Scientific Research Review

RESEARCH Open Access

Nonlinear Drivers of Extreme Rainfall in Barishal, Bangladesh: Evidence from an Integrated Statistical and Machine Learning Framework

Md Omar Farukh[1,*]

** Correspondence: Applied Science in Organizational Management and Leadership, Eastern Florida State College, USA; mdomar.farukh.01@gmail.com*

1 *Affiliation 1; e-mail@e-mail.co Applied Science in Organizational Management and Leadership, Eastern Florida State College, USA; mdomar.farukh.01@gmail.com m*

Abstract: Understanding the drivers of extreme rainfall is essential for assessing climate risks in vulnerable coastal regions such as Barishal, Bangladesh. While previous studies have primarily focused on average rainfall patterns, comparatively little attention has been given to rainfall extremes and land–atmosphere interactions. This study addresses this gap by integrating statistical modeling, machine learning, and explainable artificial intelligence (XAI) to identify the key climatic drivers of extreme rainfall. Monthly climatic data from 1984 to 2025 (N = 504) were obtained from the NASA POWER database, including temperature, relative humidity, wind speed, solar radiation, and soil moisture. Extreme rainfall events were defined using the 90th percentile of the rainfall distribution, a widely adopted threshold for identifying high-impact precipitation events while maintaining sufficient observations for robust statistical analysis. Logistic regression and Random Forest models were employed to capture both linear and nonlinear relationships, and SHAP (SHapley Additive exPlanations) was used to evaluate the contribution of individual predictors. The results indicate that temperature ($\beta = 3.501$, $p = 0.021$) and soil moisture ($\beta = 49.788$, $p < 0.001$) significantly increase the likelihood of extreme rainfall, whereas solar radiation ($\beta = -1.064$, $p < 0.001$) has a significant negative effect. The Random Forest model outperformed logistic regression, achieving an AUC of 0.976 compared with 0.958, an overall accuracy of 85.6%, and an F1-score of 0.783. SHAP analysis identified soil moisture as the most influential predictor, followed by humidity and temperature. Overall, the findings provide an interpretable and robust framework for understanding extreme rainfall dynamics and highlight the potential of soil moisture as an early warning indicator for climate risk management and flood preparedness in Bangladesh.

Keywords: Extreme rainfall prediction, Soil moisture, NASA POWER data, Bangladesh climate modeling

Check for updates

Citation: Farukh, M. O. (2026). Nonlinear Drivers of Extreme Rainfall in Barishal, Bangladesh: Evidence from an Integrated Statistical and Machine Learning Framework. *Scientific Research Review*, 1(1), 20-31. https://doi.org/10.66348/srr.26.a65

Received: 2026-05-05
Revised: 2026-07-26
Accepted: 2026-08-30
Online Available: 2026-09-12
Published: 2026-09-30

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Under changing climatic conditions, extreme rainfall events have become frequent and intense and have become a significant threat to human life, infrastructure, and agricultural systems globally (Alexander et al., 2006). These events are particularly critical in South Asia, a region highly vulnerable to climate variability due to its dense population, monsoon-dependent economy, and complex hydro-climatic systems (Singh et al., 2021). Bangladesh, as a low-lying deltaic country, is especially exposed to rainfall-induced hazards such as floods, waterlogging, and crop damage, making accurate understanding and prediction of rainfall dynamics a national priority (Rahman & Lateh, 2017; Bhuyan et al., 2018). Rainfall is a highly complex phenomenon, with strong seasonality, non-linear relationships, and chaotic behavior due to interactions between atmospheric and land surface processes (Dhanya & Kumar, 2011; Zhang et al., 2018). Traditional statistical approaches, including time series models

such as ARIMA and regression-based techniques, have been widely applied to analyze rainfall trends and variability (Masum et al., 2022; Rahman & Lateh, 2017). Although these techniques can offer useful information on long-term trends, they in most cases fail to capture any nonlinear relationship and interaction between multiple climatic variables which limits their predictive capabilities in highly dynamic systems.

Over the past few years, machine learning (ML) methods have been paying their dues as far as rainfall prediction is concerned due to their capability to model complex and nonlinear relationships without necessarily making strict assumptions about the processes that generate the data at hand (Mohammed et al., 2020; Cramer et al., 2017). Studies have demonstrated the superior performance of ML models such as Random Forest, Support Vector Machines, and ensemble techniques in capturing rainfall variability across different climatic regions (Meenal et al., 2021; Hasany et al., 2024). Hybrid and feature-enhanced ML models also have been further improved in predictive accuracy by incorporating diverse climatic inputs in tropical and monsoon-dominated regions only, as the name itself indicates (Liu, 2025). Besides, more recent developments in deep learning and ensemble models have demonstrated encouraging results in predicting extreme weather events, such as heavy rainfall, with better classification accuracy and stability (Amnuaylojaroen, 2026; Heydarizad et al., 2025). Although these developments have been made, a major constraint of most existing studies is that they have mostly focused on the average rainfall prediction but not on extreme events. Extreme rainfall, representing the tail behavior of precipitation distributions, is fundamentally different in its underlying mechanisms and impacts, yet remains relatively underexplored in empirical modeling frameworks (Ramos Filho et al., 2021). Additionally, most studies emphasize atmospheric variables such as temperature and humidity while overlooking the role of land surface processes, particularly soil moisture, which plays a crucial role in land–atmosphere feedback mechanisms influencing precipitation extremes.

The second emerging issue in climate modeling is the uninterpretability of advanced machine learning models. Although complex models tend to have high predictive accuracy, they do not reveal much about the relative importance and interaction of predictors. Recent developments in explainable artificial intelligence (XAI), including SHAP and partial dependence methods, have addressed this limitation by enabling transparent interpretation of model outputs (Dikshit & Pradhan, 2021; Labe & Barnes, 2021). These approaches have shown promise in identifying dominant climate drivers and uncovering nonlinear relationships, though their application in rainfall and extreme precipitation studies remains limited (Ibebuchi, 2025).

Moreover, as applied to the context of Bangladesh, most of the available research has been limited to trend analysis and forecasting using traditional statistical or machine learning models, with little to no integration of interpretable frameworks and land–atmosphere interactions (Rahman et al., 2026; Hasany et al., 2024). This indicates a clear gap in the literature regarding the combined use of statistical modeling, machine learning, and explainability techniques to investigate the drivers of extreme rainfall events. To fill these gaps, this research paper will design an integrated analytical framework that integrates both statistical and machine learning methods to model extreme rainfall events in Barishal, Bangladesh. In particular, extreme rainfall is defined in terms of percentile-based threshold, and logistic regression and Random Forest models are used to capture linear and nonlinear relationships among climatic variables. The work also integrates explainable AI methods to measure the relative significance of predictors, as well as to unravel underlying mechanisms that explain extreme rainfall.

Threefold contributions of this study are its main contributions. First, it changes the emphasis on average rainfall to extreme rainfall events that offer a more relevant viewpoint on climate risk assessment. Second, it explicitly includes land surface variables, especially soil moisture, to account land-atmosphere interactions that have in the past been over-

looked. Third, it integrates interpretable machine learning techniques with traditional statistical models, offering both predictive accuracy and scientific interpretability. In sum, the study can be utilized in the growing body of literature on climate analytics because it provides a solid, interpretable, and data-driven framework of understanding the dynamics of extreme rainfall in monsoon-dominated regions.

2. Data and Methods

2.1. Study Area and Data Source

The empirical study is based in Barishal, a coastal region in the southern part of Bangladesh, which is characterized by a tropical monsoon climate and high exposure to hydrological extremes. Due to its low-lying deltaic topography and proximity to the Bay of Bengal, the region is particularly vulnerable to heavy rainfall, flooding, and waterlogging.

Monthly climatic data covering the period from 1984 to 2025 were obtained from the NASA POWER (Prediction Of Worldwide Energy Resources) database. This dataset consists of uniform, satellite-based meteorological variables that are suitable for long-term climatic studies. The variables considered in this study include total precipitation (rainfall), near-surface air temperature, relative humidity, wind speed, solar radiation, and soil moisture. The final dataset comprised 504 monthly observations

2.2. Data Processing and Variable Construction

The raw data were initially structured in a wide format and subsequently transformed into a panel structure indexed by year and month. All climatic variables were merged into a unified dataset using (YEAR, MONTH) as key identifiers. To preserve temporal ordering, the dataset was sorted chronologically, and a time index, t = 1, 2, ..., 504, was constructed.

Seasonality was accounted for by converting the month variable into a categorical factor, allowing the model to capture seasonal variations. The data were then divided into training and testing sets to evaluate out-of-sample predictive performance. Specifically, the initial 400 observations were used for model estimation, whereas the remaining 104 observations were reserved for model testing.

Let X_t = (tempt, humidity_t, wind_t, solar_t, soil_t) denote the vector of explanatory variables at time t.

2.3. Definition of Extreme Rainfall

To focus on extreme precipitation events, a binary indicator variable was constructed based on a percentile threshold approach. Let $Q_{0.9}$ denote the 90th percentile of rainfall in the training dataset. The extreme rainfall variable Y_t is defined as:

$$Y_t = \begin{cases} 1, if\ rain_t > Q_{0.9} \\ 0, otherwise \end{cases} \quad (1)$$

This formulation ensures that approximately 10% of the observations are classified as extreme events, capturing the upper tail of the rainfall distribution. Such threshold-based definitions are widely used in climate risk analysis to identify high-impact precipitation events. The use of percentile-based thresholds provides a consistent and statistically robust approach for defining extreme events across different climatic regions. Although higher thresholds, such as the 95th or 99th percentile, are often adopted to identify more severe extremes, the 90th percentile is widely considered appropriate for representing high-impact events while retaining a sufficient number of observations for reliable statistical analysis. Previous research has also demonstrated a close relationship between selected percentile thresholds and the recurrence of extreme events, supporting their use in climate-related studies (Camuffo et al., 2020).

2.4. Logistic Regression Model

To model the probability of extreme rainfall occurrence, a binary logistic regression model is employed. Let $P(Y_t=1|X_t)$ denote the conditional probability of an extreme rainfall event at time ttt. The logistic model is specified as:

$$logit\big(P(Y_t = 1)\big) = ln\left(\frac{P(Y_t=1)}{1-P(Y_t=1)}\right) = \beta_0 + \beta_1 temp_t + \beta_2 humidity_t + \beta_3 wind_t + \beta_4 solar_t + \beta_5 soil_t + \sum_{m=2}^{12} \gamma_m Dm\,, t \quad (2)$$

where Dm, t presents monthly dummy variables capturing seasonal effects.

The logistic regression model provides an interpretable framework in which the coefficients β_i quantify the marginal effect of each predictor on the log-odds of extreme rainfall. Logistic regression was selected because it provides a transparent statistical framework for assessing the direction and magnitude of associations between climatic variables and the probability of extreme rainfall occurrence. During model estimation, evidence of potential quasi-complete separation was observed due to strong seasonal patterns and sparse extreme events. Therefore, a Firth penalized logistic regression model was additionally estimated as a robustness check to assess the stability of the estimated effects. The results of the robustness analysis were consistent with those obtained from the standard logistic regression model.

2.5. Random Forest Model

To capture nonlinear relationships and complex interactions among climatic variables, a Random Forest (RF) classification model is employed. Random Forest is an ensemble learning algorithm that constructs multiple decision trees using bootstrap samples of the training data and combines their predictions through majority voting.

The Random Forest prediction is given by

Formally, le $\{T_b(X_t)\}_{b=1}^{B}$ tdenote a collection of B decision trees. The Random Forest prediction is given by:

$$\{\widehat{Y}\}_t = \{majority\ vote\}\{\ T_1(X_t), T_2(X_t), \dots, T_B\ (X_t) \quad (3)$$

In this study, *B=500* trees were used. At each split, a random subset of predictors was considered, reducing the correlation among individual trees and improving the model's generalization performance.

Random Forest was selected because it can effectively capture nonlinear relationships and higher-order interactions among climatic variables without requiring parametric assumptions regarding the underlying data-generating process. This makes it particularly suitable for modelling complex climate systems where relationships among meteorological variables are often nonlinear. Furthermore, Random Forest provides robust predictive performance while allowing model interpretation through explainable artificial intelligence techniques such as SHAP, thereby complementing the inferential capability of the logistic regression model.

2.6. Model Evaluation Metrics

To evaluate the predictive performance of the models, several evaluation metrics were used, including the confusion matrix, accuracy, sensitivity, specificity, precision, F1-score, Cohen's kappa, and the Receiver Operating Characteristic (ROC) curve with its corresponding Area Under the Curve (AUC).

Let $\widehat{Y_t}$ denote the predicted class and Y_t the observed class. The confusion matrix is defined in terms of:

- **True Positives (TP):** correctly predicted extreme rainfall events
- **True Negatives (TN):** correctly predicted non-extreme rainfall events
- **False Positives (FP):** non-extreme rainfall events incorrectly classified as extreme
- **False Negatives (FN):** extreme rainfall events incorrectly classified as non-extreme

Based on these quantities, the following performance metrics were computed.

Accuracy

$$Accuracy = \frac{TP+TN}{FP+FNTP+TN}$$

Sensitivity (Recall)

$$Sensitivity = \frac{TP}{TN + FN}$$

Specificity

$$Accuracy = \frac{TP + TN}{FP + FNTP + TN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

F1-score

$$F1 = \frac{2 \times (Precision \times Sensitivity)}{Precision + Sensitivity}$$

Cohen's Kappa

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the observed agreement and p_e is the expected agreement by chance.

Given the imbalanced nature of the dataset (approximately 10% extreme rainfall events), relying solely on accuracy may provide a misleading assessment of model performance. Therefore, sensitivity, specificity, precision, F1-score, and the ROC curve were additionally used to provide a more comprehensive evaluation of the classification performance.

The ROC curve is constructed by plotting the True Positive Rate against the False Positive Rate, where

$$\text{FPR} = \frac{\text{FP}}{FP + TN}$$

The Area Under the Curve (AUC) is defined as

$$\text{AUC} = \int_0^1 TPR(FPR - 1(x))dx.$$

A higher AUC value indicates better model performance in distinguishing between extreme and non-extreme rainfall events.

2.7. Model Interpretability and Explainable Artificial Intelligence

To enhance the interpretability of the machine learning model, explainable artificial intelligence (XAI) techniques were employed. In particular, SHapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDP) were used to quantify the contribution of each predictor and to visualize their marginal effects on extreme rainfall prediction.

2.8. SHAP Values

SHAP values are based on cooperative game theory and provide a unified measure of feature importance by attributing the contribution of each feature to the model prediction. For a given feature iii, the SHAP value $\boldsymbol{\varphi_i}$ is defined as:

$$\boldsymbol{\varphi_i} = \sum_{\boldsymbol{S} \subseteq \boldsymbol{F}\ \{i\}} \left(\frac{|\boldsymbol{S}|!\ (|\boldsymbol{F}| - |\boldsymbol{S}| - \mathbf{1})!}{|\boldsymbol{F}|!} \right) [\, \boldsymbol{f}(\boldsymbol{S} \cup \{\boldsymbol{i}\}) - \boldsymbol{f}(\boldsymbol{S})]$$

Where, $\boldsymbol{F}$ represents the set of all features and $\boldsymbol{S}$ denotes a subset of features excluding the feature i. The SHAP value quantifies the average marginal contribution of each feature across all possible feature combinations.

Global feature importance is obtained by computing the mean absolute SHAP values:

$$\boldsymbol{Importance_t} = \boldsymbol{E}(|\boldsymbol{\varphi_t}|)$$

3. Results

3.1. Descriptive Resuslts

The time series of monthly rainfall shows substantial temporal variation with a strong monsoonal pattern. Rainfall is concentrated mainly between June and September, with July and August showing the highest rainfall levels and greater variability. The seasonal distribution indicates that rainfall extremes are strongly associated with the monsoon period.

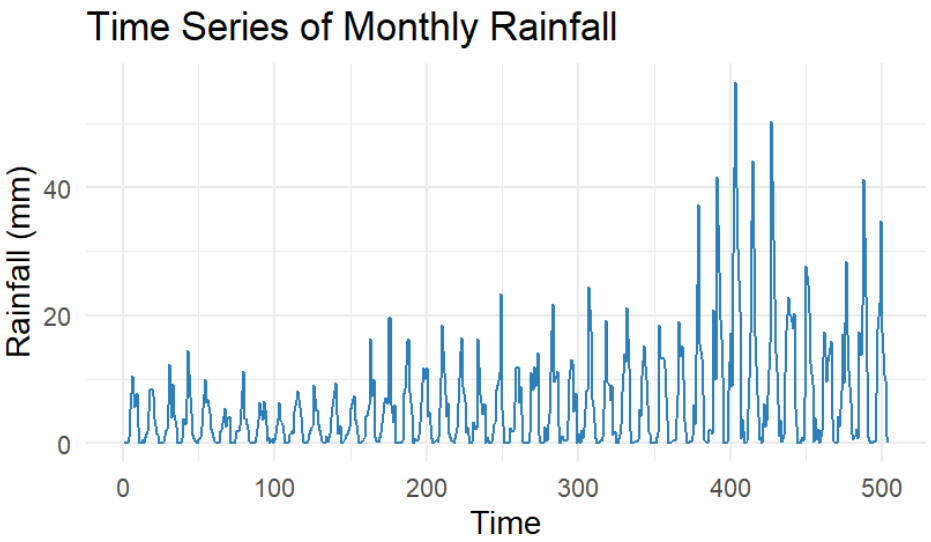


Figure 1: Time series of monthly rainfall (1984–2025)

The time series of monthly rainfall indicates that there is a significant temporal variation with strong peaks during the monsoon seasons. Although the level and the rate of extreme rainfall events have been relatively low and constant in the earlier years, the intensity and frequency of extreme rainfall events increases noticeably after the mid-2000s. This indicates that the trend involves an increase in variability and extreme events as opposed to a linear trend.

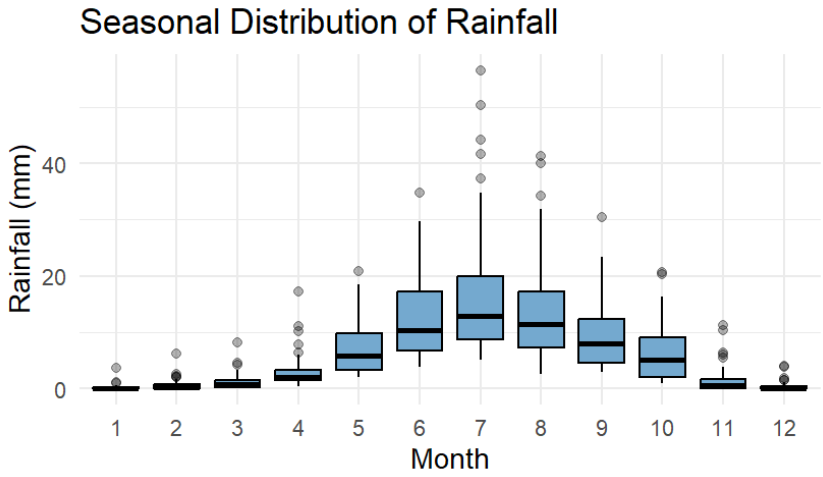


Figure 2: Seasonal distribution of rainfall

The seasonal distribution shows that there is a strong monsoonal pattern with rainfall being focused between June and September. July and August are the months with the highest median rainfall, and there is a high interquartile range and a number of extreme values that exceed 40 mm. Conversely, the dry season months have very low rainfall with very low dispersion, indicating the prevalence of seasonal dynamics.

3.2. Logistic Regression Results

The logistic regression results indicate that temperature, solar radiation, and soil moisture are significant predictors of extreme rainfall. Temperature has a positive association with extreme rainfall ($\beta = 3.501$, $p = .021$), while solar radiation has a significant negative association ($\beta = -1.064$, $p < .001$). Soil moisture shows the strongest positive association ($\beta = 49.788$, $p < .001$). Humidity and wind speed are not statistically significant.

The odds ratios and corresponding 95% confidence intervals are presented in Table 1. Evidence of quasi-complete separation was observed in the standard logistic regression, particularly among the seasonal variables. Therefore, a Firth penalized logistic regression was estimated as a robustness check. The Firth model produced consistent results for the main climatic predictors, with temperature and soil moisture remaining positively associated with extreme rainfall and solar radiation remaining negatively associated.

Table 1. Logistic Regression Results for Extreme Rainfall

Predictor	β	Odds Ratio	95% CI for OR	p
Temperature	3.501	33.20	2.09– 4.4975	.021
Humidity	−0.254	0.78	0.39–0.558	.476
Wind speed	1.913	6.78	0.65–4.0726	.125
Solar radiation	−1.064	0.35	0.18– 0.3728	< .001
Soil moisture	49.788	4.19×10^{21}	7.06×10^{12}– 2.56×10^{21}	< .001

Note. β = logistic regression coefficient; OR = odds ratio; CI = confidence interval. Monthly dummy variables were included in the model to account for seasonal effects but are not displayed in the table.

3.3. Random Forest Model Performance

The Random Forest model, consisting of 500 trees, achieved an out-of-bag error rate of 6.25%. On the independent test dataset, the model achieved an accuracy of 86.54% and an AUC of 0.9768. The model also achieved a precision of 96.55%, recall of 68.29%, specificity of 98.41%, F1-score of 0.800, balanced accuracy of 0.8335, and Cohen's Kappa of 0.703. These results indicate strong discrimination between extreme and non-extreme rainfall events.

Table 2. Random Forest Classification Performance

Performance Metric	Value
Accuracy	86.54%
Precision	96.55%
Recall (Sensitivity)	68.29%
Specificity	98.41%
F1-score	0.800
Balanced Accuracy	0.8335
Cohen's Kappa	0.703
AUC	0.9768
OOB Error Rate	6.25%

Note. Precision, recall, specificity, and F1-score are reported with extreme rainfall (class 1) treated as the positive class.

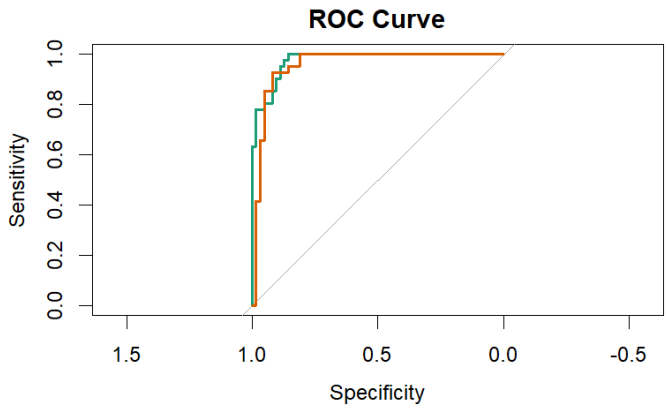


Figure 3: ROC curve comparing Logistic Regression and Random Forest

The ROC analysis further confirms model performance. The Random Forest achieves an AUC of 0.976, outperforming logistic regression (AUC = 0.958), indicating superior discriminative ability and better handling of nonlinear relationships.

3.4. Feature Importance Analysis

SHAP-based feature importance analysis identifies soil moisture as the most influential predictor of extreme rainfall, with a mean absolute SHAP contribution of 0.274. This is followed by humidity (0.050), temperature (0.042), wind speed (0.021), solar radiation (0.021), and month factor (0.019). The substantially higher contribution of soil moisture indicates its dominant role in the Random Forest predictions.

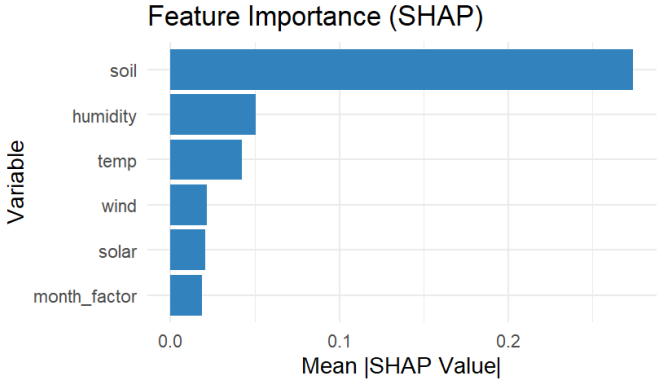


Figure 4: Feature importance based on SHAP values

SHAP-based feature importance reveals that soil moisture is the dominant predictor of extreme rainfall, with a mean absolute contribution of 0.274.

4. Discussion

This research provides new empirical evidence on the drivers of extreme rainfall in a monsoon-dominated coastal region, highlighting the importance of nonlinear interactions and land–atmosphere feedback processes. The results show that extreme rainfall cannot be explained entirely by atmospheric conditions but is also strongly influenced by land surface processes, particularly soil moisture.

The finding that soil moisture is the most important predictor is consistent with previous studies emphasizing the significance of land–atmosphere coupling in precipitation processes (Dhanya & Kumar, 2011; Ibebuchi, 2025). In tropical climates, soil moisture controls evapotranspiration and moisture availability in the boundary layer, which in turn influence convective processes. Higher soil moisture increases latent heat flux, thereby enhancing atmospheric instability and increasing the likelihood of heavy rainfall events. Similar mechanisms have been widely reported in studies investigating climate variability and nonlinear precipitation processes.

The positive association between temperature and extreme rainfall aligns with the thermodynamic theory of precipitation intensification. Warmer air masses can hold more moisture, thereby increasing the likelihood of extreme precipitation events, a phenomenon widely documented in climate change literature (Alexander et al., 2006; Singh et al., 2021). However, the interaction between temperature and land surface conditions appears to be crucial, suggesting that temperature alone is insufficient to explain rainfall extremes without considering the underlying surface processes.

The negative relationship between solar radiation and extreme rainfall reflects reduced incoming solar radiation during periods of increased cloud cover and convective activity. This finding is consistent with established meteorological understanding, whereby enhanced cloud formation reduces solar radiation while simultaneously promoting precipitation processes.

The superior performance of the Random Forest model compared with logistic regression supports existing evidence that machine learning models are more effective in capturing nonlinear relationships in climatic data (Cramer et al., 2017; Meenal et al., 2021; Hasany et al., 2024). While logistic regression provides an interpretable statistical framework, it is limited in modelling complex interactions among variables. During model estimation, evidence of quasi-complete separation was observed, which may affect the stability of some coefficient estimates. However, the Firth penalized logistic regression produced results that were consistent with the standard logistic regression, indicating that the principal findings of this study are robust. In contrast, ensemble learning approaches such as Random Forest effectively capture higher-order dependencies, leading to improved predictive performance.

Despite the strong predictive performance, the results also highlight challenges associated with imbalanced data, where extreme rainfall events constitute a relatively small proportion of the observations. This may lead to asymmetric classification performance, which is a common issue in extreme event prediction studies (Amnuaylojaroen, 2026). The inclusion of precision, recall, F1-score, and AUC provides a more comprehensive evaluation of model performance under class imbalance, while the use of AUC offers a threshold-independent measure of model discrimination.

From a regional perspective, these findings have important implications for Bangladesh. Previous studies have primarily focused on trend analysis and average rainfall patterns (Rahman & Lateh, 2017; Bhuyan et al., 2018), with limited attention given to the drivers of extreme rainfall events. The present study extends this literature by demonstrating that soil moisture plays a critical role in extreme rainfall occurrence. In a deltaic environment such as Barishal, where drainage capacity and water retention are key concerns, elevated soil moisture may substantially amplify flood risks. Furthermore, soil

moisture has the potential to serve as an early warning indicator, allowing disaster management agencies to anticipate periods of elevated flood risk and improve preparedness measures.

Methodologically, the integration of explainable machine learning techniques enhances the interpretability of predictive models, addressing a key limitation in climate modelling (Dikshit & Pradhan, 2021; Labe & Barnes, 2021). By combining statistical models with explainable artificial intelligence techniques, this study provides a more comprehensive understanding of the mechanisms underlying extreme rainfall while maintaining both predictive accuracy and model interpretability.

Nevertheless, several limitations should be acknowledged. The use of monthly aggregated data may mask short-term variability and the dynamics of extreme rainfall events. In addition, the monthly observations exhibit temporal dependence, which is an inherent characteristic of climatic time-series data and should be considered when interpreting the results. Furthermore, the percentile-based definition of extreme rainfall, although widely adopted (Ramos Filho et al., 2021), may not fully capture physically defined rainfall extremes across different climatic settings. Future research could incorporate higher temporal resolution data and investigate spatial heterogeneity across different regions of Bangladesh.

Overall, the findings highlight the importance of nonlinear interactions and land–atmosphere feedback processes in explaining extreme rainfall dynamics. This study contributes to the growing literature by providing an interpretable, data-driven framework that integrates statistical modelling, machine learning, and explainable artificial intelligence to improve the understanding of extreme rainfall in climate-vulnerable regions.

5. Conclusion

This study examined the drivers of extreme rainfall in a monsoon-dominated coastal region using an integrated framework that combines statistical modelling, machine learning, and explainable artificial intelligence techniques. By focusing on extreme rainfall events rather than average precipitation, the study provides a more policy-relevant perspective for understanding climate risk.

The results demonstrate that extreme rainfall is driven by complex and nonlinear interactions among climatic variables. Among all predictors, soil moisture emerged as the most influential factor, highlighting the critical importance of land–atmosphere feedback mechanisms. Temperature and solar radiation were also identified as significant predictors, reflecting the thermodynamic and atmospheric conditions associated with precipitation extremes. Furthermore, the superior predictive performance of the Random Forest model compared with logistic regression confirms the importance of modelling nonlinear relationships in complex climate systems.

From a practical perspective, these findings have important implications for climate risk management in Bangladesh and other monsoon-dominated regions. The strong influence of soil moisture suggests that antecedent wet conditions may serve as an effective early warning indicator of extreme rainfall risk. Incorporating soil moisture information into forecasting systems could improve early warning capabilities and strengthen disaster preparedness, particularly in flood-prone coastal regions.

Methodologically, this study contributes to the existing literature by integrating interpretable machine learning techniques with traditional statistical models, thereby providing both predictive accuracy and model transparency. The proposed analytical framework can be extended to other regions and climatic variables to improve the understanding of extreme weather dynamics.

Finally, understanding extreme rainfall requires moving beyond traditional linear assumptions by explicitly considering nonlinear interactions and land surface processes. The integrated framework presented in this study provides a robust foundation for future

research and supports the development of more effective climate adaptation and disaster risk reduction strategies.

Declarations

Ethics and Guidelines: Not Applicable.

Consent to participate: Not Applicable.

Consent to publish: The authors have provided consent to publish.

Competing interests: The authors declare no competing interests.

Data availability statement: Data will be made available on reasonable request from the corresponding author.

Funding: This research received no external funding.

Clinical Trial Number: Not Applicable.

Declaration of using generative AI: During the preparation of this work the author(s) used ChatGPT in order to correct the grammatical errors. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content.

Author Contributions: M.O.F.: Conceptualization, methodology, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript of the published article.

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres, 111*(D5). https://doi.org/10.1029/2005JD006290

Amnuaylojaroen, T. (2026). Enhancing extreme weather event classification ensemble machine learning modelling: An application to mainland Southeast Asia region. *Theoretical and Applied Climatology, 157*(1), 8. https://doi.org/10.1007/s00704-025-05925-x

Bhuyan, M. D. I., Islam, M. M., & Bhuiyan, M. E. K. (2018). A trend analysis of temperature and rainfall to predict climate change for northwestern region of Bangladesh. *American Journal of Climate Change, 7*(2), 115–134. https://doi.org/10.4236/ajcc.2018.72009

Camuffo, D., Becherini, F., & della Valle, A. (2020). Relationship between selected percentiles and return periods of extreme events. *Acta Geophysica, 68*, 1201–1211. https://doi.org/10.1007/s11600-020-00452-x

Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications, 85*, 169–181. https://doi.org/10.1016/j.eswa.2017.05.029

Dhanya, C., & Kumar, D. N. (2011). Multivariate nonlinear ensemble prediction of daily chaotic rainfall with climate inputs. *Journal of Hydrology, 403*(3-4), 292–306. https://doi.org/10.1016/j.jhydrol.2011.04.006

Dikshit, A., & Pradhan, B. (2021). Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of the Total Environment, 801*, 149797. https://doi.org/10.1016/j.jhydrol.2011.04.009

Hasany, N., Salman, H., & Farman, A. (2024). Evaluating rainfall prediction models on time series data from Bangladesh and Pakistan: A comparative approach. *Journal of Computing & Biomedical Informatics, 7*(2). https://jcbi.org/index.php/Main

Heydarizad, M., Pumijumnong, N., Minaei, M., Salari, P., Sorí, R., & Mohammadabadi, H. G. (2025). Exploring stable isotope patterns in monthly precipitation across Southeast Asia using contemporary deep learning models and SHapley Additive exPlanations (SHAP) techniques. *Isotopes in Environmental and Health Studies, 61*(5), 547–568. https://doi.org/10.1080/10256016.2025.2508811

Ibebuchi, C.C. Uncertainty in machine learning feature importance for climate science: a comparative analysis of SHAP, PDP, and gain-based methods. *Theor Appl Climatol* **156**, 476 (2025). https://doi.org/10.1007/s00704-025-05703-9

Labe, Z. M., & Barnes, E. A. (2021). Detecting climate signals using explainable AI with single-forcing large ensembles. *Journal of Advances in Modeling Earth Systems, 13*(6), e2021MS002464. https://doi.org/10.1029/2021MS002464

Liu, G. (2025). Tropical climate prediction method combining random forest and feature fusion. *International Journal of Low-Carbon Technologies, 20*, ctae281. https://doi.org/10.1093/ijlct/ctae281

Masum, M. H., Islam, R., Hossen, M. A., & Akhie, A. A. (2022). Time Series Prediction of Rainfall and Temperature Trend using ARIMA Model. *Journal of Scientific Research, 14*(1), 215-227. https://doi.org/10.3329/jsr.v14i1.54973

Meenal, R., Michael, P. A., Pamela, D., & Rajasekaran, E. (2021). Weather prediction using random forest machine learning model. *Indonesian Journal of Electrical Engineering and Computer Science, 22*(2), 1208–1215. https://doi.org/10.11591/ijeecs.v22.i2.pp1208-1215

Mohammed, M., Kolapalli, R., Golla, N., & Maturi, S. S. (2020). Prediction of rainfall using machine learning techniques. *International Journal of Scientific and Technology Research, 9*(1), 3236–3240.

Rahman, M. R., & Lateh, H. (2017). Climate change in Bangladesh: A spatio-temporal analysis and simulation of recent temperature and rainfall data using GIS and time series analysis model. *Theoretical and Applied Climatology*, *128*(1), 27–41. https://doi.org/10.1007/s00704-015-1688-3

Rahman, N. M. F., Haque, N., Asadullah, M., Hossain, A. B. M. Z., Hossain, M. S., Uddin, M. J., Hassan, S. Q., & Baten, M. A. (2026). Machine learning-based temperature prediction across diverse ecosystems for the Boro season in Bangladesh. *Scientific Reports*. https://doi.org/10.1038/s41598-026-46341-z

Ramos Filho, G. M., Coelho, V. H. R., Freitas, E. d. S., Xuan, Y., & Almeida, C. d. N. (2021). An improved rainfall-threshold approach for robust prediction and warning of flood and flash flood hazards. *Natural Hazards*, *105*(3), 2409–2429. https://doi.org/10.1007/s11069-020-04405-x

Singh, B. B., Singh, M., & Singh, D. (2021). An overview of climate change over South Asia: Observations, projections, and recent advances. In *Practices in Regional Science and Sustainable Regional Development: Experiences from the Global South* (pp. 263–277). Springer. https://doi.org/10.1007/978-981-16-2221-2_12

Zhang, Z., Zhang, K., & Khelifi, A. (2018). *Multivariate time series analysis in climate and environmental research*. Springer. https://doi.org/10.1007/978-3-319-67340-0

Disclaimer: All statements, viewpoints, and data featured in the publications are exclusively those of the individual author(s) and contributor(s), not of SRR and/or its editor(s). SRR and/or the editor(s) absolve themselves of any liability for harm to individuals or property that might arise from any concepts, methods, instructions, or products mentioned in the content.