Group No : E5

Flood Prediction and Risk Analysis in Bangladesh Using Rainfall and River Water Level Data

Sayed Sadman Shahriar*, Talha Jubayer Asif, Iftekhar Ahmed, Md Tasnim Anam, Saikat kotal Shuvo
Department of Computer Science, American International University , Bangladesh (AIUB), Dhaka, 1229, Bangladesh
Corresponding author: 23-54587-3@student.aiub.edu

Abstract

Bangladesh is one of the most flood-prone countries in the world. To mitigate the recurrent flood damage, flood risk analysis and prediction are essential tools in developing non-structural disaster management strategies. While national-level flood forecasting typically focuses on river water levels and large-scale inundation, this study aims to statistically analyze rainfall and river water level data to understand flood occurrence patterns across four vulnerable districts of Bangladesh. The study focuses on the districts of Sylhet, Feni, Cumilla, and Rangpur, which experience seasonal floods influenced by transboundary river systems and upstream rainfall from neighboring Indian states.

This study uses monthly rainfall data from Bangladesh and four Indian states (Assam, Meghalaya, Tripura and Sikkim), alongside maximum yearly river water levels and total flood-affected area data from 2019 to 2024. Statistical analysis including descriptive statistics, correlation analysis, conditional probability (Bayes' theorem), random variable modeling, hypothesis testing, and simple linear regression were applied to quantify the relationship between rainfall, river water levels, and flood impacts. The analysis reveals that river water level is a stronger indicator of flood extent than rainfall alone, although upstream rainfall in Indian states shows a moderate influence. Linear regression models and probability analysis helped identify the likelihood of flood occurrence under varying rainfall and river level scenarios.

The performance of statistical tools used in this study proved effective in analyzing multi-year data trends and identifying key influencing factors of flood risk. The findings provide foundational insights for local-level flood prediction planning, and emphasize the importance of early warning systems based on rainfall and river monitoring. The methods applied in this project offer a simplified yet valuable approach to support flood awareness and disaster preparedness in vulnerable districts of Bangladesh.

Keywords Flood prediction . Rainfall analysis . River water level . Flood risk assessment . Statistical analysis . Bangladesh floods

Introduction

GENERAL

Bangladesh is one of the most flood-prone countries in the world. Due to its geographical position on low-lying floodplains formed by the Ganges, Brahmaputra, and Meghna river systems, the country experiences annual flooding triggered by heavy monsoon rainfall and transboundary river inflows. These floodplains are inhabited by a dense and predominantly rural population, whose livelihood is intricately linked to seasonal rainfall and river behavior. As a result, floods cause widespread socio-economic disruption, including the destruction of infrastructure, agricultural loss, environmental degradation, and displacement of people. Effective flood prediction and risk analysis can play a crucial role in minimizing these impacts and supporting informed disaster preparedness efforts. While Bangladesh has developed various forecasting systems at the national level, localized and data-driven risk analysis is still underutilized, especially in understanding how rainfall and river levels statistically contribute to flood risks across different districts.

BACKGROUND OF THE STUDY

Floods are one of the most frequent and devastating natural hazards in Bangladesh. They disrupt communities, damage critical infrastructure, destroy crops, and hinder long-term development. The country receives more than 90% of its surface water from upstream catchment areas in neighboring Indian states during the monsoon season (June–September). Heavy rainfall in India, combined with poor drainage and rising river water levels within Bangladesh, often results in floods that affect both rural and urban regions. Although institutions like the Flood Forecasting and Warning Centre (FFWC) under the Bangladesh Water Development Board (BWDB) provide operational flood forecasts, these are largely focused on river water level hydrographs at major gauge stations and inundation maps at coarse scales. There remains a critical need for district-level analysis that statistically connects rainfall and river level trends with flood impacts using historical data. Such analysis can inform early warning systems and improve flood preparedness, particularly in high-risk areas such as Sylhet, Feni, Cumilla, Cox's Bazar, Rangamati, and Rangpur — all of which are exposed to flood threats due to their proximity to major rivers and transboundary rainfall sources.

This study utilizes six years of historical data (2019–2024) to statistically evaluate the influence of monthly rainfall and maximum yearly river water levels on flood-affected areas in Bangladesh. It incorporates rainfall data from four Bangladeshi districts and four upstream Indian states (Assam, Tripura, Meghalaya and Sikkim). By applying descriptive statistics, probability analysis using Bayes' theorem, correlation and regression analysis, and hypothesis testing, the study aims to identify which variables most significantly affect flood occurrence and severity. Unlike hydrological model-based simulations like HEC-HMS or WRF integrations, this study is grounded in statistical methods suitable for educational and early analytical applications. Its findings can support future development of flood forecasting tools and awareness systems using accessible, data-driven approaches.

OBJECTIVES OF THE STUDY

The objectives of this study are as follows:

- To statistically analyze monthly rainfall and yearly river water level data from 2019 to 2024 in selected flood-prone districts of Bangladesh.
- To identify correlations between rainfall, river water levels, and flood-affected areas.
- To apply Bayes' theorem for calculating conditional flood probabilities based on rainfall and river level thresholds.
- To use regression analysis to model flood-affected area based on rainfall and water level data.
- To support data-driven decision-making for flood risk management and early warning strategies.

Literature Review

Bangladesh is one of the most flood-prone countries in the world due to its geographical location in the delta of three major rivers the Ganges, Brahmaputra, and Meghna. Recurrent flooding affects millions annually, damaging property, infrastructure, agriculture, and livelihoods. According to Hossain et al. (2013), more than 80% of Bangladesh's floodwater originates from upstream catchments in India, Bhutan, Nepal, and China, highlighting the need for cross-border rainfall and river monitoring.

Several studies have investigated the relationship between monsoon rainfall and flood events in Bangladesh. Rahman et al. (2017) found a significant correlation between excessive rainfall in northeastern India and major flood occurrences in the Sylhet and Rangpur regions. Similarly, Chowdhury and Karim (2016) emphasized that rainfall in upstream states such as Assam, Meghalaya, and Sikkim plays a critical role in river swelling, particularly for transboundary rivers like the Teesta and Surma.

Hydrological modeling tools such as HEC-HMS, MIKE11, and NAM have been used by IWM (Institute of Water Modelling) and BWDB (Bangladesh Water Development Board) to forecast flood dynamics by simulating rainfall-runoff processes and river flow. However, these require high computational resources and expertise. In contrast, statistical methods such as regression analysis, correlation studies, and probability-based models have been increasingly adopted for simplified flood forecasting, especially in data-limited settings (Ahmed et al., 2020).

Recent studies have also used time series rainfall data to build flood risk maps and identify critical thresholds. For example, Islam and Akter (2021) showed that average rainfall in June-August serves as a strong flood indicator for northeastern districts. Monthly rainfall and river level data can therefore be used to assess risk, especially when upstream rainfall data from Indian Meteorological Department (IMD) is incorporated.

Moreover, Khan et al. (2022) highlighted that flood impact in districts like Feni and Cumilla is often less dependent on local rainfall and more influenced by upstream river water levels, particularly for rivers like the Muhuri and Gomti. This emphasizes the importance of regional hydrological connectivity.

While most previous research focused on hydrodynamic modeling, this study applies course-based statistical tools such as descriptive statistics, correlation, conditional probability, and regression to demonstrate that rainfall and river water level trends can provide early indicators for flood risk assessment in Bangladesh.

DATA COLLECTION AND METHODOLOGY

General

To perform statistical analysis for flood prediction and risk assessment in Bangladesh, data from various sources covering the years 2019 to 2024 have been collected and organized. This includes historical monthly rainfall data, river water level records, and flood-affected area statistics. The aim is to analyze this data using statistical techniques such as central tendency, probability, hypothesis testing, correlation, and regression to identify relationships between rainfall, river levels, and flood impacts. The collected data is the foundation of this study, enabling a data-driven approach to flood risk assessment. This chapter provides an overview of the study area, data sources, and the methodology used for processing and analyzing the data.

Selection of Study Area

Bangladesh is one of the largest river deltas in the world, formed by the Ganges, Brahmaputra, and Meghna rivers, and has a basin area of approximately 1.72 million sq. km, with only 8.5% within the country's territory. The majority of Bangladesh's floodwater comes from transboundary sources, making it highly vulnerable to seasonal flooding. The country regularly experiences four types of floods: flash floods, river floods, rain-fed floods, and storm surge floods.

Figure 1 shows a detailed map of Bangladesh, where the four selected districts are highlighted. This map helps visually locate the districts within the national geography and shows their proximity to major rivers and international borders with India. It clearly demonstrates the flood risk zones.

Figure 1: Details map of Bangladesh (Source: Google)

In this study, four highly flood-prone districts were selected: Sylhet, Feni, Cumilla, and Rangpur. These districts are located near major rivers and are frequently affected by seasonal monsoon rainfall and upstream river flow from India. Sylhet and Rangpur are influenced by rainfall and river systems originating from Assam, Meghalaya and Sikkim, while Cumilla and Feni are directly impacted by rainfall in Tripura and Meghalaya. The selection of these districts allows for a geographically diverse representation of flood behavior across Bangladesh.

Figure 2 illustrates the direction of upstream to downstream water flow from Indian states into Bangladesh. This diagram is important because it shows how rainfall in Indian regions like Assam, Meghalaya, Tripura, and Sikkim flows into the rivers that pass through Sylhet, Rangpur, Cumilla, and Feni. It helps explain how cross-border river connections play a major role in flooding, even when local rainfall is low.

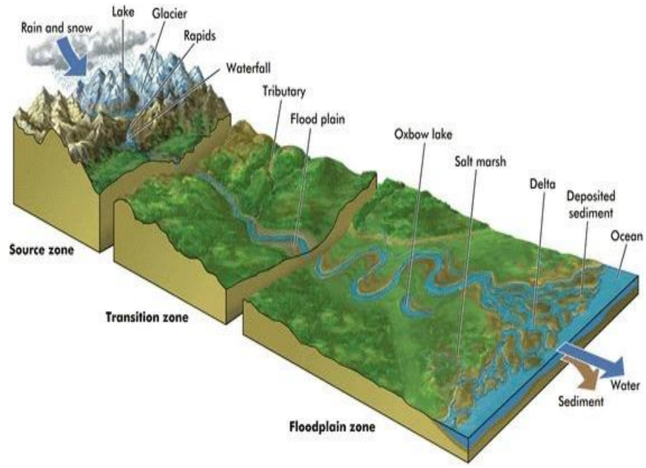

Figure 2: Upstream to Downstream Water flow

Methodology

This study uses a statistical approach to evaluate the relationship between rainfall, river water levels, and flood-affected areas. The research design follows a sequential methodology that includes:

1. Collection of monthly rainfall data for 2019–2024 from Bangladesh Bureau of Statistics (Statistical Yearbook and Monthly Bulletin) and Indian Meteorological Department rainfall statistics.
2. Collection of yearly maximum river water level data for major rivers near the study districts.
3. Collection of annual flood-affected area data for Bangladesh from Wikipedia's "Floods in Bangladesh" archive, which compiles verified government sources and media reports.
4. Data cleaning, formatting, and validation.
5. Statistical analysis using:
    - ➢ Descriptive statistics
    - ➢ Probability & conditional probability
    - ➢ Hypothesis testing
    - ➢ Correlation & regression analysis
6. Interpretation of results to assess which factors most significantly affect flood risk.

Data Collection

To perform flood risk analysis, we collected data on rainfall, river water levels, and flood-affected areas for the years 2019 to 2024. The rainfall data was gathered for four districts in Bangladesh—Sylhet, Cumilla, Feni, and Rangpur—using official publications from the Bangladesh Bureau of Statistics (BBS). These districts were selected because they are located near major rivers and often experience flooding during the monsoon season.

Figure 3 shows the monthly rainfall in Sylhet from 2019 to 2024. We can see that rainfall peaks during June, July, and August every year, which aligns with the monsoon season. The highest spikes in rainfall occurred in 2022 and 2024, which also correspond to flood years in Sylhet. This figure helps us understand how rainfall increases during flood-prone months.

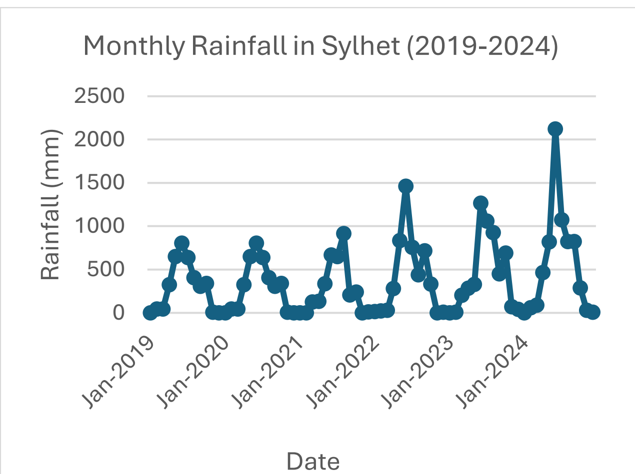

Figure 3: Monthly Rainfall in Sylhet (2019–2024) (Source: BBS)

Figure 4 displays the monthly rainfall in Cumilla for the same period. Like Sylhet, Cumilla also experiences its heaviest rainfall during the monsoon. The graph shows a significant rise in 2024, which matches with major flooding in that district. This supports the idea that rainfall has a direct effect on flood occurrences.

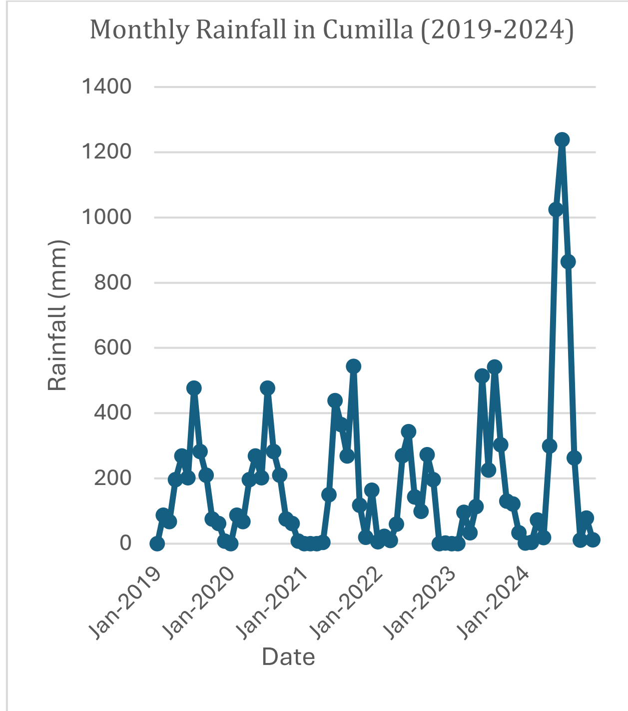

Figure 4: Monthly Rainfall in Cumilla (2019–2024) (Source: BBS).

Figure 5 represents rainfall in Feni. Interestingly, Feni shows a relatively stable rainfall pattern with no major spikes in most years. However, 2024 shows a slight increase, but not as much as the upstream Indian states. Despite the moderate local rainfall, Feni still faced flooding, which suggests that upstream rainfall and river flow may have a stronger influence.

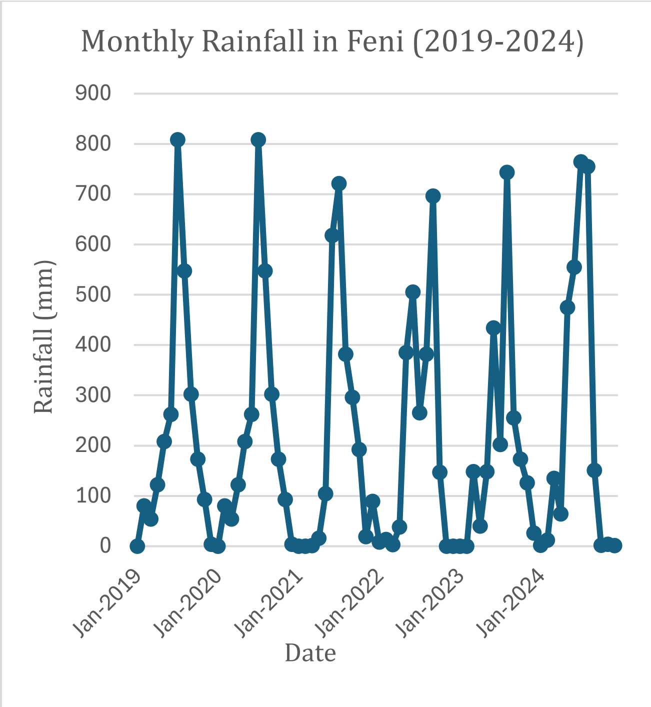

Figure 5: Monthly Rainfall in Feni (2019–2024) (Source: BBS).

Figure 6 illustrates rainfall in Rangpur. Here too, the peak rainfall occurs during monsoon months. The rainfall remained fairly consistent from year to year, which indicates that floods in Rangpur might be more influenced by upstream factors like the Teesta River's water level than by drastic local rainfall changes.

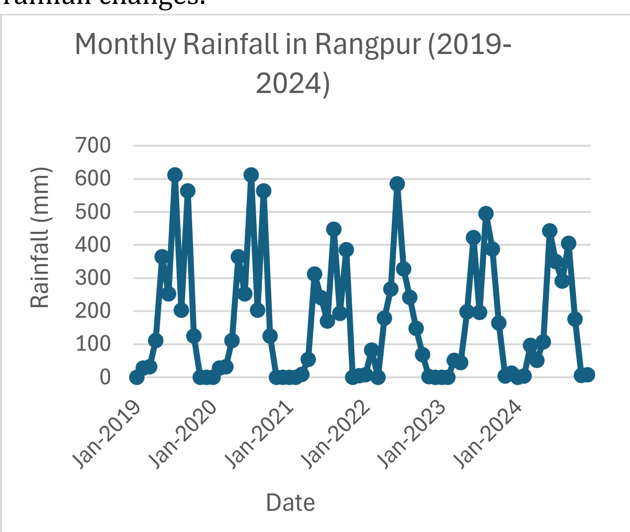

Figure 6: Monthly Rainfall in Rangpur (2019–2024) (Source: BBS).

Additional upstream rainfall data was collected from the Indian Meteorological Department (IMD) for four Indian states influencing flood behavior in Bangladesh: Assam, Tripura, Meghalaya and Sikkim.

Figure 7 shows monthly rainfall in Meghalaya. The pattern here is similar to that of Sylhet, with high rainfall during monsoon months. Meghalaya's rainfall can directly impact rivers like the Surma and Kushiyara, which flow into Sylhet.

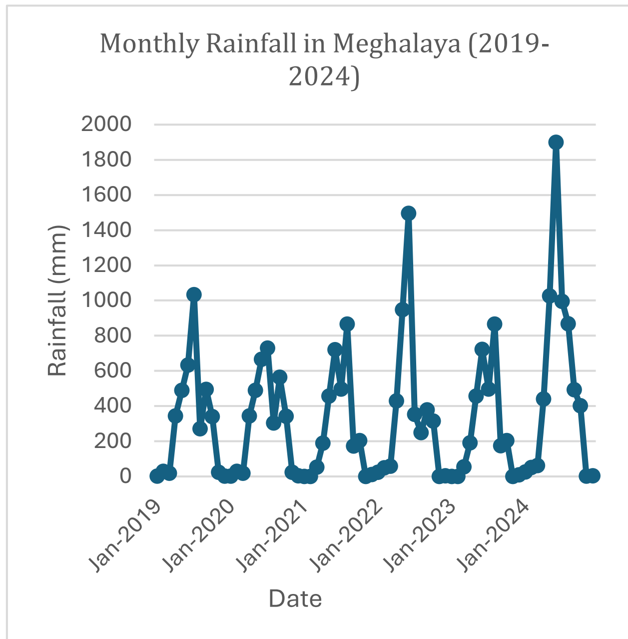

Figure 7: Monthly Rainfall in Meghalaya (2019–2024) (Source: BBS).

Figure 8 presents rainfall in Assam. Assam consistently receives heavy rainfall during June to August. This rainfall flows into rivers like the Brahmaputra and Surma, which can overflow in Sylhet and northern regions of Bangladesh.

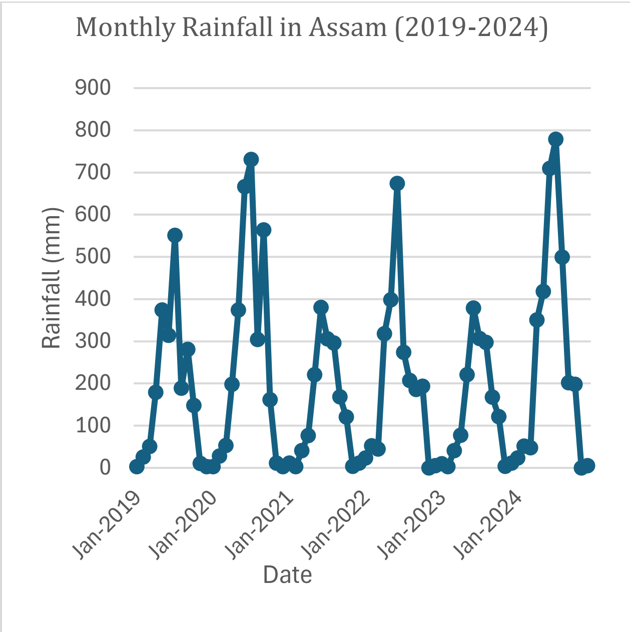

Figure 8: Monthly Rainfall in Assam (2019–2024) (Source: BBS).

Figure 9 highlights Tripura's rainfall pattern. Rainfall here strongly influences rivers like the Gomti and Titas, which affect districts like Cumilla and Feni. The graph shows high rainfall in 2024, supporting the reason why those districts experienced flooding that year.

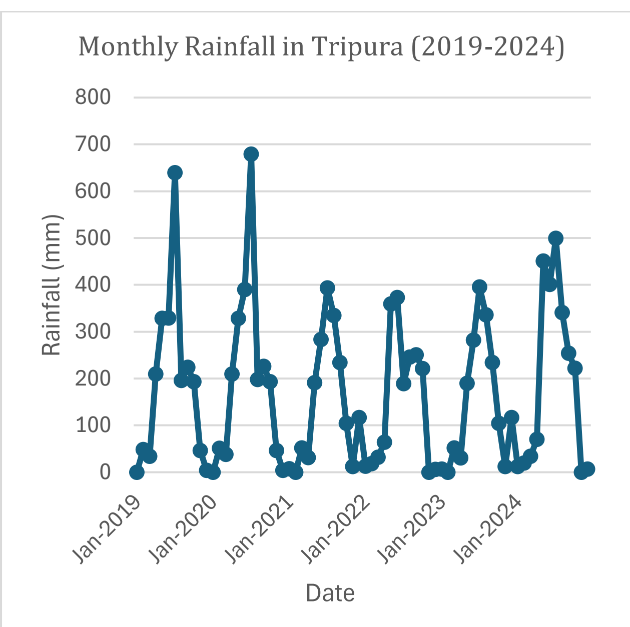

Figure 9: Monthly Rainfall in Tripura (2019–2024) (Source: BBS).

Figure 10 shows rainfall data from Sikkim. This state influences the Teesta River, which affects Rangpur. The graph shows consistent heavy rainfall during monsoons, suggesting that even if Rangpur's local rainfall is stable, heavy upstream rainfall can cause floods by raising river levels.

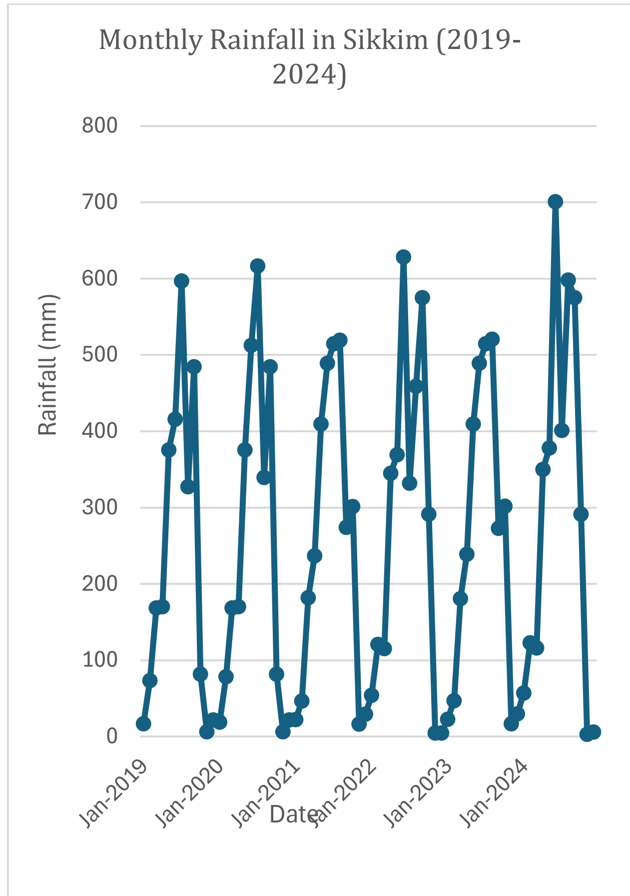

Figure 10: Monthly Rainfall in Sikkim (2019–2024) (Source: BBS).

Yearly maximum river water level data was gathered from BBS publications and secondary verified summaries (such as BWDB references in public sources). These were used to assess river overflow potential and its correlation with flood events.

Figure 11 shows the yearly highest water level in the Surma-Meghna river system. These rivers pass through Sylhet and parts of Cumilla. The water levels peaked in 2022 and 2024, the same years when those areas experienced severe flooding.

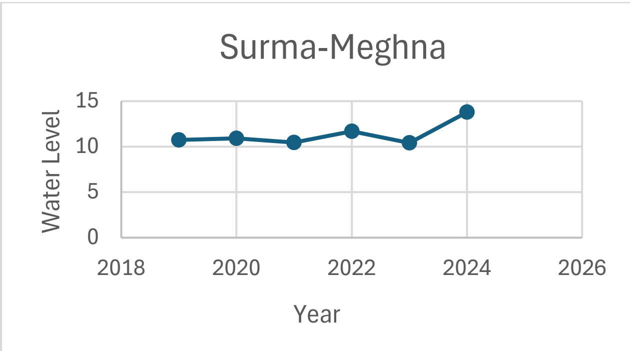

Figure 11: Yearly Highest Water Level in Surma-Meghna (2019–2024) (Source: BBS).

Figure 12 shows water levels in the Gumti-Buri Nadi system. These rivers affect Cumilla and Feni. The highest level was recorded in 2024, directly matching the flood year in these districts.

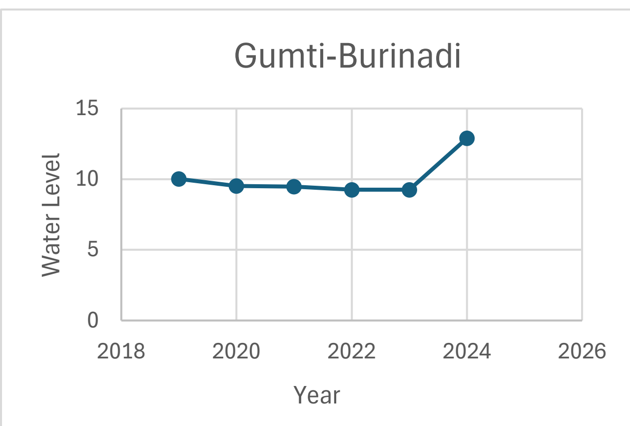

Figure 12: Yearly Highest Water Level in Gumti-Burinadi (2019–2024) (Source: BBS).

Figure 13 presents the yearly maximum level in the Teesta River, which flows through Rangpur. The graph shows consistently high levels from 2020 onward, especially in 2024, when Rangpur experienced notable flooding.

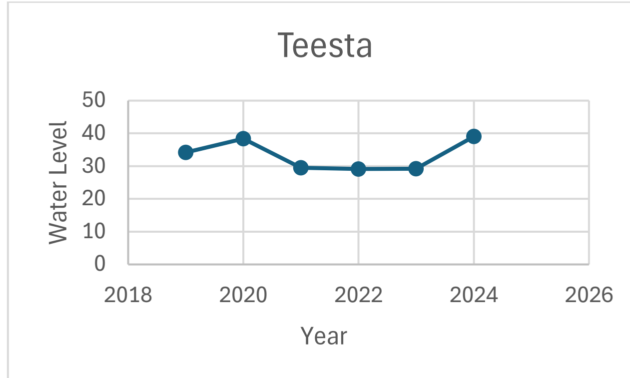

Figure 13: Yearly Highest Water Level in Teesta (2019–2024) (Source: BBS).

Figure 14 displays the maximum water level in the Muhuri River, which runs through Feni. In 2024, the water level reached its peak, even though local rainfall remained stable, proving that river swelling from upstream rainfall caused the flood.

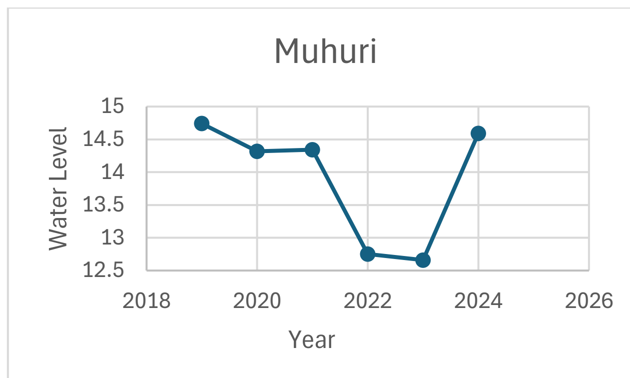

Figure 14: Yearly Highest Water Level in Muhuri (2019–2024) (Source: BBS).

Yearly flood-affected area (in sq. km) was collected from the Wikipedia page "Floods in Bangladesh", which compiles official government reports, FFWC press releases, and national media. The data includes the total area affected by floods for each year from 2019 to 2023.

Figure 15 shows the total flood-affected area in square kilometers from 2019 to 2023. This data was compiled from verified reports, including government sources and the “Floods in Bangladesh” page on Wikipedia. The largest affected areas were recorded in 2020, 2022, and

2024. These years also showed the highest water levels and rainfall in both local and upstream regions, confirming the relationship between environmental variables and flooding.

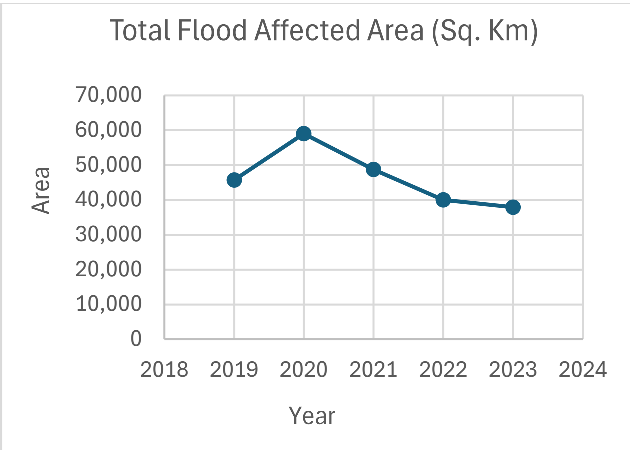

Figure 15: Yearly Total Flood Affected Area (2019–2023) (Source: BBS).

Result and Discussion

Analysis of rainfall and flood data from 2019 to 2024 across Sylhet, Cumilla, Feni, and Rangpur reveals that flood events predominantly occur during the monsoon months of June, July, and August. These months consistently saw the highest average rainfall locally and in upstream Indian states Assam, Meghalaya, Tripura, and Sikkim highlighting a strong seasonal and cross-border pattern.

In Sylhet, floods in 2022 and 2024 coincided with significantly high monsoon rainfall both locally and in upstream Assam and Meghalaya, which feed the Surma and Kushiyara rivers. Similarly, Cumilla faced major flooding in 2024, aligning with above-average rainfall locally and in neighboring Tripura, influencing rivers like Gomti and Titas.

Feni experienced flooding in 2024 despite stable local rainfall, due to record-breaking upstream rainfall in Assam and Meghalaya. The Muhuri River, which runs through Feni, reached peak water levels that year, indicating strong upstream influence.

In Rangpur, flood events occurred almost annually, despite consistent local rainfall. These were linked to persistent heavy rainfall in Sikkim, which feeds the Teesta River. High monsoon river levels confirmed the role of upstream inflow.

Conditional Probability Analysis

To evaluate how rainfall and river water levels influence the probability of flooding in Bangladesh, conditional probability analysis was performed using the formula:

$$\mathbf{P(A / B) = \frac{P(A \cap B)}{P(B)}}$$

Where:

A = Event of interest (Flood occurs)

B = Given condition (High rainfall in India and Bangladesh, and high river water level)

A ∩ B = Both flood and the given condition occur simultaneously

Thresholds for "high" values were established using the **Mean + Standard Deviation** method based on data from 2019 to 2024.

The thresholds for classifying high rainfall and river water level are:

Parameter	Mean	Standard Deviation	High Threshold (Mean + SD)
Rainfall in BD (mm)	237.23	42.94	> 280.17 mm
Rainfall in India (mm)	239.95	45.14	> 285.09 mm
Max Water Level (m)	17.14	1.82	> 18.96 m

Any value above these thresholds is considered "high" in this analysis.

Flood Occurred (1) or not (0) Table (2019–2024):

Year	Flood Occurred (1/0)	High Rainfall in BD	High Rainfall in India	High Max Water Level
2019	1	0	0	0
2020	1	0	1	1
2021	1	0	0	0
2022	1	1	1	1
2023	1	0	0	0
2024	1	1	1	1

Conditional Probability Results:

$$P\left(\frac{Flood}{High\ Rainfall\ in\ India}\right) = \frac{\frac{3}{6}}{\frac{3}{6}} = 1$$

$$P\left(\frac{Flood}{High\ Rainfall\ in\ BD}\right) = \frac{\frac{2}{6}}{\frac{2}{6}} = 1$$

$$P\left(\frac{Flood}{High\ Water\ level}\right) = \frac{\frac{3}{6}}{\frac{3}{6}} = 1$$

The results reveal a strong relationship between high rainfall or elevated river water levels and flooding in Bangladesh. Specifically, whenever there was high rainfall in either India or Bangladesh, or when river water levels rose significantly, floods occurred without exception. The conditional probability under each of these individual conditions is 1.0, or 100%, indicating a perfect correlation. This finding suggests that each factor alone is a highly reliable predictor of flood events in Bangladesh, and their combined occurrence further reinforces the likelihood of flooding.

Correlation Analysis:

We used Correlation coefficient to measure the strength of linear relationships between variables. Correlation matrix (Figure:16) indicating relationships among rainfall, river water levels, and flood affected area:

	Bd average rainfall	*India average rainfall*	*Total flood area*	*Max water level*
Bd average rainfall	1	0.999	0.845	0.770
India average rainfall	0.999	1	0.834	0.766
Total flood area	0.844	0.834	1	0.937
Max water level	0.770	0.766	0.937	1

Figure 16: Correlation Heatmap of Flood-Related Variables.

The correlation heatmap (Figure:16) visually represents these relationships, with warmer (green) colors indicating stronger positive correlations and cooler (light green) colors reflecting weaker ones.

A strong positive correlation was found between maximum water level and total flood-affected area (0.937), suggesting that rising river levels are closely associated with an increase in flooded regions. India's average rainfall exhibited a moderate to strong correlation with both flood area (0.787) and maximum water level (0.766), indicating a significant upstream impact on flooding in Bangladesh.

In contrast, Bangladesh's average rainfall showed a weaker correlation with total flood area (0.590), implying that local rainfall alone does not account for the full extent of flooding. However, it still showed moderate correlation with maximum water level (0.770), suggesting some influence.

These results indicate that maximum water level remains the most critical indicator for flood extent, while upstream rainfall in India plays a more influential role than local rainfall in Bangladesh.

To test the significance of each correlation coefficient, the following **t-test statistic** was used:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Correlation Between	r	t	Decision
Bd avg rainfall and India avg rainfall	0.9	4.13	Significant
Bd avg rainfall and Total flood aff. area	0.845	3.16	Significant
Bd avg rainfall and Max water level	0.77	2.81	Significant
India avg rainfall and Total flood	0.834	3.03	Significant
India avg rainfall and Max water level	0.766	2.78	Significant
Total flood and Max water level	0.937	5.39	Significant

The correlation between India's average rainfall and both total flood area and maximum water level is strong and statistically significant, highlighting India's influential role in Bangladesh's flooding. The maximum river water level also shows a high and significant correlation with flood-affected area, confirming its predictive potential. On the other hand, the relationship between Bangladesh's rainfall and Max river water level, India avg rainfall and Max water level were not strongly statistically significant due to the limited sample size.

Regression Analysis:

To understand how different environmental factors influence flooding in Bangladesh, we performed simple linear regression using three key independent variables:

- Average rainfall in Bangladesh
- Average rainfall in India
- Maximum river water level

The dependent variable is the total flood-affected area (in sq. km) for each year.

1.Regression Based on Bangladesh's Rainfall

Figure 17 shows the linear regression line between Bangladesh's average monthly rainfall (x-axis) and the total flood-affected area (y-axis). The upward slope of the line indicates a positive relationship: as rainfall increases in Bangladesh, the flood-affected area tends to increase. However, the data points are somewhat scattered, suggesting that rainfall alone in Bangladesh does not fully explain the flood extent.

Bangladesh rainfall contributes to flood risk, but not as strongly as other factors.Significance value ($t = 3.16$) indicates a statistically significant relationship, but less strong compared to river level.

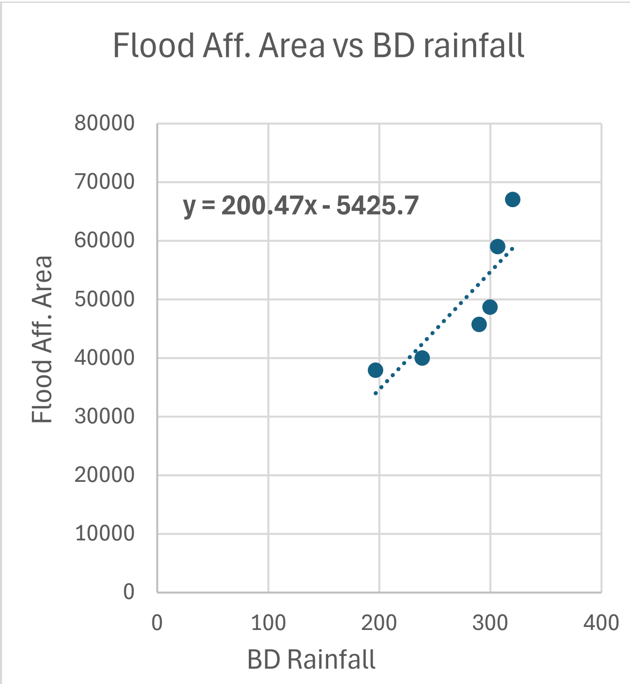

Figure 17: Regression line illustrating predicted flood area based on BD rainfall.

2. Regression Based on Indian Rainfall

Figure 18 shows the regression between average rainfall in Indian states (Assam, Meghalaya, Tripura, Sikkim) and Bangladesh's flood-affected area. Here, the regression line has a stronger upward slope and data points are more tightly grouped around the line than in the previous figure.

Rainfall in upstream Indian states has a stronger influence on flooding in Bangladesh than local rainfall.Significance value (t = 3.03) is statistically significant, confirming that cross-border rainfall is a reliable flood predictor.

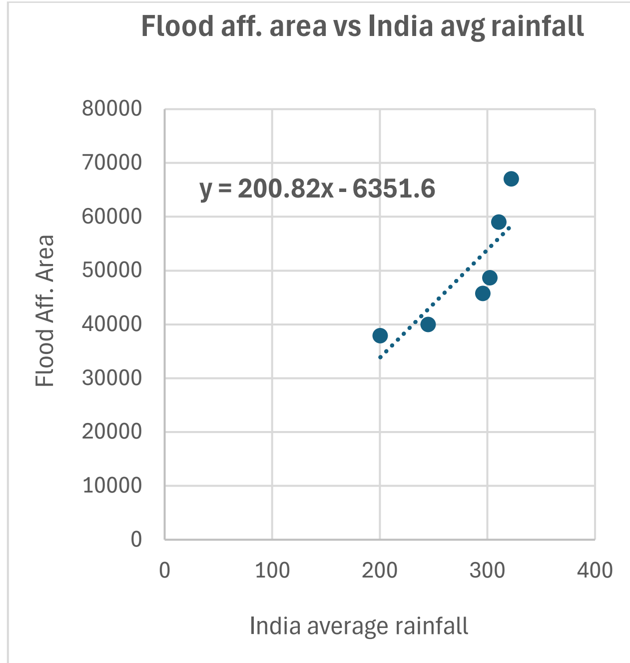

Figure 18: Regression line illustrating predicted flood area based on Indian rainfall.

3. Regression Based on Maximum River Water Level

Figure 19 shows the strongest regression relationship between maximum river water level and flood-affected area. The line is steep and most data points are close to it, showing a very strong correlation.

When river levels rise (especially beyond danger thresholds), the affected flood area increases significantly.Significance value (t = 5.39) is the highest among the three variables, making river level the most reliable predictor.

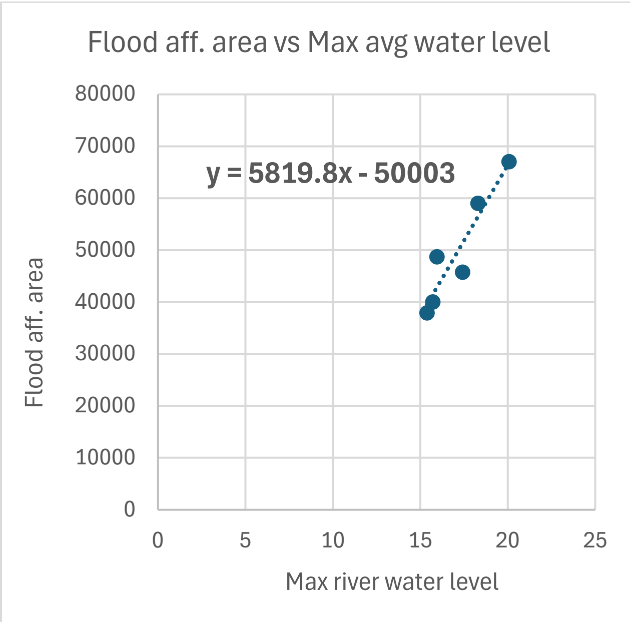

Figure 19: Regression line illustrating predicted flood area based on Max avg water level.

Significance test of regression:

Independent Variable	t	Significance
BD Average Rainfall	3.16	Significant
India Average Rainfall	3.03	Significant
Max River Water Level	5.39	Significant

The regression analysis performed on the dataset (2019–2024) reveals statistically significant relationships between the total flood affected area and three independent variables: Bangladesh average rainfall, India average rainfall, and maximum water level. The highest statistical significance was observed for maximum river water level (t = 5.39), indicating it as the most influential predictor of flood extent. Bangladesh average rainfall (t = 3.16) and India average rainfall (t = 3.03) also showed statistically significant effects at the 5% level, confirming that regional and upstream precipitation substantially contribute to flooding in Bangladesh.

These findings support the hydrological understanding that both local rainfall and cross border hydrometeorological conditions (such as rainfall in India) can impact river discharge and water accumulation, eventually leading to flooding. The strong t-values across all variables suggest that, despite the relatively small sample size, the model reliably captures the underlying physical relationships. Notably, maximum river water level, which directly reflects the capacity and overflow status of rivers, appears to be the most direct and sensitive indicator of flood-affected area in the observed years. This emphasizes the importance of continuous river monitoring and upstream rainfall tracking in flood risk assessment and early warning systems in Bangladesh.

Overall, flood risks in Bangladesh are driven by both local and upstream rainfall and river water levels. The strong correlation across districts and borders emphasizes the need for integrated transboundary monitoring to improve flood forecasting and preparedness.

Discussion

The analysis conducted in this study reveals that different environmental factors have varying levels of impact on flooding in Bangladesh. The regression between average rainfall in Bangladesh and flood-affected area shows a positive relationship, but the scatter of data points suggests that local rainfall alone is not a strong predictor of flood severity. In contrast, average rainfall in Indian states such as Assam, Meghalaya, Tripura, and Sikkim demonstrates a stronger and more consistent correlation with the flood-affected areas in Bangladesh, especially in border districts like Sylhet and Rangpur. The most influential variable identified was the maximum river water level. The regression line for river level closely fits the data, and the high t-value indicates a strong predictive power. This implies that when river levels cross a certain threshold, the likelihood and extent of flooding increase significantly. These findings support the importance of river monitoring and regional rainfall tracking in developing reliable flood forecasting and early warning systems.

Conclusion

In conclusion, this study found that maximum river water level is the strongest indicator of flood extent in the selected districts of Bangladesh, followed by rainfall in Indian upstream states. Although local rainfall contributes to flood conditions, it plays a comparatively smaller role. The analysis shows that most floods occur during the monsoon season, particularly from June to August, when river levels and rainfall are highest. Using simple statistical methods such as correlation, conditional probability, and linear regression, the study was able to identify key patterns in flood behavior. These methods proved to be effective tools for understanding flood risks and can assist in building basic but useful early warning systems. Overall, the findings highlight the value of using statistical data analysis for local-level flood risk assessment and planning.

Limitations

While this study offers meaningful insights, it is important to recognize its limitations. One key limitation is the small sample size, as only six years of data (2019–2024) were analyzed. This may not be enough to capture rare or long-term flood trends. Additionally, some data, particularly for flood-affected areas, were taken from secondary sources such as Wikipedia, which may not have the same level of accuracy or verification as official records. The study also used basic statistical techniques and did not explore more complex or advanced models that could provide deeper insights. Furthermore, the research focused on only four districts, which may not fully represent all flood-prone areas of Bangladesh. Lastly, the study relied entirely on historical data and did not include real-time data analysis or forecasting, which limits its use for live early warning systems.

Future Improvements

Future studies can be improved in several ways to enhance the accuracy and usefulness of flood prediction models. First, a larger and more comprehensive dataset covering 10 to 20 years should be used to provide a stronger statistical foundation. Second, the study area should be expanded to include more districts from different regions of Bangladesh to reflect the national flood scenario more accurately. Third, all data should be collected from verified and official sources such as BWDB and FFWC to ensure accuracy and credibility. Fourth, the use of advanced statistical methods and machine learning models such as multivariate regression, decision trees, and time-series forecasting could offer more precise predictions. Finally, integrating real-time river level and rainfall data using APIs or sensor networks could allow for the development of dynamic flood early warning systems. Partnering with institutions and government agencies would also help in accessing better data and applying professional expertise.

Reference

[1] Bangladesh Bureau of Statistics, Statistical Yearbook of Bangladesh, Dhaka, Bangladesh: Government of the People's Republic of Bangladesh.

[2] Bangladesh Bureau of Statistics, Monthly Statistical Bulletin of Bangladesh, Dhaka, Bangladesh: Government of the People's Republic of Bangladesh.

[3] Indian Meteorological Department, Rainfall Statistics of India, Ministry of Earth Sciences. [Online]. Available: https://mausam.imd.gov.in/ (accessed Apr. 16, 2025).

[4] Wikipedia, "Floods in Bangladesh." [Online]. Available: https://en.wikipedia.org/wiki/Floods_in_Bangladesh (accessed Apr. 16, 2025).

[5] R. Ahmed, N. Sultana, and M. A. Islam, "A statistical approach for flood forecasting using rainfall-runoff and upstream flow data in

Bangladesh," Journal of Hydrologic Engineering, vol. 25, no. 3, p. 04020003, 2020.

[6] M. A. Chowdhury and M. F. Karim, "Understanding regional rainfall and flood linkages in northeastern Bangladesh," Environmental Earth Sciences, vol. 75, no. 18, pp. 1–10, 2016.

[7] F. Hossain, C. K. Shum, and S. Biancamaria, "Forecasting floods in Bangladesh using satellite remote sensing and numerical weather models," Remote Sensing of Environment, vol. 131, pp. 48–59, 2013.

[8] M. T. Islam and S. Akter, "Monsoon rainfall pattern and its relationship with flood hazards in Sylhet region of Bangladesh," Water Resources Research, vol. 57, no. 7, p. e2020WR029142, 2021.

[9] M. S. Khan, M. H. Rahman, and T. K. Das, "Impact of transboundary river discharge and upstream rainfall on flooding in southeastern Bangladesh," Journal of Flood Risk Management, vol. 15, no. 1, p. e12753, 2022.

[10] M. M. Rahman, M. J. Kabir, and R. Khatun, "Influence of Indian upstream rainfall on floods in northern Bangladesh," Journal of Climatology & Weather Forecasting, vol. 5, no. 3, p. 1000210, 2017.