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.

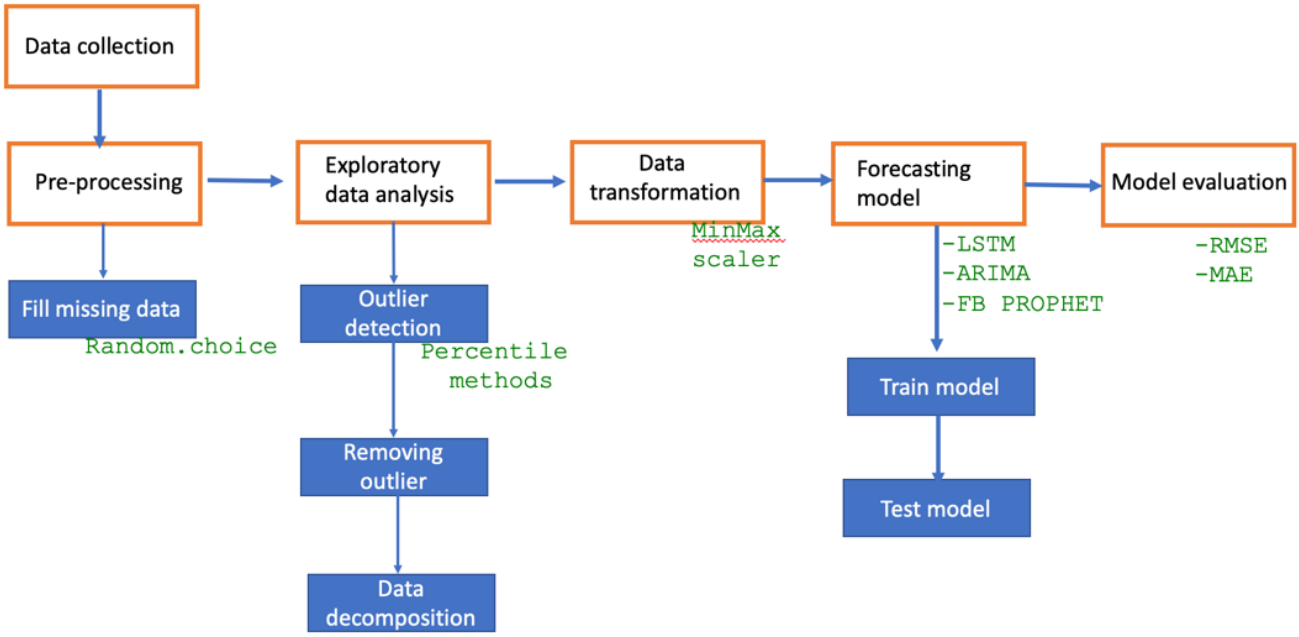


Figure 1. Model pipeline for flood forecasting using time series data of rainfall and streamflow.

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)

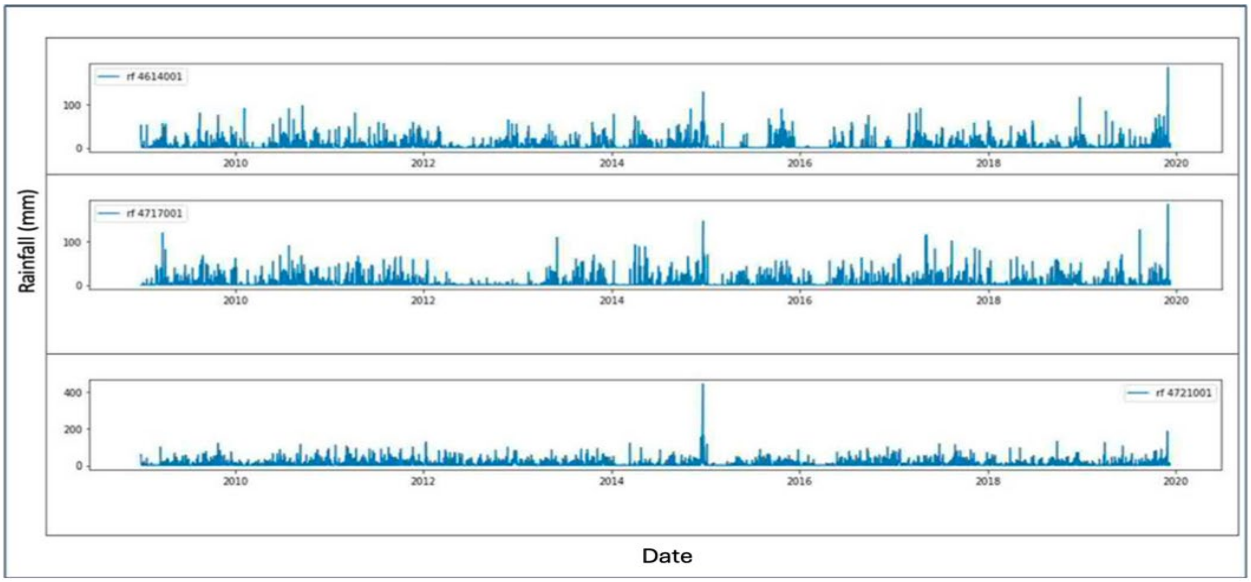


Figure 2. Rainfall intensity across three stations from 2009 to 2019.

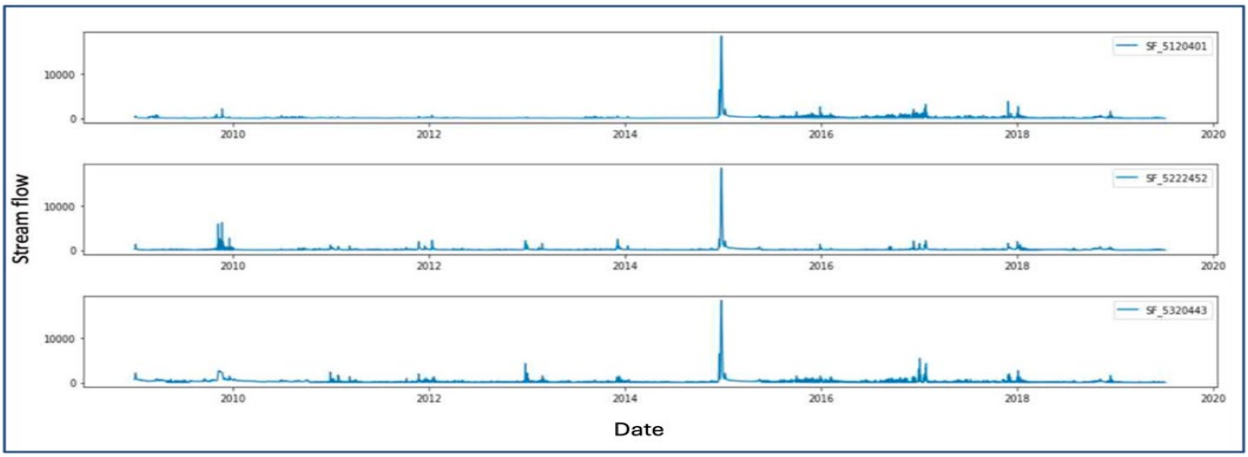
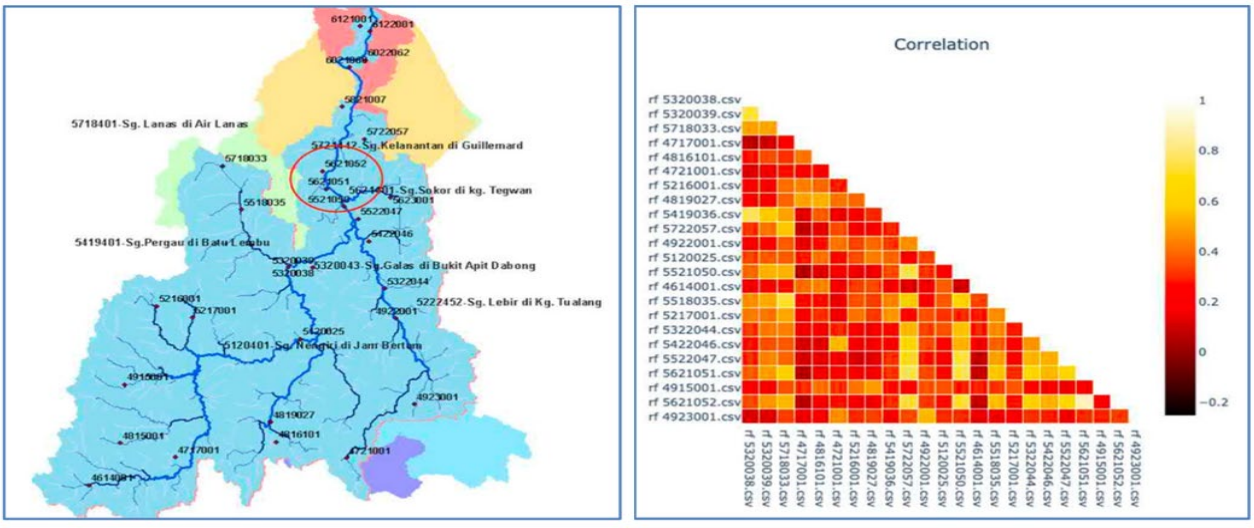


Figure 3. Stream flow pattern for 10 years of observation at 3 measuring stations from 2009 to 2019.



(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)

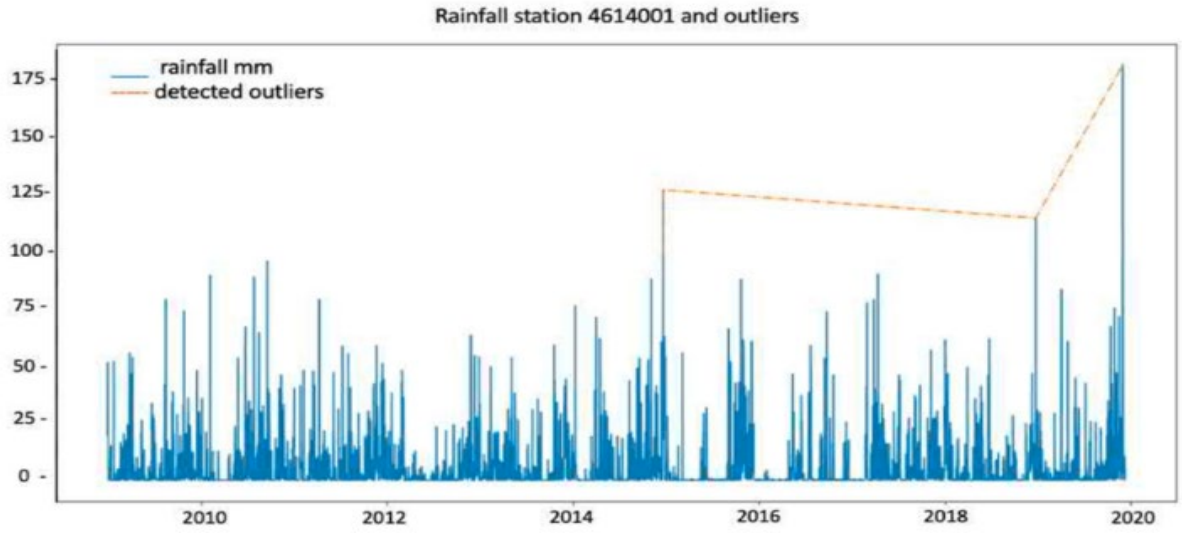


Figure 5. Graph representation of outlier detection.

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.

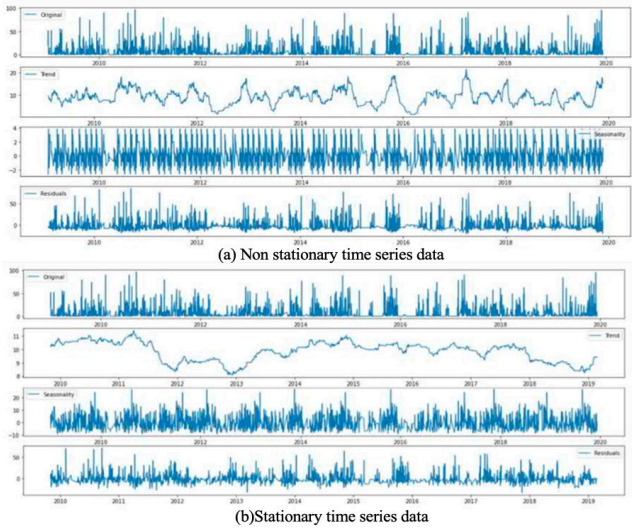


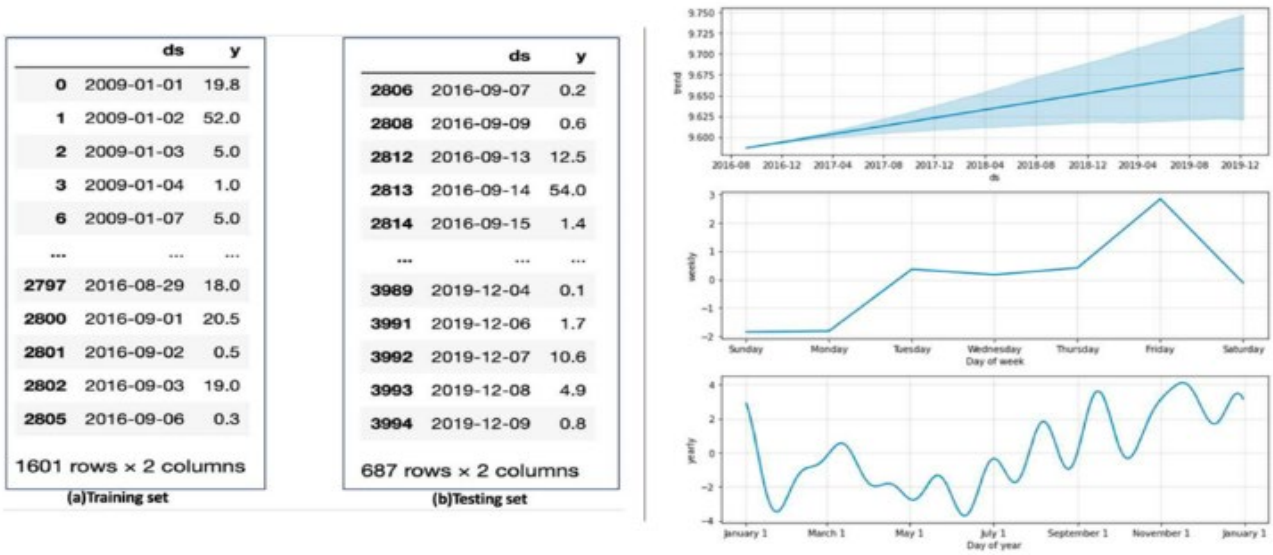
Figure 9. Transformation of (a)non-stationary time series data to (b)stationary time series.

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)

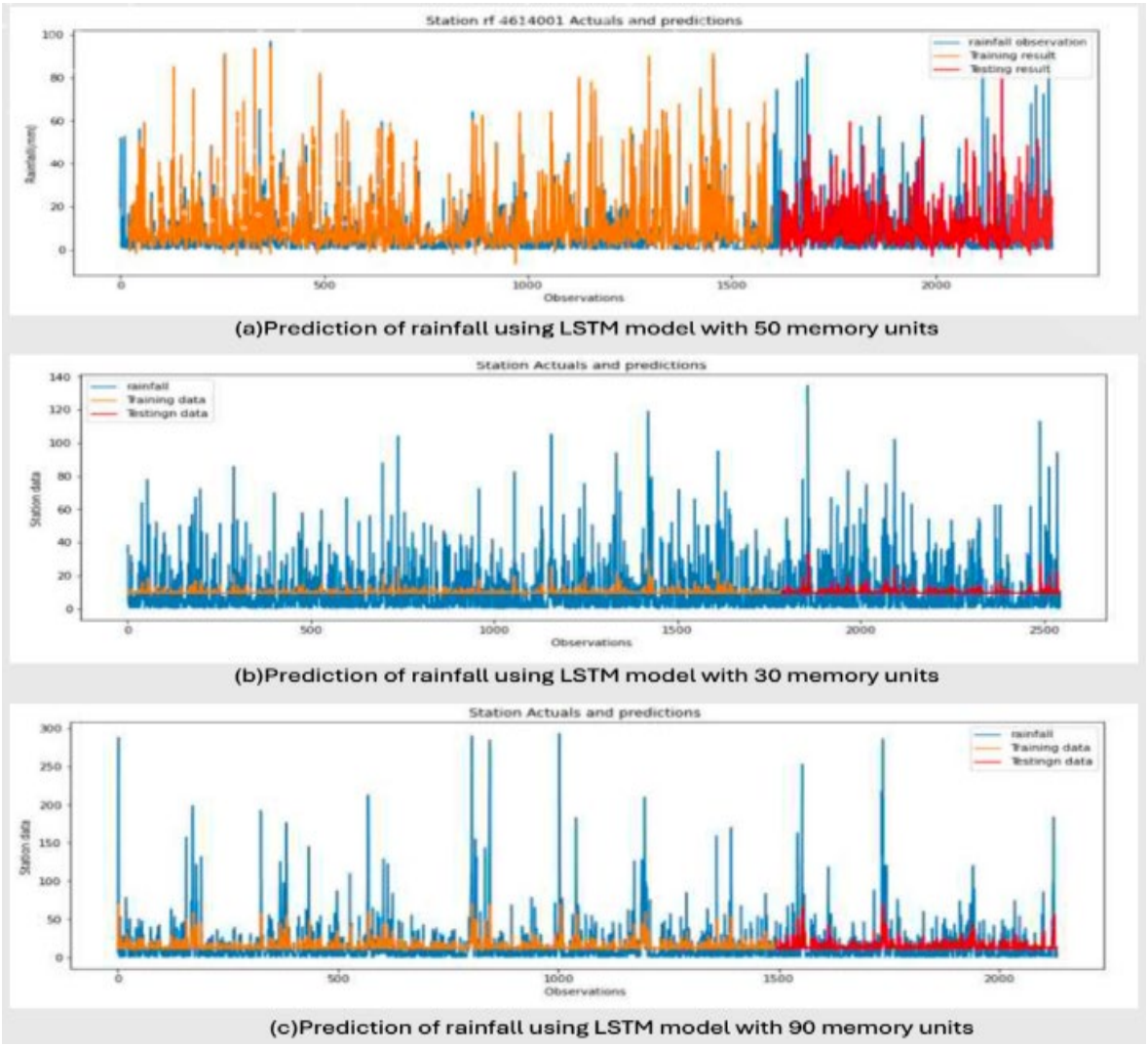


Figure 11. Prediction of rainfall using LSTM model with (a) 50 memory units, (b) 30 memory units, and (c) 90 memory units.

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)