



# International Journal of Research Publication and Reviews

Journal Homepage: [www.ijrpr.com](http://www.ijrpr.com) ISSN 2582-7421

## Rainfall Prediction System Using Machine Learning

*D. Esther<sup>1</sup>, J. Steffi<sup>2</sup>*

<sup>1</sup>II MCA, Reg.No: 25081205300112011, Department of Computer Applications and Research Centre, Sarah Tucker College(Autonomous), Tirunelveli.

<sup>2</sup>Assistant Professor, Department of Computer Applications and Research Centre, Sarah Tucker College(Autonomous), Tirunelveli.

### ABSTRACT

Forecasting rain is one of the main reasons people use weather data and machine learning. Accurate rainfall forecasts help farmers plan, manage water supplies, run transportation systems, and support other weather-dependent activities. This project uses machine learning to predict if rain is coming and how much measured in centimeters. The system depends on four main weather factors: temperature, humidity, pressure, and wind speed. The dataset gets cleaned up—filling in missing bits, removing duplicates, and building the weather features we need. We evaluate classification and regression tasks using machine learning methods like Decision Tree, Random Forest, and XGBoost. Random Forest scored the highest accuracy at 82.44%. For predicting rainfall amounts, the Random Forest Regressor scored an MAE of 0.2454, RMSE of 0.6045, and an  $R^2$  of 0.2693 among the best results. The system pulls in real-time weather and forecasts using the OpenWeather API, tailored to the user's state and city. The three-day forecast is made by crunching current weather data through trained machine learning models. The results show up in a live Streamlit web app with weather cards, predictions, graphs, and forecast details. MySQL holds the prediction results for easy future reference and analysis.

**Keywords:** Rainfall Prediction, Machine Learning, Random Forest, Classification, Regression, OpenWeather API, Streamlit, MySQL, Weather Forecasting, Data Analysis.

### INTRODUCTION

Rainfall matters a lot it shapes farming, water supplies, travel, and daily life. Getting accurate rainfall forecasts helps people and organizations plan better for upcoming weather. Traditional rainfall prediction methods[9] usually depend on weather data and stats.[7] As machine learning gets smarter, historical weather data helps reveal how different weather factors link to rainfall. This project builds a machine learning-powered rainfall prediction system[1]. The system uses temperature, humidity, pressure, and wind speed as its main inputs. Two prediction tasks are run: one to classify whether rain is expected. The second task is rainfall regression, which forecasts expected rainfall in centimeters. It also pulls in real-time weather data using the OpenWeather API. Users enter their state and city, and the system pulls up the matching weather info. Once trained, the models analyze those values and produce rainfall forecasts[10].

### METHODOLOGY

The Rainfall Prediction System uses machine learning to predict when it will rain and how much falls, based on weather data. The system mainly uses these inputs: temperature, humidity, pressure, and wind speed. The approach includes cleaning and preparing data, training the model, testing its performance, gathering live weather info, and making predictions.

#### 1.Data Preprocessing

The dataset includes various weather measurements gathered specifically for rainfall analysis. The needed weather features were picked and adjusted into a format that works well for machine learning. The system uses four key inputs: temperature, humidity, pressure, and wind speed. We figured out the average temperature by taking the max and min values and plugging them into this formula:

$$\text{Temperature} = (T_{\text{max}} + T_{\text{min}}) / 2$$

We filled in missing numbers with smart stats and scrubbed duplicates to clean up the dataset.

#### 2. Target variable generation

We created two target variables for the predictions. The first one, Rain, helps classify rainfall. If the precipitation is above zero, it's marked as rainy; otherwise, it's dry.

$$\text{Rain} = 1, \text{ if } \text{PRECTOTCORR} > 0$$

$$\text{Rain} = 0, \text{ if } \text{PRECTOTCORR} = 0$$

Rainfall\_cm measures how much rain fell, in centimetres. The precipitation value was converted to centimeters using:

$$\text{Rainfall\_cm} = \text{PRECTOTCORR} / 10$$

### 3. Machine Learning Model Development

The available dataset was divided into separate training and testing portions for model evaluation[8]. For rainfall classification, three algorithms were considered: Decision Tree, Random Forest, and XGBoost[4]. For estimating the amount of rainfall, Linear Regression, Random Forest Regressor, and XGBoost Regressor were evaluated[3]. The models used weather-related inputs such as temperature, humidity, pressure, and wind speed. Among the classification models, Random Forest achieved an accuracy of 82.44%[5] and was selected for the final prediction system. For rainfall amount estimation, the Random Forest Regressor produced an MAE of 0.2454, an RMSE of 0.6045, and an  $R^2$  value of 0.2693, making it the selected regression model.

### 4. Real-Time Weather Data Collection

The system integrates the OpenWeather API to obtain real-time weather information based on the State and City entered by the user. The API provides Temperature, Humidity, Atmospheric Pressure, and Wind Speed. The retrieved weather values are processed according to the same format used during model training. The processed values are then provided as input to the trained classification and regression models.

### 5. Rainfall Prediction

The trained Random Forest classifier predicts whether rainfall is expected or not. The classification output is represented as Rain or No Rain. The Random Forest Regressor predicts the expected rainfall amount in centimetres. The two prediction results are displayed together to provide a clear understanding of the expected rainfall condition and quantity.

### 6. Model Evaluation

The classification model's performance was assessed based on accuracy. Accuracy is determined using the formula :

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

In this context, TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative.

The regression models were assessed based on Mean Absolute Error ( MAE), Root Mean Square Error ( RMSE), and  $R^2$  score.

$$\text{MAE} = \sum |y_i - \hat{y}_i| / n$$

$$\text{RMSE} = \sqrt{(\sum (y_i - \hat{y}_i)^2) / n}$$

The  $R^2$  score indicates how effectively the model accounts for the variability in the rainfall values.

### 7. Three-day forecasting

The system also offers a three-day rainfall forecast based on weather data retrieved from the OpenWeather API. The available forecast records are organized by future dates, and the average values for Temperature, Humidity, Pressure, and Wind Speed are computed for each day. These daily weather values are provided to the trained machine learning models[6] to forecast the rainfall conditions and amounts for each of the three days.

### 8. Result visualization and data storage

The prediction results are displayed via an interactive Streamlit web application. Weather parameters, rainfall likelihood, expected rainfall volume, and three-day forecast outcomes are presented through appropriate charts and visual aids. The system also saves weather conditions and prediction results in a MySQL database. This enables the generated predictions to be preserved for future reference and analysis.

## RESULTS AND DISCUSSION

We tested the system using both classification and regression approaches. The experiment shows how well the selected ML models handle the four weather parameters.

### Current Weather Overview

Based on the retrieved real-time weather parameters, the trained machine learning models generate the rainfall prediction for the selected location. For the Chennai test case, the classification model predicted the occurrence of rainfall, while the regression model estimated the expected rainfall amount as 1.24 cm. The combined prediction provides both rainfall occurrence and estimated rainfall quantity for the current weather condition.

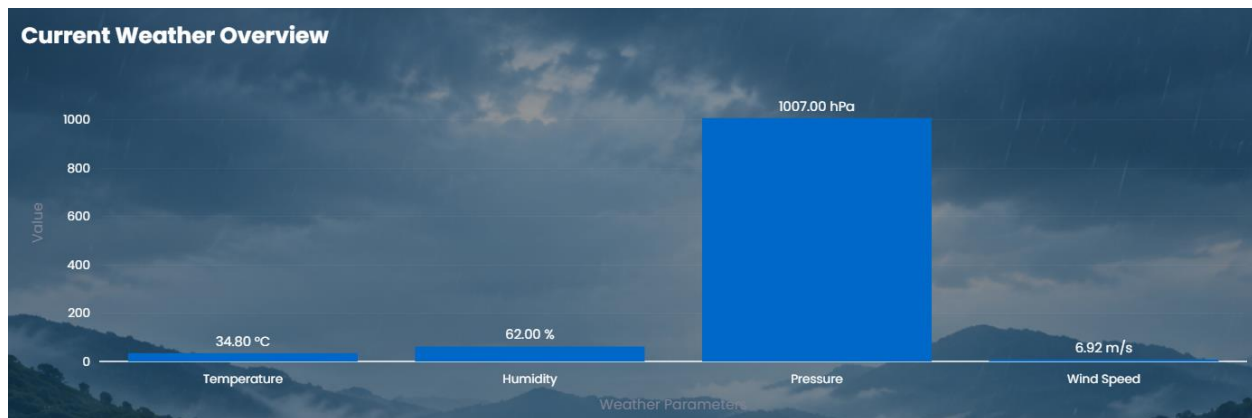


Fig 1: Current Weather Overview

### Current Weather

The developed system retrieves real-time weather conditions for the selected location through the OpenWeather API. For Chennai, Tamil Nadu, the system recorded a temperature of 34.80 °C, humidity of 62.00%, atmospheric pressure of 1007.00 hPa, and wind speed of 6.92 m/s. These parameters represent the current atmospheric conditions and are used as the input features for the rainfall prediction models.

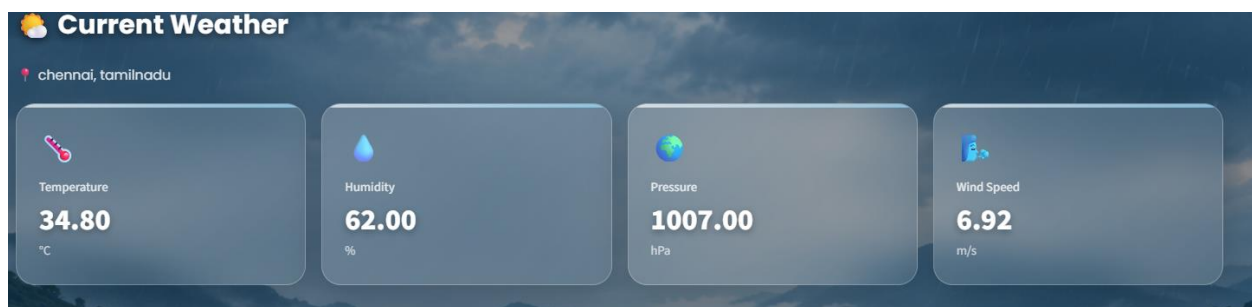


Fig 2: Current Weather

### Rainfall Prediction Result

The proposed system generates the final rainfall prediction by processing the real-time weather parameters obtained from the selected location. For the Chennai test case, the machine learning model predicted **Rain**, with an estimated rainfall amount of **1.24 cm**. The prediction combines rainfall occurrence classification with rainfall amount estimation, providing a comprehensive output based on the current atmospheric conditions. This result demonstrates the practical integration of real-time weather data with the trained machine learning models for rainfall prediction.

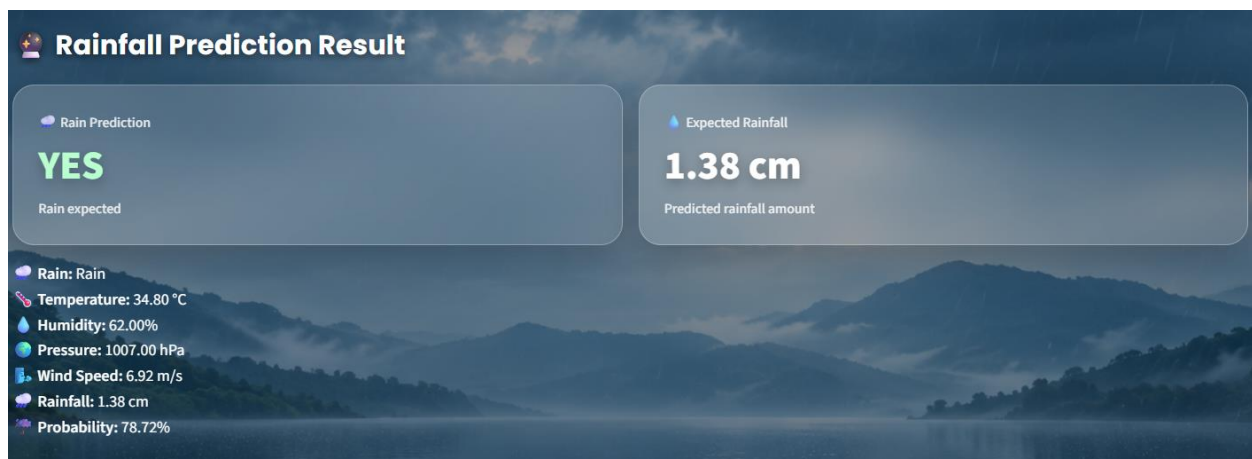


Fig 3: Rainfall Prediction Result

### Rain Probability

The rainfall probability represents the confidence of the classification model in predicting the occurrence of rainfall. For the selected Chennai weather condition, the system obtained a rain probability of 78.72% and a no-rain probability of 21.28%. The probability visualization provides a clear representation of the model output and helps users interpret the likelihood of rainfall under the observed weather conditions.

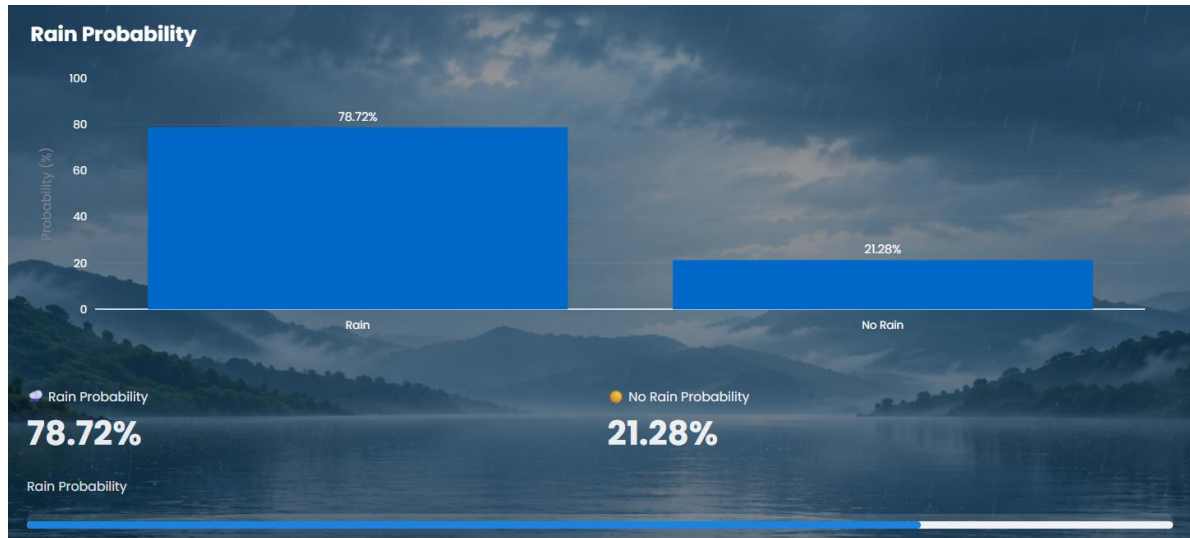


Fig 3: Rain Probability

### Three-day rainfall forecast details

The second tab of the proposed system provides rainfall forecasts[2] for the upcoming three days based on forecast weather data obtained from the OpenWeather API. The temperature, humidity, pressure, and wind speed values for each forecast day are processed by the trained machine learning models to determine rainfall occurrence and estimate the expected rainfall amount. The predicted results are presented day-wise through a clear and user-friendly interface, enabling users to observe the expected rainfall conditions over the three-day period.



Fig 4: Three-day rainfall forecast details

### Three-day Temperature and Humidity Trend

#### (a) Three-Day Temperature Forecast

The system displays the expected temperature for the upcoming three days based on forecast weather data. The day-wise values help users understand the temperature variation during the forecast period.

#### (b) Three-Day Humidity Forecast

The system presents the expected humidity levels for the upcoming three days. The day-wise humidity values provide a clear view of atmospheric moisture conditions during the forecast period.

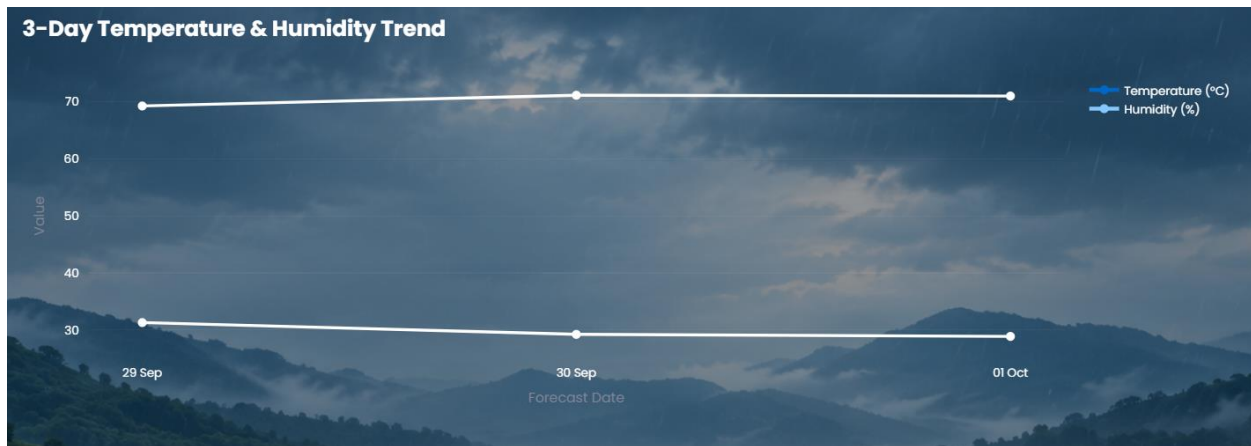


Fig 5: 3-Day Temperature and Humidity Forecast

#### Predicted Rainfall-Next 3 Days in Graph

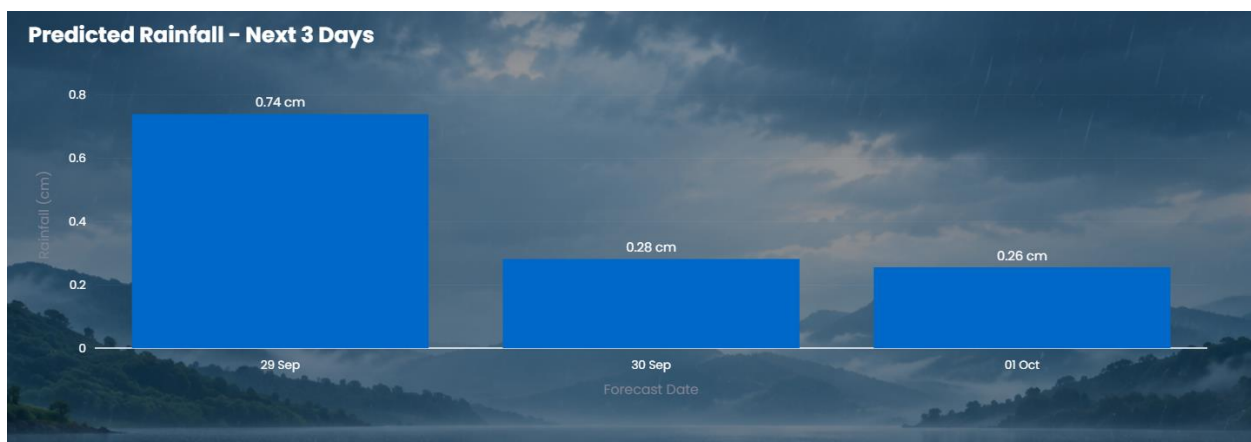


Fig 6: Predicted Rainfall in Graph

#### Forecast Summary

The forecast summary presents the predicted weather and rainfall conditions for the upcoming three days. It combines the forecasted weather parameters with the machine learning prediction results in a clear day-wise format. This provides users with a concise overview of the expected weather and rainfall conditions.

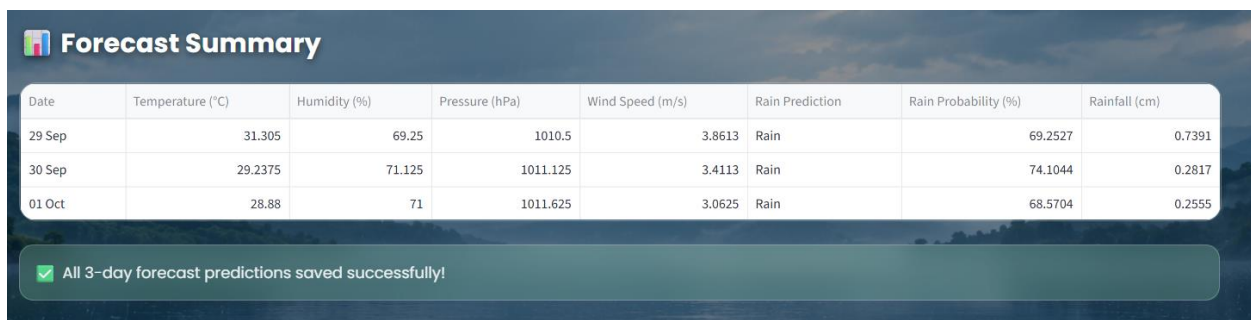


Fig 7: Forecast Summary

## CONCLUSION

The proposed Rainfall Prediction System successfully integrates machine learning with real-time weather data to predict rainfall occurrence and estimate rainfall amount. The system uses temperature, humidity, pressure, and wind speed as the primary input features and provides both classification and regression-based predictions. Among the evaluated models, Random Forest provided the best performance for both rainfall classification and rainfall amount prediction.

The integration of the OpenWeather API enables the system to process real-time weather conditions, while the three-day forecasting module provides predicted rainfall conditions for upcoming days. The Streamlit-based interface presents the results in a clear and user-friendly manner, and the MySQL database enables systematic storage of prediction records. Overall, the developed system demonstrates the practical application of machine learning for real-time rainfall prediction and short-term weather forecasting.

## REFERENCES

1. A. Khan, M. S. Khan, M. A. Khan, S. M. Khan, and A. Ahmad, "Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India," *Results in Engineering*, vol. 22, art. 102093, 2024, doi: 10.1016/j.rineng.2024.102093.
2. A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. Davila Delgado, and L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Machine Learning with Applications*, vol. 7, art. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.
3. M. A. Azeem and S. Dev, "A performance and interpretability assessment of machine learning models for rainfall prediction in the Republic of Ireland," *Decision Analytics Journal*, vol. 12, art. 100515, 2024, doi: 10.1016/j.dajour.2024.100515.
4. P. Mishra, A. M. G. Al Khatib, S. Yadav, S. Ray, A. Lama, B. Kumari, D. Sharma, and R. Yadav, "Modeling and forecasting rainfall patterns in India: a time series analysis with XGBoost algorithm," *Environmental Earth Sciences*, vol. 83, art. 163, 2024, doi: 10.1007/s12665-024-11481-w.
5. P. S. Yu, T. C. Yang, S. Y. Chen, C. M. Kuo, and H. W. Tseng, "Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting," *Journal of Hydrology*, vol. 552, pp. 92–104, 2017, doi: 10.1016/j.jhydrol.2017.06.020.
6. S. Waghere, A. Biradar, R. Zambre, A. Golande, A. Hingmire, and V. Bhutnal, "A Flask-Integrated Machine Learning Framework for Regional Rainfall Prediction in Indian Districts," in *Smart Trends in Computing and Communications (SmartCom 2026)*, Lecture Notes in Networks and Systems, vol. 1995, pp. 130–140, 2026, doi: 10.1007/978-3-032-27163-1\_12.
7. T. Pham, L. M. Le, T.-T. Le, K.-T. T. Bui, V. M. Le, H.-B. Ly, and I. Prakash, "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmospheric Research*, vol. 237, art. 104845, 2020, doi: 10.1016/j.atmosres.2020.104845.
8. T. Talan, "Machine learning-based rainfall prediction across temporal scales: model benchmarking and explainability analysis," *Stochastic Environmental Research and Risk Assessment*, vol. 40, art. 100, 2026, doi: 10.1007/s00477-026-03245-8.
9. V. Geethalakshmi, S. Kokilavani, S. P. Ramanathan, G. A. Dheebakaran, N. K. Sathyamoorthy, and N. Maragatham, "Monthly rainfall hind-cast using machine learning algorithms for Coimbatore, Tamil Nadu," *MAUSAM*, vol. 73, no. 1, 2022, doi: 10.54302/mausam.v73i1.5077.
10. Y. Z. Teoh and Y. L. Loo, "A Fundamental Framework for Analysis of Rainfall Prediction Features Significance," *Journal of Informatics and Web Engineering*, vol. 4, no. 2, pp. 36–52, 2025, doi: 10.33093/jiwe.2025.4.2.3.