MATEMATIKA, MJIAM, Volume 41, Number 3, 261–281

Forecasting Flash Floods in Puchong Using Data-Driven Machine Learning Techniques

Lim Xiao Wei, Muhammad Fadhil Marsani*, **Adila Aida Azahar, Vikneswari Someetheram and Mohd Shareduwan Mohd Kasihmuddin**

School of Mathematical Sciences, Universiti Sains Malaysia,
11800 USM, Penang, Malaysia.
*Corresponding author: fadhilmarsani@usm.my

Article history
Received: 27 December 2024
Received in revised form: 24 July 2024
Accepted: 24 September 2025
Published online: 15 November 2025

Abstract Flash floods are highly destructive in Malaysia. Mainly endangering lives, damaging infrastructure and disrupting the economy. Reliable and precise forecasting is essential to reduce their effects and enhance preparedness. This study investigates the application of Machine Learning (ML) algorithms such as Logistic Regression (LR), Random Forest (RF) and Long Short-Term Memory (LSTM) to develop a robust flash flood prediction model. Historical water level and rainfall data recorded every 15 minutes during the flash flood at IOI Puchong Jaya on December 15−16 2023 were utilized for model training. Lag features were created to capture temporal dependencies and enhance predictive accuracy. Among the tested models, RF outperformed other models by achieving an accuracy of 0.50 and a perfect recall of 1.00 successfully detecting all flash flood events with an impressive computation time of just 0.01 seconds. Further real-world validation during the May 2024 flash floods demonstrated RF's reliability in predicting actual flood occurrences while minimizing false alarms. The findings underscore the potential of RF as a powerful tool for flash flood prediction. Its integration into flash flood warning systems can provide emergency responders with timely and accurate predictions. This research highlights the significance of ML-driven approaches in addressing the challenges of flash flood management and offers valuable insights for future disaster mitigation strategies.

Keywords Machine Learning, Time series, Predictive Models, Statistical Method, Flood Forecasting.

Mathematics Subject Classification 62-07, 60G25, 62M20.

1 Introduction

Flooding is a recurring natural disaster in Malaysia that categorized into two primary types: flash floods and monsoon floods [1]. These events significantly disrupt lives, infrastructure and the economy that makes flood management a critical concern. In contrast, monsoon floods driven by prolonged heavy rainfall during the northeast monsoon season (November to March) that can last for weeks to months [2]. Both types of floods cause widespread devastation with flash floods being more common in urban areas and monsoon floods predominantly affecting rural regions in states like Kelantan, Terengganu and Pahang [3]. The distinct landscape and weather patterns of Malaysia increase its risk of floods. The tropical climate with its consistent high rainfall coupled with rapid urbanization has increased the risk of flash floods. Sudden intense rainfall in urban areas overwhelms drainage systems leading to water accumulation and flash floods [1]. Sanyal and Lu [4] describe flash floods as resulting from heavy rainfall over a short duration causing high river discharge and localized flooding. Urbanization, deforestation and poorly planned infrastructure developments have reduced natural water absorption and increased surface runoff and worsening the severity of floods [5]. Monsoon floods predominantly impact rural areas along major rivers, such as the Kelantan and Pahang rivers with devastating effects on agricultural activities and the livelihood of affected communities. Conversely, flash floods frequently occur in rapidly developing urban areas including Kuala Lumpur and Selangor. The rapid urbanization of floodplains particularly in Selangor has heightened the vulnerability of these areas to hazardous flash floods making effective flood management strategies imperative [6]. In response to these challenges, Malaysia has implemented various mitigation measures including the construction of flood retention ponds, river deepening projects and early warning systems.

Machine Learning (ML) has become a robust option for flood prediction that overcomes the challenges associated with traditional hydrological models. Unlike traditional model which often require a detailed understanding of complex physical processes. ML methods rely solely on historical data to capture the nonlinear relationships between variables [7–9]. This capability allows ML models to achieve higher predictive accuracy and adapt to diverse flood scenarios effectively. In recent years, ML-based approaches have attained significant prominence in flood prediction studies, particularly in Malaysia and on a global scale [10–14]. Researchers have successfully applied a variety of ML algorithms including Logistic Regression, Neural Networks, Random Forests, and Long Short-Term Memory networks to enhance prediction accuracy and decision-making. By leveraging advanced ML techniques, flood prediction models can provide timely and precise forecasts that offers critical insights to emergency responders. This study builds on these advancements exploring the application of ML methods to improve flash flood prediction and management strategies.

Flood prediction is a critical aspect of disaster management and various ML algorithms have been implemented to advance forecasting accuracy. Logistic Regression (LR) is a frequently applied approach in spatial studies for predicting floods due to its flexibility and ability to accommodate multiple data types, including scale, nominal, and categorical variables. Unlike many traditional statistical methods, LR does not require strict assumptions, making it a versatile tool in flood studies [15]. Tehrany *et al.* [16] demonstrated LR's ability to assimilate meteorological and hydrological data for future flood forecasting. However, while Lopez and Rodriguez [17] found LR effective in predicting non-occurrence of flash floods, its performance in predicting occurrences was only acceptable. In comparative analyses, LR has consistently shown competitive results, outperforming models like Decision Tree (DT) and Support Vector Classification (SVC) in recall accuracy [18] and performing on par with K-nearest neighbors (KNN) models [19].

Random Forest (RF) has emerged as another robust ML algorithm in hydrology known for its ensemble approach that combines multiple decision trees to mitigate overfitting while maintaining predictive precision [20]. RF offers high accuracy, rapid computational performance and resilience against missing data and noise that makes it ideal for flood prediction [21]. Studies have shown RF to outperform Artificial Neural Networks (ANN), Support Vector Machines (SVM), and regression models in various hydrological applications [19, 20]. Despite its strengths, RF's complexity and the need for extensive parameter tuning pose challenges in interpretability and training time [24].

Long Short-Term Memory (LSTM) networks have revolutionized flood prediction, particularly in time-series analysis. As an evolution of Recurrent Neural Networks (RNN), LSTM overcomes the vanishing gradient problem and captures complex temporal and nonlinear patterns effectively [25]. Its ability to handle inputs and outputs of varying lengths makes it particularly suited for dynamic flood prediction scenarios [26]. Studies have consistently highlighted LSTM's superior performance over traditional ML models and statistical techniques, such as ARIMA, in capturing time-series data patterns for flood forecasting [27, 28]. For instance, [29] found LSTM to be the most effective method in identifying and predicting flood dynamics compared to RF, SVM, and basic RNNs. This study leverages the unique strengths of LR, RF, and LSTM models aiming to enhance flash flood prediction by combining their respective capabilities. By addressing the limitations and harnessing the advantages of these algorithms, this research seeks to design a strong and accurate method for predicting floods framework to aid disaster management and mitigation efforts.

The research aims to evaluate the viability of predictive models for accurate and reliable flash flood prediction, emphasizing the identification of the most suitable model with high accuracy and recall scores. Besides, the research also applies and assesses predictive models, including Logistic Regression, Random Forest, and long Short-Term Memory, for predicting flash flood occurrences using water level and rainfall data. Besides, accurate predictive models will be developed with the aims at enhancing early warning systems, supporting informed decision-making, and mitigating risks to lives and infrastructure.

2 Methodology

2.1 Study Area

The core aspect of the research was carried out at Puchong Jaya City in Selangor, Malaysia. Bandar Puchong Jaya, a highly populated township in Puchong, Selangor, Malaysia. It is a significant economic hub with a large population, well-established residential neighbourhoods, and substantial industrial activity. IOI Puchong Jaya is a bustling part of Bandar Puchong Jaya in Selangor, Malaysia. It is a commercial and entertainment hub within the township. Despite rapid development, IOI Puchong Jaya faces occasional flash flood issues. Flash floods have occurred multiple times here, with documented incidents from 2006, including the most recent on May 12, 2024. The possible reasons for flash flood occurrence are poor drainage system, low elevation of the IOI Mall area, and water overflow due to extreme rainfall [30]. This analysis employed hydrological data, such as rainfall and water level, obtained from gauging stations managed by the Department of Irrigation and Drainage Malaysia (DID). The hydrologic gauging stations involved in this study are Station Sg. Klang At USJ 1 that provides water level data in meters (m), and Station IOI Puchong Jaya that provides rainfall data in millimeters (mm). Table 1 lists the gauge station used in data collection, and the location of the stations is plotted in figure 1. The Station Sg. Klang At USJ 1 and Station IOI, Puchong has an elevation of 10m and 27m, respectively. The average elevation of the IOI Puchong area has a topography ranging from 10 to 50 metres illustrated in figure 2. With an average elevation of just 38 meters, this area is susceptible to floods triggered by rising river levels during heavy rain. The Klang River and its drainage infrastructure lack the capacity to manage surplus water, as evidenced by the 16 December 2023 flood incident.

The hydrological data collection focused specifically on the periods surrounding a flash flood event on 16 December 2023. Data were collected for the day before and the day of the flash flood occurrence. Around 192 observations were collected over two days at 15-minute intervals. On 16 December 2023, Subang Jaya City Council's (MBSJ) Disaster Management Unit announced a level-three danger warning for the IOI Puchong mall area at 6.45 p.m. [30]. Thus, the data points from 6.45 p.m. until 8.15 p.m. on 16 December 2023 were labelled as '1' to indicate the flash flood occurrence. Before further analysis, the data is pre-processed to ensure that it is clean and does not contain any null value.

Table 1: The list of gauge stations used in data collection.

No.	Station Name	Latitude	Longitude	Station Type	Open Date	State/Basin/District
1	BPTL0033 - Sg. Klang At USJ 1	3.050117	101.610108	Water Level	30/11/2023	Selangor/Sungai Klang/Petaling
2	BPTL0005 - IOI Puchong Jaya	3.04275	101.621417	Rainfall	9/5/2007	Selangor/Sungai Klang/Petaling

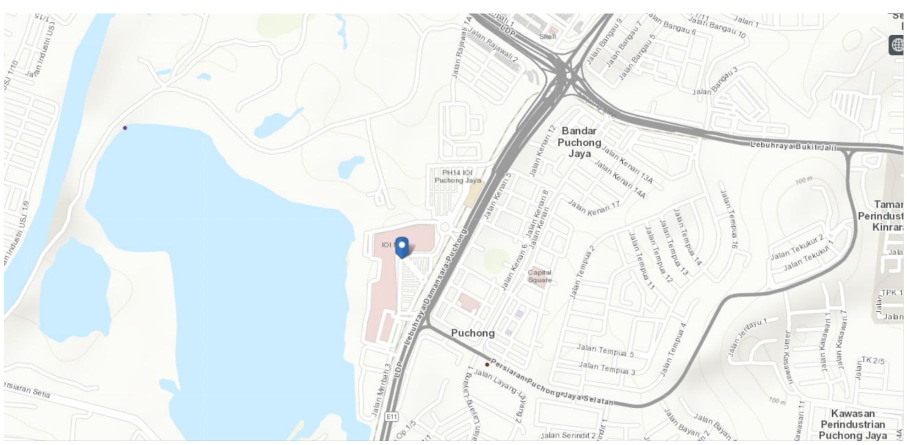

Figure 1: The location of the gauge stations used in data collection.

Figure 2: The elevation map of IOI Puchong Jaya Area.

2.2 Logistics Regression

Logistics regression analysis is a regression method that establishes the relationship between continuous or categorical explanatory (independent) variables and categorical response (dependent) variables. Logistic regression is a binary classification algorithm in which the response variable is a binary variables (0/1) where 1 denotes the presence of a given scenario (success), and 0 denotes the absence (failure). In the study, the logistics regression is used to predict the flash flood occurrence where 1 denote flash flood occur and 0 is not. The idealized linear logistic regression model is estimated using (1).

$$P_f = \frac{1}{1+e^{-Z}}; Z = \rho_0 + \rho_{1X_1} + \rho_{2X_2} + LL + \rho_{nX_n}, \tag{1}$$

where P_f, is the probability of flash flood occurrence, Z is the value of determined from $-\infty$ to $+\infty$, ρ_0 is the intercept of the LR model, the $\rho_i(i = 0, 1, 2, ..., n)$ are the slope coefficients of the LR model, and $X_i(i = 0, 1, 2, ..., n)$ are the independent variables. The prediction process of flash flood occurrence using LR involves the probability calculated. If the probability is higher than 0.5, it identifies that the flash flood will occur at that timepoint (label= '1'), and if it's 0.5 or lower, its determined that no flash flood is likely to occur (label= '0'). The architecture of LR is demonstrated in Figure 3.

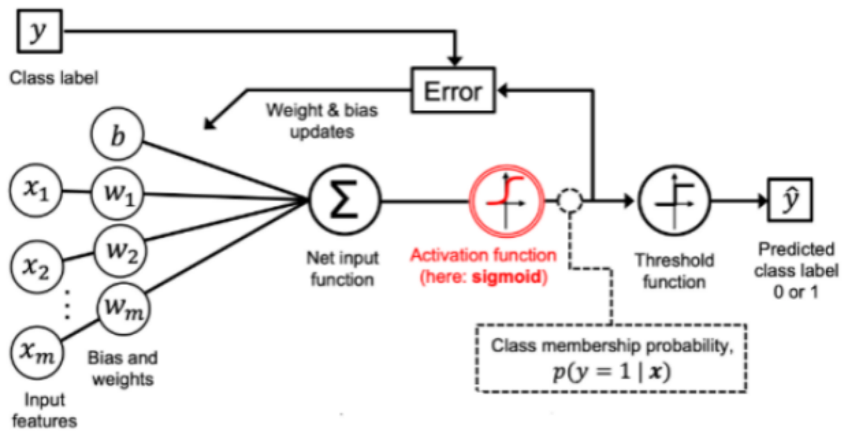

Figure 3: Architecture of Binary Logistics Regression model

2.3 Random Forest

The Random Forest (RF) technique is a well-established method for supervised classification that is applicable to both classification and regression issues. RF is an ensemble learning method combining bootstrapping principles and several decision trees on various subsets of the given dataset [31]. A bootstrapping or bagging technique ensures that each tree is trained using randomly resampled training sets from the original dataset. The technique introduces the randomness in the model and assures the decorrelation among trees. In each decision tree, the bootstrapped dataset is recursively split until a leaf node (the bottom of decision trees) is left, and it finds the best split by maximizing information gain. RF also employs random feature selection. In training each decision tree, a portion of predictor variables is utilized to partition an internal node following splitting guidelines. The standard splitting criteria in classification problems are entropy and information gain [32]. At each internal node of the decision tree, entropy (Equation 2) and information gain (Equation 3) are calculated. The split with lower entropy or higher information gain is selected and stop when achieving a homogeneous node.

$$Entropy = \sum_{i=1}^{n} p_i \log_2 p_i, \tag{2}$$

$$\text{Information Grain=1-Entropy}, \tag{3}$$

where n is the number of unique classes and pi is the prior probability of each given class. In the classification problem, each decision tree in Random Forest predicts the class independently. The final decision or prediction result is decided by a majority vote from all the individual trees. The conceptual scheme of RF model for classification process is demonstrated in Figure 4.

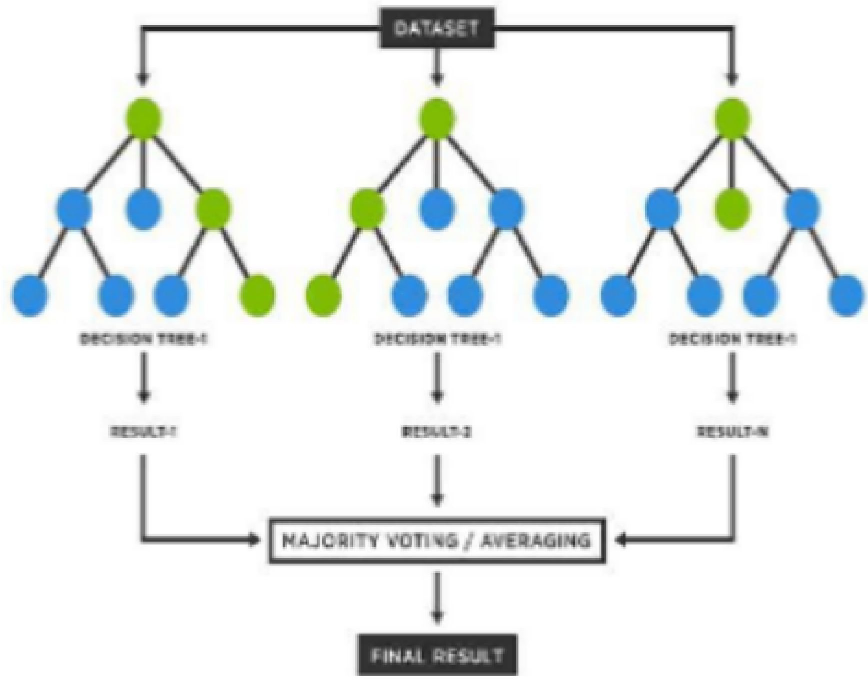

Figure 4: Architecture of Random Forest

2.4 Long Short-Term Memory (LSTM)

The Long-Short-Term Memory (LSTM) is an enhanced version of the recurrent neural network (RNN). It addresses the challenge of vanishing gradients encountered in traditional RNNs and effectively captures long-term dependency information. The structure of the LSTM model is in the form of a chain. The key element of the LSTM is a memory cell that can retain its state throughout time. LSTM is equipped with three gates for managing the cell state. The input gate permits or blocks the input signal to modify the memory cell state. The output gate regulates the transmission of the memory cell state to the output. The forget gate manages how much information the memory cell retains or discards for its next state. Figure 5 shows the internal structure of the LSTM network.

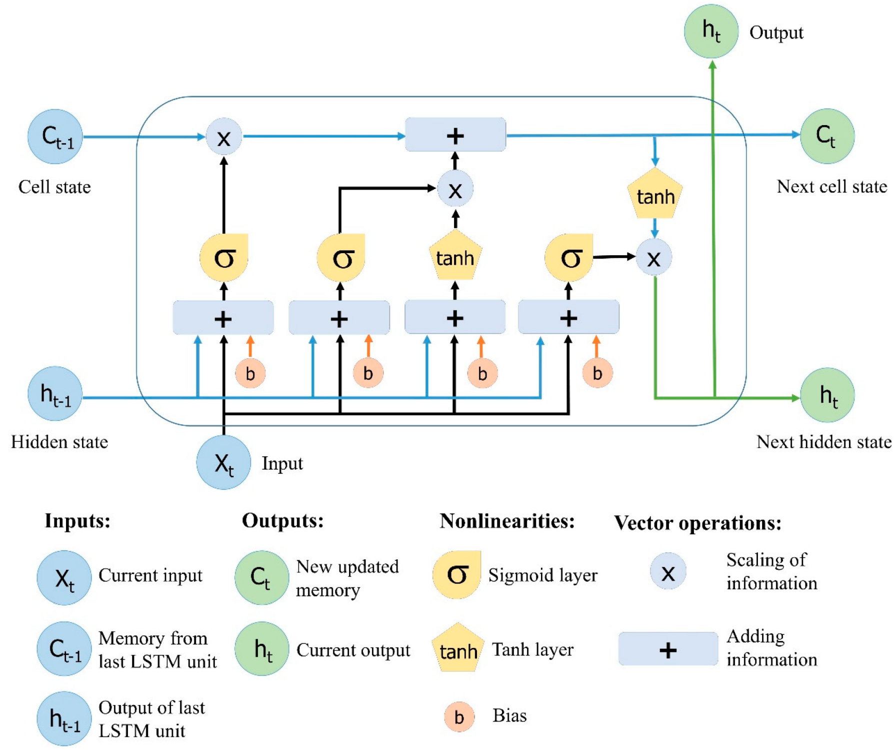

Figure 5: Internal Structure of LSTM Network.

The LSTM network begins by recognizing and removing unneeded information from the cell with the forget gate, which is controlled by the sigmoid function [33]. This function evaluates the previous output (h_{t-1}) and current input (X_t) to decide which parts of the prior output should be discarded. The forget gate (f_t) is a vector with values between 0 and 1, representing each element in the cell state (c_{t-1}).

$$f_t = \sigma(W_f[h_{t-1,X_t}] + b_f), \tag{4}$$

where σ is the sigmoid function, and W_f and bf are the weight matrices and bias, respectively, of the forget gate.

Next, the sigmoid and tanh layers decide which information from the new input (X_t) should be stored in the cell state and update it accordingly. The sigmoid layer assigns a value of 0 or 1 to update or ignore new information, while the tanh layer weights the values on a scale from -1 to 1. These values are multiplied to update the cell stat. The new memory is combined with the old memory (C_{t-1}) to form the new cell state (C_t).

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i), \tag{5}$$

$$N_t = \tanh(W_n[h_{t-1}, X_t] + b_n), \tag{6}$$

$$C_t = C_{t-1}f_t + N_t i_t. \tag{7}$$

C_{t-1} and C_t are the cell states at time $t-1$ and t, while W and b are the weight matrices and bias, respectively, of the cell state. Then, the sigmoid layer determines which parts of the cell state are included in the output. The output from the sigmoid gate (O_t) is multiplied by the new values generated by the tanh layer from the cell state (C_t), which range from -1 to 1. Finally, the output values (h_t) are derived from the output cell state (O_t) in a filtered form.

$$O_t = \sigma(W_o[h_{t-1}, X_t] + b_0), \tag{8}$$

$$h_t = O_t \tanh(C_t). \tag{9}$$

The architecture of LSTM network mainly involves one LSTM layer, one fully connected layer with dropout rate=0.2 and output layer with dense. The input of model is time lagged rainfall and water level while the output is predicted probability (PP) which take value between zero and 1. A threshold of 0.5 is used, such that the non-flash flood event when PP is equal to or lower than 0.5, and PP greater than 0.5 indicates a flash flood occurred. In the LSTM layer, the number of hidden units of LSTM layer is 200 and the activation function is Rectified Linear Unit (ReLU). The activation function in dense layer is sigmoid where it squashes the output to a probability between 0 and 1, indicating the likelihood of the positive class. The LSTM also utilizes Binary Cross-Entropy as the loss function, ideal for binary classification tasks, and employs Adam, an adaptive learning rate optimization algorithm recognized for its efficiency and effectiveness. Figure 6 shows the proposed methodology of the study.

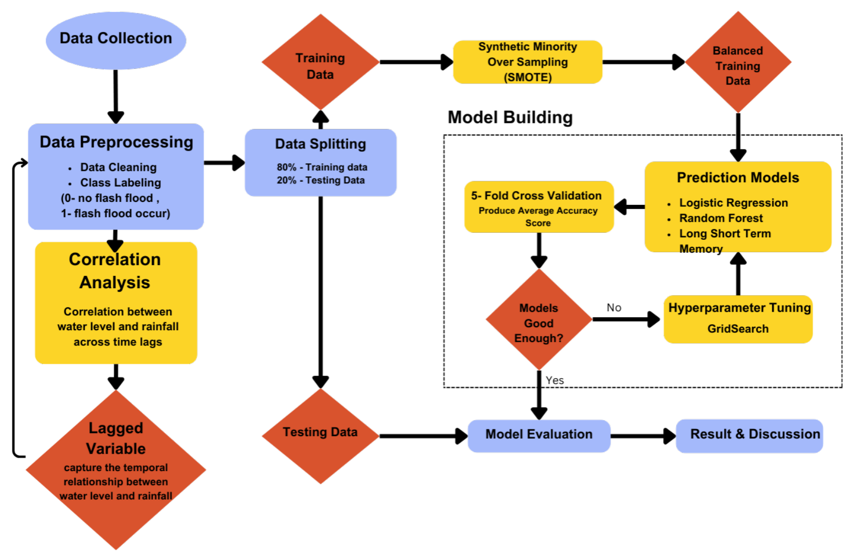

Figure 6: Proposed methodology of the present study.

In this study, the models are built with the default hyperparameters and built again using the optimal combination of hyperparameters obtained from hyperparameter tuning. The list of hyperparameters for each model is listed in Table 2. The optimal combination of hyperparameters is determined based on the highest accuracy by using K Fold Cross Validation.

Table 2: List of hyperparameters for each model

Model	Default hyperparameters	Hyperparameters Tuning
Logistic Regression	C=1.0 solver='lbfgs'	C: [0.01, 0.1, 1, 10, 100], solver: ['lbfgs', 'liblinear']
Random Forest	n_estimators=100 random_state=13 max_depth=0	n_estimators: [50, 100, 200] random_state:[13] max_depth: [4, 6, 10, 12]
LSTM	Batch_size=32 Epochs=50	batch_size:[32, 64,128] epochs: [5, 10, 20]

3 Performance Metrices

This part of the study focuses on the evaluation metrics used to evaluate the performance of classification models. Non-flash floods are categorized as negative instances 0, and flash floods are classified as positive instances 1. Accuracy represents the proportion of correctly predicted cases out of all examined cases. An increased accuracy score demonstrates improved prediction performance. The formula for accuracy can be found in the next equation (10).

$$\text{Accuracy} = \frac{\text{True of Correct Prediction}}{\text{Total Number of Predictions}}. \tag{10}$$

Recall is the proportion of actual positive instances that are correctly predicted by the model (Equation 11). It measures the accuracy of model in identifying positive cases. Higher recall indicates fewer false negatives, which is important when the cost of missing positive instances is high. The recall of the study is also known as the hit rate (HR) in the flash flood warning field, according to Zhao *et al.* [34]. The formula for specificity used in the study is shown in Equation 11.

$$\begin{aligned} \text{Recall} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}}, \\ &= \text{Hit Rate}, \\ &= \frac{\text{Hits}}{\text{Hits} + \text{Misses}}. \end{aligned} \tag{11}$$

where "Hits" is the number of observed flash flood events predicted correctly and "Misses" is the number of observed flash flood events predicted incorrectly. Recall is important in evaluating the performance of flash flood prediction models because flash floods are the key event in the study, and the cost of mispredicting flash floods is high.

F1 score in (12) is the harmonic mean of precision and recall, providing a single metric that balances both measures. High F1 Score indicates high recall and high precision.

$$\textbf{F1 Score} = \mathbf{2} \cdot \frac{\text{Precision Recall} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{12}$$

Precision measures the proportion of correctly predicted flash flood events among all the instances predicted as flash flood events (Equation 13). It focuses on the accuracy of positive predictions made by the model. Precision is important when the cost of incorrectly predicting an event is high. The greater the precision value, the lower the false alarm rate, indicating that fewer false alarms are generated.

$$\text{Precision} = \frac{\text{True Posistives}}{\text{True Positives} + \text{False Positives}}. \tag{13}$$

4 Results & Discussion

In this study, the flash floods occurrence in the IOI Puchong Area is predicted using three machine learning models: Logistic Regression, Random Forest, and Long Short-Term Network. The evaluation metrics, including accuracy, precision, recall, and F1-score were calculated to assess the performance of each model. The machine learning models were implemented using Python on Jupyter Notebook. This section presents a comparative analysis of the prediction capabilities of these models, offering insights into their effectiveness and applicability in enhancing flood prediction in the IOI Puchong Jaya Area.

4.1 Time-lagged correlations between rainfall and water level

Two days of records of rainfall and water level at the gauge stations nearest to IOI Puchong were plotted and analyzed in Figure 7. There were 2 rain events on 15 December and 16 December, respectively. Observing the peak of rainfall and of the water levels, there was a lag between the two peaks. According to the news article from Sinar Harian, the flash flood on 16 December 2023 was alerted around 6:45 p.m., which was 15 minutes after the gauge station recorded the highest rainfall value of 30 mm over the two days, as shown in Figure 8 [30]. Besides, it was observed that the water level did not increase immediately following the peak rainfall at 6:30 p.m. Instead, the highest water level was observed at 7:30 p.m., indicating a one-hour lag after the peak rainfall event. These observations suggest a time lag between the rainfall event and the hydrological response, characterized by the increase in water level.

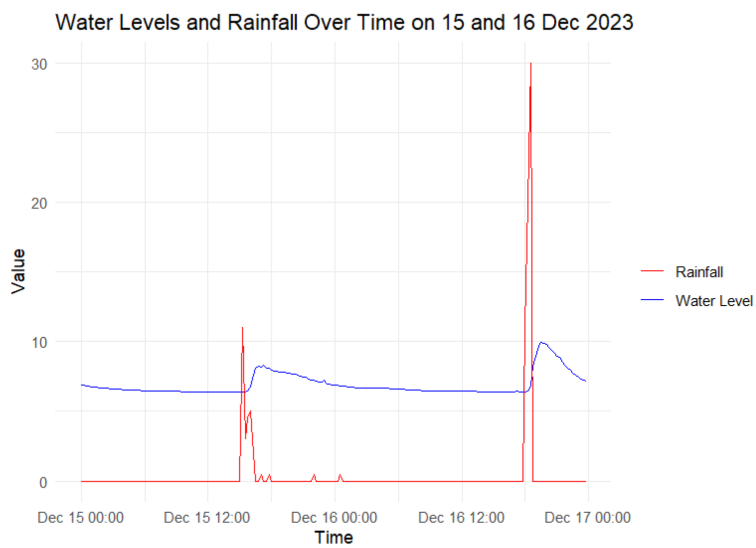

Figure 7: Rainfall and water level data for IOI Puchong area on 15 December and 16 December 2023

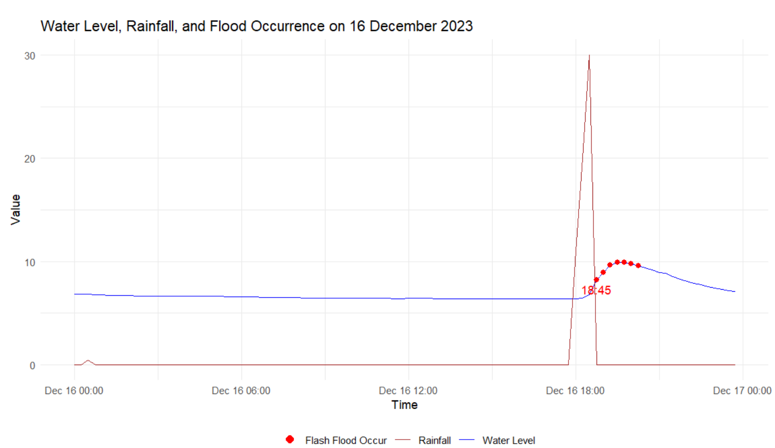

Figure 8: Rainfall and water level data for IOI Puchong area on 16 December 2023

4.2 Correlation Analysis

The general pattern in Figure 9 shows a sharp increase in correlation leading up to the peak, followed by a gradual decline. This pattern is consistent with the expected hydrological response where the water level rises after rainfall and then gradually decreases as the water drains away or evaporates. The highest correlation is observed at a 5th time lag, which corresponds to 75 minutes (1 hour and 15 minutes). This indicates that the water level is most strongly affected by rainfall that occurred 75 minutes earlier. After the peak at 5th time lag, the correlation decreases gradually. This suggests that the impact of rainfall on water levels diminishes over time. Figure 10 examines the correlation between water level and rainfall on 16 December 2023 across time lags. The highest correlation is observed at a 5th time lag, which corresponds to 75 minutes. This is consistent with the earlier overall dataset analysis, reinforcing the conclusion that the water level responds most strongly to rainfall 75 minutes prior. After reaching the peak, the correlation slightly declines but remains relatively high up to 8th time lag (120 minutes). This suggests that the impact of rainfall on water level persists but weakens after the peak response time. Both plots showed that the time lag in hydrological response to rainfall, with the highest impact observed around 75 minutes after the rainfall event. This information is crucial for flood prediction and early warning systems, as it helps in predicting water level rises following significant rainfall.

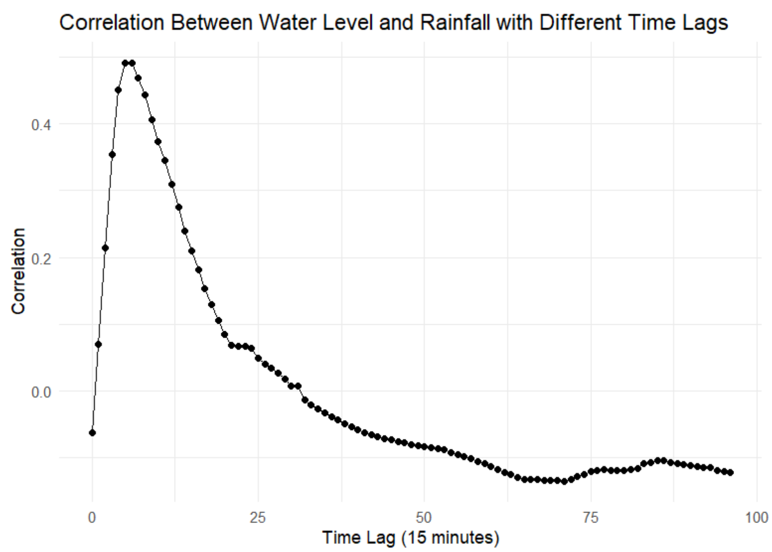

Figure 9: Correlation Between Water Level and Rainfall with Different Time Lag

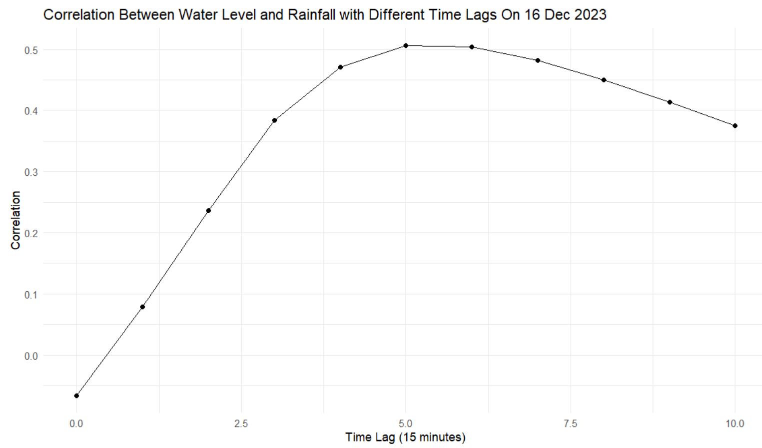

Figure 10: Correlation Between Water Level and Rainfall with Different Time Lags on 16 December 2023 Particularly

A correlation analysis revealed that the correlation between water level and rainfall increases from the 0th lag and reaches its maximum at the 5th lag. Several lag features were created to account for this lag effect in the model, each representing cumulative rainfall over different time windows. For instance, rain_lag_1 represents the cumulative rainfall for the current time and the previous 15 minutes; rain_lag_2 represents the cumulative rainfall for the current day and the previous 30 minutes; similar terms are created up to the 75 minutes. The lagged variables were then used as input variables to build the three models, Logistic Regression (LR), Random Forest (RF) and Long Short-Term Memory (LSTM) models.

4.3 Model Building

Three machine learning models, Logistic Regression (LR), Random Forest (RF) and Long Short-Term Memory models were developed to predict the flash flood occurrence. The models were built using default parameters using the dataset that had included the lagged features. Their performance with default parameters was evaluated. A hyperparameter tuning was carried out for each model to improve the predictive performance. For each model, the tuned model with the higher accuracy was selected and used for training and validation. The chosen model was then tested and evaluated using various performance metrics.

4.3.1 Using Default Hyperparameters

The models were built using the default hyperparameters and went through 5- fold cross-validation to assess their performance in predictive capability. Table 3 shows the average scores of each model during 5-fold cross-validation.

Based on the results in Table 3, all models generally demonstrated high accuracy. The LSTM model achieved slightly higher accuracy than the LR and RF models. Additionally, LSTM showed the highest recall, precision, and F1 scores, followed by LR. RF performed relatively poorly with the lowest recall and very low precision scores, indicating weaknesses in correctly predicting observed flash flood events and a tendency to generate false alarms.

Table 3: Performance of LR, RF and LSTM models using default hyperparameters

Model	Average			
	Accuracy	Recall	F1 Score	Precision
LR	0.97	0.80	0.67	0.60
RF	0.97	0.60	0.43	0.33
LSTM	0.98	0.88	0.91	0.99

4.3.2 Hyperparameter Tuning

This section presents the result after hyperparameter tuning using Grid Seach strategy for each model, with the accuracy calculated respectively. Table 4 shows the set of the best hyperparameters that represent the best model configuration obtained with the best accuracy for LR, RF and LSTM.

Table 4: Performance of LR, RF and LSTM models using tuned hyperparameters

Model	Accuracy Score	Best parameters	
RF	0.97	Max Depth	6
		Number of Estimators	50
LR	0.99	Strength of the regularization, C	10
		Solver	Lbfgs
LSTM	0.98	Batch Size	32
		Epochs	20

4.4 Training Result

The flood prediction models are built with logistic regression, random forest and long short-term memory algorithm with optimal hyperparameters. The models were trained with oversampled data from Synthetic Minority Oversampling Technique (SMOTE) and, then validated using 5-fold cross-validation techniques. The average score for each performance metric is tabulated in Table 5.

Table 5: Performance of LR, RF and LSTM models before and after Hyperparameters Tuning

Model	Average							
	Before Tuning				After Tuning			
	Accuracy	Recall	F1 Score	Precision	Accuracy	Recall	F1 Score	Precision
LR	0.97	0.80	0.67	0.60	0.99	0.80	0.73	0.70
RF	0.97	0.60	0.43	0.33	0.97	0.60	0.49	0.43
LSTM	0.98	0.88	0.91	0.99	0.98	1.0	0.98	0.95

After tuning, there were no changes in the accuracy scores for each model. LSTM improved its recall score to 1, while the recall scores for the other two models remained unchanged from before tuning. Additionally, RF and LR increased their precision scores by about 10% each, indicating an enhanced ability to avoid false alarms. Although the precision of LSTM decreased after tuning, it remained high at 0.95, the highest among the three models. The improvements in LSTM's recall score and in LR and RF's precision scores were reflected in the increased F1 scores of all three models.

4.5 Model Evaluation

This section will show the result of model testing after training. This section focuses on identifying the capability of machine learning models in forecasting flash floods occurrence [35]. Moreover, to identify the effective model for this critical predicting task among three models. Table 6 highlights the results of the three models evaluated across multiple metrics.

All three types of models, LR, RF and LSTM achieved an accuracy greater than 90%. However, given that most of the dataset consists of negative samples (cases where flash floods did not occur), and our primary focus is on detecting flash flood occurrences, the accuracy of detecting non-flash flood events is less relevant. Therefore, precision and recall, which measure the detection of positive samples, are more crucial. It is observed that all 3 models had the same performance in testing phase. They were all high in recall value but only achieving precision of 0.5. These 3 models had ability to predict the flash flood events correctly, but it also had 50% probability in giving false alarm about flash flood events.

As the main purpose of detecting flash floods, where timely and accurate alerts can save lives and properties, the models are further compared based on their computation speed. RF carried out and completed the classification task the fastest. With only 0.01 second difference, LR is behind RF model. While, considering the complex mechanism of LSTM, it was not surprised that LSTM took the longest computational time needed.

Table 6: Evaluation metrics of LR, RF and LSTM models

Model	TP	FN	TN	FP	Accuracy	Recall	F1 Score	Precision	Computational Time (s)
LR	1	0	36	1	0.97	1	0.67	0.5	0.02
RF	1	0	36	1	0.97	1	0.67	0.5	0.01
LSTM	1	0	36	1	0.97	1	0.67	0.5	3.27

Having high precision is crucial in minimizing unnecessary alerts and resource allocation, particularly when the cost of false alarms is high. On the other hand, high recall is important when missing a flash flood event (false negative) is dangerous. It ensures that the model identifies as many actual flash floods as possible. Even if it may raise the false alerts., recall is considered more vital. This is because missing a flash flood event could cause severe consequences, whereas false alarms, though undesirable, might be more manageable. Considering the criteria and the computational speed, RF stands out from another 2 models as it performs as good as other models but with the fastest speed.

4.6 Prediction of Flash Flood Occurrence on 12 May 2024

A flash flood once again hit the IOI Puchong Jaya Area on 12 May 2024. An examination of the three models' performance is provided in this section from Section 4.5 in predicting flash flood occurrence using historical water levels and rainfall on 11 and 12 May 2024. The data source is the same as the dataset used for building the model. The lagged rainfall variable was created using the approach in Section 4.1, assuming the optimal time lags are from 1st to 5th lags. The results were summarized in Table 7.

Table 7: Comparison of Three Models on Predicting Flash Flood Occurrence on 12 May 2024

Model	TP	FN	TN	FP	Accuracy	F1 Score	Precision	Recall
LR	6	0	166	15	0.91	0.44	0.29	1
RF	6	0	167	14	0.93	0.46	0.3	1
LSTM	6	0	165	16	0.91	0.43	0.27	1

Generally, all models achieved perfect recall score (1) and high accuracy as in testing. It indicated that these 3 models can predict actual flash flood events very well without missing any. The result proved that these 3 models are viable in predicting flash flood events and giving early warning. However, these models all had low precision values. Based on Table 7, LSTM had the lowest precision (0.27), indicating it was the most likely to produce a false alarm, as compared to other models. While RF has a slightly higher precision score. As compared with their performance in Section 4.2, overall, the magnitude evaluation metrics dropped down slightly, except recall.

In selecting the best model for predicting flash flood occurrence, there are a few criteria that need to be considered. Firstly, it's noteworthy that none of the models missed positive instances (flash flood events). The models are proved good in predicting true flood events. Besides, a false alarm may would like to be prevented by relevant authorities. A false alarm of flash flood occurrence not only incur immediate economic and social costs but can also have long-term implications on public safety, community resilience, and trust in disaster management systems. RF has the highest precision, and it predicted fewer number of false positives, more accurate in giving flash flood warning and less false alert. Incorporating the result learned from section 4.3, RF executed the classification task faster. Consequently, an accurate and fast flash flood alert can be given. On the other hand, RF model has the highest F1 score, often achieve a better balance between precision and recall. This balance is important because it minimizes false positives while still able to capture a significant number of actual flash flood occurrences.

In summary, RF showed a better balance of precision and recall, higher accuracy, and overall better performance metrics for predicting flash floods in the Puchong Area. Despite predicting some non-flash flood events wrongly, RF performs very well in predicting true flash flood events, which is the primary focus of the study. These factors collectively suggested that RF was likely to provide more reliable predictions. Therefore, the recommended choice among the three machine learning models in this specific prediction task in Puchong is RF model.

5 Conclusion

Flash flood is one of the severe natural disasters in Malaysia, and flash flood occurrence will cause serious damage to the economy and human life. Thus, a reliable flash flood prediction is needed to be integrated into the early warning system to mitigate the consequences of flash floods and help in disaster management. Thus, this research addresses the gap by applying predictive models to predict flash flood occurrence. The research aims to contribute ideas on how to apply and identify which machine learning models is the most reliable alternative in predicting flash floods.

The study primarily aims to apply and identify the most accurate and reliable predictive models to predict flash flood occurrence in Puchong using historical water level and rainfall data. The selected study area is IOI Puchong Jaya, as its a flash flood-prone area and is highly impacted once hit by a flash flood. 3 types of ML algorithms are used: Logistic Regression, Random Forest and Long Short-Term Memory. Due to the imbalance class in the dataset, Synthetic Minority Oversampling was applied to training data to remove the bias towards the major class. The models went through hyperparameter tuning to obtain the optimal combination of hyperparameters that improve the accuracy of models. The performance of each model was then validated through 5-fold cross-validation and evaluated in terms of accuracy, recall, F1 score, and precision. The models also applied on predicting the flash flood event in May 2024 and were evaluated. The results showed that Random Forest is the most accurate and reliable predictive model as compared to LSTM and logistic regression because it can correctly predict all true flash flood events while producing fewer false alarms in the fastest speed.

The study also reveals the number of time lags need to be included in input variables. Through the correlation analysis, it was known that the water level in IOI Puchong Jaya area will be affected by the previous rainfall and affected the most by rainfall that occurred 75 minutes earlier. The time lag effect was then incorporated into the model by creating the time-lagged variables that represents cumulative rainfall over different time windows. Thus, there are 5 new lagged variables created and used in modelling. In summary, the optimal time lag effect of rainfall on water level is lag 1 to 5 and has been included in the predicting flash flood.

In nutshell, the lagged variables up to 5th lag (75 minutes) are created and used in model building to improve the accuracy of models. Besides, the research proved that ML is an effective approach in flood prediction using rainfall and water level data. This research reveals that the RF stands out from other ML algorithms in predicting flash flood occurrence in Puchong. The application of the RF algorithm aligns with the primary goal of accurate flood prediction, which is to provide reliable early warnings to ensure people have as much time as possible to evacuate.

For future work, meteorological variables such as temperature, wind speed, and topographical factors can be incorporated into the model. More tuning on the parameters of the model.

Acknowledgments

All the authors gratefully acknowledged the financial support from the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code:
FRGS/1/2022/STG06/USM/03/1

References

[1] Saad, M. S. H., Ali, M. I., Razi, P. Z., Ramli, N. I. and Jaya, R. P. Exploring the Factors and Impacts of Flash Floods Vulnerability in Various Areas of Malaysia: A Content Analysis. *Disaster in Civil Engineering and Architecture*. 2024. 1(1):55-82.

[2] Chang, J. H. W., Kong, S. S. K., Sentian, J., Dayou, J. and Chee, F. P. Synoptic analysis and mesoscale numerical modelling of heavy precipitation: A case study of flash flood event in Kota Kinabalu, Malaysia. *Meteorology and Atmospheric Physics*. 2020. 132:181-201.

[3] Maqtan, R., Othman, F., Wan Jaafar, W. Z., Sherif, M. and El-Shafie, A. A scoping review of flash floods in Malaysia: current status and the way forward. *Natural Hazards*. 2022.114(3): 2387-2416.

[4] Sanyal, J. and Lu, X. X. Application of remote sensing in flood management with special reference to monsoon Asia: A review. *Natural Hazards*. 2004. 33: 283301.

[5] Rosmadi, H. S., Ahmed, M. F., Mokhtar, M. B., & Lim, C. K. Reviewing challenges of flood risk management in Malaysia. *Water*. 2023. 15(13).

[6] Buslima, F., Omar, R., Jamaluddin, T., & Taha, H. Flood and flash flood geo-hazards in Malaysia. *International Journal of Engineering and Technology (UAE)*. 2018. 7: 760764.

[7] Someetheram V., Marsani M. F., Kasihmuddin M. S. M. and Zamri N. E.Hybrid least squares support vector machine for water level forecasting. *Mathematical Modeling and Computing*. 2025. 12(2): 384400.

[8] Marsani, M. F., Someetheram, V., Mohd Kasihmuddin, M. S., Mohd Jamaludin, S. Z., Mansor, M. A. and Badyalina, B. The application of seasonal autoregressive integrated moving average (SARIMA) model in forecasting Malaysia mean sea level. In *AIP Conference Proceedings*. 2024. 3123(1):070005.

[9] Marsani, M. F., Someetheram, V., Mohd Kasihmuddin, M. S., Mohd Jamaludin, S. Z., Mansor, M. A., Badyalina, B. and Jan, N. A. M. Enhancing rainfall prediction in SVM by incorporating the autocorrelation function (ACF). In *AIP Conference Proceedings*. 2024. 3123(1):070006.

[10] Che Hamid, H. E., Mat Razali, N. A., Ismail, M. N., Khairuddin, M. A. and Mohd Isa, M. R. A review paper: Forecasting of flood in Malaysia using machine learning. In *AIP Conference Proceedings*. 2022. 2617(1).

[11] Sankaranarayanan, S., Prabhakar, M., Satish, S., Jain, P., Ramprasad, A. and Krishnan, A. Flood prediction based on weather parameters using deep learning. *Journal of Water and Climate Change*. 2020.11(4):1766-1783.

[12] Vidya, S., Gayathri, S., Dilsha, M. and Krithika, S. Rainfall based flood prediction in Kerala using machine learning.*Int J Intell Syst Appl Eng*. 2024.12(16s):141-4.

[13] Ghorpade, P., Gadge, A., Lende, A., Chordiya, H., Gosavi, G., Mishra, A. and Shaikh, N. Flood forecasting using machine learning: A review. In *2021 8th international conference on smart computing and communications IEEE*. 2021. 32-36.

[14] Idan, H. M. and Hussein, K. Q. Comparison study between selected techniques of (ML, SVM and Deep Learning) regarding prediction of Flooding in Eastof Iraq. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*. 2021.12(14):2893-2904.

[15] Pradhan, B. Flood susceptible mapping and risk area delineation using logistic regression,GIS and remote sensing. JOSH. 2010. 9:118.

[16] Tehrany, M. S., Kumar, L., Jebur, M. N. and Shabani, F. Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods. *Geomatics, Natural Hazards and Risk*. 2019. 10(1):79-101.

[17] Viteri López, A. S. and Morales Rodriguez, C. A. Flash flood forecasting in So Paulo using a binary logistic regression model. *Atmosphere*. 2020.11(5): 473.

[18] Lawal, Z. K., Yassin, H., and Zakari, R. Y. Flood prediction using machine learning models: a case study of Kebbi state Nigeria. *Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) IEEE*. 2021. 1-6.

[19] Hadi, F. A. A., Mohd Sidek, L., Ahmed Salih, G. H., Basri, H., Sammen, S. S., Mohd Dom, N., and Najah Ahmed, A. Machine learning techniques for flood forecasting. *Journal of Hydroinformatics*. 2024. 26(4): 779-799.

[20] Schoppa, L., Disse, M., and Bachmair, S. Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*. 2020.590: 125531.

[21] Mosavi, A., Ozturk, P. and Chau, K. W. Flood prediction using machine learning models: Literature review. *Water*. 2018. 10(11):1536.

[22] Saha, A. and Pal, S.C. Application of machine learning and emerging remote sensing techniques in hydrology: A state-of-the-art review and current research trends. *Journal of Hydrology*. 2024. 130907.

[23] Someetheram, V., Marsani, M. F., Mohd Kasihmuddin, M. S., Mohd Jamaludin, S. Z. and Mansor, M. A. Double decomposition with enhanced least-squares support vector machine to predict water level. *Journal of Water and Climate Change*. 2024.15(6):2582-2594.

[24] Wu, Y., Zhang, Z., Qi, X., Hu, W. and Si, S. Prediction of flood sensitivity based on Logistic Regression, eXtreme Gradient Boosting, and Random Forest modeling methods. *Water Science & Technology. 2024*.89(10):2605-2624.

[25] Waqas, M. and Humphries, U. W. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*. 2024. 102946.

[26] Hakim, D. K., Gernowo, R. and Nirwansyah, A. W. Flood prediction with time series data mining: Systematic review. *Natural Hazards Research*. 2023. 4(2):194-220.

[27] Atashi, V., Gorji, H. T., Shahabi, S. M., Kardan, R. and Lim, Y. H. Water level forecasting using deep learning time-series analysis: A case study of red river of the north. *Water*. 2022.14(12), 1971.

[28] Zhu, Y., Feng, J., Yan, L., Guo, T. and Li, X. Flood prediction using rainfall-flow pattern in data-sparse watersheds. *IEEE Access*. 2020.8: 39713-39724.

[29] Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A. and Kwon, H. H. Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*. 2021.35(12): 4167-4187.

[30] Jaafar, N. IOI Mall Puchong ditenggelami banjir. in *Sinar Harian*. 2023.

[31] Daho, M. E. H. and Chikh, M. A. Combining bootstrapping samples, random subspaces and random forests to build classifiers. *Journal of Medical Imaging and Health Informatics*. 2015. 5(3):539-544.

[32] Schonlau, M. and Zou, R. Y. The random forest algorithm for statistical learning. *The Stata Journal*. 2020. 20(1):3-29.

[33] Le, X. H., Ho, H. V., Lee, G. and Jung, S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*. 2019.11(7): 1387.

[34] Zhao, G., Liu, R., Yang, M., Tu, T., Ma, M., Hong, Y. and Wang, X. (2022). Large-scale flash flood warning in China using deep learning. *Journal of Hydrology*. 2022. 604:127222.

[35] Someetheram, V., Marsani, M. F., Kasihmuddin, M. S. M., Jamaludin, S. Z. M., Mansor, M. A. and Zamri, N. E. Hybrid double ensemble empirical mode decomposition and K-Nearest Neighbors model with improved particle swarm optimization for water level forecasting. *Alexandria Engineering Journal*. 2025. 115:423-433.