

**Original Article**

## Machine Learning Approaches to Rainfall Forecasting in Nigeria: A Comparative Study of RF, SVR, XGBoost, and DNN Models

Gbenga O. Ogunsanwo, Ismail O. Ogunshina, Olumuyiwa B. Alaba, Ayoade A. Owoade, and Godwin O. Odulaja

Department of Computer Science, College of Science and Information Technology, Tai Solarin University of Education, Ogun State, Nigeria.

Corresponding email: [ogunsanwogo@tasued.edu.ng](mailto:ogunsanwogo@tasued.edu.ng)

DOI: <https://doi.org/10.71428/PJS.2026.0206>

### Abstract

Rainfall prediction is critical for agricultural planning, water resource management, and disaster preparedness, particularly in regions vulnerable to climate variability. Traditional forecasting techniques often struggle to capture the non-linear and complex dynamics of meteorological data. This study developed machine learning models to improve rainfall prediction using historical data obtained from the Nigerian Meteorological Agency, which includes climatic variables such as temperature, wind speed, humidity, and prior rainfall. After preprocessing and normalization, the dataset was divided into training (80%) and testing (20%) subsets. Four algorithms, such as Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Random Forest (RF), and Deep Neural Network (DNN), were implemented and evaluated using regression metrics such as  $R^2$ , Mean Squared Error, and Mean Absolute Error. The results show that RF and XGBoost achieved moderate accuracy, with  $R^2$  values below 0.87. SVR performed better, reducing error significantly (MAE = 0.6839, MSE = 13.8717,  $R^2$  = 0.893). The DNN model outperformed all others, achieving very low error (MAE = 0.2334, MSE = 0.1956) and near-perfect accuracy ( $R^2$  = 0.9985). These findings demonstrate the superior capacity of deep learning approaches to model complex rainfall patterns compared to traditional ensemble and regression methods, while SVR remains a reliable alternative. The study recommends that stakeholders in meteorology and agriculture adopt advanced machine learning models for improved decision-making. Future research should incorporate high-resolution satellite data and ground station records for hybrid modeling and employ techniques such as SHAP and LIME.

**Keywords:** Rainfall, Prediction, Machine Learning, Ensemble, Deep Neural Network

### 1 Introduction

Rainfall is a vital meteorological phenomenon that profoundly affects both human society and the natural ecosystem. In Nigeria, where agriculture, water resource management, and disaster preparedness are closely tied to seasonal rainfall patterns, accurate prediction is essential for

sustainable development. However, forecasting rainfall remains a formidable challenge due to the complex, non-linear interactions among atmospheric variables, topography, and regional climate dynamics (1).

Traditional statistical models and basic machine learning techniques often fall short in capturing these intricacies, especially during extreme weather events such as floods and droughts. The consequences of poor rainfall prediction are far-reaching: disrupting agricultural productivity, weakening disaster response systems, and threatening food and water security. In regions with limited meteorological infrastructure and diverse climatic zones, such as Nigeria, the challenge is further compounded by sparse high-quality data and inconsistent spatial coverage (2). Accurate rainfall forecasting plays a pivotal role in policy making, infrastructure planning, and climate resilience. It enables informed decisions in irrigation scheduling, flood control, and early warning systems. Recent advances in machine learning have shown promise in overcoming traditional limitations by learning complex patterns from historical data. However, the effectiveness of these models depends heavily on proper data preprocessing, feature selection, and hyperparameter optimization.

There is a growing consensus among researchers that adaptive and robust machine learning approaches, particularly those capable of generalizing across different regions and timescales, are urgently needed to improve forecast accuracy and reliability. Recent studies have demonstrated that models such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) outperform conventional methods in tropical settings like Nigeria (1,2). Many researchers have worked on predicting rainfall; for instance, Khan et al. compared four machine learning models, which are: Logistic Regression, Decision Tree, Multi-Layer Perceptron (MLP), and Random Forest for rainfall prediction in Aligarh City using a 10-year NASA POWER dataset. Logistic Regression and MLP performed best, with Logistic Regression achieving 82.80% accuracy and MLP showing

superior precision (82.80%), recall (79.40%), and F1 score (80.30%) for positive rainfall cases. The study highlights MLP's strength in detecting rainfall events and supports its use in climate-sensitive forecasting (3). Aizansi et al. (2024) applied MLP and LSTM models to predict monthly rainfall in the Republic of Benin using 63 years of data (1959–2021) from six weather stations, sourced from METEOBENIN and the Climate Data Store (CDS). Model performance was evaluated using RMSE, MAE, MAPE,  $R^2$ , and NSE. Results showed that MLP slightly outperformed LSTM, with RMSE ranging from 39.35–72.41 mm and  $R^2$  up to 0.912. LSTM showed comparable accuracy but slightly higher error margins. The study confirms the effectiveness of deep learning models for long-term rainfall forecasting in West Africa (4). Ridwan et al. evaluated four machine learning models—Bayesian Linear Regression, Boosted Decision Tree Regression (BDTR), Decision Forest Regression, and Neural Network Regression—for rainfall forecasting in Terengganu, Malaysia, using data from 10 weather stations. BDTR emerged as the most accurate model, especially under Method 1 (ACF), achieving high  $R^2$  values across daily (0.55–0.97), weekly (0.84–0.99), 10-day (0.80–0.99), and monthly (0.91–0.99) predictions. Cross-validation and hyperparameter tuning significantly improved performance. Method 2 (Projected Error) also favored BDTR, confirming its reliability for weekly rainfall error correction. Moreover, it was confirmed that input features consistently enhanced model accuracy (5). Zhang et al. proposed a hybrid deep learning model, which is SCSSA-CNN-BILSTM, for monthly rainfall prediction using data from Xi'an City (1996–2020). The model integrates SCSSA for optimization, CNN for feature extraction, and BILSTM for temporal dependency modeling. The study compared SSA-CNN-BILSTM, CNN-BILSTM, and SCSSA-BILSTM; the proposed model achieved the highest accuracy (6). Peeriga et al. proposed a Bi-Directional LSTM (Bi-LSTM) model for real-time rainfall prediction

in agricultural contexts, leveraging AI and IoT technologies. The model is trained on historical weather data from the India Meteorological Department, incorporating key variables such as temperature, humidity, wind speed, and precipitation. The model performance was assessed using RMSE, MSE, and MAE, demonstrating strong predictive accuracy. Descriptive statistics (mean, standard deviation, min, max) were used to characterize the dataset and highlight its variability, which is essential for reliable forecasting. The study underscores the effectiveness of Bi-LSTM in capturing temporal dependencies critical for accurate rain prediction in precision agriculture (7). Aiyelokun et al. explored the use of Random Forest (RF) to correct biases in Climate Forecast System Reanalysis (CFSR) data for water resource modeling in Nigeria's Ogun River basin. By integrating monthly climate variables such as relative humidity, temperature, and rainfall from both CFSR and nearby weather stations, the RF model generates bias-reduced datasets (RF-CFSR) that closely align with observed seasonal patterns. The approach significantly improves accuracy metrics, including mean absolute error and correlation coefficients, making RF a viable alternative to traditional bias correction methods like quantile mapping (8). Despite the growing global interest in applying machine learning (ML) techniques to rainfall prediction, studies conducted in Nigeria remain limited in both scope and methodological depth. Most existing Nigerian research has relied on traditional ML models such as Random Forest (RF), Support Vector Regression (SVR), and XGBoost, with occasional exploration of Deep Neural Networks (DNN). While these

approaches have achieved moderate success, several critical gaps persist. Notably, datasets produced by Nigerian meteorological agencies have not been systematically leveraged, with many studies depending instead on generalized or international datasets. This lack of localized data integration reduces the contextual accuracy and practical relevance of current models, underscoring the need for a comparative study that evaluates advanced ML approaches using authentic Nigerian rainfall datasets. In response to this need, this study develops and evaluates four machine learning algorithms for predicting rainfall amounts using historical meteorological data from Nigeria. The goal is to enhance predictive accuracy, support proactive resource planning, and contribute to climate-informed decision-making in agriculture and disaster management.

## 2 Methodology

This study employs a data-driven approach to develop a rainfall prediction model using SVM, XGBoost, and RF algorithms. The process includes data acquisition, preprocessing, feature extraction, model training, evaluation, and comparative analysis.

### Data Collection for the study

Rainfall and meteorological data were obtained from the Nigerian Meteorological Agency (NiMet) (<https://nimet.gov.ng/datarequest>) covering a 4-year period at Ikeja's Murtala Muhammed Station. The key features include rain rate, rainy days, evaporation, precipitation, and minimum/maximum average temperatures. The study's workflow is illustrated in Figure 1.

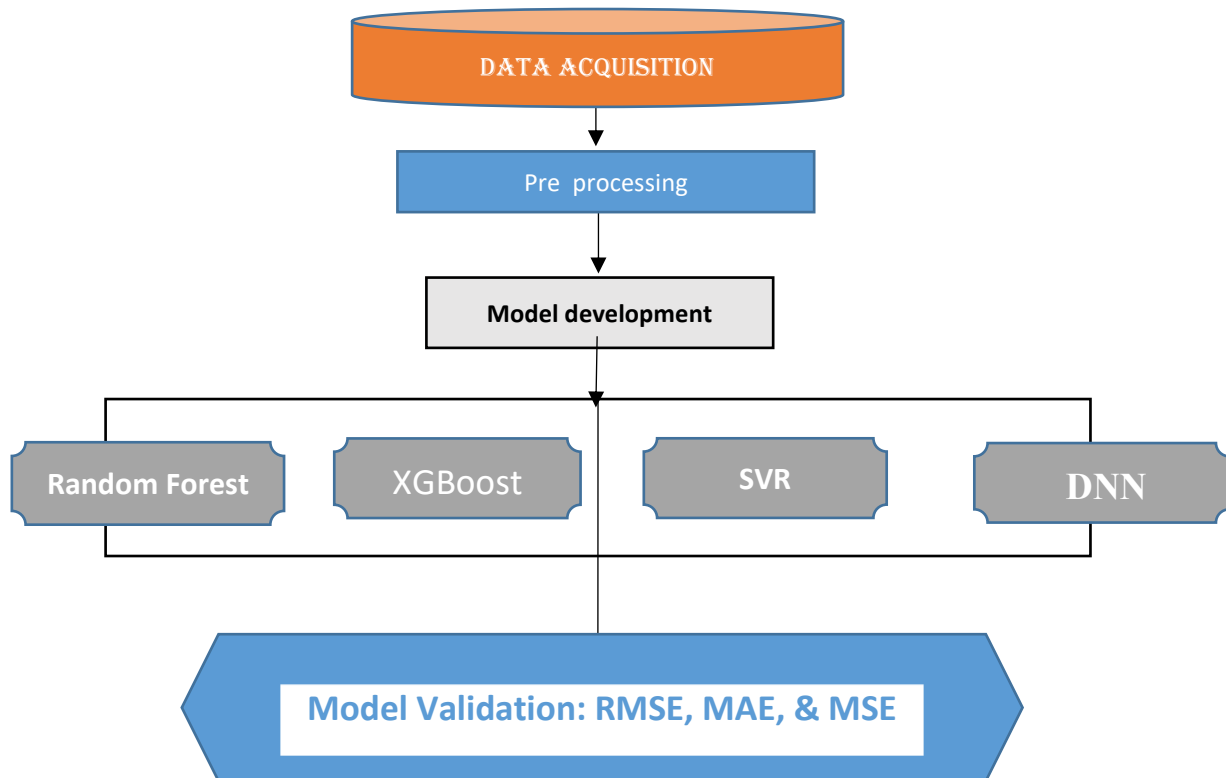


Figure 1: workflow of the study

### Data Preprocessing

The raw dataset underwent cleaning to address missing values, outliers, and inconsistencies. Feature scaling through normalization or standardization was applied to ensure consistency across variables, particularly for magnitude-sensitive algorithms. For time series modeling, temporal features were converted into lag sequences to capture dependencies. Finally, the data was partitioned into training and testing sets for model evaluation.

### Model Development

This study follows a structured machine learning workflow to build accurate rainfall prediction models. Four algorithms were implemented: SVR, XGBoost, RF, and DNN.

**SVR** is a regression-based extension of Support Vector Machines (SVM), designed to predict continuous outcomes. Unlike traditional regression, SVR employs an  $\epsilon$ -insensitive loss function, which ignores errors within a defined threshold and

emphasizes fitting data within a margin of tolerance rather than minimizing all deviations (9) as shown in equation 1

$$y = \sum_{i=1}^N (\alpha - \alpha_1^*) \cdot (x_i, x) + b \quad (1)$$

Where:  $y$ : Predicted output value for a given input  $x$ ,  $N$ : Total number of training data points  $\alpha_i, \alpha_1^*$ : Lagrange multipliers obtained during training. They control the influence of each support vector.  $x_i$ : Support vector (a training sample).  $x$ : The input data point we want to predict for, and  $b$ : Bias term (intercept), also learned during training.

**XGBoost** is a powerful ensemble learning algorithm that constructs DT sequentially to minimize prediction error and enhance model accuracy. It is widely recognized for its scalability, speed, and superior performance in structured data tasks (10), see equation (2).

$$\hat{y}_{i+1} = \frac{1}{T} \sum_{k=1}^{ff} f_k F \quad (2)$$

Where:

$\hat{y}_i$  : The predicted value for the i-th observation/sample

i: The index/identifier for a specific data point or observation

T: Total number of boosting rounds/iterations

k: The iteration/round index in the boosting process

fF: Total number of trees/weak learners in the final ensemble

fk: The k-th weak learner (decision tree) in the ensemble

F: Appears to be a notation for the final ensemble or the space of all possible tree functions

**RF** is an ensemble learning technique that constructs multiple decision trees and aggregates their outputs to enhance prediction accuracy and reduce overfitting. By using bootstrap sampling and random feature selection, it ensures model robustness and generalization across diverse datasets (11), see equation (3)

$$\hat{y}_{RF} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t \quad (3)$$

Where:

$\hat{y}_{RF}$ : The final prediction made by the Random Forest model.

T: The total number of decision trees in the forest.

$\hat{y}_t$  : The prediction made by the t individual decision tree  $h_t(x)$ .

$\frac{1}{T}$  : The normalization factor to compute the average (mean) prediction from all trees

**DNN** is a multilayered artificial neural network designed to capture complex nonlinear relationships between variables. Its depth enables hierarchical feature learning, making it highly effective for tasks like rainfall prediction involving multiple interacting meteorological factors (12), see equation (4)

$$\hat{y} = f(W^{(L)} \cdot a^{(L-1)} + b^L) \quad (4)$$

Where:

$\hat{y}$ : The final prediction made by the DNN

L: Total number of layers in the network

$W^{(L)}$  : Weights of the last layer

$a^{(L-1)}$ : Activations from the second to the last layer

$b^{(L)}$ : Bias term of the last layer

f : Activation function (e.g., sigmoid, ReLU, softmax, depending on the task)

### Model Validation

**MSE** is a widely used loss function in model training, particularly for regression tasks. It calculates the average of the squared differences between predicted and actual values, helping quantify prediction accuracy by penalizing larger errors more heavily (13), see equation (5)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

Where:

$y_i$  represents the observed (actual) value for the i-th data point.

$\hat{y}_i$  represents the predicted value for the i-th data point.

n is the total number of samples (or data points) in the dataset.

**MAE** is a widely used metric for evaluating regression models. It measures the average absolute difference between predicted and actual values, with lower MAE indicating higher prediction accuracy, see equation (6)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Where:

$y_i$  or alternatively  $y_i, y_j$ ,  $MSL_o$ : Actual (observed) value for the i-th (or t-th, j-th) data point

$\hat{y}_i$  or alternatively  $\hat{y}_t, \hat{y}_j$ ,  $MSL_p$ : Predicted value for the i-th (or t-th, j-th) data point

n: total number of data points (samples/observations).

( $R^2$ , or the coefficient of determination, measures how well a regression model explains the variance in the dependent variable. A higher  $R^2$  indicates that the model's predictors account for a greater proportion of the outcome variability, reflecting better fit and predictive power (14), see equation (7)

$$R^2 = 1 - \frac{\sum_{i=1}^n (R_o - R_p)^2}{\sum_{i=1}^n (R_o - \bar{R}_o)^2} \quad (7)$$

Where:

$R_o$ : Observed daily rainfall

$R_p$ : Predicted daily rainfall

$\bar{R}_o$ : Mean of observed rainfall

### 3 Results and Discussions

The sample weather dataset used for Nigerian rainfall prediction includes features such as 'Date', 'Evaporation', 'MaxTemp', 'MinTemp', 'RainRate', and 'Rainfall', as illustrated in Figure 2.



	Date	Evaporation	MaxTemp	MinTemp	RainRate	Rainfall
0	2021-01-01	2.0	34.0	24.1	0.0	0.0
1	2021-01-02	2.1	34.2	25.0	0.0	0.0
2	2021-01-03	2.0	33.6	23.2	0.0	0.0
3	2021-01-04	2.0	34.4	23.7	0.1	1.9
4	2021-01-05	2.0	34.0	21.9	0.0	0.0

Figure 2: Sample of weather dataset

### The Results of the Dataset Split

The dataset used was divided into two 80% training and 20% testing. The visualization of the split is shown in Figure 3

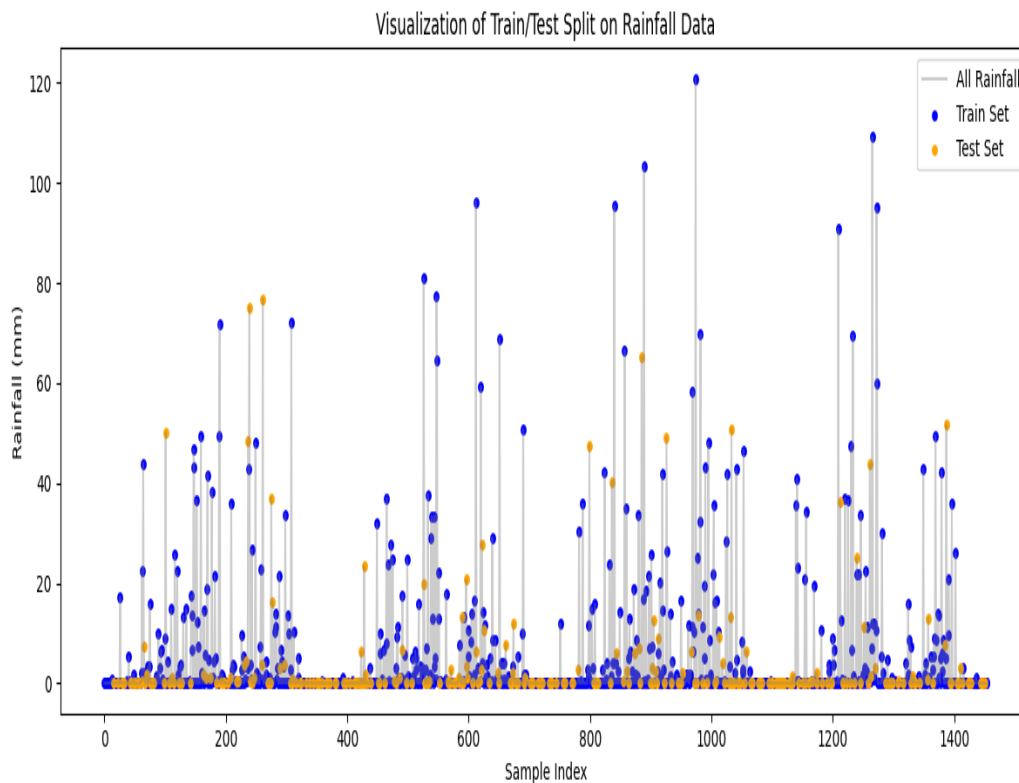


Figure 3: Data split

### Rainfall Predictive Model Developed for Nigeria

The result of the RF model developed for rainfall prediction in Nigeria demonstrates strong performance across key evaluation metrics, as seen in Table 1. These results indicate that the RF model effectively captures the underlying patterns in the meteorological data, with a relatively low error margin and a high degree of explained variance. An  $R^2$  value of 0.8601 suggests that approximately 86%

of the variability in rainfall can be accounted for by the model's predictors, which reflects a robust fit. This aligns with findings from Obisesan, who reported RF as the top-performing model for predicting meteorological variables in tropical regions, achieving high  $R^2$  values and low MAE across temperature, humidity, and wind speed forecasts (15).

Table 1: RF Model development result

1	MSE	18.2274
2	MAE	1.1334
3	$R^2$	0.8601

The scatter plots visualize how well each model's predictions align with the actual rainfall values in the test set. The red dashed line represents the ideal scenario where predicted rainfall exactly matches actual rainfall. Points that fall directly on this line indicate perfect predictions. Points above the red dashed line: These points represent instances where

the model over-predicted the rainfall, while the points below the red dashed line: These points represent instances where the model under-predicted the rainfall. The RF developed shows some spread around the line, indicating that while it captures the general trend, there are instances of both over- and under-prediction as seen in Figure 4

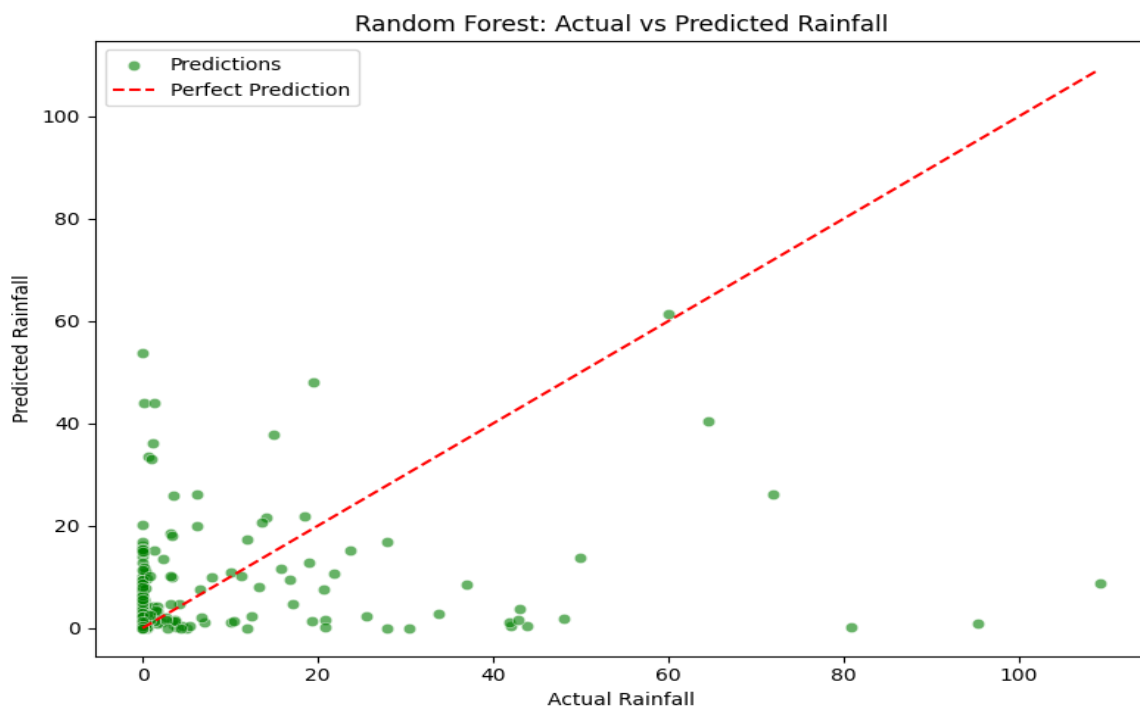


Figure 4: RF Scatter Plot

### XGBoost Regressor

The evaluation of the model for rainfall prediction revealed an MSE value of 20.59, an MAE value of 1.0391, and  $R^2$  value = 0.8420. This demonstrates a high level of predictive accuracy and model reliability. The MSE value of 20.59 indicates a low average squared error, suggesting the model effectively minimizes large prediction deviations.

The MAE value of 1.0391 reflects a small average absolute error, meaning predictions are, on average, just over 1 mm off actual rainfall values, and the  $R^2$  value of 0.8420 shows that 84.2% of the variance in rainfall is explained by the model's features, which is considered strong in environmental forecasting as seen in Table 2.

Table 2: XGBoost Regressor Evaluation

1	MSE	20.59
2	MAE	1.0391
3	$R^2$	0.8420

### The Scatter Plot

The scatter plots visualize how well each model's predictions align with the actual rainfall values in the test set. XGBoost: Similar to Random Forest, the points are somewhat spread out, suggesting reasonable but not perfect predictions as seen in Figure 5.

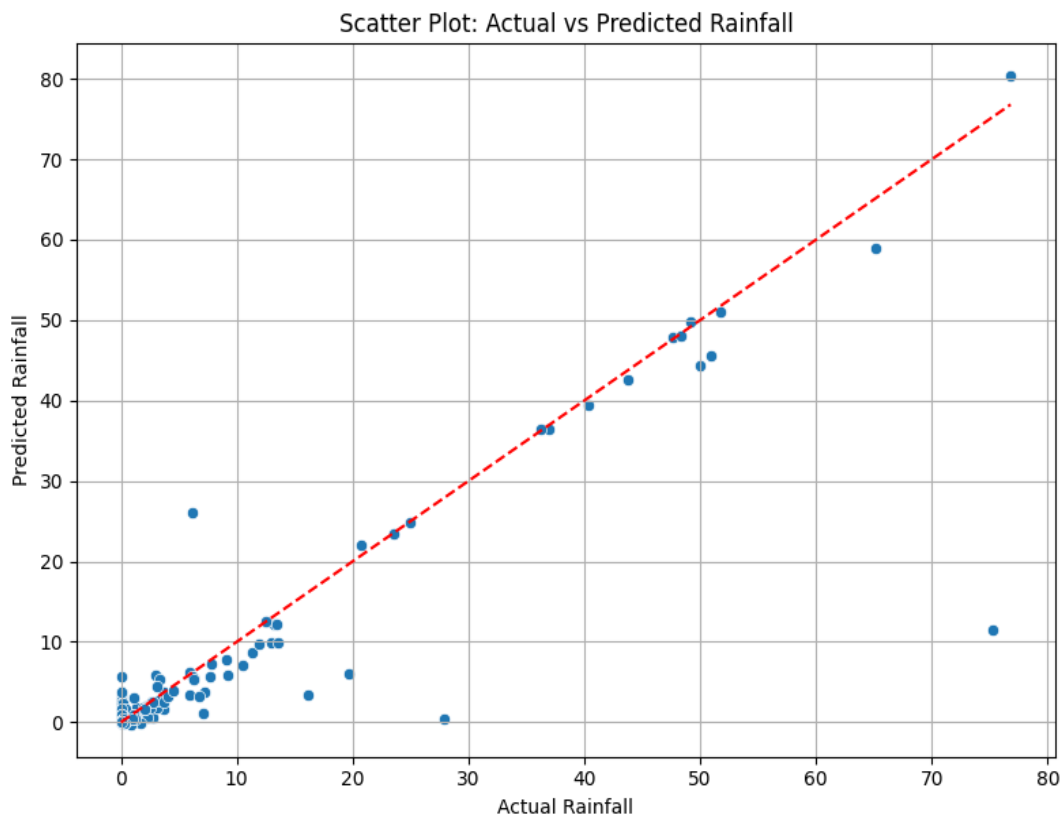


Figure 5: XGBoost scatter Plot

### SVR Model

The SVR model evaluated for rainfall prediction in Nigeria demonstrates excellent performance, with MSE = 13.8717, MAE = 0.6839, and  $R^2 = 0.8935$ .

These metrics reflect a highly accurate model with minimal error and strong explanatory power. The MSE value of 13.87 indicates a low average squared error, suggesting the model effectively

minimizes large deviations. The MAE value of 0.68 shows that predictions deviate from actual values by less than 1 mm on average, which is highly precise for meteorological forecasting.  $R^2$  value of

0.8935 reveals that nearly 89% of the variance in rainfall is explained by the model, underscoring its robustness and reliability as seen in Table 3.

Table 3: SVR Evaluation

1	MSE	13.87
2	MAE	0.68
3	$R^2$	0.8935

### The Scatter Plot

The scatter plots visualize how well each model's predictions align with the actual rainfall values in the test set. The points for the SVR model appear to be slightly more clustered around the line compared to Random Forest and XGBoost, suggesting potentially better performance as seen in Figure 6.

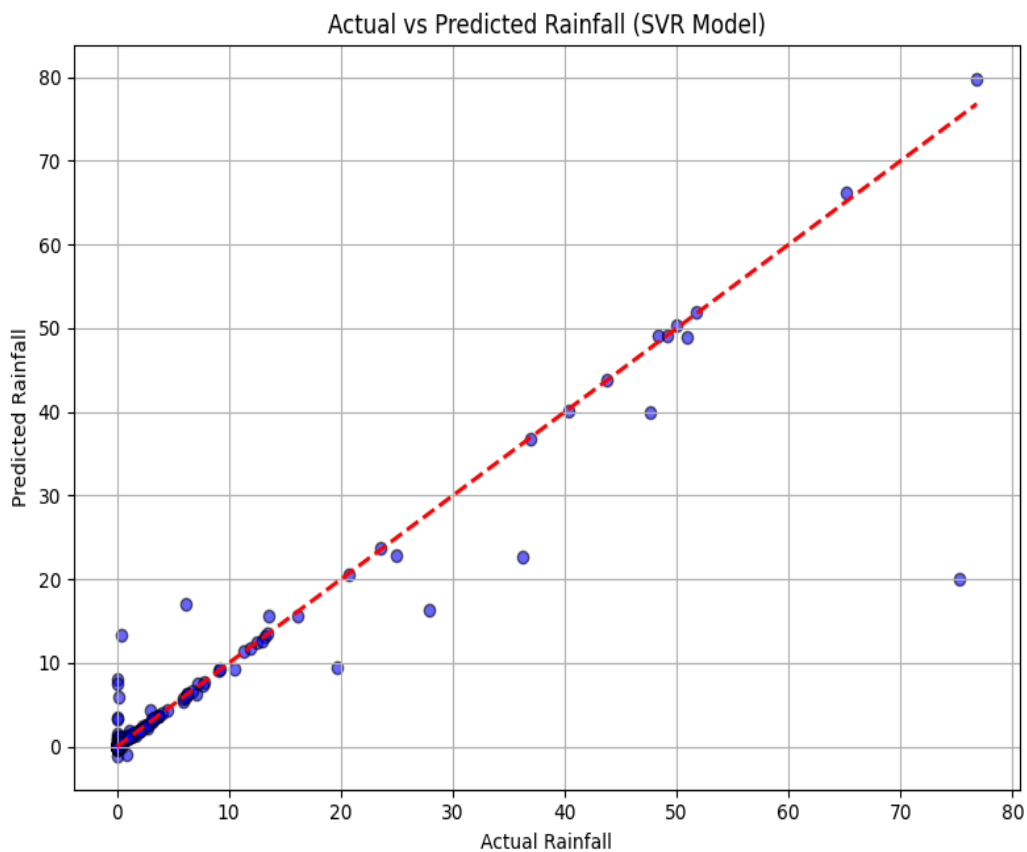


Figure 6: SVR Scatter Plot

### DNN Model

The DNN model evaluated for rainfall prediction in Nigeria demonstrates exceptional performance, with  $MSE = 0.2193$ ,  $MAE = 0.2868$ , and  $R^2 = 0.9983$ . These metrics reflect near-perfect accuracy, indicating that the model captures almost all variability in the rainfall data with minimal error. The  $MSE$  value of 0.2193 is extremely low,

suggesting the model rarely makes large prediction errors.  $MAE$  value of 0.2868 indicates that, on average, predictions deviate from actual rainfall values by less than 0.3 mm.  $R^2$  value of 0.9983 implies that 99.83% of the variance in rainfall is explained by the model, an outstanding fit as seen in Table 4.

Table 4 DNN Evaluation

1	MSE	0.2193
2	MAE	0.2868
3	$R^2$	0.9983

### The Scatter Plot

The scatter plots visualize how well each model's predictions align with the actual rainfall values in the test set. The points for the DNN model are very tightly clustered around the red dashed line, indicating a very high level of accuracy in its predictions. This reveals the  $R^2$ , which was very close to 1 for the DNN as seen in Figure 7.

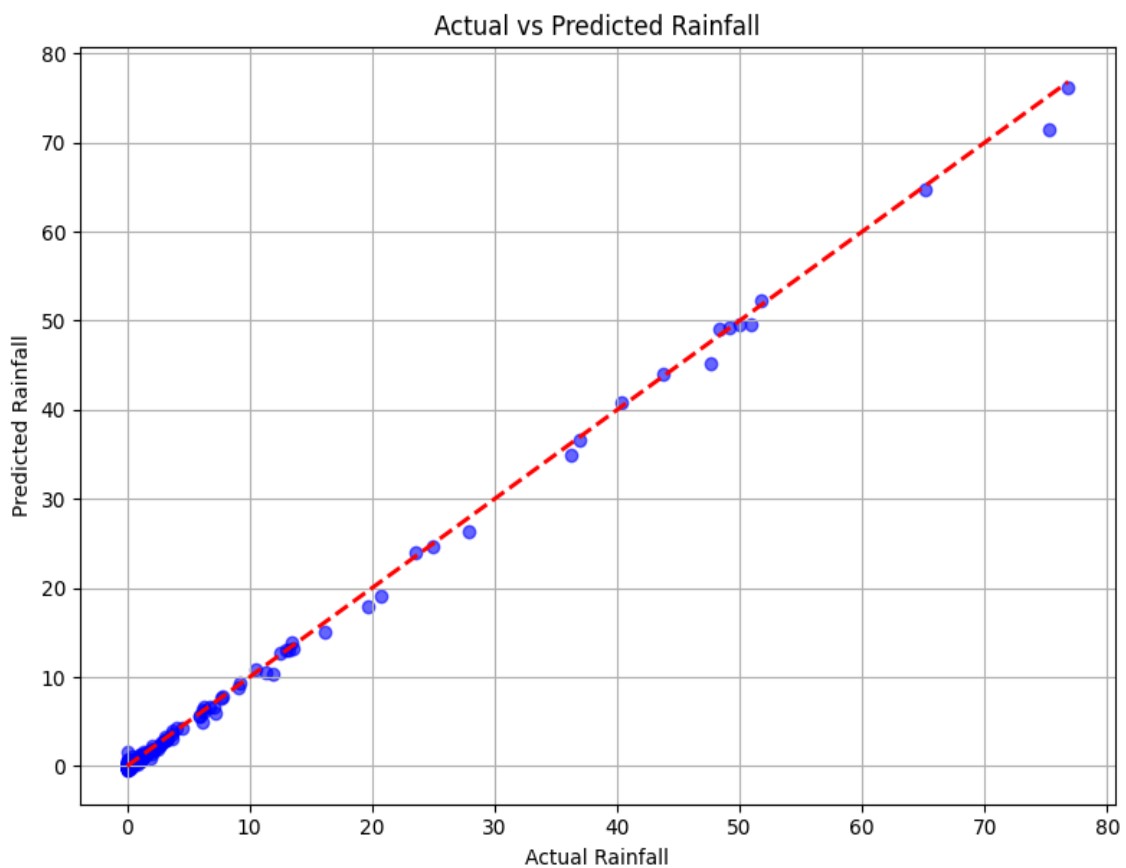


Figure 7: DNN scatter Plot

### The DNN training loss and Validation Loss

The plot shows the training and validation loss MSE for the DNN model over 50 epochs, where both training and validation loss decrease rapidly,

indicating that the model is learning. The overall loss values are very low by the end of training, indicating a good fit to the data, as seen in Figure 8

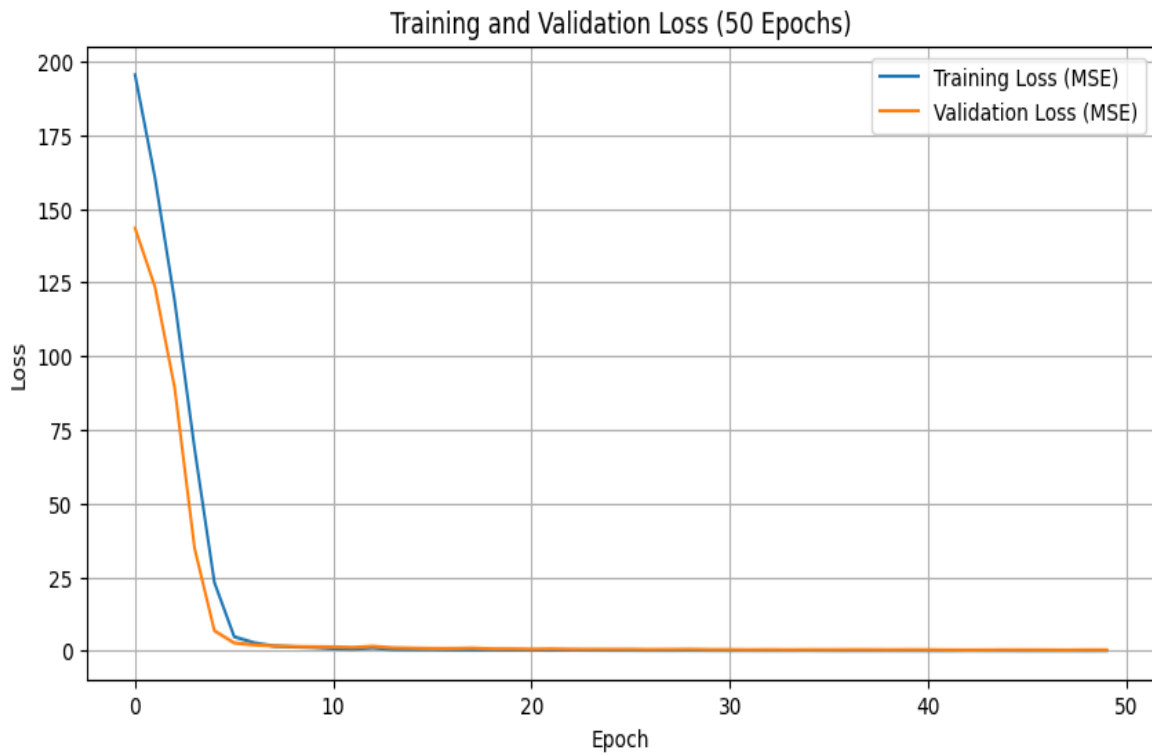


Figure 8: DNN training loss and Validation Loss

### The Correlation Heatmap of weather features

The correlation matrix helps to identify which features have strong linear relationships with each other and with the target variable (Rainfall). This information is valuable for understanding the data and can inform feature selection for the models. The Model reveals Positive values (closer to 1): the high positive correlation between MaxTemp and Evaporation (0.87) means that on hotter days, evaporation tends to be higher. The strong positive correlations among the Rainfall\_Lag features and Rainfall\_Rolling features with Rainfall show that

past rainfall and recent average rainfall are good indicators of current rainfall. Negative values (closer to -1), the negative correlations between Rainfall and Evaporation (-0.19) and MaxTemp (-0.27) suggest that higher temperatures and evaporation are weakly associated with lower rainfall. Values close to 0: Indicate a weak or no linear relationship. The correlation between MinTemp and RainRate (0.02) is very close to zero, suggesting almost no linear relationship between the minimum temperature and the rate of rainfall, as seen in Figure 9.

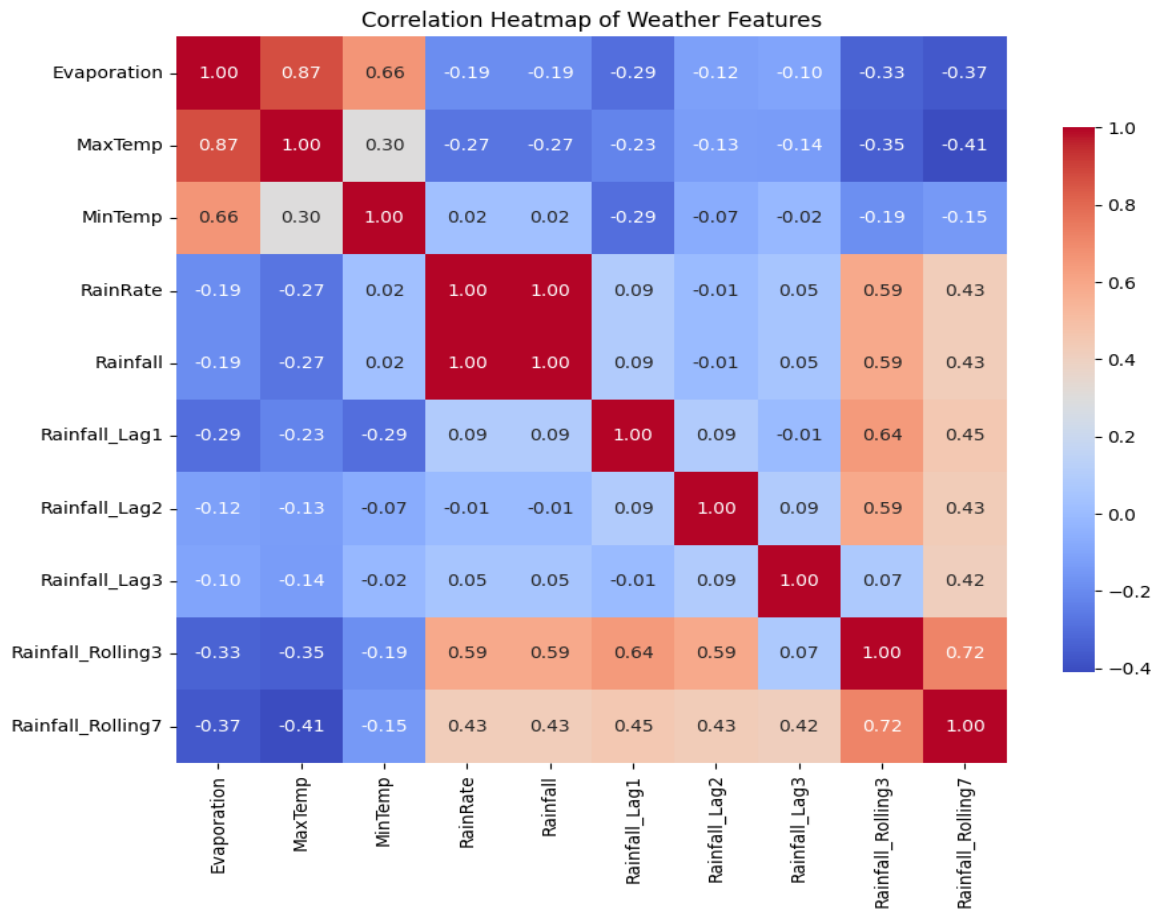


Figure 9: Correlation Heatmap of weather features

### Features Importance

Feature importance is a technique that assigns a score to input features based on how useful they are at predicting the target variable. A higher score means the feature has a greater impact on the model's predictions. The analysis confirms that temporal features like recent rainfall and rolling averages are highly influential in predicting rainfall, which is expected for time-series data. Rainfall\_Rolling3 is the most important feature, meaning the average rainfall over the past 3 days is

the strongest predictor of current rainfall. Rainfall\_Lag1 and Rainfall\_Lag2 are also very important, indicating that the rainfall from the previous one and two days significantly influences the prediction. MaxTemp and MinTemp have some importance, but less than the rainfall-related features. Rainfall\_Rolling7, Rainfall\_Lag3, and Evaporation have the lowest importance scores among the features included in this model, as seen in Figure 10.

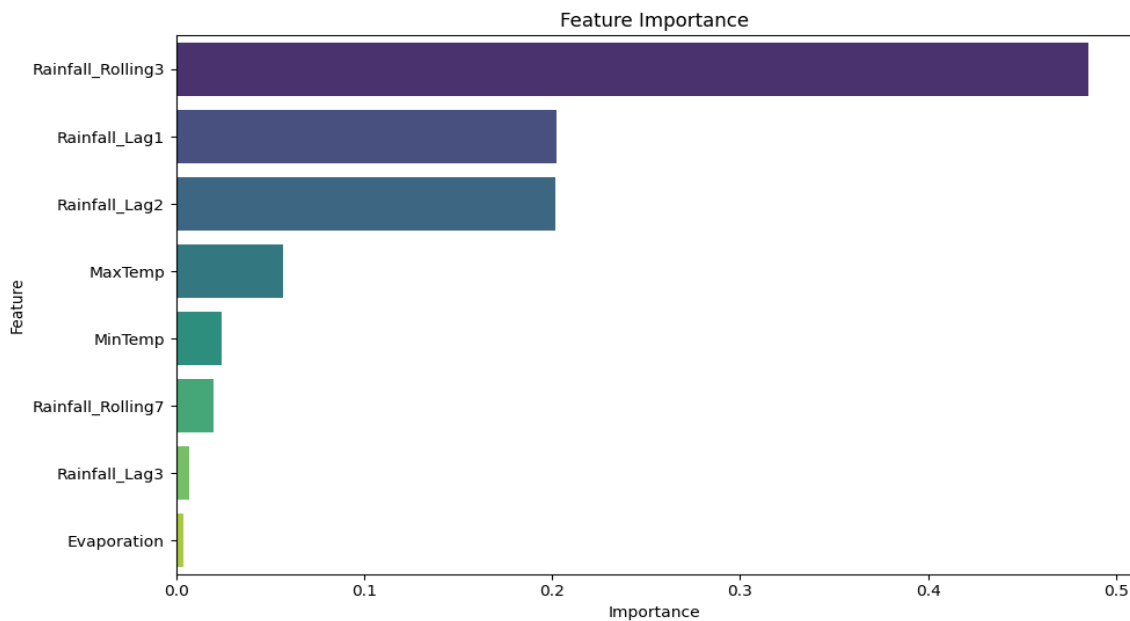


Figure 10: Feature Importance

### Prediction With DNN Model

The DNN model appears to be the best-performing model based on the evaluation metrics and scatter plot visualizations, suggesting it is the most suitable

for rainfall prediction in this context. The predicted rainfall values for the next day and today provide actionable insights for planning, as seen in Figure 11.

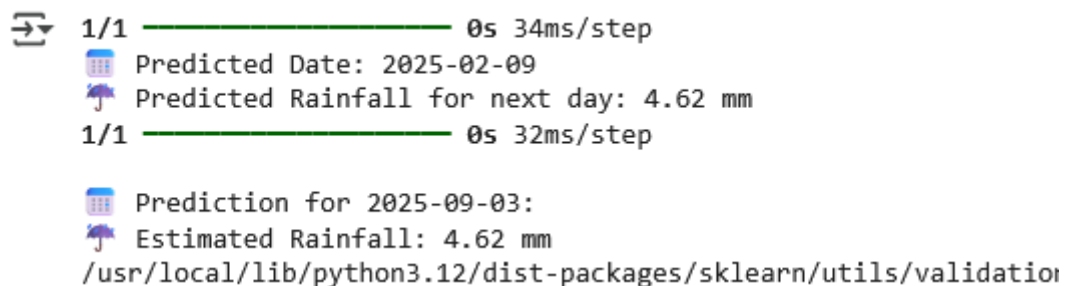


Figure 11: Rainfall prediction

### 4 Comparison of Model

The comparative evaluation of four machine learning models, which are RF, XGBoost, SVR, and DNN for rainfall prediction in Nigeria, reveals distinct performance strengths, with DNN emerging as the most accurate, as seen in Table 5. The metrics used include MAE, MSE, and the coefficient of determination ( $R^2$ ). DNN shows exceptional performance with the lowest MAE and MSE and an  $R^2$  of 0.9985, indicating near-perfect prediction accuracy. This aligns with recent findings by El Hafyani et al., who reported that

deep learning models, especially when optimized, outperform traditional ML algorithms in monthly precipitation forecasting. SVR also performs well, with an  $R^2$  of 0.8935 and relatively low error values (16). Lazri et al. found SVR to be highly effective for short-term rainfall estimation using satellite data, particularly in regions with nonlinear climatic patterns. XGBoost and RF deliver solid results, though slightly less accurate than SVR and DNN (17). In a comparative study carried out by Budiarto & Supriyanto, XGBoost and RF were top performers for daily rainfall prediction using

automatic rain gauge data, but their  $R^2$  values were notably lower (0.505 and 0.4485, respectively),

suggesting that model tuning and data quality significantly affect outcomes (18).

Table 5: Results of comparison of models developed

	MAE	MSE	$R^2$
RF	1.1334	18.2274	0.8601
XGBoost	1.0391	20.5891	0.8420
SVR	0.6839	13.8717	0.8935
DNN	0.2334	0.1956	0.9985

## 5 Conclusion

The study successfully developed and applied machine learning models to predict rainfall. The findings highlight the critical role of selecting and fine-tuning models for effective rainfall forecasting. While ensemble methods such as RF and XGBoost demonstrate consistent and dependable performance, SVR offers a commendable trade-off between simplicity and precision. Notably, DNNs emerge as the most powerful approach, indicating that deep learning should be at the forefront of future rainfall prediction systems, particularly in regions like Nigeria, where climatic patterns are highly complex and variable. This aligns with the broader comparative analysis by Kumar et al., who found that although RF and XGBoost are robust and interpretable, deep learning models, especially DNNs and LSTMs, are superior in modeling intricate temporal and nonlinear dynamics in rainfall data. These insights reinforce the present study's conclusion: DNNs significantly outperform other models in predictive accuracy. Future research should incorporate high-resolution satellite data and ground station records for hybrid modeling, and employ techniques such as SHAP, LIME, or attention mechanisms to enhance the interpretability of deep learning predictions for policymakers and meteorologists

**Conflict of Interest:** The authors declare that there are no conflicts of interest

**Funding:** NIL

## References

- Ojo OS, Ogunjo ST. Machine learning models for prediction of rainfall over Nigeria. *Scientific African*. 2022. <https://doi.org/10.1016/j.sciaf.2022.XXXX>
- Gulma UL, Hassan LU, Bala G. Analyzing and predicting rainfall patterns: A comparative analysis of machine learning models in Sokoto, Nigeria. *Int J Sci Res Arch*. 2024;13(1):121–126. <https://doi.org/10.30574/ijrsra.2024.13.1.1627>
- Khan MUS, Khan MS, Hussain A, Azamathulla HM. Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India. *Results Eng*. 2024;102093. Available from: <https://doi.org/10.1016/j.rineng.2024.102093>
- Aizansi AN, Ogunjobi KO, Ogou FK. Monthly rainfall prediction using artificial neural network: Case study of Republic of Benin. *Environ Data Sci*. 2024;3:e11. Available from: <https://doi.org/10.1017/eds.2024.10>
- Ridwan WM, Mohd N, Mohd N, Mohd Z, Mohd A. Evaluation of machine learning models

- for rainfall forecasting in Terengganu, Malaysia. *Heliyon*. 2020;6(11):e05481. Available from: <https://doi.org/10.1016/j.heliyon.2020.e05481>
- 6 Zhang Y, Li H, Wang J, Chen X. Hybrid deep learning model SCSSA-CNN-BILSTM for monthly rainfall prediction: Case study of Xi'an City, China (1996–2020). *Atmos Res*. 2024;290:107246. Available from: <https://doi.org/10.1016/j.atmosres.2024.107246>
- 7 Peeriga R, Rinku DR, Bhaskar JU, Nagalingam R, Aldosari FM, Albarakati HM, Alharbi AA, Jaffar AY. Real-time rain prediction in agriculture using AI and IoT: A Bi-Directional LSTM approach. *Eng Technol Appl Sci Res*. 2024;14(4):15805–15812. Available from: <https://doi.org/10.48084/etasr.8011>
- 8 Aiyelokun O, Ogunsanwo G, Pham QB, Salau O, Eguagie-Suyi P. Bias reduction of CFSR data with random forest for applications in water resources modelling in Ogun River. *Hydrol Sci J*. 2025;70(8):1308–1327.
- 9 Awad M, Khanna R. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. New York: Apress; 2015. [https://doi.org/10.1007/978-1-4302-5990-9\\_4](https://doi.org/10.1007/978-1-4302-5990-9_4)
- 10 Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowledge Discovery and Data Mining. 2016. p. 785–794. <https://doi.org/10.1145/2939672.2939785>
- 11 Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- 12 Mienye ID, Swart TG. A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*. 2024;15(12):755. <https://doi.org/10.3390/info15120755>
- 13 Hodson TO, Over TM, Foks SS. Mean squared error, deconstructed. *J Adv Model Earth Syst*. 2021;13:e2021MS002681. <https://doi.org/10.1029/2021MS002681>
- 14 Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021;7:e623. <https://doi.org/10.7717/peerj-cs.623>
- 15 Obisesan OE. Machine learning models for prediction of meteorological variables for weather forecasting. *Int J Environ Climate Change*. 2024;14(1):234–252. <https://doi.org/10.9734/IJECC/2024/v14i13829>
- 16 El Hafyani M, El Himdi K, El Adlouni SE. Improving monthly precipitation prediction accuracy using machine learning models: A multi-view stacking learning technique. *Front Water*. 2024;6:1378598. <https://doi.org/10.3389/frwa.2024.1378598>
- 17 Lazri M, Labadi K, Mohia Y, Absi R, Attaf Y, Alouache D. Comparison between SVR and SVM in rainfall estimation from remote sensing data. In: Kulkarni AJ, Cheikhrouhou N, editors. *Intelligent Systems for Smart Cities*. Singapore: Springer Nature; 2024. p. 295–312. Available from: [https://doi.org/10.1007/978-981-99-6984-5\\_19](https://doi.org/10.1007/978-981-99-6984-5_19)
- 18 Budiarto I, Supriyanto A. Performance comparison of manual and automatic rain gauge using XGBoost and Random Forest regression. *Electron J Educ Soc Econ Technol*. 2025;6(1). <https://doi.org/10.33122/ejeset.v6i1.955>