Prediction of Flood Using Ensemble Machine Learning Methods in Bangladesh.

BY

GALIB MD. BIN HASAN SIAM
ID: 201-15-3391
AND

RIFA AKILA
ID: 201-15-3390

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Md Atik Asif Khan Akash
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project titled **"Prediction of Flood Using Ensemble Machine Learning Methods in Bangladesh"**, submitted by **GALIB MD. BIN HASAN SIAM** ID: 201-15-3391 and **RIFA AKILA** ID: 201-15-3390 to the Department of Computer Science and Engineering, Daffodil International University, has been accredited as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *January 21, 2024*.

BOARD OF EXAMINERS

Chairman

Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Raja Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mayen Uddin Mojumdar
Senior Lecturer
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Mohammed Nasir Uddin
Professor
Department of Computer Science & Engineering
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Shah Md Tanvir Siddiquee, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of the project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Siddiquee

Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Md Atik Asif Khan Akash
Lecturer
Department of CSE
Daffodil International University

Submitted by:

galib

GALIB MD. BIN HASAN SIAM
ID: 201-15-3391
Department of CSE
Daffodil International University

Rifa

RIFA AKILA
ID: 201-15-3390
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We are really grateful and wish our profound indebtedness to **Shah Md Tanvir Siddiquee, Assistant Professor, Department of CSE** Daffodil International University, Dhaka. Deep knowledge & keen interest of our supervisor in the field of “Machine Learning” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Shah Md Tanvir Siddiquee,** and Head**,** department of CSE, for his kind help to finish our project and to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Flooding occurs when a water body engulfs the mainland, disrupting the regular lives of the inhabitants. Bangladesh faces a persistent risk of flooding due to its geographical location and the impacts of climate change. Annual floods consistently disrupt the ordinary rhythm of life in the country, causing the impoverished residents to lose their homes, crops, and, tragically, loved ones. This ongoing threat renders these individuals more susceptible. In our study, we gathered a comprehensive set of climate data spanning 74 years, ranging from 1949 to 2022, from the Bangladesh Meteorological Department. This research aims to alleviate the destructive impacts of flooding by employing ensemble machine learning techniques. We have used two ensemble approach Bagging and Stacking and utilize two models for each approach. Different combination of six algorithms namely Decision Tree, Random Forest, Xtreme Gradient Boosting, AdaBoost, Support Vector Machine and Logistic Regression are used to develop those four models. All our four model demonstrate robust performance for predicting flood in different regions of Bangladesh. Our highest accuracy is obtained 97.22% with the Bagging approach model where we have used Decision Tree, Random Forest, Xtreme Gradient Boosting as base classifier. The precision, recall, F1-score, and ROC-AUC for this model were respectively 0.92, 0.91, 0.92 and 0.95. We have also evaluated Matthews Correlation Coefficient (MCC) and Brier Score. Those are respectively 0.9 and 0.028. These results signify the potential of our model to play a significant role in flood forecasting, showcasing its effectiveness in predicting and mitigating the impact of floods.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Bangladesh faces annual flooding due to the convergence of the Padma, Brahmaputra, and Meghna rivers, profoundly affecting residents' daily lives and economic well-being. Addressing the repercussions of floods is both a scientific and moral imperative. The primary goal is to employ machine learning for discriminating flood prediction in different regions of Bangladesh. By integrating diverse datasets covering geography, weather, and 74-year flood history, the study uncovers hidden patterns, enhancing understanding of cyclical flood nature and identifying trends.

The research's fundamental tenet lies in integrating extensive datasets to comprehend the causes of floods in Bangladesh. The complex riverine systems and low-lying regions, influenced by seasonal changes and climatic anomalies, regulate flood intensity. Seven and a half decades of historical flood data offer valuable insights into long-term trends, acting as a temporal lens. The machine learning model serves as a tactical tool for predicting and preventing floods, functioning as an early warning system. This proactive approach not only reduces immediate losses but also strengthens the social fabric for ongoing flood threats.

To achieve this objective, the study incorporates Ensemble approaches such as Bagging and Stacking. This study has created a total of four Ensemble models with those two approaches. Six algorithms such as Decision Tree, Random Forest, Xtreme Gradient Boosting, AdaBoost, Support Vector Machine and Logistic Regression are utilized to develop those 4 models. These techniques enhance model resilience and prediction accuracy by leveraging their collective intelligence. The prediction model, bolstered by 74 years of climate data and advanced techniques, aims to contribute significantly to the discourse on catastrophe preparedness and mitigation. Subsequent sections delve into methodological foundations, data synthesis, and the anticipated transformative impact.

1.2 Motivation

Bangladesh routinely faces devastating floods, impacting the nation's socioeconomic fabric with significant material losses, and disrupting daily life. As computer science students, we aim to leverage our skills to address the challenges posed by floods in our region. Our objective is to create a reliable ensemble model that can recognize floods in Bangladesh.

Motivated by the humanitarian emergency caused by frequent floods, our academic background in computer science equips us to make a substantial contribution to mitigating the complex issues arising from natural catastrophes. By integrating diverse datasets related to climate and past flood events, our proposed model seeks to identify patterns and connections not easily discernible through traditional analytical methods. This predictive capacity can enhance disaster preparedness, enabling communities to strengthen defenses against impending floods and mitigate harm to vulnerable populations.

The moral imperative driving our research is the potential to improve the lives of those affected by floods, offering communities greater resilience and preparedness. The technology-based solution proposed in this study aims to break the destructive cycle of floods, guiding communities toward adaptive readiness and reducing overall vulnerability.

Our motivation also stems from a commitment to social responsibility and sustainable growth. Floods contribute to a cyclical pattern of poverty, hindering long-term prospects for affected areas. By proactively using machine learning to anticipate floods, this research project seeks a paradigm shift where technology innovation plays a pivotal role in developing resilient and sustainable communities. The potential for computational methods to bring about constructive social change serves as a beacon of hope, guiding the study toward a beneficent and meaningful conclusion.

1.3 Research Objectives

This research is guided by objectives focusing on flooding dynamics in Bangladesh:

1. **Synthesis of Datasets:**
   Meticulously combine diverse datasets (geography, meteorology, 74-year flood history) for a holistic understanding of flooding factors.
2. **Develop Predictive Model:**
   Utilize Ensemble Machine Learning methods to create a sophisticated model, optimizing diverse algorithms for enhanced predictive accuracy in distinct Bangladesh regions.
3. **Evaluate and Validate Model:**
   Rigorously test the model against historical floods and diverse climatic scenarios, ensuring reliability and generalizability.
4. **Ethical Considerations and Impact Assessment:**
   Address ethical concerns and assess societal impact, ensuring alignment with principles of equity, accessibility, and community empowerment in disaster prediction.

1.4 Problem Statement

The research contextualizes challenges posed by recurrent floods in Bangladesh through clear and concise problem statements:

1. **Inadequate Predictive Capabilities:**
   Bangladesh faces challenges due to limited predictive capabilities, hindering effective flood anticipation. Traditional approaches fall short in discerning complex patterns. This research uses machine learning to enhance accuracy, improving disaster preparedness.
2. **Limited Temporal Perspective:**
   Current methodologies lack a comprehensive temporal view, impeding the understanding of long-term climate patterns. Incorporating 74 years of data, this

research aims to bridge temporal gaps, providing a nuanced exploration of flooding's cyclical nature in Bangladesh.

3. **Insufficient Integration of Ensemble Methods:**

   Existing flood prediction models lack sufficient integration of Ensemble Machine Learning methods, compromising predictive efficacy. This research addresses this gap by systematically incorporating diverse algorithms, enhancing the model's robustness.

4. **Ethical and Social Considerations:**

   Ethical implications of using predictive models in disaster preparedness require scrutiny. This research emphasizes equitable resource distribution, information accessibility, and community empowerment. It aims to ensure a positive, ethical, and socially responsible societal impact.

1.5 Research Questions

How does synthesizing diverse datasets contribute to understanding factors influencing flooding in Bangladesh?

Can Ensemble Machine Learning methods enhance the predictive accuracy of flood forecasting in Bangladesh?

How does the developed model perform in forecasting floods across different climatic scenarios and historical occurrences?

What ethical considerations are associated with deploying machine learning for disaster preparedness, and how can they be addressed in flood-prone regions?

In what ways does integrating 74 years of climate data improve the understanding of flooding patterns in Bangladesh?

To what extent does the predictive model contribute to sustainable development and reduce vulnerability to flooding in Bangladesh?

1.6 Report Layout

Chapter 1: This chapter aims to give a proper understanding of the introduction, motivation, and objectives of this study. This chapter mentions the research question and problem statement.

Chapter 2: A thorough exploration of the background study establishes the contextual framework for the research.

Chapter 3: Provides a detailed account of the methodology employed in the study.

Chapter 4: An in-depth discussion and comparison of the results derived from the proposed model are presented.

Chapter 5: Delves into the examination of the social, environmental, and ethical aspects associated with the study.

Chapter 6: Addresses the study's limitations, outlines potential avenues for future research, and encapsulates the conclusions drawn from the overall study.

CHAPTER 2

Background

2.1 Related Works

Khabat, et al.[8] employs MCDM techniques (VIKOR, TOPSIS, and SAW) and machine learning methods (NBT and NB) for flood susceptibility modeling in China's Ningdu Catchment. Twelve flood conditioning factors, including NDVI, lithology, land use, and others, are used. Variance Inflation Factors, Information Gain Ratio, and Multicollinearity diagnosis tests are employed to identify effective factors and assess their correlation. Evaluation metrics such as ACC, Kappa, RMSE, MAE, and AUC indicate strong predictive capability for all models, with the NBT model consistently performing well (AUC = 0.98). The study concludes that the NBT model is a promising tool for assessing flood hazards and facilitating proper planning in flood-prone areas.

Miah Mohammad Asif, et al.[9] employs Binary Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), and Decision Tree Classifier to accurately predict floods in Bangladesh. The dataset includes monthly rainfall index and yearly flood occurrence data from 34 stations, covering the years 1980-2020. Preprocessing involves organizing, formatting, feature encoding, and partitioning the data (80:20 ratio for training and testing). Classification is done on 16 columns, including station, year, 12 months, and flood index. The SVC model exhibits the highest accuracy (0.8409) and precision (0.7647), outperforming Binary Logistic Regression. Additionally, the SVC model achieves accuracy (0.8088), precision (0.4667), and recall (0.5833). The study suggests that SVC is a robust model for flood prediction in Bangladesh.

Abu Reza Md Towfiqul, et al.[10] utilized advanced ensemble machine learning models (ANN, SVM, RF, RS, and Dagging) for developing flood susceptibility models. Evaluation of flood conditioning factors includes information gain ratio and multi-collinearity tests. The Topographic Wetness Index (TWI) is utilized to represent wetness variation in the basin. Rainfall data from four meteorological stations in Bangladesh are obtained and interpolated using kriging. Flood susceptibility models are reclassified using the quantile reclassifying approach. Hybrid ensemble machine learning algorithms surpass conventional models, showing enhanced performance in flood susceptibility prediction. Validation and comparison of models using statistical measures (Freidman test, Wilcoxon signed-rank test, t-paired test) reveal good predictive ability with an AUC above 0.80 for all models.

Bahram, et al.[11] examined various methods, including frequency ratio, analytical hierarchy process, bivariate statistics, and logistic regression, for flood susceptibility assessment. Machine learning models CART, MDA, and SVM are utilized to create a flood susceptibility map. The ensemble modeling approach combines predictions using weighted averaging based on AUC statistics, addressing limitations of individual models. The study, covering a watershed area of about 126 km^2 with 51 flooded points, incorporates categorical and numerical variables like land use and slope. The ensemble model surpasses individual models, demonstrating the effectiveness of combining different techniques for flood susceptibility mapping. Researchers partitioned training and testing datasets, repeating the process until achieving a desired AUC value of at least 80.

Shafapour, et al.[12] introduces an innovative flood susceptibility assessment method combining Support Vector Machine (SVM) and Frequency Ratio (FR) to enhance accuracy in the upper Kelantan basin, Malaysia. The FR model conducts bivariate statistical analysis, and its weights are integrated into SVM analysis for comprehensive flood susceptibility evaluation. Using a flood inventory map with 155 locations, the study area is divided into training (70%) and validation (30%) datasets. Diverse variables, including elevation, curvature, geology, and others, contribute to flood susceptibility assessment. The ensemble FR and SVM method is compared rigorously with the Decision Tree (DT) model, showcasing superior success and prediction rates (88.71% and 85.21%, respectively). Flood susceptibility maps demonstrate spatial resemblance, affirming the method's efficiency through ROC curve analysis and indicating its applicability for flood susceptibility mapping.

Abu El-Magd et al.[13] introduces a hybrid approach using a naïve Bayes (NiB) machine learning algorithm and hydrologic indices for flash flood vulnerability assessment in coastal regions. Hydrologic indices (STI, SPI, TWI, SD, curvature, slope angle, and SA) serve as controlling factors. The study utilizes 189 locations from Wadi Ghoweiba and surrounding areas, processed in ArcGIS (10.5) and R computing environment. The hybrid machine learning (HML) model, combining NiB and hydrologic indices, achieves a superior accuracy of 90.8%, surpassing the NiB model (87.7%). The study identifies sediment accumulation linked to low topography and gentle slope, with stream networks contributing significantly to sediment erosion near high lands.

Motta et al.[14] done methodology adhered to the Cross Industry Standard Process for Data Mining (CRISP-DM) framework, covering phases such as business understanding, data understanding, data preparation, modeling, evaluation, and deployment. To address class imbalance, resampling techniques (over-sampling, under-sampling, and SMOTE ENN) were employed. Feature selection methods did not significantly improve model performance. Data preparation involved cleaning, treating missing values, and handling outliers using nearby weather stations, recent observations, and k-nearest neighbors. Data were sourced from internal reports by the Lisbon city council, the fire department, and the national weather authority, covering flood events between January 2013 and December 2018. The dataset includes geospatial coordinates, timestamps, event types, event severity, and weather data from three meteorological stations.

Isaac Kofi, et al. [15] employs four state-of-the-art machine learning (ML) algorithms (LSTM, XGBoost, RF, and Extra Trees) to implement distinct flood prediction models. Performance is evaluated using multiple statistical metrics. The dataset from the meteorological agency in Ghana contains rainfall details, with rainfall level as the main feature. Data preprocessing involves noise removal, handling inconsistency, and imputing missing values with averages. The dataset is normalized using the Min-Max function and partitioned into train (80%) and test (20%) sets, with the training set further split into 15% for validation. Experimental results showcase the potential of the flood prediction models—LSTM, XGBoost, RF, and Extra Trees—demonstrating promising performance in efficient and effective flood prediction.

Ganguly et al.[16] used three machine learning algorithms (linear regression, random forest, and artificial neural network) to predict flood-affected households in Bangladesh. Linear regression outperformed the other two, and its assumptions were considered. Predictors were rescaled using Z-score standardization, and feature selection employed stepwise selection with AIC. K-fold cross-validation addressed data limitations. Performance metrics (RMSE, MAE, Correlation Coefficient) and a Paired T-Test compared algorithms. Data sources included Bangladesh's latest disaster report, Statistical Yearbook 2015, District Statistics, and Bangladesh Water Development Board. The dataset had 29 predictors in five classes, including economic indicators. Linear regression's R2 value was 0.8, Adjusted R^2 was 0.74, and F-Test p-value was 2.589e-^{12}, indicating statistical significance.

Gauhar et al.[17] employed various correlation coefficients for feature selection and used the k-nearest neighbors (k-NN) algorithm for flood prediction. Z-score normalization was applied for data scaling. The k-NN algorithm utilized Euclidean distance (p=2) and varied k from 2 to 9, employing a uniform weight function. Weather data for 65 years from Kaggle and Bangladesh Meteorological Department, comprising 20544 instances and 32 districts, was used. Key attributes included Rainfall, Cloud Coverage, Relative Humidity, Minimum Temperature, and Wind Speed. The k-NN algorithm achieved a high testing accuracy of 94.91%, with an average precision of 92.00% and an average recall of 91.00%, indicating its effectiveness in predicting floods in Bangladesh.

Mahfuzur, et al. [18] This paper proposes a flood susceptibility assessment approach in Bangladesh using statistical, machine learning, and multi-criteria decision analysis methods (ANN, LR, FR, and AHP). Flood inventory data, remote sensing-derived causative factors, hydrological modeling, and secondary data are integrated for a flood hazard map. Model performance is evaluated using the area under the receiver operating curve (AUROC), with the LR-FR model having the highest predictive power (AUROC

88.10%). AHP determines weighting factors for factors contributing to flood occurrence, offering insights into their importance. The study has significant implications for disaster management in Bangladesh, a highly populous country prone to climate-induced adversities like flooding.

2.2 Scope of the Problem

Recent flood prediction studies often use various machine learning methods, with ensemble methods showing superior efficiency. However, reliable works, especially in Bangladesh, a flood-prone region, are limited. Existing studies lack a dataset spanning 74 years, crucial for comprehensive flood prediction models.

This research addresses these gaps by developing a sustainable flood prediction model with two ensemble methods: Bagging and Stacking with six different algorithms to mitigate Bangladesh's vulnerability to frequent flooding.

Overcoming identified limitations and incorporating a more extensive temporal dataset, this research aims to significantly contribute to flood prediction in Bangladesh. The following sections will have detailed discussion about the methodology, dataset, and comprehensive analysis of the proposed flood prediction model.

2.3 Challenges

Developing an ensemble model for anticipatory disaster preparedness in Bangladesh faces several challenges that must be addressed for successful implementation and impact.

1. Data Complexity and Integration:

Managing diverse datasets, including geographical, meteorological, and 74 years of historical flood data, poses challenges in handling various formats, scales, and temporal resolutions. Ensuring coherence and reliability demands advanced preprocessing and integration strategies.

2. Algorithmic Complexity and Ensemble Integration:

Incorporating Ensemble Machine Learning introduces algorithmic complexity. Balancing strengths and weaknesses of individual algorithms to enhance predictive accuracy requires sophisticated optimization within the ensemble framework, demanding thorough experimentation and validation.

3. Ethical Considerations and Societal Impact:

Deploying machine learning for disaster preparedness presents ethical challenges. Ensuring equitable resource distribution, information accessibility, and community empowerment demands a nuanced understanding of societal dynamics and a comprehensive ethical framework.

4. Temporal Depth and Long-Term Analysis:

While 74 years of climate data offer a unique temporal perspective, extracting meaningful insights requires advanced analytical methodologies. Balancing long-term trend analysis with relevance to contemporary climatic changes is crucial to avoid succumbing to data noise.

5. Model Evaluation and Generalizability:

Robust evaluation of the predictive model's effectiveness across diverse climatic scenarios and geographical regions within Bangladesh is challenging. Balancing model complexity with generalizability is essential for real-world applicability, requiring rigorous testing against historical flood occurrences.

Addressing these challenges requires a multidisciplinary approach, incorporating expertise in data science, ethics, climatology, and community engagement. Proactively acknowledging and navigating these challenges enhances the research methodology, contributing meaningfully to the discourse on disaster resilience in flood-prone regions.

CHAPTER 3

Research Methodology

3.1 Research Subject

This study investigates the effectiveness of ensemble methods (Bagging, Stacking) with a focus on optimizing predictive accuracy. It synthesizes expansive datasets, including geographical features and 74 years of historical flood data, forming the foundation for developing and evaluating a machine learning predictive model. The research extends beyond technical aspects to address ethical dimensions in deploying machine learning for disaster preparedness. Aligned with clear objectives and addressing identified problem statements, it contributes substantively to discussions on resilient disaster management and sustainable development in flood-prone regions through machine learning innovation and a comprehensive dataset.

3.2 Dataset Preparation

The compilation and preparation of datasets play a pivotal role in advancing scientific research, especially in domains such as climate science and disaster management. In this study, we meticulously curated a comprehensive dataset spanning a significant period of 74 years (from 1949 to 2022) to investigate the intricate relationship between climate variables and flood occurrences in Bangladesh. This dataset amalgamates weather data sourced from the Bangladesh Meteorological Department (BMD) and flood occurrence information obtained from diverse outlets, creating a robust resource for further analysis and modeling.

The initial 65 years of the dataset (1949-2013) were acquired from a primary source, specifically the Bangladesh Meteorological Department [17]. This portion of the dataset serves as a foundational base, providing a historical context for understanding climate patterns in the region. The climate data obtained from BMD covers crucial parameters, including Rainfall, Cloud Coverage, Relative Humidity, Minimum Temperature, Wind Speed, among others. The wealth of information contained within these records facilitates a detailed examination of long-term climate trends, offering insights into the climatic conditions that have shaped Bangladesh over the decades.

To augment the historical dataset, we undertook a meticulous process of collecting flood occurrence data for specific months and years. This involved sourcing information from an array of outlets, including annual flood reports, newspapers, research papers, and other relevant sources. By merging this flood occurrence data with the existing weather data from BMD, we created an enriched dataset that not only spans an extensive temporal range but also incorporates real-world instances of flood events.

The subsequent phase of dataset development focuses on the years 2014 to 2022, encompassing an additional nine years. For this period, we again turned to the Bangladesh Meteorological Department as a reliable source for climate data. Simultaneously, flood occurrence information was gathered from a diverse set of sources, such as news media, articles, and reports from non-governmental organizations (NGOs). This multi-sourced approach ensures a comprehensive and up-to-date representation of flood events in the more recent years, allowing for the exploration of evolving patterns and trends.

In total, our dataset comprises 24,108 instances, each capturing a unique combination of climate variables and flood occurrence information. This expansive dataset covers most of Bangladesh, enabling a granular analysis of regional variations in climate and flood dynamics. The inclusion of districts as a geographical unit adds a spatial dimension to the dataset.

The dataset's attributes, ranging from Rainfall and Cloud Coverage to Relative Humidity and Wind Speed, offer a multifaceted view of the climatic conditions influencing flood events. Researchers and practitioners in climate science, hydrology, and disaster management can leverage this dataset to formulate evidence-based strategies for mitigating the impact of floods in Bangladesh.

In conclusion, our dataset preparation endeavors represent a meticulous process of integrating historical climate data with real-world flood occurrence instances. By incorporating information from multiple sources and spanning over seven decades.

Table 3.2.1: Dataset description with its unit.

Name	Description	Unit
Station_Name	Name of the respective weather station	Not applicable
Year	Year of the recorded data	Not applicable
Month	Month of the recorded data	Not applicable
Max_Temp	Maximum temperature of the recorded month	Celsius
Min_Temp	Minimum temperature of the recoded month	Celsius
Rainfall	Rainfall of the recorded month	Centimeter
Relative_Humidity	The amount of atmospheric moisture present in the air relative to the amount that would be present if the air were saturated.	Percentage
Wind_Speed	The rate of at which air is moving	Meters per second
Cloud_Coverage	The mass of cloud covers the sky	Okta
Bright_Sunshine	The total hour of sunlight is stronger than the threshold.	Hours per day
Station_Number	Unique number of each station.	Not applicable
X_COR	X coordinate of the station	
Y_COR	Y coordinate of the station	
LATITUDE	Latitude of the station	Degree
LONGITUDE	Longitude of the station	Degree
ALT	Altitude of the station	Meter
Period	Combination of year and each month	Not applicable
Flood?	Is there any flood or not	1 for yes

3.3 Data Preprocessing

Data preprocessing is one of the most important tasks in any machine learning model development involving a systematic array of operations aimed at refining raw data into a form amenable to comprehensive analysis and modeling. The initial phase of this process entails a meticulous examination of the dataset to identify and rectify discrepancies such as missing values, outliers, and inconsistencies. Addressing missing values is executed through methods like imputation or removal, ensuring the dataset's integrity. Outliers, defined as data points significantly deviating from the norm, are treated judiciously to prevent undue influence on the model; this involves either correction or application of appropriate statistical techniques. After addressing missing values and outliers, the dataset undergoes feature scaling and normalization to standardize the scale of numerical attributes. This step is pivotal in ensuring that all variables contribute equitably to the model's learning process.

Missing Value Handle: The dataset spans 74 years, encompassing monthly weather data for Bangladesh from 1949 to 2022. Despite a low percentage of missing values, each instance represents a complete month of weather observations, precluding their removal. Given the dynamic nature of weather conditions, KNN Imputation is employed to fill missing values, preserving temporal coherence. This method leverages neighboring data points to approximate original values, ensuring the dataset's fidelity. We have used the value of n_neighbors = 2. The formula of KNN Imputation goes as follows:

$$\hat{X}_i = \frac{1}{k} \sum_{j \in N_i} X_j$$

Where:

- $\hat{X}_i$ is the imputed value for the missing data point i.
- N_i represents the set of k nearest neighbors to the missing data point i.
- X_j denotes the observed values of the neighbors in N_i.

- k is the number of nearest neighbors considered for imputation.

Scaling: Feature scaling is a vital preprocessing step in machine learning, ensuring that numerical features are standardized to a common range. This normalization prevents biases in model training, where certain features might disproportionately influence the outcome. By achieving uniform scales, feature scaling enhances algorithm convergence, promotes model robustness, and contributes to better generalization across diverse datasets. We have used Standard Scaler. The Standard Scaler is a widely employed technique in feature scaling for machine learning, designed to transform numerical data into a standard distribution with a mean of 0 and a standard deviation of 1. This process is crucial for ensuring that all features contribute equally to model training, particularly in algorithms sensitive to varying magnitudes of input features.

The formula for standardization using the Standard Scaler is given by:

$$X_{\text{std}} = \frac{X - \text{mean}\,(X)}{\text{std}\,(X)}$$

Here, X_{std} represents the standardized values, X denotes the original feature values, mean (X) is the mean of the feature, and std (X) is the standard deviation of the feature.

By applying the Standard Scaler, the data distribution is centered around zero, and the spread of values is normalized.

Feature Engineering: We have a total of 19 columns in our dataset. So, to reduce the dimensions of the data and fetch the more accurate results from the dataset we must employ feature engineering techniques. In this study we have used the Foreword Selection and Backward Elimination method. First, we done Foreword Selection and the Backward Elimination separately. Then we intersect the results and obtain the optimal solutions.

Forward selection is a feature engineering method that systematically builds a model by adding one feature at a time, selecting the most impactful predictor at each step. This iterative process aims to enhance model performance by including the most relevant features incrementally. Mathematically, at each step k, the algorithm selects the feature X_i

that maximizes a chosen criterion, often a performance metric such as accuracy or information gain:

$$X_k = \arg \max_{X_i \in \text{ remaining features}} \text{Criterion } (X_i \mid X_1, X_2, \dots, X_{k-1})$$

The process continues until a predefined stopping criterion is met, such as achieving optimal model performance or reaching a specified number of features.

Conversely, backward elimination starts with the full feature set and systematically removes the least significant feature in each iteration. The process continues until a stopping criterion is met, resulting in a reduced set of features that optimally contribute to the model's performance. Mathematically, at each step k, the algorithm eliminates the feature X_i that minimizes a chosen criterion:

$$X_k = \arg \min_{X_i \in \text{ remaining features}} \text{Criterion } (X_i \mid X_1, X_2, \dots, X_{k-1})$$

3.4 Algorithm Details

Decision Tree: The decision tree method determines which characteristic offers the optimal data split at each node. This is decided by a dividing criterion, such mean squared error for regression or Gini impurity for classification.

Root node: The starting point, representing the entire dataset.

Internal nodes: Represent decisions based on feature values.

Branches: Connect nodes, indicating possible outcomes of decisions.

Leaf nodes: Terminal nodes representing final classifications or predictions.

Formulas:

Entropy: Measures uncertainty or impurity in a dataset:

$$\text{Entropy}\ (S) = -\sum p(c) \cdot \log_2\ (p(c))$$

- S : dataset
- c : classes within S
- $p(c)$: proportion of data points in class c

Information Gain: Measures the reduction in entropy after a split:

$$\text{Gain}\ (S, A) = \text{Entropy}\ (S) - \sum \left(\frac{|S_v|}{|S|}\right) \cdot \text{Entropy}\ (S_v)$$

- S : dataset
- A : attribute used for splitting.
- S_v : subsets of S based on A 's values.

Gini Index: Alternative to entropy, measuring impurity:

$$\text{Gini}\ (S) = 1 - \sum p(c)^2$$

- S : dataset
- $p(c)$: proportion of data points in class c

Random Forest Classifier: Developed by *Leo Breiman [1]* it is an ensemble learning method that combines the predictions from multiple individual decision trees to improve the overall accuracy and robustness of the model. It is a collection of unpruned classification or regression trees formed through the random sampling of training data. During the induction process, features are randomly selected. The prediction is determined by combining the individual predictions of the ensemble, employing a majority vote for classification, or averaging for regression. This ensemble technique, involving the random selection of both samples and features, enhances the robustness and predictive performance of the model by aggregating diverse perspectives from the constituent trees. *[2]*

Random Forest Algorithm Steps:

1. Randomly select N bootstrap samples from the training data (with replacement).

2. For each bootstrap sample, grow a decision tree by recursively selecting the best split from a random subset of features at each node.

3. Repeat steps 1 and 2 to create a forest of decision trees.

4. For classification, each tree "votes" for a class, and the majority class is chosen as the final prediction. For regression, the predictions from all trees are averaged.

The final prediction of a Random Forest for a given input is determined by a majority vote of its constituent trees. If T is the number of trees in the forest, and C_i is the predicted class by the i-th tree, then the final predicted class final C final is given by:

$$C_{final} = argmax_c \left(\sum_{i=1}^{T} 1(C_i = c) \right)$$

XGboost Classifier: Developed by *Tainqi Chen et.al [3]* XGBoost is based on the gradient boosting framework. It minimizes a loss function by adding weak learners, and each new learner corrects the errors of the previous one. The key idea is to fit a new model to the residuals (the differences between the actual and predicted values) of the existing model. XGBoost incorporates regularization techniques to prevent overfitting. It includes both L1 (Lasso) and L2 (Ridge) regularization terms in the objective function to control the complexity of the individual trees. The objective function of XGBoost is a combination of a loss function and a regularization term. The goal is to minimize this objective function during the training process. The general form of the objective function is:

$$Objective = \sum_{i=1}^{n} \ell(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)$$

Where, $\ell(y_i, \hat{y}_i)$: Loss function measuring the difference between the true label y_i and the predicted label $\hat{y}_i$ for the i-th observation. And $\Omega(f_k)$: Regularization term for the k-th tree, controlling the complexity of the tree.

AdaBoost Classifier: Developed by *Yoav Freund et al.[4]* AdaBoost, short for Adaptive Boosting, is an ensemble learning method that combines the predictions of multiple weak learners (typically shallow or weak classifiers) to create a strong classifier. The idea behind AdaBoost is to sequentially train weak learners, assigning higher weights to the misclassified samples in each iteration. This focuses the subsequent weak learners on the examples that were previously difficult to classify correctly, improving overall performance.

Error of Weak Learner (epsilon_t):

$$\epsilon_t = \frac{\sum_{i=1}^{N} w_i \cdot \mathbb{I}(h_t(x_i) \neq y_i)}{\sum_{i=1}^{N} w_i}$$

- N is the number of training examples.
- w_i is the weight assigned to the i-th example.
- $h_t(x_i)$ is the prediction of the weak learner for the i-th example.
- y_i is the true label of the i-th example.
- $\mathbb{I}$(condition) is the indicator function (1 if true, 0 if false).

Classifier Weight (alpha_t):

$$\alpha_t = \frac{1}{2} \cdot \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

Update Weights:

$$w_{t+1,i} = w_{t,i} \cdot \exp\left(-\alpha_t \cdot y_i \cdot h_t(x_i)\right)$$

Where:

- $w_{t+1,i}$ is the updated weight for the i-th example.
- α_t is the weight of the weak learner.
- y_i is the true label of the i-th example.
- $h_t(x_i)$ is the prediction of the weak learner for the i-th example.

Final Classifier (H(**x**)) :

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t \cdot h_t(x)\right)$$

Where:

- T is the number of weak learners.
- α_t is the weight of the weak learner.
- $h_t(x)$ is the prediction of the $t^{\text{-th}}$ weak learner.

Support Vector Machine: SVMs aim to find a hyperplane (a line in 2D, a plane in 3D, and so on in higher dimensions) that best separates two classes of data points, maximizing the margin between them**.**

Formulas:

1. Hyperplane Equation:

$$w^T \cdot x + b = 0$$

- w : weight vector, determining the orientation of the hyperplane
- x : input data point
- b : bias term, adjusting the hyperplane's position

2. Margin Width:

$$\gamma = \frac{2}{\parallel w \parallel}$$

- γ : margin width

- $\| w \|$: Euclidean norm of the weight vector

3. Optimization Problem:

$$\text{Minimize } \| w \|^2$$

Subject to: $y_i \cdot (w^T \cdot x_i + b) \geq 1$ for all data points (x_i, y_i)

- y_i : class label (+1 or -1)

4. Lagrangian Formulation:

$$L(w, b, \alpha) = \frac{1}{2} \| w \|^2 - \sum \alpha_i \cdot (y_i \cdot (w^T \cdot x_i + b) - 1)$$

- α_i : Lagrange multipliers

Logistic Regression : Logistic Regression is a statistical method designed for predicting binary outcomes. It estimates the probability of a certain event occurring based on a set of independent variables. Unlike linear regression, it models the relationship between the independent variables and the log-odds of the event. Logistic Regression employs a sigmoid function to map predicted probabilities to values between 0 and 1, ensuring that the output is interpretable as probabilities.

Formula:

The logistic regression model is represented by the formula:

$$\text{logit } (P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

- logit (P) : The log odds of the event occurring (the log of the probability of the event happening divided by the probability of it not happening).
- β_0 : The intercept term, representing the value of the log odds when all independent variables are 0 .
- $\beta_1, \beta_2, \dots, \beta_n$: The coefficients or weights assigned to each independent variable. These represent the change in log odds associated with a one-unit increase in the corresponding variable.

- $X_1, X_2, \ldots, X_n$: The independent variables, which are the factors used to predict the outcome.

Bagging : Developed by *Leo Breiman. [5]* Bagging is an ensemble learning technique that aims to improve the stability and accuracy of machine learning algorithms. It works by training multiple instances of a base classifier on different subsets of the training data. The term "bagging" is derived from the notion of bootstrap sampling, where random samples are drawn with replacement from the original dataset. The basic idea behind bagging is to reduce overfitting and variance by combining predictions from multiple models, thereby achieving better generalization performance on unseen data. Bagging is effective when the base classifier has high variance or tends to overfit the training data. Bagging is the process of generating many versions of a predictor and aggregating them into a single predictor. This aggregation technique uses a majority vote for class forecasts and averages the predictions for numerical results. By making bootstrap duplicates of the initial learning set and using them as new learning sets, multiple predictor versions are produced. Bagging can result in considerable increases in accuracy, according to experiments on actual and simulated datasets that use subset selection, regression trees, and classification in linear regression. The primary cause of this improvement is the prediction method's intrinsic instability, as alterations to the learning set might have observable effects. For binary classification, the bagging classifier's final prediction can be determined as follows:

$$\hat{y}_{\text{final}} = \text{sign}\left(\sum_{i=1}^{N} \hat{y}_i\right)$$

where:

- $\hat{y}_{\text{final}}$ is the final predicted class.
- N is the number of base classifiers.
- $\hat{y}_i$ is the predicted class of the i-th base classifier.

Stacking : Developed by *David H. Wolpert et. al [7].* To enhance prediction performance, stacking is an ensemble learning strategy that builds a meta-model, also known as a meta-classifier, by combining many basic models. When it comes to classification challenges, the stacking classifier employs the predictions made by many base classifiers as input characteristics for a higher-level classifier that's also known as the meta-classifier. The ultimate prediction is then made by this meta-classifier.

Base Classifiers: These are the individual models that make predictions on the input data. They can be of different types, utilizing diverse algorithms, and may have varying strengths and weaknesses.

Meta-Classifier: This is the higher-level classifier that takes the predictions of the base classifiers as input features and makes the final prediction. It can be any classifier, ranging from decision trees to more complex models like random forests or even neural networks.

Workflow:

- Initially, training and testing sets of the input data are separated.
- Subsets of the training set are created, and each subset is utilized to train a different base classifier.
- A new feature matrix is produced by combining the predictions made by each base classifier on the testing set.
- The meta-classifier is trained using both the new feature matrix and the old features.
- The trained meta-classifier is used to make the final prediction.

The Stacking process involves combining the predictions of multiple base classifiers to create a feature matrix, which is then used to train the meta-classifier. Let's denote the predictions from the base classifiers as $P_1, P_2, \ldots, P_n$, where n is the number of base classifiers.

The feature matrix X_{meta} for training the meta-classifier is constructed as follows:

$$X_{\text{meta}} = \left[X_{\text{original}}, P_1, P_2, \ldots, P_n\right]$$

Here, X_{original} represents the original feature matrix of the training set.
The meta-classifier is trained using the feature matrix X_{meta} and the corresponding true labels.

3.5 STATISTICAL ANALYSIS

Accuracy: Accuracy is a common metric used to evaluate the performance of a classification model. It measures the ratio of correctly predicted instances to the total number of instances in the dataset. In mathematical terms, if you have a binary classification problem (two classes, often denoted as positive and negative), the formula can be expressed as:

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

True Positives (TP): The number of instances correctly predicted as positive.

True Negatives (TN): The number of instances correctly predicted as negative.

False Positives (FP): The number of instances incorrectly predicted as positive.

False Negatives (FN): The number of instances incorrectly predicted as negative.

Precision: Precision is a metric used in the field of machine learning and statistics to evaluate the performance of a classification model, particularly in binary classification problems. Precision measures the accuracy of the positive predictions made by the model, indicating how many of the predicted positive instances are actually relevant or true positives. Precision is calculated using the following formula:

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

A higher precision value indicates that the model has fewer false positives, meaning that when it predicts a positive instance, it is more likely to be correct.

Recall: Recall is a metric used in the field of machine learning and statistics to evaluate the performance of a classification model, particularly in the context of binary classification problems. It measures the ability of a model to correctly identify all relevant instances within a dataset. Recall is also known as sensitivity, true positive rate, or hit rate.

The formula for Recall is given by:

$$Recall = \frac{TP}{TP + FN}$$

A high recall value indicates that the model is good at capturing most of the positive instances in the dataset. However, it may come at the cost of a higher number of false positives.

F1 Score: The F1 score is a metric commonly used in machine learning and statistics to evaluate the performance of a binary classification model. It is particularly useful when the class distribution is imbalanced, meaning that one class significantly outnumbers the other. The F1 score is calculated using precision and recall, which are two other important metrics in classification evaluation. Precision is the ratio of true positive predictions to the total number of positive predictions (both true positive and false positive), while recall is the ratio of true positive predictions to the total number of actual positive instances.

The formula for precision (P), recall (R), and the F1 score is as follows:

$$F1\ score = 2 \times \frac{P \times R}{P + R}$$

The F1 score is the harmonic mean of precision and recall. It ranges from 0 to 1, with 1 being the best possible score. The harmonic mean is used instead of the arithmetic mean to ensure that the F1 score gives more weight to lower values. This is particularly important when dealing with imbalanced datasets, where the impact of misclassifying the minority class is often more significant.

ROC-AUC: Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) are metrics commonly used to evaluate the performance of binary classification models.

The ROC curve is a graphical representation of a binary classification model's performance across different threshold settings. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The ROC curve helps visualize the trade-off between sensitivity and specificity at different classification thresholds. AUC measures the area under the ROC curve. It provides a single scalar value that summarizes the performance of a classification model across all possible classification thresholds. AUC ranges from 0 to 1, where a higher AUC indicates better model performance. An AUC of 0.5 suggests that the model's performance is equivalent to random chance, while an AUC of 1.0 indicates perfect discrimination.

MCC: The Matthews Correlation Coefficient (MCC) is a metric used to assess the quality of binary classification models, particularly when dealing with imbalanced datasets. It takes into account true positives, true negatives, false positives, and false negatives to provide a balanced measure of a model's performance.

The MCC is calculated using the following formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The MCC ranges from -1 to +1, where +1 indicates a perfect prediction, 0 indicates no better than random prediction, -1 indicates total disagreement between prediction and observation.

Brier score: The Brier score is calculated as the mean squared difference between the predicted probabilities and the actual outcomes. It provides a measure of the overall accuracy of a set of probabilistic predictions, with lower scores indicating better performance.

Here's the formula for calculating the Brier score:

$$Brier\ Score = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{P_i} - \mathrm{O_i})^2$$

Where N is the number of observations or instances. Pi is the predicted probability for the i-th observation. Oi is the actual outcome for the i-th observation. It is a binary variable, taking the value of 1 for a successful outcome and 0 for an unsuccessful outcome. The Brier score ranges from 0 to 1, with 0 indicating perfect accuracy and 1 indicating perfect inaccuracy. Lower Brier scores reflect better calibration and accuracy of the probabilistic predictions.

3.6 PROPOSED METHODOLOGY

This study presents a comprehensive approach for flood prediction utilizing ensemble machine learning methodologies. The initial phase involves the amalgamation and organization of 74 years' worth of climate data. After dataset acquisition, the independent and target variables are delineated, followed by a randomized partitioning of the dataset into a 90:10 ratio for training and testing purposes.

The pivotal stage in any machine learning investigation, data preprocessing, is then undertaken. Within the dataset, a sole categorical feature denoted as "Station Names" is identified, describing the names of observed weather station locations. This column is subsequently omitted due to the presence of an alternative feature, "Station Number" serving as a unique identifier for each respective "Station Name." Addressing missing values is accomplished through the application of KNN Imputer methods, and normalization is executed using the standard scaler.

Feature engineering is then implemented, incorporating forward selection and backward elimination techniques. Subsequently, two ensemble methods such as Bagging and Stacking are deployed with six classifiers, namely Decision Tree (DT), Random Forest Classifier (RFC), XGBoostClassifier (XGB), AdaBoost Classifier, Support Vector Classifier (SVC) and Logistic Regression (LR) creating two model with each ensemble approach, total of 4 ensemble models ware proposed in this study. The performance of these models is evaluated using metrics such as accuracy, precision, recall, ROC-AUC, and

F1-score. Additionally, the Mathies Correlation Coefficient (MCC) and Brier score are considered for an encompassing assessment of overall model performance.

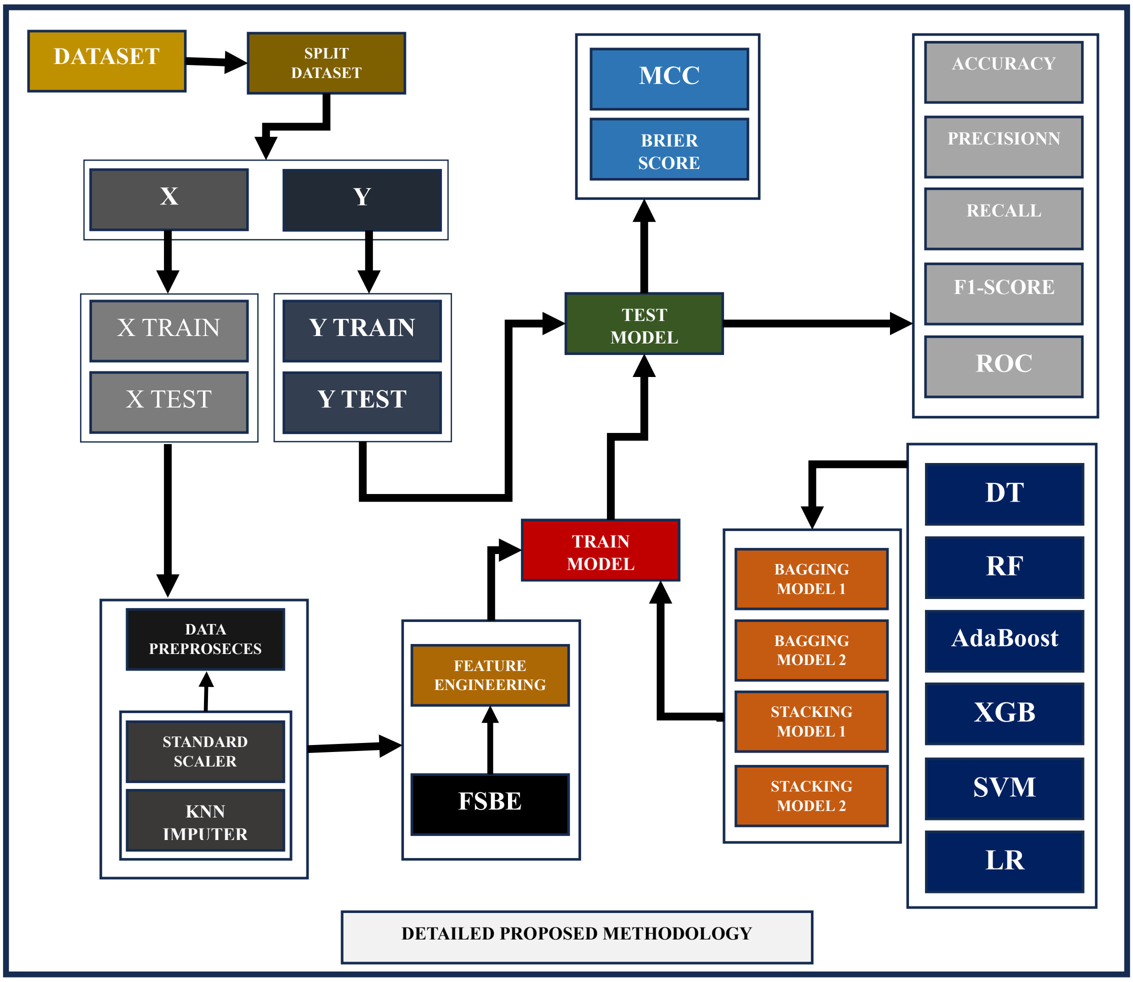

Figure 3.6.1: Detailed proposed methodology.

CHAPTER 4

Experimental Results and Discussion

4.1 Model Performance

Bagging Model 1: This model has XGBoostClassifier (XGB), AdaBoost Classifier, Support Vector Classifier (SVC) as base classifier. Majority voting was calculated among them. This model achieved accuracy of 95.44% and TN = 1958, TP = 343, FP = 50, FN = 60.

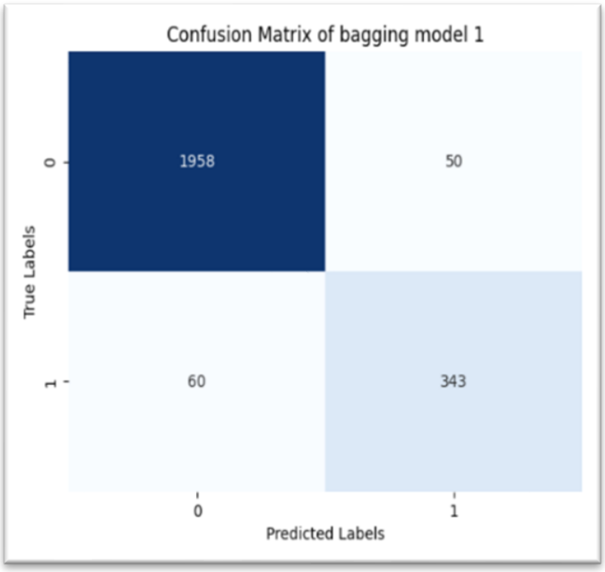

Figure 4.1.1: Confusion matrix of Bagging model 1.

Bagging Model 2: This bagging model has Decision Tree (DT), Random Forest Classifier and XGBoostClassifier (XGB) as base classifier and achieved 97.22% accuracy TN = 1977, TP = 367, FP = 31, FN = 36.

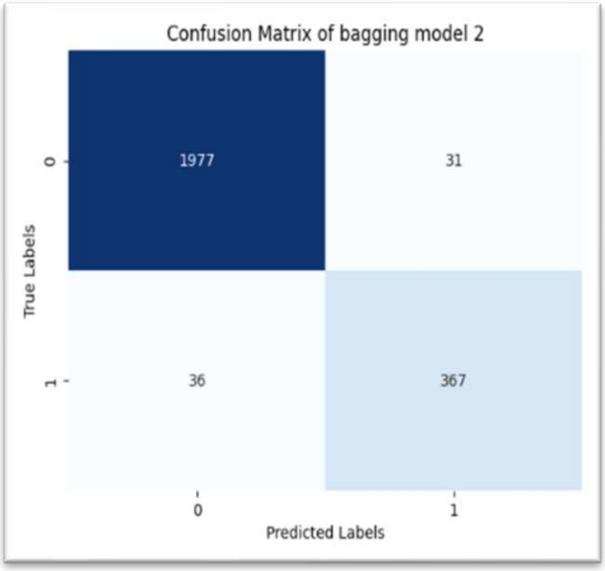

Figure 4.1.2: Confusion matrix of Bagging model 2.

Stacking Model 1: This model has Decision Tree (DT), Random Forest Classifier and XGBoostClassifier (XGB) as base classifier. While final estimator was Logistic Regression. This model showed accuracy of 96.89% and TP = 358, TN = 1978, FP = 30 FN = 45.

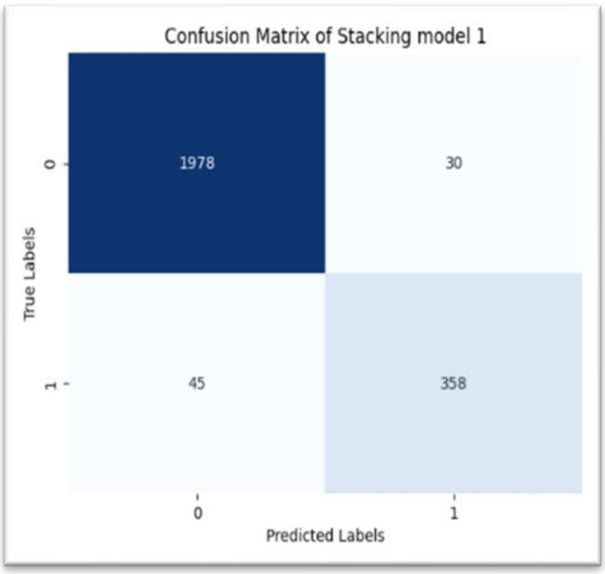

Figure 4.1.3: Confusion matrix of Stacking model 1.

Stacking Model 2: This model showed accuracy of 94.90% and TN = 1942, TP = 346, FP = 66, FN = 57. While this model has XGBoostClassifier (XGB), AdaBoost Classifier, Support Vector Classifier (SVC) and Decision Tree as final estimator.

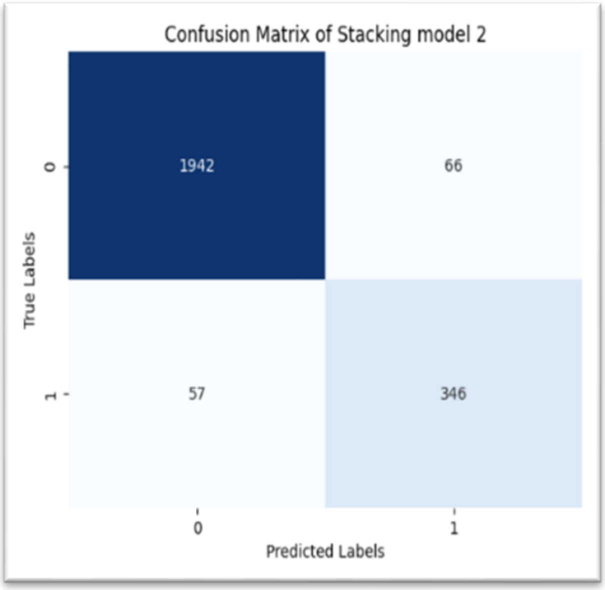

Figure 4.1.4: Confusion matrix of Stacking model 2.

4.2 Comparative Analysis

Comparison between all models:

Accuracy: Bagging model 1 and Stacking model 1 achieved highest accuracy with 97.22% and 96.89% respectively. The other two models Bagging 2 and Stacking 2 also have satisfactory accuracy with 95.44% and 94.9% respectively.

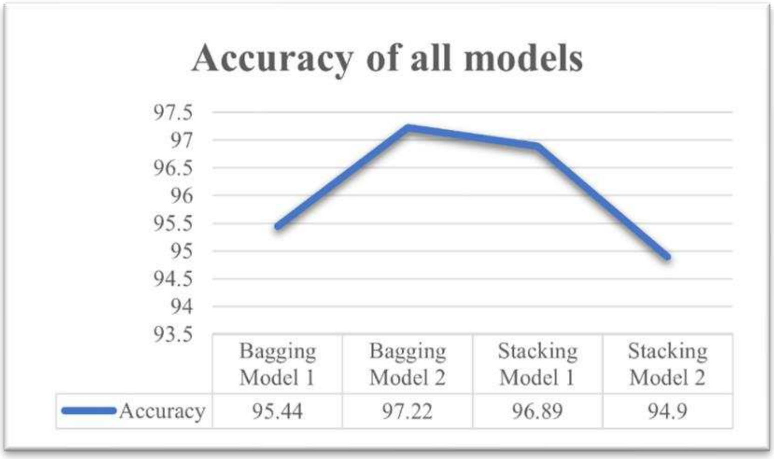

Figure 4.2.1: Accuracy of all models.

Recall: With recall values averaging 0.91, all three models (Bagging Model 1, Bagging Model 2, and Stacking Model 1) show promise in identifying affirmative cases. With a recall of 0.86, Stacking Model 2 is a little less sensitive in detecting positive cases.

Figure 4.2.2: Recall of all models.

Precision: All models (Bagging Model 1, Bagging Model 2, and Stacking Model 1) achieve a high precision of 0.92, indicating accurate positive predictions about 92% of the time. Stacking Model 2 has a slightly lower precision of 0.84, but it still demonstrates a respectable level of accuracy in identifying positive instances.

Figure 4.2.3: Precision of all models.

F1 Score: Bagging Model 2 excels with the highest F1-score of 0.92, indicating a superb balance between precision and recall. Stacking Model 1 follows closely with a strong F1-score of 0.91. Bagging Model 1 demonstrates a good balance at 0.86, while Stacking Model 2 has a slightly lower F1-score of 0.85, suggesting a somewhat larger trade-off between precision and recall.

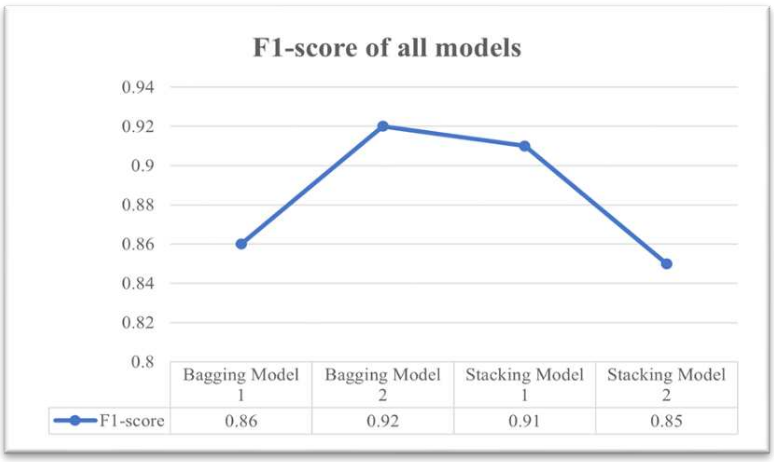

Figure 4.2.4: F1-score of all models.

ROC-AUC: Bagging Model 2 excels with the highest ROC value of 0.95, indicating superior discrimination ability. Stacking Model 1 closely follows with a strong ROC of

0.94. Both Bagging Model 1 and Stacking Model 2 achieve good ROC values of 0.91, reflecting effective discrimination, though not as high as Bagging Model 2.

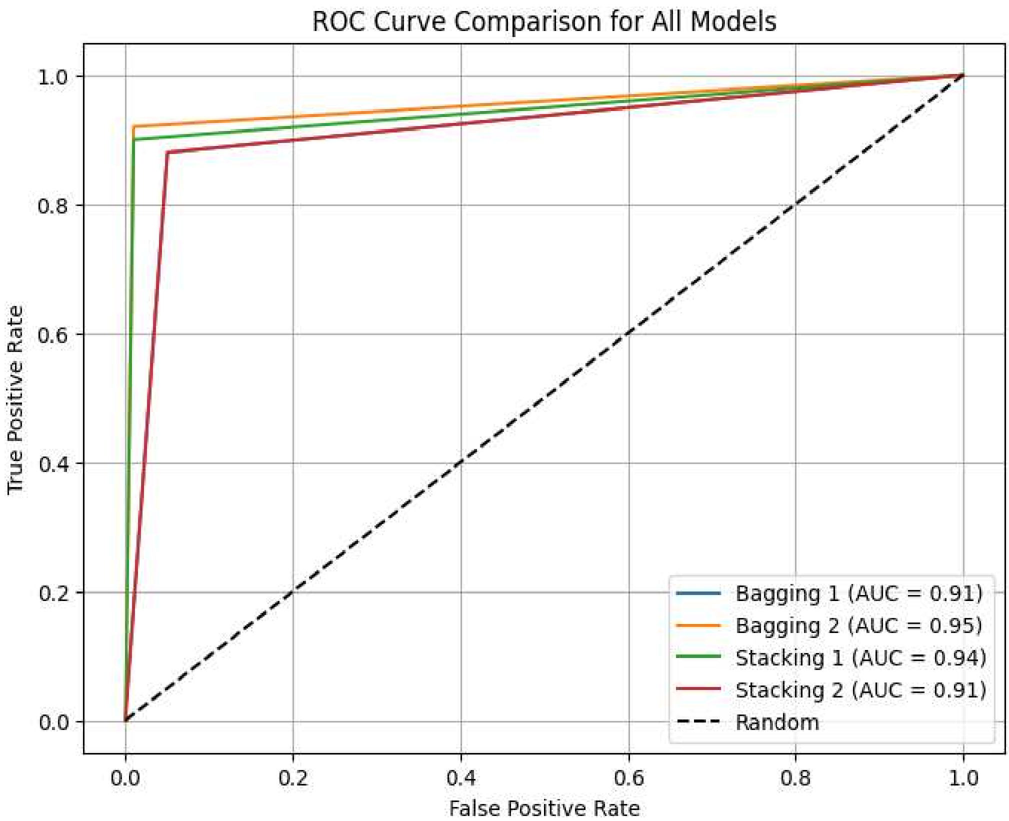

Figure 4.2.5: ROC curve of all models.

Brier Score: Bagging Model 2 has the lowest Brier Score (0.028), indicating the most accurate probability predictions. Stacking Model 1 follows closely with a Brier Score of 0.031. Bagging Model 1 has a moderate Brier Score of 0.046, while Stacking Model 2 has the highest Brier Score of 0.051, suggesting less accurate probability estimates.

Figure 4.2.6: Brier Score of all models.

MCC: Bagging Model 2 stands out with the highest MCC (0.9), indicating a robust overall performance. Stacking Model 1 closely follows with a high MCC of 0.89. Bagging Model 1 has a moderate MCC of 0.83, while Stacking Model 2 has the lowest MCC at 0.82, suggesting a slightly weaker overall performance.

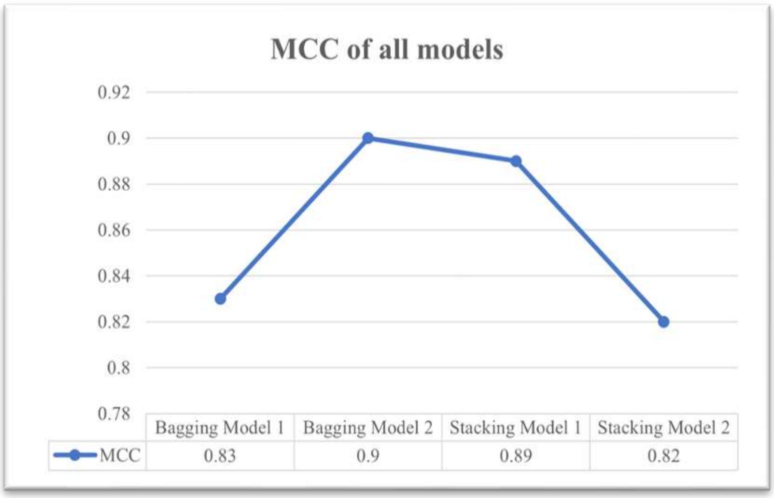

Figure 4.2.7: MCC of all models.

Comparison with other studies: *Miah Mohammad Asif et al. [9]* utilized various machine learning models to predict floods, achieving an accuracy of 86.76%, precision of 0.76, recall of 0.67, and a maximum ROC score of 0.87. In a separate study, *Gauhar et al. [17]* employed the K-Nearest Neighbors algorithm on a similar dataset, obtaining notable metrics with the highest accuracy, precision, recall, and F1 score at 94.91%, 92.50, 91.00, and 92.00, respectively, along with the highest ROC-AUC value of 0.96.

In contrast, our study surpasses these results, achieving an accuracy of 97.22%, recall of 0.91, precision of 0.92, and an F1 score of 0.92. Additionally, introducing Matthews Correlation Coefficient (MCC) and Brier Score yielded values of 0.9 and 0.028, respectively, showcasing superior performance. The study's predictive model demonstrated the highest ROC-AUC value of 0.95, highlighting advancements in flood prediction accuracy and related metrics compared to previous research efforts.

4.3 RESULT DISCUSSION

Four different models: Bagging Model 1, Bagging Model 2, Stacking Model 1, and Stacking Model 2 models have been evaluated based on various metrics such as accuracy, precision, recall, F1-score, ROC (Receiver Operating Characteristic), MCC (Matthews Correlation Coefficient), and Brier Score.

Bagging Model 1 exhibits a strong predictive performance with an accuracy of 95.44%, indicating a high level of correctness in its classifications. With a precision of 0.92, it demonstrates accurate positive predictions, while a recall of 0.91 underscores its ability to capture positive instances effectively. The F1-score of 0.86 signifies a balanced trade-off between precision and recall. The ROC value of 0.91 reflects good discrimination ability, and the Matthews Correlation Coefficient (MCC) of 0.83 indicates a robust overall performance. Additionally, the low Brier Score of 0.046 suggests well-calibrated probability estimates.

Bagging Model 2 surpasses its counterpart with an impressive accuracy of 97.22%, showcasing enhanced overall predictive capabilities. The precision, recall, and F1-score, all at 0.92, indicate consistent and accurate positive predictions. With a high ROC value of 0.95, Bagging Model 2 demonstrates improved discrimination ability. The MCC of 0.9 signifies a strong overall performance, and the Brier Score of 0.028 suggests accurate probability predictions, making Bagging Model 2 the standout performer among the models discussed.

Stacking Model 1 achieves a commendable accuracy of 96.89%, indicating a high level of correctness in its predictions. With a precision of 0.92, it maintains accuracy in positive predictions, though the recall of 0.89 is slightly lower than the bagging models. The F1-score of 0.91 demonstrates a good balance between precision and recall. The ROC value of 0.94 suggests strong discrimination ability, and the MCC of 0.89 indicates a robust overall performance. The Brier Score of 0.031 reflects accurate probability predictions, making Stacking Model 1 a reliable choice.

Stacking Model 2, while exhibiting a respectable performance, lags behind the other models with an accuracy of 94.9%. The precision of 0.84 indicates that positive predictions are correct about 84% of the time, and the recall of 0.86 highlights a reasonable ability to capture positive instances. The F1-score of 0.85 strikes a balance between precision and recall. With a ROC value of 0.91, Stacking Model 2 shows good discrimination ability, although not as high as Bagging Model 2. The MCC of 0.82 suggests a slightly weaker overall performance, and the higher Brier Score of 0.051 indicates less accurate probability predictions compared to the other models.

In summary, Bagging Model 2 appears to be the most robust model among the four, boasting the highest accuracy, precision, recall, F1-score, ROC, MCC, and the lowest Brier Score. Stacking Model 1 also performs well but with a slightly lower accuracy and recall compared to Bagging Model 2. Bagging Model 1 follows closely in terms of performance. Stacking Model 2, while still achieving respectable metrics, falls behind the other models, especially in terms of accuracy and MCC.

TABLE 4.3.1: Overall results of all models.

Model	Accuracy	Precision	Recall	F1-score	ROC	MCC	Brier Score
Bagging Model 1	95.44	0.92	0.91	0.86	0.91	0.83	0.046
Bagging Model 2	97.22	0.92	0.91	0.92	0.95	0.9	0.028
Stacking Model 1	96.89	0.92	0.89	0.91	0.94	0.89	0.031
Stacking Model 2	94.9	0.84	0.86	0.85	0.91	0.82	0.051

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Accurate flood prediction models, exemplified by the Bagging 2 and Stacking 1, profoundly impact Bangladeshi society. They enhance early warning systems, empower proactive disaster response, and improve infrastructure resilience. The economic impact is mitigated through informed risk management, fostering stability in agriculture and businesses. Socially, these models promote community resilience, cohesion, and engagement in disaster preparedness. While acknowledging challenges, the holistic implementation of these models contributes significantly to public safety, economic stability, and community well-being in flood-prone regions of Bangladesh.

5.2 Impact on Environment

Accurate flood prediction models, such as the Bagging 2 and Stacking 1, positively impact Bangladesh's environment. They aid in preserving ecosystems and biodiversity by preventing catastrophic floods, contributing to sustainable water resource management, optimizing agricultural practices, and mitigating soil erosion. These models offer a valuable tool for balancing ecological effects, emphasizing the importance of a holistic and adaptive approach to ensure positive and sustainable environmental outcomes in Bangladesh.

5.3 Ethical Aspect

The ethical dimensions of implementing accurate flood prediction models in Bangladesh are vital, requiring a careful balance between technological advancements, human welfare, and environmental preservation. Key ethical considerations include ensuring equitable benefits distribution, transparent communication, data privacy protection, prevention of misuse, and addressing the environmental impact. Engaging marginalized communities, communicating models functioning transparently, and refining models for accuracy contribute to ethical practices. Striking the right balance in sensitivity, specificity, and environmental preservation is imperative for responsible and equitable deployment of flood prediction models in the face of climate-related challenges.

5.4 Sustainable Plan

To address the challenges in flood prediction, a sustainable plan for implementing accurate models in Bangladesh must be strategic and holistic. Key components include community engagement and education, capacity building, continuous model improvement, integration with early warning systems, data governance, privacy protection, ecosystem conservation, and policy integration. This comprehensive approach ensures long-term effectiveness, community resilience, and environmental consciousness, contributing to a sustainable strategy for mitigating the impacts of climate-related disasters.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

In this study we have utilized four ensemble models of two approaches namely Bagging and Stacking. Data acquisition for this study combines publicly available 65 years of climate data from 1949 to 2013 and 9 years of recently added data from 2014-2022. Which is gathered from Bangladesh Meteorological Department (BMD). Data processing of this study includes KNN imputer for missing values and standard sealer for normalize the data. In this study feature engineering method forward selection and backward elimination was also used to reduce the dimension of the data. Then four model were trained on test data, and it was evaluated on different performance and evaluation metrics. This study achieved 9.7.22% of accuracy on Bagging model 2 which uses DT, RF and XGB as base classifier. This study supports that ensemble models can predict flood on the Bangladesh region more accurately and robustly.

6.2 Conclusions

The study underscores the transformative potential of accurate flood prediction models in Bangladesh. Leveraging machine learning, including the Random Forest Classifier (RFC) and Bagging, these models offer significant societal benefits by enhancing disaster preparedness and response. Their precision empowers communities, fostering shared responsibility and proactive interventions to minimize losses. Environmental considerations are addressed through a sustainable plan, promoting informed decision-making for ecological conservation. Ethical commitments ensure equitable benefits, transparent communication, and data privacy protection. As Bangladesh navigates climate challenges, these insights pave the way for a resilient and sustainable future, emphasizing continuous improvement and collaboration.

6.3 Implication for Further Study

Refine Historical Data Dependency: Develop methodologies for continuous refinement in response to evolving climate patterns.

Improve Generalizability: Explore region-specific models or identify commonalities for broader applicability.

Address Data Availability Impact: Improve data quality in regions with limited or unreliable data. Explore alternative data sources or augmentation techniques.

Real-time Data Integration: Incorporate real-time data for dynamic calibration addressing climate change dynamics.

Geographic Expansion: Extend research to diverse regions for comprehensive model insights.

Socioeconomic Integration: Include demographic and economic factors for holistic flood risk perspective.

In-depth Ethical Analyses: Conduct stakeholder-driven analyses for ethical model deployment.

On-the-Ground Validation: Implement and refine proposed plans through pilot programs for practical integration.

These implications guide future studies in refining models, improving generalizability, addressing data challenges, and ensuring ethical considerations, ultimately advancing flood prediction accuracy and applicability.

Reference

[1] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.

[2] Ali, Jehad, et al. "Random forests and decision trees." *International Journal of Computer Science Issues (IJCSI)* 9.5 (2012): 272.

[3] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

[4] Freund, Yoav, and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences* 55.1 (1997): 119-139.

[5] Breiman, Leo. "Bagging predictors." *Machine learning* 24 (1996): 123-140.

[6] Breiman, Leo. "Arcing classifier (with discussion and a rejoinder by the author)." *The annals of statistics* 26.3 (1998): 801-849.

[7] Wolpert, David H., and William G. Macready. "An efficient method to estimate bagging's generalization error." *Machine Learning* 35 (1999): 41-55.

[8] Khosravi, Khabat, et al. "A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods." *Journal of Hydrology* 573 (2019): 311-323.

[9] Syeed, Miah Mohammad Asif, et al. "Flood prediction using machine learning models." *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 2022.

[10] Islam, Abu Reza Md Towfiqul, et al. "Flood susceptibility modelling using advanced ensemble machine learning models." *Geoscience Frontiers* 12.3 (2021): 101075.

[11] Choubin, Bahram, et al. "An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines." *Science of the Total Environment* 651 (2019): 2087-2096.

[12] Tehrany, Mahyat Shafapour, Biswajeet Pradhan, and Mustafa Neamah Jebur. "Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method." *Stochastic environmental research and risk assessment* 29 (2015): 1149-1165.

[13] Abu El-Magd, Sherif Ahmed, Ali Maged, and Hassan I. Farhat. "Hybrid-based Bayesian algorithm and hydrologic indices for flash flood vulnerability assessment in coastal regions: machine learning, risk

prediction, and environmental impact." *Environmental Science and Pollution Research* 29.38 (2022): 57345-57356.

[14] Motta, Marcel, Miguel de Castro Neto, and Pedro Sarmento. "A mixed approach for urban flood prediction using Machine Learning and GIS." *International journal of disaster risk reduction* 56 (2021): 102154.

[15] Nti, Isaac Kofi, et al. "Enhancing flood prediction using ensemble and deep learning techniques." *2021 22nd International Arab Conference on Information Technology (ACIT)*. IEEE, 2021.

[16] Ganguly, Kishan Kumar, Nadia Nahar, and BM Mainul Hossain. "A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh." *International journal of disaster risk reduction* 34 (2019): 283-294.

[17] Gauhar, Noushin, Sunanda Das, and Khadiza Sarwar Moury. "Prediction of flood in Bangladesh using K-nearest neighbors algorithm." *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2021.

[18] Rahman, Mahfuzur, et al. "Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis." *Earth Systems and Environment* 3 (2019): 585-601

[19] Géron, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[20] Bell, Jason. *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons, 2020.

Plagiarism report :

3390_3391

ORIGINALITY REPORT

25%	19%	12%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	2%
3	Submitted to Democritus University Student Paper	1%
4	www.researchgate.net Internet Source	1%
5	Submitted to University of Lancaster Student Paper	1%
6	Noushin Gauhar, Sunanda Das, Khadiza Sarwar Moury. "Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm", 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021 Publication	1%
7	www.mdpi.com Internet Source	1%