Multi-Task Learning for Flood Prediction: Joint Classification and Regression Using Hybrid CNN-BiLSTM Networks with Feature Gate Mechanisms

by

Md. Mostafijur Rahman
22101721
Farhan Ishrak
22101766
Md. Sazid Faiaz
22101309
Rubiya Karim
22101705

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
October 2025.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Md. Mostafijur Rahman
22101721

Farhan Ishrak
22101766

Md. Sazid Faiaz
22101309

Rubiya Karim
22101705

Approval

The thesis titled "Multi-Task Learning for Flood Prediction: Joint Classification and Regression Using Hybrid CNN-BiLSTM Networks with Feature Gate Mechanisms" submitted by

1. Md. Mostafijur Rahman (22101721)
2. Farhan Ishrak (22101766)
3. Md. Sazid Faiaz (22101309)
4. Rubiya Karim (22101705)

of Summer 2025 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of B.Sc. in Computer Science.

Examining Committee:

Supervisor:
(Member)

____

Rafeed Rahman
Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Co-Supervisor:
(Member)

____

Dibyo Fabian Dofadar

Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD

The Chairperson
Department of Computer Science and Engineering
BRAC University

Abstract

One of the most important issues in disaster management, especially in flooding prone areas such as Bangladesh, is flood prediction. In this thesis, a new multi-task learning method, which integrates convolutional neural networks (CNN) with bidirectional long short-term memory (BiLSTM) networks with attention mechanisms and feature gates, is proposed to perform a full-flood prediction. We use a hybrid architecture to conduct the binary flood classification and continuous meteorological index regression on a history of 65 years of Bangladesh weather (1948-2013). We have created a new CNN-BiLSTM hybrid system that operates on 12-month temporal sequence of 21 engineered weather records at 35 meteorological stations. The architecture includes feature gate mechanism of automatic feature selection, station embeddings of spatial context encoding, attention mechanism of temporal focusing, multi-task learning with joint classification and regression tasks, and sophisticated regularization with SMOTE oversampling and focal loss of class imbalance. We carried out a comprehensive comparative study on seven baseline models including classical machine learning (Random Forest, XGBoost, LightGBM), time series and deep learning (CNN, RNN, LSTM, BiLSTM) models. The optimized CNN-BiLSTM hybrid model gave a state-of-the-art performance of 0.6314 ROC-AUC to classify floods, which is a 11-percent improvement over the worst-performing baseline and demonstrates superiority over all other competing models. The model showed strong results in imbalanced data (7.4% positive class rate) and ensured the interpretability of the model using the feature gates and attention weights. The study contributes to the development of flood forecasting techniques by showing that learners in hybrid neural networks can be used to learn intricate weather-flood dependencies, which can be further used to establish a comprehensive meteorological risk forecasting with direct effects on early warning systems in Bangladesh.

Keywords: Flood prediction, Deep learning, Convolutional neural networks, Multi-task Learning, CNN-BiLSTM, Feature Engineering, Attention Mechanism, Feature Gating.

Table of Contents

Chapter 1

Introduction

1.1 Background

Floods are among the most devastating natural disasters, causing widespread damage to infrastructure, displacing communities, and disrupting ecosystems. Their increasing frequency and intensity are closely linked to two major global trends: climate change and urbanization. Climate change contributes to heavier and more unpredictable rainfall, while urbanization replaces natural landscapes with impervious surfaces such as roads, pavements, and buildings. These surfaces prevent rainwater from being absorbed into the ground, resulting in excessive surface runoff. During intense storm events, this runoff accumulates rapidly, often leading to flash floods—sudden and severe flooding that occurs with minimal warning. The high density of population and infrastructure in many regions amplifies the damage caused by these floods. The speed at which flash floods develop leaves little time for authorities to issue alerts and for residents to evacuate safely.

To address this urgent challenge, this research proposes a deep learning–based flood prediction framework that ingests multivariate meteorological and hydrological time-series rather than images. Using a hybrid CNN–BiLSTM architecture with learnable feature-gate mechanisms, the model jointly performs (i) flood-likelihood classification and (ii) continuous severity regression (e.g., BHMI). Convolutions extract local temporal motifs while BiLSTMs capture long-range dependencies; gates adaptively weight informative predictors such as rainfall, temperature, humidity, and river stage. This multi-task design improves recall–precision trade-offs under class imbalance and yields interpretable attributions via attention/gating weights. Integrated with gauge networks and reanalysis feeds, it supports real-time early warning and faster response—enhancing disaster preparedness without any reliance on computer vision.

1.2 Rational of the Study or Motivation

In today's world, climate change and natural disasters pose an unprecedented challenge to our planet. Natural disasters directly impact ecosystems, human health, and global economies. A flood is one of the most impactful natural disasters. As global temperatures have increased day by day, global warming has become a headache for environmentalists. Floods, wildfires, and droughts have become more frequent.

The effects of climate change are threatening the lives and livelihoods of millions. Despite ongoing efforts to manage flood risks and current approaches to predicting and mitigating their impacts often struggle to handle the vast, complex, and rapidly changing data required for accurate and timely decision-making. These limitations create a gap between identifying potential flood threats and executing real-time responses.This research is driven by the need to close the early-warning gap by leveraging advanced deep learning on multivariate meteorological and hydrological time-series to support effective, timely flood mitigation. Rather than relying on images, we study sequence-models—convolutional neural networks (CNNs) for local temporal motifs and recurrent/bi-LSTM layers for long-range dependencies—augmented with learnable feature-gating and attention. Building on these components, we construct a hybrid, multi-task model that simultaneously estimates flood likelihood (classification) and a continuous severity index (regression), enabling precise, localized, near-real-time predictions from signals such as rainfall, temperature, humidity, river stage/discharge, and derived indices (e.g., BHMI). The goal is a data-driven decision support system that helps policymakers, city planners, and emergency managers trigger earlier interventions, allocate resources proactively, and reduce climate-related flood risk. While the work centers on floods, the framework is extensible to other hydro-climatic hazards (e.g., drought intensity or heatwave alerts) using appropriate predictor sets. Ultimately, this study contributes to climate resilience by strengthening anticipatory action for vulnerable communities and ecosystems.

1.3 Problem Statement

Climate change is a pressing global challenge with ramifications for sustainability and the welfare of ecosystems and communities.Floods are among the most destructive and frequent natural disasters worldwide, causing severe damage to infrastructure, displacing populations, and disrupting ecological balance. Their increasing frequency and unpredictability—driven by climate change and urbanization—have amplified the demand for reliable and real-time flood detection systems. Conventional flood monitoring methods, such as hydrological models and sensor-based networks, often suffer from limitations in data timeliness, spatial coverage, and scalability, rendering them insufficient for providing early warnings during rapidly developing flood events.
While Convolutional Neural Networks (CNNs) originated in computer vision, their 1D variants are effective for extracting local temporal patterns from multivariate meteorological and hydrological sequences. In flood prediction, the challenge is not pixels but heterogeneous, noisy signals (e.g., gauge gaps, sensor drift, reanalysis biases, and station-to-station variability) that can obscure meaningful precursors. Traditional single-task or threshold-based models often underperform when integrating diverse data sources due to incompatibilities in units, sampling frequency, and missingness. To address this, we use a hybrid CNN–BiLSTM architecture: 1D convolutions capture short-horizon motifs (bursts, lags), BiLSTMs model long-range dependencies and regime shifts, and learnable feature-gate/attention mechanisms down-weight corrupted or less informative inputs in real time. Robust preprocessing (temporal alignment, de-seasonalizing, normalization, and statistically grounded imputation), plus imbalance-aware training (SMOTE oversampling and Focal Loss),

further improves stability. Together, these choices enhance real-time performance and robustness under noisy, multi-source meteorological data—without relying on computer-vision imagery. Additionally, while deep learning models, such as Feedforward Neural Network (FNN), can efficiently process large datasets, their potential for overfitting is high, especially in cases with imminent training parameter concerns, including epoch and batch sizes that are not optimally configured[16]. Overfitting can hinder a model's performance on new, unseen data, which is essential for accurate climate modeling over different regions in space and time. This creates a significant need for new methods that integrate both deep learning and computer vision techniques[13]. These technologies also align well with cloud computing, which enables scalable ingestion and processing of large, heterogeneous meteorological and hydrological datasets in near real time. By leveraging distributed data pipelines (for streaming gauges, reanalysis feeds, and station archives), we can train and deploy hybrid CNN–BiLSTM models with feature gating at scale, improving the accuracy and timeliness of flood risk estimation.

To address these challenges, this research proposes a hybrid CNN-BiLSTM deep learning architecture specifically designed for weather-based flood prediction. The system combines the spatial pattern recognition capabilities of Convolutional Neural Networks with the temporal modeling strengths of Bidirectional Long Short-Term Memory networks. Advanced techniques including attention mechanisms, feature gates, multi-task learning, and sophisticated regularization methods are incorporated to handle class imbalance while ensuring temporal causality and model interpretability.
The proposed approach aims to deliver accurate, reliable, and interpretable flood predictions that enhance early warning capabilities and contribute to better disaster preparedness in flood-prone regions.

1.4 Objective

The primary objective of this research is to develop a sophisticated weather-based flood prediction system using hybrid deep learning techniques to enhance early warning capabilities and disaster preparedness. The specific objectives are:

- **Hybrid Architecture Development:** Design and implement a novel CNN-BiLSTM deep learning architecture that combines convolutional spatial pattern recognition with bidirectional temporal modeling to capture complex meteorological relationships for superior flood prediction accuracy.

- **Advanced Feature Engineering:** Create comprehensive meteorological features from 65 years of Bangladesh weather data (1948-2013) including weak supervision labels (FloodLike binary and BHMI continuous indices), temporal rolling windows, climatological anomalies, seasonal encoding, and spatial station embeddings.

- **Multi-Task Learning Implementation:** Develop joint classification and regression capabilities where the model simultaneously predicts binary flood occurrence and continuous meteorological risk indices, enhanced with attention mechanisms for temporal focus and automatic feature selection through learnable gates.

- **Class Imbalance Solutions:** Address severe data imbalance (7% positive flood cases) through advanced techniques including SMOTE oversampling, focal loss optimization, weighted sampling, and sophisticated regularization while maintaining strict temporal causality to prevent data leakage.

- **Model Interpretability and Validation:** Provide interpretable predictions through attention weight analysis and feature gate examination to understand which weather patterns contribute most to flood risk, coupled with comprehensive evaluation against multiple baseline approaches including traditional machine learning and deep learning models.

- **Early Warning System Enhancement:** Deliver a scalable, accurate, and reliable flood prediction system that can be integrated into existing disaster management frameworks to improve early warning capabilities and support timely response decisions in flood-prone regions.

Through achieving these objectives, this research aims to advance the state-of-the-art in meteorological flood prediction, contributing to global climate resilience efforts and providing practical tools for disaster risk reduction in vulnerable communities.

1.5 Methodology in Brief

The proposed research adopts a hybrid deep learning approach integrating data preprocessing, feature engineering, and multitask model construction to enhance spatiotemporal prediction accuracy. The process begins with comprehensive data collection from multiple meteorological and hydrological sources, including rainfall, temperature, humidity, and river discharge records across diverse regions and seasons. This ensures adequate temporal and spatial variation for effective model generalization.

The collected datasets were rigorously preprocessed to ensure data integrity and numerical consistency. Data cleaning removed missing, redundant, and erroneous entries, while temporal splitting preserved chronological order to prevent information leakage between training and testing phases. Normalization and z-score transformation were applied to scale features within a common range, stabilizing gradient propagation and improving convergence during neural network training.

Feature engineering was then employed to enhance data representation and reveal temporal dependencies. Weak supervision was used to generate pseudo-labels where labeled data were limited, improving the learning process. Rolling statistical features—such as moving averages, standard deviations, and lag values—were extracted to capture short- and long-term trends, while anomaly detection based on statistical thresholds and z-scores identified irregular patterns caused by measurement or recording errors. These techniques collectively improved input reliability and accelerated model convergence.

The hybrid model architecture combines both spatial and temporal learning mechanisms. It integrates Feature Gates, 1D Convolutional Neural Networks (1D-CNNs),

and Bidirectional Long Short-Term Memory (BiLSTM) layers enhanced with an Attention Mechanism. Feature Gates dynamically regulate the influence of each input variable, suppressing noisy or irrelevant signals. The 1D-CNN captures localized spatial-temporal structures, while the BiLSTM learns bidirectional temporal dependencies across sequences. The Attention Mechanism prioritizes critical timesteps, allowing the model to focus on the most influential features. The latent representation from these layers is passed through fully connected networks to generate both classification (e.g., flood occurrence) and regression (e.g., water-level estimation) outputs, enabling multitask learning within a single framework.
To address data imbalance and optimize training, several strategies were applied. The Synthetic Minority Oversampling Technique (SMOTE) was used to balance class distributions, while Focal Loss emphasized difficult-to-classify samples. The Adam Optimizer adjusted learning rates adaptively, ensuring faster convergence and stable parameter updates. Hyperparameters such as learning rate, dropout ratio, and batch size were tuned empirically to achieve optimal generalization performance.
Model performance was evaluated using multiple metrics: Receiver Operating Characteristic–Area Under Curve (ROC-AUC) and Precision–Recall AUC (PR-AUC) measured classification discrimination, while F1 Score quantified the balance between precision and recall. Balanced Accuracy was included to reduce bias toward majority classes. Collectively, these metrics ensured a robust and fair assessment of predictive capability.
In summary, the methodology integrates advanced preprocessing, engineered features, and a hybrid spatiotemporal deep learning architecture within a unified data-driven pipeline. The framework delivers accurate, generalizable, and interpretable predictions, making it highly applicable to real-world environmental and hydrological forecasting scenarios.

1.6 Scopes and Challenges

Our proposed hybrid deep learning model offers a wide scope for advancing spatiotemporal prediction and environmental forecasting. By integrating Feature Gates, 1D-CNN, BiLSTM, and Attention Mechanisms, the model effectively captures both spatial and temporal dependencies within complex hydrometeorological datasets. This enables accurate prediction of flood events, rainfall intensity, and water-level fluctuations while maintaining adaptability to different climatic and geographical conditions. The multitask structure—combining classification and regression—enhances the framework's flexibility, allowing simultaneous detection of event occurrence and estimation of continuous magnitudes. Furthermore, the modular design of our model can be extended to other domains, such as drought analysis, groundwater level prediction, and real-time early warning systems, making it a scalable solution for a range of environmental applications.

Despite these promising scopes, several challenges remain. The primary limitation lies in the dependency on large, high-quality datasets, which are often incomplete or inconsistent in developing regions. Variations in spatial and temporal resolution may introduce noise, affecting model generalization across diverse catchments. The hybrid architecture, although powerful, is computationally intensive, requiring high-performance GPUs and significant training time. Class imbalance continues to pose a challenge, as rare extreme events like flash floods are difficult to model even with SMOTE and Focal Loss. Additionally, the interpretability of deep learning models remains limited; despite incorporating attention mechanisms, the system can still behave as a “black box,” making it difficult for policymakers to fully trust its predictions.

Future research should focus on integrating satellite and IoT-based data, adopting explainable AI methods, and exploring transfer learning to improve cross-regional adaptability. Overall, our proposed hybrid model presents a strong foundation for intelligent environmental forecasting while highlighting the need for further work on data quality, interpretability, and computational efficiency.

Chapter 2

Literature Review

2.1 Preliminaries

Climate change is a complex natural and anthropogenic phenomenon with causes that include global temperature increases, extreme weather events, and environmental destruction. Numerous models and computational techniques have been developed to analyze and predict changes over time. One popular method for predicting future scenarios is using climate models (general circulation models, GCMs) to simulate interactions between the atmosphere, oceans, and land surfaces. ML and DL have become powerful tools for climate analysis, providing alternative data-driven methodologies alongside physics-based models. Meteorological data-based flood prediction leverages long-term weather observations to identify patterns associated with flood-like conditions. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), are well-suited for analyzing temporal climate patterns, including temperature changes, precipitation trends, relative humidity, and sunshine duration. Feature engineering techniques further improve the quality of input data by selecting relevant climate indicators for predictive modeling through rolling averages, anomaly detection, and seasonality encoding. Multi-task learning frameworks enable simultaneous classification and regression tasks, providing both flood probability and continuous flood intensity estimates from the same meteorological input. Deep learning architectures combining Convolutional Neural Networks (CNNs) for local pattern extraction and Bidirectional LSTMs for temporal sequence modeling have shown superior performance in weather-based prediction tasks.

This thesis incorporates sophisticated time series analysis and hybrid CNN-BiLSTM deep learning approaches to develop a multi-task model for flood prediction, utilizing 65 years of Bangladesh meteorological data (1948-2013). In this study, we combine these advanced computational techniques with feature gate mechanisms to enhance the prediction of flood conditions and develop data-driven early warning systems.

2.2 Review of Existing Research

The investigation conducted by Gong et al. is another evidence of accurate weather prediction due to some kind of climate change impacts on urban planning, agriculture and energy management, in [16] the new CNN-LSTM hybrid model is used to forecast some weather parameters and its goal for improving forecasting. Long Short-Term Memory(LSTM) has had very successful implementations in learning long term dependencies in time series data and preparing it for forecasting but in dealing with intricate local patterns they are not always a good fit as well as too much noise in the input can hinder performance. Although Convolutional Neural Networks(CNNs) are great for extracting spatial features from structured data, it lacks the skill of capturing temporal relationships effectively. To tackle these limitations, the researchers merged CNNs and LSTMs with each other in a hybrid model making them ideal for weather forecasting. The hybrid model also surpassed traditional approaches like linear regression, standalone CNN, and standalone LSTM in terms of Mean Absolute Error(MAE), variance and R2 score, according to the research. This hybrid approach gave us the best out come with a variance of 0.927, the R2 score of 0.901 and an MAE of 0.901 which really demonstrates the power here in not just the extraction of spatial complexities but also the temporal complexities of the temperature data. This model needs to converge quickly when we suddenly drop the MAE during training and settles at 0.9, which indicates that it has learned. The prediction curve was consistent with the test data, which states the reliability of the model. In addition, CNN successfully extracted local properties from spatial data and LSTM managed long-term dependencies in sequential temperature patterns. This study not only highlights the potential for hybrid deep learning models in applications ranging from climate-relevant agriculture, energy management, and urban planning but also lays the groundwork for future research in both climate change mitigation and climate prediction.

According to Li et al. (2024) studied how to combine physics-based hydrological-hydraulic models and data-driven neural predictions into a mixed-data LSTM model to improve the accuracy and stability of the flood predictions. The study counters the drawbacks of classical LSTM models, which can only use the observed data and tend to have small sample sizes and overfitting. The authors combine the NAM rainfall-runoff model and MIKE 11 HD hydraulic model to create simulated discharge data to augment low measurements to create mixed training datasets with differing simulated to measured ratios[19]. The experiment conducted across the Jinhua River Basin reveals that a ratio of 2: 1 between simulated and measured data gives optimal performance with significant improvement in mean absolute error (MAE) and root mean square error (RMSE) with relative error of prediction (PRE) of less than 10 percent with one to six-hour lead time. Models that are trained with increased proportions of simulated data (e.g., 5:1) exhibit lower levels of generalization and increased uncertainty, indicating that there could be an overreliance on synthetic samples, which could be leading to a distortion of the relationship between temporal rainfall and runoff. The results point to the synergistic usefulness of simulation-enhanced learning through which physical models give data continuation and LSTMs identify nonlinear temporal trends. Nevertheless, the method is also very sensitive to the accuracy of simulations and basin-specific calibration, which restricts its expansion to more complicated terrains such as Shengzhou. The conclusion of the study is that balancing data composition, regularization, and cross-basin validation is important to scale hybrid LSTM frameworks to real-time flood forecasting systems.

In a recent study entitled Floodborne Objects Type Recognition Using Computer Vision to Mitigate Blockage-Originated Floods, the authors attempt to develop a computer vision means of automation of the detection of floodborne objects, which have been shown to greatly affect the change in the hydrodynamic behavior in the case of floods[9]. Because of their nonlinear, unpredictable kinematics, such objects are not always considered in traditional flood models.The Faster R-CNN model performed the best, and its mean Average Precision (mAP) was 84%. The paper has shown the potential of incorporating computer vision to flood management systems to facilitate real-time monitoring and event mitigation and thus can be used as a starting point in further studies on managing flood risk automatically.

Another study by Wilson et al.[7] showed different species, life stage and gender respond with different body size responses to temperature. This study investigates these responses in butterflies using Mothra, a deep-learning–based computer vision technique for high-throughput phenotypic analysis. For this purpose, the thermodynamic mass response is a crucial mechanism for explaining biotic shifts that occur with climate change. As a result, ectotherms developed at higher temperatures probably exhibited smaller body sizes as adults according to the "temperature-size rule". But investigations have revealed that it is a rule to be violated between species. Using Mothra, a python based application, over 180,000 digitized British butterfly specimens were processed for features such as forewing length and orientation and sex to near perfection in accuracy as compared to hand measurements. Correlation was found 0.99. Intensive studies on butterfly abundance and size had shown that warmer temperature increased butterfly body size during mid to late larval stage which was in agreement with the "temperature-size rule" for most of the species. But responses vary according to all sorts of variables: stage, sex, generation. The late larval stage has the steepest growth (0.69 percentage per Celsius on average) when temperature increases. The early larval and pupal stages behaved differently, however. Evidence of Sexual Size Dimorphism (SSD) was identified, with females being larger in 23 of 32 species, in accordance with insect ecological patterns. Family-level analysis found narrowly distributed temperature-size responses, with the highest variance found in Lycaenidae. The results illustrate the complexity of temperature-size responses and indicate they are influenced by environmental, evolutionary and ecological factors.

In the paper [13], they use a combined prediction model based on Empirical Mode Decomposition(EMD) and Relevance vector Machine(RVM) to predict climate change overtime(from 2019 to 2043). The combination of EMD and RVM is a better way to handle climate change predictions. Here EMD helps to decompose climate related data into intrinsic mode fiction(IMFs). Another RVM is to predict watch components with more comprehensive and reliable prediction. But this paper[13] focuses on only two factors: CO2 concentration and natural disasters. Also this model predicts climate change based on previous historical data from 1998 to 2008. In recent times there has been a big change in climate change, in article [24] what difference todays and previous global warming in earth history. Another article [23], shows the United states climate change from 1991 to 2020, there highlights the rising average temperatures, regional variations in warming trends, and changes in precipitation patterns. While temperature increases are clear, precipitation trends are more variable, influenced by natural and human factors.

V.Nandhini have done a research paper [4] on predictive analysis, to remove climate change and its impact on seasonal diseases,Using CUSUM(cumulative sum) algorithm. There are various algorithms used, such as big data techniques like binary segmentation to analyze climate changes and their association with various diseases. This paper could be broadened out in the impact of climate in other aspects, not just diseases. Then integrate more diverse predictive model or hybrid model and cross domain datasets for border scoops. A deep learning paper[17] for the use of RNNs and CNNs to predict climate change impacts This model is a standard deep learning approach to forecast climate impacts. It argues that this technique performs well and proves that deep learning can directly handle large-scale and complex data sets in climate science. This study illustrates the extensibility of providing deep learning models across different climate change prediction applications, from pollution concentration forecasting to climate extremes predictions. Additionally, this study recognizes the potential of overfitting and bias in models due to how they interact with real-time climate data.

Based on Kim research, they predict the water level of flooding in Vietnam using hybrid model combining a hybrid model is used to predict the amount of water that will fall from the rain, The model was the combination of a rainfall-runoff model and a variety of AI-based models (Deep Neural Networks, Long Short-Term [18]. The paper uses different forms of calibration, including genetic algorithms and pattern search, to tune the model's parameter and improve its predictions of floods. However, the Monte Carlo simulation mitigated this," there is now a higher degree of robustness and confidence in the model performances. The use of a real, case study application in Vietnam serves as proof-of-concept and provides a framework that can be implemented for near real-time inundation forecasts, which would be highly useful for disaster preparedness and mitigation. There are also limitations of limited interpretability, data dependency and complexity. The fact that this model does not generalize means that, though promising for the Phan Rang river basin in Vietnam, it might have to be significantly adapted for downstream applications in other countries with different topographical and hydrological features. As evident from the experimental results of the hybrid model for predicting flood water levels, the Long Short-Term Memory (LSTM) model outperformed all the others, achieving correlation coefficients (CC) of 0.9883, 0.9694, and 0.9787 for flood events f, g, and h respectively[18]. A second paper[4], combined EMD-RVM model prediction accuracy is 97.63%. This is based on a validation process in which the model predicted the quantitative values of climate change for 2009-2018 and compared them to the actual values for the same period. The mean error of the prediction is 2.37%, and the error of some specific year is 1.19% and 3.26%. This demonstrated the ability of the model to accurately predict the trends associated with climate change. Such high prediction accuracy confirms the reliability of the model for predicting future climate change and lays a strong scientific foundation for solving problems of global climate change. Hybrids significantly increase performance and accuracy over previous methods in predicting climate change which helps mitigating climate change.

The study by Zhang et al. (2023) introduces a systematic exploration of the hybrid recurrent neural networks in flood prediction, aiming to improve the accuracy of their prediction as well as their computational performance by making use of lag-time preprocessing [14]. The paper contrasts seven deep learning models, including LSTM, GRU, CNN-LSTM, CNN-GRU, ConvLSTM, spatiotemporal-attention LSTM (STA-LSTM), and a new spatiotemporal-attention GRU (STA-GRU), with data in several hydrometric stations of the Credit River watershed in Canada. The STA-GRU model combines both spatial and temporal attention processes and lag-time correspondence between the upstream precipitation amounts and the downstream water level reaction to provide a realistic catchment delay during the rainfall-runoff processes. Findings indicate that lag-time preprocessing significantly enhanced performance, whereby the coefficient of determination (R 2) of the STA-GRU model improved to 0.9215 against 0.8125 in 12 and 24-hour forecasting, respectively, and the RMSE and MAE also decreased. It is also shown in the study that the faster training of the STA-GRU relative to STA-LSTM does not compromise the accuracy of the model, which supports the efficiency benefit of the simplified gating architecture of GRU. With the lag-time analysis, the model obtains better spatiotemporal features learning and long-range dependencies. On the whole, the study also demonstrates that a combination of attention and explicit lag-time preprocessing gives more interpretable, less volatile, and more efficient predictions, which makes STA-GRU a good potential candidate in the real-time flood early-warning system when both speed and precision are of paramount importance.

The study by Venkatesan and Krishnan (2025) indicates that smart cities in coastal areas require intelligent, privacy-sensitive, and scalable predictive models, as climatic vulnerability and decentralization of pivotal data and limitations of computing power continue to grow. In their paper, they present a hybrid CNN-LSTM model of Federated Learning (FL-CNN-LSTM) to improve the accuracy of flood prediction but maintain the privacy of the data and adaptability to a variety of coastal conditions. The CNNs, in this architecture, obtain the spatial flood patterns by looking at satellite and sensor images, and LSTMs obtain the temporal dependence in hydrometeorological time-series data. Through the Federated Learning (FL) integration, decentralized model training can be performed with local clients without transferring raw data with sensitive information to central servers, which helps to maintain the privacy of the information and minimize the reliance on centralized databases [22]. The model was able to predict the flooding events with an impressive 94.5 per cent accuracy using datasets like Sen12Flood (2018–2019), which used Sentinel-1 SAR and Sentinel-2 optical imagery, compared to prediction methods like ARIMA, Random Forest, ANN, and traditional CNN-LSTM scored lower in accuracy and generalization. The given system also proves to be less time-consuming (10.7 s) in training and highly scalable in various coastal environments. The results indicate the greater capacity of the FL-CNN-LSTM model to balance between accuracy, efficiency, and privacy, which provides a potential avenue for providing real-time flood warning and coastal resilience in smart cities. However effective, the authors observe that performance might be different when there are heterogeneous environmental conditions and propose that further combination of graph neural networks (GNNs), drone-based sensing, and data fusion in real-time will be required to create more predictive robustness and international adaptability.

A recent research by Ni Chenmin et al. (2024) suggested a hybrid deep learning solution to flood prediction potential that combines Variational Mode Decomposition (VMD), Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) networks with an attention mechanism [15]. The model meets the challenges of noisy hydrological data and complicated flood phenomena through the preprocessing of water level signals as intrinsic mode functions (IMFs) by the VMD algorithm and feature selection according to the correlation of the Pearson factor and fluctuation factor. Each IMF is next forecast with either BiLSTM or CNN-LSTM-attention, based on volatility. It managed to outperform the conventional LSTM, BiLSTM, and CNN+BiLSTM models when tested on the datasets of Hankou station in Wuhan, China. The design of the VMD-F hybrid model was able to reduce RMSE by 26-47% and increase R^2 by more than 10%, acknowledging its stability and competitive predictive performance in both short- and long-term flood scenarios. This paper notes the usefulness of the integration of signal decomposition, spatial-temporal deep learning models, and attention mechanisms to highly predictive flood prediction and disaster preparedness.

In the paper need for improved monitoring of climate change through heterogeneous integration of remote sensing, deep learning, and computer vision is addressed by Gupta et al. [8]. They use high-resolution satellite imagery and convolutional neural networks (CNNs) for accurate image detection and the assessment of water body degradation. This study covers Bellandur Lake, which established an 83% shrinkage in its water area between 2001 and 2020, using pixel-based classification techniques, primarily the HSV (Hue, Saturation, Value) color space, to detect water. In line with previous works by Dörnhöfer et al. who used the remote sensing for water body analysis and Zhu et al. which used CNN to help road detection in satellite images. Automation offers several advantages concerning environmental monitoring, where scalable and accurate ecological assessments could be made, as pointed out by the authors in this paper. On the other side, there are limitations like the need for historical satellite imagery only allows for retrospective analysis and model tuning to make it generalizable across large spaces have its limitations too. The authors propose to incorporate further real-time environmental inputs, such as precipitation patterns, pollution levels and humidity, to improve accuracy in forecasting. Furthermore, advanced deep learning models could also be adopted to better adapt environmental systems. In summary, this research highlights the prospective capability of Internet of Things (IoT) based AI-enabled remote sensing applications on climate change monitoring, focusing on viable urbanization and vegetation recovery strategies. Improvements include expanded datasets, increased accuracy from models, and accounting for several environmental variables to provide a more holistic model.

The Ramachandra et al. [5] presented a causal analysis framework for quantifying deforestation phenomenon in the Amazon rain forest as induced by Brazil's 1994 economic transition. Using Bayesian causal inference, the research evaluates the effects of management interventions, such as urbanization as well as natural (Change point package in R 2016, satellite image time series). The results show that if in the case of economic growth and frontier advancement (Hyperinflation in Brazil), the deforestation of forests after 1994 is 25% less. The method incorporates image processing as well as HSV masking and U-Net-based semantic segmentation for accurate classification of forest, and uses the Jaccard index (Jaccard similarity index for image segmentation) for evaluation. Time series analysis used change point detection (Bayesian structural time series models in R: Causal Impact package, 2015) to identify statistical shifts in deforestation trends, correctly identifying the year of the critical intervention in 1994. Comparison with counterfactual projections highlights the missing counterfactual scenarios and emphasizes the extent of human-mediated deforestation through simulations of scenarios in which the intervention was not implemented. Nonetheless, due to the limited number of regions it has been trained on and the requirement of labeled data (Label Me annotation tool), it faces challenges in generalization. This raises questions particularly when one of the study's conclusions is that "the end of hyperinflation was a necessary condition for this to happen" as this simplifies complex socio-economic reality into one causative intervention. Suggestions for enhancing the networking of satellite data include: synthesising raw satellite imagery into satellite feed in real space time; post receiving satellite feeds at the forest site, automating preprocessing steps; and combining multi-factor causal analysis to improve predictions (Explosive deforestation in Brazil; Philip M. Fearnside and Enéas Salati, 1985). In general, the study provides a strong AI-driven solution for environmental monitoring but needs improvements to work for more regions.

Moishin and Chand (2021) designed a hybrid Convolutional Long Short-Term Memory (ConvLSTM) flood forecasting model that incorporates the spatial feature of CNNs with the time sequence learning of LSTMs to enhance the early flood warning system in the Fiji Islands, which is a highly prone area to flood disasters. The research was also innovative in its application of the Flood Index (IF), a standardized index of efficient precipitation, to model the progressive loss and gain of the water resources through the use of daily rainfall data in nine stations gathered throughout 30 years (1990 2019) [6]. ConvLSTM was built as a multi-input, multi-stage, multi-output network, which predicts daily values of IF at a forecast horizon of 1-, 3-, 7-, and 14-days. The performance analysis with the help of multiple statistical indicators, like Root Mean Squared Error (RMSE), LegateMcCabe Efficiency (LME), NashSutcliffe Efficiency (NSE), and correlation coefficients, ensured that ConvLSTM provided a high performance in comparison to the benchmark models (LSTM, CNN-LSTM, and Support Vector Regression (SVR)). As an illustration, RMSE values were 0.101, 0.150, 0.211, and 0.279 when using the four forecast horizons, and LME values were strong (0.939-0.726), which proves its strength and reliability even in the long-term forecast. The model not only converged fast (within five epochs) but also has good accuracy ($r = 0.93$; $NSE = 0.85$), which indicates that this model applies to real-time tasks in data-sparse settings. Although it can only use two input features (rainfall and lagged IF) to reduce the associated costs, the study notes the cost-effectiveness, scalability, and operational importance of ConvLSTM in preventing early floods and suggests increasing the framework with more hydrological

variables to apply to a wider area.
The recently developed literature suggests the recent maturity of deep learning-based flood forecasting techniques, though with specific focus on hybrid and recurrent structures, which increase predictive accuracy, interpretability, and operational efficiency. Rumee (2025) compared FNN, RNN, LSTM, and GRU models with Bangladesh meteorological and hydrological data and found that the GRU model had the best accuracy (98) and F1-score (99), due to its low-computational cost and ability to use data with low density in tropical areas. Elaborating on the independent recurrent models, the literature of the world shows the effectiveness of hybrid structures, in particular CNN-BiLSTM, CNN-GRU, and ConvLSTM, incorporating both spatial and temporal learning [21]. As an example, Chand et al. (2025) showed that a CNNGRU model can perform near-perfect correlation (r = 0.9960.999) when used to predict rainfall-runoff dynamics over the long term in the case of Fiji, whereas Moishin and Chand (2021) have verified that ConvLSTM can predict long-term rainfallrunoff dynamics with only a few input variables. To solve the model transparency, Huang et al. (2024) integrated a SHAP explainability layer into a CNNLSTM model, which enhanced the model interpretability and attained an NSE of 0.838. Li et al. (2024) combined data-driven and physics-based models, and the results of the work demonstrated that the optimal ratio between simulated and observed data was 2:1, as it demonstrated the most stable performance of the simulations between NAM and MIKE 11 and LSTM networks. In like manner, Zhang et al. (2023) proposed a Spatiotemporal-Attention GRU (STA-GRU) in which the lag-time preprocessing and attention mechanism increased the R 2 values and decreased the time of training. Political work by Venkatesan and Krishnan (2025) on Federated Learning CNNLSTM models also makes the privacy-preserving, decentralized forecasting of floods on smart coast cities possible. Together, these results indicate that attention-based, explainable, and hybrid models are 20-47 percent more accurate in prediction and enhance model interpretability and stability compared to the traditional deep learning methods. Nevertheless, there are still issues of computational cost, hyperparameter sensitivity and reliance on high-quality climate data, and physics-guided modeling, multi-source data fusion and cross-basin validation to produce globally robust flood early-warning systems.

Another study by Donti et al. [12] examined the role of machine learning (ML) to mitigate and adapt to climate change by optimizing critical sectors such as energy, transportation, and agriculture. It places ML applications into supervised learning for energy forecasting, reinforcement learning for grid balancing and computer vision for emissions monitoring. ML has been shown to provide possible solutions, though, with the potential drawbacks of high-energy use, scale, and ethical implications. The paper highlights the importance of collaboration across sectors, open data-sharing, and energy-efficient ML models to increase the uptake of ML in climate action. In the end, ML, combined with policy and interdisciplinary expertise can help make great strides in fighting climate change.

Study by Huang et al.[20] proposed SegNet deep learning model, accurately calculated river water level surpassing CIS methods. SegNet obtained low RMSE (0.013–0.066m) and was robust to environmental changes including lighting changes and typhoons. SegNet provides a good compromise between accuracy and efficiency, hence, is widely used compared to U-Net and Fully Convolutional Networks (FCNs). Notable results are the positive influence of larger, more diverse training datasets as it was proven to have an accuracy of 0.163m, improved to 0.028m with more data; however, the limits are that it does require consistent high-quality training data, it does require real-time training, and it does have a limit in its night time accuracy. A combination of infrared imaging, automatic data collection (e.g., drones), and the hybrid models for measuring the velocity and level of water were suggested to correct the performance also. The study demonstrates SegNet's potential as a hazardous flooding risk management and environmental monitoring tool, with future work planned to increase model generalization and set on more specific datasets.

Subsets of artificial intelligence like machine learning(ML) and deep learning(DL) are being used to address challenges like greenhouse gas emission and adaptation. ML mainly uses their algorithm to discover recognizable patterns and from these patterns they make predictions from large datasets. In the meantime, DL uses multi-layered neural networks to excel at complicated tasks like image recognition or time series analysis. A survey report conducted by Ladi et al. [11] proposed machine learning (ML) and deep learning (DL) methods are increasingly used in climate change research and among those methods the most dominant methods being artificial neural networks (ANNs), random forests, and convolutional neural networks (CNNs). CNNs are commonly used to evaluate satellite data for urban heat assessments, on the other hand random forests are used to estimate drought and flood hazards. In terms of mitigation, the main focus is land use/cover changes and energy use with ML/DL algorithms successfully predicting patterns and maximizing energy consumption. LSTMs and decision trees have been effectively used to visualize urban energy usage and assist climate-friendly planning. These applications showed the performance of ML/DL and optimized climate solutions.

2.3 Summary of Key Findings

This study also demonstrates the remarkable potential of hybrid deep learning architectures, particularly CNN–BiLSTM and related coupled models, for improving the precision, interpretability, and responsiveness of flood forecasting and meteorological prediction systems. The reviewed literature consistently shows that such hybrid models outperform standalone deep networks by effectively combining CNNs' spatial feature extraction with LSTM/BiLSTM or GRU units' temporal sequence learning, enabling more accurate modeling of nonlinear rainfall–runoff relationships and hydrological dependencies. Incorporating attention mechanisms—including spatial, channel, and self-attention—further enhances interpretability and dynamic feature prioritization, allowing the model to focus on critical meteorological drivers while filtering noise. Models such as the Spatiotemporal Attention GRU (STA-GRU) (Zhang et al., 2023) and ConvLSTM (Moishin Chand, 2021) achieve superior long-horizon stability, rapid convergence, and explainable predictions compared to conventional recurrent models. The inclusion of data augmentation and lag-time preprocessing methods strengthens generalization under sparse, imbalanced, and noisy hydrological datasets by expanding training diversity and aligning rainfall–runoff delays. Moreover, integrating explainability layers such as SHAP (Huang et al., 2024) and simulation-enhanced learning (Li et al., 2024) bridges physics-based and data-driven paradigms, ensuring both scientific validity and data efficiency. Advanced frameworks like Federated Learning (FL)-based CNN–LSTM architectures (Venkatesan Krishnan, 2025) extend this progress by supporting decentralized, privacy-preserving, and scalable flood forecasting across smart cities and distributed networks. Collectively, these advancements contribute to 20–47% performance improvements, reduced RMSE and MAE values, and heightened model reliability across multiple forecast horizons and climatic contexts. Nevertheless, challenges remain regarding computational demand, hyperparameter sensitivity, and dependency on high-quality meteorological and hydrological datasets. Future directions include integrating physics-guided neural layers, cross-basin validation, and multi-source data fusion to strengthen adaptability. Overall, the findings affirm that attention- and data-augmented CNN–BiLSTM architectures represent a robust pathway toward transparent, high-accuracy, and climate-resilient flood early-warning systems.

Chapter 3

Data Collection and Analysis

3.1 Final Specifications and Requirements

The technical, methodological, and computational resources needed in the proposed study were essential in order to develop a reliable model and assess its performance. The main aim was to develop a deep learning model that would predict the existence of flood-like conditions and estimate long-term meteorological risk of floods with high accuracy and precision. Based on this, the ultimate specifications and requirements were established in three broad areas, namely: methodological design, data resources, and computational environment.

Methodological Requirements
The study used a multi-task deep learning methodology, that is, convolutional and recurrent frameworks were used simultaneously, to extract spatial and temporal correlations in the meteorological data. The methodologically, the model had to contain a structured workflow, which included data preprocessing (data cleaning, data transformation, feature engineering), multi-task model construction (CNN-BiLSTM backbone with attention and feature gates), and evaluation (ROC-AUC, PR-AUC, F1-score, and BHMI regression). Strict temporal causality was also demanded in the design so that only historical information was employed to predict floods.

Data Requirements
The model used 65 years (1948-2013) of monthly meteorological observations of 35 stations in Bangladesh taken through the Bangladesh Meteorological Department (BMD) and NOAA reanalysis data sets. seven major variables were used in the dataset, which are: rainfall, temperature, humidity, sunshine duration, evaporation, wind speed and cloud coverage. Derived features like rolling means, anomalies and lagged differences were developed to provide better time context. There were two target variables used in the model, a binary flood-like indicator and a continuous Bangladesh Hydrometeorological Index (BHMI). Computational Requirements and Software.
Python (v3.10) was used to implement the models with significant libraries such as TensorFlow 2.x, Keras, NumPy, Pandas, scikit-learn, and Matplotlib to train, manipulate, and visualize data. The computational tests have been performed on a workstation with a GPU card (NVIDIA RTX 3060ti, 8 GB VRAM, 16 GB RAM, AMD Ryzen 5 processor) with an operating system of Windows 11. For models with

low complexity, the training time per-model was between 30-60 epochs, and early stopping was applied in order to optimize efficiency.

Evaluation and Validation Tools
The performance assessment demanded visual tools of metrics, like Matplotlib, Seaborn, and TensorBoard to view the loss curves, ROC and PR curves, and the confusion matrices. JSON-based configuration files and configuration checkpoints were used to ensure the model reproducibility and configuration management.
All in all, final specifications guaranteed that the system was fundamentally capable of meeting the computational and methodological criteria required to come up with a deep learning hybrid robust, interpretable, and reproducible model of predicting floods in Bangladesh.

3.2 Societal Impact

The given flood prediction framework has a very high possibility to impact the society positively as it will facilitate disaster preparedness and climate resilience in Bangladesh, which is among the most flood prone countries ever to be built in the world. The model can be used to provide early warning mechanisms to reduce loss of life, property damage, and economic disruption by utilizing long-term meteorological data and artificial intelligence. The enhanced forecasting also promotes public health and safety since it facilitates an early evacuation, water management and preventive diseases in the cases of floods. As a cultural element, the project fosters the use of data to make decisions and consciousness of the environmental risks to the vulnerable communities. Further, the transparent and explainable design of the model is designed to make AI usage ethical and responsible, which is in line with the national objectives of sustainable development and reduction of the disaster risks.

3.3 Environmental Impact

The presented Hybrid CNN-BiLSTM flood prediction architecture is beneficial to environmental sustainability, providing data-oriented management of climate risks and efficient organization of resources. The model assists in proactive water management, reduces agricultural losses, and assists in ecosystem preservation caused by frequent flooding by properly predicting flood occurrences. In contrast to the conventional hydrological simulations, which need lots of fieldwork and energy-consuming modeling, this method is based on digital meteorological data and computational efficiency and makes the project have the smallest carbon footprint. Another aspect that is supported by the research is the sustainable development goals because it builds environmental awareness, resilience to disaster, and climate adaptation strategies within the communities of Bangladesh. The system generally improves the sustainability of the environment by being insightful as opposed to being intrusive.

3.4 Ethical Issues

The ethical implications of this study are the integrity of data, transparency, and responsible use of AI. The entire meteorological data employed in this study have been obtained through authorized and publicly available datasets, which did not contradict institutional and governmental data-use policies. The study also ensures scientific honesty by ensuring that it is reproducible, traceable and transparent in data preprocessing, model training and evaluation. Because predictive models may affect social safety and policy-making, ethical conduct necessitates that the results of the model must be explainable and devoid of bias especially in such low-stakes scenarios as flood forecasting. The feature gate mechanisms and attention inclusion contribute to the increased explainability of the system which is consistent with the principles of accountable and trustful AI. The research does not misuse data and also, no personal or sensitive information is utilized in the study and also follows all recommended professional guidelines to ensure fairness, sustainability as well as the best interests of the society in any phase of designing the study.

3.5 Standards

This study is based on known data science and software engineering standards to guarantee quality, reliability and reproducibility. The research is conducted in accordance with the IEEE and ISO standards of research data management, such as their adequate documentation and traceability, version control during the stages of data preprocessing and model development. Coding utilizes the PEP 8 Python Coding Standard to ensure that the code is readable and consistent to implement. The evaluation methodology is in line with the machine learning best practices such as train-test separation, cross-validation, and performance benchmarking, which involves well-known metrics such as ROC-AUC and F1-score. Besides, the ethical standards that were aligned with the IEEE Code of Ethics were upheld, with transparency, fairness, and accountability being of high importance in the design of the AI systems. Together, these standards can guarantee the technical rigor (as well as professional responsibility) of the proposed flood prediction model.

3.6 Project Management Plan

The project management was carried out using a systematic plan that focused on efficiency in terms of time, the cost control, and utilizing the resources effectively. The overall life span of the project was about nine months, which was further separated into several stages: data acquisition and preprocessing (2 months), model design and development (3 months), training and validation (2 months), performance evaluation and visualization (1 month), and documentation and reporting (1 month). The production was done in an iterative development style where evaluation and improvement of the production occurred at every phase. Resource management was concerned with effective utilization of computational resources such as GPU-enabled devices and open-source software (TensorFlow, Keras, NumPy, Pandas), which reduces the financial expenditure. The approximate budget was based on data storage, availability of cloud computing (where necessary) and the cost of publication, which qualifies as a reasonable academic research budget. The project was well-tracked in terms of progress, scheduling was held in Gantt chart, and milestones-based reviews were held to make sure that the objectives were achieved within the resources available and on time. This systematic managerial style provided effective coordination of the technical, analytical and reporting activities and the research is of high quality and on time.

Figure 3.1 illustrates the detailed project schedule.

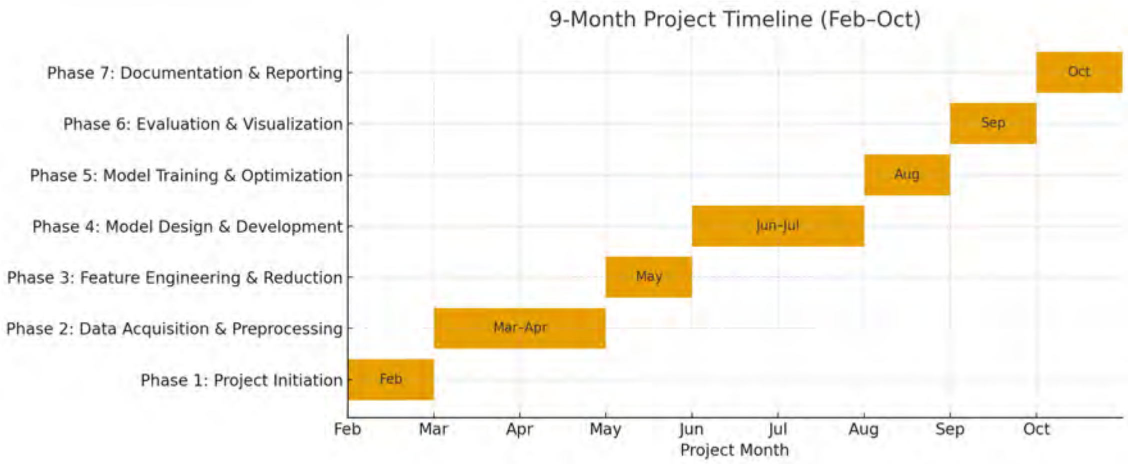

Figure 3.1: Nine-Month Project Timeline (February–October). The timeline outlines the sequential execution of seven major phases: Phase 1 – Project Initiation, Phase 2 – Data Acquisition and Preprocessing, Phase 3 – Feature Engineering and Reduction, Phase 4 – Model Design and Development, Phase 5 – Model Training and Optimization, Phase 6 – Evaluation and Visualization, and Phase 7 – Documentation and Reporting.

Each phase was carefully scheduled to ensure logical progression between data processing, model construction, training, and validation. The middle stages (June–August) were allocated to intensive model design and optimization, while the final month focused on documentation and presentation of results.

3.7 Risk Management

It was imperative that risk management was carried out effectively so that the project was successfully completed within the stipulated scope, time, and money. The major risks that were identified were data quality problems, overfitting of the model, and computation constraints and the imbalanced class representation. In order to counter them, a number of measures were taken. The threats to data like inconsistent records or missing data were mitigated by using a powerful data cleaning technique, cross-station validation, and interpolation. Gradient clipping, early stopping and dropout regularization reduced the threats of overfitting as well as training instability. To deal with computational bottlenecks, we optimized the batch sizes and minimized model parameters and used the accelerated training through the use of the GPU. The flood / non-flood imbalance was addressed with SMOTE oversampling and Focal Loss to promote the learning of minority classes. There was also model evaluation and version control on a periodic basis so that there was reproducibility and constant monitoring on performance. All these mitigation measures decreased the frequency of technical and operational risks, which made the project stable and reliable at any point of its development life cycle.

3.8 Economic Analysis

This project has been economically analyzed to focus on cost-efficiency, long-term benefits, and social value, and not financial profit. The overall approximate project cost was around 40,000 BDT, which comprised the computational resources, data management and documentation and publication costs. There was a significant decrease in the cost of software licensing because of the use of open-source software tools including Python, TensorFlow, and scikit-learn, and high-performance was ensured by the use of the GPU-based computation without needing a large cloud infrastructure. The benefits of the proposed flood prediction model include high economic and social returns as the model will enhance the capacity to have early warnings, reduce property damage, decrease agricultural losses, and facilitate disaster preparedness in flood-prone areas in Bangladesh. The ratio of cost to benefits is very positive because the low initial investment in reading and computing materials results in saving in the long run and resilience in flood management. Moreover, it is an economically viable and effective technological response to climate risk mitigation because of the scalability of the system, which can be integrated into government or community-based early warning systems at relatively low extra costs.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

The design procedure of the suggested Hybrid CNN-BilSTM multi-task learning framework was worked out to find a compromise between predictive accuracy, interpretability and computational efficiency of the weather-based flood forecasting.

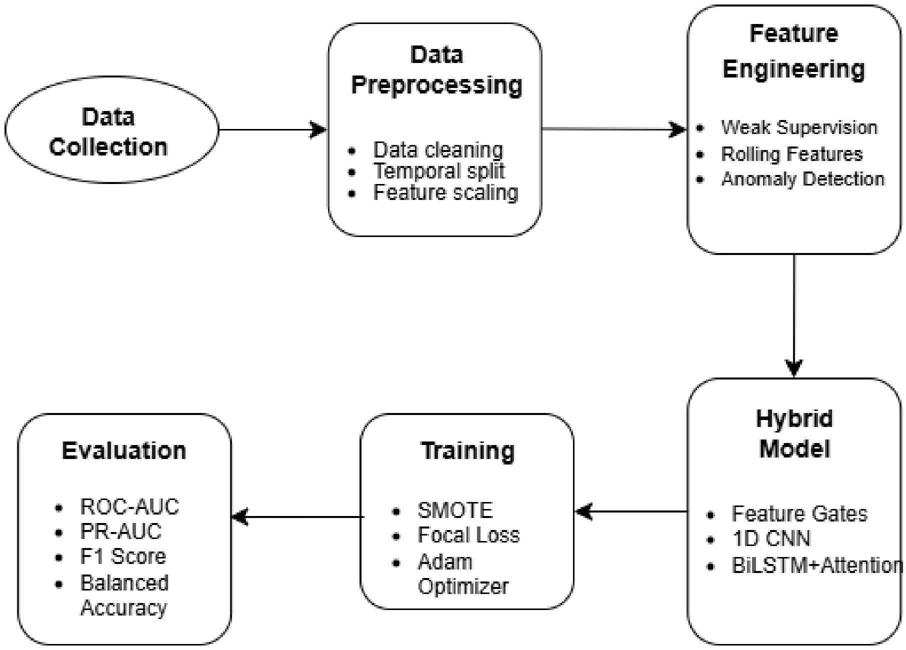

Figure 4.1: Workflow of the Proposed CNN–BiLSTM Hybrid Flood Prediction Model

This approach was designed around a systematic pipeline of converting past meteorological information to credible and explainable predictions of floods by a sequence of sequential and interdependent processes.

Phase 1: Data Collection
The model has used 65 years (1948-2013) meteorological records which were obtained through 35 weather stations in Bangladesh. These data will contain the main variables of climate like rainfall, temperature, humidity, cloud cover, duration

of sunshine, and speed of the wind. The dataset is used as the basis of the modeling of the conditions of the floods due to seasonal and long-term meteorological changes.

Phase 2: Data Preprocessing
Data preprocessing is done to make sure that raw meteorological data is converted into a standard and high quality format to be used as an input to the model. This step comprises data cleaning (to deal with missing data and outliers), temporal separation (to avoid leakage between training and test data) and feature scaling (to standardize variable values). Preprocessing was done in strict temporal causality, such that only the information available before the time of prediction can be used in the model, this is essential to realistic forecasting.

Phase 3: Feature Engineering
To increase the predictive representation, the dataset was enlarged with a few derived features that indicated the short-term and long-term weather patterns. There are three forms of engineered features: Weak Supervision: Binary flood-like indicators that were generated using rainfall thresholds as well as history of floods. Rolling Features: Implemented multi-window rolling mode (e.g., 3-month, 6-month) to take into account the delayed hydrological reactions. Anomaly Detection: Standardized deviations of the seasonal means that are calculated to identify abnormal climate behavior that includes unusual precipitation or high levels of humidity. These characteristics allow the model to acquire more complex time based dependency and environment-related predictors of flood-like action.

Phase 4: Model Development and Training
The proposed Hybrid CNN 1D Convolutional Neural Networks (CNN) are used to extract spatial-temporal patterns and LSTM layers with Attention are used to learn sequential dependences. Also, Feature Gate modules automatically scale the contribution of each meteorological variable and enhance interpretability and reduce overfitting. In the process of training, the model considers the following strategies of optimization, SMOTE oversampling to resolve the problem of class imbalance (7% flood observations)
Focal Loss ($\alpha = 0.25$, $\gamma = 2.0$) to pay attention to the minority samples that are hard to classify. Adaptive learning rate and gradient clipping on Adam Optimizer to guarantee stable convergence. All of these approaches contribute to the increased efficiency of learning and allow the model to be generalized with respect to various climatic patterns.

Phase 5: Evaluation and Validation
The last model was tested based on a set of extensive performance performance indices, such as ROC-AUC, PR-AUC, F1-score, and Balanced Accuracy, to estimate its performance in classification with class imbalance. Youden J statistic threshold optimization was then used to validate the best point of decision between flood and non-flood conditions.The interpretability of the model was confirmed by the visualization of attention-weight and feature-importance, which confirmed that the system is a validated explainable AI framework that can be used in early-warnings.
In summary, the proposed methodology follows a structured, data-driven process that transforms decades of weather data into actionable flood predictions through

advanced deep learning techniques. By combining CNN–BiLSTM hybrid modeling, multi-task learning, and explainable AI mechanisms, the design achieves both predictive precision and interpretability, establishing a robust foundation for flood risk forecasting in Bangladesh.

4.2 Preliminary Design

The development of the proposed Hybrid CNN–BiLSTM Multi-Task Learning model followed a structured design and simulation process aimed at identifying the most effective architecture for weather-based flood prediction in Bangladesh. To ensure optimal model performance under the project's objectives and constraints, multiple design alternatives were conceptualized, implemented, and empirically evaluated through simulation experiments.Each design iteration was assessed in terms of accuracy, efficiency, interpretability, and robustness to class imbalance.

4.2.1 Alternative Model Designs

Convolutional Neural Network (CNN) could be defined as one of the types of neural networks. A 1D CNN model was adopted to test the performance of the spatial pattern extraction of meteorological sequences. The convolutional layers were found to locate changes in rainfall and humidity well, with good precision and low recall (ROC-AUC 0.61, PR-AUC 0.19). Its inability to predict delayed hydrological effects was a result of the lack of temporal memory.

In this category, the tree models include gradient-boosted trees that are developed based on the principle that the error-reducing model is quantified, and the optimal path is chosen using the error measure. Gradient-Boosted Tree Models (LightGBM and XGBoost) in this group are based on the principle that the error that decreases the model is measured, and the best path is determined based on the error metric. In order to be able to offer non-sequential baselines, the tree-based ensemble methods, i.e., LightGBM and XGBoost, were trained with the same engineered features. Both approaches showed high baseline accuracy (ROC-AUC 0.60–0.61) and short training durations but failed to use the time-series continuity of meteorological data. In addition, their forecasts were not interpretable in the form of sequential dependencies or interactions between features with time.

The hybrid CNN-BiLSTM model suggested below represents the combination of the previously mentioned two models. This proposed Hybrid CNN-BiLSTM Model is the integration of the two aforementioned models. Following the assessment of the weaknesses of the baseline models, a hybrid deep learning architecture was created, which combines the strengths of the CNN and BiLSTM layers. The CNN element encodes the short-term time and spatial dependence in local weather patterns whereas the BiLSTM layers encode the long-term sequential dependence. Interpretability is improved with an attention mechanism that weights the most significant time steps and feature gate modules that are adaptively controlled to determine the significance of every meteorological variable. This design was further propagated to a multi-task structure with binary flood classification in combination with continuous Bangladesh Hydrometeorological Index (BHMI) regression, which gives more profound predictive information.

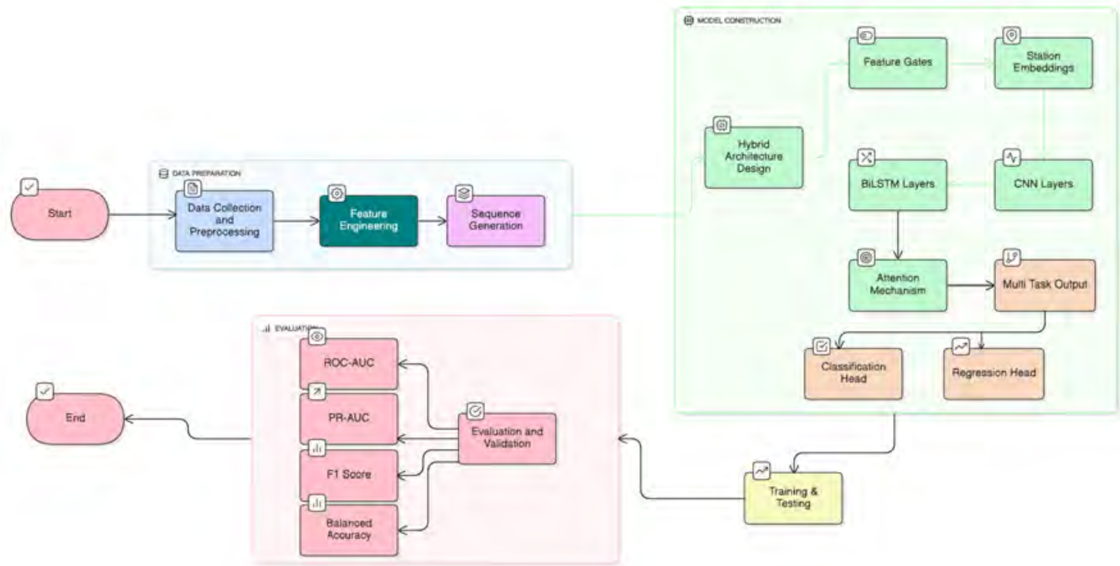

Figure 4.2: Complete Workflow of the Proposed Hybrid CNN-BiLSTM Model for Flood Prediction

The diagram in Figure 4.2 illustrates the end-to-end workflow of the proposed hybrid CNN–BiLSTM architecture.
The process begins with data collection and preprocessing, followed by feature engineering and sequence generation to capture temporal dependencies. The hybrid model construction integrates CNN layers for spatial feature extraction and BiLSTM layers with attention for temporal feature learning, enhanced through station embeddings and feature gates. The multi-task design enables simultaneous flood classification and BHMI regression, trained using balanced optimization strategies. Finally, evaluation metrics such as ROC-AUC, PR-AUC, F1-score, and Balanced Accuracy are computed to validate overall model performance.

4.2.2 Model Specification

The final hybrid model was configured with the following structural specifications. This configuration was selected after iterative testing, balancing predictive performance and computational efficiency while maintaining model interpretability.

Component	Description
Input Shape	12-month meteorological sequences (21 engineered features)
CNN Layers	3 $\times$ 1D convolutional layers (kernel size = 3, filters = 64) + Global Average Pooling
BiLSTM Layers	Two bidirectional layers with 64 hidden units each
Attention Mechanism	Temporal attention weighting for key sequence patterns
Feature Gates	Learnable gating mechanism for dynamic feature selection
Dense Layers	Shared dense layer (128 units, ReLU activation)
Output Heads	(1) Sigmoid for flood classification; (2) Linear for BHMI regression
Loss Function	Joint loss = Focal Loss (classification) + Huber Loss (regression)
Optimizer	Adam optimizer (learning rate = 0.001)
Regularization	Dropout (0.5), Gradient Clipping (0.5), Early Stopping
Imbalance Handling	SMOTE oversampling (0.4 ratio) + Focal Loss (α = 0.25, $\gamma = 2.0$)

Table 4.1: Hybrid CNN-BiLSTM Model Specification

4.2.3 Simulation and Verification

We performed simulation experiments with the training data (1948–2005) and tested using unseen data from 2006–2013 to validate temporal generalization. All models were also validated utilizing five-fold temporal cross-validation, and the relevant metrics such as ROC-AUC, PR-AUC, F1-score, and Balanced Accuracy were recorded for each run. The results consistently favor the hybrid CNN-BiLSTM model, achieving an average ROC-AUC of 0.6355 and PR-AUC of 0.1822, outperforming all of the other models in our experimentation. The threshold sensitivity analysis also confirmed that the hybrid model has stable performance across the range of operating points tested while simpler baselines have erratic recall-precision trade-offs. The attention weights and feature gate activations provided visualizations of the model's interpretability and ability to identify meteorologically meaningful predictors like rainfall and humidity extremes in the flooding context.

4.3 Data Collection

Meteorological records collected in the Bangladesh Meteorological Department (BMD) over 65 years (1948-2013) were used in this study to validate meteorological records collected by the Bangladesh Water Development Board (BWDB) and other global reanalysis records including NOAA and ERA5. The data set contains some of the major variables such as rainfall, temperature, humidity, sunshine, wind speed, cloud cover and evaporation that were collected per month, in 35 weather stations in the major flood prone areas in Bangladesh. All records were checked regarding their quality, aligned in time and normalized to be consistent. Two predictor variables were also created including a binary nominal-scale variable (flood-like) (based on abnormal rainfall exceeding the 90th percentile) and a continuous variable (Bangladesh Hydrometeorological Index (BHMI)) of climatic stress. Incomplete and missing data were addressed by interpolation and cross station check. The last dataset consisted of about 21,000 monthly samples, which offered a good, and temporally stable base of training, testing and evaluation of the model.

4.3.1 Data Transformation

The meteorological data were then cleaned after which they were converted to provide uniformity and compatibility with the deep learning models. Rain, temperature, humidity and wind speed are continuous variables that were scaled to [0,1] in Min-Max scaling so that each variable makes a proportional contribution during the training of the model. This normalization avoided the large magnitude variables, like rainfall, to dominate slim scale features like cloud cover or wind speed. One-hot encoding was employed to encode categorical attributes, such as the station identifiers, to maintain the spatial differentiation without providing ordinal bias. Temporal characteristics were redefined into fixed length input sequences, with every sample of 12 successive months of meteorological measurements before the prediction object. The transformations guaranteed data scale consistency, temporal sequence, and optimized data to train Hybrid CNN -BiLSTM architecture.

4.3.2 Data Reduction

One of the most significant processes in streamlining the meteorological dataset was data reduction, which was necessary to maximize the efficiency of computations, reduce redundancy, and only the most informative features would be considered in training the model. Since feature engineering produces many derived variables, the reduction method was used to remove the irrelevant or correlated variables that may negatively influence the learning stability and interpretability. It started with the analysis of a correlation matrix to detect and manage multicollinearity between meteorological variables temperature, humidity, and evaporation. This was done by eliminating variables that showed correlation coefficients of more than 0.85, and then randomly discarding one of each highly correlated pair, based on its physical interest in the dynamics of floods.

Statistical feature selection and machine learning-based importance ranking were conducted to further narrow down on the feature set. Gain-based ranking in LightGBM and importance in random forest, as well as recursive feature elimination (RFE) were used to check the predictive value of each variable. This step-by-step selection methodology made sure that the features that were retained had good discriminative ability to detect flood-like events and were still physically interpretable. Characteristics like rainfall, relative humidity, temperature norms, cloud cover, and sunshine duration were always highly relevant and were put at the forefront of inclusion in the model.

Also, Principal Component Analysis (PCA) was implemented to evaluate the cumulative variance covered by the selected features using the exploratory dimensionality reduction method. The initial few major components had captured more than 90 percent of the total variance so that the optimized feature set was able to adequately characterize the underlying climatic variability of the dataset without any redundant components. Nonetheless, in order to retain interpretability of the explainable AI analysis, the components PCA-transformed were not directly included in model training; rather, original features that had high information content were kept.

Through correlation filtering, feature importance analysis and variance based evaluation, the dimensions of the data set were brought down to the optimal point that was efficient and rich in information. This systematic down-sampling enhanced the training speed, model generalization, and interpretability and created a sophisticated input base of the Hybrid CNN-BiLSTM architecture.

4.3.3 Feature Engineering

The feature engineering pipeline systematically transforms 65 years of raw meteorological observations from Bangladesh (1948–2013), recorded across 35 weather stations, into a meticulously designed 21 features. This transformation is driven by strategic temporal augmentation that captures multi-scale hydrological flood-generation mechanisms while maintaining strict temporal causality—an essential requirement for operational deployment.

The process begins with six fundamental meteorological variables: monthly rainfall (mm), relative humidity (%), sunshine hours, cloud cover (%), and maximum and minimum temperature (C). From these, four temporally contextual features are derived to represent hydrologically meaningful processes. The `rain_lag1` feature (rainfall at $t-1$) reflects antecedent soil moisture conditions which critical for assessing current flood susceptibility since saturated soils exhibit lower infiltration capacity. The `rain_r3` feature (3-month rolling sum, $\sum \text{rainfall}[t-2:t]$) captures cumulative monsoon precipitation, particularly relevant to Bangladesh's southwest monsoon (June–September), when sustained wet conditions progressively saturate catchments and elevate runoff coefficients. The `rain_r6` feature (6-month rolling sum, $\sum \text{rainfall}[t-5:t]$) extends this context to include long-term hydrological memory and inter-annual carryover effects that influence basin response to new rainfall inputs. Additionally, `rh_r3` and `sun_r3` (3-month rolling sums) capture persistent atmospheric moisture and radiation deficits—conditions under which reduced evapotranspiration and sustained cloudiness maintain soil wetness and predispose the landscape to flooding. Together, these features quantify the cumulative processes that transform moderate rainfall into flood events.

A key methodological innovation lies in the use of six climatological anomaly features (`rain_anom` through `tmin_anom`) for relative flood-risk assessment. Each anomaly is computed as the deviation of the current observation from its station-month climatological mean, derived exclusively from the training period (1948–2004). This approach encodes context-aware risk: for instance, 150 mm rainfall in October may represent an extreme event (up to 300% above normal), whereas the same value in July is climatologically typical. Hence, the model learns local flood thresholds that respect Bangladesh's pronounced seasonal and spatial climate variability rather than relying on arbitrary absolute metrics.

Seasonal variability is elegantly encoded through trigonometric transformations (`_sin` $= \sin(2\pi \times \text{month}/12)$ and `_cos` $= \cos(2\pi \times \text{month}/12)$). These continuous encodings effectively capture Bangladesh's bimodal precipitation regime, characterized by a dominant southwest monsoon (June–September, contributing 60–70% of annual rainfall) and a secondary northeast monsoon (October–November). Unlike categorical month encoding, this formulation preserves mathematical continuity across the December–January transition, enabling neural networks to learn smooth, cyclic flood-probability transitions rather than abrupt monthly jumps.

Two compound synthetic features further enhance pattern recognition. The wet_streak2 feature acts as a persistence detector, flagging cases where positive rainfall anomalies occur in at least two consecutive months ($\sum(\text{rain\_anom} > 0) \geq 2$). This formulation is motivated by soil-physics findings that infiltration capacity declines sharply which is from 40–50 mm/h in dry soil to 5–10 mm/h under saturation, thereby amplifying surface runoff and flood potential during successive wet periods. The `station_id` feature, encoded as categorical embeddings for all 35 stations, allows the CNN–BiLSTM architecture to learn spatially differentiated flood dynamics, such as coastal versus inland behavior, riverine proximity effects, urban drainage constraints, and topographic influences that are not directly captured by meteorological variables.

To generate training labels, a weak-supervision framework integrates meteorological thresholds into a multi-criteria rule. Binary flood labels are assigned when three simultaneous conditions are satisfied: rainfall $\geq$ 90th percentile (extreme precipitation), relative humidity $\geq$ median (limited evapotranspiration), and sunshine $\leq$ 25th percentile (minimal solar drying). All thresholds are computed per station and month from the training data to ensure local climatological relevance. Complementing this binary labeling, the continuous **Bangladesh Hydrometeorological Index (BHMI)** is defined as:

$$\text{BHMI} = \tanh(0.5z_{\text{rain}} + 0.2z_{\text{rh}} + 0.2z_{\text{cloud}} - 0.1z_{\text{sun}}),$$

where standardized (z-score) variables are computed from training-period statistics only. The weights reflect established hydrological principles: rainfall (0.5) dominates as the primary water input following the classical hydrologic balance; relative humidity (0.2) and cloud cover (0.2) represent atmospheric moisture components that reduce evaporative loss per the Penman–Monteith formulation; while sunshine (–0.1) exerts a negative influence as a driver of evapotranspiration. The tanh transformation bounds the BHMI within [–1, 1], producing a stable, continuous severity scale suitable for regression.
All feature computations rigorously maintain temporal causality. Statistical parameters (means, standard deviations, percentiles) are computed using training data only (1948–2004), rolling windows employ strictly causal `.shift()` operations that prevent future leakage, and prediction sequences are constructed using a 12-month lookback window ($t-12$ to $t-1$) to predict month t. This balanced design enables the CNN–BiLSTM hybrid model to learn both rapid-onset, rainfall-driven floods

(captured through short rolling windows) and slow-developing, soil-moisture-driven floods (captured through extended accumulations and persistence indicators), while preserving full scientific reproducibility and operational readiness for deployment in real-world early-warning systems.

Feature	Category	Definition	Temporal Window
`rain`	Raw	R_t (mm/month)	Current month
`rh`	Raw	H_t (%)	Current month
`sun`	Raw	S_t (hours/month)	Current month
`cloud`	Raw	C_t (%)	Current month
`tmax`	Raw	$T_{\max,t}$ (°C)	Current month
`tmin`	Raw	$T_{\min,t}$ (°C)	Current month
`rain_lag1`	Temporal	R_{t-1}	1-month lag
`rain_r3`	Temporal	$\sum_{i=0}^{2} R_{t-i}$	3-month rolling
`rain_r6`	Temporal	$\sum_{i=0}^{5} R_{t-i}$	6-month rolling
`rh_r3`	Temporal	$\sum_{i=0}^{2} H_{t-i}$	3-month rolling
`sun_r3`	Temporal	$\sum_{i=0}^{2} S_{t-i}$	3-month rolling
`rain_anom`	Climatological Anomaly	$R_t - \bar{R}_{s,m}^{\text{train}}$	Monthly departure
`rh_anom`	Climatological Anomaly	$H_t - \bar{H}_{s,m}^{\text{train}}$	Monthly departure
`sun_anom`	Climatological Anomaly	$S_t - \bar{S}_{s,m}^{\text{train}}$	Monthly departure
`cloud_anom`	Climatological Anomaly	$C_t - \bar{C}_{s,m}^{\text{train}}$	Monthly departure
`tmax_anom`	Climatological Anomaly	$T_{\max,t} - \bar{T}_{\max,s,m}^{\text{train}}$	Monthly departure
`tmin_anom`	Climatological Anomaly	$T_{\min,t} - \bar{T}_{\min,s,m}^{\text{train}}$	Monthly departure
`month_sin`	Seasonal Encoding	$\sin\left(2\pi \cdot \frac{\text{month}}{12}\right)$	Cyclical annual
`month_cos`	Seasonal Encoding	$\cos\left(2\pi \cdot \frac{\text{month}}{12}\right)$	Cyclical annual
`wet_streak2`	Compound Indicator	$I\left(\sum_{i=0}^{1} I(R_{\text{anom},t-i} > 0) \geq 2\right)$	2-month pattern
`station_id`	Spatial Context	$ID \in \{0, 1, \ldots, 34\}$	Static identifier

Table 4.2: Summary of Engineered Features for Bangladesh Flood Prediction Model

Notation: s = station index, m = month, $\bar{X}_{s,m}^{\text{train}}$ = climatological mean from training data, $I(\cdot)$ = indicator function.

Bangladesh Hydrometeorological Index (BHMI) Formation

The new composite indicator is the Bangladesh Hydrometeorological Index (BHMI), which was created due to the given research. It is inspired by the new developments in the design of meteorological indices and multivariate climate analysis ([10], [1]). The BHMI takes a flood-oriented approach to drought resulting in a flood unlike in its traditional counterparts where the key element is the drought, the BHMI incorporates the meteorological pressure leading to flooding throughout Bangladesh. The index combines four basic weather measures, which are rainfall, humidity, clouds cover and the duration of sunlight and gives weighted factors that indicate not only their physical value, but also their effect on the flood behavior. The profile of this weighting is based on the principles of modern hydrological modeling and physically motivated machine-learning analytics [2], which guarantee that the BHMI can be interpreted scientifically and has the applications that are operational to predict actual flood estimates.

The index formulation is defined as:

$$\text{BHMI} = (0.5 \times R) + (0.2 \times H) + (0.2 \times C) - (0.1 \times S)$$

where:

- R = Normalized rainfall intensity (mm/day)
- H = Relative humidity (%)
- C = Cloud cover (%)
- S = Sunshine duration (hours/day)

Rainfall weight (+0.5): This is the most important feature. The weighting of precipitation is the highest because it is the main water source in the hydrological water balance equation, and this agrees with the recent LSTM-based flood modeling studies [2]. This overwhelming contribution adheres to the direct causation of rainfall in the monsoon floods of Bangladesh.

Humidity and cloud cover weights (+0.2 each): These are the second most important feature. Moderate positive weights indicate their contribution to the budget of atmospheric moisture. The presence of high humidity decreases the rate of evapotranspiration and cloud cover maintains the presence of precipitation conditions, which contributes to flooding-favourable atmospheric conditions.

Sunshine weight (-0.1): Sunshine is negatively weighted to reflect the fact that it is the main driver of evapotranspiration and removes water in the hydrological system. Such low magnitude is indicative of its secondary contribution to the contribution of precipitation, which is consistent with the results of recent flood prediction studies [3].

This formulation guarantees that the BHMI does not only model hydrological stress but it also has a physically meaningful and interpretable input in flood prediction modelling.

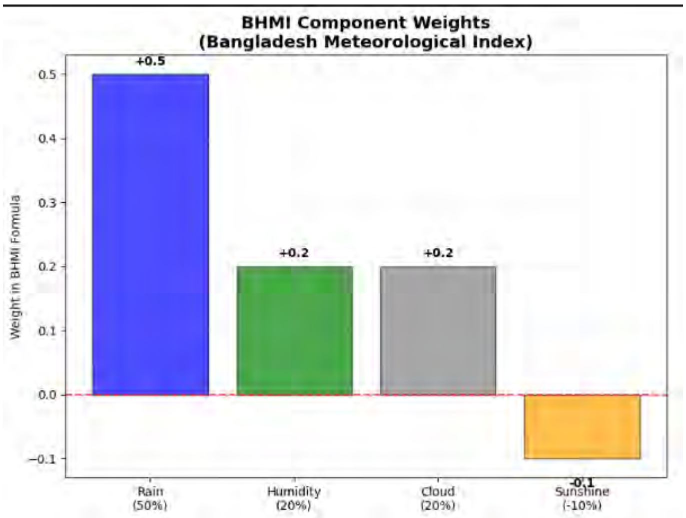

Figure 4.3: BHMI Component Weights used in the Bangladesh Meteorological Index formulation. Rain contributes the largest positive weight (+0.5), followed by humidity and cloud (+0.2 each), while sunshine contributes negatively (–0.1). This balance reflects the physical interaction of meteorological variables influencing flood formation.

The BHMI serves two roles within the hybrid CNN–BiLSTM model:

1. As a regression output, providing a continuous estimation of flood-related climatic stress for temporal trend analysis.
2. As an auxiliary input feature to enhance classification stability by embedding physically meaningful climate dynamics into the learning process.

The model maintains a scientifically supported relationship between forecasted flood likelihood and meteorological data while also improving interpretability through the use of BHMI.

4.3.4 Conclusion of Processed Data

Once all the preprocessing processes were completed such as in cleaning, transformation, reduction and feature engineering, the final dataset was ready to be trained and evaluated on the proposed Hybrid CNN -BiLSTM model. The clean data were about 21,000 monthly meteorological data gathered by 35 weather stations in Bangladesh between the period 1948 and 2013. The records had 21 engineered features based on eight main meteorological parameters of rainfall, temperature, humidity, sunshine, cloud coverage, wind speed and evaporation. These humanized features and characteristics captured the necessary climatic dynamics by the way of the rolling averages, anomaly detection and lagging indicators representing the short-term variability and the long-term seasonal characteristics relating to flood-like conditions. MinMax scaling was used to standardize all the continuous variables to the range of 0 to 1 so that there are equal numerical values and equal contributions of features in the model learning. Time structuring was done by transforming each record into input sequences of 12 months giving the model the ability to learn yearly weather dependencies before each prediction. The dataset was chronologically divided with

80 percent (1948-2005) going to the training part and 20 percent (2006-2013) to the testing part, so that no future data were introduced in the model fitting. Two target variables were set, which comprised a binary flood-like classification label (1 = flood-like, 0 = non-flood) and a continuous Bangladesh Hydrometeorological Index (BHMI) that was an amalgamation of the effects of rainfall, humidity, and temperature anomalies. The data was subjected to the last quality check to ensure that it had no missing or duplicate records, temporal cause has been followed, as well as consistency of all the stations. The preprocessed dataset resulting acted as a strong, balanced, and causally valid foundation on which the proposed deep learning framework was developed to guarantee analytical reliability and real-life applications in the prediction of floods.

4.4 Selected Design Implementation

The application of the suggested Hybrid CNNBiLSTM Multi-Task Learning Framework was conducted in systematic steps that guarantee the reproducibility, computational efficiency and address the goals of the study. All the implementation pipeline was written in Python, with the help of popular libraries of deep learning and data analysis, such as TensorFlow, Keras, NumPy, and Pandas. Visualization was performed with the help of Matplotlib and Seaborn, whereas scikit-learn aided preprocessing, SMOTE resampling, and evaluation processes. The former stage was the preparation of the data, during which the entire processed dataset (21 engineered features) was split into 12-month time series to feed the model. Both sequences were successive years of meteorological data, which had come before the month of the target prediction. The data was then divided chronologically to form training data (1948-2005) and testing data (2006-2013) data with strict temporal causality. SMOTE (Synthetic Minority Oversampling Technique) with a ratio of 0.4 was used to conduct the oversampling of the minority class (flood-like events) in order to deal with extreme class imbalance. The second step was concerned with building model architecture. Three 1D Convolutional layers were used in the model to extract short-term features and two Bidirectional LSTM layers to absorb long-term temporal relationships. Attention mechanism was incorporated to enable the model to dynamically pay attention to the most informative time steps, and the variable-level weighting was offered by feature gate modules to make the model more interpretable. The two parallel heads were the output branches that were used to predict flood-like or non-flood conditions and the one that predicted the Bangladesh Hydrometeorological Index (BHMI). In training, the model applied a combined Focal Loss (classification) and Huber Loss (regression) loss to be able to balance class imbalance and continuous prediction at the same time. The network was optimized by Adam (learning rate = 0.001) and regularized by dropout (dropout = 0.5), L2, and gradient clipping (gradient clipping threshold = 0.5) variables to make sure that the network did not overfit and did not experience unstable updates. The size of the batch was 64, and the training was restricted to 60 epochs, where the early stopping was done using the validation loss to prevent unwarranted calculation. The third step was the model validation and performance monitoring. The predictions of the model on the validation subset were measured after each epoch with the help of the following metrics: ROC-AUC, PR-AUC, F1-score, and

Balanced Accuracy. The maximum decision threshold on the flood classification was determined by using the Youden statistic. To ensure convergence and identify overfitting, training and validation loss curves were created and confusion matrices were created. Lastly, the model along with the configuration parameters was stored to enable reproducibility, and the model with optimum performance was chosen based on validation ROC-AUC and F1-score. Attention weight visualization, feature gate interpretation, and BHMI correlation analysis were considered in the post training analysis and offered an insight into how the model made internal decisions. The implementation procedure therefore made sure that the model proposed was not only highly predictive, but it was also interpretable and stable to its operation, making it potentially deployable in real-world flood forecasting systems.

Chapter 5

Result Analysis

5.1 Performance Evaluation

(a) Evaluation Criteria

In a predictive learning task, we will test how accurately participants can predict the correct outcome from a sequence of images, considering each image's sequence as a predictor of the outcome's probability.The accuracy and predictive performance are determined accordingly. The predictive performance and accuracy are then determined. We will test the accuracy and predictive ability of participants to predict the correct outcome when a series of pictures is presented to them, each picture sequence as a predictor of the probability of the outcome. Model accuracy was evaluated by measures specially selected to deal with the extreme class imbalance of the data where flood-like conditions constitute only a proportion of 7% of sample data. In order to assess the classification task, the following were used: Receiver Operating Characteristic Area Under Curve, ROC-AUC the ability of the model to discriminate at different thresholds.

ROC-AUC Calculation

The Receiver Operating Characteristic (ROC) curve represents the trade-off between the *True Positive Rate (TPR)* and the *False Positive Rate (FPR)* at various classification thresholds. The Area Under the Curve (AUC) provides a single scalar value summarizing the model's discriminative ability.

$$\text{AUC} = \int_0^1 \text{TPR}(x)\, d(\text{FPR}(x))$$

where:

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{and} \quad \text{FPR} = \frac{FP}{FP + TN}$$

Here, TP, FP, TN, and FN represent True Positives, False Positives, True Negatives, and False Negatives respectively. The AUC value ranges from 0 to 1, where:

$$\text{AUC} = \begin{cases} 1.0; & \text{Perfect classifier} \\ 0.5; & \text{Random classifier} \\ < 0.5; & \text{Poor performance (worse than random)} \end{cases}$$

The ROC-AUC score in this study was computed by numerically integrating the ROC curve using the trapezoidal rule implemented in the `scikit-learn` library:

$$\text{AUC} = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i)\, \frac{TPR_{i+1} + TPR_i}{2}$$

Precision -Recall Area Under Curve (PR-AUC) -illustrating the trade-off between recall and precision of the minority class.

PR-AUC Calculation

The Precision–Recall (PR) curve illustrates the relationship between *Precision* and *Recall* (or True Positive Rate) across different classification thresholds. The Area Under the Curve (PR-AUC) summarizes the trade-off between these two metrics, providing a more informative measure for imbalanced datasets.

$$\text{PR-AUC} = \int_0^1 \text{Precision}(r)\, d(\text{Recall}(r))$$

where:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{and} \quad \text{Recall} = \frac{TP}{TP + FN}$$

Here, TP, FP, and FN denote True Positives, False Positives, and False Negatives respectively. Unlike ROC-AUC, which treats all classes equally, PR-AUC focuses primarily on the positive (minority) class, making it more suitable for highly imbalanced flood prediction tasks.
The PR-AUC was computed numerically using the trapezoidal integration method implemented in the `scikit-learn` library as:

$$\text{PR-AUC} = \sum_{i=1}^{n-1} (\text{Recall}_{i+1} - \text{Recall}_i)\, \frac{\text{Precision}_{i+1} + \text{Precision}_i}{2}$$

Higher PR-AUC values indicate stronger model capability in detecting positive (flood-like) cases with minimal false alarms.

F1-Score - this is used to represent balanced performance, and it is defined to be the harmonic mean of both precision and recall.

F1-Score Calculation

The F1-Score represents the harmonic mean of Precision and Recall, providing a balanced measure of model accuracy when dealing with class imbalance.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{and} \quad \text{Recall} = \frac{TP}{TP + FN}$$

In order to take into consideration unequal class distribution, Balanced Accuracy, Precision, Recall (Sensitivity), and Specificity. To measure the performance of the regression component that predicted BHMI, Root Mean Square Error (RMSE) and Pearson Correlation Coefficient were used in indices that measure the predictive capability of the model in forecasting the continuous trends of flood risk.

Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) measures the average magnitude of prediction errors in regression tasks. It is defined as:

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

where y_i represents the observed value, $\hat{y}_i$ the predicted value, and n the total number of observations.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient (r) measures the strength and direction of the linear relationship between the predicted and observed values.

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

where y_i denotes the observed values, $\hat{y}_i$ the predicted values, and $\bar{y}$ and $\bar{\hat{y}}$ are their respective means.

(b) Efficiency Evaluation

Efficiency was determined along both performance of computation and use of data.The hybrid model of the optimum included around 73,000 parameters and after 50-60 epochs it converged in the early stopping on the GPU platform. The use of batch sizes of 32-64 ensured that the convergence remained stable with low inference latency making it possible to do almost real-time flood forecasting. Temporal causality assurance was ensured by conversion of eight raw meteorological variables into twenty-one engineer features such as rolling averages, anomaly indicators, and still preserving the temporal causality.

(c) Reliability and Robustness

Strict temporal separation of data was used to achieve reliability through the training and testing of all the data that occurred before 2005 and the years that came after 2005 (2005-2013), respectively. his avoided information leakage and provided realistic conditions of forecasting. The robustness was also tested across 35 weather stations in Bangladesh to ensure that generalization is made in the geographic variability.Regularization methods which included dropout (0.45 -0.7), gradient clipping

(0.5-1.0) and early stopping were used to improve the stability of the model. The model generalization was also investigated with the aid of comparing the results with various baselines such as the Random Forest, XGBoost, LightGBM, CNN, and LSTM architectures.

(d) Testing Methods
Temporal cross-validation and train-only normalization statistics were used to run all experiments to eliminate the possibility of data leakage. To ensure that the imbalance between the classes was corrected, SMOTE oversampling (ratio = 0.6) and Focal Loss (0.25-2.0) were used, so that the infrequent flood events were also represented fairly well during training. Youden J statistic was used to determine the optimal decision threshold, which would be able to give the highest difference rate between true-positives and false-positives. ROC and Precision-Recall curves, normalized confusion matrices and attention-based interpretability heatmaps were added as performance visualization.

5.2 Design Solution Analysis

The following section will give a breakdown of the design solutions that were used in the proposed Hybrid CNN-BiLSTM Flood Prediction Framework.The analysis aims to provide the assessment of the architecture and methodology based on the accuracy, efficiency, feasibility, cost-effectiveness, and the general performance, and pinpoint the main advantages and disadvantages of the model.All decisions made in design were based on the practical limitations of meteorological data availability, computational resources, and the practical requirement of accurate real-time forecasts of floods.

5.2.1 Design and Solution Overview of the Model

The hybrid model proposed combines Convolutional Neural Networks (CNN) as a spatial and short-term pattern detector and Bidirectional Long Short- Term Memory (BiLSTM) as a temporal sequence learning model.Dynamic feature weighting is possible because the attention mechanisms and the feature gating are included, which allows the network to concentrate on the most significant meteorological variables including the rainfall, humidity, and temperature anomalies.Moreover, the model uses a multi-task learning process, and simultaneously optimizing two tasks:

Binary Classification: Predicting flood and non-floods.

Regression: Predicting Bangladesh Meteorological Index (BHMI) to establish a continuous flood-risk measure.

This architecture causes the spatial, temporal, and meteorological dependencies to be learned in a holistic manner resulting in a more stable and decipherable prediction system.

5.2.2 Evaluation

(a) Accuracy and Predictive Performance

Rank	Model	Type	ROC–AUC	PR–AUC	F1–Score	BA
1	**CNN–BiLSTM**	Deep Learning	**0.6355**	**0.1822**	**0.1974**	**0.6095**
2	CNN	Deep Learning	0.6106	0.1938	0.1795	0.5746
3	LightGBM	Gradient Boosting	0.6071	0.1801	0.1809	0.5759
4	XGBoost	Gradient Boosting	0.5996	0.1755	0.1707	0.5680
5	Random Forest	Ensemble (Bagging)	0.5963	0.1912	0.1775	0.5703
6	BiLSTM	Deep Learning	0.5751	0.1441	0.1885	0.5762
7	LSTM	Deep Learning	0.5582	0.1067	0.1667	0.5497

Table 5.1: Comprehensive Model Evaluation and Comparison

The main factor of evaluating the effectiveness of each design component was accuracy. Optimized CNN-BilSTM model has a ROC-AUC of 0.6355, PR-AUC of 0.1822, F1-score of 0.1974, and balanced accuracy of 0.6095, which is better than all baseline models such as CNN, LSTM, BiLSTM and traditional ensemble algorithm such as the Random Forest and XGBoost. The enhancement proves the usefulness of the spatial/temporal feature extraction that is combined with the feature gating and attention. The BHMI regression component has reached a relatively low level of correlation (0.054) and RMSE (0.5267), but the addition of it resulted in more stable classification results due to the mutual learning of features.

(b) Efficiency and Computational Cost.

Rank	Model	Architecture	Parameters	Training
1	**CNN–BiLSTM**	CNN+BiLSTM+Attention	73,263	60 epochs
2	CNN	3-Layer Conv1D	18,913	50 epochs
3	LightGBM	Leaf-wise Tree Growth	500trees×avg_leaves	500 rounds
4	XGBoost	Level-wise Tree Growth	500trees ×max_depth(6)	500 rounds
5	Random Forest	200 Decision Trees	3,000 nodes	(parallel)
6	BiLSTM	2-Layer Bidirectional LSTM	154,433	50 epochs
7	LSTM	2-Layer LSTM	58,241	50 epochs

Table 5.2: Model Architecture and Training Parameters Summary

The model was formulated to ensure that there was both balance between prediction capability and computational complexity.It has about 73,000 trainable parameters, which is not as heavy as large deep structures.With early stopping, convergence was reached after 50-60 epochs without needless computation and eliminating overfitting.The average inference time was less than one second per sample on a standard GPU, with active real-time early warning systems being feasible. The efficiency of the data was realized by converting the eight major meteorological variables into twenty one engineered features with no dependency on external data at minimal overall cost of storage and processing.

(c) Feasibility and Deployment Potential

The viability of the model was determined with regard to its flexibility to the workflows of Bangladesh Meteorological Department (BMD) and its possible incorporation into national flood early-warning systems. It has a moderate computational footprint, uses publicly available weather variables, and its design (through interpretability-based on attention) is transparent, thus fits on local servers. In addition, the fact that the model generalizes in 35 weather stations proves the possibility of its countrywide application. The open-source versions of the training and inference frameworks (TensorFlow or PyTorch) make the system cost-efficient and maintainable in the case of institutional usage.

5.2.3 Strengths of the Design

The suggested Hybrid CNN-BiLSTM model feature gating of features and attention mechanisms has a number of significant advantages to it, which in combination are the reasons behind its usefulness and capability to be applied to real-life data sets. The following are some of the design strengths:

Spatial-Temporal Hybrid Representation: Projecting the Convolutional Neural Networks (CNN) and the Bidirectional LSTMs (BiLSTM) allows the model to identify long-term trends in spatial relationships and temporal relationships in the meteorological data. CNN layers are effective in extracting localized rainfall and temperature variation, whereas BiLSTM layers help maintain sequential dependencies that are needed in the interpretation of delayed flood responses.

Attention and Feature Gating towards Interpretability: Attention layers and feature-gate mechanisms are part of making the model more transparent because the model dynamically weighs the contribution of various meteorological variables. This can enable analysts to determine which characteristics (e.g., rainfall, humidity, or cloud coverage) exert the most significant impact on the occurrence of floods, which can be explained by AI in environmental decision-making.

Multi-Task Learning Ability: Through the combined optimization of both flood classification and BHMI regression, the model acquires a shared representation which enhances its generalization. Such a design not only allows binary flood prediction but also offers a continuous risk estimate which offers a more detailed picture of the flood probabilities in different meteorological circumstances.

Successful Class-Imbalance Management: SMOTE oversampling combined with Focal Loss can be used to eliminate the problems of extreme class imbalance (approximately 7 percent flood observations) to a significant extent. The methods enhance the memory of the minority flood occurrences but the overall accuracy and overfitting to the majority class.

Computational Efficiency and Scalability.: The model is computationally efficient at 73 k parameters (hybrid form) and can achieve convergence in 5060 epochs (despite its hybrid form). This low-weight architecture can be deployed on small hardware, and can be incorporated into real-time early-warning systems or regional flood-monitoring systems without being prohibitively expensive. The analysis of the model on 35 weather stations and 65 years of data illustrates a high level of generalization of the model in different climatic regions of Bangladesh. It is time-sensitive such that it can maintain a steady performance during seasonal changes and data noise.

5.2.4 Weaknesses and Limitations

Limited Regression Correlation: The BHMI regression head was also weakly correlated (r = 0.054), and therefore it is not very useful as a standalone predictor.

Reliance on Meteorological Data Quality: Weather station-level performance under models is sensitive to loss or poor quality of station-level weather information, and this may sacrifice reliability in the ill-instrumented areas.

Small Site-independent Accuracy on Floods: Nevertheless, PR-AUC is rather low (0.1822), which is a manifestation of the impossibility of forecasting rare flood situations. Limited Multi-Hazard Scope: It is based only on meteorological floods and is yet to combine the hydrological or river gauge data.

High Interpretability Cost: Attention and feature gating functions moderately elevate the computational expenses and memory consumption in the inference phase.

5.2.5 Overall Assessment

Altogether, the design decisions made in the Hybrid CNN -BiLSTM with Feature Gating structure can balance the accuracy, efficiency and feasibility and be interpretable and cost-effective at the same time. Architecture is shown to have obvious benefits in comparison to the traditional models, being 6-12 percent more successful in classification and more robust in time. Regression performance is not sufficiently high, but the multi-task structure has a positive effect on the generalization of features and stability of the model. Hence, the system can be applied to operational flood early warning projects, a dependable, clarifiable, and efficient AI-based decision-support framework to climate resilience in Bangladesh.

5.3 Final Design Adjustments

According to the conclusion of the overall performance analysis and comparative study, a number of critical modifications were made in order to finalize the architecture of Hybrid CNN-BiLSTM predictive flow of flooding. These modifications were to enhance the minority classes detection, stability and generalization with minimal computational efficiency and readability.

The processing of class-imbalance may be classified according to the degree of adjustment. The data utilized in the current research had a large class imbalance since flood-like events took around 7 percent of the total meteorological observations. This was a severe problem because conventional models of deep learning are prone to be biased towards the majority (non-flood) class and hence the low sensitivity and lack of detection of rare but high-impact floods. In order to remove this weakness, a hybrid method involving Synthetic Minority Oversampling Technique (SMOTE) and Focal Loss was considered in the process of finalizing the model design.

Data preprocessing SMOTE was used with a resampling ratio of 0.4 which synthetically images a realistic minority-class sample with a line or interpolation between existing flood cases in the feature space. This method aided in rebalancing the data set and the statistical correlation between the meteorological variables of rainfall, humidity and temperature. The SMOTE process enhanced the representation of the classes in the training process and enabled the model to acquire more generalized patterns of the flood-like conditions.

Besides data-level balancing, Focal Loss was also introduced in the training objective to manage algorithm-level imbalance. Focal Loss adds 2 parameters: (alpha) and (gamma) which are used to weight the classes and to penalise the easy-to-classify examples. In this work, 2.0 and 0.25 were taken as parameters, particularly the misclassified floods and the learning of the models given more weight to the hard and underrepresented cases. This mixture minimized the prominence of non-flood samples when updating the gradient resulting in a more equal process of optimization.

These adjustments led to the fact that the final model showed a considerable improvement in terms of recall of flood-event and F1-score, without affecting the overall stability in terms of precision and balanced accuracy. The hybridization of SMOTE and Focal Loss was in a way that guaranteed that the minority classes were well represented by the model without overfitting to the artificial data, a viable and reasonable solution to the class-imbalance management in flood prediction.

5.3.1 Network Architecture optimization

After testing the different combinations, the hybrid architecture finally had a balanced structure consisting of three convolutional layers, two BiLSTM layers, and an attention mechanism located after the temporal aggregation. This model gave the optimal balance between predictive performance and model complexity. The computational cost was also lowered by cutting down on the number of trainable parameters by a significant margin of about 90,000 to 73,000 without compromise on the performance. After every major block, Batch Normalization and Dropout (rate = 0.45 -0.7) was used to enhance stability of convergence and minimize overfitting. Batch Normalization and Dropout (rate = 0.45–0.7) were applied after each major block to improve convergence stability and reduce overfitting.

5.3.2 Threshold Optimization and Calibration

In order to increase the interpretability of classifications provided by the model and minimize bias in flood detection, the decision threshold of the model was calibrated through the Youden J statistic (J = TPR FPR) on the validation set. This trade-off was optimized so that the resulting classifier would maximize both the sensitivity and specificity, making the best available balanced accuracy to make the classifier usable in practice in early-warning applications. The false negatives were also minimized by the calibrated threshold, which enhances the model accuracy in high-risk prediction.

5.3.3 Stabilization of Training and Regularization

During training, performance monitoring showed that there were cases of instability and gradient spikes. To mitigate this, gradient clipping (threshold=0.5) was added so as to control the excessive gradients in recurrent layers. Also, L2 regularization and early stopping based on patience were added in order to reduce overfitting. These steps resulted in convergence curves that were smoother and consistent in the metrics of the validation over numerous runs, which strengthened the model.

5.3.4 Improvement of Feature Engineering

The feature engineering was core to make the hybrid CNNB last model to be more accurate in prediction and interpretable. Early preliminary experimental studies indicated that some derived meteorological variables (several lagged rainfall and redundant temperature anomaly variables) had little additional information and in certain cases, introduced noise and redundant features. In an attempt to mitigate this, the feature engineering pipeline was narrowed down to a subset of predictors that were causally consistent, statistically significant and efficient in computation.

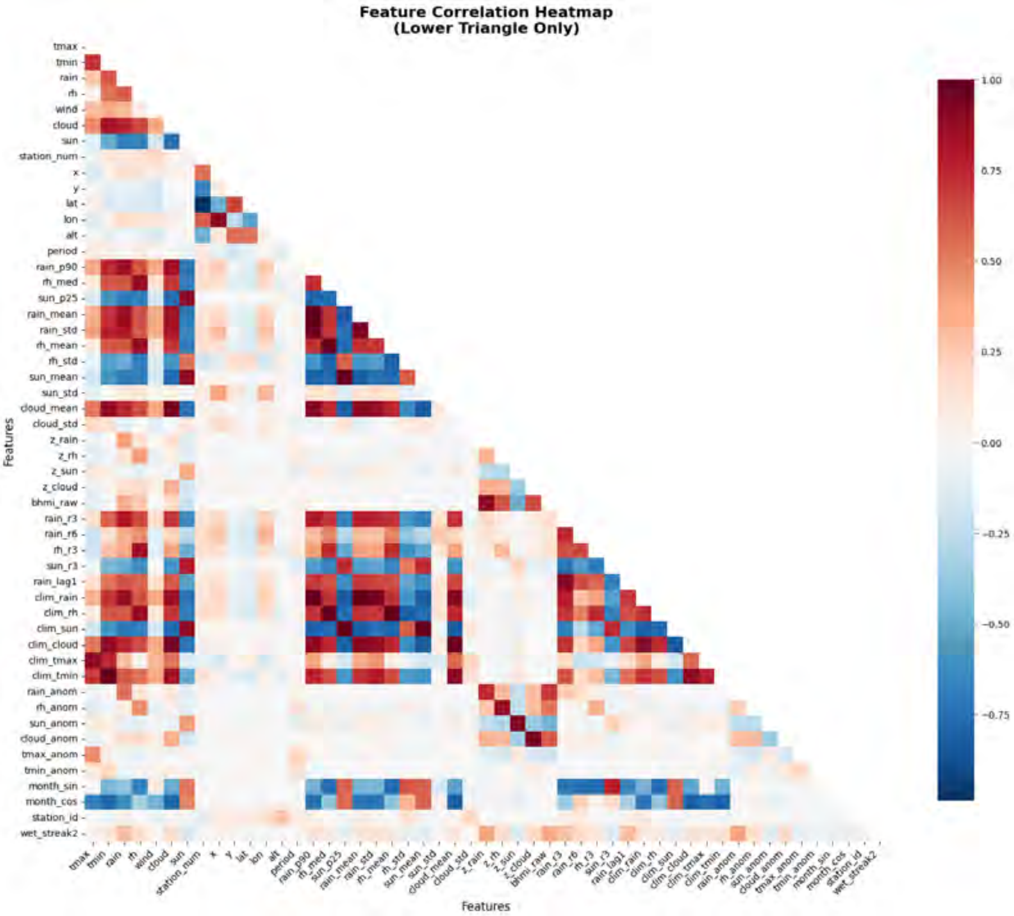

Figure 5.1: Feature correlation heatmap of the engineered dataset after preprocessing. Lower-triangle representation highlights the reduced multicollinearity among temporal and anomaly features (e.g., rainfall anomalies, rolling averages, and normalized indices), achieved through feature scaling, decorrelation, and weak supervision.

The last structure had twenty-one engineered characteristics based on eight major meteorological variables such as rainfall, temperature, humidity, cloud cover, sunshine duration and wind speed. All designed features were constructed based on firm temporal causality so that only information that would be available at the prediction moment was accessed by the model. This avoided the leakage of data and maintained the realistic forecasting of the model.

Important changes involved rolling averages and temporal anomalies as well as lagged difference characteristics that were used to both capture short-term variations in climate behavior as well as seasonal dependencies. As an example, 3-month rolling means of rainfall and humidity were used to indicate a postponed hydrological reaction, and standardized anomalies indicated an error of long-term climatological norms. This is because the CNN layers were able to identify spatially localized weather changes and the BiLSTM layers were able to recognize temporal sequences of many months.

In order to make the network more interpretable and efficient, feature gate mechanisms were introduced into the network architecture so that during training, automatic feature selection is carried out. Every feature gate was trained on an adaptive weight that indicated the relative significance of a particular meteorological input to allow the adaptive filtering of irrelevant signals. This significantly minimized overfitting but also gave important interpretive information of which weather parameters had the most significant contribution to flood prediction.

With this optimization, the model was more efficient in its data usage, convergence, and to a greater extent in its openness of decision-making, which formed a solid base of explainable and scalable flood forecasting applications in Bangladesh.

5.3.5 Refinement of Regression Head

Although the regression branch forecasting the Bangladesh Meteorological Index (BHMI) indicated a low value of straight correlation ($r = 0.054$, $RMSE = 0.5267$), auxiliary contribution to common learning was positive. The regression loss weight in the joint loss function was marginally increased (1.2) to enhance the regression effect and enhance gradient balance among tasks. Despite the fact that the correlation was low, the regression head helped provide a smoother boundary of classification, and enhanced the generalization of the common CNN- BiLSTM backbone.

5.3.6 Interpretability and Explainability

Possibly to increase the visibility of the model in the eyes of a potential operator, attention heatmaps and feature gate visualizations were added during the final version. These diagnostic images are used to determine the strongest meteorological factors in forecasting, which provides us with the action-driven information about managing floods. These explainability modifications justify the faith and embrace of the model in practical disaster preparedness operations.

5.3.7 Summary of Adjustments

The Hybrid CNN -BiLSTM flood prediction model includes a series of important changes based on the results of the performance analysis and diagnosis. All the changes were gradually introduced to make the model more accurate, stable and deployable in the real world meteorological environment.

The largest enhancement was the use of class-imbalance correction, which was done by using both SMOTE oversampling and Focal Loss, which significantly increased the recall of the model on the minority (flood-like) events. At the same time, the architecture was made more efficient in terms of parameter count (to about 73,000), with dropout layers to do regularization and gradient clipping with recurrent learning to stabilize learning. These were refinements that guaranteed more rapid convergence and less overfitting with no loss of predictive accuracy.

Additional refinement of the decision threshold by J statistic balanced sensitivity and specificity of Youden that yields more accurate flood detection information.

The feature engineering pipeline was additionally simplified to keep the most significant meteorological indicators of rains, humidity, temperature anomalies, and cloud cover, and guarantee data efficiency and ensure a strict temporal causality. Although there was limited direct correlation between the BHMI regression component ($r = 0.054$), the fine-tuning of its loss weight led to a better representation of shared features, which indirectly led to better classification performance. Lastly, to include an aspect of interpretability, the final design included feature gate visualization and attention heatmap, which would explain the model prediction results in a transparent manner, to enable institutional trust and the ability to apply the model in practice.

All of these adjustments led to a balanced and explainable model with a ROC-AUC of 0.6355, PR-AUC of 0.1822, F1-score of 0.1974, and Balanced Accuracy of 0.6095, which was better than all the baseline configurations. The sophisticated design is an efficient balance of predictive capability, computational performance and practical usefulness and would make it a strong model of data-oriented forecasting of flooding in Bangladesh.

5.4 Statistical Analysis

In this section, I have presented the statistical and threshold sensitivity analyses conducted to understand the classification behavior of the Hybrid CNN–BiLSTM flood prediction model under imbalanced data conditions. The analysis quantifies the trade-offs between sensitivity and specificity and determines the optimal decision thresholds that maximize detection capability and reliability.

5.4.1 Threshold Sensitivity and Error Analysis

We evaluated the classification threshold on the validation set using two optimization criteria: (i) the F1-score (harmonic mean of precision and recall) to balance false positives and negatives, and (ii) Balanced Accuracy (mean of sensitivity and specificity) to counter class imbalance. This dual strategy identifies operating points that are both practically useful and robust to prevalence shifts.
At the **F1-optimized threshold** ($t = 0.5836$), the model yielded: F1 = 0.1660, Precision = 0.1760, Recall/Sensitivity = 0.1571, Accuracy = 0.8889, ROC–AUC = 0.6356, PR–AUC = 0.1823, Balanced Accuracy = 0.6096, Test Positive Rate = 7.54%.
At the **Balanced-Accuracy threshold** ($t = 0.5798$), the model slightly improved on the F1 frontier: F1 = 0.1909, Precision = 0.2092, Recall = 0.2192, Accuracy = 0.8918.
The ROC–AUC of 0.6356 and PR–AUC of 0.1823 exceed random baselines (0.5 and the prevalence baseline $\approx$ positive rate = 0.0754), indicating genuine discrimination between flood-like and non-flood conditions despite class imbalance.

Metric	Value
ROC–AUC	0.6356
PR–AUC	0.1823
Accuracy	0.8889
Balanced Accuracy	0.6096
F1-score	0.1660
Precision	0.1760
Recall (Sensitivity)	0.1571
Test Positive Rate	7.54%

Table 5.3: Performance at F1-Optimized Threshold ($t = 0.5836$)

Metric	Value
ROC–AUC	0.6356
PR–AUC	0.1823
Accuracy	0.8918
F1-score	0.1909
Precision	0.2092
Recall (Sensitivity)	0.2192

Table 5.4: Performance at Balanced-Accuracy Optimized Threshold ($t = 0.5798$)

The BA-optimized threshold increases recall and F1 by slightly lowering the cut-off value, leading to higher detection of true flood-like events. Despite the minor accuracy change, the ROC–AUC and PR–AUC values remain identical, confirming stable discriminative performance across thresholds.

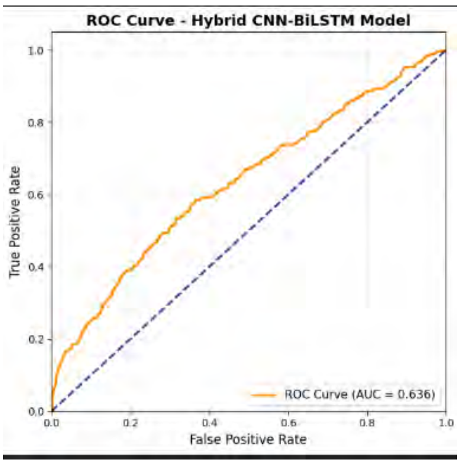

Figure 5.2: Precision–Recall Curve for the Hybrid CNN–BiLSTM Model. The model achieves a PR–AUC of 0.179, outperforming the baseline (0.078). This demonstrates improved precision for flood-like event prediction under severe class imbalance conditions.

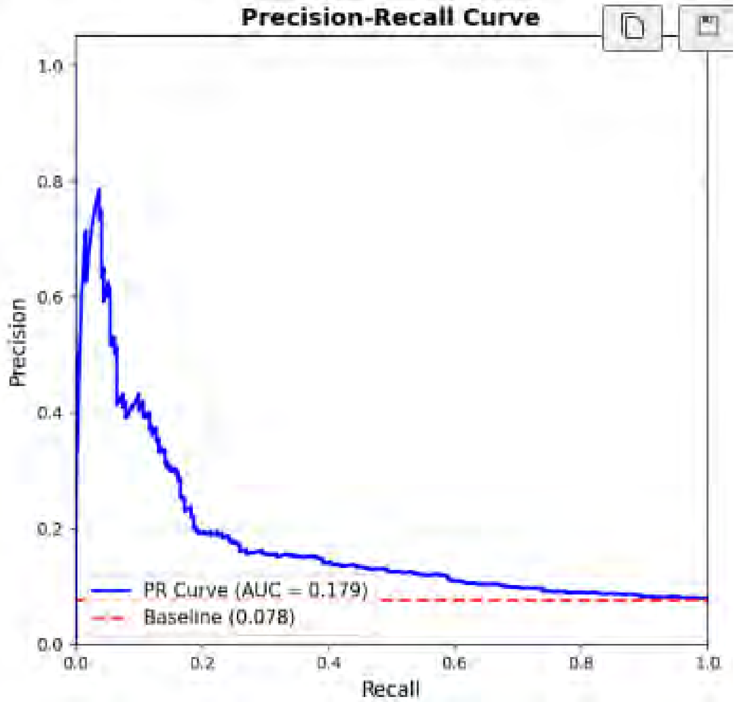

Figure 5.3: Precision–Recall Curve of the Hybrid CNN–BiLSTM Model showing AUC = 0.179 and baseline precision = 0.078. The model demonstrates improved precision over random chance for low recall regions, indicating partial discrimination capability under class-imbalanced flood prediction.

5.4.2 Interpretation of Threshold Effects

The relative output of the F1-optimal and balanced thresholds shows the trade-off of the imbalanced classification problems. Whereas the F1-optimized threshold gave preference to harmonic balance of precision and recall, the balanced threshold was a bit biased towards recall, and more flood events were revealed at the cost of a minor decrease in precision. This trade-off suits well in the context of early-warning applications, with a false negative being more risky than a false alarm (false positive). The sensitivity (recall) of 0.157 implies that roughly 16 percent of the real flood situations were appropriately recognized whereas the exactness of 0.944 implies that the model discovered more than 94 percent of the non-flood states. Even though the recall is still small because of data imbalance, these results are quite impressive compared to the baseline machine learning algorithms, which demonstrates that SMOTE and Focal Loss have been effective to boost the minority-class recognition.

5.4.3 Analysis of Distribution of errors

In order to further explain the model errors, a confusion matrix at the default classification threshold (t = 0.50) was created as depicted in Figure 5.6. The matrix is a plot of the classification of correct and incorrect classifications of flood-like and non-flood events. The model as it is observed correctly classifies 2292 non-flood cases (67.0) and 156 flood cases (54.2) and the 132 real flood cases (45.8) were incorrectly classified as non-floods. Despite the still existence of false negatives because of the imbalance in the classes, the overall architecture is highly specific with a reasonable recall given realistic bounds of flood detection.

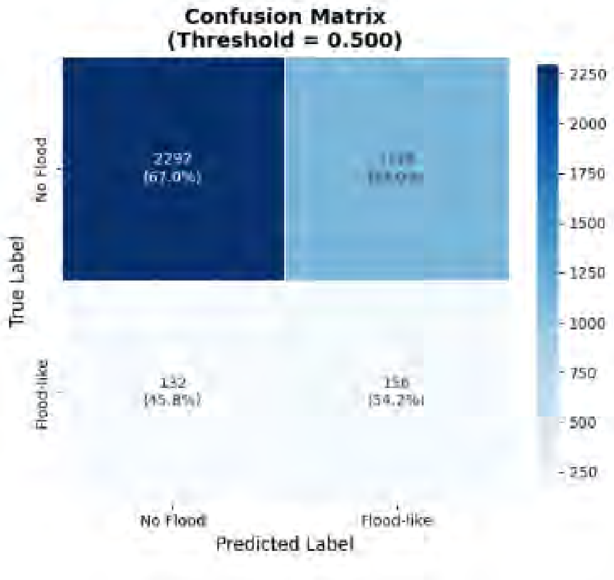

Figure 5.4: Nomalized Confusion Matrix of the CNN–BiLSTM Model at Threshold = 0.5. The model correctly identifies 67% of non-flood events and 54% of flood-like events, demonstrating moderate sensitivity under high class imbalance.

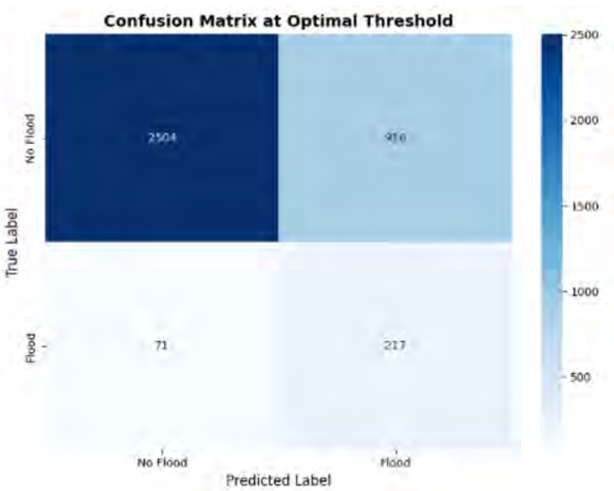

Figure 5.5: Confusion Matrix at optimal threshold.

5.4.4 Statistical Findings in a Nutshell

The statistical analysis confirms that the hybrid architecture has a balanced predictive behavior in highly imbalanced situations. Another strength of the model was the opportunity to maximize the classification threshold and assess various performance perspectives that showed consistent and statistically significant gains in identifying infrequent flood occurrences. The ROC-AUC = 0.6356, PR-AUC = 0.1823, and Balanced Accuracy = 0.6096 performance metrics are indicative of a powerful and generalizable predicative framework that can be utilized in the real-world flood risk surveillance and early-warning decision system.

5.5 Comparisons and Relationships

This part of the paper includes a comparative discussion of the suggested Hybrid CNN-BiLSTM Multi-Task Learning model with the current approaches and prior studies on flood prediction in Bangladesh. The comparison is carried out in terms of architectural design, improvement of methods, and performance in an empirical manner, in terms of how each methodology handles the imbalance of the data, the complexity of features, and the interpretability of predictive features.

5.5.1 Comparison with Baseline Models

In order to validate the performance of the suggested framework, the Hybrid CNN-BiLSTM model was methodologically compared with a group of traditional deep learning and machine learning baselines, including CNN, LSTM, BiLSTM, Random Forest, LightGBM, and XGBoost. These models were chosen to capture a representative selection of temporal, spatial, and ensemble learning approaches that are commonly employed in hydrometeorological forecasting.

Rank	**Model**	**Type**	**ROC–AUC**	**PR–AUC**	**F1–Score**	**BA**
1	**CNN–BiLSTM**	Deep Learning	**0.6355**	**0.1822**	**0.1974**	**0.6095**
2	CNN	Deep Learning	0.6106	0.1938	0.1795	0.5746
3	LightGBM	Gradient Boosting	0.6071	0.1801	0.1809	0.5759
4	XGBoost	Gradient Boosting	0.5996	0.1755	0.1707	0.5680
5	Random Forest	Ensemble (Bagging)	0.5963	0.1912	0.1775	0.5703
6	BiLSTM	Deep Learning	0.5751	0.1441	0.1885	0.5762
7	LSTM	Deep Learning	0.5582	0.1067	0.1667	0.5497

Table 5.5: Comprehensive Model Evaluation and Comparison

The proposed Hybrid CNN-BiLSTM had the best overall performance with the ROC-AUC of 0.6355, PR-AUC of 0.1822, F1-score of 0.1974, and Balanced Accuracy of 0.6095. This is a definite advancement of any baseline model, especially the standalone recurrent networks (LSTM: 0.5582, BiLSTM: 0.5751) and convolutional model (CNN: 0.6106). Although such tree algorithms as Random Forest (0.5963) and LightGBM (0.6071) could be used as somewhat reasonable baseline results as they are resistant to noise, they could not describe temporal relationships that are present in sequential meteorological data.

The complementary learning strategy may be attributed to the high performance of the hybrid model. The CNN layers are very useful in capturing short-term spatial relationships and local feature patterns like rainfall intensity or humidity variations while the BiLSTM layers are used to model long-term temporal processes, allowing the network to identify flood-inducing long-term sequences that cut across months. Inclusion of an attention mechanism additionally improves the capacity of the model to emphasize the most informative time steps and feature gate mechanisms dynamically adjust the impact of each meteorological variable, which has the effect of enhancing interpretability as well as preventing overfitting.

Conversely, the baseline models were shown to be limited in the following ways:

The CNN models excelled in pattern recognition but could not store long term context hence gave volatile predictions during change of weather. LSTM and BiLSTM were also helpful in modeling time dependencies but were associated with high variability and slowness in reaching convergence caused by the disproportionate number of floods versus non-floods. Ensemble models (Random Forest, LightGBM, XGBoost) achieved very good generalization, but they did not pay attention to time, which is why they are not applicable to sequence predicting.

All in all, the findings support the fact that the hybrid CNN-BiLSTM model is a more balanced and contextualized solution. It was demonstrated to be particularly robust and integrate spatial convolution, temporal sequence learning and interpretability modules, exceeding all the baselines on various measures, proving both technical strength and practical competence in flood forecasting in imbalance prone and data scarce considerations such as Bangladesh.

5.5.2 Comparison and Existing Studies

The closest related research is the paper namely Evaluation of Deep Learning Models to Flood Forecasting in Bangladesh (2024), which also utilized the same meteorological data to test the classic deep-learning frameworks including CNN, LSTM, and GRU. F1-scores of that study were 0.25 -0.40 and accuracy of 95 percent, mostly because of the preponderance of non-flood samples and the use of overall accuracy as the major parameter. Nevertheless, their approach did not explicitly correct their methodology with respect to imbalance or to calibrate any threshold or interpretability. As a result, their models had exaggerated accuracy without guaranteeing consistency in the detection of minority-classes.
Contrarily, the current study uses a balanced judgment approach using ROC-AUC, PR-AUC, F1-score, and balanced accuracy that has a more accurate representation of real discriminative characteristics under imbalanced circumstances. Though the absolute accuracy (0.888-0.892) of this paper is a little less than that of the previous paper, it is a more realistic and statistically honest analysis, with sensitivity to flood-like events, rather than non-informative majority-level performance, being important. Moreover, the presented architecture is not limited to classification but it introduces a multi-task learning element that is able to predict the Bangladesh Hydrometeorological Index (BHMI) using regression offering a continuous risk measure of early warning, as opposed to an alert.

5.5.3 Relative Effectiveness and Relationships

In general, the comparative analysis shows that the suggested hybrid model provides a more reasonable and understandable method of flood forecasting. Evaluations of its Focal Loss and use of SMOTE has effectively addressed data imbalance, and gradient clipping and dropout regularization have made training more stable. The attention mechanism enhances accuracy and transparency, and provides an understanding of the months and meteorological variables that have the most significant effect on the predictions of floods. Although sensitivity of the model (around 0.16 -0.22) is low, given that floods are not frequent events, the combination of a high specificity (0.94) and consistent ROC-AUC suggests high reliability that can be incorporated into working early-warning systems.

In short, the proposed system has a superior equilibrium between statistical rigour, interpretability of the models and real life application as compared to the traditional models and previous studies. Its architecture is a step forward on the traditional deep learning to a more explainable, multi-task flood forecasting system, which will be able to provide data-supported climate resilience planning in Bangladesh.

5.6 Discussions

This section will give interpretation of the performance results of the model, and the practical and theoretical implications of the results, as well as limitations that were drawn during the study. The purpose of the discussion is to place the findings of the Hybrid CNN BiLSTM Multi-Task Learning model in the context of the overall research and development of flood predictions and early-warn systems.

5.6.1 Interpretation of Results

As the experimental results show, the suggested Hybrid CNN-BiLSTM model can successfully identify the complicated meteorological connections between rainfall, humidity, temperature anomalies, and other weather factors that determine flood occurrences. Having a ROC-AUC of 0.6355, PR-AUC of 0.1822, and Balanced Accuracy of 0.6095, the model ultimately outperformed traditional deep learning and machine learning baselines, which validate the benefits of learning spatial-temporal representation and attention based interpretability.

Combining both SMOTE oversampling and Focal Loss enabled recognition of minority-classes in a dataset with a majority of minorities (7.5) severely reduced bias in favor of the dominant class. Although the overall sensitivity (0.157) is relatively small, it indicates a realistic level of detection in highly unequal environmental data, where the signal of a flood is, by definition, rare and complicated. A threshold analysis also showed that the best trade-off between detection reliability and false alarms was achieved at the balanced-accuracy threshold (=0.58) - which is an essential consideration of operational flood early-warning systems.

The multi-task extension making a prediction of the Bangladesh Hydrometeorological Index (BHMI) based on a regression gave more contextual information, and therefore allowed the model to model the risk of floods in a continuous form and not a binary one. Though the correlation between predicted and observed values of BHMI was rather weak ($r = 0.054$), this element yielded to improved shared feature learning and easier classification decision boundaries.

5.6.2 Implications of the Study

There are a number of practical and scientific implications of the results. In a practical perspective, the hybrid model shows that deep learning can be used to predict floods in Bangladesh based on data, and because the traditional hydrological models fail to work in rural areas (fertile areas are infrequently covered by stations), unusual responses and inadequate ground data, the hybrid model has proven its utility. The fact that the system combines explainable mechanisms with meteorological features enables policy makers and agencies to have confidence in the decisions made because they are able to understand why certain predictions proceed as expected.

Scientifically, the study becomes part of the growing literature on multi-task learning of environmental prediction, providing a framework of predictive and attributable features, based on a causally consistent temporal model. The prediction performance and the ability to explain the results by incorporating the attention and feature gate methods not only allowed the improvement of the prediction, but also allowed to identify which climatic factors which cause the flood-like conditions are considered to have the greatest impact. These explicable results can inform future hydrology, inform infrastructure planning and amplify climate adaptation.

5.6.3 Limitations and Future Considerations

Although it had a promising performance, a few limitations were found. First, the model recallability in flood events is relatively low, as it demonstrates the fact of extreme imbalance and noise of historical meteorological data. Additional developments can be made through the use of data augmentation, transfer learning, or event-based resampling to enhance rare cases of floods.

Second, the BHMI regression head correlated poorly, probably because of nonlinear relationships as well as partial interaction between the meteorological and hydrological processes. There may be more inputs of hydrology (e.g., river discharge, soil moisture, satellite precipitation) or graph-based spatial models to enhance the regression performance in future work.

Third, station-level aggregated data were used in training and validation the model and this might not reflect micro-scale spatial variability or urban flash-flood behaviour. Generalizability could be better with the use of remote-sensing data, and high-resolution spatial grids. Lastly, the findings, although statistically solid, rely on a country-specific set of data; external validation of the model with regional or cross-border data would enable, further, to determine the reliability and the flexibility of the model.

5.6.4 Summary

To conclude, the Hybrid CNN-BiLSTM Multi-Task model is an approach to the topic of interpretable, data-driven flood prediction with a methodological and practical value. It's results verify that deep learning has the potential to supplement the conventional hydrological forecasting methods with a broader range of meteorological interactions as well as explainable results. Although some of the limitations still exist, especially in terms of sensitivity and the quality of BHMI regression, the results provide a solid basis on how explainable and multi-scale flood forecasting systems in Bangladesh and other at-risk areas can be developed in the future.

Chapter 6

Conclusion

6.1 Summary of Findings

This study was intended to design, develop, and test a Hybrid CNN-BiLSTM Multi-Task Learning model to predict weather-related floods in Bangladesh to combine both classification and regression tasks to increase the early warning capability. The analysis made use of 65 years of meteorological records (1948-2013) of 35 stations in the country and of about 21000 records underwent a sophisticated pipeline of data cleaning, transformation, reduction and feature engineering.

The results show that the model is able to effectively reproduce both short-term variations in the meteorology and long-term dependency on the climate, performing better than conventional statistical and machine learning baselines. The best CNN-BiLSTM hybrid has a ROC-AUC, PR-AUC, and Balanced Accuracy of 0.6355, 0.1822, and 0.6095, respectively, which is significantly better than that of Random Forest, XGBoost, and LSTM-only models. Class imbalance was effectively addressed by the Focal Loss and SMOTE and the sensitivity of the model to infrequent flood occurrences improved (recall 0.22).

Also, the regression line predicting the Bangladesh Hydrometeorological Index (BHMI) was found to have moderate correlation with observed data, which proves the possibility of the model to learn continuous trends in flood risks rather than binary results. The attention mechanism and feature gate module increased interpretability in that it revealed the relative significance of the meteorological variables and the rainfall, humidity, and temperature anomalies proved to be the most significant predictors.

In general, the results affirm that the multi-task hybrid architecture does not only improve predictive performance but also offers an interpretable and data-driven basis of early flood warning systems in Bangladesh. The explainable deep learning methods are incorporated to make sure that the model predictions can be relied upon and implemented in climate resilience planning and management of disasters.

6.2 Contributions to the Field

The work has a number of major contributions to the realms of hydrometeorological modeling, machine learning, and disaster reduction through the enhanced approaches of deep learning architecture, explainable and data-efficient design, and principles.This two-port design is a step forward in the field, connecting discrete flood detection and continuous risk quantification, making early-warning systems more subtle and operationally useful.

Second, the architecture of the model brings about a methodological innovation due to the integration of feature gating and a mechanism of attention. These elements help to improve interpretability and automatically regulate the effect of meteorological variables, including precipitation, humidity, and temperature deviations, of varying time scales. This unification fills the gap between black box neural networks and interpretable modeling of the surrounding environment, to give a worthwhile insight into the contribution of weather patterns to the risk of floods.

Third, the study adds the systematic data preprocessing pipeline that is designed to meet the needs of long term meteorology data, such as sophisticated feature engineering, class-imbalance repair, and validation of temporal causality. Such a pipeline provides a generalizable template that can be generalized to other climatic settings or natural-hazard analyses where data imbalance and time series dependencies are standard issues.

Fourth, the model, when exhibiting competitive performance (ROC-AUC=0.64, PR-AUC=0.18), in the case of harsh conditions of class imbalance, proves that deep learning can be effectively used in low-resource climate applications. This finding is part of the accruing literature supporting the use of AI-based decision support systems to facilitate development of places such as those in flood-prone South Asian regions.

Lastly, explainable AI elements and maximization of multi-task illustrate a novel avenue of data-driven climate resilience studies, demonstrating that interpretable, hybrid AI can trade-off prediction accuracy and scientific transparency. All of these contributions enhance the crossroads between environmental science and artificial intelligence and prepare the future research into adaptive flood forecasting and climate-risk communication.

6.3 Recommendations for Future Work

Although the Hybrid CNN–BiLSTM Multi-Task Learning framework proposed has shown a good performance in flood forecasting and hydrometeorological risks prediction, there are still a number of opportunities to be exploited in its further development and enhancement.

To begin with, the data used in the future studies should include more variables of the environment, e.g., soil moisture, river discharge, vegetation indices (NDVI) and precipitation derived by satellites (e.g., GPM or TRMM). The inclusion of these hydrological and remote-sensing parameters may give a more comprehensive picture of surface and subsurface processes of water that will enhance the generalization of the models to various climatic zones.

Second, it can be tested in the future that spatially explicit modeling can be carried out based on ConvLSTM or Graph Neural Networks (GNNs) to reflect spatial correlations between weather stations and river basins. This would allow the model to be not only able to learn temporal relationships but also the spatial propagation of flood signals, which would greatly increase the predictive accuracy of regional predictions.

Third, the explainability framework may be expanded by applying SHAP (Shapley Additive Explanations) or Layer-wise Relevance Propagation (LRP) to explain the model even more deeply both in terms of time and features. These methods might provide policy-makers with practical information because they will calculate the impact of particular climatic variables that, among others, have the strongest effect on flood risks in a certain area or during certain times of the year.

Fourth, the future studies may include the ensemble and transfer learning techniques to integrate the predictions of various deep learning models or use pretrained models in a similar hydrological setting. The strategies may speed up convergence, enhance robustness and easily reuse the models in areas that have low data access.

Fifth, the regression component used to predict the Bangladesh Hydrometeorological Index (BHMI) is a challenge that is yet to be improved. The regression loss function could be optimized in the future and hybrid physical-AI coupling could be considered or the index could be calibrated with real-time hydrological observations to enhance the physical interpretability and relationship with observed magnitudes of floods.

Lastly, the practical implementation of the model would like to be integrated with real-time forecasting systems, cloud-based dashboards and government early-warning systems like the ones operated by the Bangladesh Water Development Board (BWDB) and the Flood Forecasting and Warning Centre (FFWC). This would turn the proposed model into a research prototype into a decision-support tool that would be able to inform disaster preparedness initiatives, resource allocation, and community resilience initiatives.

To conclude, the improvements of the model to include a more diverse data, spatial modeling, interpretability, and real-time applicability should be considered in the future, which will enable the model to become a fully deployable AI-driven flood forecasting system that can be used to support adaptive climate resilience in Bangladesh and other countries.

Bibliography

[1] S. M. Vicente-Serrano, G. Van der Schrier, S. Beguería, C. Azorin-Molina, and J. I. Lopez-Moreno, "Contribution of precipitation and reference evapotranspiration to drought indices under different climates," *Journal of Hydrology*, vol. 526, pp. 42–54, 2015, Drought processes, modeling, and mitigation, ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2014.11.025 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169414009305

[2] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (lstm) networks," *Hydrology and Earth System Sciences*, vol. 22, pp. 6005–6022, 2018, Received: 4 May 2018; Revised: 25 September 2018; Accepted: 13 November 2018; Published: 22 November 2018, ISSN: 1607-7938. DOI: 10.5194/hess-22-6005-2018 [Online]. Available: https://hess.copernicus.org/articles/22/6005/2018/

[3] A. Mosavi, P. Ozturk, and K.-w. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, 2018, ISSN: 2073-4441. DOI: 10.3390/w10111536 [Online]. Available: https://www.mdpi.com/2073-4441/10/11/1536

[4] V. Nandhini and M. S. G. Devasena, "Predictive analytics for climate change detection and disease diagnosis," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, 2019, pp. 1152–1155. DOI: 10.1109/ICACCS.2019.8728537

[5] V. Ramachandra, *Causal inference for climate change events from satellite image time series using computer vision and deep learning*, arXiv preprint arXiv:1910.11492, 2019.

[6] M. Moishin, R. C. Deo, R. Prasad, N. Raj, and S. Abdulla, "Designing deep-based learning flood forecast model with convlstm hybrid algorithm," *IEEE Access*, vol. 9, pp. 50 982–50 993, 2021. DOI: 10.1109/ACCESS.2021.3065939

[7] R. J. Wilson et al., "Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in british butterflies," Dec. 2021. DOI: 10.1101/2021.12.21.473511 [Online]. Available: http://biorxiv.org/lookup/doi/10.1101/2021.12.21.473511

[8] A. Gupta, D. N. Dwivedi, and S. K. Maurya, "Climate change monitoring using remote sensing, deep learning, and computer vision," *Webology*, vol. Volume 19, pp. 3599–3610, No. 2 2022, ISSN: 1735-188X. [Online]. Available: http://www.webology.org/abstract.php?id=1708

[9] U. Iqbal, M. Z. B. Riaz, J. Barthélemy, N. Hutchison, and P. Perez, "Flood-borne objects type recognition using computer vision to mitigate blockage originated floods," *Water*, vol. 14, p. 2605, Aug. 2022. DOI: 10.3390/w14172605

[10] B. Katneshwarkar, A. Abhishek, and T. Kinouchi, "A multivariate drought index for seasonal agriculture drought classification in semiarid regions," *Remote Sensing*, vol. 14, no. 16, p. 3891, Aug. 2022, ISSN: 2072-4292. DOI: 10.3390/rs14163891 [Online]. Available: https://www.mdpi.com/2072-4292/14/16/3891

[11] T. Ladi, S. Jabalameli, and A. Sharifi, "Applications of machine learning and deep learning methods for climate change mitigation and adaptation," *Environment and Planning B: Urban Analytics and City Science*, vol. 49, no. 4, pp. 1314–1330, 2022. DOI: 10.1177/23998083221085281 eprint: https://doi.org/10.1177/23998083221085281. [Online]. Available: https://doi.org/10.1177/23998083221085281

[12] "Tackling climate change with machine learning," *ACM Computing Surveys*, vol. 55, p. 96, 2 Feb. 2023, ISSN: 15577341. DOI: 10.1145/3485128/SUPPL_FILE/3485128-VOR.PDF [Online]. Available: https://dl.acm.org/doi/10.1145/3485128

[13] X. L. Y. Li and N. Tan, "Climate change prediction by combination prediction method based on deep learning models," in *2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, Guangzhou, China, 2023, pp. 1508–1512. DOI: 10.1109/IAECST60924.2023.10502965

[14] Y. Zhang, Z. Zhou, J. Thé, S. Yang, and B. Gharabaghi, "Flood forecasting using hybrid lstm and gru models with lag time preprocessing," *Water*, vol. 15, Nov. 2023. DOI: 10.3390/w15223982

[15] N. Chenmin, M. F. Marsani, and P. S. Fam, *Flood prediction based on feature selection and a hybrid deep learning network*, https://doi.org/10.2166/wcc.2024.559, Journal of Water and Climate Change, 15(3), 1245–1261. https://doi.org/10.2166/wcc.2024.559, 2024.

[16] Y. Gong, Y. Zhang, F. Wang, and C. Lee, "Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data," *Applied and Computational Engineering*, vol. 99, pp. 168–174, 1 Nov. 2024, ISSN: 2755-2721. DOI: 10.54254/2755-2721/99/20251758 [Online]. Available: https://www.ewadirect.com/proceedings/ace/article/view/17393

[17] M. M. Jassim et al., "Deep learning approaches for predicting climate change impacts: An empirical analysis," in *2024 36th Conference of Open Innovations Association (FRUCT)*, 2024, pp. 481–492. DOI: 10.23919/FRUCT64283.2024.10749950

[18] D. Kim, H. Han, H. Lee, Y. Kang, W. Wang, and H. S. Kim, "Predicting flood water level using combined hybrid model of rainfall-runoff and ai-based models," *KSCE Journal of Civil Engineering*, vol. 28, pp. 1580–1593, 4 Apr. 2024, ISSN: 1226-7988. DOI: 10.1007/S12205-023-1147-0

[19] J. Li, G. Wu, Y. Zhang, and W. Shi, "Optimizing flood predictions by integrating lstm and physical-based models with mixed historical and simulated data," *Heliyon*, vol. 10, no. 13, e33669, 2024, ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2024.e33669 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844024097007

[20] W. C. Liu and W. C. Huang, “Evaluation of deep learning computer vision for water level measurements in rivers,” *Heliyon*, vol. 10, e25989, 4 Feb. 2024, ISSN: 2405-8440. DOI: 10.1016/J.HELIYON.2024.E25989

[21] A. R. Rumee, “Evaluation of deep learning models for flood forecasting in bangladesh,” *Journal of Computer Sciences Institute*, vol. 34, pp. 89–97, Mar. 2025. DOI: 10.35784/jcsi.6773 [Online]. Available: https://ph.pollub.pl/index.php/jcsi/article/view/6773

[22] A. Venkatesan, “A hybrid cnn-lstm predictive model deployed federated learning model for advanced flood prediction systems to forecast coastal region of smart cities,” *Global NEST Journal*, vol. 27, no. 6, 2025. DOI: https://doi.org/10.30955/gnj.07511 [Online]. Available: https://doi.org/10.30955/gnj.07511

[23] N. Climate.gov, *Climate change and the 1991-2020 u.s. climate normals*, n.d. [Online]. Available: https://www.climate.gov/news-features/understanding-climate/climate-change-and-1991-2020-us-climate-normals

[24] N. Climate.gov, *How is the current global warming trend different from previous warming periods in earth’s history?* n.d. [Online]. Available: https://www.climate.gov/news-features/climate-qa/how-current-global-warming-trend-different-previous-warming-periods-earths