. Due to the construct comprising high-order LVs and the presence of some LVs that do

As stated in the previous section, PLS-PM consists of two models: the internal model, which describes the relationship between latent variables, and the outer model, which describes the relationship between latent variables and their manifest variables. The internal model in this research is shown in Eq. (4).

$$
\begin{aligned}
&LV_{SDGs} = \beta_1 LV_{soc} + \beta_2 LV_{eco} + \beta_3 LV_{env} + \beta_4 LV_{gov} \\
&LV_{soc} = \beta_{11} LV_{Goal1} + \beta_{12} LV_{Goal2} + \beta_{13} LV_{Goal3} + \beta_{14} LV_{Goal4} + \\
&\beta_{15} LV_{Goal5} \\
&LV_{eco} = \beta_{21} LV_{Goal7} + \beta_{22} LV_{Goal8} + \beta_{23} LV_{Goal9} + \\
&\beta_{24} LV_{Goal10} + \beta_{25} LV_{Goal17} \\
&LV_{env} = \beta_{31} LV_{Goal6} + \beta_{32} LV_{Goal11} + \beta_{33} LV_{Goal12} + \\
&\beta_{34} LV_{Goal13} + \beta_{35} LV_{Goal14} + \beta_{36} LV_{Goal15} \\
&LV_{gov} = \beta_{41} LV_{Goal16}
\end{aligned}
\tag{4}
$$

The outer model describes the relationship between LVs and their associated MVs. In this SDGs model, each Goal is represented by an LV, and each LV has its own indicators or MVs. Therefore, it is essential to verify the availability of data (indicators) for the implementation of the SDGs in Indonesia. This paper utilizes provincial-level data from 2023, encompassing 34 provinces, as listed in the annual SDGs report by Statistics Indonesia (BPS, 2024). Out of 198 indicators published, only 75 indicators that available its data on provincial level. Based on the availability of data indicator, we propose the outer model showed in Eq. (5).

$$
\begin{aligned}
&LV_{Goal_1} = \lambda_{11} X_{1.2.1} + \lambda_{12} X_{1.4.1} + \lambda_{13} X_{1.4.2} + error_{Goal1} \\
&LV_{Goal2} = \lambda_{21} X_{2.1.1} + \lambda_{22} X_{2.1.2} + \lambda_{23} X_{2.2.1} + \lambda_{24} X_{2.2.2} +
\end{aligned}
$$

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

$\lambda_{25}X_{2.3.1} + \lambda_{26}X_{2.3.1.(a)} + error_{Goal2}$

$LV_{Goal3} = \lambda_{31}X_{3.1.2.(a)} + \lambda_{32}X_{3.3.3} + \lambda_{33}X_{3.3.3.(a)} + \lambda_{34}X_{3.3.5.(a)} + \lambda_{35}X_{3.3.5.(b)} + \lambda_{36}X_{3.4.1.(b)} + \lambda_{37}X_{3.4.1.(c)} + \lambda_{38}X_{3.5.1.(b)} + \lambda_{39}X_{3.8.1.(a)} + \lambda_{310}X_{3.a.1} + \lambda_{311}X_{3.b.1} + \lambda_{312}X_{3.b.3} + \lambda_{313}X_{3.c.1} + error_{Goal3}$

$LV_{Goal4} = \lambda_{41}X_{4.1.1.(a)} + \lambda_{42}X_{4.1.2} + \lambda_{43}X_{4.1.2.(a)} + \lambda_{44}X_{4.2.2} + \lambda_{45}X_{4.3.1} + \lambda_{46}X_{4.3.1.(a)} + \lambda_{47}X_{4.4.1.(a)} + \lambda_{48}X_{4.5.1} + \lambda_{49}X_{4.6.1.(a)} + \lambda_{410}X_{4.a.1} + \lambda_{411}X_{4.c.1} + error_{Goal4}$

$LV_{Goal5} = \lambda_{51}X_{5.3.1} + \lambda_{52}X_{5.5.2} + \lambda_{53}X_{5.a.1} + \lambda_{54}X_{5.b.1} + error_{Goal5}$

$LV_{Goal6} = \lambda_{61}X_{6.2.1} + \lambda_{62}X_{6.3.2.(a)} + \lambda_{63}X_{6.6.1} + error_{Goal6}$

$LV_{Goal7} = \lambda_{71}X_{7.1.1} + \lambda_{72}X_{7.1.2.(a)} + error_{Goal7}$

$LV_{Goal8} = \lambda_{81}X_{8.1.1} + \lambda_{82}X_{8.1.1.(a)} + \lambda_{83}X_{8.2.1} + \lambda_{84}X_{8.3.1} + \lambda_{85}X_{8.5.1} + \lambda_{86}X_{8.5.2} + \lambda_{87}X_{8.5.2.(a)} + \lambda_{88}X_{8.7.1} + \lambda_{89}X_{8.7.1.(a)} + \lambda_{810}X_{8.9.1.(b)} + error_{Goal8}$

$LV_{Goal9} = \lambda_{91}X_{9.1.2.(a)} + \lambda_{92}X_{9.2.1} + \lambda_{93}X_{9.2.1.(a)} + \lambda_{94}X_{9.2.2} + \lambda_{95}X_{9.b.1.(a)} + error_{Goal9}$ (5)

$LV_{Goal10} = \lambda_{101}X_{10.1.1} + \lambda_{102}X_{10.1.1.(a)} + \lambda_{103}X_{10.3.1.(a)} + error_{Goal10}$

$LV_{Goal11} = \lambda_{111}X_{11.1.1.(a)} + \lambda_{112}X_{11.6.2.(b)} + error_{Goal11}$

$LV_{Goal12} = \lambda_{121}X_{12.3.1.(a)} + \lambda_{122}X_{12.5.1.(a)} + error_{Goal12}$

$LV_{Goal13} = \lambda_{131}X_{13.1.1} + error_{Goal13}$

$LV_{Goal14} = \lambda_{141}X_{14.5.1} + error_{Goal14}$

$LV_{Goal15} = \lambda_{151}X_{15.1.2.(a)} + error_{Goal15}$

$LV_{Goal16} = \lambda_{161}X_{16.7.2.(a)} + \lambda_{162}X_{16.7.2.(b)} + \lambda_{163}X_{16.7.2.(c)} + \lambda_{164}X_{16.9.1} + \lambda_{165}X_{16.9.1.(a)} + \lambda_{166}X_{16.9.1.(b)} + error_{Goal16}$

$LV_{Goal17} = \lambda_{171}X_{17.1.1} + \lambda_{172}X_{17.1.1.(a)} + \lambda_{173}X_{17.8.1} + error_{Goal17}$

5. Result and Discussion

This section discusses the analysis and interpretation of results using PLS-PM for SDGs in Indonesia. The analysis of the results of Higher-Order PLS-PM model must be performed sequentially. The first step is to analyze the results from the outer model then the inner model, because before we look at the value of the latent variable in the inner model, we first need to look at the measuring tool, it is the indicators contained in the outer model. Subsequently, the results of the inner model provide the structural model, which illustrates the relationships between latent variables (LVs) and other relevant factors. Once the parameters for the outer and inner models are obtained, model validation is necessary by assessing the model's Goodness of Fit (GoF).

5.1 Analysis on the Result of Outer Model

The outer model in this SDGs model work in formative mode, where the indicators or MVs form the latent variables. There are two important aspects that can be obtained from the outer model analysis; correlations and loadings. However, before analyzing the result of the outer model, we must evaluate the outer model obtained.

In the outer model using formative mode, evaluation can be done by obtaining multicollinearity, which indicates that there are indicators that are similar to other indicators. There are indications of multicollinearity in the outer model of the SDGs in Indonesia, as shown by the VIF values that exceed 5. Table 1 presents the VIF values for each indicator. It can be seen that indicators 4.2.2, 4.4.1.a, 8.2.1, 8.7.1, 8.7.1.a, 16.9.1, 16.9.1.a, and 16.9.1.b have VIF values greater than 5, which are highlighted in red. This indicates that these indicators are highly similar to other indicators, suggesting that they can be removed. After the high-value VIF indicators were removed and the PLS-PM algorithm was rerun, no multicollinearity was detected; thus, the result of outer model is significant. Note that there are now only 66 indicators remaining due to evaluation of the outer model, which show that the eliminated indicators are similar to others.

Table 1. The Value of VIF.

Indicators	VIF	Indicators	VIF	Indicators	VIF	Indicators	VIF
1.2.1	1.173	3.b.3	1.652	7.1.1	1.491	10.1.1.a	1.059
1.4.1	1.375	3.c.1	2.960	7.1.2.b	1.491	10.3.1.a	1.076
1.4.2	1.568	4.1.1.a	3.361	8.1.1	2.027	11.1.1.a	1.014
2.1.1	1.956	4.1.2	2.623	8.1.1.a	3.615	11.6.2.b	1.014
2.1.2	1.533	4.2.2	10.091	8.2.1	2.972	12.3.1.a	1.051
2.2.1	2.956	4.3.1	3.589	8.3.1	7.887	12.5.1.a	1.051
2.2.2	3.699	4.4.1.a	10.257	8.5.1	3.486	13.1.1	1.000
2.3.1	1.451	4.5.1	1.979	8.5.2	2.952	14.5.1	1.000
2.3.1.a	1.550	4.6.1.a	2.881	8.5.2.a	2.897	15.1.2.a	1.000
3.1.2.a	2.591	4.a.1	1.924	8.7.1	46.778	16.7.2.a	2.604
3.3.3	2.207	4.c.1	3.076	8.7.1.a	50.235	16.7.2.b	1.123
3.3.5.a	3.242	5.3.1	1.274	8.9.1.b	1.515	16.7.2.c	1.123
3.3.5.b	1.885	5.5.2	1.213	9.1.2.a	1.571	16.9.1	8.833
3.4.1.b	1.428	5.a.1	1.182	9.2.1	1.461	16.9.1.a	111.928
3.5.1.b	3.448	5.b.1	1.360	9.2.1.a	2.088	16.9.1.b	100.667
3.8.1.a	2.397	6.2.1	1.325	9.2.2	2.178	17.1.1	1.057
3.a.1	2.207	6.3.2.a	1.244	9.b.1.a	1.771	17.1.1.a	1.129
3.b.1	2.461	6.6.1	1.574	10.1.1	1.110	17.8.1	1.184

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

The outer model also provides the loading values that represent the relationships between LVs and their MVs. Loadings indicate the direction and magnitude of the indicator's contribution of the latent variable. As we can see in Figure 2, most of the loading value is positive. This indicates that the relationship between MV and its correlated LV is positively related, increasing the value of the indicator will increase the value of the latent variable. In Goal 1, indicator 1.2.1* has the largest loadings compared to other indicators, which indicates that the percentage of people living below the national poverty line is the most influencing factor in the achievement of no poverty goal. Indicator 1.1.1* and 14.1.1* show the highest loading values. In other words, these indicators are the ones that most influence its corresponding LV. The next highest loading value is 2.1.1 which is the prevalence of undernourishment, making it the most influencing factor in Goal 2. Indicators in Goal 3 do not show loadings value indicator 3.c.1 or density and distribution of health workers is the most influencing indicator. Indicator 4.1.2 and indicator 5.3.1 namely level of completion of high school and proportion of women whose age at first marriage was before the age of 18 years old are the most influencing factor in Goal 4 and Goal 5, respectively.

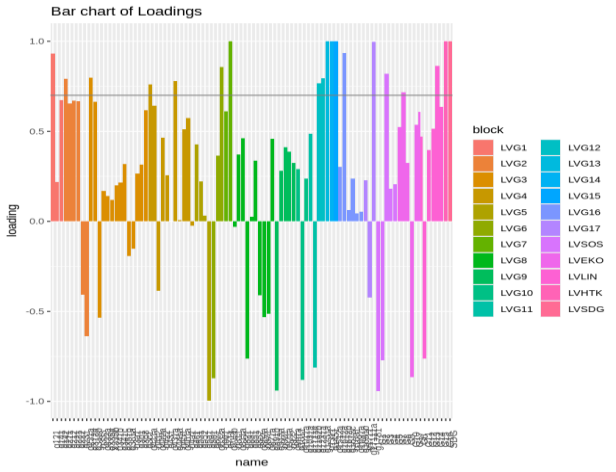


Figure 2. Loadings.

The percentage of households that have access to adequate sanitation become the most influencing factor in achieving Goal 6, indicated by the loadings value of indicator 6.2.1 being the highest among all the indicators. Subsequently, the higher the ratio of household gas usage, the higher the achievement of Goal 7. This is shown from indicator 7.1.2.b has a positive relationship with Goal 7 and is the indicator with the highest value of loadings. Indicators in Goal 8, Goal 9 and Goal 10 do not show a large loadings value. Indicator 8.2.1, 9.b.1.a, and 10.3.1.a is the most influencing factors of achievement in each correlated Goals, which are GDP growth rate per worker, proportion of exports of high-tech industrial products and freedom index. Indicator 11.1.a is the percentage of household that have access to affordable living giving the high value of loadings in Goal 11, in contrast to indicator 12.3.1.a is the percentage of food waste which shows a small loading value but is the most influencing factor in Goal 12. In addition, Goal 13, 14, and 15 each only have 1 indicator and its loading is all positive, which means that each indicator has a direct positive effect on the achievement of correlated Goal.

However, there are negative loadings in the outer model of SDGs in Indonesia. Negative loadings indicate that the indicator has an opposite relationship with the latent variable; that is, increasing the value of the indicator reduces the value of the latent variable. There are 2 indicators in Goal 2 that have negative loadings; 2.3.1 and 2.3.1. In addition, indicator 5.b.1 in Goal 5 also has negative loading. Goal 3 and 8 have the most indicators with negative loadings, namely indicator 3.3.3, 3.3.5.a, 3.8.1.a and 3.a.1 and indicators 8.1.1, 8.5.1, 8.5.2.a and 8.7.1. Similarly, indicators 9.1.2.a, 10.1.1.a, 12.5.1.a, and 17.1.1.1.a have an opposite relationship with the Goal. This indicates that the higher the value of indicators such as malaria incidence, unmet need of health service, the percentage of children aged 5-17 who work and etc. the lower the achievement of SDGS in Indonesia.

The results of the outer model were found to be significant, as shown by evaluating the model by checking for the multi-collinearity. Each indicator with VIF value is greater than 5 is indicated to have multi-collinearity, so it must be removed. After re moving the indicator, the PLS-PM algorithm using 66 indicators was run again so that significant correlations and loadings were obtained.

5.2 Analysis on the Result of Inner Model

After evaluating the quality of the outer model, the results of the inner model can be examined. In analyzing this model, correlations are calculated to determine the strength and direction of the relationship between two variables, as well as the path coefficient, which indicates the relationship between one LV and another. Similar to the outer model, it is essential to evaluate the inner model before analyzing its results. This evaluation involves examining the outcomes of each regression within the model. The quality of the inner model can be assessed based on two aspects: the R^2 coefficient of determination and the redundancy index.

The R^2 indicates the amount of variance in the endogenous latent variable, which is explained by its independent latent variables. The model is considered good if the R^2 is high ($R^2 > 0{,}50$). It is shown in Table. 2 that the inner model of SDGs in Indone sia is considered good, since the value of all the R^2 are greater than 0.6. Redundancy measures the ability of a set of independent latent variables to explain variation in the dependent latent variable. The redundancy of SDGs LV has a value of approximately 0.865, indicating that the four dimensions account for 86.5% of the variability in the SDGs. Based on the evaluation of the inner model, it is determined that the inner model is both good and significant, allowing for further analysis of the inner model of the SDGs in Indonesia.

Figure. 3 representing the obtained path coefficient by SDGs model using PLS-PM. Path coefficient represents the strength and direction of relationship between latent variable. As shown in the figure, some path coefficient values are negative, indicating that the relationship between these two LVs should be in the opposite direction. However, to identify the factors that most influence the intended LV, is seen from the absolute value. Therefore, although the path coefficient of Goal 3 on the social dimension is negative,

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

Table 2. Inner Model Quality.

	Type	R^2	Communality	Redundancy
Social Dimension	Endogenous	0.999	0.879	0.878
Economic Dimension	Endogenous	0.996	0.805	0.805
Environment Dimension	Endogenous	0.986	0.479	0.476
Law and Governance Dimension	Endogenous	1.000	0.863	0.863
SDGs	Endogenous	1.000	0.865	0.865

which means that if Goal 3. "Good Health and Well-Being" increases, the social dimension decreases, Goal 3 is the most influential factor in achieving SDGs in the social dimension after Goal 5 "Gender Equality" and Goal 1 "No Poverty". In the economic dimension, Goal 7 "Affordable and Clean Energy" is the most influencing factor, followed by Goal 8 "Decent Work and Economic Growth" and Goal 9 "Industry, Innovation and Infrastructure". Goal 10 "Reduced Inequalities" and Goal 17 "Partner ships for the Goals" have a very small path coefficient values, which indicates that those Goals only make a small contribution to the achievement of the SDGs in the economic dimension, but that does not mean they are not important. Then, Goal 13 "Climate Action" is the most influential factor in the achievement of SDGs in the environment dimension, because it shows the largest path coefficient value. Although Goal 6 shows a negative path coefficient, its absolute value shows the second largest value among other Goals in the environment dimension, which means that Goal 6 "Clean Water and Sanitation" is the second most influencing factor in the achievement of SDGs in Indonesia, especially in the environment dimension. The law and governance dimension only have one correlated LV, Goal 16 "Peace, Justice and Strong Institution", and its path coefficient value is positive and significant.

In terms of the second-order and third-order LV, the path coefficient between the four dimensions and the SDGs itself. The four dimensions give the same portion of contribution to the achievement of SDGs in Indonesia, which means that the difference in the path coefficient values of the four dimensions is very small. Of the four dimensions, the social dimension is the most influential factor on achieving SDGs in Indonesia. The other three dimensions contribute the same, with only slightly different numbers. The achievement of SDGs in Indonesia in 2023 is evenly distributed across all dimensions, not only excelling in one or two other dimensions. This is a good thing because sustainability requires a balance of progress across all dimensions.

The Higher-Order PLS-PM approach also allows us to predict the SDGs score as well as its four dimensions for each of the 34 provinces in Indonesia. This process allows us to rank each province by assigning it score. The latent SDGs score shows the value of the achievement of SDGs in each province based on the data availability and data processing. A high value of SDGs indicates that the quality of SDGs in the specific province is good, while a low and negative values indicates that the quality of the SDGs is low. Figure 4 shows that the province of DKI Jakarta in on top of the list, showing that the quality of SDGs in the capital of Indonesia is the best out of all the province. Provinces in Java and

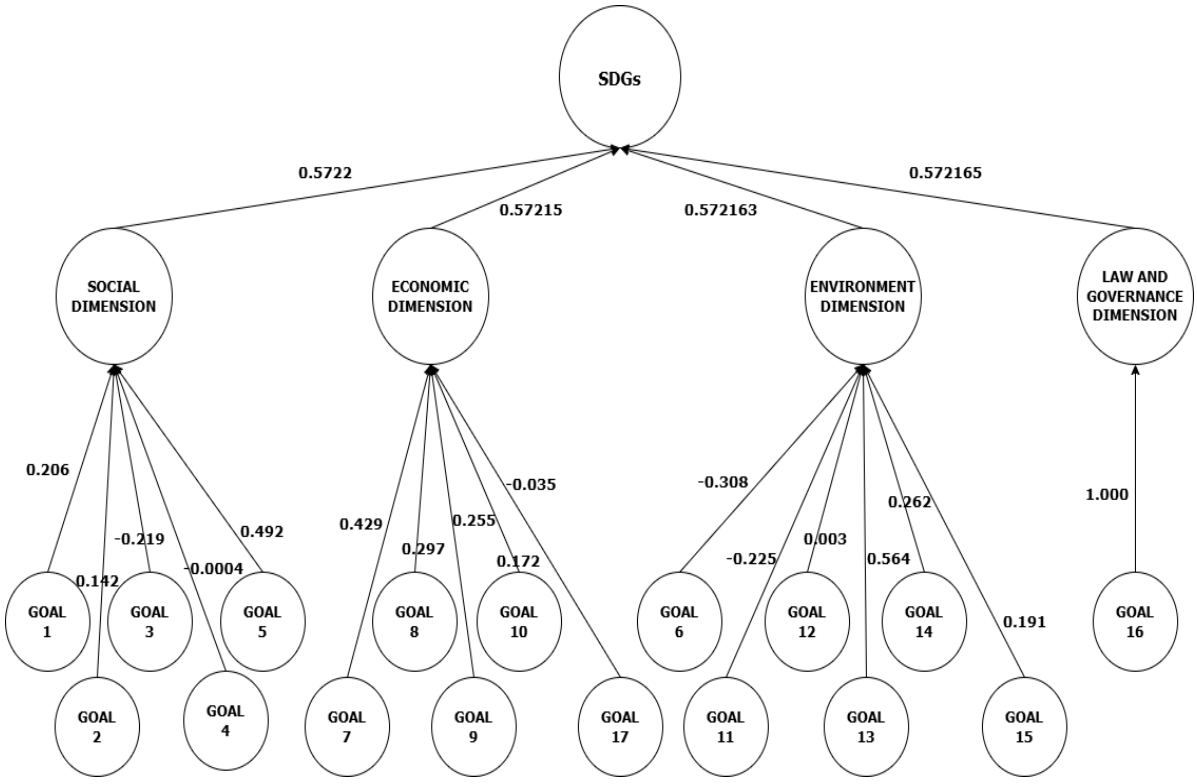


Figure 3. Path Coefficients in the Inner Model

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

Kalimantan island are mostly on the top 10 ranking, while provinces in Sumatera and Sulawesi island are on the mid-rankings. It can be also seen that provinces in the east part of Indonesia are majorly on the bottom 10 ranking. This indicates that the sustainable development performance in Indonesia is uneven. It can be concluded that there needs to be equality in achieving SDGs in Indonesia, so that it is not only centered on one region that is superior compared to others.



Figure 4. Ranking of 34 Provinces in Indonesia according to the SDGs Achievements in 2023

5.3 Model Validation

The analysis of the outer and inner models led to the value of parameter, which are loadings and path coefficients for the SDGs mathematical model.

$$LV_{SDGs} = 0.5722 \cdot LV_{soc} + 0.57215 \cdot LV_{eco} + 0.572163 \cdot LV_{env} + 0.572165 \cdot LV_{gov}$$

$$LV_{soc} = 0.260 \cdot LV_{Goal1} + 0.142 \cdot LV_{Goal2} - 0.219 \cdot LV_{Goal3} - 0.0004 \cdot LV_{Goal4} + 0.492 \cdot LV_{Goal5}$$

$$LV_{eco} = 0.429 \cdot LV_{Goal7} + 0.297 \cdot LV_{Goal8} + 0.255 \cdot LV_{Goal9} + 0.172 \cdot LV_{Goal10} - 0.035 \cdot LV_{Goal17} \quad (6)$$

$$LV_{env} = -0.308 \cdot LV_{Goal6} - 0.225 \cdot LV_{Goal11} + 0.003 \cdot LV_{Goal12} + 0.564 \cdot LV_{Goal13} + 0.262 \cdot LV_{Goal14} + 0.191 \cdot LV_{Goal15}$$

$$LV_{gov} = 1.000 \cdot LV_{Goal16}$$

The validation of the model obtained in Eq. 6 can be validated by calculating its Goodness of Fit (GoF). GoF is measured as the geometric mean of the average communality and the average R^2 value. The Goodness of Fit (GoF) index represents the average predictive power of the entire model. The GoF index can be used to evaluate the performance of both the inner and outer models can be evaluated. There are differences of opinion among scientists regarding the number that could be considered a good GoF value. In Boubker & Douayri (2020), the GoF index is divided into 4 criteria: (1) no fit ($GoF < 0.1$); (2) small ($0.1 \leq GoF < 0.25$); (3) medium ($0.25 \leq GoF \leq 0.36$); (4) large ($GoF \geq 0.36$). In this model, the GoF value was 0.6455, which means that the prediction power of the model is of 64.55%. The obtained model is expected to be strongly valid for predicting the relationship between LV and MV.

In Henseler & Sarstedt (2013), show that GoF cannot be the only standard of model validity. Although shows validity for both inner and outer models,or the formative mode, it is also impor tant to check the R^2, which describes how the endogenous latent variable is explained by exogenous latents and each indicator forms the latent. From Table 3, it can be seen that the value of R^2 is greater that 0.5, which means that it is considered good. This indicates that the proposed model is fit to the data.

6. Conclusions

The use of Partial Least Square Path Modeling (PLS-PM) proved to be efficient in forming the mathematical modeling for Sustainability Development Goals in Indonesia. The latent variables and manifest variables (indicators) form the path model for SDGs in third-order construct. The latent variables consist of 17 Goals and its corresponding four dimensions in SDGs where they form the latent variable for SDGs itself. The manifest variables consist of indicators in every Goal, based on the annual report of SDGs implementation from 34 provinces in 2023. From 298 indicators on the report, only 77 indicators are available for data at provincial level. It is evident that the lack of data is a crucial factor in this context, as the goal cannot be fully measured due to data loss.

The Higher Order PLS-PM framework facilitates the analysis of relationships between indicators, Goals, and their dimensions in measuring the implementation of the SDGs. In this study, we propose two types of models in PLS analysis: the outer model and the inner model. The outer model expresses the relationship between manifest variables with its latent variable, which means what it refers to in this concept is the relationship between

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

indicators and its related Goal. By using this model, we obtain loading values that serve as parameters to assess the strength and direction of relationships. The inner model shows the relationship between latent variables. In this model, it is arranged in third order, where the first order is the 17 Goals, then the second order is dimensions in SDGs, namely, social, economic, environment, law and governance dimension and third order is the SDGs itself. This inner model work in formative-formative mode, where the lower order LV form the higher one. The inner model describes the relationship between each Goal with its corresponding dimension and dimensions with latent SDGs. The approach used in this study is Mixed Two Step Approach.

Although the loadings and path coefficients were obtained, the results cannot be directly analyzed. It is necessary to evaluate the model first, which sequentially starts from the outer model and then the inner model. Evaluation of the outer model of SDGs in Indonesia found that there is multicollinearity indicated by the VIF value of the MV which is more than 5, so the mV must be removed and re-run the algorithm with 66 MV. This means that 11 out of 77 SDGs indicators were eliminated to obtain accurate results in measuring SDGs in Indonesia. The inner model is evaluated to determine whether it is significant and appropriate, by considering the R^2 coefficient determination and redundancy index. The R^2 and redundancy of the model are good number, close to 1.00, meaning that the inner model is considered good. Thus, the next step is to analyze the results of the outer model and inner model.

Based on our model, the loading values in the outer model are either positive or negative. The positive loadings mean that increasing indicator it related Goal, while negative loadings apply the other way around. MV that has the largest loadings value compared to other loadings indicating that the indicator is the most influencing factor in the achievement of its Goal. Note that the MV with the smallest loadings does not mean that it is a factor that does not provide influence; the indicators make a small contribution to its related Goal.

The results of the inner model are presented in the form of path coefficients, which represent the strength and direction of relationships between latent variables. In this context, these coefficients allow us to identify the most influential Goal within each dimension and the dimension with the highest value in the implementation of the SDGs. The path coefficient of each dimension towards the SDGs is the same, but it differs slightly in small amount of number. This means that all dimensions make the same contribution to the achievement of the SDGs, there is no one dimension that is superior to the others and causes the quality of one another is too lag behind. This is a good thing because there is balance and sustainability in all dimensions.

Goal 5 “Gender Equality” excels all the Goals in social dimension, followed by Goal 1 “No Poverty” then Goal 3 “Good Health and Well-Being”. The quality of SDGs in Indonesia is largely influenced by the factors of equal rights for both men and women, poverty and the health of its citizens. For the economic dimension, Goal 7 “Affordable and Clean Energy” and Goal 8 “Decent Work and Economic Growth” have the highest value of path coefficients. In other words, Indonesia’s implementation on SDGs has showed a good performance in achieving these goals. Two Goals in environment dimension, Goal 13 “Climate Action” and Goal 6 “Clean Water and Sanitation” are the most influencing factor in the achievement of SDGs in Indonesia, as indicated by the largest path coefficient value. A small path coefficient value does not mean that it is not an influencing factor; however, they contribute little to the achievement of the SDGs. This means that Goal 4 “Quality Education”, Goal 17 “Partnerships for the Goals” and Goal 12 “Responsible Production and Consumption” need to get more attention by the government so that those Goals can also make a major contribution to the quality of SDGs in Indonesia.

From the score of each LV, the quality of SDGs in each province in Indonesia can be obtained based on the data. It is found that DKI Jakarta, DI Yogyakarta, and Kalimantan Timur occupy the top three in the rank of quality of SDGs in Indonesia. Meanwhile, the bottom ranks are occupied by the eastern part of Indonesia, such as Maluku, Papua, Maluku Utara and Papua Barat. The scores from these provinces also have a considerable difference, indicating that there is significantly different progress in achieving SDGs in each province.

Overall, based on the model obtained, the index of each LV shows the quality of the SDGs, dimension and Goals in Indonesia alongside 34 provinces in it. The prediction power of the model proposed by calculating the GoF is 64.55%, which indicates that the model fits the data. Thus, the results of this study may provide insights for policymakers in planning sustainable development in Indonesia to achieve the SDGs by 2030. Due to data unavailability, the analysis and interpretation of this model are applicable only to data with 66 SDGs indicators in Indonesia. There should be more complete data for other indicators mentioned or review of the indicators on the SDGs by the Indonesian government so that the measurement of SDGs progress is tangible and close to what is happening in reality.

7. Acknowledgments

The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

8. References

Alisjahbana, A.S., Murniningtyas, E. (2020) Tujuan Pembangunan Berkelanjutan (SDGs) di Indonesia. Unpad Press. Available online, accessed December 27, 2024.

Bappenas (2023) Laporan Pelaksanaan Pencapaian Tujuan Pembangunan Berkelanjutan (TPB/SDGs) Tahun 2023. Available online, accessed January 9, 2025.

Bihun, R., Lytvyn, V., Oleksiv, N. (2022) Mathematical modeling of tourism development in terri torial communities, Technol. Audit Prod. Reserves 2(2/64): 21–30.

Boubker, O., Douayri, K. (2020) Dataset on the relationship between consumer satisfaction, brand attitude, brand preference and purchase intentions of dairy product: The case

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)

of the Laayoune Sakia El Hamra region in Morocco, DataBrief. 32, 106172.

BPS (2024) Indonesian Sustainable Development Goals Indicators 2024. Available online, accessed January 22, 2025.

Cataldo, R., Corbisiero, F., Delle Cave, L., Grassia, M.G., Marino, M., Zavarrone, E. (2019) The Quality of Life in the Historic Centre of Naples: the use of PLS-PM Models to measure the Well-Being of the Citizens of Naples, Ital. Stud. Qual. Life, 111–125.

Cataldo, R., Crocetta, C., Grassia, M.G., Lauro, N.C., Marino, M., Voytsekhovska, V. (2021) Methodological PLS-PM framework for SDGs system, Soc. Indic. Res. 156, 701–723.

Cataldo, R., Cascella, C., Grassia, M.G., Lauro, C.N., Voytsekhovska, V. (2024) A PLS Hierarchical Path Modeling approach to analyze and address gender equality in the EU countries, Socio-Econ. Plan. Sci. 96, 102076.

Crocetta, C., Antonucci, L., Cataldo, R., Galasso, R., Grassia, M.G., Lauro, C.N., Marino, M. (2021) Higher-order PLS-PM approach for different types of constructs, Soc. Indic. Res. 154, 725–754.

de Morais Fernandes, G.H., do Valle Junior, R.F., de Melo Silva, M.M.A.P., Mendes, R.G., Fernandes, L.F.S., Fernandes, A.C.P., Pacheco, F.A.L., Pissarra, T.C.T., de Melo, M.C., Valera, C.A. (2023) A structural equation model of land capability and related drivers: Implications for land degradation forecasting in a tropical river basin, Land Degradation & Development 34(6): 1778–1794.

Henseler, J., Sarstedt, M. (2013) Goodness-of-fit indices for partial least squares path modeling, Comput. Stat. 28, 565–580.

Maia, R.V.D.L., Brigido, W.J.H., De Oliveira, J.M.P. (2024) Influencing factors on internal auditing maturity using PLS-PM, IEEE Access.

Pu, X., Cheng, Q. (2024) Unraveling the impacts of multiscale landscape patterns and socioeco nomic development on water quality: A case study of the National Sustainable Development Agenda Innovation Demonstration Zone in Lincang City, Southwest China, J. Hydrol. Reg. Stud. 51, 101660.

Rajani F., Boluk K.A. (2022) A critical commentary on the SDGs and the role of tourism, Tourism and Hospitality 3 (4): 855-860.

Sanchez, G. (2013) PLS Path Modeling with R, Trowchez Editions.

Sun, H., Shi, Y., Li, C., He, S., Bai, Y., Zhao, P., Qiu, D., Liu, J., He, S. (2024) Factors control ling the heavy metal ion activity in soil contaminated by microplastics with different mulch durations: Partial least squares path model, J. Hazard. Mater. 464, 133027.

Yang, Z., Liu, V., Lyu, C. (2024) Exploring social sharing value: Effects on customer attitudes and behaviors in restaurant livestreaming, Behavioral Sciences 14(7), 621.

Zafeer, H.M.I., Maqbool, S., Rong, Y., Maqbool, S. (2024) Mapping the relationship and influence of school internal factors with an eye towards students' science academic outcomes, Heliyon 10(19).

DOI: https://doi.org/ 10.22452/mjs.vol45sp2.12
Malaysian Journal of Science 45 (Special Issue 2): 114-122(November 2026)