# A machine learning approach to predict the heatwave in Bangladesh region due to global warming

By

Md. Rakibul Hasan
20201155
Nur-E-Jannat
21301744
Kanak Roy Shanda
21301743
Umma Faria Rahman
20301346

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
June 2025

# Declaration

It is hereby declared that

1. The thesis submitted is 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:**

Rakibul Hasan

Md. Rakibul Hasan
20201155

Nur-E-Jannat

Nur-E-Jannat
21301744

Kanak

Kanak Roy Shanda
21301743

Faria Rahman

Umma Faria Rahman
20301346

# Approval

The thesis titled "A machine learning approach to predict the heatwave in Bangladesh region due to global warming" submitted by

1. Md. Rakibul Hasan(20201155)
2. Nur-E-Jannat(21301744)
3. Kanak Roy Shanda(21301743)
4. Umma Faria Rahman(20301346)

Of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on June 18, 2025.

**Examining Committee:**

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

# Abstract

Since heatwaves are becoming more frequent and intense all over the world due to global warming, accurate heat wave predictions become essential for ensuring global safety. According to the Ministry of Environment, Forest and Climate Change, Bangladesh is one of the vulnerable countries that is being affected by global warming. To the best of our knowledge, no remarkable machine learning based research has been conducted yet on the actual dataset from Bangladesh. In Bangladesh, heatwaves typically occur during the pre-monsoon season. It has been observed in various seasons in recent years. The heatwave increased the health risk of humans as well as the death rate. Accurate early warning of heatwaves will be beneficial to keep human life safe. This study employs multiple machine learning approaches used to predict the division-wise heatwave based on meteorological data recorded by the Bangladesh Meteorological Department (BMD) from 35 stations between 2000 and 2024. Heatwave events are calculated with the operational definition of heatwaves used by BMD based on three or more consecutive days with maximum temperatures of 36°C or higher. This research explores ML binary classification models for classifying heatwaves, multi-regressor and hybrid deep learning models to provide early heatwave alerts using BMD recommended post-thresholding on forecasted temperature. The core objective of this research is to contribute to Bangladesh's heatwave forecasting system, which will help with upcoming environmental risk management initiatives.



**Keywords:** Heatwaves, Global warming, Global safety, Early warning, Forecasting, Machine learning, Meteorological variables, Feature engineering

# Acknowledgement

First of all, we are grateful to the almighty Allah that we were able to complete our thesis without any major obstacles.

Secondly, we acknowledge with gratitude to our supervisor Dr. Md. Ashraful Alam, for his guidance and support throughout the journey. Whenever we needed help, he was always there to assist us.

Finally, we might not be able to finish this without the help of our parents. We are almost done with graduation by their kind support and prayers.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Heatwave definition

A heat wave is a prolonged period of high temperatures; it occurs in specific regions during specific seasons. It can affect both urban and rural areas, and it may occur in both dry and humid climate conditions [16] . In Bangladesh, a heatwave is defined by the extended periods of extremely high temperature which is considered as 36°C, and can be accompanied by the high humidity, which makes the heat feel even worse [30]. In Europe, they consider heatwaves when the temperature exceeds the 95th percentile of normal temperature from June to August for five consecutive days [29]. Bangladesh considers 3 consecutive days or more for heatwaves [26]. WHO/UNICEF (2021): Recommends nighttime temps ≥20°C as a risk threshold for heat stress in tropical climates due to lack of recovery [12]. Also, high relative humidity and low wind speed target heat stress. If we miss any of this like Min_Temperature, Relative_Humidity, Wind_Speed, then it may mismatch with prediction heatwaves. For that reason heat stress, health risk may increase. Different countries use different thresholds and consecutive days for heatwaves like India used two consecutive days or more[15]. The threshold we used in our research is only for the Bangladesh region, following the Bangladesh Meteorological Department.

| Max_Temperature at a day (Celsius) | Intensity |
|---|---|
| 36-38°C | Mild |
| 38.1-40°C | Moderate |
| 40.1-42°C | Severe |
| >42°C | Extreme |

Table 1.1: Heatwave Classification Threshold of Bangladesh Meteorological Department

## 1.2 Motivation

Due to global warming, the whole world is affected by heatwave phenomena every year. Emission of greenhouse gases and the urban heat island effect resulting from industrialization and urbanization are major reasons for heat waves that are directly influenced by human activities. In addition, oceanographic phenomena like El Niño, shifting of the jet stream, and creation of heat domes by high pressure systems are the crucial causes of excessive temperature. Heatwaves are more intense during the summer season (April to June) in Bangladesh. Throughout the past few years, Bangladesh has experienced an increase in summer heatwaves. According to the report of the Bangladesh Meteorological Department (BMD), the level of maximum and minimum temperature is increasing, but the maximum temperature is rising rapidly. From May 2023 to May 2024, an extra 76 days of extreme heatwaves were recorded, which is hotter than 90% of normal temperatures based on the previous 30 year records [27]. With average temperatures of 40° to 42° Celsius throughout all the districts, 2024 is set to be the warmest year to date. As per the report of the Environment and Social Development Organization (ESDO), the temperature has increased by 0.5° Celsius in the last 44 years in Bangladesh, and it will reach 1.4° Celsius by 2050 [25]. Heatwaves cause various issues in day-to-day life, including health problems and even deaths. The heatwave is brutal for slum dwellers, refugees, and outdoor labourers like farmers and day labourers. A large number of people have died of heatstroke in the month of April 2024 in Bangladesh. Since newborns, babies, and young children are particularly susceptible to heat-related illnesses like heatstroke and dehydration, UNICEF has highlighted the major concerns about the current rising heatwave in Bangladesh and the risk it creates for the health and safety of children [28]. Bangladesh's electricity demand has risen because of the heavy use of air conditioners, coolers, and fans during the heatwave. Besides, the garment industry is affected by the power crisis due to the heatwave. As Bangladesh was the second-largest exporter of ready-made garments in 2023, the power crisis caused by the heatwave had a significant economic impact. The Bangladesh Rice Research Institute (BRRI) reported that 210,000 hectares of paddy were destroyed as a consequence of the heatwave in 2021 [31]. In order to survive in this adverse condition of Bangladesh and to reduce the sufferings, it is an absolute necessity to take urgent steps to emphasise regular environmental monitoring and research. As this problem is an ongoing process and a future threat to all, the research is motivated to make a potential contribution to this field. In the context of predicting tasks, ML models are chosen to get better accuracy. To mitigate the effects of climate change and adapt to it, predictions of heat waves in Bangladesh are being made division-wise because coastal areas and urban areas have different effects of heat waves in Bangladesh. It aims to ensure human safety, public health, agricultural sustainability, economic sustainability, prevention of food grain spoilage, and adaptation to climate change. In short, the paper aims to research different machine learning techniques to predict heat waves in Bangladesh division-wise, ensuring wiser decision-making for the environmentalists and communities.

## 1.3 Problem Statement

Nowadays, heatwaves have become a threat to life on earth. Excessive heat waves have occurred every 50 years in the past, but in the present days people are experiencing heat waves frequently due to global warming [9]. Current heatwave prediction systems are not excellent at observing serious effects, which is especially true in tropical regions. These areas face unique weather conditions and lack sufficient data, which makes prediction difficult. Fewer studies have been conducted to predict heat waves using machine learning. A study compared different existing machine learning models to determine which has the highest accuracy in predicting heat waves. The author searches specific keywords and selects studies from 2016 to 2021. 22% of the paper is based on classical ML algorithms, 18% depend on deep learning, 41% use regression models, and others use different ML techniques. This paper demonstrates a lack of adequate research using ML techniques compared to regression models, which are not explored thoroughly. Additionally, the amount of work to predict heat waves is low, even if a heatwave is a threat to the world [14]. Notably, there is no remarkable research found to predict heat waves in Bangladesh using machine learning. As less significant work has been conducted on heatwave prediction in many countries using ML techniques, the feasibility of the resource makes this field of study difficult and requires complex analysis. The Excess Heat Factor (EHF) is a new way to measure heat waves using only temperature as data. It calculates intensive heatwaves by averaging daily mean temperatures over three days. As the author considers only the temperature data, the system is not reliable for other regions to predict heat waves accurately [5]. A study finds better results by using the Support Vector Machines (SVM) model than Random Forest (RF) and Artificial Neural Network (ANN) in Pakistan. The limitation of the paper is not ensuring this model will give better results for other regions [11]. As a result, the existing model could not accurately predict the accurate heatwave days in different regions. Those were not considering region-wise climatological data, which is challenging to collect. Previous heatwave prediction research has two major limitations. Firstly, the papers do not take into account regional climate variations because they rely on aggregated data at the divisional level, ignoring local microclimates. Secondly, the research mainly used binary classification methods that struggle due to the extreme class imbalance. These issues arise from data limitations such as missing records and methodological flaws. Although classic machine learning algorithms like logistic regression and random forests are commonly used for heatwave classification, their inability to manage temporal dependencies and rare events makes them unsuitable for operational forecasting. Additionally, deep learning approaches have data limitation problems and extreme data imbalances in regional heatwave predictions compared to binary classical ML models.

This research aims to address these challenges by investigating how machine learning models can leverage historical meteorological data to accurately predict heatwaves in Bangladesh with the optimal solution of an early warning system, considering the growing impacts of global warming. This research proposes an idea to predict heat waves in Bangladesh by categorising a seasonal and sub-seasonal (S2S) forecasting system where ENSO effects are integrated slowly. As S2S involves complex atmospheric, marine, and land data that are slowly integrated with phenomena like heatwaves and often influenced by the long-term time scale, it is challenging work to predict heatwaves more accurately than the short-term time scale. [6] Due to the daily maximum temperature in this dataset, predictions are constrained to daily resolution, making short-term forecasting feasible. This research explores ML binary classification models for classifying heatwaves, multi-regressor and hybrid deep learning models to provide early heatwave alerts using BMD recommended post-thresholding on forecasted temperature. This method gives meteorologists detailed forecasts while considering the growing impact of climate change on the frequency and intensity of heatwaves in Bangladesh. As mentioned earlier, classical ML algorithms are generally used for binary classification to predict heatwave days, which has greater limitations. This research tackles these challenges by evaluating different machine learning techniques to provide an effective solution that follows World Meteorological Organization (WMO) guidelines.

## 1.4 Research Objectives

The paper is aimed at predicting heatwaves division-wise in Bangladesh using different machine learning techniques, incorporating historical meteorological datasets from 35 weather stations of the country in different areas. The objectives of the research are:

1. To understand the definition of heatwaves for the Bangladesh region and compare the heatwave thresholds according to the national and international meteorological departments.
2. To learn how meteorological variables and dynamically engineered features influence the formation of heat days and forecasting.
3. To explore different approaches for heatwave prediction using different machine learning techniques, which satisfy the rules of operational agencies (ECMWF, BMD).
4. To evaluate the results of the binary classification models, regression, and hybrid deep learning models through different metrics.
5. To gain knowledge about the limitations of extreme class imbalance scenarios for the Bangladesh region and identify the increasing costs of false alarms.
6. To explore how the different approaches dynamically predict and compare heatwave streaks from forecasting temperature.
7. To extend the framework for a seasonal and sub-seasonal (S2S) forecasting system, incorporating ENSO anomalies and engineered features of ENSO.
8. To develop a heatwave prediction dashboard for BMD to show graded alerts based on predicted heatwave streaks from the most optimal performance model to predict heatwaves in Bangladesh using historical meteorological variables by incorporating dynamic feature engineering.

# Chapter 2

# Literature Review

The purpose of this section is to critically review the various techniques used in previous relevant work in the field of machine learning on heatwave prediction. This research analyzes different machine learning techniques used for the target result achieved and shows how the prediction of heatwaves in Bangladesh has its own specific challenges due to the feasibility of the definite weather dataset and the necessary research papers in this field using machine learning tools.

A study identifies some limitations to improve the advance predictions of summer heatwaves in central Europe by using statistical and machine learning techniques to forecast heat waves more than 2 weeks in advance. Averaging latitude and longitude boxes results in the missing of important information because the initial dataset was too large. The author suggested that the size of the boxes be small to improve the accuracy. Besides, the author hypothesized that deep learning models could perform better in larger datasets than classical machine learning techniques in this climate data field. Additionally, the proposed classical ML model depends on daily input data, which generates data scarcity; the performance is lower with longer lead times [19].

The authors face challenges in defining a GNN model to predict regional heatwaves. By using the daily historical weather records from ground weather stations, the performance of the GNN model is reliable. Mentionably, because of capturing dynamic values for daily data, the computational complexity is higher in this proposed GNN model. The Graphic Neural Network (GNN) model couldn't result efficiently in real-time application to predict regional heatwaves, and the model may not be universally applicable due to variations in geographical location and weather conditions. To improve the GNN model, the author suggested incorporating hybrid modelling and reliable features for extreme weather events [17].

A study used daily temperature data from 27 locations in Iran to predict heatwave days through a hybrid machine learning model. Key predictors included temperature, humidity, wind components, and geopotential height at multiple pressure levels. After applying principal component analysis for dimensionality reduction, the AdaBoost Regression with Decision Tree (ABR-DT) model outperformed both Decision Tree and Random Forest in terms of CC, RMSE, and MAE. The approach demonstrates the effectiveness of hybrid ML models in improving heatwave prediction for climate adaptation [13].

A study investigates a significant change in the relationship between oceanatmospheric variables and heatwave days (HWDs), particularly in the lower atmosphere at 925 hPa. The SVM-based (Support Vector Machine) model gave high accuracy in predicting heat waves in Pakistan. SVM achieved a high $R^2$ value of 0.89, indicating this has strong predictive capability, while RF (random forest) had $R^2$ values of 0.84 and ANN (Artificial Neural Network) had $R^2$ values of 0.73. The normalized root mean square error (NRMSE%) for SVM was significantly lower at 32.6%, compared to 152% for RF and 53.5% for ANN, providing higher accuracy of the Support Vector Machine algorithm. Although the SVM-based model provides robust forecasting for heatwaves in Pakistan, the author suggested further optimization using the application of advanced machine learning techniques [11].

The researchers used SST, soil moisture and 200 hPa geopotential height anomalies to develop a machine learning model to forecast daily maximum temperature (Tmax) 10 days in advance over India (March–June). The AdaBoost with MLP model performed best in April and May, outperforming traditional methods like persistence and CFS, but showed poor performance in March and June. Although MLP is popular in climate prediction, its sensitivity to architecture (such as neurons and activation functions) limits generalizability. The study suggests hybrid models that combine numerical weather prediction with deep learning to improve results [18].

The study focused on predicting heat waves during the summer season over the north, northwest, central, and coastal regions. Only Tmax data had been used in this paper for real-time prediction of heatwaves. For this reason, the author highlights the necessity of research on this topic in the future, considering the other value to predict the real-time heatwave properly [10].

There is another study based on an extensive dataset comprising over 11,000 observations from nearly a thousand sites in eight different countries. The research is done in spring to make the results robust and minimize the seasonal fluctuations. In this study, the daily highest temperature threshold is set at the 90th percentile of common summer temperatures, avoiding the fixed temperature. So that more regional fluctuations and temporal shifts in extreme heat events could be captured. For analyzing the data, the researchers used a probability model that considers the interaction not only of the direct effects of temperature but also of the surrounding landscape [23].

A study introduces the Excess Heat Factor (EHF) to better quantify heatwave intensity and related health impacts. Unlike traditional methods relying only on maximum temperature, EHF combines two components: EHI-sig (for short-term heat spikes) and EHI-accl (for long-term acclimatization). The strong correlation between EHF and health outcomes highlights its value as a standardized metric for heatwave assessment and its integration into public health strategies and warning systems [7].

A study uses numerical models and data from radar and satellites to predict high-impact weather events. Although tropical cyclones track forecasts, it is still challenging to anticipate storm intensity. Integrating real-time data into models creates challenges, like predicting sudden weather events such as thunderstorms and flash floods. The uncertainty in multiple prediction models makes reliable forecasting is difficult, especially for localized, small-scale events like tornadoes. Despite advancement, more accurate data assimilation and model resolution are still required to improve short-term forecasts of severe [3].

In a study, the LTSM networks perform better for capturing temporal dependencies to predict heat waves using Recurrent Neural Network (RNN) based on deep learning. Because of the limitations in complexity and computation of LTSM, it requires additional resources to implement in a real-time system. Additionally, overfitting the model to the training data can be created for poor generalization to unseen data because the architecture of the model is too complex. Lastly, there is a limitation to generalizing various geographical regions considering climate conditions in the model [22].

A study explores how the study applies a machine learning technique called MARS (Multivariate Adaptive Regression Splines) to define heatwave thresholds that are more closely linked to real health outcomes. Unlike traditional methods that rely on fixed temperature cutoffs such as labeling any two consecutive days over 33°C as a heatwave this approach lets the data speak for itself. By analyzing emergency room visit records, the study identifies specific “tipping points” where health risks rise sharply, like a two-day average maximum temperature of 32.58°C or a heat index of 79.64. The research also reveals how heat vulnerability varies across age groups, occupations, and even times of the year, emphasizing the need for personalized, multi-level heat warning systems [8].

In short, advanced machine learning (ML) techniques can find patterns in huge datasets that traditional models might overlook. It offers a more sophisticated and dependable way to forecast heatwaves. Extreme weather event forecasts can be made more accurate and reliable by using advanced machine learning (ML) to improve data assimilation, optimize ensemble predictions, and improve probabilistic forecasts.

A study aims to define heatwaves in Bangladesh based on extreme day and night temperatures lasting at least three days, showing a strong link between hot nights and increased death rates. Using national mortality data, the definition is validated, and predictive patterns like low soil moisture and wind changes 10–30 days before are identified, offering potential for early warnings. However, limitations include short forecast windows, lack of real-time soil data, limited historical records, and underexplored urban vulnerability [6].

A study highlights that machine learning models like Random Forest, CNN, and LSTM are widely used in heatwave prediction, with some studies exploring hybrid models and alternative data sources like search trends. While regression models have shown good results using weather variables, most existing approaches are region-specific, data-heavy, and lack real-time or socio-economic integration [14].

These limitations suggest that light weight models like XGBoost or Random Forest, as used in this research, may offer a more scalable solution for a real-time, localized heat wave forecasting system in Bangladesh.

In a study from China that focused on heatwave classification, researchers used an optimized binary XGBoost model to forecast heatwave days using hourly meteorological station data. The model integrated predictors such as dew point, solar radiation, and wind gusts to capture fine-grained environmental interactions. It performed significantly better than traditional logistic regression and SVM models, especially under high humidity conditions. However, the study noted a major limitation: the performance degraded when applied to other regions outside the training dataset, suggesting overfitting to regional climate norms [20].

This emphasizes the importance of incorporating generalization techniques for models intended for cross-regional deployment like in Bangladesh. These limitations suggest that simpler yet powerful models like XGBoost or Random Forest, as used in this research, may offer a more scalable solution for a real-time, localized heat wave forecasting system in Bangladesh.

A study in India used a hybrid CNN-LSTM model to forecast daily maximum temperature anomalies during the pre-monsoon season. CNN captured spatial patterns, while LSTM addressed temporal trends. The model performed well in April and May, outperforming persistence and CFS forecasts, but struggled in other months, showing limits in generalization. [24].

Traditional RNN-based models like LSTM are widely used in time-series forecasting, but newer methods are pushing the boundaries further. A 2022 study explored the use of attention-based deep learning, inspired by the Transformer architecture, to predict the severity and duration of heatwaves across Asia. By dynamically focusing on the most relevant weather signals such as sudden changes in humidity or wind pressure the model achieved more stable and accurate forecasts than conventional LSTM or CNN-LSTM setups [21].

While this approach is more complex, it opens a future direction for this research to consider attention mechanisms that improve interpretability and robustness in heatwave forecasting.

# Chapter 3

# Methodology

This research aims to develop a robust heatwave prediction system for Bangladesh using machine learning models. Due to the socio-economic and health risks posed by heatwaves, accurate forecasting is crucial for early warning systems. The study leverages capturing temporal dependencies and seasonal patterns in historical weather data by dynamic feature engineering. In this research, dynamically engineered features are used in the models for more robust predictions that outperform conventional physics rules. This study uses a machine learning (ML) regression approach to predict heatwave streaks from forecasted temperatures in Bangladesh. It addresses the problems of traditional binary classification methods when dealing with extreme class imbalance. Firstly, the paper assesses classic ML classifiers, including XGBoost, Random Forest, Logistic Regression, and Decision Tree for detecting heatwaves. Analyzing their performance using accuracy scores, classification reports, and confusion matrices. Secondly, a hybrid CNN-LSTM model is applied. CNNs extracted spatial patterns from input features, while LSTMs captured temporal dependencies to predict heatwaves over several days. The architecture included stacked convolutional layers and bidirectional LSTMs. The model was trained with dropout regularization to avoid overfitting. Understanding the limits of direct classification and deep learning approaches for the Bangladesh region, this research adopts a hybrid strategy that combines continuous temperature predictions with changing post-processing thresholds. This helps identify heat waves based on intensity, duration, and streak detection. After that, compare models like Random Forest Regressor, CNN-LSTM, XGBoost Regressor, and Decision Tree Regressor (DTR) using metrics such as MAE, RMSE, and $R^2$. The research further processes the multi-regression outputs t+5 days with adaptive thresholds to turn probabilistic forecasts into practical heatwave alerts for the most recent lead time. This allows meteorologists to consider how climate change affects heatwave patterns. This two-step approach connects detailed temperature predictions with clear extreme weather detection, providing a scalable solution for climate resilience planning in Bangladesh which satisfies the climate science forecasting laws by the World Meteorological Organization (WMO).

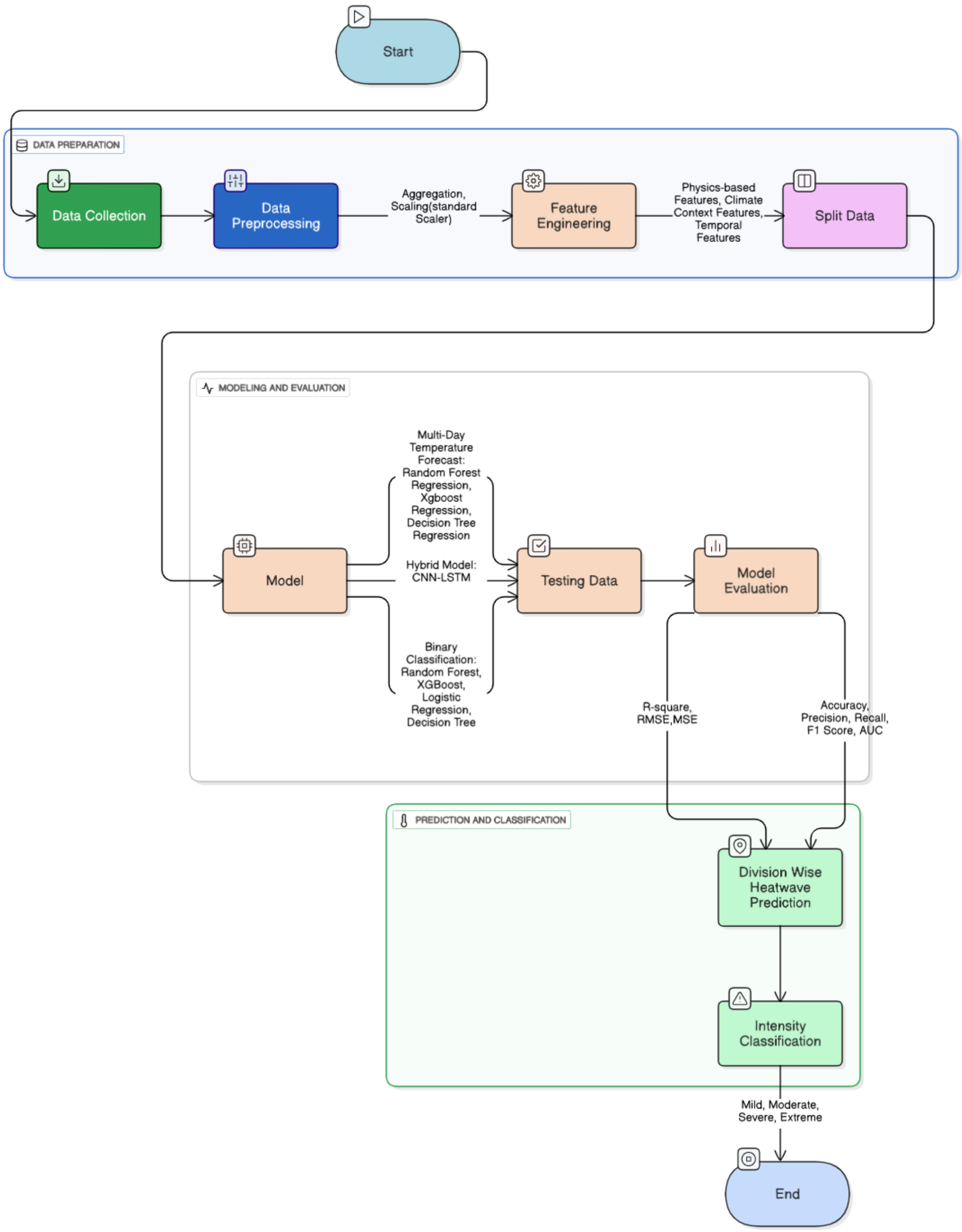


Figure 3.1: Workflow Diagram

# Chapter 4

# Dataset

## 4.1 Dataset Collection

Bangladesh Meteorological Department (BMD) provided the dataset used in this study, weather-related data collected from 35 stations across Bangladesh, spanning the years 2000 to 2024. The dataset contains 306,178 entries and 13 columns, providing comprehensive historical meteorological information for each station on a day, which is crucial for heatwave prediction modelling division-wise.
Key fields include:

- Station: Name of the weather stations across Bangladesh.
- Year, Month, Day: Temporal details of each observation
- Max_Temperature, Min_Temperature: Recorded daily extremes (°C)
- Relative Humidity: Measured in percentage (%)
- Rainfall: Total precipitation (mm)
- Sunshine_Hours: Duration of bright sunlight
- Cloud_Amount: Degree of cloud coverage
- Wind metrics in knots

## 4.2 Dataset Preprocessing

The dataset contains weather data collected from 35 different stations across Bangladesh over the years 2000 to 2024. As we mainly predict the heatwave by division wise for capturing regional trends, thus I group these station data into division for each day's value of all these parameters. The raw dataset underwent several preprocessing steps to ensure data quality and compatibility for machine learning models. The key steps included:

### 4.2.1 Initial Exploration and Cleaning

Initial data analysis validated the dataset (319,620 rows, 12 columns), assessed missing values, and computed descriptive statistics (central tendencies, dispersion, distribution) to optimize it for further analysis.

### 4.2.2 Handling Missing Value

The initial dataset comprised 34,659 missing values. Data cleansing involved addressing missing values, outliers, and inconsistencies. In this study, interpolation was employed for short data gaps, with consideration given to the time elapsed between non-missing values. Specifically, backward and forward interpolation methods were applied within each gap. Gaps exceeding 30 days , all rows within that long gap are dropped to prevent the introduction of significant errors from interpolating over extended periods of missing data. After handling missing values, the dataset was verified to contain 306,178 rows and 12 columns, as confirmed by shape and sample data checks.

### 4.2.3 Standardizing Column Names

To maintain consistency throughout the analysis and avoid potential errors due to naming discrepancies, all column names were standardized by converting them to lowercase and stripping extra spaces. This ensured uniform referencing in code and facilitated the seamless execution of transformations and model pipelines.

### 4.2.4 Mapping Stations to Divisions

This division-wise distribution of stations is crucial for analyzing regional patterns in heatwave occurrences across Bangladesh. The weather stations in the dataset were mapped to their respective divisions. Mymensingh Division is represented by 1 station, while Sylhet Division has 2 stations. Rangpur Division contains 3 stations, as does Rajshahi Division. Dhaka Division is represented by 4 stations, and Barisal Division also contains 4 stations. Khulna Division is represented by 5 stations and Chattogram Division has the highest number, with 9 stations.

### 4.2.5 Aggregate Data by Division and Date

Groups the DataFrame by both 'Division' and 'Date' and calculates the mean of all specified weather columns as index = False ensures that 'Division' and 'Date' remain as regular columns rather than becoming a multi-index. This effectively creates daily average weather data for each division.

### 4.2.6 Scaling

Use standardscaler to scale the features. standardize the numerical features to have a mean of 0 and a standard deviation of 1. Scaling numerical features ensures that no single feature dominates the machine learning model due to differences in magnitude. When working with machine learning algorithms, some models are sensitive to the scale of the input features. If one feature (like Max Temperature) has much larger values than another (like Wind Speed), the model will be biased toward the larger feature. Models like Linear Regression, Logistic Regression also benefit from feature scaling because they rely on optimization methods that involve gradient descent, which can be slower or less stable when features have different scales.

### 4.2.7 Create Target Column

A binary target column, is heatwave on the next day, was created using a threshold of 36°C, where 1 indicates a heatwave (max temp >= 36°C) and 0 indicates no heatwave (max temp < 36°C). This column was used as the target for classification and for multi-step prediction, creating Lagged Target Variables for forecasting temp then used earlier is heatwave or no heatwave target models use multitarget. Shifts the 'Max_Temperature' column forwarding by i days. Tmax+1 will be the max temperature of the next day, Tmax+2 the day after, and so on. These are the target variables for multi-step prediction. As previously mentioned, the 36°C threshold was chosen to enable early warnings and provide a realistic approach to enhance model sensitivity and statistical robustness for accurate heatwave prediction by using most recent temperature forecasting lead time 1 to make this multi-step prediction realistic.

### 4.2.8 Train-Test Split

For binary classification models and regression models, a 80:20 train-test split was used for both models, with 2000-2019 data for training and 2020-2024 for testing. This robust time series partitioning prevents future data leakage.

## 4.3 Feature Engineering

Extensive feature engineering was conducted to derive meaningful inputs for binary classification and regression models. In order to anticipate heatwaves, the generated characteristics encompass temporal patterns, thermodynamic interactions, and climate influences, all of which are difficult to manually describe using straightforward physics-based criteria. The engineered features are outlined below:

### 4.3.1 Physics-Based Features

The following indices were calculated in order to measure the combined effects of climatic variables on heat stress:

- Heat Index (HI): Estimates perceived temperature by combining air temperature and relative humidity.

  To account for the perceived temperature, the Heat Index ($HI$) was calculated. A simplified regression formula, based on the foundational work of Steadman on apparent temperature [1], was used. The formula is given by:

$$HI = 0.5 \times (T_F + 61.0 + ((T_F - 68.0) \times 1.2) + (RH \times 0.094)) \tag{4.1}$$

  where $T_F$ is the maximum air temperature in degrees Fahrenheit and $RH$ is the relative humidity in percent.

- Wet-Bulb Temperature (WB): Reflects the lowest temperature achievable via evaporative cooling, critical for assessing human heat stress. Used the following formula:

  In this study, the Wet-Bulb Temperature ($T_w$) was calculated using the empirical formula proposed by Stull [4]. The formula is given by:

$$\begin{aligned} T_w = T \cdot \arctan[0.151977 \cdot (RH + 8.313659)^{1/2}] \\ + \arctan(T + RH) - \arctan(RH - 1.676331) \\ +0.00391838 \cdot (RH)^{3/2} \cdot \arctan(0.023101 \cdot RH) - 4.686035 \end{aligned} \tag{4.2}$$

  where $T$ is the air temperature in degrees Celsius and $RH$ is the relative humidity in percent. This approximation is known for its computational efficiency.

- Apparent Temperature: Represents the ”feels-like” temperature by incorporating wind speed and humidity.

  The Apparent Temperature ($AT$), which estimates the perceived ”feels-like” temperature, was calculated. The model, adapted from the work of Steadman [1], incorporates air temperature, humidity, and wind speed. The calculation first requires determining the water vapor pressure, $e$ (in hPa):

$$e = \frac{RH}{100} \times 6.105 \times \exp\left(\frac{17.27 \times T}{237.7 + T}\right) \tag{4.3}$$

  where $T$ is the temperature in degrees Celsius and $RH$ is the relative humidity (%). The Apparent Temperature is then calculated as:

$$AT = T + 0.33 \times e - 0.70 \times WS - 4.00 \tag{4.4}$$

  where $WS$ is the wind speed in m/s. This formulation is notably used by the Australian Bureau of Meteorology.

- Temp-Humidity Interaction: Captures the multiplicative effect of high temperature and humidity.

  To capture the combined effect of temperature and humidity on perceived heat, an interaction term was engineered. This feature was created by calculating the product of the maximum temperature and the relative humidity:

  $$\text{Temp\_Humidity\_Interaction} = T_{max} \times RH \tag{4.5}$$

  where $T_{max}$ is the maximum temperature and $RH$ is the relative humidity. This term is designed to model the synergistic effect where high temperatures are more impactful in the presence of high humidity, a relationship that a linear model might not capture from the individual features alone.

- Wind Cooling Effect: Measures the cooling influence of wind under high-temperature conditions.

  The physiological effect of wind is highly dependent on the ambient temperature. In warm conditions, wind can provide a cooling sensation by increasing the rate of convective heat loss from the body. The principles governing this interaction are fundamental to modern thermal comfort indices [2].

  To model this relationship, a "Wind Cooling Effect" feature was engineered to capture the interaction between wind speed and temperature:

  $$\text{Wind\_Cooling\_Effect} = WS \times (30 - T_{max}) \tag{4.6}$$

  where $WS$ is the wind speed and $T_{max}$ is the maximum temperature (°C). This provides the model with a feature representing the temperature-dependent effect of wind.

### 4.3.2 Climate Context Features

Feature selection considered large-scale climate variability:

- ENSO (Southern Oscillation) phases, such as La Niña, Neutral, and El Niño are incorporated for each year.
- ENSO-anomaly interactions: Cross-features between ENSO anomalies and temperature-based variables (ENSO_Tmax=ENSO_anom× Max_Temperature).

### 4.3.3 Temporal Features

To capture how things change over time, we put together the engineered features:

- Lagged features (1-7 days): It captures short-term memory effects in weather (if the past 3 days were extremely hot, today is more likely to be a heatwave day), which helps the model to detect accumulated heat stress over multiple days. (Max_Temperature_lag1, Heat_Index_lag1)

- Rolling features (three- and five-day averages): Rolling-window stats gently smooth the day-to-day noise, letting longer weather trends rise to the surface. They make it easy to tell quick ups and downs from stretches of real heat that last for days. The feature mainly reduces the impact of single-day outliers, like a sudden temperature drop on a mild rainy day. (Max_Temperature_roll3, Heat_Index_roll3)

- Cyclical encoding: As heat waves follow the rhythm of the seasons, wrapped by calendar features like day of year and month in sine and cosine waves. As a result, the model captures December (12) and January (1) as neighbours instead of letting the raw number. It can then spot yearly trends, such as how Bangladeshi summers tend to peak between April and June.

- Climatological anomalies: Compared daily temperatures against long-term station-specific averages to capture unwanted heatwave events. Identifying extreme deviations, such as a day with a +5°C anomaly is more critical than a naturally hot day in summer. It distinguishes between normal seasonal heatwaves and unwanted heatwaves.

  Calculation:

  $$T_{\text{max\_Anomaly}} = \text{Max\_Temperature}_{\text{observed}} - \text{Climatology\_Tmax} \tag{4.7}$$

  Where Climatology_Tmax is the average Max_Temperature for each day_of_year across all years.

  Together, these engineered feature inputs help to learn the algorithms of the complicated, curved links to boost accuracy over physics rules.

# Chapter 5

# Applied Model and Analysis

## 5.1 Binarry Classification Approach

### 5.1.1 XGBoost

XGBoost is an open-source distributed machine learning library that makes use of gradient boosted decision trees, which is a boosting algorithm for supervised learning that makes use of gradient descent. It has been hailed for its efficiency, speed and scalability with large data sets.
XGBoost was configured with 100 boosting rounds and tuned for class imbalance using the scale_pos_weight parameter. This technique scales the gradient updates for the minority class during training, effectively compensating for imbalanced class distributions. Additionally, SMOTE was used prior to training to synthetically balance the dataset. The model was evaluated using the logarithmic loss (logloss) function.
The XGBoost model achieved a 96% accuracy, with a precision of 0.51, recall of 0.70, and an F1-score of 0.59 for the heatwave class. The ROC-AUC score of 0.9603 indicates excellent class separation. Compared to the Random Forest, XGBoost slightly improved recall while maintaining a similar level of precision. This performance affirms the model's capability in handling imbalanced classification problems and highlights its effectiveness in operational contexts requiring a balance of precision and recall.

### 5.1.2 Random Forest

An ensemble learning method that constructs multiple decision trees during training and then aggregates their predictions to classify the instances as 0 or 1.
The Random Forest model was constructed with 100 decision trees and class_weight= 'balanced' to manage class imbalance. Training was performed on the SMOTE-balanced dataset to further reinforce minority class representation. While tree-based models do not require feature scaling, scaling was retained for consistency in preprocessing.
This model achieved high accuracy of 96%, with precision and recall for the heatwave class at 0.56 and 0.67, respectively. The F1-score (0.61) and ROC-AUC (0.9619) were substantially higher than the Decision Tree, reflecting improved overall balance and predictive power. The confusion matrix shows that while false positives

were reduced compared to the Decision Tree, some heatwave days were still missed. Nonetheless, Random Forest provided one of the best performances among all models tested.

### 5.1.3 Logistic Regression

Logistic regression is a supervised machine learning algorithm in data science. It refers to the classification algorithms that predict a discrete or categorical outcome. Therefore, it is very much used for classification purposes, especially those tasks involved with binary classification.

Logistic Regression was implemented as a baseline model due to its interpretability, efficiency, and well-established performance in binary classification tasks. To address the significant class imbalance in the dataset—where heatwave days were far fewer than non-heatwave days—two techniques were employed. First, class weights were adjusted using class_weight='balanced' to penalize the misclassification of the minority class (heatwave events) more heavily. Second, SMOTE (Synthetic Minority Over-sampling Technique) was applied to the training set to synthetically generate additional samples of the heatwave class, thereby providing the model with a more balanced distribution of class instances. All input features were standardized using a StandardScaler to ensure uniform scaling across variables, which is essential for the convergence and performance of Logistic Regression.
The model was evaluated on the test set for the Dhaka division and achieved an accuracy of 83%, with a remarkably high recall of 0.98 for the heatwave class. This indicates that the model was successful in identifying almost all actual heatwave days. However, the precision was relatively low at 0.21, resulting in a high number of false positives, which lowered the F1-score for the heatwave class to 0.34. The confusion matrix further highlights this trade-off, showing 80 true positives, 2 false negatives, and 310 false positives. Despite this, the model achieved a ROC-AUC score of 0.9698, demonstrating excellent discriminatory ability between the two classes.

### 5.1.4 Decision Tree

Using the Decision tree which is a machine learning technique, we can do binary classification, which means that the outcome has only two possibilities.
The Decision Tree model was implemented using a maximum depth of 5 to reduce the risk of overfitting, particularly important in datasets with noise or complex patterns. To mitigate the effects of class imbalance, class weights were computed and applied during model training. Additionally, SMOTE was used on the training set to ensure a balanced representation of heatwave and non-heatwave days.
The Decision Tree achieved an accuracy of 84%, with a recall of 0.95 for the heatwave class, indicating strong sensitivity in detecting heatwave events. However, precision was low (0.21), resulting in a high number of false positives. The F1-score for the heatwave class was 0.35, and the ROC-AUC score was 0.8944, suggesting moderate class separation. While the model is interpretable and easy to visualize, its performance indicates limitations in handling class overlap and data complexity.

## 5.2 Deep Learning Approach

### 5.2.1 CNN LSTM (Long Short-Term Memory)

A CNN-LSTM (Convolutional Neural Network–Long Short-Term Memory) is a hybrid deep learning architecture that combines the temporal sequence modeling capabilities of Long Short-Term Memory (LSTM) networks with the feature extraction capabilities of Convolutional Neural Networks (CNNs). When there are both temporal and spatial patterns in the data, this model performs especially well in time series prediction tasks. As required for time series modeling with LSTM layers, the CNN-LSTM model's input data is in 3D shape (samples, time_steps, features).

To identify heatwave streaks and forecast future maximum temperature values, a hybrid convolutional neural network and long short-term memory (CNN-LSTM) model was used. This combination makes use of the advantages of both architectures: LSTM layers discover long-term temporal dependencies in the data, while CNN layers identify local trends and short-term variations. In order to prepare the data for CNN-LSTM models, it was reshaped into a 3D format. To ensure all variables contribute equally to the mode, the features are flattened and scaled after the data has been reshaped into sequences. Splits the data into sequences of past observations (nsteps) and assigns the next time point's value as the target. StandardScaler was used to apply standardization because deep learning models are sensitive to feature magnitude. Every input feature was scaled to have a unit variance and zero mean. Within the sliding window, the Conv1D layer finds short-term patterns throughout the feature space. Long-term dependencies are learned by the LSTM layer from the filtered sequence. Predictions for multiple time steps (multi-step forecast) are made using a dense layer. Five time steps was selected as the window size. This indicates that the model balances short-term and mid-range dependencies by predicting future temperatures using weather data from five consecutive days.

An LSTM layer and a 1D Convolutional Layer (Conv1D) were combined to create a hybrid neural network during the modeling stage, which took advantage of the weather data's long-term temporal dependencies as well as short-term local patterns. In order to reduce overfitting, the architecture started with a Conv1D layer that had 64 filters and a kernel size of 2. This was followed by a dropout layer (30%). Then, sequential patterns were captured using an LSTM layer with 50 units, which was again followed by a dropout layer (30%). The output prediction came from the last Dense layer. Mean Squared Error (MSE) was used as the loss function and the Adam optimizer was used to compile the model. A learning rate scheduler that maintained the learning rate constant for the first 10 epochs and then decayed it by 10% per epoch was used to improve training stability and convergence. In order to avoid overfitting and unnecessary training, early stopping was also implemented with a 10-epoch patience. A batch size of 32 was used to train the model, and 20% of the training data was used as validation. To track performance, a training loss curve was plotted.

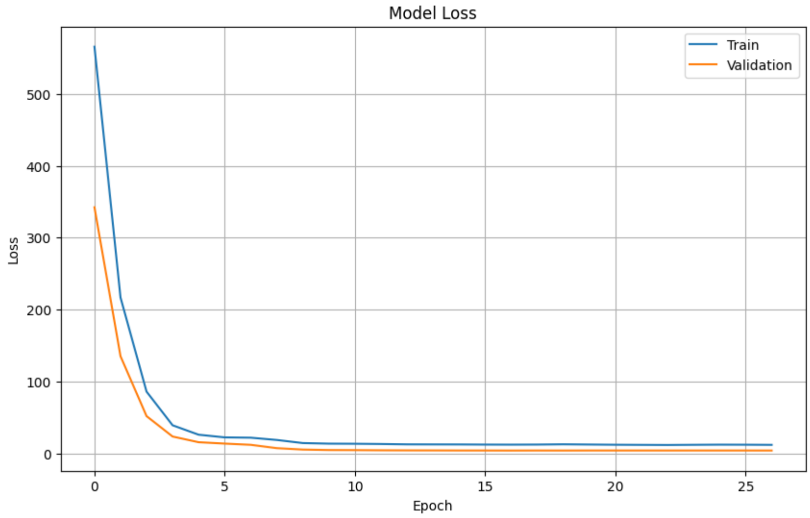


Figure 5.1: Model Loss Curve

Training Loss (Blue Line): This indicates the error or loss for your model on the training data.

Validation Loss (Orange Line): This indicates the error or loss for your model on the validation data, which represents data the model has not seen during training.

The performance of the proposed CNN-LSTM hybrid model for predicting maximum temperature in the Bangladesh region was evaluated using standard regression metrics: Coefficient of Determination ($R^2$ Score), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The $R^2$ score of approximately 0.685 indicates that the model explains about 68.5% of the variance in the maximum temperature values. This demonstrates a moderate to strong predictive capability, especially considering the high variability inherent in meteorological data. The relatively low RMSE suggests that the model's average prediction error is around 2.23°C, which is acceptable for many practical applications involving temperature forecasting. After predicting max temperature sequences, I apply a post-processing rule-based method: if 3+ consecutive days have predicted max_temp  36°C, it's labeled a heatwave streak, and the severity is labeled accordingly (36/38/40/42°C thresholds).

## 5.3 Regressor Approach

### 5.3.1 Random Forest Regression

The Random Forest Regressor is an ensemble machine learning algorithm that makes predictions by averaging the outputs of several decision trees. The RFR algorithm is strong against overfitting, and this model works well with both linear and non-linear relationships. This model also works well with data that is noisy or not complete. It works especially well for predicting heat waves because it can find complicated relationships in weather data, give feature importance scores, and this variable works well across different regions.
RandomForestRegressor is used to predict the temperature 5 days in advance. Important features such as Min_Temperature, Relative_Humidity, Rainfall, Sun-

shine_Hours, Cloud_Amount, Wind_Speed, enso enomaly, as well as engineered variables like Heat_Index, cos_month, and lagged temperature values were used as predictors. The target variable was the maximum temperature for the next 5 days (Tmax_t+1 to Tmax_t+5). RandomForestRegressor predicts the 5-day temperature using Direct Multi-Output Strategy. Multi-Output implemented using MultiOutputRegressor. A standard RandomForestRegressor is built to predict a single numerical value. In general it works like, give it a set of input features and it outputs one number. However, to predict five different target: `MultiOutputRegressor` creates and trains **five completely separate and independent Random Forest models** behind the scenes:

- **Model 1 (The T+1 Specialist)**: It trains a full Random Forest model using your input features (`X_train`) with the sole purpose of learning how to predict the `Tmax_t+1` column from your target data (`y_train`).
- **Model 2 (The T+2 Specialist)**: It trains a second Random Forest model. It uses the exact same input features (`X_train`) but this time its only goal is to learn how to predict the `Tmax_t+2` column.
- **Model 3, 4, and 5**: This process is repeated, creating a specialist model for each of the five days you want to forecast.

The data was split into training and testing set with an 80:20 ratio. This make sure that the model evaluation was done on unseen data. A RandomForestRegressor with 100 trees (n_estimators=100) and a fixed random state (random_state=42) was use in this mode. To handle multi target, MultiOutputRegressor used for multi-step forecasting. The model was trained on the training data (X_train, y_train) and evaluated on the test set (X_test, y_test) using performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and $R^2$ Score to assess its predictive capability.
The Random Forest Regressor produced an $R^2$ score of 0.7525. Approximately 75.25% of the variance in maximum temperature is explained by the model. This shows strong predictive capability of this model. The RMSE of 1.99°C indicates that the model's average prediction error is under 2°C, which is generally acceptable for regional climate forecasting. Mean square error 3.97 is relatively low overall error across the testing data.

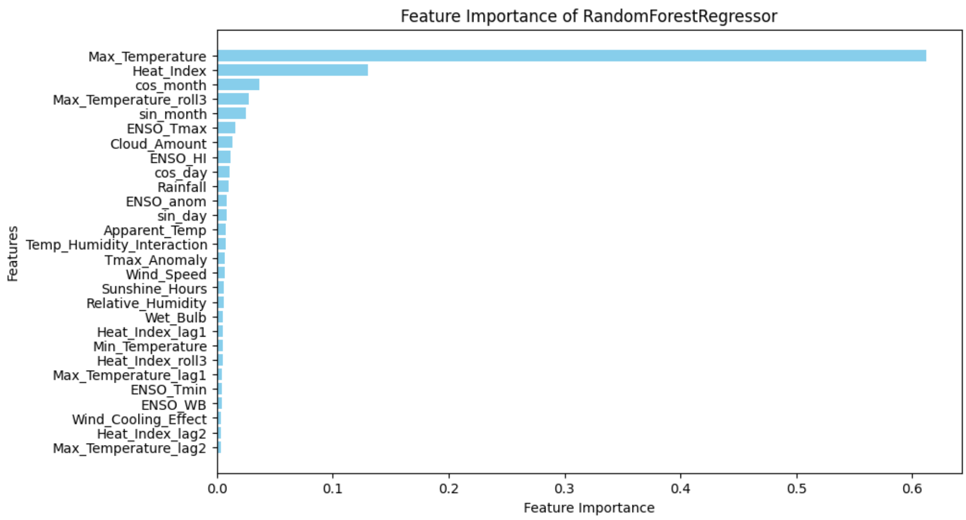


Figure 5.2: Feature Importance of RandomForestRegressor

The current day's maximum temperature is the most significant predictor for the next day's temperature, This shows the correlation in temperature data with prediction. Also, second important feature Heat Index is etimates perceived temperature by combining air temperature and relative humidity. Cyclical encodings of the month (cos_month / sin_month) indicate the importance of seasonal patterns. In addition, A 3-day rolling average of maximum temperature effectively captures trends for predicting sustained phenomena like heatwaves. Finally,though having a modest impact, ENSO-related variables, represent important long-term global climate signals.

### 5.3.2 XGBoost Regression

XGBoost is a fast, scalable ensemble algorithm using gradient-boosted decision trees. It sequentially builds trees to correct errors and optimize a loss function. It's well Known for modeling complex non-linear relationships and reducing overfitting via regularization. It effectively addresses non-linear interactions and handles missing values for improved interpretability and confidence in climate decision-making.
Like Random Forest Regression, MultiOutputRegressor use for to enable multi-step forecasting of maximum temperatures (Tmax_t+1 to Tmax_t+5). Same feature set used in XGBoost model.

The dataset was split into training and testing sets using an 80:20 ratio. The model was trained with key parameters such as n_estimators=100, learning_rate=0.1, and random_state=42 to ensure consistent and reproducible results. Although exhaustive hyperparameter tuning was not performed, the default configuration already yielded promising outcomes. The model's performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the $R^2$ Score to measure goodness-of-fit and prediction accuracy.
The XGBoost Regressor achieved the following results on the test dataset: Mean Squared Error (MSE): 3.8734, Root Mean Squared Error (RMSE): 1.9681 , $R^2$ Score: 0.7584

These results indicate that the model can explain approximately 75.84% of the variance in the future maximum temperature values, demonstrating strong predictive performance in a complex, real-world forecasting scenario. The relatively low RMSE ( 1.97°C) suggests that the model's average prediction error is within an acceptable range for climate forecasting applications.

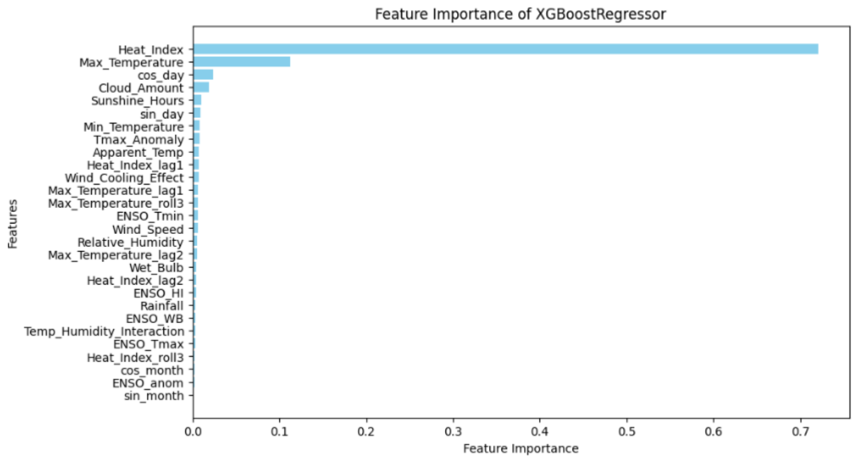


Figure 5.3: Feature Importance of XGBoostRegressor

In this model, Heat_Index is the most dominant predictor of future maximum temperatures, contributing over 85% of the total model influence. This reinforces the need to consider thermal comfort variables—such as Heat_Index and Apparent_Temp—rather than raw meteorological variables alone when modeling heatwave risks. Additionally, cyclical encoding of time through cos_day and cos_month enabled the model to learn seasonal patterns, further improving its forecasting ability.

### 5.3.3 Decision Tree Regressor

A Decision Tree Regressor (DTR) is a non-parametric supervised learning algorithm used for regression tasks. It works by recursively splitting the data space into regions based on input features, forming a tree-like structure. Each internal node represents a decision based on a feature value. Each leaf node holds a prediction (in regression, a continuous value). The model selects the best feature and threshold to split the data at each node by minimizing a loss function (typically mean squared error in regression).

Like Random Forest Regression, MultiOutputRegressor use for to enable multi-step forecasting of maximum temperatures (Tmax_t+1 to Tmax_t+5). Same feature set used in this model.
The Decision Tree Regressor produced a Mean Squared Error (MSE) of 8.5177, a Root Mean Squared Error (RMSE) of approximately 2.92°C, and an $R^2$ Score of 0.4688, indicating that the model was able to explain only about 46.88% of the variance in the future maximum temperature values. These results reflect a relatively lower predictive performance compared to ensemble models like Random Forest and XGBoost. The high RMSE suggests that the model's average prediction error is nearly 3°C, which may be less reliable for operational heatwave forecasting.

# Chapter 6

# Results Discussion

## 6.1 Accuracy Score

### 6.1.1 Binary classification Approach

#### XGBoost

The XGBoost model achieved a 96% accuracy, with a precision of 0.51, recall of 0.70, and an F1-score of 0.59 for the heatwave class. The ROC-AUC score of 0.9603 indicates excellent class separation. Compared to the Random Forest, XGBoost slightly improved recall while maintaining a similar level of precision. This performance affirms the model's capability in handling imbalanced classification problems and highlights its effectiveness in operational contexts requiring a balance of precision and recall.

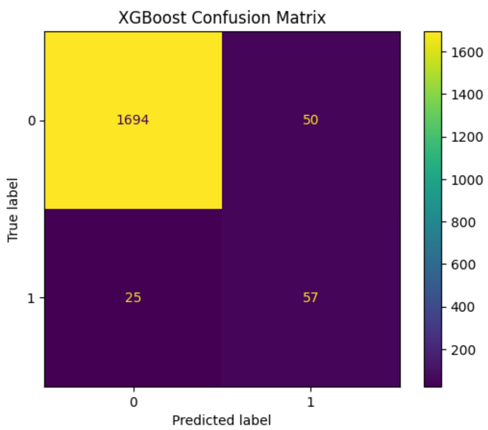


Figure 6.1: XGBoost confusion matrix for Dhaka division

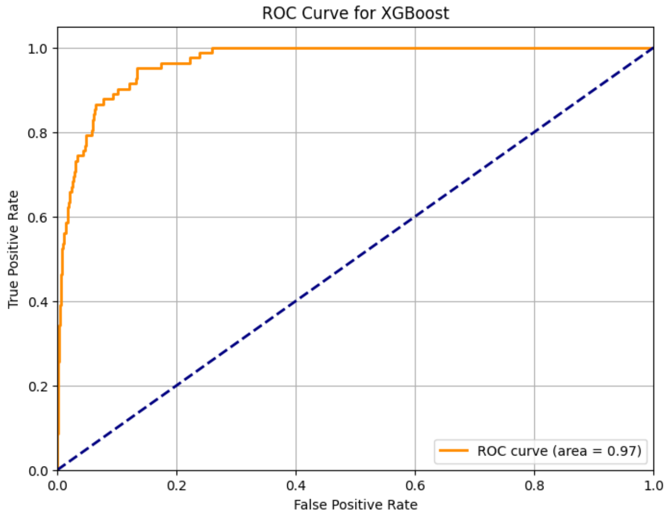


Figure 6.2: ROC Curve of XGBoost for Dhaka division

### Random Forest

This model achieved high accuracy of 96%, with precision and recall for the heatwave class at 0.56 and 0.67, respectively. The F1-score (0.61) and ROC-AUC (0.9619) were substantially higher than the Decision Tree, reflecting improved overall balance and predictive power. The confusion matrix shows that while false positives were reduced compared to the Decision Tree, some heatwave days were still missed. Nonetheless, Random Forest provided one of the best performances among all models tested.

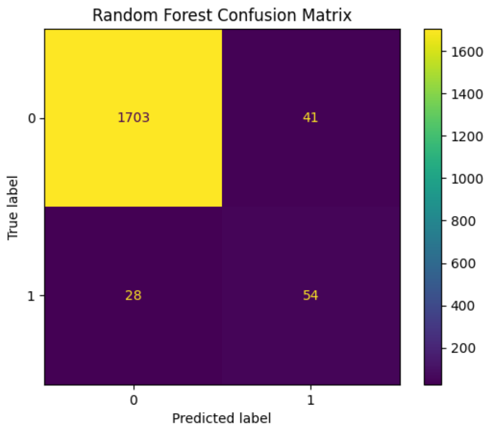


Figure 6.3: Random Forest confusion matrix for Dhaka division

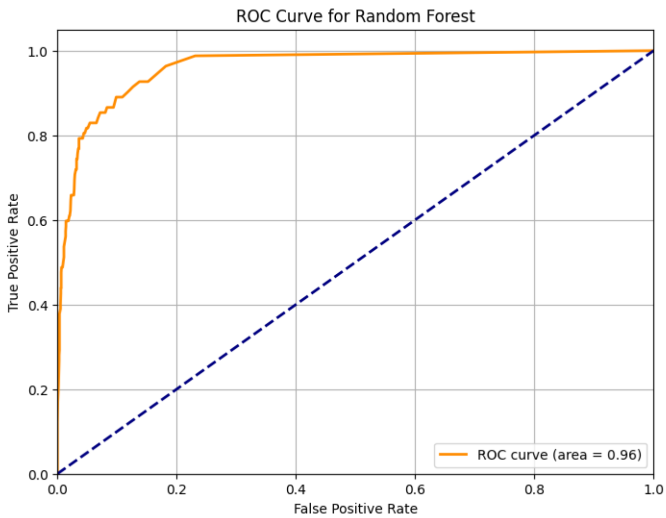


Figure 6.4: ROC Curve of Random Forest for Dhaka division

### Logistic Regression

The model was evaluated on the test set for the Dhaka division and achieved an accuracy of 83%, with a remarkably high recall of 0.98 for the heatwave class. This indicates that the model was successful in identifying almost all actual heatwave days. However, the precision was relatively low at 0.21, resulting in a high number of false positives, which lowered the F1-score for the heatwave class to 0.34. The confusion matrix further highlights this trade-off, showing 80 true positives, 2 false negatives, and 310 false positives. Despite this, the model achieved a ROC-AUC score of 0.9698, demonstrating excellent discriminatory ability between the two classes.

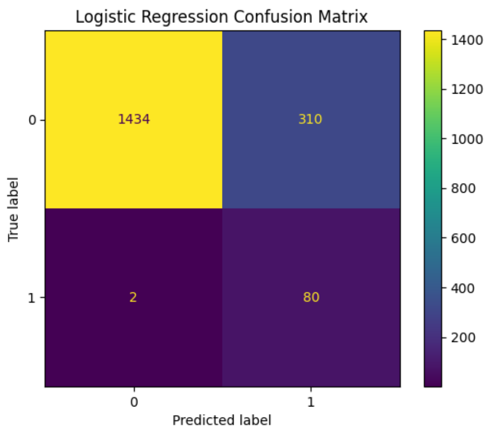


Figure 6.5: Logistic Regression confusion matrix for Dhaka division

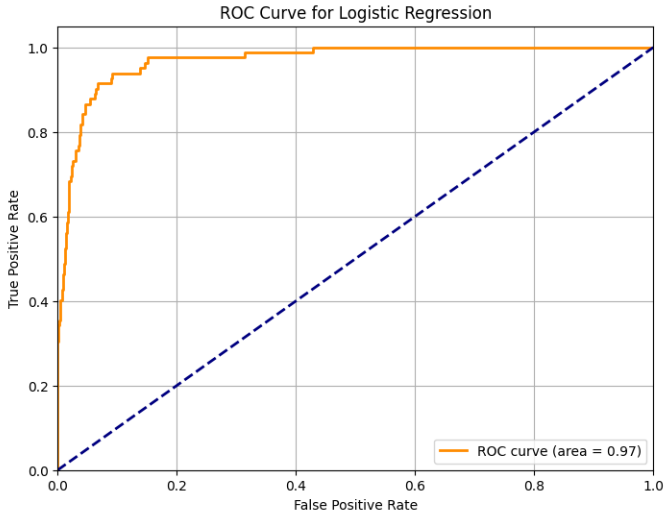


Figure 6.6: ROC Curve of Logistic Regression for Dhaka division

## Decision Tree

The Decision Tree achieved an accuracy of 84%, with a recall of 0.95 for the heatwave class, indicating strong sensitivity in detecting heatwave events. However, precision was low (0.21), resulting in a high number of false positives. The F1-score for the heatwave class was 0.35, and the ROC-AUC score was 0.8944, suggesting moderate class separation. While the model is interpretable and easy to visualize, its performance indicates limitations in handling class overlap and data complexity.

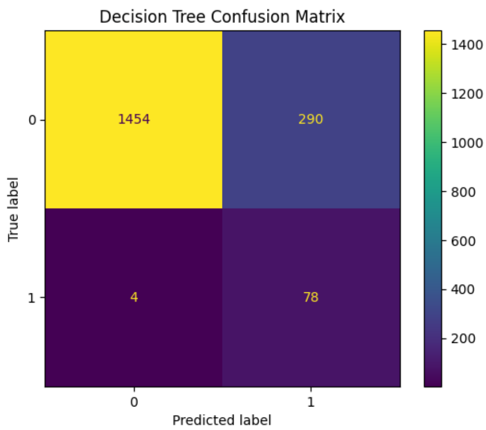


Figure 6.7: Decision Tree confusion matrix for Dhaka division

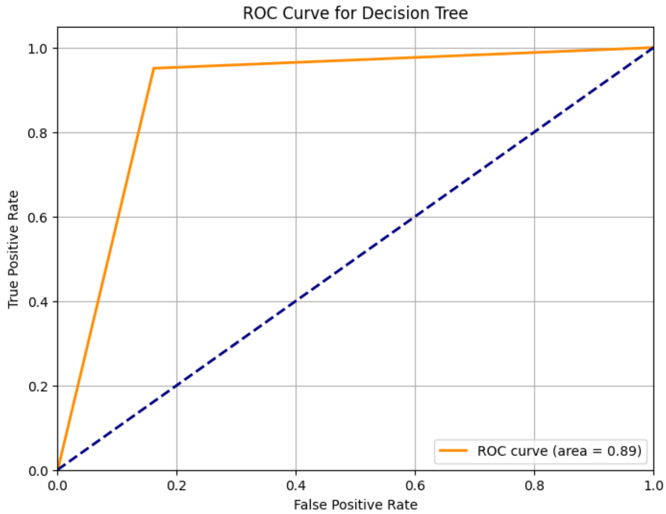


Figure 6.8: ROC Curve of Decision Tree for Dhaka division

### Overall result

Here is the overall result for all divisions of all the models using binary classification. The below picture combinedly shows the Precision, Recall, F1-Score, ROC-AUC, Accuracy for both heatwave and non-heatwave days (0=non-heatwave, 1=heatwave) days for all divisions.

| Division | Model | Precision (0) | Precision (1) | Recall (0) | Recall (1) | F1 (0) | F1 (1) | ROC-AUC | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Dhaka | XGBoost | 0.99 | 0.53 | 0.97 | 0.7 | 0.98 | 0.6 | 0.9668 | 0.96 |
| | Random Forest | 0.98 | 0.57 | 0.98 | 0.66 | 0.98 | 0.61 | 0.962 | 0.96 |
| | Logistic Regression | 1 | 0.21 | 0.82 | 0.98 | 0.9 | 0.34 | 0.9698 | 0.83 |
| | Decision Tree | 1 | 0.21 | 0.83 | 0.95 | 0.91 | 0.35 | 0.8944 | 0.84 |
| Rajshahi | XGBoost | 0.98 | 0.54 | 0.94 | 0.75 | 0.96 | 0.63 | 0.9529 | 0.92 |
| | Random Forest | 0.97 | 0.57 | 0.95 | 0.7 | 0.96 | 0.63 | 0.9471 | 0.93 |
| | Logistic Regression | 0.99 | 0.29 | 0.77 | 0.96 | 0.87 | 0.44 | 0.9553 | 0.79 |
| | Decision Tree | 0.99 | 0.31 | 0.8 | 0.93 | 0.88 | 0.46 | 0.9194 | 0.81 |
| Khulna | XGBoost | 0.98 | 0.59 | 0.94 | 0.85 | 0.96 | 0.7 | 0.9669 | 0.93 |
| | Random Forest | 0.98 | 0.59 | 0.94 | 0.81 | 0.96 | 0.68 | 0.9625 | 0.93 |
| | Logistic Regression | 1 | 0.34 | 0.8 | 0.99 | 0.89 | 0.51 | 0.9726 | 0.82 |
| | Decision Tree | 0.99 | 0.37 | 0.83 | 0.95 | 0.9 | 0.53 | 0.9339 | 0.84 |
| Sylhet | XGBoost | 0.99 | 0.36 | 0.97 | 0.64 | 0.98 | 0.46 | 0.9537 | 0.96 |
| | Random Forest | 0.99 | 0.31 | 0.97 | 0.52 | 0.98 | 0.39 | 0.9442 | 0.96 |
| | Logistic Regression | 1 | 0.08 | 0.74 | 1 | 0.85 | 0.15 | 0.9709 | 0.75 |
| | Decision Tree | 1 | 0.08 | 0.74 | 0.93 | 0.85 | 0.14 | 0.8525 | 0.75 |
| Barishal | XGBoost | 1 | 0.35 | 0.97 | 0.83 | 0.98 | 0.49 | 0.9358 | 0.97 |
| | Random Forest | 1 | 0.39 | 0.97 | 0.86 | 0.98 | 0.53 | 0.9569 | 0.97 |
| | Logistic Regression | 1 | 0.11 | 0.83 | 1 | 0.91 | 0.19 | 0.9887 | 0.83 |
| | Decision Tree | 1 | 0.11 | 0.85 | 0.97 | 0.92 | 0.2 | 0.9099 | 0.85 |
| Chattogram | XGBoost | 1 | 0.34 | 0.99 | 0.67 | 0.99 | 0.45 | 0.9129 | 0.99 |
| | Random Forest | 1 | 0.4 | 0.99 | 0.67 | 0.99 | 0.5 | 0.8269 | 0.99 |
| | Logistic Regression | 1 | 0.12 | 0.94 | 1 | 0.97 | 0.21 | 0.9888 | 0.94 |
| | Decision Tree | 1 | 0.15 | 0.97 | 0.73 | 0.98 | 0.25 | 0.8493 | 0.96 |
| Rangpur | XGBoost | 0.99 | 0.38 | 0.97 | 0.62 | 0.98 | 0.47 | 0.9486 | 0.96 |
| | Random Forest | 0.98 | 0.38 | 0.97 | 0.5 | 0.98 | 0.43 | 0.9388 | 0.96 |
| | Logistic Regression | 1 | 0.11 | 0.73 | 0.98 | 0.84 | 0.19 | 0.9489 | 0.74 |
| | Decision Tree | 1 | 0.1 | 0.72 | 0.95 | 0.83 | 0.18 | 0.8636 | 0.72 |
| Mymensingh | XGBoost | 0.99 | 0.29 | 0.98 | 0.53 | 0.99 | 0.38 | 0.9521 | 0.97 |
| | Random Forest | 0.99 | 0.3 | 0.98 | 0.43 | 0.99 | 0.35 | 0.9416 | 0.97 |
| | Logistic Regression | 1 | 0.05 | 0.67 | 1 | 0.8 | 0.09 | 0.9547 | 0.68 |
| | Decision Tree | 1 | 0.05 | 0.72 | 0.9 | 0.83 | 0.1 | 0.8077 | 0.72 |

Figure 6.9: All Division recision, Recall, F1-Score, ROC-AUC, Accuracy

Here is the comparison bar chart for all models-

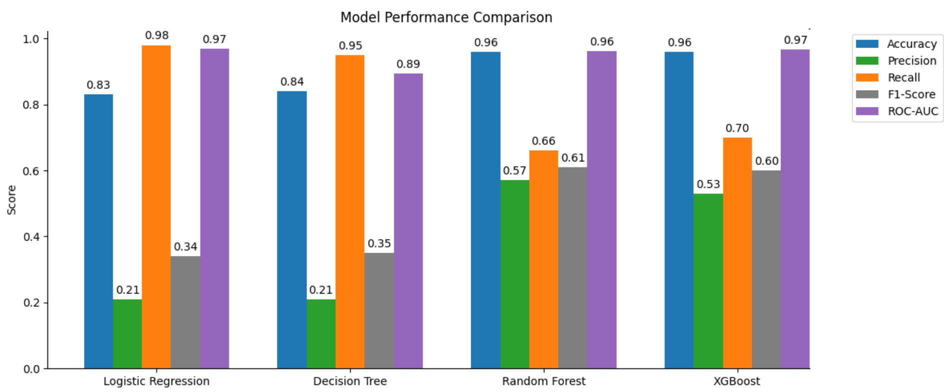


Figure 6.10: Overall comparison results on Binary Classification

## 6.2 Regressor and Deep Learning Hybrid Approach

The performance of the proposed CNN-LSTM hybrid model for predicting maximum temperature in the Bangladesh region was evaluated using standard regression metrics: Coefficient of Determination ($R^2$ Score), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The $R^2$ score of approximately 0.685 indicates that the model explains about 68.5% of the variance in the maximum temperature values. MSE value is 4.983 and RMSE value is 2.33 .

The Random Forest Regressor produced an $R^2$ score of 0.7525, meaning approximately 75.25% of the variance in maximum temperature is explained by the model. This reflects strong predictive capability, especially given the variability of climate data. The RMSE of 1.99°C indicates that the model's average prediction error is under 2°C, which is generally acceptable for regional climate forecasting. The MSE of 3.97 supports the conclusion of relatively low overall error across the testing data.

The XGBoost Regressor achieved the following results on the test dataset: Mean Squared Error (MSE): 3.8734, Root Mean Squared Error (RMSE): 1.9681 , $R^2$ Score: 0.7584 These results indicate that the model can explain approximately 75.84% of the variance in the future maximum temperature values, demonstrating strong predictive performance in a complex, real-world forecasting scenario. The relatively low RMSE ( 1.97°C) suggests that the model's average prediction error is within an acceptable range for climate forecasting applications.

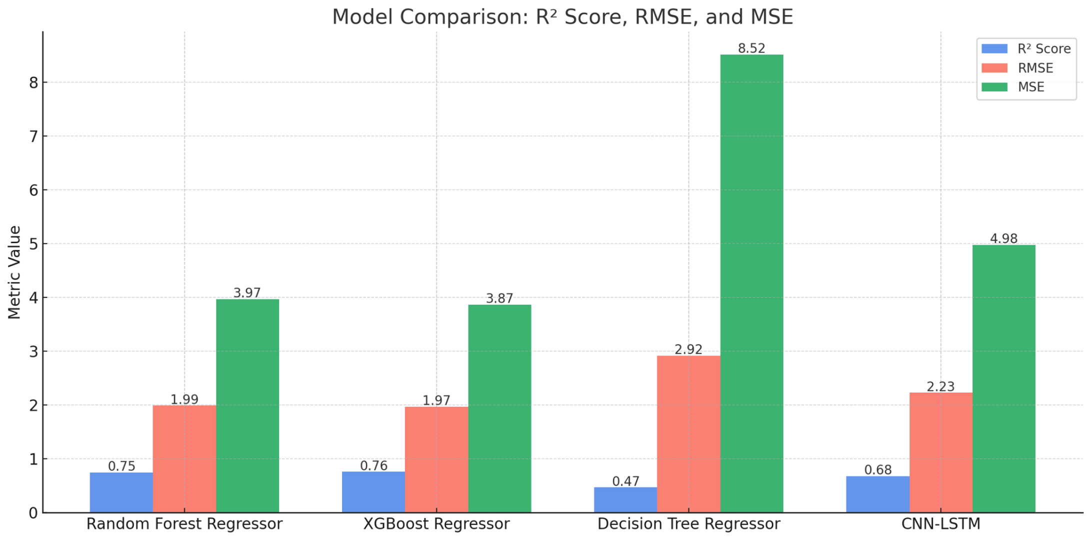


Figure 6.11: Overall comparison results on regressor approach

The Decision Tree Regressor produced a Mean Squared Error (MSE) of 8.5177, a Root Mean Squared Error (RMSE) of approximately 2.92°C, and an $R^2$ Score of 0.4688, indicating that the model was able to explain only about 46.88% of the variance in the future maximum temperature values. These results reflect a relatively lower predictive performance compared to ensemble models like Random Forest and XGBoost.

## 6.3 Result Comparison

In our study two distinct modeling strategies applied to the task of heatwave prediction in Bangladesh:

- Classification + rule-based detection: Binary classification models (Binary Model Name) that predict whether a day is a heatwave or not and then use rule-based logic to identify heatwaves.
- Temperature regression + rule-based detection: Regression models (CNN-LSTM, baki 3 tah regression er) that predict future temperatures and then use rule-based logic to identify heatwaves.

**Direct Classification Models**

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| XGBoost | 96% | 0.53 | 0.7 | 0.60 | 0.9668 |
| Random Forest | 96% | 0.57 | 0.66 | 0.61 | 0.9620 |
| Logistic Regression | 83% | 0.21 | 0.98 | 0.34 | 0.9698 |
| Decision Tree | 84% | 0.21 | 0.95 | 0.35 | 0.8944 |

Table 6.1: All results of Binary Classification Approach

Logistic Regression and Decision Tree models achieved very high recall (>95%), capturing most heatwave events, but at the cost of low precision ( 0.21)—many false positives. Ensemble models (XGBoost and Random Forest) offered a better balance between precision and recall, leading to higher F1-scores. All models performed well in terms of ROC-AUC, indicating strong separability between classes. However, class imbalance (few heatwave days compared to non-heatwave days) remained a significant challenge, even after using SMOTE and class weights.
CNN-LSTM hybrid deep learning model predicted future maximum temperatures (next 5 days) and then applied a rule-based approach to identify heatwaves. The CNN-LSTM achieved moderate-to-strong correlation with actual temperature values, explaining 68.5% variance. Despite the complexity of meteorological data, the model maintained reasonable accuracy ( 2.23°C error). This approach allowed greater flexibility, enabling heatwave classification and severity categorization (mild, moderate, etc.) via temperature thresholds rather than rigid binary output.
Tree-Based Regression Models we use Random Forest Regressor, XGBoost Regressor, Decision Tree Regressor . Both XGBoost and Random Forest regressors outperformed the Decision Tree Regressor significantly. These models were able to explain >75% of the variance in max temperature values and provided low prediction errors (<2°C). Their post-processing logic effectively identified heatwave streaks and classified severity levels, adding interpretability and granularity to the results.
Initially, binary classification models were implemented to directly predict heatwave days. While these models achieved high accuracy and recall, the inherent class imbalance, relatively few heatwave days compared to non-heatwave days resulted in high false positive rates, poor precision and F1-scores, and limited ability to classify heatwave severity levels. These challenges motivated a strategic pivot in the modeling approach: instead of directly classifying heatwaves, the methodology

| Model | $R^2$ Score | RMSE | MSE |
|---|---|---|---|
| Random Forest Regressor | 0.7572 | 1.99°C | 3.97 |
| XGBoost Regressoer | 0.7584 | 1.97°C | 3.87 |
| Decision Tree Regressor | 0.4688 | 2.99°C | 8.51 |
| CNN-LSTM | 0.6822 | 2.23°C | 4.98 |

Table 6.2: $R^2$ Score, RMSE, MSE results of Regressor Approach Models

shifted to forecasting maximum temperatures using regression models, followed by a rule-based post-processing logic where sequences of three or more consecutive days with predicted temperatures  36°C were labeled as heatwave streaks, with further classification into severity levels (mild, moderate, severe, extreme) based on defined thresholds. This revised strategy offered several advantages it produced more stable and interpretable predictions, avoided direct dependency on skewed class labels, enabled nuanced classification of heatwave severity, and generalized better across divisions due to improved temporal modeling, such as capturing seasonal patterns and autocorrelation in temperature data.
The regression + rule-based method provided better operational utility, especially in climate-sensitive regions like Bangladesh. While binary classification provided initial insights, the hybrid regression approach—particularly with XGBoost and Random Forest regressors provide superior accuracy, interpretability, and practical application for heatwave event classification and intensity assessment.
For the Bangladesh region, heatwaves are rare but deadly; only 5% of days are heatwaves (Class 1 ), creating severe imbalance. This imbalance proves particularly intractable in this result analysis. High ROC-AUC scores ($> 0.96$) reflect excellent non-heatwave day detection, not heatwave prediction. Logistic Regression/DT—high recall but unusable precision (80-95% false alarms). This research applied mitigation attempts such as class weighting to penalise misclassifying heatwaves and SMOTE oversampling for synthetic heatwave examples. Imbalance remains irrecoverable—no model achieves both high precisions, This isn't a modelling failure—it's a data reality because heatwaves affect $< 1$ in 20 days.
For XGBoost, misses 25/82 true heatwaves (30% undetected).On the other hand, for logistic regression, flags 310 false alarms to catch 80 true heatwave. For Bangladesh, XGBoost is missing 30% of heatwaves. Flooding alerts with false alarms in logistic regression are both unacceptable operational outcomes. As a result, this research looks beyond binary classification to resolve this trade-off for the Bangladesh region. This research pivoted to a regression and post-thresholding approach to avoid the imbalance entirely by predicting temperatures first. Random Forest Regressor able to explain >75% of the variance in max temperature values and provided low prediction errors (<2°C).

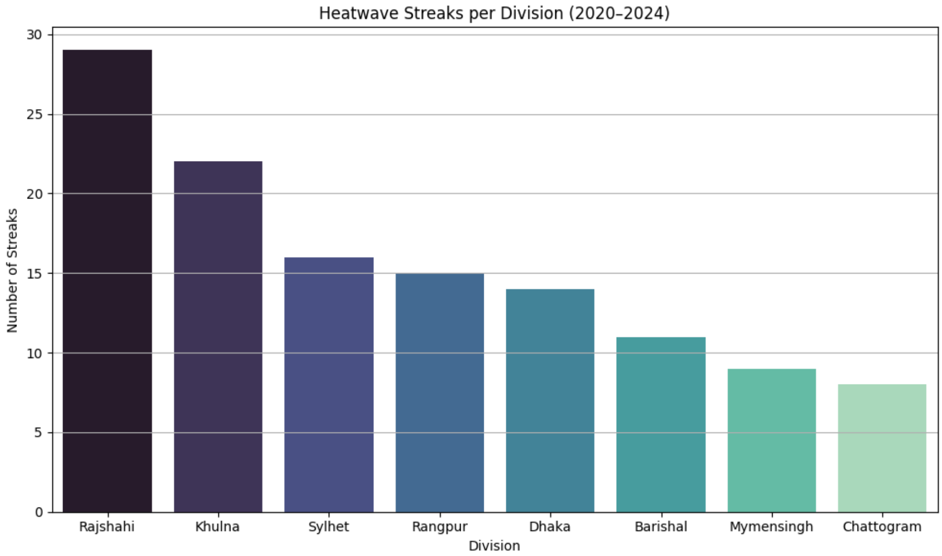


Figure 6.12: Actual streak count by division (2020-2024)

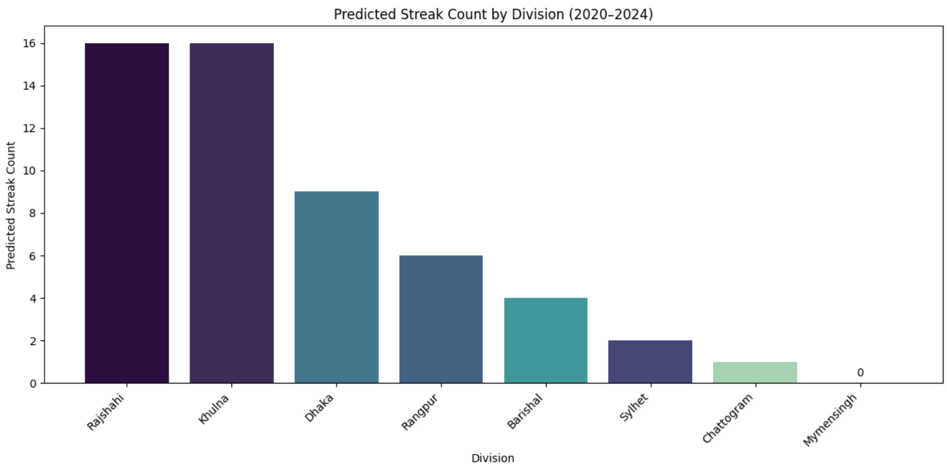


Figure 6.13: Predicted streak count by division (2020-2024)

# Chapter 7

# Limitations

## 7.1 Data-Related Limitations

- Class Imbalance: In binary classification, heatwave days, which make up less than 5% of the dataset, cause models to bias toward ”no heatwave” case compared to “heatwave” days, as heatwaves are rare events in this region.
- Sparse Station Coverage: Only one weather station represents Mymensingh division, which leads to fewer taring data sample making unreliable streak predictions for this region. As a result, no streaks are predicted for the Mymensingh division in the multi-step regressor approach. Additionally, missing data gaps, such as humidity and wind speed, in historical records reduce feature robustness.
- Overfitting Risk: Deep learning requires large datasets; our small sample size leads to overfitting (LSTM validation loss divergence).

## 7.2 Model Limitations

- Binary classification: The extreme class imbalance scenario in Bangladesh causes more false alarms using binary classification ML models. As a result, the heatwave prediction result will be mostly wrong, with increasing costs for EWS.
- Multi-regressor approach: Despite a decent $R^2$, RFR may overfit due to high feature dimensionality. RFR could misclassify borderline heatwave days. For XGBR, small datasets (single-station Mymensingh) lead to high variance. Small data changes drastically alter tree structure, making predictions unreliable for operational use in DTR model.
- Deep learning: Fewer data sample.

## 7.3 S2S Forecasting Gaps

ENSO features improve lead time but fail to capture hyper-local phenomena, such as urban heat islands in Dhaka.

## 7.4 Operational Challenges

- Real-World Deployment: The model assumes real-time data availability. Delays in station updates degrade forecast accuracy. There is no integration with health data, such as mortality and hospitalization rates.

## 7.5 Climate Change Uncertainty

- Non-Stationary Climate: The model is trained on historical data from 2000 to 2019, but rising baselines may make thresholds outdated. For example, 36°C could become 38°C by 2050.

# Chapter 8

# Conclusion and Future Work

## 8.1 Conclusion

Despite the growing threat of heatwaves, no remarkable research has applied advanced ML models using meteorological data and geographical factors for forecasting in Bangladesh. The research fills that gap by leveraging historical weather data for precise predictions. The research highlights the effectiveness of Multi-step approach using the machine learning models and hybrid deep learning model for heatwave prediction in Bangladesh, where regional temperature, humidity, and environmental variations are influenced by the proximity to the Bay of Bengal. A region-specific forecasting approach is emphasized for better accuracy, and integrating the effects of climate phenomena like ENSO phases (La Niña, Neutral, and El Niño) provide crucial long-term insights into atmospheric and oceanic conditions, thereby enhancing the predictive capability of the models. In a nutshell, this study will make a substantial contribution to future developments in heatwave forecasting and disaster management, not only within Bangladesh but also serving as a valuable model for other vulnerable regions globally.

## 8.2 Future Work

In future, Expanding the dataset to include a greater volume and diversity of historical meteorological, geographical, and potentially socio-economic data will enable the models to capture more intricate patterns and improve generalization. Additionally, future work includes developing a unified, multi-target hybrid model with a shared encoder. This approach would allow for simultaneous end-to-end prediction, integrating both machine learning and deep learning techniques seamlessly to enhance overall prediction accuracy and efficiency. These advancements aim to develop a robust early warning framework, contributing to heatwave forecasting and disaster management in the Bangladesh region. Moreover, our idea is to develop a web interface that implements the best-performing model at the backend and could be deployed as a prototype real-time warning system in collaboration with an organization like the BMD. We hope that this project will be helpful for future research and development.

# Bibliography


[1] R. G. Steadman, “A universal scale of apparent temperature,” *Journal of Climate and Applied Meteorology*, vol. 23, no. 12, pp. 1674–1687, 1984. DOI: 10.1175/1520-0450(1984)023<1674:AUSOAT>2.0.CO;2.

[2] A. P. Gagge, A. P. Fobelets, and L. G. Berglund, “A standard predictive index of human response to the thermal environment,” *ASHRAE Transactions*, vol. 92, no. 2B, pp. 709–731, 1986.

[3] A. Hollis, *'Modelling and Understanding High Impact Weather': Extended Abstracts of the Third CAWCR Modelling Workshop, 30 November-2 December 2009, Melbourne, Australia.* Citeseer, 2009.

[4] R. Stull, “Wet-bulb temperature from temperature and relative humidity,” *Journal of Applied Meteorology and Climatology*, vol. 50, no. 11, pp. 2267–2269, 2011. DOI: 10.1175/JAMC-D-11-0143.1.

[5] J. R. Nairn and R. J. Fawcett, “The excess heat factor: A metric for heatwave intensity and its use in classifying heatwave severity,” *International journal of environmental research and public health*, vol. 12, no. 1, pp. 227–253, 2015.

[6] H. Nissan, K. Burkart, E. Coughlan de Perez, M. Van Aalst, and S. Mason, “Defining and predicting heat waves in bangladesh,” *Journal of Applied Meteorology and Climatology*, vol. 56, no. 10, pp. 2653–2670, 2017.

[7] J. Nairn, B. Ostendorf, and P. Bi, “Performance of excess heat factor severity as a global heatwave health impact index,” *International journal of environmental research and public health*, vol. 15, no. 11, p. 2494, 2018.

[8] J. Park and J. Kim, “Defining heatwave thresholds using an inductive machine learning approach,” *Plos one*, vol. 13, no. 11, e0206872, 2018.

[9] X. Gu, Q. Zhang, J. Li, V. P. Singh, J. Liu, P. Sun, and C. Cheng, “Attribution of global soil moisture drying to human activities: A quantitative viewpoint,” *Geophysical Research Letters*, vol. 46, no. 5, pp. 2573–2582, 2019.

[10] R. Mandal, S. Joseph, A. Sahai, R. Phani, A. Dey, R. Chattopadhyay, and D. Pattanaik, “Real time extended range prediction of heat waves over india,” *Scientific reports*, vol. 9, no. 1, p. 9008, 2019.

[11] N. Khan, S. Shahid, T. B. Ismail, and F. Behlil, “Prediction of heat waves over pakistan using support vector machine algorithm in the context of climate change,” *Stochastic Environmental Research and Risk Assessment*, vol. 35, pp. 1335–1353, 2021.

[12] World Health Organization (WHO) and United Nations Children's Fund (UNICEF), "Technical brief on protecting maternal, newborn and child health from the impacts of climate change," WHO/UNICEF, Tech. Rep., 2021, Recommendation on nighttime temperatures 20°C for heat stress risk in tropical climates (Page 12). [Online]. Available: https://www.who.int/publications/i/item/9789240035852.

[13] S. B. H. S. Asadollah, N. Khan, A. Sharafati, S. Shahid, E.-S. Chung, and X.-J. Wang, "Prediction of heat waves using meteorological variables in diverse regions of iran with advanced machine learning models," *Stochastic environmental research and risk assessment*, pp. 1–16, 2022.

[14] W. P. F.S. Sourjah, "Heat wave prediction using machine learning techniques: A review," *International Journal of Computer Applications*, vol. 184, pp. 33–40, 2022.

[15] A. Sharma, G. Andhikaputra, and Y.-C. Wang, "Heatwaves in south asia: Characterization, consequences on human health, and adaptation strategies," *Atmosphere*, vol. 13, no. 5, p. 734, 2022.

[16] S. Byagar, A. A. Deshmukh, K. Wanjale, V. Wadne, N. Ranjan, and R. Gangarde, "Heat wave prediction using recurrent neural networks based on deep learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1s, pp. 612–619, 2023.

[17] P. Li, Y. Yu, D. Huang, Z.-H. Wang, and A. Sharma, "Regional heatwave prediction using graph neural network and weather station data," *Geophysical Research Letters*, vol. 50, no. 7, e2023GL103405, 2023.

[18] J. Ratnam, S. K. Behera, M. Nonaka, P. Martineau, and K. R. Patil, "Predicting maximum temperatures over india 10-days ahead using machine learning models," *Scientific Reports*, vol. 13, no. 1, p. 17 208, 2023.

[19] E. Weirich-Benet, M. Pyrina, B. Jiménez-Esteve, E. Fraenkel, J. Cohen, and D. I. Domeisen, "Subseasonal prediction of central european summer heatwaves with linear and random forest machine learning models," *Artificial Intelligence for the Earth Systems*, vol. 2, no. 2, e220038, 2023.

[20] G. Wu, C. Zhang, R. Zhao, P. Qin, and Y. Qin, "Asymmetries of the lag between air temperature and insolation in gauge observations and reanalyses over china," *Atmospheric Research*, vol. 288, p. 106 729, 2023.

[21] X. Zhang, C. Zhong, J. Zhang, T. Wang, and W. W. Ng, "Robust recurrent neural networks for time series forecasting," *Neurocomputing*, vol. 526, pp. 143–157, 2023.

[22] S. Byagar, A. A. Deshmukh, K. Wanjale, V. S. Wadne, N. Ranjan, and R. Gangarde, "Heat vave prediction using recurrent neural networks based on deep learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, pp. 612–619, 2024.

[23] D. Cortés-Guzmán, D. E. Bowler, and P. Haase, "Spatial and temporal effects of heat waves on the diversity of european stream invertebrate communities," *Science of The Total Environment*, p. 176 229, 2024.

[24] Y. Gong, Y. Zhang, F. Wang, and C.-H. Lee, "Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data," *arXiv preprint arXiv:2410.14963*, 2024.

[25] S. Hossain, "Temperatures may soar past 45°c by 2030: Study," *The Business Standard*, 2024. [Online]. Available: https://www.tbsnews.net/bangladesh/environment/temperatures-may-soar-past-45degc-2030-study-890261.

[26] G. M. Rashid, A. Akhter, and M. Mallik, "The study of heat wave condition in bangladesh during 1990 to 2019," *The ATMOSPHERE*, vol. 10, 2024.

[27] M. B. Rashid, A. Sultana, S. Q. Hassan, E. Kuya, K. Parding, and H. O. Hygen, "Changing climate of bangladesh," Copernicus Meetings, Tech. Rep., 2024.

[28] M. S. Yett, "Children are at high risk amid countrywide heatwave in bangladesh in bangladesh," *United Nations*, 2024. [Online]. Available: https://bangladesh.un.org/en/267046-children-are-high-risk-amid-countrywide-heatwave-bangladesh.

[29] N. Korhonen, O. Hyvärinen, V. Kollanus, T. Lanki, J. Jokisalo, R. Kosonen, D. S. Richardson, and K. Jylhä, "The probabilistic skill of extended-range heat wave forecasts over europe," *Natural Hazards and Earth System Sciences*, vol. 25, no. 6, pp. 1865–1879, 2025.

[30] M. S. Miah, S. Ullah, and M. H. Siddique, "Heat waves in bangladesh: Understanding the threats and finding solutions,"

[31] *Brri annual report for july 2020-june 2021*, Bangladesh Rice Research Institute, 2021. [Online]. Available: https://brri.portal.gov.bd/sites/default/files/files/brri.portal.gov.bd/page/6d86d72e_953c_4f74_9157_3cfeaa80c2b1/2021-10-13-05-28-667ef1c70f3d1cbb1bbcf9953b4bd4c3.pdf.