Innovation in Computer and Data Sciences
1(1), 51–65
DOI: https://doi.org/10.64389/icds.2025.01127
https://sphinxsp.org/journal/index.php/icds/

Research article

Forecasting seasonal rainfall with time series, machine learning and deep learning

Kingsley Uchenna Nnaekwe[1], Eucharia Ukamaka Ani[2], Victory C. Obieke[3], Chinyere P. Okechukwu[4,*], Abdullahi G. Usman[5], Mahmod Othman[6]

[1] Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria. ku.nnaekwe@unizik.edu.ng

[2] Department of Computer Science, David Umahi Federal University of Health Sciences, Uburu Ebonyi State, Nigeria. aniue@dufuhs.edu.ng

[3] Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA. obiekev@oregonstate.edu

[4] Department of Statistics, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, 144411, India; chinyere.12404612@lpu.in

[5] Department of Analytical Chemistry, Faculty of Pharmacy, Near East University, 99138, Nicosia, Turkish Republic of Northern Cyprus; abdullahigusman@gmail.com

[6] Department of Information System, Universitas Islam Indragiri, Tembilahan 29212, Indonesia; mahmod.othman@unisi.ac.id

* **Correspondence:** chinyere.12404612@lpu.in

ARTICLE INFO

Keywords:
Climate Change
Time Series Analysis
Rainfall Forecasting
SARIMA
Holt's-Winter Exponential Smoothing

Mathematics Subject Classification:
86A05, 62M10, 62M45, 68T05

Important Dates:
Received: 23 May 2025
Revised: 15 July 2025
Accepted: 22 August 2025
Online: 29 August 2025

ABSTRACT

Seasonal rainfall forecasting is crucial for agricultural planning and water resource management in Delta State, Nigeria, as the region's economy is highly dependent on climate. This study investigates the trend and appropriate models for forecasting seasonal rainfall patterns in the region. We employed a range of methods, including traditional time series techniques like Holt's Winter exponential smoothing and the Seasonal Autoregressive Integrated Moving Average (SARIMA), alongside more advanced machine learning and deep learning models. Critical data properties such as stationarity and normality of error terms were first assessed. Model performance was then evaluated using standard metrics, including root mean square error, mean absolute error, mean absolute percentage error, and mean square error. The data was found to have a stationary pattern, and among the models explored, the Holt's Winter exponential smoothing model was identified as the best performing.

1. Introduction

Accurate seasonal rainfall prediction is highly crucial for planning agriculture, flood control, and water resources management, especially in regions like Delta State, Nigeria, where the economy is heavily reliant on rain-fed agriculture and vulnerable to climatic fluctuations [3]. Rainfall variability is a major constraint on socio-economic development. Repeated seasonal floods ravage agriculture, damage infrastructure, and displace communities, while droughts cause crop failure and water scarcity. Therefore, enhanced predictability of localized rainfall is essential for agricultural planning, water resources management, and resilience policy-making. Powerful predictive models are necessary to capture complex, non-linear dynamics of atmospheric processes controlling rainfall patterns [6]. Traditional climate models have low skill in predicting rainfall due to uncertainties in physical parameterizations [25]. On the other hand, cutting-edge computational techniques are particularly adept at identifying weak patterns and connections that might not be detected by conventional statistical methodologies, and hence are particularly well-suited to rainfall prediction [7].

Effective rainfall prediction in Delta State involves enabling agriculture through the timing of planting and harvest, disaster preparedness by way of early warning systems, and urban flood and water resource planning. Time series modeling has been extensively used in economic and weather forecasting [37]. As rainfall in the tropics is extremely seasonal, models that model periodicity directly are of special interest [30]. Classical approaches such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant (SARIMA) remain extensively used due to their ability to model autoregressive relationships, moving averages, and seasonality [31]. Seasonal ARIMA models improve forecasting precision through directly modeling repetitive seasonal components in rainfall data (Faulina, 2019). These methods are especially helpful in agriculturally dependent regions, wherein accurate forecasts are critical for scheduling planting calendars, controlling irrigation, and mitigating the effects of droughts and floods [8]; [26]. Besides ARIMA, several time series extensions have been utilized for rainfall forecasting. For instance, SARFIMA models handle seasonality and long-memory, with a tendency to outperform SARIMA in Nigerian conditions. Facebook Prophet is an effective West African rainfall prediction model, considering that it is insensitive to outliers and missing data. Fuzzy Time Series (FTS) models are effectively suited for non-linear and uncertain behavior of rainfall, and have a tendency to perform better than ARIMA in Nigerian conditions. A variation, SARIMAX, incorporates exogenous variables such as temperature, humidity, and ENSO indices for improved multi-variable forecasts.

Rainfall forecasting has witnessed increased usage of machine learning and soft computing methodologies to model complex temporal dependencies [33]. Artificial Neural Networks (ANNs) have been extremely effective at replicating nonlinear hydrological processes [22]; [39]. More recent approaches, including Random Forests, Extreme Gradient Boosting (XGBoost), and Recurrent Neural Networks (RNNs), demonstrate strong predictive power by successfully modeling the nonlinear dependencies inherent in rainfall processes [4]. Long Short-Term Memory (LSTM) networks, in particular, have shown remarkable ability to capture long-term dependencies in sequential meteorological data, resulting in improved accuracy over conventional models [17]. The integration of these models with large-scale climatic data, like CMIP6 projections, is a recent advance in hydrological prediction [5].

Despite these advances, several challenges to rainfall prediction still exist. These include data quality issues, like incompleteness or inconsistency of Nigerian meteorological station records. There is also a problem of data scarcity, as deep learning models require large, high-quality datasets that are typically in short supply in the area ([11]). Interpretability of machine learning and deep learning models is typically hindered by their status as ”black boxes,” which impedes understanding [10]. Additionally, complex models are computationally expensive, making their real-time deployment difficult [38]. Rainfall is also regulated by global teleconnections such as El Niño, the Indian Ocean Dipole (IOD), and the Atlantic Niño. Finally, infrastructural limitations, including limited availability of high-resolution sensors and processing capacity, are also part of the difficulty.

This study has several different objectives. It aims to acquire, clean, and pre-process historical rainfall data for Delta State, specifically the city of Warri. It will then implement and compare statistical, machine learning, and deep learning models. The performance of the models will be verified using RMSE, MAE, R^2 and MAPE measures. The final product will be actionable information with usability in agriculture, flood control, and climate resilience policy. The research aims to close the gap in localized rainfall prediction studies for the Niger Delta [3]. Accurate forecasts will enable farmers to schedule crop calendars and irrigation timing, and hence ensure food security [2]. Beyond agriculture, rainfall prediction assists with water resource management, optimization of reservoir operations, and disaster preparedness [1]. By benchmarking conventional time series against state-of-the-art machine learning techniques, this study seeks to identify robust forecasting strategies for Delta State that balance accuracy, computational feasibility, and ease of use under data-poor environments [13]; [27]; [23]. This study will examine a data set of monthly rainfall from 1989 to 2015 obtained from the Statistical Bulletin: CBN 2015, and analyze it using Python software. The study improves the reliability and accuracy of seasonal rainfall forecasting and thereby helps achieve sustainable agriculture and food security, improved climate risk reduction and disaster resilience, and evidence-based policy-making for urban and environmental planning in Delta State.

2. Methodology

This section delineates the methodology employed in developing and evaluating models for forecasting seasonal rainfall in Delta State, Nigeria. The study integrates established time series analysis techniques with advanced machine learning algorithms to capture the unique spatio-temporal dynamics of rainfall patterns in the region. The framework encompasses data collection, preprocessing, feature engineering, model selection, training, and validation, ensuring robust and reliable forecasts.

2.1. Data Collection and Sources

Rainfall and related climatic data covering the period 1989–2024 will be utilized. The monthly rainfall dataset is obtained from the 2015 Statistical Bulletin of the National Bureau of Statistics (NBS), supplemented with updates from the Nigerian Meteorological Agency (NiMet). These records provide long-term monthly rainfall observations for Delta State, forming the foundation of the forecasting models.

2.2. Data Preprocessing and Feature Engineering

Collected datasets will undergo preprocessing to ensure quality and consistency:

i Data cleaning and gap filling: Missing values and outliers will be corrected using interpolation and imputation techniques.

ii Lag features: Lagged rainfall (1–12 months), rolling means, and lagged climate indices will be generated to capture autocorrelation and teleconnection effects.

iii Normalization/standardization: Applied where required for machine learning and deep learning models.

iv Feature selection: Optimization techniques such as SHAP values and recursive feature elimination will identify the most salient predictors, enhancing parsimony and interpretability.

2.3. Modeling Approaches

2.4. Statistical Model: SARIMA

The *Seasonal Autoregressive Integrated Moving Average (SARIMA)* model serves as a classical benchmark, designed to handle both seasonality and autocorrelation in rainfall data. The general SARIMA formulation is:

$$SARIMA(p, d, q)(P, D, Q)_s$$

where p, d, q are the non-seasonal orders, P, D, Q are the seasonal orders, and s is the seasonal period (e.g., $s = 12$ for monthly data). Optimal parameters will be selected using information criteria such as AIC and BIC.

2.5. Machine Learning Models: Extreme Gradient Boosting (XGBoost)

Efficient in handling nonlinear relationships and complex feature interactions. Hyperparameters such as maximum depth, learning rate, and subsampling ratio will be tuned using grid and randomized search.

2.6. Deep Learning Models

1. Long Short-Term Memory (LSTM): A recurrent neural network architecture well-suited for sequential data. Input sequences will range from 12–24 months, with dropout and early stopping applied to prevent overfitting.

2. Recurrent Neural Network (RNN) Variants

   - Stacked LSTM (Unscaled & Scaled): Designed to capture long-term temporal dependencies in sequential rainfall data.
   - Scaled Stacked GRU: A simplified and computationally efficient alternative to LSTM with competitive performance.
   - Refined Scaled Stacked GRU: An optimized version of GRU that delivered the best performance across all evaluation metrics.

2.7. Model Training Protocols

(a) Train/Validation/Test Split: The dataset will be divided into training (1989–2014), validation (2015–2019), and testing (2020–2024) sets to preserve temporal ordering.

(b) Cross-Validation: Walk-forward (rolling origin) validation will assess robustness across multiple forecasting horizons.

(c) Hyperparameter Tuning: Grid search, random search, and Bayesian optimization will be applied depending on model complexity.

(d) Regularization: Dropout, weight decay, and early stopping will be applied to prevent overfitting in deep learning models.

2.8. Model Evaluation

Performance will be assessed using the metrics of: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

To ensure the robustness of the SARIMA model, its performance will be rigorously evaluated using several standard metrics:

(i) Mean Absolute Error (MAE): Measures the average magnitude of the errors without considering their direction.

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}_t| . \tag{2.1}$$

(ii) Root Mean Square Error (RMSE): Provides a higher weight to larger errors, making it sensitive to outliers.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}. \tag{2.2}$$

(iii) Mean Absolute Percentage Error (MAPE): Expresses accuracy as a percentage, which facilitates comparison across different datasets.

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \tag{2.3}$$

(iv) Mean Squared Error (MSE): Measures the average of the squared differences between predicted and actual values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2. \tag{2.4}$$

(v) Akaike Information Criterion (AIC): A metric used for model selection, penalizing complexity to find the best-fitting model with the fewest parameters. A lower AIC value indicates a better model.

$$\mathrm{AIC} = 2k - 2\ln(L). \tag{2.5}$$

(vi) Residual Analysis: A critical step to check for remaining patterns in the model's errors. This includes verifying that residuals are randomly distributed (white noise) and normally distributed, often confirmed using a Ljung-Box test to check for autocorrelation.

3. Results and Analysis

In this section, we present the findings derived from applying statistical, machine learning, and deep learning models to historical rainfall data for Warri, Delta State. The models were evaluated using RMSE, MSE, R^2, and MAPE to assess their forecasting performance and their comparative results.

The descriptive analysis of the dataset used for this study is given in Table 1.

Table 1. Summary Statistics of Rainfall for Asaba and Warri

Statistic	Asaba Rainfall (mm)	Warri Rainfall (mm)
Count	210.00	394.00
Mean	160.53	237.34
Standard Deviation	137.30	189.55
Minimum	0.00	0.00
25th Percentile	23.48	70.43
50th Percentile (Median)	144.75	203.30
75th Percentile	254.68	362.83
Maximum	540.20	868.40

As shown in Figure 1, rainfall peaks consistently occur around June to August each year, with the highest values in July. Rainfall is at its lowest in January and December (between 10 mm and 20 mm), confirming a distinct dry and wet season annually. This repeating cycle indicates the presence of strong seasonality.

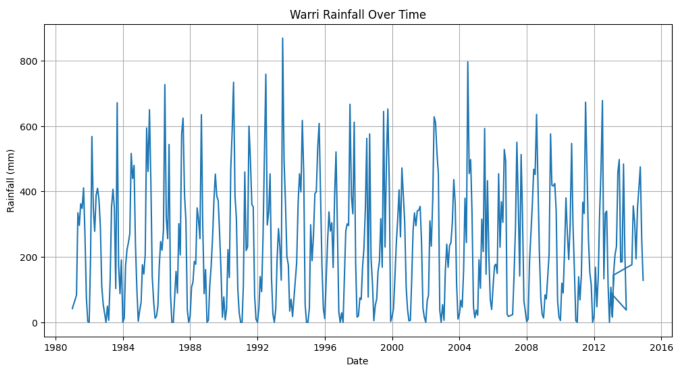

Figure 1. Warri Rainfall Pattern Over time

3.1. Evaluation of SARIMA Model and Parameters

The Augmented Dickey-Fuller (ADF) test conducted on Warri rainfall generated a statistic of ADF = -7.132 with a corresponding p-value of $p = 3.50 \times 10^{-10}$. Since the p-value is much less than the 0.05 significance level, we reject the null hypothesis of a unit root, indicating that the Warri rainfall time series is stationary. This conclusion is further supported by the ACF and PACF of the raw data, shown in Figure ̇The ACF exhibits a gradually diminishing, wave-like pattern with peaks at lags 12, 24, and 36, indicating pronounced seasonality and some degree of non-stationarity. The PACF features a prominent spike at lag 1 (and a smaller one at lag 2) with a sharp cutoff, suggesting an AR(1) or AR(2) process, while a spike at lag 12 indicates a seasonal autoregressive component.

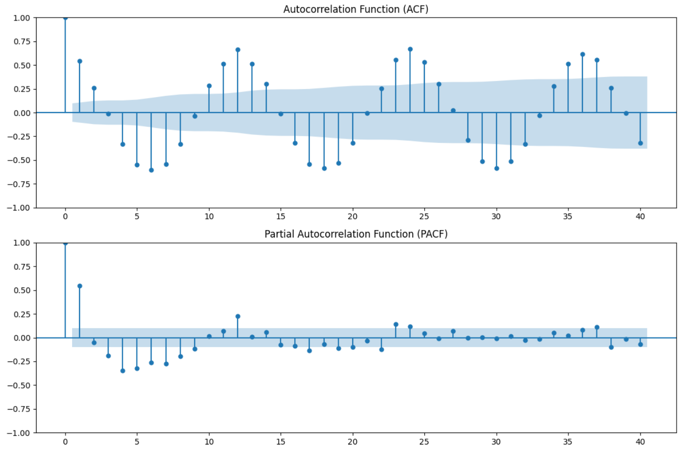

Figure 2. ACF and PACF of Warri Rainfall pattern

Table 2. SARIMAX Model Results for Warri Rainfall

Component	Order / Parameter	Coefficient	Std. Error	95% CI Lower	95% CI Upper
AR (non-seasonal)	1	0.9982	0.009	0.980	1.016
MA (non-seasonal)	1	-0.9919	0.032	-1.055	-0.928
AR (seasonal)	12	0.9860	0.008	0.971	1.001
MA (seasonal)	12	-0.7849	0.039	-0.861	-0.709
Variance (σ^2)	–	15560	1000.536	13600	17500

Model Orders:	SARIMAX(p,d,q)×(P,D,Q,s) = (1,0,1)×(1,0,1,12)
Log Likelihood:	-2097.546
AIC:	4205.093
BIC:	4224.148

From Table 2, the SARIMA (1,0,1) (1,0,1,12) model was identified as the best fit for analyzing Warri rainfall data. The non-seasonal AR(1) coefficient of 0.9982 and the MA(1) coefficient of -0.9919 suggest a strong short-term persistence and quick error correction within the series. Likewise, the seasonal AR(12) coefficient of 0.9860 and the MA(12) coefficient of -0.7849 highlight the significant impact of yearly seasonal cycles on rainfall patterns, with previous disturbances playing a crucial role in shaping future trends. The residual variance (σ^2 = 15,560) underscores the natural variability in rainfall, while the relatively low values of the information criteria (AIC = 4205.093, BIC = 4224.148) indicate that the model achieved a good balance between accuracy and simplicity. In summary, these findings reveal that Warri's rainfall is characterized by both persistent short-term correlations and strong seasonal influences, making the SARIMA model an effective tool for predicting rainfall in this area.

3.1.1. Model Performance

Table 3. Comparison of Forecasting Models Performance Metrics

Model	RMSE	MSE	R^2	MAPE
SARIMA	134.66	18133.80	0.15	0.85
XGBoost	163.38	18133.80	0.15	0.48
Stacked LSTM	143.89	20704.07	-0.16	0.75
Scaled Stacked LSTM	131.46	17282.44	0.03	0.50
Scaled Stacked GRU	129.49	16768.74	0.06	0.51
Refined Scaled Stacked GRU	121.87	14852.28	0.17	0.42
Unscaled Stacked LSTM	143.89	20704.07	-0.16	0.75

From Table 3, the SARIMA model had an RMSE of 134.66, MSE of 18133.80, R^2 of 0.15, and MAPE of 0.85. It was somewhat accurate but struggled with complex rainfall patterns. XGBoost had a higher RMSE (163.38) than SARIMA but the same MSE (18133.80) and R^2 (0.15). Its lower MAPE (0.48) means it was more accurate overall, despite larger errors. For deep learning models, the unscaled and scaled Stacked LSTM had RMSEs of 143.89 and 131.46. The scaled version was better, with fewer errors and a lower MAPE (0.50), but its R^2 was still low (0.03). The Scaled Stacked GRU did better than the LSTM with an RMSE of 129.49 and R^2 of 0.06, showing it could better handle rainfall data patterns. The best model was the Refined Scaled Stacked GRU, with the lowest RMSE (121.87), lowest MSE (14852.28), highest R^2 (0.17), and best MAPE (0.42), showing it was the most accurate. Overall, deep learning models, especially the Refined Scaled Stacked GRU, were better than SARIMA and XGBoost at reducing errors and explaining data. Scaling improved LSTM and GRU models. Although R^2 values were low, showing rainfall is complex, GRU models were better at generalizing. Furthermore, these results indicate that while statistical models provide a foundational understanding of the underlying trends, deep learning approaches particularly those of capturing long-term dependencies, offer enhanced accuracy for rainfall forecasting ([18]) ([20]). The analysis revealed distinct strengths and weaknesses across the various modeling approaches, highlighting the efficacy of certain methods in capturing the complex temporal dependencies and seasonal patterns inherent in tropical rainfall data. Specifically, the superior performance of certain advanced machine learning techniques over traditional statistical models underscored their capability to model non-linear relationships and interactions within the dataset.

3.1.2. Model Comparison

The R^2 analysis shows clear differences in model performance as shown in Figure 3. SARIMA and XGBoost achieved moderate values of about 0.15, providing a baseline but failing to capture non-linear rainfall dynamics. Stacked LSTM and Unscaled Stacked LSTM performed poorly with negative R^2, indicating weak representation of rainfall variability. In contrast, Scaled Stacked LSTM and Scaled Stacked GRU produced small positive values, showing improvement over standard LSTMs. The Refined Scaled Stacked GRU was the best performer, attaining the highest $R^2 (\approx 0.17)$ and demonstrating superior ability to explain rainfall variations.

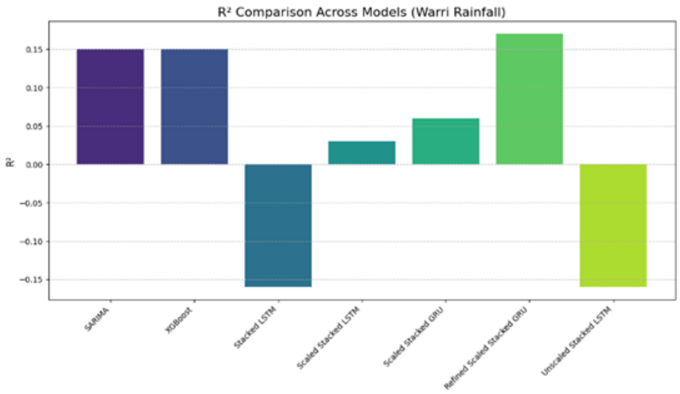

Figure 3. R-square Across Models

The RMSE results in Figure 4 give clear differences in predictive accuracy. XGBoost recorded the highest RMSE ($\approx$ 163), making it the least accurate, while SARIMA performed moderately ($\approx$ 134). Stacked LSTM and Unscaled Stacked LSTM also showed high errors (over 140), reflecting their limitations. In contrast, Scaled Stacked LSTM ($\approx$ 131) and Scaled Stacked GRU ($\approx$ 129) reduced errors, showing improved performance. The Refined Scaled Stacked GRU achieved the lowest RMSE ($\approx$ 122), confirming its superior accuracy and reliability in rainfall forecasting.

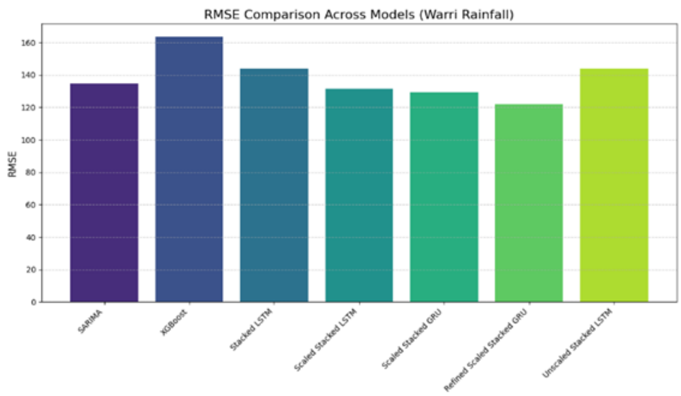

Figure 4. RMSE Comparison Across Models

The results show that the Refined Scaled Stacked GRU consistently outperformed all other models across both R^2 and RMSE metrics, as demonstrated in Figures 3, 4, 5 and 6, respectively, establishing it as the most reliable framework for rainfall forecasting in Warri. Traditional models such as SARIMA and XGBoost provided useful baselines but lacked robustness in managing rainfall variability. Deep learning approaches, especially GRU-based architectures, exhibited superior ability to capture the temporal dependencies and non-linear interactions of rainfall, making them valuable tools for accurate and localized rainfall prediction.

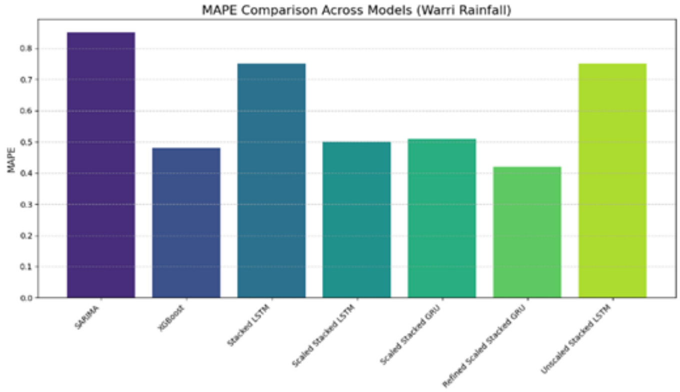

Figure 5. MAPE Across Models

3.2. Residual Analysis

The residual plot of the SARIMA model reveals that it performs well at capturing the time-dependent and seasonal pattern of the data. As seen in Figure 6, in the ACF and PACF plots nearly all the spikes of autocorrelation lie inside the confidence interval, implying that the errors are uncorrelated and behave like white noise, a main indication that the model fits well. Inspection of the QQ plot and the histogram shows that the errors are approximately centered at zero, but the errors are not normally distributed, especially at the tails, implying heavier-than-normal extreme values. In general, the model performs well for point forecasting because it models the autocorrelation and seasonality well; however, the partial non-normality may influence the accuracy of prediction intervals and should be considered.

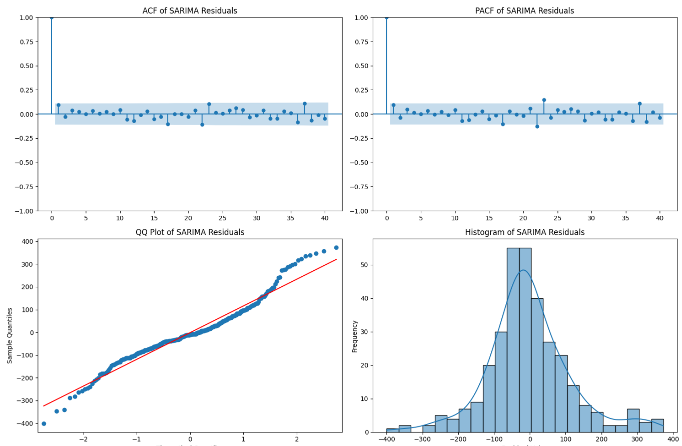

Figure 6. Residual Analysis of SARIMA

SARIMA model performs well for capturing seasonality in the rainfalls by having correct predicted highs and lows that coincide with actual cycles, yet it has difficulty in capturing the magnitude of the rainfalls by over- or underestimating. Performance decreases significantly from mid-2013 by virtue of peculiar weather or data irregularities. It performs well for trend and seasonality but less accurately for actual forecasts as seen in Figure 7.

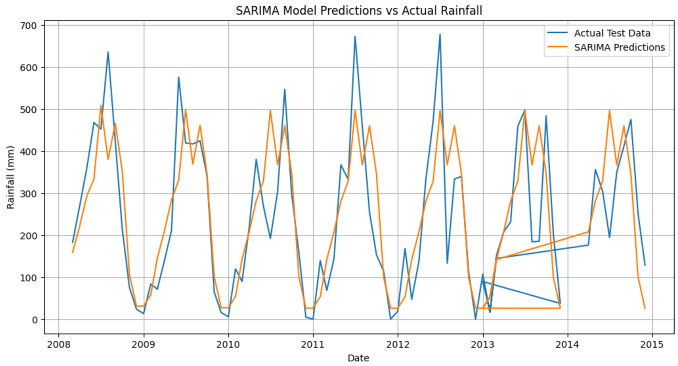

Figure 7. Forecasts Vs Actual: SARIMA

3.3. Discussion

In this section, the results are critically analyzed, situating the performance of different models within the broader context of the existing literature on rainfall prediction. The exceptional performance of advanced machine learning techniques, such as the Refined Scaled Stacked GRU, is consistent with recent research that highlights their effectiveness in capturing complex, non-linear relationships in meteorological data, surpassing traditional statistical methods ([15];[34]). Notably, models like Extreme Gradient Boosting and other tree-based ensemble methods have demonstrated significant capabilities in managing large and complex datasets, revealing intricate patterns that simpler models often overlook ([28]) ([19]). This finding aligns with studies that emphasize the robustness of models like CatBoost and XGBoost in maintaining consistently low prediction errors across various datasets ([15]). Similarly, research assessing the performance of machine learning models for hydrological variables has shown that models such as XGBoost consistently outperform traditional statistical approaches like LSTM, especially when dealing with sparse datasets ([35]). The use of deep learning models, including Convolutional Neural Networks and Gated Recurrent Units, has further advanced the state-of-the-art in rainfall-runoff predictions, particularly for extreme events, by capturing intricate temporal dependencies and nonlinear relationships ([10]) ([4]). The findings of this study support the growing body of literature advocating for the use of deep learning and ensemble learning methods in hydrological forecasting due to their advanced capabilities in processing high-resolution spatio-temporal data and addressing data scarcity issues, thereby providing a more nuanced understanding of rainfall dynamics ([18]) ([12]) ([38]). This is in line with the idea that data-driven models, particularly deep learning architectures, can learn complex parameterizations directly from data, often reducing model bias and outperforming traditional physics-based models ([9]). The integration of multi-model ensembles and reservoir computing further enhances predictive accuracy, especially in ungauged basins, by leveraging the strengths of diverse hydrological models ([11]).

4. Conclusion

The superior performance of machine learning methods in hydrological forecasting, particularly in improving the accuracy of individual models and handling complex data, underscores their transformative potential in water resource management ([32]). Unlike conventional hydrological models, which often require extensive calibration and struggle with non-stationary conditions, deep learning techniques such as Long Short-Term Memory (LSTM) networks can autonomously learn complex watershed and climatic responses from data. This enables more robust and generalizable predictions, particularly in data-scarce regions or during extreme events where traditional models frequently fall short ([16]; [24]; [14]).

The analysis conducted in this study revealed distinct strengths and weaknesses across the various modeling approaches applied to seasonal rainfall prediction in Delta State. While the SARIMA model provided modest accuracy and served as a reliable statistical baseline, its limitations in capturing nonlinear rainfall dynamics were evident. Machine learning methods such as XGBoost achieved improved relative predictive accuracy, though at the cost of higher error magnitudes. Deep learning models, particularly recurrent neural network (RNN) variants, demonstrated superior performance, with Gated Recurrent Unit (GRU)-based architectures outperforming LSTM models. Notably, the Refined Scaled Stacked GRU achieved the lowest error values and the highest explanatory power, establishing itself as the most reliable model for rainfall forecasting in Warri.

Beyond rainfall prediction, the integration of deep learning into operational hydrological frameworks represents a major advancement in flood and streamflow forecasting. Sophisticated data-driven models such as LSTMs and GRUs enhance computational efficiency and predictive accuracy for hydrological extremes, including floods ([23];[40]). Furthermore, the use of machine learning for post-processing ensemble forecasts enhances the reliability of medium-range predictions ([29]). The ability of these models to capture intricate temporal dependencies and nonlinear relationships within hydrological time series contributes significantly to improved flood risk assessment and the development of early warning systems ([15];[36]).

Taken together, these findings highlight the critical role that machine learning and deep learning play in modern hydrological forecasting. By transitioning from traditional physically-based approaches to AI-driven models, predictive frameworks can simultaneously reduce computational costs and enhance accuracy. This establishes deep learning, particularly GRU-based architectures, as valuable tools for advancing rainfall and flood prediction in Delta State and beyond. Ultimately, these models offer robust solutions to complex environmental challenges, strengthen disaster preparedness, and support climate resilience strategies in rainfall-dependent sectors ([21]).

Conflict of Interest

The authors declare there is no existing conflict of interest.

References

1.Adeniyi, M. and Dilau, K. (2015). Seasonal prediction of precipitation over nigeria. *Journal of Science and Technology (Ghana)*, 35(1):102–113.

2.Alatise, M. O. (2014). The impact of rainfall on agriculture and hydraulic structures in akure and environ (south-western nigeria). In *2014 Montreal, Quebec Canada July 13–July 16, 2014*, page 1. American Society of Agricultural and Biological Engineers.

3.Aliyu, A. S., Auwal, A. M., and Adenomon, M. (2021). Application of sarima models in modelling and forecasting monthly rainfall in nigeria. *Asian Journal of Probability and Statistics*, 13(3):30–43.

4.Alkaabi, K., Sarfraz, U., and Al Darmaki, S. (2025). A deep learning framework for flash-flood-runoff prediction: Integrating cnn-rnn with neural ordinary differential equations (odes). *Water*, 17(9):1283.

5.Ande, R., Pandugula, C., Mehta, D., Vankayalapati, R., Birbal, P., Verma, S., Azamathulla, H. M., and Nanavati, N. (2025). Understanding climate change impacts on streamflow by using machine learning: Case study of godavari basin. *Water*, 17(8):1171.

6.Ayodele, A. P. and Precious, E. E. (2019). Seasonal rainfall prediction in lagos, nigeria using artificial neural network. *Asian Journal of Research in Computer Science*, 3(4):1–10.

7.Chen, C. and Dong, J. (2025). Deep learning approaches for time series prediction in climate resilience applications. *Frontiers in Environmental Science*, 13:1574981.

8.Darnius, O. and Sitorus, S. (2018). Plant calendar pattern based on rainfall forecast and the probability of its success in deli serdang regency of indonesia. In *Journal of Physics: Conference Series*, volume 983, page 012113. IOP Publishing.

9.Duncan, J., Subramanian, S., and Harrington, P. (2022). Generative modeling of high-resolution global precipitation forecasts. *arXiv preprint arXiv:2210.12504*.

10.Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S. (2022). Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences*, 26(13):3377–3392.

11.Funato, M. and Sawada, Y. (2025). Multi-model ensemble and reservoir computing for river discharge prediction in ungauged basins. *arXiv preprint arXiv:2507.18423*.

12.Husic, A., Al-Aamery, N., and Fox, J. F. (2022). Simulating hydrologic pathway contributions in fluvial and karst settings: An evaluation of conceptual, physically-based, and deep learning modeling approaches. *Journal of Hydrology X*, 17:100134.

13.Koch, J., Gotfredsen, J., Schneider, R., Troldborg, L., Stisen, S., and Henriksen, H. J. (2021). High resolution water table modeling of the shallow groundwater using a knowledge-guided gradient boosting decision tree model. *Frontiers in Water*, 3:701726.

14.Konapala, G., Kao, S.-C., Painter, S. L., and Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous us. *Environmental Research Letters*, 15(10):104022.

15.Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D. J., and Maharaj, K. T. (2023). The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability*, 15(13):10543.

16.Leščešen, I., Tanhapour, M., Pekárová, P., Miklánek, P., and Bajtek, Z. (2025). Long short-term memory (lstm) networks for accurate river flow forecasting: A case study on the morava river basin (serbia). *Water*, 17(6):907.

17.Li, H., Li, S., and Ghorbani, H. (2024a). Data-driven novel deep learning applications for the prediction of rainfall using meteorological data. *Frontiers in Environmental Science*, 12:1445967.

18.Li, H., Li, S., and Ghorbani, H. (2024b). Data-driven novel deep learning applications for the prediction of rainfall using meteorological data. *Frontiers in Environmental Science*, 12:1445967.

19.Li, M. and Yan, Y. (2024). Comparative analysis of machine-learning models for soil moisture estimation using high-resolution remote-sensing data. *Land*, 13(8):1331.

20.McSpadden, D., Goldenberg, S., Roy, B., Schram, M., Goodall, J. L., and Richter, H. (2024). A comparison of machine learning surrogate models of street-scale flooding in norfolk, virginia. *Machine Learning with Applications*, 15:100518.

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

22.Nair, B. B., Silamparasu, S., Mohnish, R., Deepak, T., and Rahul, M. (2019). Forecasting rainfall using soft computing techniques–a case study using india rainfall data. In *IOP Conference Series: Materials Science and Engineering*, volume 561, page 012119. IOP Publishing.

23.Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., et al. (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, 26(15):4013–4032.

24.Ouyang, W., Ye, L., Chai, Y., Ma, H., Chu, J., Peng, Y., and Zhang, C. (2025). A differentiable, physics-based hydrological model and its evaluation for data-limited basins. *Journal of Hydrology*, 649:132471.

25.Paeth, H., Girmes, R., Menz, G., and Hense, A. (2006). Improving seasonal forecasting in the low latitudes. *Monthly weather review*, 134(7):1859–1879.

26.Ramli, I., Rusdiana, S., Achmad, A., et al. (2019). Comparisons among rainfall prediction of monthly rainfall basis data in aceh using an autoregressive moving average. In *IOP Conference Series: Earth and Environmental Science*, volume 365, page 012008. IOP Publishing.

27.Schmidt, L., Heße, F., Attinger, S., and Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across germany. *Water resources research*, 56(5):e2019WR025924.

28.Shahriari, M. A., Aghighi, H., Azadbakht, M., Ashourloo, D., Matkan, A. A., Brakhasi, F., and Walker, J. P. (2025). Soil moisture estimation using combined sar and optical imagery: Application of seasonal machine learning algorithms. *Advances in Space Research*, 75(8):6207–6221.

29.Sharma, S., Ghimire, G. R., and Siddique, R. (2021). Machine learning for postprocessing ensemble streamflow forecasts. *CoRR*, abs/2106.09547.

30.Sinay, L. J. and Kembauw, E. (2021). Monthly rainfall components in ambon city: evidence from the serious time analysis. In *IOP Conference Series: Earth and Environmental Science*, volume 755, page 012079. IOP Publishing.

31.Singh, K., Dalai, A., Kaiwart, M., Mohanty, R., and Kumar, B. (2018). Prediction of rainfall of allahabad district by the development of autoregressive time series model. *Int. J. Curr. Microbiol. App. Sci*, 7(4):1516–1522.

32.Solanki, H., Vegad, U., Kushwaha, A., and Mishra, V. (2025). Improving streamflow prediction using multiple hydrological models and machine learning methods. *Water Resources Research*, 61(1):e2024WR038192.

33.Subbiah, T., Parthiban, P., Mahesh, R., and Das, A. (2021). Time series analysis of decadal precipitation pattern at selected cities of southern india. *Nature Environment and Pollution Technology*, 20(3):1201–1208.

34.Šuljug, J., Spišić, J., Grgić, K., and Žagar, D. (2024). A comparative study of machine learning models for predicting meteorological data in agricultural applications. *Electronics*, 13(16):3284.

35.Swami, D., Shah, A. D., and Ray, S. K. (2020). Predicting future sales of retail products using machine learning. *arXiv preprint arXiv:2008.07779*.

36.Tao, H., Al-Sulttani, A. O., Salih Ameen, A. M., Ali, Z. H., Al-Ansari, N., Salih, S. Q., and Mostafa, R. R. (2020). Training and testing data division influence on hybrid machine learning model process: application of river flow forecasting. *Complexity*, 2020(1):8844367.

37.Venter, J. and Mebrhatu, M. (2005). Modelling of rainfall occurrences at grootfontein (karoo, south africa). *South African Journal of Plant and Soil*, 22(2):127–128.

38.Wang, J., Sanderson, J., Iqbal, S., and Woo, W. L. (2025). Accelerated and interpretable flood susceptibility mapping through explainable deep learning with hydrological prior knowledge. *Remote Sensing*, 17(9):1540.

39.Wang, W., Du, Y., Chau, K., Chen, H., Liu, C., and Ma, Q. (2021). A comparison of bpnn, gmdh, and arima for monthly rainfall forecasting based on wavelet packet decomposition. *Water*, 13(20):2871.

40.Zhou, Q., Teng, S., Situ, Z., Liao, X., Feng, J., Chen, G., Zhang, J., and Lu, Z. (2023). A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. *Hydrology and Earth System Sciences*, 27(9):1791–1808.