Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Machine Learning Modeling for Rainfall Prediction in Bangladesh

Md. Sahin Alom
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
sahinalom0x@gmail.com

Md. Tilak Ahamed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
tilakahmmed864@gmail.com

Md. Fahad Khan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mfahad2412@gmail.com

Md. Zulkar Nain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
joysarkar99008@gmail.com

Ahammad Hossain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
ahammadstatru@gmail.com

A.H.M. Rahmatullah Imon
Center for Interdisciplinary Research
Varendra University
Rajshahi, Bangladesh
director.cir@vu.edu.bd

***Abstract*— The increasing unpredictability of rainfall poses significant risks to environmental sustainability and socio-economic stability. This study focuses on forecasting rainfall in eight districts of Bangladesh—Barishal, Chattogram, Dhaka, Khulna, Rajshahi, Rangpur, Mymensingh, and Sylhet using monthly data from January 1970 to December 2022, collected from the Bangladesh Agricultural Research Council (BARC). Time series models, including SARIMA(1,1,1)(1,1,1)1212, are applied to capture seasonal patterns. Additionally, advanced machine learning models such as Long Short-Term Memory (LSTM) networks are employed to detect non-linear trends and improve prediction accuracy. Results indicate significant seasonal fluctuations in rainfall, especially during the monsoon season. These findings underscore the necessity for data-driven water management strategies and climate adaptation policies. The combination of traditional statistical models with modern machine learning techniques ensures robust and precise forecasting. Future research could extend this approach by including other climatic variables and using ensemble models for enhanced accuracy. This study offers critical insights for policymakers in climate-sensitive regions.**

Keywords—Climate Variability, Rainfall Modeling and Forecasting, Seasonal Trends, Climate Change Adaptation, Time Series Analysis.

I. INTRODUCTION

Climate change presents significant challenges to environmental sustainability, socioeconomic development, and human well-being, with rainfall being a critical climatic variable that influences agriculture, water resources, and disaster management [4], [5]. In Bangladesh, where livelihoods are highly dependent on seasonal rainfall, increasing variability and unpredictability pose severe risks, including floods, droughts, and food insecurity [2], [10]. Given the country's vulnerability to climate change, accurate rainfall prediction is crucial for effective planning in agriculture and disaster management. This study focuses on forecasting rainfall using monthly data from January 1970 to December 2022 for eight key districts in Bangladesh: Barishal, Chattogram, Dhaka, Khulna, Rajshahi, Rangpur, Mymensingh, and Sylhet. The research employs time series models, specifically SARIMA, along with advanced machine learning techniques such as Long Short-Term Memory (LSTM) networks, which can capture historical patterns and seasonal variations for more precise rainfall prediction [6], [7]. Recent studies highlight that hybrid models combining statistical and machine learning approaches significantly improve forecasting accuracy, particularly in climate-sensitive regions [3], [8]. By providing deeper insights into rainfall variability, this study aims to assist policymakers in developing strategies for climate adaptation, water resource management, and disaster risk reduction [1], [9]. The outcomes of this research will contribute to enhancing agricultural productivity, minimizing the impacts of extreme weather events, and supporting sustainable development efforts in Bangladesh, ultimately fostering resilience in vulnerable communities.

II. MATERIALS AND METHODS

A. Study Area

This study focuses on Bangladesh, a tropical South Asian country divided into eight administrative divisions: Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet. Rainfall data from all available meteorological stations across these divisions, sourced from the Bangladesh Agricultural Research Council (BARC), were utilized. These stations represent diverse climatic zones, ensuring a comprehensive analysis of rainfall trends across the country. This approach captures spatial variations and offers a comprehensive understanding of Bangladesh's climate dynamics.

B. Data Preprocessing

Data preprocessing in this study involved cleaning and preparing 50 years of climatic rainfall data, ensuring its quality and reliability for accurate modeling and analysis. The following preprocessing steps were applied:

1. Handling Missing Data: Missing values are prevalent in long-term climatic datasets and can adversely impact model accuracy if not addressed. In this study, stations with over 30% missing data were

excluded to prevent bias, while remaining gaps were filled using interpolation techniques to maintain dataset integrity.

2. Outlier Detection and Treatment: The Box Plot method was used for outlier detection and treatment to ensure data accuracy. Values outside Q1 − 1.5 × IQR or Q3 + 1.5 × IQR were identified as outliers; errors were excluded, while extreme events were retained. Outliers were adjusted to the nearest threshold, preserving dataset integrity for accurate analysis.

3. Stationarity Test: Stationarity is essential in time series analysis, as it ensures constant statistical properties over time. The Augmented Dickey-Fuller (ADF) test was used in this study, where a p-value ≤ 0.05 confirmed stationarity; non-stationary series were transformed through differencing or detrending to ensure reliable modeling and accurate predictions.

4. Data Splitting: Training vs. Testing, The data were cleaned, preprocessed, and split into a training set (80%) and a testing set (20%). The training set was used for model training and parameter adjustment, while the testing set ensured unbiased evaluation of the model's generalization on unseen data.

III. METHODOLOGY

This study applied four distinct algorithms- ARIMA, SARIMA, GARCH, and LSTM—to model and forecast rainfall trends in Bangladesh. Each algorithm was selected for its unique capability to address different characteristics of time-series data, enabling a comprehensive analysis of climatic variations.

A. ARIMA, SARIMA and GARCH Forecasting

The ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal ARIMA), and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models are foundational approaches in time series forecasting, offering robust solutions for analyzing rainfall trends in Bangladesh, where climatic patterns are highly complex. ARIMA is ideal for non-seasonal data, incorporating dependencies through its autoregressive (AR) component, differencing (I) to ensure stationarity, and moving average (MA) for residual error modeling. SARIMA extends ARIMA by addressing seasonal patterns with additional differencing and autoregressive components, making it effective for recurring climatic phenomena such as monsoons. Meanwhile, GARCH models capture time-dependent variance, which is crucial for analyzing extreme rainfall events.

Forecasting Steps

- Stationarity Check: Statistical tests, including KPSS and ADF, were conducted to evaluate stationarity. When necessary, differencing techniques were applied to stabilize the series.

- Parameter Identification: Parameters for ARIMA and SARIMA were determined using ACF and PACF plots, while AIC and BIC ensured optimal selection. For GARCH, volatility clustering guided parameter choices.

- Model Selection and Estimation: Various configurations were tested, with RMSE, MAE, and MSE assessing predictive accuracy. Maximum Likelihood Estimation (MLE) was used for precise GARCH parameter estimation.

- Residual Diagnostics: Tests such as the Jarque-Bera test for normality and Durbin-Watson for autocorrelation ensured model adequacy.

- Forecasting: The models generated point forecasts and confidence intervals. GARCH forecasts also identified periods of high or low rainfall variability.

This methodology enhances rainfall forecasting accuracy in Bangladesh, aiding policy planning, optimizing agricultural resource allocation, and improving risk mitigation strategies for extreme weather events.

B. LSTM Forcasting

Long Short-Term Memory (LSTM) is a specialized type of recurrent neural network (RNN) designed to handle sequential and time-series data effectively. Traditional RNNs suffer from issues such as the vanishing gradient problem, which makes it challenging to capture long-term dependencies in data. LSTM addresses these issues using a unique architecture that includes a memory cell, which maintains information across time steps, helping the model retain important features over long periods. LSTMs use three primary gates to control the flow of information: the forget gate, which determines what information to discard from the cell state; the input gate, which decides what new information to add to the cell state; and the output gate, which regulates the output and updates the hidden state. These gates, controlled by learnable weights, use sigmoid and tanh activation functions to regulate information flow and retention effectively.

Forecasting Steps:

- Data Preprocessing: The rainfall data was normalized to a scale of 0 to 1 to improve the performance and convergence of the LSTM model.

- Model Construction: An LSTM model was constructed with one or more hidden layers. Hyperparameters such as the number of neurons, learning rate, and batch size were tuned to optimize model performance.

- Model Training: The LSTM model was trained using backpropagation through time (BPTT) with the Adam optimizer and mean squared error (MSE) as the loss function.

- Model Evaluation: The trained model was evaluated using statistical metrics such as RMSE, MAE, MSE, and R-squared. The results were compared with the SARIMA model to assess relative performance.

- Forecasting: Rainfall forecasts were generated using the trained LSTM model. Both short-term and long-term forecasts were produced to capture different forecasting horizons.

This approach highlights LSTM networks' capability to model complex, non-linear temporal patterns in rainfall data.

IV. RESULT AND DISCUSSION

In this study, we employed four statistical models—ARIMA, SARIMA, GARCH, and EGARCH—along with a deep learning-based LSTM model to forecast rainfall patterns in Bangladesh. The primary objective was to identify the most effective algorithm in capturing the temporal dynamics and seasonal variations of rainfall data.

A. Time Series Plot

A time series plot is a graphical representation of data points in a time-ordered sequence, used to observe underlying trends, seasonal patterns, and potential anomalies in a dataset. In this study, a time series plot of monthly rainfall data for Bangladesh from January 1970 to December 2022 was used to visually examine fluctuations in rainfall over time.

.

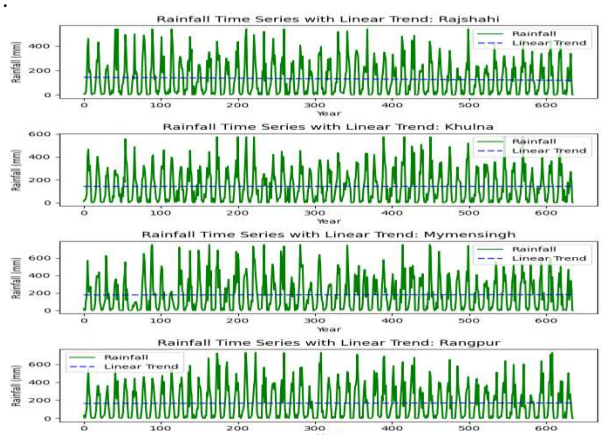

Fig. 1. Time Series with Linear Trend for Rainfall of Rajshahi, Khulna, Mymensingh and Rangpur.

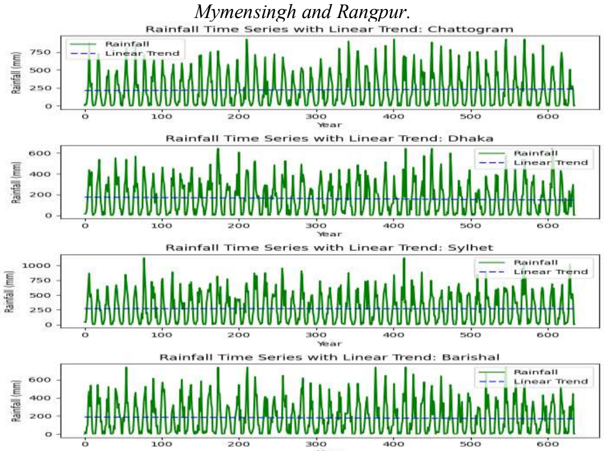

Fig. 2. Time Series with Linear Trend for Rainfall of Chattogram, Dhaka, Sylhet and Barishal.

B. Model Performance Evaluation

The performance of each model was assessed using four statistical metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of Determination (R^2). These metrics comprehensively evaluate the predictive accuracy and reliability of the models.

Model	RMSE	MSE	MAE	R^2
ARIMA	120.106	14425.532	103.143	-0.01
SARIMA	66.476	4419.06	46.43	0.82
GARCH	119.757	14341.797	102.563	-0.004
LSTM	73.112	5345.503	52.974	0.642

TABLE I. Evaluation of Rainfall, Region: Rajshahi

Model	RMSE	MSE	MAE	R^2
ARIMA	251.7	63353.382	201.859	-0.001
SARIMA	115.52	13345.004	77.515	0.863
GARCH	277.391	76945.817	202.658	-0.216
LSTM	143.505	20593.631	97.463	0.675

TABLE II. Evaluation of Rainfall, Region: Chitagong

Model	RMSE	MSE	MAE	R^2
ARIMA	157.974	24955.777	115.858	-0.243
SARIMA	81.153	6585.763	55.01	0.801
GARCH	141.648	20064.077	116.527	0
LSTM	89.023	7925.165	62.687	0.616

TABLE III. Evaluation of Rainfall, Region: Khulna

Model	RMSE	MSE	MAE	R^2
ARIMA	162.212	26312.646	138.585	-0.119
SARIMA	91.765	8420.75	63.876	0.782
GARCH	153.569	23583.539	126.475	-0.003
LSTM	97.421	9490.911	74.019	0.618

TABLE IV. Evaluation of Rainfall, Region: Dhaka

Model	RMSE	MSE	MAE	R^2
ARIMA	176.628	31197.696	156.773	-0.003
SARIMA	97.831	9571.092	65.870	0.752
GARCH	176.442	31131.830	154.565	-0.001
LSTM	112.509	12658.490	85.032	0.608

TABLE V. Evaluation of Rainfall, Region: Mymensingh

Model	RMSE	MSE	MAE	R^2
ARIMA	316.563	100212.498	272.634	-0.453
SARIMA	133.883	17924.748	90.099	0.82
GARCH	263.068	69204.84	226.889	-0.003
LSTM	159.277	25369.334	109.336	0.633

TABLE VI. Evaluation of Rainfall, Region: Sylhet

Model	RMSE	MSE	MAE	R^2
ARIMA	191.806	36789.634	144.299	-0.071
SARIMA	119.375	14250.429	76.707	0.745
GARCH	229.814	52814.552	153.923	-0.537
LSTM	149.92	22476.24	110.023	0.362

TABLE VII. Evaluation of Rainfall, Region: Rangpur

Model	RMSE	MSE	MAE	R^2
ARIMA	178.276	31782.356	154.756	-0.009
SARIMA	104.221	10862.156	65.965	0.755
GARCH	177.48	31499.403	152.613	-0.001
LSTM	114.446	13098.084	81.912	0.602

TABLE VIII. Evaluation of Rainfall, Region: Barishal

C. Performance Comparison

Among the models evaluated, SARIMA delivered good overall performance by effectively capturing the seasonal patterns and trends in rainfall data. In contrast, LSTM provided average results, demonstrating reasonable capabilities in handling non-linear relationships, but falling short of SARIMA in predictive accuracy. The following is a summary of each model's performance:

- ARIMA: Produced stable results for non-seasonal patterns but underperformed compared to SARIMA due to its inability to model seasonal components effectively.
- SARIMA: Achieved strong results, particularly in scenarios with clear seasonal variations, due to its ability to model seasonality through seasonal differencing.

- LSTM: Delivered average performance, showing potential in modeling non-linear patterns but struggling with the sparse and highly variable nature of rainfall data.
- GARCH: The models were suitable for capturing volatility and heteroscedasticity but were less effective for long-term forecasting due to their limited handling of non-linear and seasonal dynamics.

D. Recommended Model

Among the ARIMA, SARIMA, LSTM, and GARCH models, the best results were achieved with the Seasonal ARIMA (SARIMA) model.

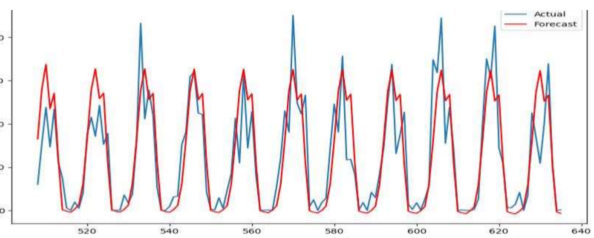

Fig. 3. SARIMA Forecast vs Actual for Rainfall in Rajshahi

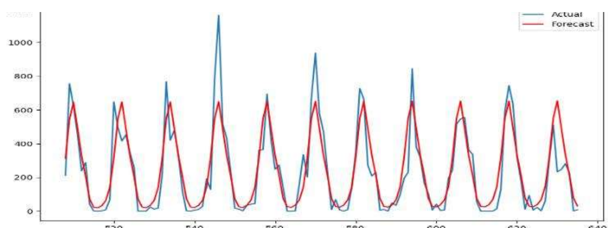

Fig. 4. SARIMA Forecast vs Actual for Rainfall in Chattogram

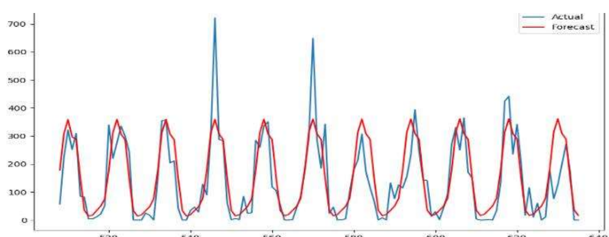

Fig. 5. SARIMA Forecast vs Actual for Rainfall in Khulna

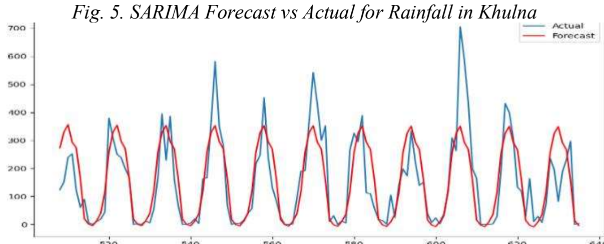

Fig. 6. SARIMA Forecast vs Actual for Rainfall in Dhaka

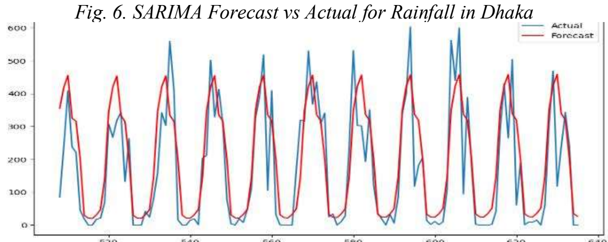

Fig. 7. SARIMA Forecast vs Actual for Rainfall in Mymensingh

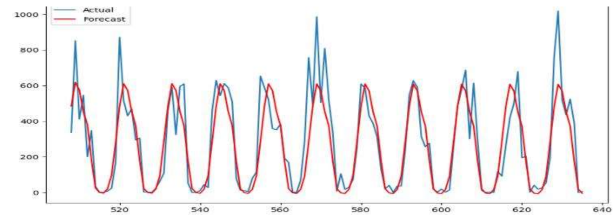

Fig. 8. SARIMA Forecast vs Actual for Rainfall in Sylhet

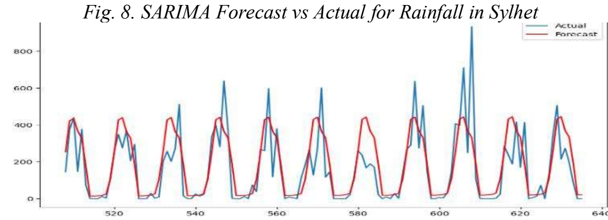

Fig. 9. SARIMA Forecast vs Actual for Rainfall in Rangpur

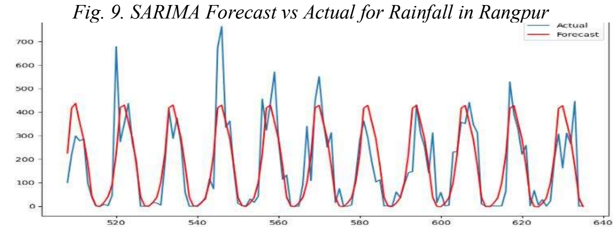

Fig. 10. SARIMA Forecast vs Actual for Rainfall in Barishal

Overall, SARIMA proved to be the most reliable model for rainfall forecasting, while LSTM and GARCH models showed potential for specific aspects, such as non-linear dynamics and volatility modeling.

References

[1] N. Jahan Ria, J. F. Ani, M. Islam, and A. K. M. Masum, "Standardization of rainfall prediction in Bangladesh using machine learning approach," in *Proc. 12th Int. Conf. on Computing Communication and Networking Technologies (ICCCNT)*, 2021.

[2] M. M. H. Khan, M. R. U. Mustafa, M. S. Hossain, et al., "Short-term and long-term rainfall forecasting using ARIMA model," *Int. J. of Environmental Science and Development*, vol. 14, no. 5, pp. 292-298, 2023.

[3] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, et al., "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Machine Learning with Applications*, vol. 7, p. 100204, 2022.

[4] E. A. Hussein, M. Ghaziasgar, C. Thron, M. Vaccari, and Y. Jafta, "Rainfall prediction using machine learning models: Literature survey," in *Artificial Intelligence for Data Science in Theory and Practice*, pp. 75-108, 2022.

[5] J. Hussain and C. Zoremsanga, "A survey of rainfall prediction using deep learning," in *Proc. 2021 3rd Int. Conf. on Electrical, Control and Instrumentation Engineering (ICECIE)*, pp. 1-10, IEEE, 2021.

[6] C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *Proc. 2020 Int. Conf. on Electronics and Sustainable Communication Systems (ICESC)*, pp. 92-97, IEEE, 2020.

[7] A. U. Rahman, S. Abbas, M. Gollapalli, et al., "Rainfall prediction system using machine learning fusion for smart cities," *Sensors*, vol. 22, no. 9, p. 3504, 2022.

[8] B. Kumar, K. Atey, B. B. Singh, et al., "On the modern deep learning approaches for precipitation downscaling," *Earth Science Informatics*, vol. 16, no. 2, pp. 1459-1472, 2023.

[9] A. Parmar, K. Mistree, and M. Sompura, "Machine learning techniques for rainfall prediction: A review," in *Proc. Int. Conf. on Innovations in Information Embedded and Communication Systems*, vol. 3, 2017.

[10] H. Farman, N. Islam, S. A. Ali, et al., "Advancing rainfall prediction in Pakistan: A fusion of machine learning and time series forecasting models," *Int. J. of Emerging Engineering and Technology*, vol. 3, no. 1, pp. 17-24, 2024.