Pre-Thesis I

Inspiring Excellence

Implementing Machine Learning Techniques in order to Forecast Floods in Bangladesh based on Historical Data

Supervisor

Dewan Ziaul Karim

Lecturer, Department of Computer Science and Engineering, BRAC University

Authors (Group ID: T2230033)

Sadiq Uddin Bhuiyan ~ 19201018

Arman Zaman ~ 19201005

S. M. Toufique ~ 19201141

Ahmed Lateef ~ 19241016

Jubaer Islam ~ 19341002

Department of Computer Science and Engineering

BRAC University

Abstract

Flooding is a complex phenomenon that, due to its nonlinear and dynamic character, is difficult to anticipate. As a result, the prediction of floods has emerged as a critical area of study in the field of hydrology. Numerous researchers have handled this topic in various ways, spanning from physical models to image processing, however the time steps and precision are insufficient for all applications. This report looks at machine learning approaches for forecasting weather conditions and criteria and assessing the associated margins of uncertainty. The evaluated outputs enable more accurate and precise flood prediction for a variety of applications, including transportation systems.

Introduction

The South Asian population eagerly await the arrival of the monsoon season, every summer. Between June and October, the region experiences more than 70% of its annual rainfall. Rains that are unusually heavy nearly always signal calamity. Occasional droughts occur when rain falls too little or too late. When there is too much rain, vast swaths of land are washed away. This year has already been disastrous for sections of Bangladesh and India. Rivers break their banks due to unusually heavy rainfall in May and June. By June 22nd, 83% of Sylhet and 90% of Sunamganj, two districts in Bangladesh's north-east that are home to 6 million people, were entirely submerged. Authorities and humanitarian workers are desperately attempting to reach more than 9 million people throughout Bangladesh and the neighbouring Indian regions of Assam and Meghalaya. At least a hundred of them are believed to have perished, with roughly 30 of them dying in Bangladesh. In the following days, the death toll is virtually expected to rise. However, considering the intensity of the floods, it is less than what would have been predicted. Heavy rain and flash floods, for example, killed more than 180 people in Belgium and Germany in July 2021, countries far richer and less populated than Bangladesh. The number of deaths linked with such calamities has dropped considerably in Bangladesh. Cyclone Bhola killed an estimated 300,000 to 500,000 people in 1970. Cyclone Amphan, predicted to be the most severe cyclone to form in the Bay of Bengal in two decades, killed roughly 30 people in 2020. Flooding in Bangladesh during the latter week of June 2020 was caused by a protracted and intensified severe monsoon and upstream flooding, affecting 5.4 million people in the country's northern, central, and north-eastern regions. Around 37% of the country's entire territory was flooded, impacting 33 districts, and it was deemed the country's longest flooding event in the past 22 years. People in numerous regions saw repeated floods till the beginning of October 2020 as a result of monsoon rains and severe rainfall upstream. Housing, access to clean and safe water, hygiene and sanitation facilities, and access to livelihoods were all severely damaged in the majority of the impacted areas. According to the Bangladesh Ministry of Disaster Management and Relief (MoDMR) report, dated 2 August 2020, the lengthy floods impacted about 5,448,271 people in 33 districts, with 1,059,295 households marooned and 41 individuals killed. Furthermore, the Ministry of Agriculture (MoA) reported that 83,000 hectares of paddy fields, 125,549 hectares of agricultural land and USD 42 million in crops were damaged. The floods also caused moderate to severe damage to livestock and fisheries. According to the Department of Livestock Service (DLS), the industry lost USD 74.5 million in livestock, as well as 16,537 hectares of grassland. According to the Department of Public Health and Engineering (DPHE), 928,60 tube-wells and 100,223 latrines were destroyed. The Water Development Board's north zone office reported that rivers degraded 3,745 hectares of land in eight flood-affected districts in Rangpur division. The monsoon rains, along with prolonged flooding and the COVID-19 epidemic, exacerbated the population's situation.

How has Bangladesh mitigated the impact of harsh weather?

Floods regularly inflict severe damage to numerous infrastructure and socioeconomic system elements, resulting in large direct and indirect economic losses. River flow has complicated behaviour that is influenced by soil qualities, land use, temperature, river basin, snowfall, and other geophysical factors. It is essential to correctly estimate floods and create flood mapping as a consequence to plan for emergency responses. It is currently a popular study area in natural disaster prediction and risk management. Physical, statistical, and computational intelligence/machine learning algorithms are the most popular forms of prediction models.

Problem Statement

Floods all over the world have had a significant negative impact on the economic and social life. The impacts of the coronavirus epidemic that engulfed the world last year and wreaked havoc on the economy and health grew worse. In Vietnam, there have been 90 fatalities and at least 34 persons have gone missing. In Laos, 100 villages were flooded due to the submergence of at least 10,000 hectares of cropland. In Cambodia, there were 25 fatalities and 40.000 displaced people. According to disaster management authorities, flooding in Bangladesh has affected close to 1 million people in 13 districts since late June of last year. Numerous northern districts of the nation were flooded as a result of the country's severe rains and those in the river catchments of the neighbouring country of India. While another 7,31,958 people are still without access to safe drinking water, more than 3.3 million people have been evacuated from flood-affected areas. 41 of the 93 fatalities that have occurred since June 2020 have been children, with drowning accounting for the majority of these deaths.

One would anticipate that the recovery processes would be well-known given that the effects of natural disasters like floods are largely consistent from year to year. Each severe flood, however, brings back the same problems. Floods have a wide range of effects, including both social and economic ones. It's critical for water management, environmental challenges, and social safety to predict river water levels after significant rain. For these purposes, mathematical models based on statistical analysis or physical considerations have been created. The forecasts they offer are time-consuming and imperfect in both situations. Floods have been avoided and managed for thousands of years. But because floods are a natural occurrence, it is challenging to predict them solely using statistical techniques. Therefore, using machine learning tools will be a better strategy to predict floods and minimise their risks.

Research Objectives

Weather prediction in the past needed to be performed using human brain power. Calculators and then computers made those prediction calculations much easier. In the modern era, machine learning is the most advanced and accurate method of weather prediction. Our goal with this research is to create and review a predictive model for floods in Bangladesh. Being able to know beforehand which areas are more flood-prone under certain conditions can help mitigate the overall damage they do, will help at-risk areas prepare in anticipation of floods, and can be used to inform where relief aid is most needed. Our objectives for this research are:

- To understand and utilise different machine learning algorithms
- To develop a good model for flood prediction using machine learning
- To evaluate the model
- To recommend changes and improvements for the model

Detailed Literature Review

The methodologies and techniques used in flood prediction have been evaluated in a few studies, including one by Devia, Ganasri, and Dwarakish [1], who analysed hydrological models and explored the three different types of flood prediction models. They are the data-driven model, the conceptual model, and the physical principle-based model. Each model's benefits and shortcomings were discussed. The authors of this work, however, emphasised several types and uses of physically based models, such as SWAT and the SHE/MIKE SHE models, with a particular focus on physical principle-based models. Data-driven models, on the other hand, are not covered in detail. In their assessment and comparison of several hydrological and hydrodynamic model types, Teng, Jakeman, Vaze, Croke, Dutta, and Kim [6] further divided them into 1D model, 2D model, and 3D model categories. According to a recent survey of research articles, only one piece [2] expressly reviews data-driven models based on ML approaches. This paper compared the most recent ML techniques for flood prediction models, but there was no discussion of the significance of input parameters.

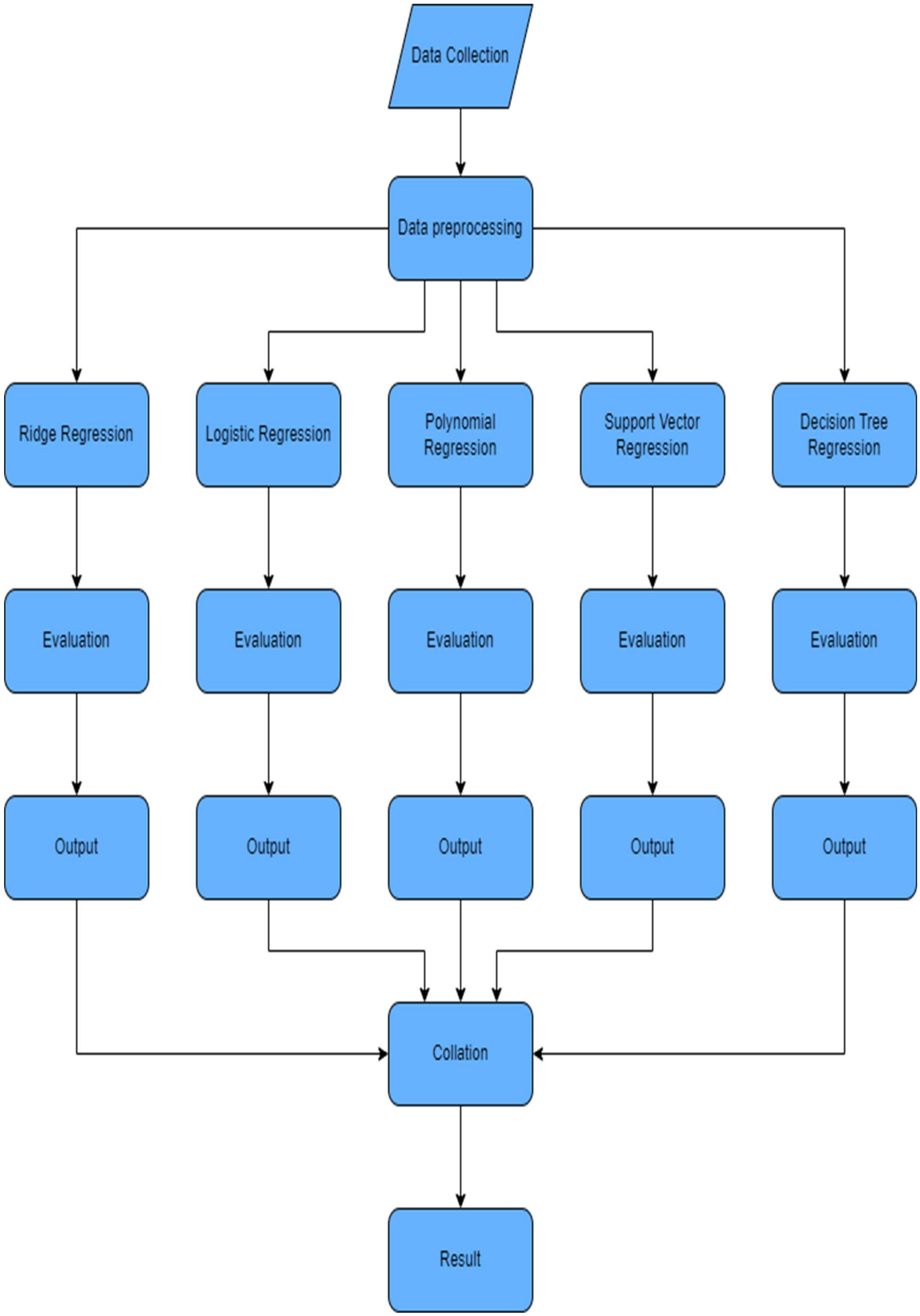
In order to increase the accuracy of flood predictions, Mosavi, Ozturk, and Chau [2] proposed four ways. The first strategy is hybridization, which is combining two or more machine learning (ML) techniques or combining ML techniques with physically based models. The second tactic is an algorithm ensemble, where the model compiles and executes a number of related algorithms before selecting the one that produces the most accurate results. The uncertainty of the forecast will be reduced by using the ensemble technique. A third technique to increase the quality of ML algorithms is algorithm optimization. For example, to improve ANN, employ BPN and FFMLP for model optimization. And finally, data decomposition is an additional method for raising prediction accuracy. The accuracy of the predictions will also increase as the dataset quality does. Because they frequently outperform individual models, model optimization and hybridization inside ML approaches have thus gained significant popularity [3]. Contrarily, hybridization within the ML approach, ensemble, and model optimization continue to be more popular than hybridization with physically based models. The expense of computation and model complexity may be a factor in the decline in popularity.

In order to estimate flash flood property damage in the Southeast US (SEUS), this study proposed by Atieh Alipour et al., presented a risk-based and physically informed model. It did this by using a range of significant elements, including geographic, socioeconomic, and meteorological aspects [4]. RF was chosen as the main model. Using data from several data sources for a significant number of flash flood incidents between 1996 and 2017, the model was trained and evaluated. Regression and classification are the two distinct ways that RF has been used. The quantity of property damage was projected in the regression mode after we evaluated whether or not the flash flood caused any property damage in the classification mode. The performance of both classifier and regression models was assessed using a variety of statistical metrics, and the findings showed that the established framework was reliable. The results of this study point to the applicability and precision of the RF model for estimating property damage caused by flash flood occurrences across a wide area. Researchers are urged to create probabilistic models for projecting flash flood damage in the future. Additionally, the model might include other predictors such watershed features.

The objective of the work proposed by Yanlai Zhou et al., is to combine the XAJ and MCQRNN models to improve the precision and dependability of short-term FPD forecasts. This study aims to investigate and assess the efficacy of the XAJ model paired with the

MCQRNN on short-term (twelve hours) flood forecasts at various horizons in order to increase the predictability and reliability of FPD forecasts [1]. Real-world applications drive the need for machine learning and conceptual model hybridization in order to improve the precision and dependability of short-term FPD forecasting. In order to produce short-term FPD forecasts, this study suggested a hybrid rainfall-runoff model (i.e., XAJ-MCQRNN) that combined the Xinanjiang (XAJ) conceptual model and the MCQRNN. Its aptitude for effective learning and precise predicting is examined and confirmed.

Working Plan



Conclusion

In some years, due to heavy rainfall and upstream water movement, floods can be devastating. Examples of the intensity of massive flooding in Bangladesh in terms of duration and damage are in 1988, 1998, 2004, 2007, and 2017. One of these natural disasters has also been identified as the 2020 floods because a significant number of people will experience its effects over an extended period of time. The most crucial thing to learn is how to lessen flood-induced vulnerability and better manage floods while causing the minimum amount of damage to people's lives and property. In order to do that, it's critical to be aware of impending flooding in advance so that the appropriate safety measures can be taken. To predict the occurrence of a flood, it is necessary to keep a number of factors in mind. Rainfall, river basin water levels, the geography of the area, etc. are a few examples. By combining one or more of the aforementioned factors and using various methods, such as Artificial Neural Networks (ANN), Deep Neural Networks (DNN), etc., current research has produced very high accuracy rates for predicting these floods.

The climate is a variable that changes over time, and as a result, so do other variables that depend on it. For instance, Bangladesh's geography and industry have undergone significant change over the past 40 years, both of which have an impact on climate change. As a result, both the amount of rainfall and the water level have changed over time. Bangladesh's population has grown over the years along with the number of housing units as well. Additionally, Bangladesh has a very high population density, which results in a large number of building structures. The country's drainage system is evolving as a result of the country's shifting geographic makeup over time. Due to the rainwater collection and discharge onto a nearby river, the water level is reliant on the drainage system. Therefore, every element that contributes to the occurrence of a flood is connected to every other element. By taking the time to gather these data for the respected location and determining the correlation between the factors, the aforementioned study was completed.

Due to the rainwater collection and discharge onto a nearby river, the water level is reliant on the drainage system. Therefore, every element that contributes to the occurrence of a flood is connected to every other element. By taking the time to gather these data for the respected location and determining the correlation between the factors, the aforementioned study was completed.

References

- [1] Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., & Xu, C.-Y. (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, *604*, 127255

- [2] Mosavi A, Ozturk P and Chau K-w 2018 Flood prediction using machine learning models: Literature review Water 10 1536

- [3] Al-Jarrah O Y, Yoo P D, Muhaidat S, Karagiannidis G K and Taha K 2015 Efficient machine learning for big data: A review Big Data Research 2 87-93

- [4] Alipour1, A. *et al.* (2020) *IOPscience*, *Environmental Research Letters*. IOP Publishing. (Accessed: January 13, 2023).

- [5] Devia G K, Ganasri B and Dwarakish G 2015 A review on hydrological models Aquatic Procedia 4 1001-7

- [6] Teng J, Jakeman A J, Vaze J, Croke B F, Dutta D and Kim S 2017 Flood inundation modelling: A review of methods, recent advances and uncertainty analysis Environmental Modelling & Software 90 201-16

- www.reliefweb.int

- www.economist.com

- www.practicalaction.org

- www.researchgate.net

- www.ifrc.org