International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 25(2), November, 2025, pp 127-141
DOI: https://doi.org/10.3329/ijss.v25i2.85776

Rainfall Modeling in Northwestern Bangladesh: A Hybrid Approach Using Distribution Fitting and Machine Learning

Md. Palash Bin Faruque[1], Saiful Islam[2], Palash Mondol[3], Manoj Roy[4] and Md. Abdul Khalek[5*]

[1]Department of statistics, University of Rajshahi, Bangladesh,
Email: mdpalashbinfaruque@gmail.com
[2]Department of statistics, University of Rajshahi, Bangladesh,
Email: saiful524349@gmail.com
[3]Department of statistics, University of Rajshahi, Bangladesh,
Email: palash.stat116@gmail.com
[4]Department of statistics, University of Rajshahi, Bangladesh,
Email: manojroy989898@gmail.com
[5]Department of statistics, University of Rajshahi, Bangladesh,
[*]Correspondence should be addressed to Md. Abdul Khalek
(Email: mak.stat09@gmail.com)

[Received July 10, 2025; Accepted October 27, 2025]

Abstract

Proper rainfall modeling is important for efficient water resource management, strategic agricultural planning and formulating disaster preparedness plans, particularly in climate-sensitive areas like northwest Bangladesh. This study develops a hybrid approach that integrates probabilistic distribution fitting with machine learning techniques to improve accuracy and forecast rainfall. For this purpose, monthly time series data collected by Bangladesh meteorological department (BMD) of three meteorological stations of northwest Bangladesh, namely, Rajshahi, Ishurdi and Bogura for the period 1964-2023 were analyzed. The time series rainfall data was evaluated models of probability distributions as Log-Normal, Log-Pearson III, Pearson III, GEV, Gamma, Weibull, and Exponential using AIC and BIC criteria, alongside Random Forest and hybrid models for forecasting. The spatial analysis identified Log-Normal (LN2) as the most suitable distribution for Rajshahi (AIC = 739.44), Log-Pearson III for Ishurdi (AIC = 768.15), and Pearson III for Bogura (AIC = 761.04). Forecasts indicate an upward trend in extreme rainfall events, with Ishurdi demonstrating the highest inter annual variability, including projected peaks exceeding 600 mm. Among the forecasting approaches, the hybrid model integrating Random Forest and probabilistic distribution fitting achieved superior performance, particularly in Ishurdi (RMSE = 127.78, MAPE = 34.71%). In contrast, the LN2 and Pearson III distributions yielded the most accurate predictions for Rajshahi and Bogura, respectively. These results highlight how regional model selection can improve rainfall forecasting accuracy and how crucial these methods are for guiding water resource management in highly variable and data-constrained environments, supporting precision agriculture, developing early warning systems, and forming climate-resilient policy.

Keywords: Rainfall, Probabilistic distribution, Machine learning model, Hybrid modeling, Random forest, Northwest Bangladesh.

AMS Classification: 62P12, 68T09.

1. Introduction

Rainfall has an important influence on socio-economic livelihood and environmental functioning of Bangladesh, in particular over its northwest region since agriculture is one of the main livelihood to a significant portion of the people. But the region is known for its instability in rainfall – both across space and through time – and is prone to droughts, floods and other climate-derived threats. These challenges will be compounded by climate change, which is predicted to increase rainfall extremes and alter seasonal timing of rainfall, with potentially critical impacts on food security, water resource management and disaster preparedness. Rainfall variability has become a significant area of research in global hydrometeorological studies because it greatly affects climate-sensitive sectors, including agriculture, water resource management, and disaster risk reduction (Allan & Soden, 2008; O'Gorman & Schneider, 2009; Trenberth, 2011). Accurate modeling of rainfall patterns, especially under conditions of climate uncertainty, remains a pressing challenge in both data-rich and data-scarce regions(C. Li et al., 2021; Noor et al., 2021; Smith et al., 2019). Over the last few years, hybrid techniques blending statistical distribution fitting with machine learning methods have picked up momentum for improving rainfall prediction accuracy and reliability, showcasing higher performance in simulating intricate hydrological processes(Aderyani et al., 2022; Morovati & Kisi, 2024; Shu & Ouarda, 2008). However, the application of such integrative methodologies remains limited in highly vulnerable regions like northwestern Bangladesh, where complex spatiotemporal rainfall patterns demand more localized and adaptive modeling frameworks(Khan et al., 2020; M. N. Rahman & Azim, 2023; Roy et al., 2022).

Rainfall is critical for various environmental, agricultural, and economic reasons. Rainfall prediction has become increasingly important in recent times. Recent research has shown that hybrid machine learning models enhance rainfall prediction by integrating satellite data, land features, and hydrologic modelling to achieve better spatiotemporal accuracy (Di Nunno et al., 2022). Some studies show that using hybrid machine learning models—like ANN-SVM for better rainfall prediction in Khulna and EEMD-ANN/SVM for more accurate thunderstorm frequency prediction in Bangladesh—works well by combining different methods and carefully chosen climate factors, resulting in much better accuracy and reliability than using single models alone (Azad et al., 2021; Shuvo et al., n.d.). ELM was better than ANN at predicting rainfall in North-Western Bangladesh, and using Bayesian Model Averaging with GCM and historical data made the predictions better, showing that pre-monsoon rainfall trends are decreasing and becoming more unpredictable, probably due to SST (Basher et al., 2020; Monir et al., 2023; Rizvee et al., 2020). Rainfall distribution analysis showed that while Makkah fits normal, gamma, and Weibull distributions and Jeddah fits only normal, Pakistan's diverse zones favour log-Pearson III and Weibull, with log-Gumbel and log-normal best for high return periods, and in northwest Bangladesh (Rajshahi, Ishurdi, Bogura), log-Pearson III fits annual maximum rainfall best, with LN2 and P3 as strong alternatives (Alzahrani et al., 2025; Haseeb et al., 2025; Khalek et al., 2023; Montes-Pajuelo et al., 2024). Although certain research has employed multiple probability distribution models or independent/hybrid machine learning models for rainfall prediction, a combined integration of both in the form of a shared framework is limited. Most of the existing research focuses on either statistical modeling of rainfall patterns in a specific area or employing machine learning techniques to predict rainfall trends. The integration of machine learning and probabilistic methods simultaneously for a robust and unified model of rainfall prediction has not been investigated before. It is important to focus on this section because most divisions in Bangladesh exhibit a declining trend in monsoon rainfall, threatening agricultural productivity and water resource management, thereby highlighting the need for effective irrigation strategies (Das et al., 2021; M. M. Rahman & Abdullah, 2022).

For filling the gap, this study aims to investigate annual maximum rainfall patterns in Rajshahi, Ishurdi, and Bogura through statistical parameters and determine the most appropriate probability distributions by AIC and BIC criteria. It also seeks to forecast rainfall up to 2033 using distribution-based, machine learning (Random Forest), and hybrid models, compare their performance using RMSE, MAE, and MAPE, and determine the most suitable model for each district to support effective regional water resource planning. The monthly maximum rainfall data is used in this study, and it integrates multiple probability distribution fittings, random forest machine learning, and a hybrid modelling approach to forecast the maximum annual rainfall across three districts. Best-fit distributions are selected using AIC and BIC, while the best predicted model accuracy is evaluated using RMSE, MAE, and MAPE to compare performance. Through the integration of probabilistic and machine learning models for rainfall forecasting, this study offers crucial insights for early warning systems and localized water resource planning. It enables policymakers to design region-specific adaptation strategies and infrastructure planning under climate uncertainty. For agriculture, the accurate prediction of rainfall extremes supports better crop calendar planning, irrigation scheduling, and drought resilience, particularly in rainfall-volatile regions like Ishurdi.

2. Study Area and Data Collection

This study is concentrated on the three high temperature meteorological stations in the northwestern region of Bangladesh, namely Rajshahi (24°22′ N, 88°36′ E), Ishurdi (24°8′57″ N, 89°3′57″ E), and Bogura (24°46′48″ N, 89°21′ E), that precisely capture each climatic condition required for rainfall modeling for the period 1964-2023. These districts are part of the Barind Tract area, famous for their peculiar agro-climatic features and sensitivity to rainfall variability.

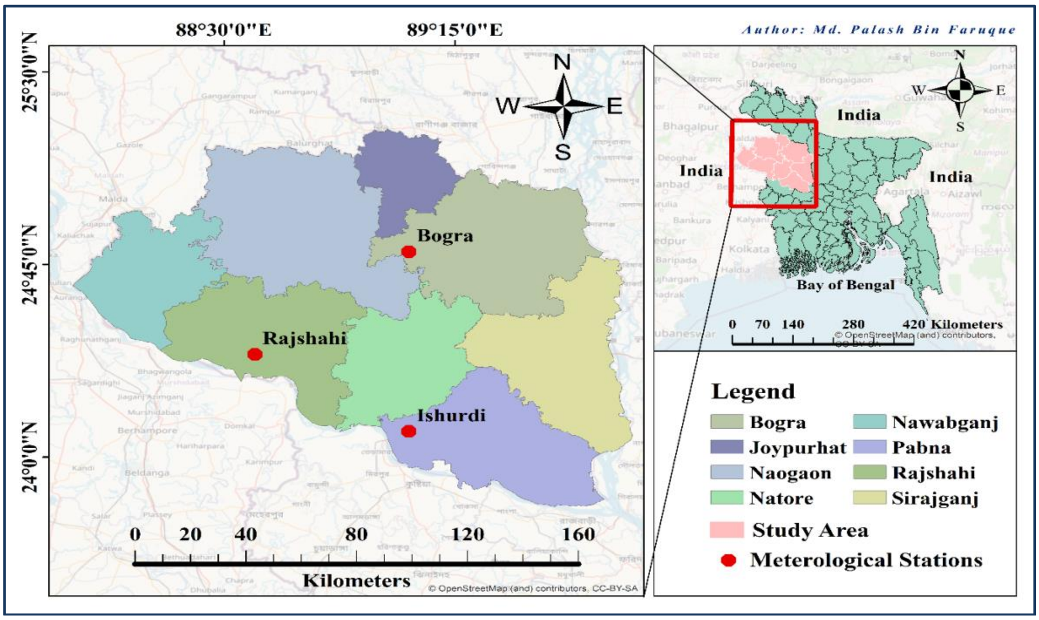

Figure 1: Geographical Map of Rajshahi Division, Bangladesh, with Study Area (Own)

Figure 1 shows meteorological stations that are representative of these districts, along with their names and locations. The Bangladesh Meteorological Department (BMD) provided monthly rainfall data from January 1964 to December 2023. For analysis, the highest value of rainfall for each dataset was utilized, and missing values were filled with the respective maximum rainfall value to avoid losing data.

3. Methodology

In this study, we use seven different unique distributions to determine and contrast the highest precipitation performance, including 3-parameter gamma (G3), Weibull (W2), Exponential (Exp), Log-Normal (LN2), Pearson type III (P3), Log-Pearson type III and the Generalized Extreme Value (GEV) distribution is also utilized in this study. Estimating the frequency distribution of extreme events has become simpler and more acceptable due to the benefits of these probability distributions (Z. Li et al., 2015). We are determined the best fitted distribution among these distributions by comparing AIC and BIC value. We also apply a supervised machine learning algorithm—Random Forest—to construct and predict rainfall by using atmospheric data.

3.1 3-Parameter Gamma (G3) Distribution

The G3 distribution is incredible for hydrological uses since it can easily show skewed and limited rainfall data. It consists of three parameters: shape (α), scale (β), and location (θ). These parameters make it easier to fit than the old 2-parameter gamma distribution, particularly in areas where rainfall indicators are not initiated at zero (Haddad et al., 2012). The probability density function (PDF) of the 3-Parameter Gamma (G3) Distribution is given by:

$$f(x;\alpha,\beta,\theta) = \begin{cases} \dfrac{(x-\theta)^{\alpha-1} e^{-\frac{(x-\theta)}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, & x > \theta \\ 0\,, & x \leq \theta \end{cases}$$

Where, $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter, θ is the location parameter, and $\Gamma(\alpha)$ is the gamma function.

To apply the G3 distribution to rainfall prediction, a regression or machine learning model that dynamically estimates the parameters (α, β, θ) of the distribution is developed based on meteorological variables such as temperature, relative humidity, wind speed, dew point, and surface pressure. By this method one may calculate probability distributions of the rainfall given existing climatic variables (Katz, 1977; Wilks, 1990).

3.2 2-Parameter Weibull (W2) Distribution

The Weibull (W2) distribution, a flexible probability distribution function, is commonly used in meteorology and hydrology to describe rainfall incidence due to its capacity to represent various data types (Makkonen, 2006). This work uses the two-parameter Weibull distribution to predict rainfall based on real-world weather data, including temperature, humidity, and wind speed. The probability density function (PDF) of the two-parameter Weibull distribution is given by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], x > 0$$

Where, x is the rainfall amount (mm), $\beta > 0$ is the scale parameter, $\alpha > 0$ is the shape parameter.

3.3 2-Parameter Exponential (2-parameter) Distribution

The two-parameter form of the Exponential distribution includes a location parameter that determines when rainfall begins, making it a more flexible option than the standard one-parameter variety. The probability density function (pdf) of the Exponential (2-parameter) Distribution is given by:

$$f(x) = \frac{1}{\lambda} exp\left[-\left(\frac{x-\xi}{\lambda}\right)\right],\ x > \xi$$

Where: x is the rainfall amount (mm), ξ is the location parameter, λ is the scale parameter (mean excess rain).

3.4 2-parameter Log-Normal (LN2)

The LN2 distribution performs best with positively skewed hydrological data such as precipitation, where values cannot be negative and extreme events occur frequently (Wilks, 2011). The LN2 model differs from the normal distribution in that it presupposes that the natural logarithm of the variable of interest (rainfall) has a normal distribution. Let r denote the observed rainfall. Then, under the LN2 assumption:

$$\ln(r) \sim N(\mu, \sigma^2)$$

Where: μ is the mean of the logarithm of rainfall, σ is the standard deviation of rainfall.
The probability density function (PDF) of the LN2 distribution is given by:

$$f(r) = \frac{1}{r\sigma\sqrt{2\pi}} exp\left[-\frac{(\ln r-\mu)^2}{2\sigma^2}\right], r > 0$$

The parameters μ and σ were estimated using atmospheric predictors (temperature, humidity, wind speed, and atmospheric pressure). The model can be expressed as:

$$\ln(r) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon$$

Where: X_i represents the i -th atmospheric variable, β_i are the corresponding regression coefficients, $\varepsilon \sim N(0, \sigma^2)$ is the error term.

3.5 Pearson type III (P3) Distribution

The P3 distribution accurately simulates the positive skewness of distribution for all variables, including precipitation and stream flow in hydrologic models. This method is ideal for analyzing rainfall frequency in semi-arid and monsoon regions with skewed data (Haan, 2002; Te Chow et al., 1988).

The probability density function (PDF) of the Pearson Type III distributions is defined as:

$$f(x) = \frac{1}{|\beta|\Gamma(\alpha)}\left(\frac{x-\xi}{\beta}\right)^{\alpha-1} exp\left[-\left(\frac{x-\xi}{\beta}\right)\right]$$

Where: ξ is the location parameter, β is the scale parameter, α is the shape parameter, and $\Gamma(\alpha)$ is the gamma function evaluated at α.

3.6 Log-Pearson type III Distribution

The LP3 distribution is extensively utilized in hydrological and meteorological studies because it can simulate positively skewed data such as precipitation (Te Chow et al., 1988). We fit the Pearson Type III distribution to the logarithm of the significant variable. Its probability density function is defined by:

$$f(x) = \frac{1}{|\beta| \mathrm{x}\Gamma(\alpha)} \left(\frac{\ln(x) - \xi}{\beta}\right)^{\alpha-1} exp\left[-\left(\frac{\ln(x) - \xi}{\beta}\right)\right]$$

Where: $x > 0$ is the rainfall value, α is the shape parameter, β is the scale parameter, and ξ is the location parameter.

3.7 Generalized Extreme Value (GEV) distribution

We used the Generalized Extreme Value Distribution to show extreme rainfall; this distribution brings together the Gumbel, Fréchet, and Weibull families in one framework (Coles et al., 2001). The GEV Distribution is very good for applications in hydrology and meteorology where block maxima or threshold exceedances are used.

The probability density function (PDF) of the GEV distribution is given by:

$$f(x) = \frac{1}{\alpha} exp\big(-(1 - ky) - exp(-y)\big) \text{ , where } = -\frac{1}{k}\log\left(1 - k\,\frac{x-\xi}{\alpha}\right), k \neq 0$$

Where ξ, $\alpha > 0$, and k are the location, scale, and shape parameters respectively.

We use a regression-based parameterization of the GEV distribution parameters to model extreme rainfall with atmospheric predictors. That is, we model the location parameter ξ as a function of atmospheric covariates:

$$\xi(t) = \beta_0 + \beta_1 T(t) + \beta_2 H(t) + \beta_3 W(t)$$

Where; T(t) is the temperature at time t, H(t) is the humidity, W(t) is the wind speed, and $\beta_0, \beta_1 \ldots \beta_3$ are regression coefficients.

This allows dynamic modeling of the GEV parameters, adjusting the distribution of rainfall to vary with changing atmospheric conditions. Parameters are estimated by maximum likelihood estimation (MLE) and standard extreme value theory (Coles et al., 2001; Katz et al., 2002).

3.8 Random Forest (RF)

In this study, rainfall prediction with low error was attempted using a machine learning approach based on the Random Forest algorithm. (Breiman, 2001) introduced RF as an ensemble learning method which works by growing many decision trees at training time and outputting the mean prediction (regression) of the individual trees. Available atmospheric data, which is complex and has a nonlinear relationship, fits well within this model due to its robustness against over fitting.

The random forest regression builds a set of B regression trees $\{T_b\}_{b=1}^{B}$ each learned on a bootstrap sample of the training data. For a new input vector x, the RF prediction is given by:

$$\hat{f}_{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

Each tree T_b is grown by choosing a random subset of m features (where m<p, and p is the total number of predictors) at every split, ensuring model diversity and reducing correlation between individual trees (Breiman, 2001).

3.9 Model evaluation

To predict rainfall, this method trains two separate models independently. After both models have made their projections, the final rainfall prediction is estimated by taking the average of their results. The performance of the rainfall prediction of Random Forest (RF), probability distribution, and hybrid model is evaluated using the root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE), which are calculated as follows:

The Hybrid Model

$$\hat{\mathbf{y}}_{\mathbf{hybrid}} = \frac{\mathbf{1}}{\mathbf{2}}(\widehat{\mathbf{y}_1} + \widehat{\mathbf{y}_2})$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error

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

Mean Absolute Percentage Error

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100$$

4. Results and Discussions

4.1 Descriptive Analysis of Annual Maximum Rainfall

The extreme precipitation behavior over the studied time was statistically characterized by analyzing the yearly maximum rainfall data for three meteorological stations: Rajshahi, Ishurdi, and Bogura. The maximum monthly rainfall value recorded for each calendar year was chosen to determine the annual maxima for each station.

Table 1: Descriptive statistics of annual maximum rainfall

Stations	Years	Mean	SD	Skewness	Kurtosis
Rajshahi	60	403.10	116.68	0.57	0.02
Ishurdi	60	422.23	166.28	1.64	4.79
Bogura	60	477.68	138.23	0.51	-0.29

The yearly maximum amount of rainfall varies a lot between the stations studied. For Bogura, we can see that there was a higher mean (477.68 mm), close to symmetrical data with a skewness of 0.51, also there exists an above average variability (SD = 138.23 mm) and less than normal kurtosis (-0.29). There was more change in Ishurdi than anywhere else, with the highest SD = 166.28 mm, kurtosis = 4.79, and skewness = 1.64, with the mean rising to 422.23 mm. Rajshahi's rainfall is consistent, with the least skewness (0.57) and lowest average and standard deviation, at 403.1 mm and 116.68 mm. Since Ishurdi has strong distributional features, it appears to be more prone to large variations in water levels. Bogura often has regular and heavy rainfall every year. Having a lower range and a lower average, Rajshahi's rainfall appears steadier, which could make hydrological changes easier to predict.

4.2 Selection Criteria for Fitting Probability Distributions

The Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) were used to determine the best probability distribution for yearly maximum rainfall. These criteria strike a balance between goodness-of-fit and model complexity; models with lower values fit better.

Table 2: Model Selection Criteria for Probability Distribution Fitting

District	Distribution	AIC	BIC
Rajshahi	Log-Normal (2P)	739.44	743.62
	Log-Pearson III	739.55	743.74
	Pearson III	739.77	743.96
	Weibull (2P)	745.63	749.82
	Gamma (3P)	751.2	755.39
	Exponential (2P)	752.96	755.06
	GEV	741.66	747.95
Ishurdi	Log-Pearson III	768.15	772.34
	Log-Normal (2P)	768.6	772.79
	Pearson III	771.77	775.96
	Gamma (3P)	781.49	785.68
	Exponential (2P)	783.72	785.82
	Weibull (2P)	783.8	787.99
	GEV	770.19	776.48
Bogura	Pearson III	761.04	765.23
	Log-Normal (2P)	761.51	765.7
	Log-Pearson III	762.02	766.21
	Weibull (2P)	765.77	769.96
	Gamma (3P)	791.71	795.9
	Exponential (2P)	800.17	802.26
	GEV	762.79	769.07

The Log-Normal (2P) distribution was the best-fitting model for the annual maximum rainfall in Rajshahi because of its moderate skewness and variability with AIC = 739.44 and BIC = 743.62. Log-Pearson III and Pearson III were nearly identical options. Among the models tested, Log-Pearson III did the best (AIC = 768.15, BIC = 772.34), since rainfall in Ishurdi is much skewed and has more frequent extreme events. Pearson III was chosen, showing AIC = 761.04 and BIC = 765.23, to represent Bogura's high average rainfall and equal-sided distribution. Lognormal and Log-Pearson III also presented similar results. The distribution fitting results underscore the spatial heterogeneity in rainfall extremes across the study region. We must calibrate probabilistic models to accommodate regional rainfall patterns, as the most appropriate distributions reflect the distinct statistical signatures of individual districts. Under a shifting climatic regime, differences are essential for sound hydrological simulation, risk analysis, and infrastructure planning.

4.3 Distribution Fitting and Diagnostic Evaluation

In order to align actual data with theoretical distributions and promote sound water resource planning, probabilistic modeling with goodness-of-fit evaluation improves the accuracy of rainfall extreme predictions.

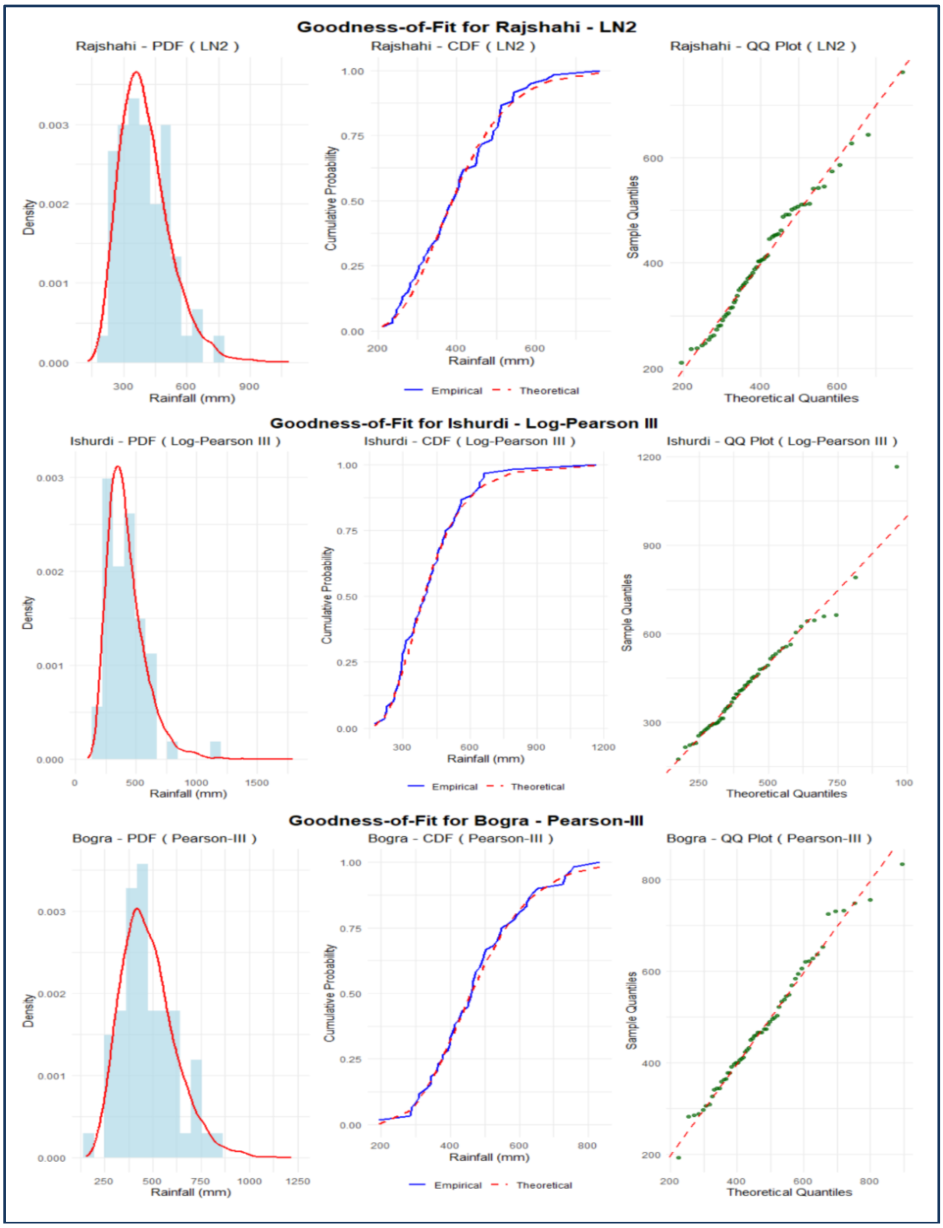

Figure-2: Goodness-of-Fit Diagnostics of Annual Maximum Rainfall

The values of precipitation in Rajshahi are described well by LN2, which lies between 200 and 1000 mm, with the maximum probability around 400 mm. On both the QQ plot and the CDFs, there is strong agreement, ending at a probability of 1. Ishurdi is suited to the Log-Pearson Type III distribution since there are 200 to 1100 mm of rainfall, with a maximum close to 400 mm. The QQ plot aligns most points closely to the 1:1 line, and the CDF curves also show extensive overlap. Pearson Type III distribution fits Bogura's data well since it ranges from 200mm to 1100mm and peaks at about 500mm. The CDF represents that the empirical and theoretical curves are very comparable, and the QQ plot demonstrates that sample quantiles are lined up closely with the theoretical curve.

This analysis confirms that each station's rainfall pattern is best represented by a distinct statistical distribution. Rajshahi corresponds with LN2, Ishurdi with Log-Pearson III, and Bogura with Pearson III, indicating variations in rainfall characteristics across the regions. Rajshahi exhibits a slightly lower peak rainfall, whereas Bogura records the highest peak. These fitted distributions improve the accuracy of hydrological modeling and are essential for effective water resource planning and risk assessment in the respective districts.

4.4 Forecasting Annual Maximum Rainfall Using Probabilistic Distributions, Machine Learning and Hybrid Approaches

Rainfall forecasts for Rajshahi, Ishurdi, and Bogura up to 2033 are illustrated in the chart with hybrid, Random Forest and probabilistic distributions. Rainfall variations from 1964 to 2023 are known from history, and predictions agree in sequence with the observations of these years. This combination of techniques gives very useful advice for adjusting water planning methods as the climate changes because it blends the strengths of risk-based modeling with the reliability of machine learning.

Rajshahi's expected highest rainfall each year changes between seasons. The Random Forest doesn't believe there is a greater than average peak, predicting only about 450 mm in 2026; however, the distribution-based prediction is higher because it estimates the peak to be around 570 mm next year. Hybrid projection, covering the years 2024 to 2033 and falling between 440 mm and 510 mm, closely follows the trends shown by both models with only a small difference.

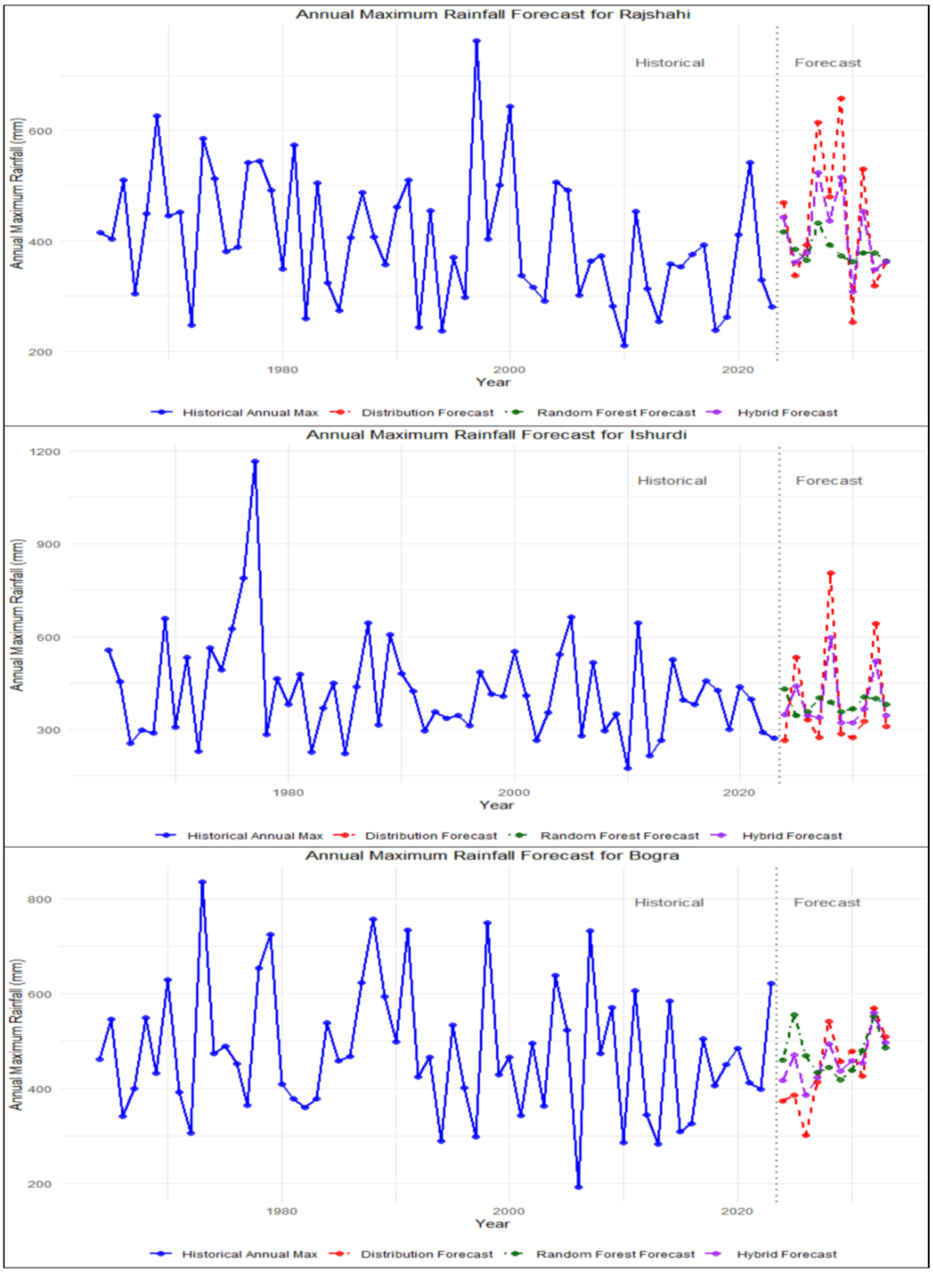

Figure-3: Historical and Forecasted Annual Maximum Rainfall Using Distributional, Machine Learning, and Hybrid Models

The Random Forest projection remains modest, barely exceeding 430 mm for the years under study, while the Ishurdi forecast predicts significant increases of over 600 mm in 2025 and 2028.

Values in this model can vary from 380 mm to 500 mm, indicating that patterns might be more stable in the future. The forecasts generated by Bogura using all three models show notable similarities. The distribution model indicates that the highest expected rainfall in 2032 will be 568 mm, while the Random Forest model predicts a range between 418 mm and 556 mm. The combination model maintains rainfall within a range of 416 mm to 560 mm, implying that annual variations in rainfall will be moderate.

Different regions have different expected patterns for rainfall. Since Rajshahi's statistics have high peaks and low Random Forest values, we can say its variability is moderate. Ishurdi suggests there might be heavy rain, with amounts forecast up to 600 mm or more, but Bogura's agreement among models implies that there will be consistent and moderate rainfall.

4.5 Comparative Evaluation of Rainfall Forecasting Models

A comparative analysis of rainfall forecasting performance across three districts utilizing distribution-based, random forest, and hybrid models. The methods are measured for predictive accuracy using standard error metrics called RMSE, MAE and MAPE.

Table 3: Performance Metrics of Rainfall Forecasting Models

District	Methods	RMSE	MAE	MAPE
Rajshahi	Distribution	96.78	73.8	21.36
	Random Forest	118.36	107.66	31.07
	Hybrid	101.67	90.24	26.07
Ishurdi	Distribution	138.95	118.24	37.72
	Random Forest	128.15	109.01	36.78
	Hybrid	127.78	103.57	34.71
Bogura	Distribution	89.33	76.02	14.99
	Random Forest	105.46	88.31	17.99
	Hybrid	94.26	82.16	16.49

The forecasting analysis revealed spatially varied model performance. The distribution-based model consistently produced the lowest forecast errors in both Rajshahi and Bogura (e.g., Rajshahi RMSE: 96.78, MAE: 73.8, MAPE: 21.36%; Bogura RMSE: 89.33, MAE: 76.02, MAPE: 14.99%). More specifically, LN2 for Rajshahi and Pearson-III for Bogura. It demonstrates that it matches what is needed in these fields. Meanwhile, the hybrid model was more accurate than Random Forest and the individual distribution models in Ishurdi, giving an RMSE of 127.78, a MAE of 103.57 and a MAPE of 34.71%. Rajshahi and Bogura, both having predictable rainfall patterns, had their forecasts most accurately predicted under the distribution-based model, as clearly shown by the results. But in Ishurdi, the hybrid model showed its superiority based on how well it handled the unpredictable rainfall. It appears that the best predictive results are found when the selection of a model is customized for local features of the data.

5. Conclusion

The study reveals that the maximum rainfall patterns and forecasting model performance are not the same in all parts of Rajshahi, Ishurdi and Bogura. In these districts, where the rainfall is predictable, LN2 for Rajshahi and Pearson III for Bogura were most accurate because of their stability. On the other hand, Ishurdi because of its stronger rains and unpredictability worked better with the hybrid model, proving it copes well with such weather conditions. Improvement in risk management and hydrological planning depends on selecting forecasting models that fit local rainfall patterns. The limitations of this study are its use of historical data up to 2023, which projects future climatic stationarity, and the limitation of its analysis to three districts, its lack of generalizability. The model could be enhanced for future research by using more climatic predictors and having a broader geographic domain. This study's findings will help with regional water resource management, since they can be used immediately to build more accurate early warning systems and support better architecture of local water networks. Using hybrid models alongside hydrological forecasting can boost the accuracy of predicting rainfall in places with much variation, such as Ishurdi. As a result, policymakers can predict what rainfall the area might experience because of climate change, helping areas that are already vulnerable become more resistant.

6. Future Scope

Further studies should cover more places to make the findings more reliable and cover a wider range of rain patterns. Involving more climatic factors such as temperature, humidity, and climate indices (such as ENSO and IOD) can make the model more accurate. You can make your forecasts better by applying CNN or LSTM models in places where there is more variation. Having climate scenario-based simulations and real-time forecasting tools allows for better choices in water management and adapting to climate change. Also, when these impact analyses are combined with hydrological models, the plans drawn up from that information are more complete. Regional organizations might play a role in helping to use these models in building infrastructure and creating policies.

Acknowledgement: The author would like to thank the reviewer for the time and effort dedicated to reviewing this paper.

Reference

[1] Aderyani, F. R., Mousavi, S. J., and Jafari, F. (2022). Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. Journal of Hydrology, 614, 128463.

[2] Allan, R. P., and Soden, B. J. (2008). Atmospheric warming and the amplification of precipitation extremes. Science, 321(5895), 1481–1484.

[3] Alzahrani, A. S., Abdelbaki, A. M., and Mobarak, B. A. (2025). Exploring the most suitable probability distribution for analyzing annual rainfall data: a case study of Makkah and Jeddah cities. Journal of Umm Al-Qura University for Engineering and Architecture, 16(1), 52–63.

[4] Azad, M. A. K., Islam, A. R. M. T., Rahman, M. S., and Ayen, K. (2021). Development of novel hybrid machine learning models for monthly thunderstorm frequency prediction over Bangladesh. Natural Hazards, 108(1), 1109–1135.

[5] Basher, A., Islam, A. K. M. S., Stiller-Reeve, M. A., and Chu, P. (2020). Changes in future rainfall extremes over Northeast Bangladesh: A Bayesian model averaging approach. International Journal of Climatology, 40(6), 3232–3249.

[6] Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

[7] Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). An introduction to statistical modeling of extreme values (Vol. 208). Springer.

[8] Das, J., Mandal, T., Rahman, A. T. M. S., and Saha, P. (2021). Spatio-temporal characterization of rainfall in Bangladesh: an innovative trend and discrete wavelet transformation approaches. Theoretical and Applied Climatology, 143(3), 1557–1579.

[9] Di Nunno, F., Granata, F., Pham, Q. B., and de Marinis, G. (2022). Precipitation forecasting in Northern Bangladesh using a hybrid machine learning model. Sustainability, 14(5), 2663.

[10] Haan, C. T. (2002). Statistical Methods in Hydrology, 2nd Edn., 496 pp. Iowa State Press, Ames, IA.

[11] Haddad, K., Rahman, A., and Stedinger, J. R. (2012). Regional flood frequency analysis using Bayesian generalized least squares: a comparison between quantile and parameter regression techniques. Hydrological Processes, 26(7), 1008–1021.

[12] Haseeb, F., Ali, S., Ahmed, N., Alarifi, N., and Youssef, Y. M. (2025). Comprehensive Probabilistic Analysis and Practical Implications of Rainfall Distribution in Pakistan. Atmosphere, 16(2), 122.

[13] Katz, R. W. (1977). Precipitation as a chain-dependent process. Journal of Applied Meteorology (1962-1982), 671–676.

[14] Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. Advances in Water Resources, 25(8–12), 1287–1304.

[15] Khalek, A., Rahaman, M., Alam, M., and Rahman, M. S. (2023). Probability Distribution Analysis for Rainfall Scenarios: A Case Study. In G Families of Probability Distributions (pp. 344–355). CRC Press.

[16] Khan, M. M. R., Siddique, M. A. B., Sakib, S., Aziz, A., Tasawar, I. K., and Hossain, Z. (2020). Prediction of temperature and rainfall in Bangladesh using long short term memory recurrent neural networks. 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 1–6.

[17] Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., and Wehner, M. (2021). Changes in annual extremes of daily temperature and precipitation in CMIP6 models. Journal of Climate, 34(9), 3441–3460.

[18] Li, Z., Li, Z., Zhao, W., and Wang, Y. (2015). Probability modeling of precipitation extremes over two river basins in northwest of China. Advances in Meteorology, 2015(1), 374127.

[19] Makkonen, L. (2006). Plotting positions in extreme value analysis. Journal of Applied Meteorology and Climatology, 45(2), 334–340.

[20] Monir, M. M., Rokonuzzaman, M., Sarker, S. C., Alam, E., Islam, M. K., and Islam, A. R. M. T. (2023). Spatiotemporal analysis and predicting rainfall trends in a tropical monsoon-dominated country using MAKESENS and machine learning techniques. Scientific Reports, 13(1), 13933.

[21] Montes-Pajuelo, R., Rodríguez-Pérez, Á. M., López, R., and Rodríguez, C. A. (2024). Analysis of probability distributions for modelling extreme rainfall events and detecting climate change: Insights from mathematical and statistical methods. Mathematics, 12(7), 1093.

[22] Morovati, R., and Kisi, O. (2024). Utilizing hybrid machine learning techniques and gridded precipitation data for advanced discharge simulation in under-monitored river basins.

Hydrology, 11(4), 48.

[23] Noor, M., Ismail, T., Shahid, S., Asaduzzaman, M., and Dewan, A. (2021). Evaluating intensity-duration-frequency (IDF) curves of satellite-based precipitation datasets in Peninsular Malaysia. Atmospheric Research, 248, 105203.

[24] O'Gorman, P. A., and Schneider, T. (2009). Scaling of precipitation extremes over a wide range of climates simulated with an idealized GCM. Journal of Climate, 22(21), 5676–5685.

[25] Rahman, M. M., and Abdullah, S. M. A. (2022). Analysis of long-term rainfall trends in Bangladesh.

[26] Rahman, M. N., and Azim, S. A. (2023). Spatiotemporal evaluation of rainfall trend during 1979-2019 in seven climatic zones of Bangladesh. Geology, Ecology, and Landscapes, 7(4), 340–355.

[27] Rizvee, M. A., Arju, A. R., Al-Hasan, M., Tareque, S. M., and Hasan, M. Z. (2020). Weather forecasting for the north-western region of bangladesh: a machine learning approach. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–6.

[28] Roy, R., Haque, D. M. E., Sifa, S. F., Tasnim, S., Mahmud, T., and Hayat, T. (2022). Empirical approach based rainfall threshold estimation for landslide occurrence in Cox's Bazar district, Bangladesh. The Dhaka University Journal of Earth and Environmental Sciences, 11(1), 81–94.

[29] Shu, C., and Ouarda, T. B. M. J. (2008). Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology, 349(1–2), 31–43.

[30] Shuvo, S. P., Hossain, M. S., Adhikary, S. K., Mondol, S. C., and Rana, M. M. (n.d.). ENHANCED PREDICTION OF RAINFALL USING A HYBRID MACHINE LEARNING APPROACH-A CASE STUDY IN KHULNA, BANGLADESH.

[31] Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J. (2019). A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction. Hydrology and Earth System Sciences, 23(8), 3247–3268.

[32] Te Chow, V., Maidment, D. R., and Mays, L. W. (1988). Solutions Manual to Accompany Applied Hydrology. McGraw-Hill.

[33] Trenberth, K. E. (2011). Changes in precipitation with climate change. Climate Research, 47(1–2), 123–138.

[34] Wilks, D. S. (1990). Maximum likelihood estimation for the gamma distribution using data containing zeros. Journal of Climate, 1495–1501.

[35] Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.