Thesis Report

Weather Prediction in Bangladesh Using Distinct Artificial Intelligence Techniques

SUBMITTED BY

Joysri Rani Das

Roll: BKH1911018F

Session: 2018-19

Year 04, Term 02

Department of Information and Communication Engineering, NSTU

THESIS SUPERVISOR

Dr. Md. Ashikur Rahman Khan

Professor

Department of Information and Communication Engineering, NSTU

DEPARTMENT OF INFORMATION AND COMMUNICATIONENGINEERING

NOAKHALI SCIENCE AND TECHNOLOGY UNIVERSITY

MAY 2024

A Thesis Report on Weather prediction in Bangladesh using distinct Artificial Intelligence Techniques.

Joysri Rani Das

BKH1911018F

Department of Information and Communication

EngineeringNoakhali Science and Technology University

DECLARATION

I hereby declare that this research report has not been submitted elsewhere for any Degree, Diploma,or Publication requirement.

Joysri Rani Das

Roll No: BKH1911018F

Session: 2018-19

ACCEPTANCE

This research proposal is submitted to the Department of Information & Communication Engineering, Noakhali Science & Technology University, Sonapur, Noakhali, in partial fulfillment of the requirements for having the B.Sc. degree in ICE.

Dr. Ashikur Rahman Khan
Department of Information and Communication Engineering
Noakhali Science & Technology University
Sonapur, Noakhali-3814

ACKNOWLEDGEMENT

I am deeply grateful to my Supervisor, Professor Dr. Md Ashikur Rahman Khan, for his unwavering support and guidance throughout my Research Work. His assistance has been invaluable in the completion of this thesis. I sincerely appreciate all of his help, and his practical insights and advice have been instrumental during my academic journey here. I am confident that the lessons learned will benefit me greatly in my future endeavors.

I would also like to extend my heartfelt thanks to my seniors and the staff members of the Information and Communication Engineering Department at NSTU, who have provided support in various ways. Additionally, I am grateful to the authorities of the Bangladesh Meteorological Department that helped us to collect data. Lastly, I am profoundly thankful to my parents and family member for their unwavering encouragement and support throughout my life. Words cannot adequately express my gratitude for everything they have done for me.

ABSTRACT

Weather forecasting holds immense significance in Bangladesh due to its heavy reliance on agriculture and vulnerability to weather-related risks. This study aims to develop reliable weatherforecast models for Rainfall, Temperature, and Humidity using artificial intelligence algorithms such as Decision Trees, Random Forests, K-Nearest Neighbors, Support Vector Machines, and Multilayer Perceptron Neural Networks. Historical weather data from the Bangladesh Meteorological Department (BMD) undergo preprocessing to handle missing values and outliers. Various AI techniques are then applied, leveraging their ability to address complex and nonlinear interactions in data. The prediction models are evaluated using performance metrics like Mean Absolute Error and Root Mean Square Error, Root Mean Square Error, R-square accuracy, R-square Value, and Relative Absolute Error. This research contributes to advancing weather forecasting in Bangladesh, benefiting sectors like agriculture, disaster management, and public safety. We divided our Temperature, Humidity, and Rainfall Datasets into two intervals: 1999-2022 which consist of 34 weather stations and 1980- 2022 consist of 26 weather stations.And we used 80% of our data for training and 20% for testing our models. We analyzed our data using both interval and did a detailed analysis of the yearly, monthly, and seasonal trends of Temperature, Rainfall, and Humidity. It also highlights the relationship between These weather parameters. From our results, we have found that Random Forest and Extreme Gradient Boosting Perform better than the Support vector machine, K-Nearest Neighbor, Multilayer neural network, and Decision Tree algorithm. And It is also noticeable that The results of the time interval 1999-2022 are better than the time interval 1980-2022.

KEYWORDS**:** Artificial Intelligence; Machine Learning; Random Forest; Decision Tree; Support Vector Machine; K-Nearest Neighbor; Extreme Gradient Boosting; MSE; RMSE; MAE; RAE; R-Square Value; R-Square Accuracy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

2.1 INTRODUCTION

The term weather describes the condition of the atmosphere at a specific location and time, which may include variables like temperature, Humidity, precipitation, wind direction and speed, atmospheric pressure, and cloud cover. The Earth's rotation, solar radiation, and interactions between air masses are just a few of the natural processes that have an impact on this dynamic and ever-changing component of our environment.

The core field of atmospheric sciences is weather forecasting, which sits at the nexus of science, technology, and the ever-present dynamics of our natural world. Understanding and predicting weather patterns has wide-ranging effects on many facets of society, having an impact on our everyday lives, business operations, and the safety of communities throughout. The subject of weather forecasting has entered a new phase recently thanks to developments in observational methods and the expansion of processing power. Meteorologists and researchers now have the ability to create ever more complex predictive models because of access to an unprecedented number of data from ground-based sensors, satellites, and remote sensing devices.

AI algorithms process historical climate data to uncover patterns and relationships, aiding in understanding long-term climate trends like temperature fluctuations and precipitation patterns. Machine learning (ML) techniques enable the development of sophisticated climate models, predicting future climate conditions.

Bangladesh, a nation heavily affected by climate change, faces challenges such as sea-level rise, storm surges, tidal flooding, and erosion due to its low-lying topography and extensive coastline.

Cyclones, flooding, and excessive rainfall have become more frequent and severe. Rising temperatures, changing rainfall, and pest outbreaks disrupt agriculture, endangering crop production, food availability, and rural incomes. Weather forecasting is critical for Bangladesh's economy, particularly its agriculture, which contributes 20% to the GDP. With 70% of the population in rural areas and 60% relying on agriculture, variations in rainfall, Humidity, wind, and temperature can impact agriculture and exports, representing 13-18% of the GDP.

The importance of weather forecasting resides not only in its immediate repercussions but also in its position as a crucial instrument for resource management, risk reduction, and well-informed decision-making. Accurate and timely weather forecasts play a crucial role in everything from helping farmers plan their crops to helping the aviation, aiding in disaster preparation to changing consumer habits.

The ability to anticipate the weather is crucial in many facets of modern life, including transportation, agriculture, disaster relief, and environmental planning. The ability to make educated decisions that can reduce hazards, save lives, and increase economic output depends on accurate and timely weather forecasts. The need for more accurate and trustworthy weather forecast technologies has never been greater due to the rising frequency and intensity of extreme weather occurrences and the expanding effects of climate change. Utilizing artificial intelligence, machine learning, and sophisticated data analytics, the ultimate goal of this research is to develop weather forecast methods. This study intends to enhance the accuracy and lead time of weather forecasts by creating and improving predictive models, improving data assimilation methods, and investigating new sources of meteorological data. The goal of this thesis is to study the state of weather forecasting, its urgent problems, and the cutting-edge methods that have the potential to improve forecast precision and lead times. This project aims to contribute to the continued development of meteorological science by exploring the methodologies, technology, and data assimilation methods that support contemporary weather prediction.

This study's ultimate goal is to offer insights and recommendations that can guide advancements in weather prediction, reducing the effects of extreme weather events, optimizing resource allocation, and enabling informed decision-making across various sectors.

The significance of this research is found in its broader implications for industries like agriculture, aviation, energy, and disaster preparedness, in addition to its potential to improve our capacity to foresee and respond to weather-related difficulties. In addition, there is an increasing demand for flexible and resilient weather prediction systems that can adapt to changing circumstances and deliver accurate forecasts in a fluid environment as climate change continues to affect global weather patterns.

1.2 PROBLEM STATEMENT

The ability to anticipate the weather is crucial for protecting lives, conserving resources, and reducing the negative effects of catastrophic weather events. Due to the complexity, nonlinearity, and dynamical nature of the Earth's atmosphere, traditional meteorological models, while sometimes effective, frequently struggle to predict weather conditions effectively. Artificial intelligence (AI) has recently made it possible to find new ways to increase the accuracy and dependability of weather forecasts. However, despite important developments in AI, there is still an urgent need to research and create unique AI methods suited to the particular difficulties of weather prediction.

This thesis' main concern is with the present weather prediction models' subpar performance and the field's scant use of cutting-edge AI technologies. This project intends to investigate, assess, and improve various AI methods particularly created for weather prediction in order to address this issue. Machine learning algorithms, deep learning models, neural networks, and other cutting-edge technologies that have shown promise in other fields but are neglected in meteorology will be included in these AI methodologies. The following issue could be resolved via weather forecasting.

i. Vulnerability to Climate Change in Bangladesh

In the 2020 edition of German Watch's Climate Risk Index, Bangladesh was ranked seventh among the Nation's most severely affected by climate-related disasters between 1999 and 2018 because of its high population density (166 million people live in 147,570 km2), dependence on extractive industries, and frequent natural disasters. Bangladesh is one of the Countries most vulnerable to the far-reaching effects of climate change, so it is urgent to address this pressing issue Impacts of Climate Change in Bangladesh.

ii. Impacts of Climate Change in Bangladesh

The frequency and severity of natural disasters, such as those brought on by greater rainfall, rising sea levels, and tropical cyclones, are predicted to grow as climate change takes hold. In Bangladesh, these occurrences pose serious risks to agriculture, public health, food and water security, and housing. An unprecedented heatwave is currently sweeping the country, causing energy and economic problems to worsen. Temperatures are made worse by the protracted dry period and low humidity levels, which cause them to rise above 40 °C (104 °F). Such heat waves put a burden on the economy's stability and energy supplies.

iii. Agricultural Challenges Due to Climate Change

An important component of Bangladesh's economy, agriculture, suffers significantly as a result of the changing environment. Alarmingly, it was predicted that most countries, including Bangladesh, will experience a 50% reduction in rain-fed agriculture output by 2020, undermining food security in a country that already struggles with famine. As agricultural production is highly correlated with temperature and rainfall, the ongoing weather changes hurt crop yields and limit the amount of arable land. For instance, Aman rice yield was significantly reduced at several growth stages when the maximum temperature was raised by just one degree Celsius. Furthermore, rice production, a mainstay in Bangladesh, is further threatened by water salinity, particularly in the coastal area.

ii. Necessity for Long-Term Climate Adaption in Bangladesh

Bangladesh needs to prepare for long-term climate adaptation in light of these difficulties. Accurate and fast weather forecasts are essential to enable proactive steps for risk mitigation, decision-making, and readiness. However, Bangladesh's dynamic climate, which is characterized by variations in temperature, rainfall, and cyclone activity, poses considerable challenges for accurate weather forecasting using conventional techniques.

iii. Role of Artificial Intelligence in Weather Prediction

This study attempts to analyze and evaluate several artificial intelligence (AI) techniques, such as decision trees, K-nearest neighbors, support vector machines , multilayer neural networks , and random forests, in order to address these issues. The objective is to improve Bangladesh's weather forecast models' accuracy and dependability, particularly for occurrences like rainfall, floods, droughts, temperature changes, and cyclones.

The project's main objective is to gather and examine historical meteorological data from Bangladesh while assessing the effectiveness of several AI algorithms. The results of this study will offer important new information about the suitability and effectiveness of AI methods for Bangladeshi climate forecasts. This information will considerably improve risk management, decision-making, and planning techniques for controlling weather-related risks and their effects.

1.3 RESEARCH OBJECTIVES

In our thesis work, we are dedicated to achieving the following objectives:

I. To analyze the historical data of various weather parameters such as Temperature, Rainfall, and Humidity.

II. To Analyze the yearly, monthly, and seasonal Trends of this parameter.

III. To find the relationship Between Temperature, Rainfall, and Humidity.

IV. To develop an intelligent model for predicting climate that will calculate climate (rainfall, temperature, and Humidity).

V. To evaluate the performance of different machine learning methods, such as MPNN, Decision Tree, SVM, KNN, and Random Forests using different performance metrics (MAE, MSE, RMSE, MAPE).

VI. To Compare several algorithms and methodologies, provide analyses and recommendations for selecting the best appropriate artificial method for weather forecasting in Bangladesh.

1.4 SCOPE OF THIS RESEARCH

This thesis proposal focuses on comprehensive weather forecasting in Bangladesh using a wide range of artificial intelligence (AI) techniques. It includes gathering and analyzing historical weather data from 1980 to 2022 obtained from reputable sources. Various AI methods such as machine learning algorithms, neural networks, and decision trees will be employed to develop advanced weather prediction models capable of forecasting multiple variables like rainfall, temperature, floods, Humidity, wind speed, and cyclones. The research also considers customized models for different regions within Bangladesh, conducts rigorous modelevaluations, and addresses challenges and potential future developments. Additionally, there's scope to develop a userfriendly weather forecasting app for public use. In this thesis, we are predicting weather in Bangladesh; there is scope to extend the research to predict weather in other regions, such as other countries, the broader Asian region, or even different continents. Here, data is taken from a time period spanning 1980-2022. important to keep in mind that there is flexibility to take into account data before 1980 or include futuredata after 2022. In the next Chapter we added some related work and analyzed the

1.5 THESIS STRUCTURE

Our thesis report is structured as follows:

Chapter 1: Introduction – In This Chapter we set the stage for the research, providing a comprehensive overview of the study's background, significance, objectives, and scope. It introduces the problem statement or research question, delineates the rationale behind the research, and outlines the structure of the thesis.

Chapter 2: Literature Review - This section critically reviews existing literature relevant to the research topic, synthesizing key findings, theories, and methodologies. It identifies gaps, contradictions, and areas for further investigation, serving as the theoretical framework for the study.

Chapter 3: Methodology—Detailing the research design, data collection methods, prediction model, performance metrics, and data analysis procedures, this Chapter elucidates how the research objectives will be achieved. It provides a clear roadmap for conducting the study and ensures its reliability and validity.

Chapter 4: Results and Analysis—In this Chapter, we Present the empirical findings derived from the research. This Chapter analyzes the historical data.It employs various statistical or qualitative techniques to explore patterns, trends, relationships, and deviations, aligning the findings with the research objectives.

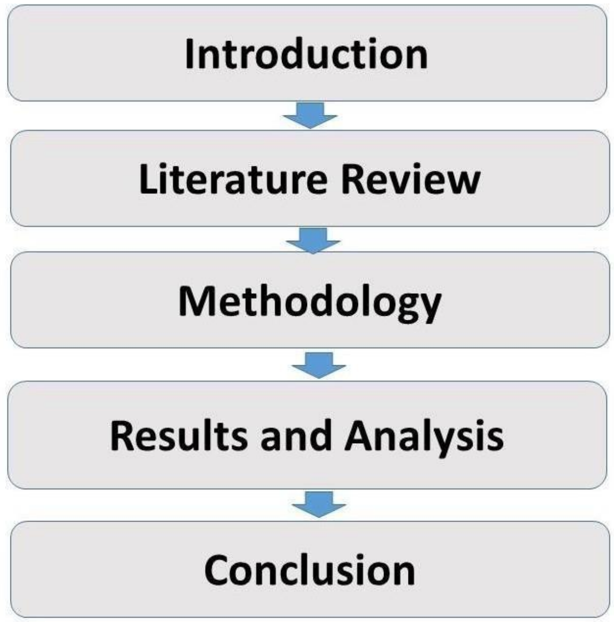


Figure 1.1: Structure of the thesis

Chapter 5: Conclusion – This Chapter include Summarizing the key findings, implications, andcontributions of the study, this Chapter offers insights into the broader significance of the research. It discusses the implications for theory, practice, and policy, highlights limitations, and suggests avenues for future research, providing a coherent closure to the thesis.

1.5 CONTRIBUTIONS

1. We utilized a comprehensive dataset spanning 43 years, providing an unprecedented depth of data for climate studies.
2. We applied six different machine learning techniques within one study, showcasing a diverse methodological approach.
3. We divided temperature, humidity, and rainfall data into two distinct intervals to capture variations over time.
4. We conducted a thorough analysis of temperature, rainfall, and humidity on yearly, monthly, and seasonal scales to identify broad climate trends.
5. We made predictions for both data intervals, offering insights into short-term and long-

term climate patterns.

6. We predicted climate patterns for eight different divisions, introducing a novel approach to regional climate forecasting.

7. We generated predictions on both a daily and monthly basis to provide detailed and comprehensive climate insights.

1.6 SUMMARY

This chapter provides a comprehensive introduction to the research, highlighting the significance of weather forecasting in Bangladesh's socio-economic context due to its vulnerability to climate change. It outlines the thesis's objectives, scope, and structure, emphasizing the urgent need for advanced AI techniques in weather prediction. Chapter also presents the problem statement, focusing on the limitations of traditional meteorological models and the potential of AI to address these challenges. Additionally, it discusses the research objectives and system requirements necessary for conducting the study.

CHAPTER – 2

LITERATURE REVIEW

2.1 INTRODUCTION

Weather prediction is vital for numerous sectors, and AI techniques offer promising avenues for enhancing forecast accuracy. This literature review aims to synthesize current research on weather prediction using AI methods. By analyzing trends, methodologies, and gaps in the literature, we seek to inform our own research and contribute to advancements in the field. Understanding existing knowledge is crucial for developing effective strategies and addressing challenges in weather forecasting. This review is structured to achieve objectives such as identifying trends, assessing methodologies, and providing a theoretical framework. Through this exploration, we aim to contribute to the broader understanding of AI applications in weather prediction.

2.2 OBJECTIVES OF THIS REVIEW

Here are the objectives of

I. Identify existing knowledge in weather prediction and AI techniques.

II. Analyze methodologies and techniques used in weather prediction.

III. Identify gaps and research questions in the literature.

IV. Assess trends and developments in weather prediction research.

V. Provide a theoretical framework for the study.

2.3 PRIOR WORK

The use of artificial intelligence (AI) in automated weather prediction has been the subject of a large amount of research over the past few decades. Numerous studies have examined various strategies and techniques to improve forecasting accuracy in this field. Expert systems, fuzzy logic, data mining, machine learning (ML), and deep learning are some of these techniques. These strategies have been used by researchers to create cutting-edge models and algorithms that can examine meteorological data and produce forecasts. The incorporation of AI technology has paved the way for developments in the sector by opening up new opportunities for enhancing the precision and effectiveness of weather forecasting systems. We will discuss few related approaches have been studied yet.

In 2019, S. Nalluri, S. Ramasubbareddy and G.Kannayaram [12] research on Weather Prediction by Using Clustering Strategies. They have used various clustering technique such as simple k-means, Hierarchical clustering, Density based clustering, Filtered clustering and Farthest first clustering. The data set they have used consists several attributes which includes formatted date, Summary, Temperature, Precipitation type, Humidity, Wind speed, and daily summary, cloud coverage pressure. After applying all AI technique, in result analysis, they have found that K-means take 0.01 second to build the model, where filtered clustering take 1 second, filtered clustering take 0.01s. Where make density clustering consumes much time than based on the other algorithms, it takes 2 seconds to build the model.

In 2007 Mohsen Hayati, and Zahra Mohebi [13] worked on Temperature fore casting Using Artificial Technique. They have focused on fore casting short term temperature system in Kermanshah city, which is located in west of Iran. For this research they have used 10 years of meteorological data from 1996 to 2006. They have collected their data from meteorological department of Kermanshah, Iran. They trained and tested their model using these data. Multilayer perceptron neural network which is an well-known Architecture in Artificial Intelligence have been used to model Short term Temperature forecasting. They have used 67% of data for train the model and remaining 33% used to test the model. They have Consider mean temperature, maximum temperature, minimum temperature, Precipitation, mean humidity, mean pressure, sunshine, radiation and evaporation as an attribute. They have found that MLP obtained minimum error which have a good prediction and reasonable prediction accuracy. They have cleaned their data to fill the missing values as the real world database are highly susceptible to noisy and missing data. They have normalized their data using z-score normalization.

Y.Radhika and M.Shashi [14] worked in 2009 to predict maximum temperature using Support vector machine. Data for this research has been collected from University of Cambridge for a period of five years in a time spanning 2003 to 2008. Between these data they have used data from 2003-2007 to build the model and data between January and July of the year of 2008 has been used to test the model. A non-linear Support Vector regression has been used to train the SVM. They have also used MLP to build model and used Backpropagation Technique to train the algorithm. They have observed that SVM perform better than MLP. The Mean Square Error in the case of MLP varies from 8.07 to 10.2 based on the order, whereas it is in the range of 7.07 to 7.56 in the case of SVM.

A similar method SVM was used in 2020 by Binh Thai Pham et al. [1] to predict Daily Rainfall. In their study then have developed and compare several AI model namely Adaptive based Fuzzy Inference System optimized with Particle Swarm Optimization (PSOANFIS), Artificial Neural Networks (ANN) and Support Vector Machine for the prediction of daily rainfall in Hoa Binh province, Viatnam. For this research they have used parameters such as maximum temperature, minimum temperature, wind speed, relative humidity and solar radiation were collected and used as input parameters and daily rainfall as an output parameter in the model. Several quality assessment metrics, including the correlation coefficient (R) and mean absolute error , skillscore (SS), probability of journal pre-proof journal pre-proof 2 detection (POD), critical success index (CSI), and false alarm ratio (FAR), were used to validate the developed models. All of the AI models produced good daily rainfall predictions, according to the data, but the SVM was shownto be the most accurate technique. Using the Monte Carlo method, it was also discovered that this strategy was the most reliable and effective prediction model when input variability was taken into account. This AI-based study will aid in the rapid and precise forecasting of everyday rainfall.

Wondmagegn Taye Abebe et al. [2] They made an effort to look at the usage of artificial intelligence (AI) models to forecast monthly rainfall at 92 meteorological stations in Ethiopia. Using geographic and periodicity component (longitude, latitude, and altitude) data gathered from 2011 to 2021, the usefulness of Artificial Neural Networks (ANNs) with Adaptive Neuro-Fuzzy Inference System (ANFIS) models in forecasting long-term monthly precipitation was examined. Mohammad Mahmudur Rahman Khan et al. [15] Their forecasting model can be used to analyze seasonal diseases in Bangladesh whose incidences are influenced by local temperature and/or rainfall as well as variations in weather patterns.

A popular approach in AI which is feed forward MLP used by Taohdul Islam et al [4]. In 2016 to forecast monthly weather in Barishal City. They have collected their data from Gopalganj Station, in Bangladesh. They have used MATLAB to trained MLP. By the model they has been tested with individual month's average rainfall percentage to obtain the confirmation of the model, and they have Normalized their collected data by using premnmx function in Matlab 2013.They have used Total of eleven parameters of eight month's data as input, that is total of 11*245 sample data and target is set as daily rainfall in mm that is 1*245 sample data. From their collected data set, they have used 70% of data as training, 15% as validation and 15% as testing. After completion the training and testing they have observed that that the trained network gives similar result with respect to the measured rainfall. By using only eight month's data give above 83% accuracy.

Besides Support Vector Regression A very important approach in machine learning The Boosting Algorithm such as, (GB), Extreme Gradient Boosting Category Boosting Adaptive Boosting and also K-NN has been used by Atik Mahabub et al. [5] in 2019. They have worked to preict the weather including wind speed, humidity, rainfall and temperature in Bangladesh. They have collected their data from BMD of 5 years, time spanning of 2012-2017. As a Statistical measure they have used MAE, MSE, MAPE, which helps them to compare the used model. IN Case of Rainfall CatBoost performed relatively better than the other algorithms in terms of error percentage, with a MAPE value of 32.93%; the other algorithms, KNN, XGBoost, GB, DTR, and SVR, performed moderately to quite satisfactorily, with MAPE values of 39.72%, 35.71%, 35.46%, 46.73%, and 35.07%, respectively. Their error percentage values approximately fall in

28% to 36%. In the research paper, we can see from the MAE, MSE, and MAPE data that DTR outperforms other algorithms in forecasting with a MAPE of just 2.73%, just as it did when forecasting wind speed. Thus, it once again demonstrates that DTR's structure and algorithmic tree creation play a crucial part in forecasting. Similar to the training and testing data, AdaBoost had the least effective performance and had the highest MAPE error percentage, which was roughly 84.54%. When compared to the performance of other algorithms, Linear Regression and Bayesian Ridge also displayed some inconsistent performance, with MAPE values of 64.31% and 64.28%, respectively. Other algorithm also performed reasonably well; their MAPE values for CatBoost, SVR, GB, XGBoost, and KNN were 36.93%, 45.91%, 43.67%, 43.07%, and 46.13%, respectively.

An App was developed by M. N. H. Khan1 et al. [16] in 2013 which will aware farmers. This app will help to mitigate the problem such as harvested rice washed out by rain, seeds from seedbed washed out due to rain, rainfall occurs just few Hours after irrigation as a result loss energy, seeds wet and damage due to rainfall when drying as a result decrease seed quality, washed out fertilizer and pesticide just after applied and so on. This will forecast the weather 2 to 3 days in the future. In weather an important perspective in Flood which damages a lot and can occur for multiple reason, In 2019 Talha Khan et al. [6] research on flash flood prediction and they have used real time data. In order to estimate hurricanes and flash floods more accurately, a unique predictive hybrid algorithm (ANN PSO) has been used in this research. When compared to the use of a single sensor, an appropriate combination of the sensors will provide the benefit of greater precision and accuracy. In This study they has examined the six process factors that were combined for the measurement and investigation of the flash flood. PIR, ultrasonic sensor, temperature sensor, CO2 sensor, rainfall module, pressure sensor, and temperature sensor have all recorded real-time data for more than Forty eight hours. They have used PSO optimization technique to optimized the neurons. In the result it is shown that Compared to current methods, AI-based systems for hurricane and flash flood early detection have operated more accurately and efficiently. They have also used MLP-PSO which provide better result than ANN-PSO. Its accuracy level is 95.15%.

Rainfall is an important parameter of weather, heavy rainfall is responsible for flood, Several Work has been done in previous to predict rainfall, In 2016 Emilcy Hern´andez et al [7]. used deep Learning Approaches to predict Rainfall. They have collected their data from from an environmental data warehouse administered by IDEA. They used MLP and ANN to build their model. In the paper, they displays the typical MSE and RMSE for each implemented method's optimal network setup. To obtain the average, each network is trained 30 times, counting times and comparing the MSE and RMSE to the test set. After one year later Imrus Salehin et al. [8] predict Rainfall using Long Short Term Method which is a very sophisticated method in AI. They have developed a model that will assist in estimating the amount of rainfall. To do this, they have collected data from six different regions. Temperature, dew point, humidity, wind pressure, wind speed, and wind direction are the six parameters we used to make their predictions. They obtained a 76% accuracy rate for their job after examining all of the data. For the best outcome, they also concentrate on a sizable dataset in long-term weather. In 2009 another work has been done to predict rainfall by M. BILGILI et al. [9] Artificial neural networks were used in this study to forecast the long-term monthly temperature and rainfall at any target point in Turkey using information from nearby monitoring stations. As training (59 stations) and testing (17 stations) data, 76 measuring stations' worth of meteorological data collected between 1975 and 2006 by the Turkish State Meteorological Service were used. In the input layer of the network, four neurons that receive input signals of latitude, longitude, altitude, and month were used. The output layer of the network made use of two neurons that provide corresponding output signals for long-term monthly temperature and rainfall.

A correlation between rainfall with temperature, wind Direction, humidity And date was developed by Nazim Osman Bushara et al. [10] They have used a numerous number of AI technique to found the correlation between these attributes. In this study, an attempt has been made to look at how rainfall in Sudan relates to crucial variables such station, wind direction, date, humidity, minimum temperature, maximum temperature, and wind speed. The relationship between rainfall and these elements has been studied. The aim of this study was, show how feature selection can be used to determine the correlations between rainfall occurrences and other weather conditions and show which classifiers can provide the most precise rainfall estimations. They have collected their data from The Central Bureau of Statistics Sudan and used monthly weather data from 2000 to 2012 for 24 meteorological sites. They employed a collection of data mining algorithms to perform feature selection and create prediction models. The investigation demonstrates that Date, Min-T, Humidity, and Wind have an impact on rainfall in Sudan, and they were able to identify the top 14 algorithms for creating rainfall prediction models. They have used difference performance metrics to make comparison among the algorithm they have used.

To predict drought status in Bangladesh in 2020 Md. Jalal Uddin et al. [11] The Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) were computed in this study using daily temperature and precipitation data from the Bangladesh Meteorological Department (BMD). Then, in order to determine the strengths of SPI and SPEI, we conducted a statistical analysis using tools like the Pearson correlation coefficient, cross-correlation, cross-wavelet transform, and root mean square error. Our results demonstrated that, despite the substantial association between the two indices, SPEI outperformed SPI because evaporative demand positively affects the definition of drought conditions in Bangladesh. Cross-wavelet transform has been used to investigate the relationship between temperature and rainfall indices, such as warm spell duration indicator (WSDI) and the maximum amount of rain (Rx5day) that occurs over five consecutive days. Our findings suggested that Bangladeshi droughts are influenced by climatic extremes like WSDI and Rx5day. Our findings suggested that in order to reach a solid judgment, multiple drought assessment indices should be used.

In 2007 to forecast Rainfall ANN and large scale climate indices was used by D. Negesh Kumar et al. [17]]This study proposes an Artificial Intelligence (AI) approach for forecasting regional rainfall in Orissa state, India, utilizing large-scale climate indices and Artificial Neural Networks (ANNs). By optimizing ANN architecture with Genetic Optimizer (GO), the study achieves reasonably good accuracy in both monthly and seasonal rainfall predictions. Data sources include climate indices and rainfall anomalies collected from reputable institutions like the Climate Analysis Centre, NCEP, USA, and the Indian Institute of Tropical Meteorology. Results demonstrate high correlation coefficients for individual months and the summer monsoon season (June to September), underscoring the effectiveness of AI-driven methods in capturing rainfall patterns. Overall, the study highlights the significance of integrating AI techniques with climatic data for improved regional rainfall forecasting, offering valuable insights for meteorological research and practical applications.

Table 2.1 Summary of the Literature Review

Serial No.	Author, Year and Reference No.	Source of Data	Implemented Methods	Observations	Limitations
1	Nalluri Sravani et al., 2019 [12]	Not found	Simple k-means, Hierarchical clustering, Density based clustering, Filtered clustering and Farthest first clustering.	Overall weather (rain, temp, Humidity, wind speed etc.)	Does not mention the source of data
2	Z. M. Mohsen Hayati et al.200713]	Weather data of ten years were collected from the Meteorological department of Kermanshah, Iran	MLP	Temperature	1.Used Limited Data Range (only 10 Years) 2. Geographic Limitations
3	M. Shashi et al. 2009[14]	The weather data of University of Cambridge for a period of five years (2003-2008)	A nonlinear support vector regression method is used to train the SVM. MLP trained with backpropagation algorithm.	Maximum Temperature	1.Limited data Used (only 2003 to 2007) 2. Few testing data
4	T.A. Demeke, 2023[2]	The rainfall data from 92 meteorological stations within the study area (Ethiopia.	Feed forward MLP applied for rainfall prediction. Levenberg–Marquardt, gradient descent, gradient descent with adaptive learning rate, gradient descent with	Rainfall	1.Doesn't mention the quality of data 2.Used only 10 years data

			momentum, adaptive learning rate, and scaled conjugate gradient are used For training		
5	M.M.R.K et al. 2020[15]	Kaggle(1901-2015)	LTSM(RNN)	Rainfall and temperature	1.Only 15 years' data 2. Used only one Algorithm 3. Used Online Data
6.	T.l.S Saha et al. 2016,[4]	Data collected from gopalgange station.	Feed-forward MLP	Rainfall in Barisal City	1.Predict only on Attribute
7	S.B. Habib et al. 2019[3]	BMD(5 years data)	SVR Linear Regression, Bayesian Ridge, (GB), Extreme Gradient Boosting Category Boosting Adaptive Boosting and Decision Tree	Wind speed, humidity, rainfall and temperature	1.Only 5 years data
8	M.N.H.K.M. A et al. 2013[16]	Not found	Not found	Temperature, humidity, Rainfall and wind	1.Forecast data for Only 2-3 Days in advance

9	M.j Uddin et al. 2020[11]	BMD	SPI and SPEI	Drought	1.Predict only Drought
10	M.A.Talha Khan et al. 2019[6]	Data has been collected from the multi-modal sensing device comprising of PIR, ultrasonic, temperature, CO2, and pressure and altimeter transducers.	Particle swarm optimization (PSO) MLP-PSO	Flash flood	1.Used Only PSO Algorithm 2.Predict only Flood
11	B.T Pham et al. 2020[1]	Data collected from meteorological department of, Vietnam	ANN ANFIS PSO SVM	Daily Rainfall	1.Predict only Daily Rainfall
12	E.H Sanchez 2016[7]	The data is taken from an environmental data warehouse administered by IDEA	MLP ANN	Rainfall	1. Used Only one algorithm 2. Predict only rainfall
13	L.Salehin et al.	BMD	ANN, LTSM, RNN	Rainfall	1.Only one Attribute

	2020[8]				
14	N.O Bhusharaet al. 2014[10]	Central Bureau of Statistics Sudan (2000 to 2012) for 24 meteorological stations	1.Gaussian Processes 2. Linear Regression 3.Multilayer Perceptron 4. Kstar 5. IBK 6. Bagging 7.Random Subspace 8.Decision Table 9.User Classifier	Correlation	1.Used low range of data
15	M. BILGILI et al., 2009[9]	Meteorological-ll department of Turkey	ANN	Rainfall	1.Predict only Rainfall
16.	D. Negesh Kumar et al. 2007[17]	Website of Climate Centre, NCEP, Indian Institute of Tropical Meteorology	Large Scale Climate Indices and AI techniques	Rainfall	1.Does not used recent Data.

2.4 RESEARCH GAP

From the literature review, we identified the following gaps:

I. Limited Range of Data Used: The studies typically use data spanning only 5 to 10 years, which may not capture long-term trends and patterns.

II. Predict Few Attributes: Most studies focus on predicting only one or two attributes, limiting the comprehensiveness and applicability of the findings.

III. Does Not Use Data from Reliable Sources: Many studies rely on data that may not be from the most reliable or authoritative sources, potentially affecting the accuracy and validity of the results.

IV. Apply Few Algorithms: There is a limited application of different algorithms, which may restrict the exploration of more effective or innovative analytical methods.

V. Geographical Limitation: The research often focuses on a small geographical area, reducing the generalizability of the findings to broader or different regions.

VI. Does Not Analyze Historical Data Broadly: The analysis often overlooks a broad examination of historical data, which could provide more robust insights.

VII. Does Not Predict Both Daily and Monthly Data: The studies typically do not predict both daily and monthly data, limiting the scope of their predictive capabilities.

VIII. No Division of Time Interval: There is a lack of studies dividing data into specific time intervals for more detailed analysis.

2.5 SUMMARY

The research highlighted several AI-based approaches for weather prediction, encompassing clustering strategies, neural networks, and support vector machines. However, certain limitations should be noted. First, the study by S. Nalluri, S. Ramasubbareddy, and G. Kannayaram lacked detailed discussion of the specific attributes and their relevance in clustering. The efficiency comparison solely based on model building time may not fully reflect the algorithm's performance. Second, while MohsenHayati and Zahra Mohebi's work on short-term temperature forecasting showed promise with MLP, the research was conducted using data from 1996 to 2006, potentially limiting its applicability to current climate conditions. Third, Y. Radhika and M. Shashi's study on maximumtemperature prediction using SVM and MLP primarily focused on 2003-2008 data, raisingquestions about the model's adaptability to changing weather patterns. Moreover, Binh Thai Phamet al.'s research on daily rainfall prediction employed several AI models but relied heavily on SVM, leaving room for exploring other algorithms' potential. Furthermore, the study by Atik Mahabub et al. on weather prediction in Bangladesh assessed various algorithms but didn't delve into the impact of changing climate scenarios on model accuracy. These limitations highlight the need for continued research and adaptation of AI models to address evolving climate dynamics and improve the accuracy and reliability of weather forecasting.

CHAPTER – 3

METHODOLOGY

3.1 INTRODUCTION

In the following chapter, we will discuss our overall Working process, including Feature Selection, Data Collection, and Data pre-processing. After that, we will discuss the variation in our data and the geographical coordinates used in our Study. Then, we will discuss various AI models that we will use in our study, like Random forest, Decision Tree, Support Vector Machine, K- Nearest Neighbor, Extreme Gradient Boosting, and Multilayer Neural Networks. Our primary goal is to develop a sophisticated and accurate intelligent model for climate prediction in Bangladesh, encompassing essential parameters like Temperature, Rainfall, and Humidity. This goal will be achieved by rigorously evaluating and comparing the performance of multiple machine learning methods, including MPNN, Decision Tree, SVM, KNN, and Random Forests, using various performance metrics, for example, Mean Absolute Error, Mean Square Error, Relative Absolute Error, Root Mean Square Error, R-square Accuracy, R-Square Value.

3.2 ARTIFICIAL INTELLIGENCE TECHNIQUES

Various artificial intelligence techniques have become a useful tool for improving weather prediction models in recent years. In Bangladesh, a country frequently affected by cyclones and floods, precise weather forecasting becomes increasingly vital for effective disaster management and optimal resource allocation. In our research, we will use various regression algorithms, including decision trees, multilayer perceptron neural networks, random forests, K nearest neighbors, and extreme gradient boosting.

A set of supervised machine learning techniques known as regression algorithms is used for predictive modeling and to comprehend the relationship between a dependent variable (the target) and one or more independent variables (features or predictors). The main goal of regression analysis is to find a mathematical function that best explains how changes in the independent variables correspond to changes in the dependent variable. Then, predictions or inferences can be drawn from this learned function, often expressed as a linear equation or a more sophisticated model. Regression algorithms are essential to studying artificial intelligence (AI) because they provide helpful tools for analyzing and predicting continuous numerical data.

In order for the system to estimate values or make predictions based on the input data, these algorithms simulate the relationship between an input variable and a target variable. Regression can take many different forms, such as linear regression for modeling linear relationships, polynomial regression for modeling nonlinear patterns, and more sophisticated methods like support vector and decision tree regression. Regression is frequently employed in situations where it is crucial to comprehend and predict numerical outputs based on input data, such as sales forecasting, stock price prediction, medical diagnosis, and other situations. Here, we will give simple basic information about the algorithm that we will use for our prediction.

3.2.1 Decision Tree

A decision tree is a popular machine-learning algorithm generally used for **classification and regression** tasks. It's a **supervised learning method** that is easy to understand and interpret, making it a valuable tool for solving many problems. A decision tree is a hierarchical structure resembling an upside-down tree. It consists of nodes and edges. The top node is called the root, and the final nodes are called the leaves. There are two types of nodes in a decision tree. One is the decision node, and the other is the leaf node. Decision Nodes (Internal Nodes) represent a test condition on one or more features.

Based on the outcome of the test, the data is split into subsets that move down the tree to the next set of nodes**. Leaf Nodes** (Terminal Nodes) represent the final predicted outcome or class label.

Once a data point reaches a leaf node, a decision is made. **Edges** represent the outcome of a test condition. Typically, there are two edges leaving each decision node, one for each possible outcome of the test. The process of dividing the data into subsets based on a feature's value is called **splitting**. The goal is to create splits that maximize the homogeneity of the data within each subset. Various methods, such as Gini impurity or information gain, are used to measure the homogeneity. The tree-building process is recursive. It continues until a stopping criterion is met, such as a maximum tree depth or the number of samples in a leaf node.

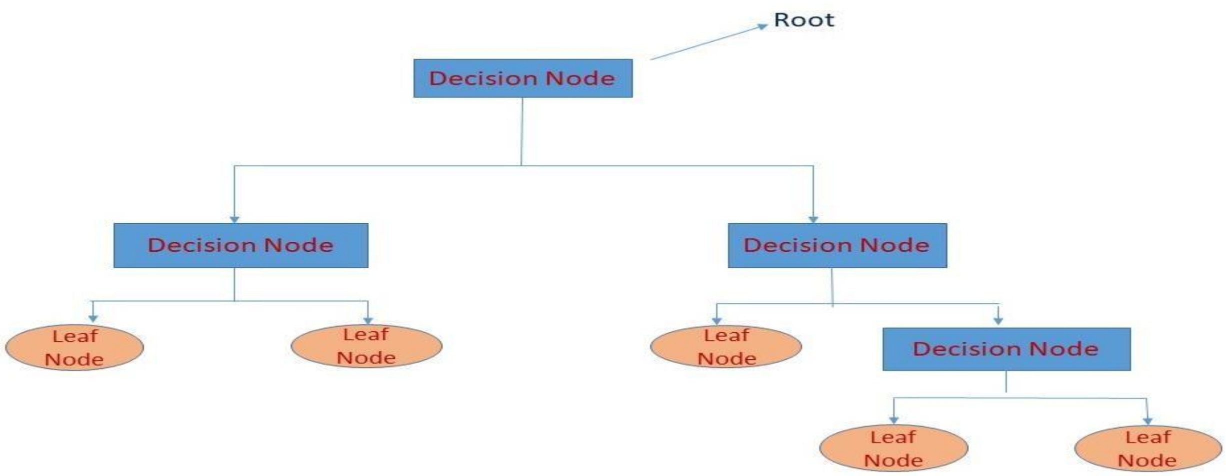


Figure 3.1 Decision Tree Algorithm Structure

3.2.2 Support Vector Machine

SVM is a versatile algorithm known for its effectiveness in both **classification** and **regression** tasks, making it well-suited for the multifaceted challenges inherent in weather forecasting. However, it is used for **Classification** problems in Machine Learning. Its core principle is to find an **optimal hyperplane** in a high-dimensional feature space that best separates data points of different classes while maximizing the margin between them. The goal of the SVM algorithm is to create the **best line or decision boundary** that can segregate n-dimensional space into classes so that we can easily put the new data points in the appropriate category in the future. This best decision boundary is called a **hyperplane.** SVM chooses the extreme points/vectors that help in creating the hyperplane. These values are called **support vectors**, and hence, the algorithm is termed a **Support Vector Machine**.

The working process of an SVM involves selecting the hyperplane that has the largest margin, meaning the largest separation between the nearest data points of different classes (support vectors). This hyperplane is chosen to **minimize classification errors** and improve **generalization** to unseen data. SVMs can handle linearly separable as well as non-linearly separable data by utilizing kernel functions that map data into higher-dimensional spaces, allowing for more complex decision boundaries. There can be multiple lines or decision boundaries to segregate the classes in n-dimensional space, but our requirement is to find out the best decision boundary that helps to classify the data points. This best boundary is called the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the Dataset, that means if we have 2 features, then the hyperplane will be a straight line. And if we have 3 features, then the hyperplane will be a 2-dimension plane.

Linear SVM:

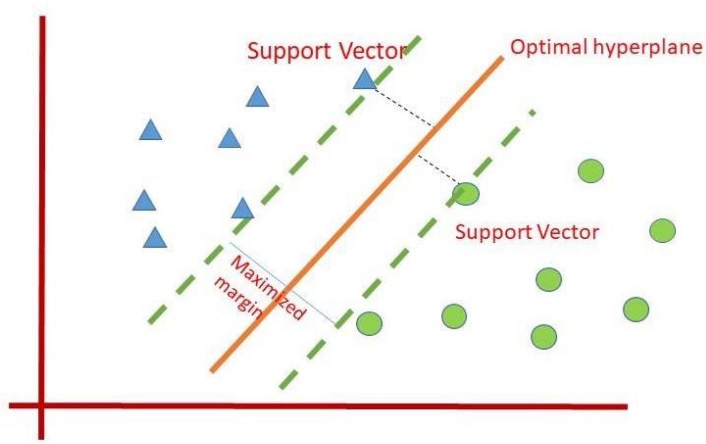


Figure 3.2: Support Vector Machine Structure for Linear Data Set

Non-Linear SVM:

If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension, z. It can be calculated as:

$$z=x^2+y^2 \tag{3.1}$$

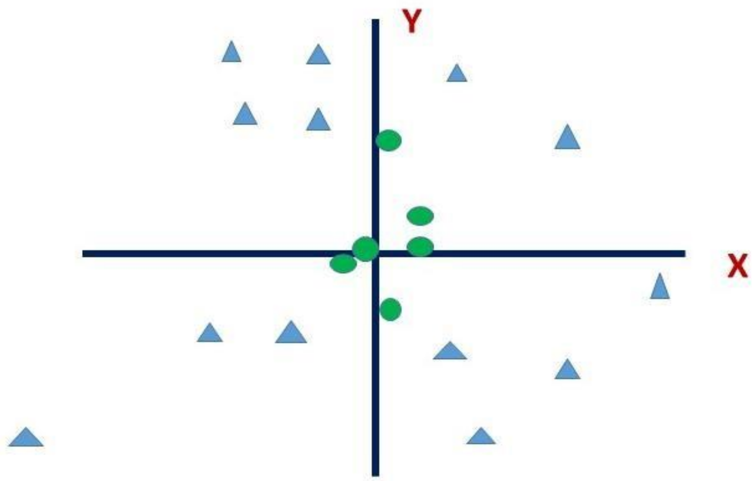


Figure 3.3: Support Vector Machine Structure for Non-Linear Data Set

One of the key strengths of SVM is its ability to handle high-dimensional datasets effectively, which aligns with our objective of analyzing complex and multi-dimensional data in our research. Furthermore, SVM offers various kernel functions, such as linear, polynomial, and radial basis function (RBF), enabling us to adapt the algorithm to different data distributions.

The advantages of SVMs lie in their ability to handle high-dimensional data efficiently, making them suitable for tasks with many features. They are also robust to outliers, as they primarily focus on the support vectors that have the most influence on the decision boundary. Additionally, SVMs often generalize well and have strong theoretical foundations, which aids in model interpretation and performance optimization. These characteristics make SVMs a valuable tool in various applications, from text classification and image recognition to bioinformatics and financial modeling. By leveraging the power of SVM, we aim to contribute to the advancement of weather prediction accuracy and reliability, ultimately enhancing our understanding of atmospheric phenomena and better equipping society to mitigate the impacts of weather-related events.

3.2.3 Multilayer Perceptron Neural Network

For many machine learning applications, such as classification, regression, and pattern recognition, a multilayer perceptron (MLP) neural network is a basic and adaptable artificial neural network architecture. It has an input layer, one or more hidden layers, and an output layer, which are all layers of interconnected neurons.

A network of interconnected nodes or neurons in each layer is necessary for an MLP to function. Each neuron receives weighted input signals from neurons in the layer below, processes those signals with an activation function, and then sends the results to neurons in the layer above. Using methods like backpropagation, the network iteratively modifies the weights during training in an effort to reduce the discrepancy between expected and actual outputs.

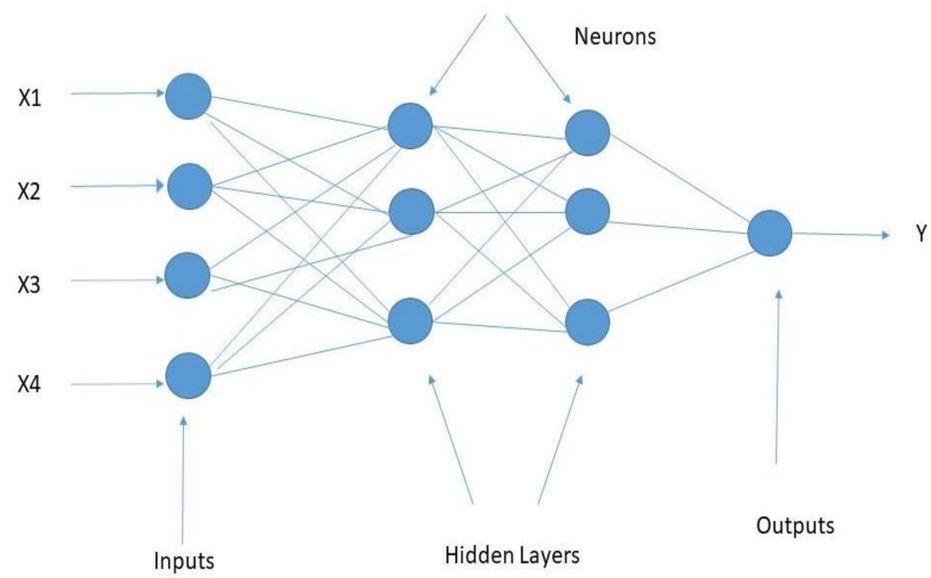


Figure 3.4: Multilayer Perceptron neural network with 2 hidden layer

The benefit of MLPs is that they can simulate intricate non-linear relationships in data, which makes them useful for a variety of jobs. Given enough data and the right architecture, they can approximate practically any continuous function and learn and recognize complex patterns. MLPs are also very adjustable because there are a number of activation functions and architectural options available to customize the network to certain issues. They are an excellent option for a wide range of applications, from image recognition and natural language processing to financial forecasting and medical diagnostics, thanks to their adaptability and capability to learn from data.

The Multilayer Perceptron neural network is essential for forecasting the weather because it makes it possible to model intricate, non-linear correlations in meteorological data. Its importance stems from its ability to increase forecasting precision, accommodate large-scale datasets, and manage temporal connections.

3.2.4 K-Nearest Neighbors

K-Nearest Neighbors is a versatile machine-learning algorithm used for both classification and regression tasks. At its core, KNN relies on the principle that similar data points tend to share similar characteristics. When a new data point needs classification or prediction, KNN calculates the distance between this point and all other data points in the training dataset using distance metrics like Euclidean or Manhattan distance.

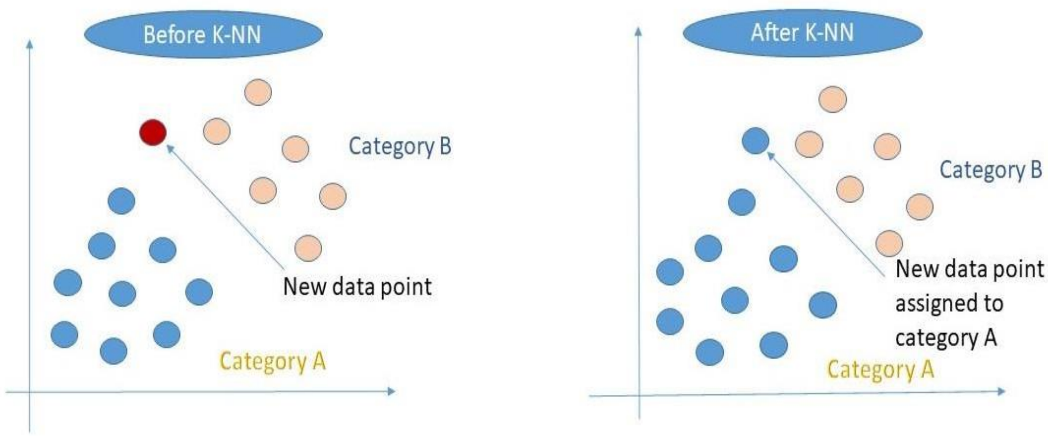


Figure 3.5: K-Nearest Neighbor Algorithm Structure

The user-defined hyperparameter 'K' plays a pivotal role in KNN. It specifies the number of nearest neighbors to consider when making predictions. A smaller 'K' results in predictions that are sensitive to local patterns but can be noisy, while a larger 'K' smooths out predictions but may overlook subtle distinctions. KNN then identifies the 'K' nearest data points with the smallest calculated distances to the new data point. These neighbors are extracted from the training dataset. For classification tasks, KNN assigns the most frequently occurring class label among these neighbors to the new data point. For regression tasks, it computes the average (or weighted average) of the target values of these neighbors to predict a continuous value.

KNN's simplicity and effectiveness make it a valuable tool in various domains, although its performance can be influenced by the choice of 'K' and the distance metric. Proper preprocessing, feature scaling, and tuning 'K' are often necessary for optimal results.

3.2.5 Random Forest

A machine learning algorithm known for its effectiveness and reliability is the random forest ensemble. Both classification and regression tasks make extensive use of it. Its name comes from the term of the construction process, which entails creating a forest of decision trees. It gets its power from integrating different decision trees to generate predictions, which makes it extremely accurate and reliable. Because of the randomness of the data sampling used to generate bootstrap data and feature selection, this approach is known as Random forest. This increases the robustness of the models by ensuring that no single feature dominates the decision-making process. Each tree is exposed to a slightly different view of the data as a result of the randomness, which helps lower the danger of overfitting. In order to generate numerous decision trees, it starts by selecting random subsets of the training data. This method is known as bagging or bootstrap aggregating. The Original data was used to build this bootstrap data. It is conceivable for the same data to exist in bootstrap data numerous times or not at all. Multiple decision trees are produced as a result of this process, and each one learns from a distinct subset of the input.

Only a random subset of features is taken into account for splitting at each node of the decision tree. Adding more randomness avoids overfitting and enhances the generalization of the models. In the case of classification, each tree casts a vote for a certain class, but in the case of regression, they offer numerical predictions. The final prediction for classification is decided by majority vote, and for regression, the results of all the forest's trees are averaged. There are many benefits of using random forests. It frequently provides good accuracy, resists overfitting, and can handle huge datasets with a variety of characteristics.

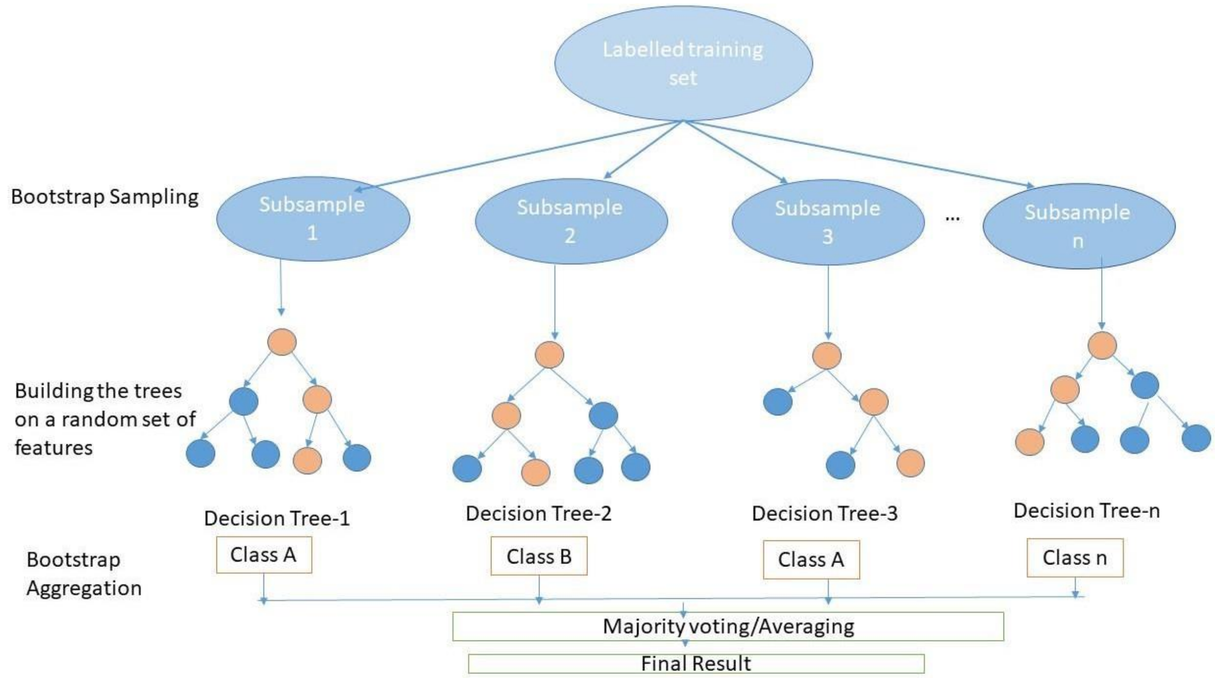


Figure 3.6: Random Forest Algorithm Structure

3.2.6 Extreme Gradient Boosting

XG Boost, or extreme Gradient Boosting, is a highly acclaimed machine learning algorithm known for its speed, scalability, and accuracy. It belongs to the ensemble learning family, utilizing gradient boosting techniques to sequentially train weak learners, typically decision trees. What sets XG Boost apart is its implementation of various enhancements, including regularization techniques like L1 and L2 regularization to prevent overfitting and an optimized tree construction algorithm that efficiently prunes unnecessary splits. This efficiency is further bolstered by the "weighted quantile sketch" method for approximate tree learning. XG Boost offers a plethora of hyperparameters for fine-tuning model behavior and performance, covering aspects like tree structure, learning rate, and regularization. Its scalability is noteworthy, supporting distributed computing frameworks for training large-scale models on massive datasets across computing clusters. XG Boost's versatility makes it a popular choice across diverse machine learning tasks, from classification and regression to ranking and recommendation systems.

3.3 WORK FLOW DIAGRAM

In our working process, the initial step is selecting the Attributes; by selecting the appropriate Attributes, we can accurately predict the weather. After selecting the features/ attributes, the next step is to collect the raw data, which we will collect from the Bangladesh Meteorological Department. Raw data provide better prediction and less error. After collecting the data, it goes under pre-processing and normalization to ensure the reliability and suitability of the collected data for the training algorithm.

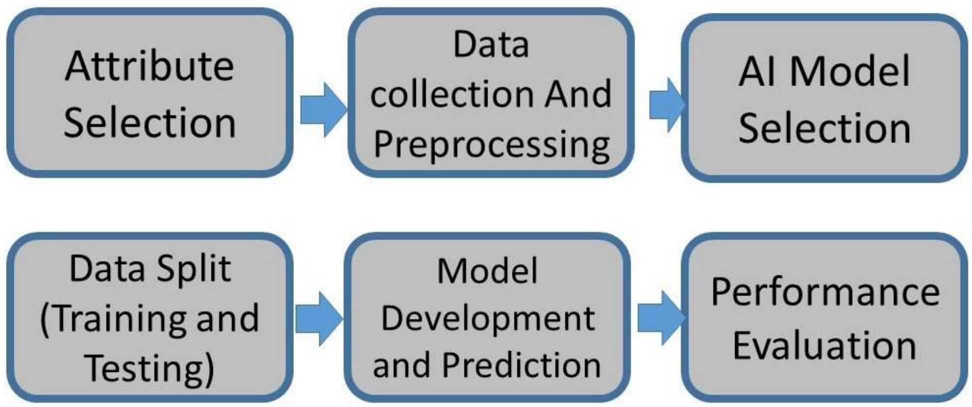


Figure 3.7: Workflow Diagram of the Weather Prediction Model

The normalized data is then divided into training and testing data. After that, training data is used to train the model, and testing data is used for the testing process. After that a Comparative analysis is conducted, considering various training algorithm such as Decision trees, Random forest, KNN, Support Vector Machines, Multilayer Perceptron Neural Network. By comparing the outcomes among these algorithms, the most effective one is identified. This chosen algorithm becomes the final model used for weather prediction. The Weather Forecast workflow diagram visually represents the entire process, showcasing the progression from attribute selection to model selection. The comprehensive methodology ensures that the collected data is processed and analyzed meticulously to achieve accurate weather predictions

3.4 IMPLEMENTATION

After the workflow planning the next step is to Implement the plan, in this Section we will describe the steps which we will implement. We will Selected our feature for the weather prediction, after that based on the selected data we will collect our data from various source which include BMD, Bangladesh Disaster Management Department and Bangladesh Water Development Board (BWDB). Then Pre-processed our data and Build a model. After that we will use various performance metrics to Evaluate the performance and select the best method

3.4.1 Feature Selection

Weather depends on various features and attributes such as temperature, humidity, wind speed, wind direction, atmospheric pressure, and precipitation (rainfall, snowfall, etc.). Solar radiation, Cloud cover, Latitude and longitude, Elevation or terrain height Land use and land cover data, Geographic coordinates of weather stations, Date and time, Seasonal information, Day, Air Quality etc.

The choice of features and attributes is crucial since it has a direct impact on the precision and effectiveness of forecasting models. By carefully selecting the most pertinent weather variables and removing redundant or irrelevant data points, we can increase forecast accuracy, decrease computational complexity, and boost model interpretability. This technique also enables meteorologists to concentrate on the primary causes of weather patterns by using valuable computational resources as efficiently as possible.

Additionally, feature selection makes forecasting models resilient and flexible in real-world weather scenarios by ensuring that they generalize well to unobserved data. Fundamentally, it simplifies the forecasting process, synchronizes models with domain knowledge, and ultimately results in more accurate and actionable weather forecasts that have broad ramifications for disaster preparedness, agriculture, transportation, and climate research.

Here Maximum and minimum temperature, Rainfall, Wind speed, Humidity, Pressure, Wind direction, dew point Data has been selected as a features for weather prediction.

3.4.2 Data Collection

Data from the Bangladesh Meteorological Department (BMD) has been gathered for our study. Data on Maximum and minimum Temperature, Rainfall, Humidity, Wind Speed, Dew Point, Wind Speed, Wind Direction, and Pressure. It included the data of 34 weather stations in Bangladesh. After gathering the data, we analyzed the data from various perspectives, like yearly average, monthly average, etc. After that, we used a variety of machine learning techniques to examine and understand the relationships and patterns in the Dataset. Our goal is to identify the algorithms that produce the most useful results in in terms of the ability to forecast and comprehend these weather-related occurrences. Due to this by improving our understanding of this mechanism, we want to contribute to the improvements in weather prediction and mitigation measures.

Table 3.1 Attributes Type and Description

Attribute	Type	Description
Maximum Temperature	Numerical	C(Degree Celsius)
Minimum Temperature	Numerical	C(Degree Celsius)
Average Temperature	Numerical	C(Degree Celsius)
Humidity	Numerical	%(Percentage)
Rainfall	Numerical	Millimeter
Pressure	Numerical	In Kilo Pascal
Wind speed	Numerical	In kilometers Per Hour
Wind Direction	Numerical	In Degree
Dew Point	Numerical	C(Degree Celsius)

Table 3.1 Shows the attributes that we have collected and their Descriptions. We Collected Each parameter from Bangladesh Meteorological Department. Here, as an attribute, we have Maximum, minimum, and mean temperature, Humidity, wind speed, wind direction, dew point, pressure, and Rainfall.

Table 3.2 Sample Datasets of Daily Temperature, Humidity, Pressure, Wind Speed, Wind Direction and Wind Direction

Station	Year	Month	Day	Max Temperat	Min Temperat	Average Temperat	Humidity	Pressure (kpa)	Wind Speed	Wind Direction	Dew Point
Barisal	1980	1	1	26.1	7.5	16.8	73	101.25	8.57	186.31	9.11
Barisal	1980	1	2	27	13.2	20.1	83	101.09	10.15	141.25	9.15
Barisal	1980	1	3	21.8	15.7	18.75	82	101.75	14.43	188.02	8.55
Barisal	1980	1	4	25.1	14.5	19.8	80	100.48	13.47	187.75	10.01
Barisal	1980	1	5	23.9	11.1	17.8	77	100.03	17.06	171.53	12.83
Barisal	1980	1	6	23.8	11	17.5	72	100.10	17.07	187.17	25.70
Barisal	1980	1	7	23.4	10.5	17.4	71	100.04	17.13	177.17	13.87
Barisal	1980	1	8	24	8.5	16.95	76	100.61	12.34	188.00	12.59
Barisal	1980	1	9	24	10.4	16.25	78	100.88	7.89	246.44	11.98
Barisal	1980	1	10	23.4	9.5	17.2	84	101.10	8.35	234.88	12.27
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
Teknaf	2022	12	27	29.4	25	22.2	85	100.6	4.788	174.88	17.12
Teknaf	2022	12	28	29.4	25.6	22.2	90	100.62	4.284	187.94	18.56
Teknaf	2022	12	29	28.8	28.6	24.45	90	100.7	4.86	196.94	18.37
Teknaf	2022	12	30	30.3	15.3	22.3	85	100.8	5.112	173	14.65
Teknaf	2022	12	31	29.3	13.4	20.2	75	100.77	5.148	181.31	13.3

Table 3.3 Sample Dataset of Monthly Rainfall

Station	Year	Month	Rainfall(mm)
Barisal	1980	1	0
Barisal	1980	2	55
Barisal	1980	3	47
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
Teknaf	2022	11	27
Teknaf	2022	12	0

On the other hand, Table 3.2 Shows the Sample of Data that we have collected. We have collected daily data on temperature, including maximum, minimum, and average temperature, humidity, wind speed, wind direction, pressure, and dew point from 1980 to 2022. We also Collected the Monthly Average Rainfall of Various Stations from 1980 to 2022, which we showed in Table 3.3. In our research, we collected data from various geographical stations, which we listed in Table 3.4. There are a total of 35 distinct geographical areas.

Table 3.4 Geographical Co-Ordinate of the Station Selected for Model Training and Testing

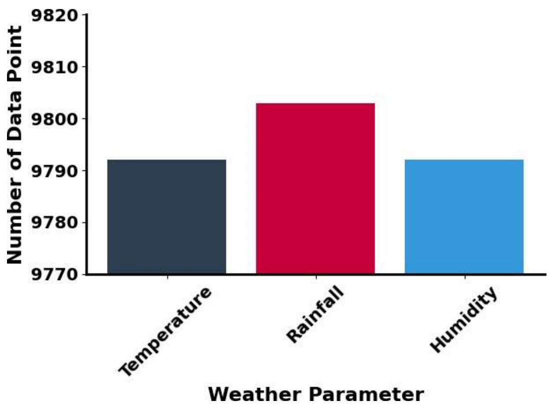
Station	Latitude	Longitude	Station	Latitude	Longitude
Barisal	22.7	90.36	Kutubdia	21.83	91.84
Bhola	22.7	90.66	Madaripur	23.17	90.18

Bogra	24.88	89.36	Maijdee Court	22.83	91.08
Chandpur	23.26	90.67	Mongla	22.43	89.66
Ambagan	22.35	91.81	Mymensingh	24.75	90.41
Chittagong	22.34	91.79	Patuakhali	22.36	90.34
Chuadanga	23.65	88.81	Rajshahi	24.35	88.56
Comilla	23.48	91.19	Rangamati	22.67	92.20
Cox's Bazar	21.46	91.98	Rangpur	25.72	89.26
Dhaka	23.78	90.39	Sandwip	22.5	91.46
Dinajpur	25.63	88.66	Satkhira	22.68	89.07
Faridpur	23.61	89.84	Sitakunda	22.64	91.64
Feni	23.01	91.37	Srimangal	24.29	91.73
Hatiya	22.29	91.13	Syedpur	25.75	88.91
Ishurdi	24.12	89.04	Sylhet	24.88	91.93
Jessore	23.17	89.22	Tangail	24.15	89.55
Khepupara	21.98	90.22	Teknaf	20.87	92.26
Khulna	22.8	89.58			

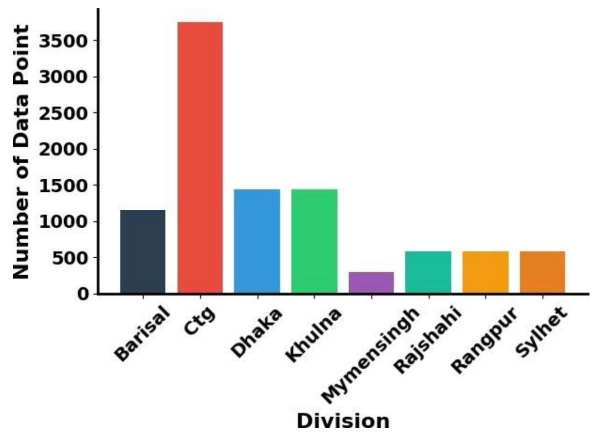
We have divided our datasets into some category, Rainfall in 2 intervals, one is 1999-2022 and another is 1980-2022. And Temperature and Humidity in 3 Interval respectively 1948-2022, 1980-2022, 1999-2022. In figure 3.8 we have added the graph of the data in interval 1999-2022. And further we have divided our datasets into 8 divisions and visualize the number of data instance on that division.

(A) On-time interval 1999-2022

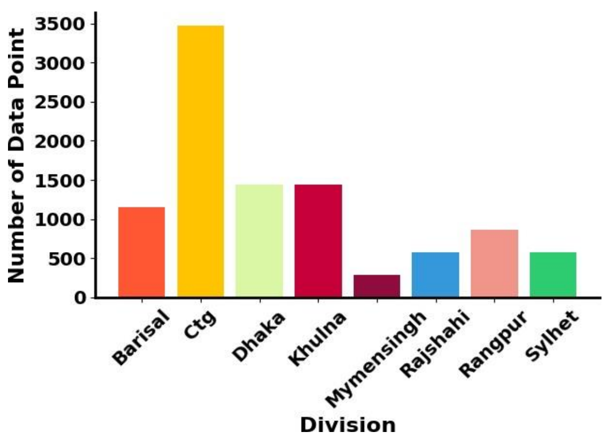
As we divided our datasets into two intervals, there were different numbers of data in each interval. Figure 3.8 shows the total amount of data in the time interval 1999-2022. In this interval, we have 9792 Temperature data, 9803 Rainfall Data, and 9709 Humidity data. We have more Rainfall Data than Temperature and Humidity. Temperature and Humidity in 8 Division named Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet, respectively 1152, 3745, 1440, 1440,289,576, 576, and 576. Chittagong Division has the highest data, and Dhaka has the second highest data. On the other hand, Rainfall in 8 Divisions, Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet, respectively 1152, 3468,1440,1440,289,576,865 and 576.



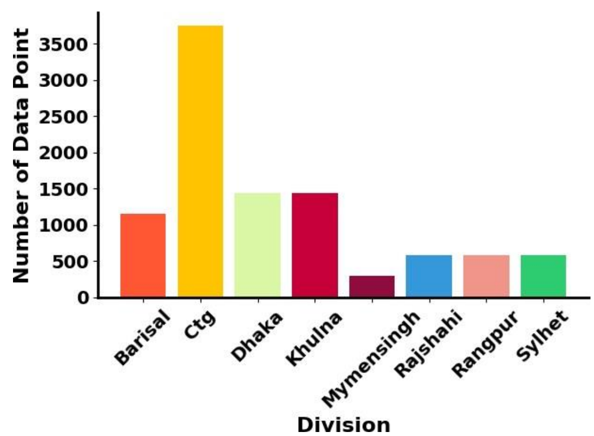
(a) Total Data of Temperature, Rainfall and Humidity



(b) Temperature Data in each Division



(c) Rainfall Data in each Division

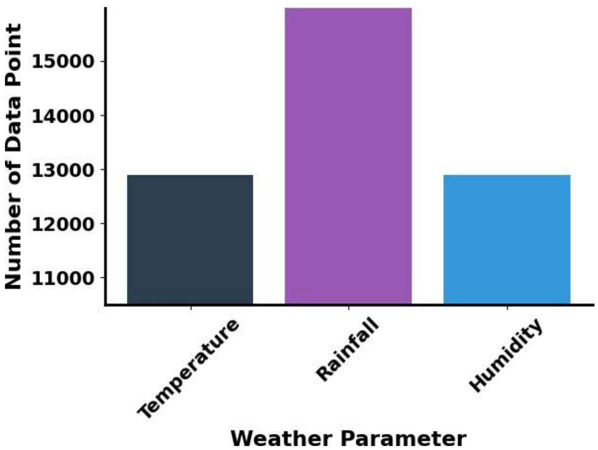


(d) Humidity Data in Each Division

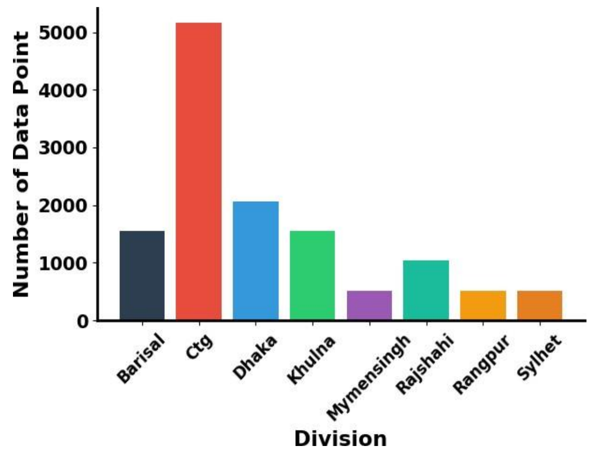
Figure 3.8: Total and Division Wise Data in the time interval 1999-2022

(B) On-time interval 1980-2022

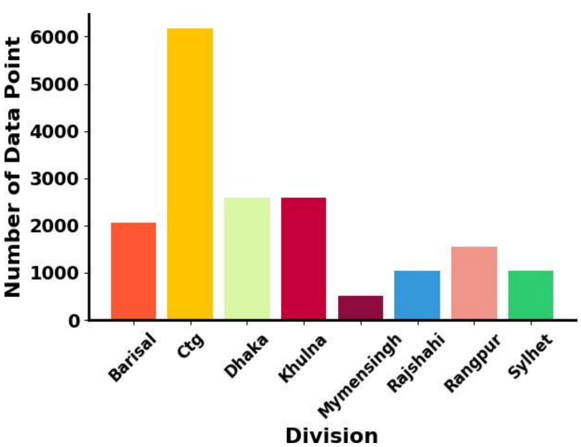
In the time interval 1980-2022, we have 12901 Temperature data, 15973 Rainfall Data, and 12901 Humidity data. And Temperature and Humidity in 8 Division Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet respectively 1548,5160, 2064, 1568,516,1032, 516 and 516. On the other hand, Rainfall in 8 Divisions, Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rangpur, and Sylhet, respectively 2065, 6269,2581,2579,289,516,1032,1549 and 1032.



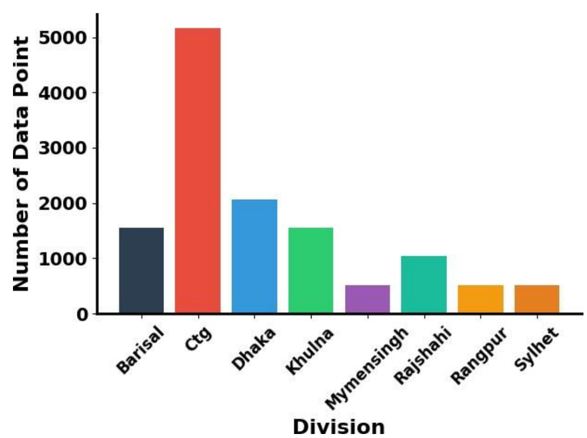
(a) Total Data of Temperature, Rainfall and Humidity



(b) Temperature Data in each Division



(c) Rainfall Data in each Division



(d) Humidity Data in each Division

Figure 3.9: Total and Division Wise Data in the time interval 1980-2022

Beside Temperature and Humidity Data we have some other weather parameter like wind speed in km/hour, Wind Direction in Degree, Pressure in KPa and Dew Point, we have collected this data from 1981 to 2022. This data is in monthly form; we have total 15091 Data of ball of this parameter.

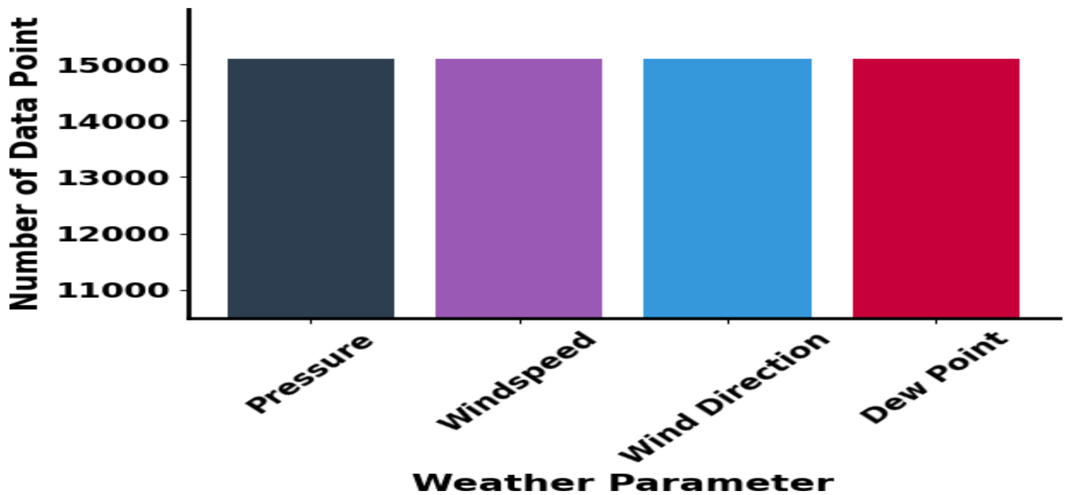


Figure 3.10: Number of Pressure, Wind Speed, Wind Direction and Dew Point Data

3.4.3 Data Pre-processing and Model formulation

The raw Dataset from BMD required a substantial amount of cleaning to convert it from a semi-structured dataset to a structured dataset in order to be ready for the building of our desired model. We used data from **1980 to 2022** to provide the most accurate findings for Bangladesh using the most recent meteorological data. The Dataset had 6 particular criteria: wind speed, rainfall, flood, drought, temperature, and cyclones.

(a) Data Pre-processing

Before running the AI model, we pre-processed our data in some steps. At first, we organized our data in a manner that made it easy for our model to train and test. Then, after loading the data sets, we used min-max normalization techniques so that the values remain in any specific range. After that, we divided our datasets into two sections, one is testing data, another one is training data. After dividing, we apply the different AI models on that data.

i) Min-Max Normalization

Min-max normalization, also known as feature scaling or min-max scaling, is a data preprocessing technique used in machine learning and statistics. Its purpose is to rescale numerical features within a specific range, typically between 0 and 1. This normalization method preserves the relative relationships between data points while ensuring that all features have the same scale.

For each feature, we must determine the Dataset's minimum (min) and maximum (max) values. Then, the range of each feature will be compared by subtracting the minimum value from the maximum value. This range defines the span of the data for that feature. For each data point in the feature, we applied the following formula to scale it to a value between 0 and 1:

$$X' = \frac{X - Xmin}{Xmax - Xmin} \tag{3.2}$$

This formula linearly scales the data, mapping the minimum value to 0 and the maximum value to 1. Intermediate values are scaled proportionally within this range. Where

- X' is the Normalized Value
- X is the original Value
- X_{min} is the minimum value of Rainfall/ Temperature
- X_{max} is the maximum value of Rainfall/ Temperature

Min-max normalization is particularly useful when The features have varying scales, and we want to bring them to a similar scale to prevent certain features from dominating others in the learning process (especially in algorithms sensitive to feature scales like gradient descent-based methods). The algorithm we're using assumes the input data is on a similar scale, like distance-based algorithms like K-nearest neighbor.

ii) Training and Testing Data

After preprocessing and cleaning to make it a structural dataset, the raw Dataset will be split into two parts, one for training data and the other for testing data. In order to build the model, the training dataset will be used to test these regression-based learning techniques. The effectiveness of certain algorithms will next be assessed for the testing dataset in order to compare their level of learning and prediction to that of the training dataset. As we have two intervals of data, in the prediction of using 1980-2022 datasets, we used 1980-2017 for training and 2018-2022 for testing So that we can evaluate our model's performance. In prediction using 1999-2022 datasets, we used 1999-2017 for training and 2018-2022 for testing. We made a comparison graph for the actual and predicted rainfall and temperature using this training dataset.

(b) Model Formulation

In formulating predictive models for rainfall prediction, various machine learning algorithms are considered, each offering distinct methodologies and advantages. Random Forest, employing an ensemble of decision trees, excels in capturing complex interactions between features and mitigating overfitting. Support Vector Machine constructs hyperplanes in high-dimensional spaces, optimizing margins for robust classification or regression. Decision Trees, with their hierarchical structure, provide interpretability and handle non-linear relationships effectively, albeit prone to overfitting. K-Nearest Neighbors relies on local similarities for prediction, offering simplicity and adaptability, albeit sensitive to parameter choices. Multilayer Neural Networks, utilizing layered architectures, learn intricate patterns but demand substantial data and computational resources. XG Boost, a gradient boosting technique, sequentially builds weak learners, leveraging gradient descent and regularization for superior performance in tabular data tasks. Each approach, tailored to the unique characteristics of the rainfall prediction task, offers insights into temporal dynamics, non-linearity, and model interpretability, crucial for accurate forecasting. After model formulation and predicting rainfall/ temperature, we used various performance metrics like MSE, MAE, RMSE, RAE, R^2-accuracy, etc. Which suggested the better or best model for predicting rainfall/ temperature using this type of Dataset.

3.4.4 Performance Evaluation

The main goal of this project is to create a model to evaluate the effectiveness and viability of using various artificial intelligence approaches for weather prediction. Metrics like Mean Absolute Error and Mean Square Error will be used to assess the model's performance. These stats give numerical measurements to assess the efficacy and correctness of the models in predicting weather patterns. The study seeks to use these evolution metrics to determine to what extent artificial intelligence can improve the precision and usefulness of weather forecasting. These metrics provide quantitative measurements to assess the model's precision and accuracy in predicting weather conditions. It must be demonstrated that our model's error % is little in comparison to how well it predicts the future in order to demonstrate its validity.

To assess the error rate in percentage here the following metrics will be used:

a. Mean Absolute Error
b. Mean Square Error
c. Root Mean Squared Error
d. Relative Absolute Error
e. R^{2} -Value
f. R^{2}-Accuracy

a. Mean Absolute Error

In statistics and machine learning, Mean Absolute Error is a commonly used metric to evaluate the accuracy of a predictive model. It provides insight into the typical magnitude of differences between observed and expected values. Unlike some other error measures that square the discrepancies, MAE calculates the absolute differences between each projected and actual value for each data point, totals them up, and divides them by the total number of data points. This approach provides a clear and sensible estimate. MAE is particularly valuable when outliers should not be heavily penalized, and the magnitude of errors is critical. It is widely employed to assess the performance of models in regression problems involving the prediction of continuous numerical quantities.

In essence, MAE represents the average absolute deviation of predictions from their actual values. Lower MAE values, indicating a closer alignment between predictions and actual data, signify a more accurate model. Since MAE uses the same units as the target variable, it is easy to comprehend and compare across various datasets and models.

The average value of all absolute errors is known as the Mean Absolute Error. The formulais:

$$MAE = \frac{1}{\mathrm{n}} \sum_{i-1}^{n} |xi - x| \tag{3.3}$$

Where:

- n = the number of errors,
- Σ = summation symbol,
- |xi – x| = the absolute errors.

b. Mean Squared Error

The metrics known as Mean Squared Error are extensively used for performance evolution in machine learning algorithms. The difference between excepted and actual data is calculated asthe average squared variance. When comparing excepted and actual values, the mean of all squared differences is employed to determine MSE. This statistic is particularly useful when greater errors need to be penalized and more severely useful because it makes greater mistakes more noticeable when the differences are squared. MSE is based on averaging the squared deviation guesses from the actual data interpretation. A model with a lower MSE is assumed to be more accurate because, on average, the predictions are according to the real data.

When analyzing regression problems, MSE is typically used to gauge how effectively models predicts continuous numerical values, MSE's ability to consider error magnitude and provide a measurement of error dispersion one of its advantage. Due to the squaring process, bigger errors are penalized mode severely. This attributes makes MSE effective in application when it is important to emphasize the impact of significant errors.

Formula:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2} \tag{3.4}$$

MSE= Mean Squared Error

y_i= Observed Values

$\hat{y}$= predicted values

N= Number of data points

c. Root mean Squared Error

Root Mean Squared Error is a statistical metric used to assess the accuracy and reliability of predictive models. It measures the average magnitude of errors or deviations between the predicted values generated by our forecasting models and the actual observed values. RMSE provides a valuable measure of how well our predictive models align with historical sales data.

In our proposal, we will use RMSE to evaluate the performance of various predictive models that we plan to develop. A lower RMSE indicates that our models are generating predictions that are closer to the actual sales figures, suggesting a higher level of accuracy and reliability. Conversely, a higher RMSE suggests that our models may have larger prediction errors. RMSE is computed by taking the square root of the mean of the squared differences between predicted and actual values. This metric is particularly valuable when it is necessary to assign greater weight to larger errors, as the squaring operation magnifies the impact of these deviations.

By regularly monitoring and analyzing the RMSE values, we can fine-tune our predictive models, make necessary adjustments to our forecasting strategies, and improve the precision of our sales forecasts over time. This, in turn, will enable us to make more informed decisions regarding inventory management, resource allocation, and overall business planning.

In summary, RMSE is a fundamental tool we will utilize in our proposal for implementing predictive analytics in sales forecasting, ensuring that our models are effective in providing accurate sales predictions and contributing to our company's operational efficiency and profitability.

The formula for calculating RMSE is as follows:

Formula:

$$RSME = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{N-P}} \tag{3.5}$$

y_i is the actual value for the i^{th} observation.

$\hat{y}$ is the predicted value for the i^{th} observation.

N is the number of observations.

P is the number of parameter estimates, including the constant.

In The Methodology Chapter we described our working procedure and few Algorithm which include Regression and optimization algorithm, which we will used in this thesis work to build model, we also described the performance metrics which will help us to evaluate the algorithm and make comparison between them. In the next Chapter we will describe about the expected outcome.

d. Relative Absolute Error

Relative Absolute Error is a performance metric commonly used in regression analysis to evaluate the accuracy of predictive models. It measures the average absolute difference between predicted values and actual values relative to the average absolute difference between the actual values and their mean.

RAE provides a relative measure of prediction accuracy, allowing for comparison across different datasets and models. A lower RAE indicates better predictive performance, as it signifies that the model's predictions are closer to the actual values relative to the variability in the data. This metric is particularly useful when comparing models across datasets with varying scales or when the absolute accuracy of predictions is more important than their direction (i.e., overestimation or underestimation). However, it's important to consider the context of the specific problem and other evaluation metrics alongside RAE to gain a comprehensive understanding of a model's performance.

e. R^2 –Value

R^2 –Value provides information about how effectively the link between the input characteristics and the target variables is captured by the model. Although more complicated models may produce falsely inflated numbers, caution is advised as higher R-squared values indicate stronger predictive potential. Assessing R-squared in conjunction with additional metrics offers a thorough comprehension of the model's performance and capacity for generalization. It may be applied to a wide range of AI applications, such as forecasting, classification, and regression. In order to improve predicted accuracy and resilience, R-squared values are frequently optimized as part of the ongoing validation and improvement of AI models. R-squared should not be interpreted in isolation, as this might lead to misunderstandings. Instead, a comprehensive evaluation of model performance is necessary during the growth of AI models. Equation of R-Square Value-

$$R^2 = 1 - \frac{\mathrm{MSE}}{\mathrm{Var}(y)} \tag{3.6}$$

Where,

- **MSE** is Mean Square Error
- $\mathrm{Var}(y)$ is the variance of the true target variable y.

f. $\mathbf{R^{2}}$ – Accuracy

The R-squared accuracy is computed to evaluate the accuracy or goodness of fit of the Regression model on the training data. It quantifies the proportion of the variance in the observed sample data, which is explained by the independent variables included in the model. A higher R-squared value closer to 1 indicates a better fit of the model to the training data, suggesting that a larger portion of the variability in rainfall/temperature is captured by the model's predictions. This R-squared accuracy metric serves as a key indicator of the model's performance in understanding and predicting rainfall/temperature patterns based on the features provided. It allows for a quantitative assessment of how well the model aligns with the actual rainfall observations, facilitating informed decisions in model evaluation and refinement.

3.4.5 System Requirement

For our research work, we used different types of tools. Some are listed below.

1. Hardware requirement: A desktop or laptop with a Core i5, 8 GB RAM, and a 4GB graphics card has been used.

2. Software requirement:

i) Python Jupiter**:** In our study, we used Python Jupiter to implement our prediction code and also for making a quality full graph

ii) Matlab: Matlab is a high-level programming language that is used broadly in research all over the world. In our research work, We used the Matlab R2018a version for making graphs, which we used to analyze our historical Data

iii) Microsoft Office 2010**:** We used this for data processing and storing.

3.5 SUMMARY

In this chapter, we have discussed our working procedure and workflow. After that, we discussed our data collection and the variation of data in the different intervals and divisions. We have visualized our diversion of data in different areas and also show the geographical coordinates of our study area. Then, we talked about Various AI model and their use in our work. Then, we discussed various performance metrics that we have used to evaluate our model. In the next chapter, we will discuss our results and analyze them.

CHAPTER 4

RESULTS AND ANALYSIS

4.1 INTRODUCTION

In environmental and climatological studies, the analysis of meteorological variables such as Rainfall, Temperature, and humidity plays a vital role in understanding the dynamics of Earth's atmosphere and its impact on various natural and human systems. This chapter delves into a comprehensive examination of these key variables, spanning different temporal scales and encompassing both descriptive analysis and predictive modeling. Rainfall, Temperature, and Humidity are fundamental components of Earth's climate system, influencing a wide array of ecological processes, agricultural practices, and socio-economic activities. Understanding the spatial and temporal variations of these variables is essential for assessing climate change impacts, predicting weather patterns, and designing effective adaptation and mitigation strategies. The primary objective of this chapter is to provide a detailed exploration of Rainfall, Temperature, and humidity data, with a focus on analyzing their yearly, monthly, and seasonal patterns. Additionally, this chapter aims to evaluate the performance of predictive models in forecasting these variables across different temporal intervals and spatial divisions.

This chapter is structured into two main sections. The first section provides a thorough examination of Rainfall, Temperature, and humidity data, elucidating their temporal trends, seasonal variations, and spatial distributions. Where we work with the Historical data. The second section delves into the evaluation of predictive models, assessing their accuracy and reliability in forecasting these variables at different temporal resolutions and spatial scales.

4.2 ANALYSIS OF HISTORICAL DATA

In this section, we conduct a comprehensive analysis of Rainfall, Temperature, and Humidity dataspanning various intervals. The dataset encompasses a significant temporal range, allowing us to discern long-term trends and seasonal patterns in these meteorological variables. We divided the Rainfall, Temperature and Humidity Datasets into two categories, so that we can analyze it more accurately and clearly. By Utilizing Python libraries such as Matplotlib and Pandas, we visualizethe data through line graphs and bar charts to facilitate a clear understanding of the observed trendsand variations.

4.2.1 Temperature Data Analysis

Our analysis delves into the temperature dynamics of Bangladesh across two distinct intervals: 1999-2022, 1980-2022. Because we have collected data from 1980-2022, but in 1980 there were few weather station, many of the stations were not established on that time. But in 1999 many new weather stations were presents, so we analyzed them into 2 section. Within these temporal spans, we meticulously examine the yearly average temperature patterns both at a country-wide level and across eight individual Season. Utilizing line graphs, we portray the temporal evolution of temperatures, adeptly capturing both short-term fluctuations and long-term trends. Moreover, we extend our analysis to explore monthly average temperature data, allowing us to illuminate seasonal variations and fluctuations. By segmenting the temperature data into these intervals, we acquire a comprehensive understanding of the climatic dynamics of Bangladesh, essential for devising robust adaptation strategies and bolstering resilience to the impacts of a changing climate. We have used both the Mean, Maximum and Minimum Temperature in our analysis.

A. Analysis of Temperature Data on Interval 1999-2022

There are 35 individual Weather Station included in this interval. We Divided the data of this interval into 8 Division named Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet, where each division consist of several weather station. We analyze this Division individually so that we can make comparison between them. There is a variation of temperature between the geographical area of this Division. That's why we Divided them into 8 Division.

(i) Analysis of Yearly Average Temperature

We made yearly average graph to understand the yearly trends and changing of the Rainfall in each year. Which give us an overview of the climate change and the changing behavior of the Rainfall in each year. Python was utilized to compute the average Rainfall for each year. This involved aggregating monthly data from all stations and then calculating the yearly average. This methodology provides a comprehensive understanding of annual rainfall trends across the study period.

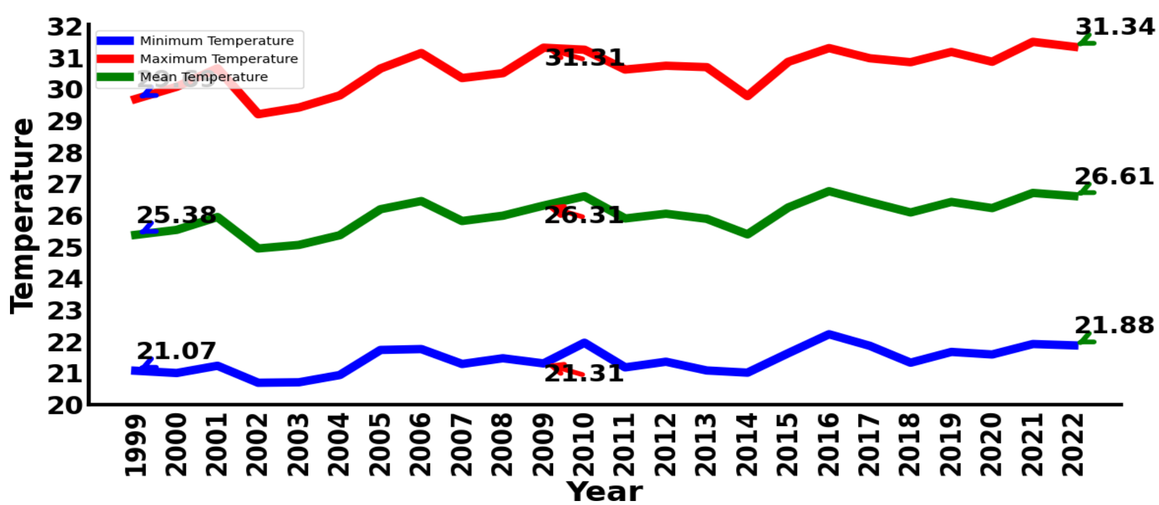


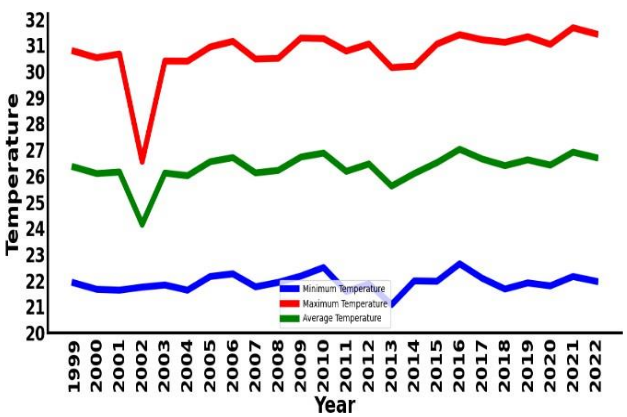
Figure: 4.1 Yearly Average Min, Max and Mean Temperature of Bangladesh

Here we will analyze the yearly average graph for Bangladesh (which contain all the 34 weather station) and 8 Individual Division (which contain the weather station lies in this Division).

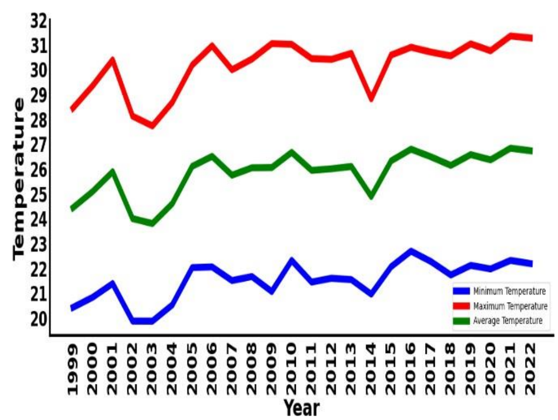
In figure 4.1 we visualize the yearly average Temperature of Bangladesh which included 35 weather station. can see that temperature is increased from 1999-2022. In 1999 temperature maximum was 29.69 Degree Celsius on the other hand in 2022 it was almost 31.34 Degree. Which is a matter of concern and an effect of global warming. Same is for average and minimum temperature.

In a 10 years' gap (1999-2009) Maximum Temperature increased 1.62 °C. From 2009 to 2022 it increased 0.03 °C. If we notice from 1999 to 2022 it increased a lot which is 1.65 °C. Same Situation Happened for the Minimum and Mean Temperature. Mean temperature was 25.38 °C in 1999, after 10 years later it has been 26.31 °C and the in 2022 it has increased again 0.30 °C. Although it has been increase almost 0.93 Degree. And Minimum Temperature also increased 0.81 °C. This analysis highlights the temperature trends over three distinct time periods—1999 to 2009, 2009 to 2022, and the cumulative change from 1999 to 2022. It delineates the varying rates of temperature increase over these periods, showcasing both short-term accelerations and long-term trends. The detailed examination of Maximum, Minimum, and Mean Temperatures provides a comprehensive understanding of the temperature dynamics, elucidating the nuances in temperature fluctuations over time. Such meticulous analysis aids in discerning patterns and drivers of temperature change, contributing to informed decision-making in climate-related matters. Overall, this graph suggests a warming trend over the years, with both minimum and mean temperatures showing an upward trajectory. The maximum temperature also exhibits variability but maintains a high value.

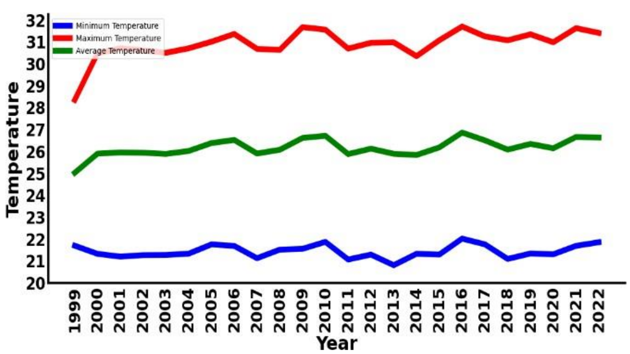
Figure 4.2 Showcasing the yearly average temperature if Eight Individual Division where each division Consist of Several Weather Station. And as their geographical Co-ordinates are different, there temperature and the variation is also different. Figure 4.2(a) Represents the yearly average Temperature of Barisal. The minimum Temperature fluctuated between approximately 20°C and 22°C from 1999 to 2022. The mean temperature gradually increased from around 26°C in 1999 to approximately 28°C by 2022. Meanwhile, the maximum temperature remained relatively stable, starting at about 32°C and showing minor variations. Figure 4.4 Shows the yearly average Temperature of Chittagong. In the early 1990s, the maximum temperature was around 28°C. By the late 1990s and early 2000s, it gradually increased to approximately 30°C. Around 2010, there was a significant spike, with the maximum temperature reaching about 31.5°C. Post-2010, it remained relatively stable, fluctuating between 30°C and 31°C. The average temperature closely mirrored the maximum temperature trend. Starting around 27°C in the 1990s, it rose to about 29°C by the early 2000s. The peak around 2010 corresponded to an average temperature of approximately 30.5°C. Post-2010, it remained consistent, varying between 29°C and 30°C. Beginning around 25°C in the 1990s, the minimum Temperature increased to about 27°C by the early 2000s. The peak around 2010 saw a minimum temperature of approximately 28.5°C. Post-2010, it fluctuated between 27°C and 28°C.



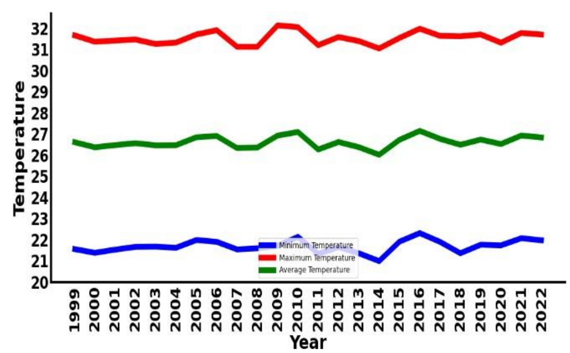
(a) Barisal



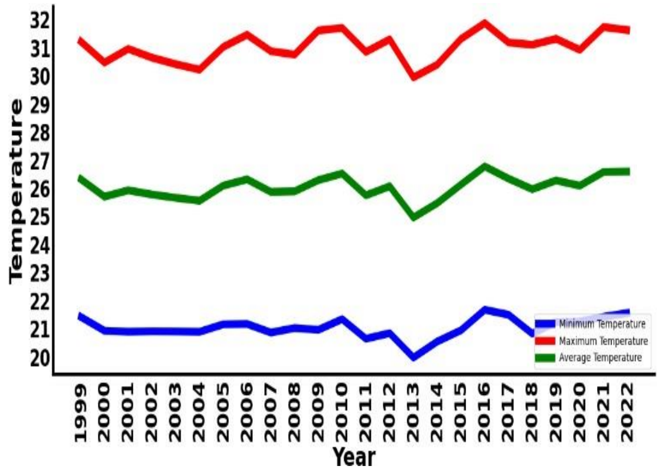
(b) Chittagong



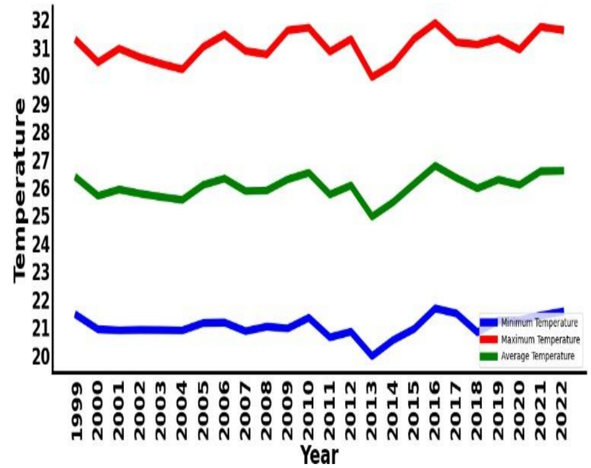
(c) Dhaka



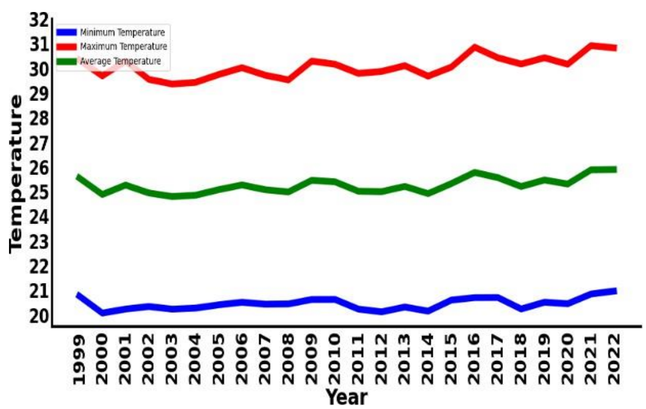
(d) Khulna



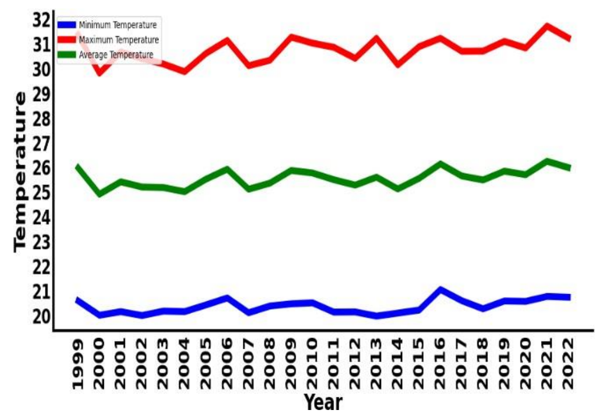
(e) Mymensingh



(f) Rajshahi



(g) Rangpur



(h) Sylhet

Figure 4.2: Yearly Average Rainfall of Eight Divisions

Figure 4.2(c), (d), (e), (f), (g), (h) Respectively Shows the yearly Average Temperature of Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet. We have noticed, almost similar changes in this graph as like Bangladesh, Barisal and Chittagong.

All divisions exhibit fluctuations in temperatures over the years. Most divisions showed an overall upward trend, indicating increasing temperatures. Barisal Division Peaks around 2004, dips around 2011 and then rises again. Chittagong Division has a Sharp increase, peaking near 2009, followed by fluctuations. Dhaka Division has Multiple peaks and troughs, with one of the highest points around 2016. Khulna Division is Relatively stable, with slight variations but no extreme spikes. Rajshahi Division had a Significant increase starting from around 2000, peaking near 2010, followed by variability. Rangpur Division: Less variation but still an overall upward trend, with a peak around 2010. Sylhet Division Considerable variability, several peaks, and troughs, but also follows an increasing trend. Mymensingh Division has experienced a rapid increase, peaking around 2016. The highest recorded average maximum temperature is approximately **32°C** (by Khulna division around 2010). The lowest recorded temperatures are near or slightly above **27°C** (by Barisal divisions in 2002). From this graph, it is seen that Khulna is a division of the highest temperature; after that, Rajshahi and Dhaka are in the next position. On the other hand, Mymnesingh and Rangpur have lower temperatures. The average line helps identify whether the overall temperature has increased, decreased, or remained relatively stable.

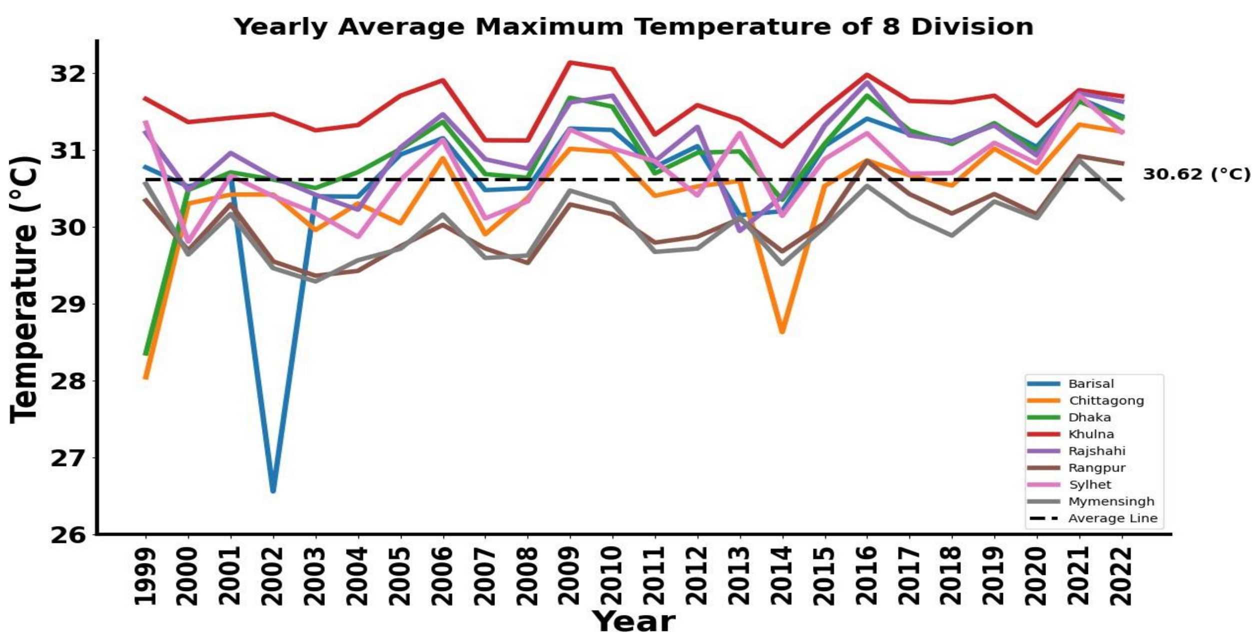


Figure 4.3: Yearly average Maximum Temperature of Eight Division together

Figure 4.4 displays the yearly average minimum temperature for eight divisions in Bangladesh named Barisal, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur, Sylhet, and Mymensingh from 1999 to 2021. While temperatures fluctuate for each division over time, there is no clear long-term trend upwards or downwards for any single division.

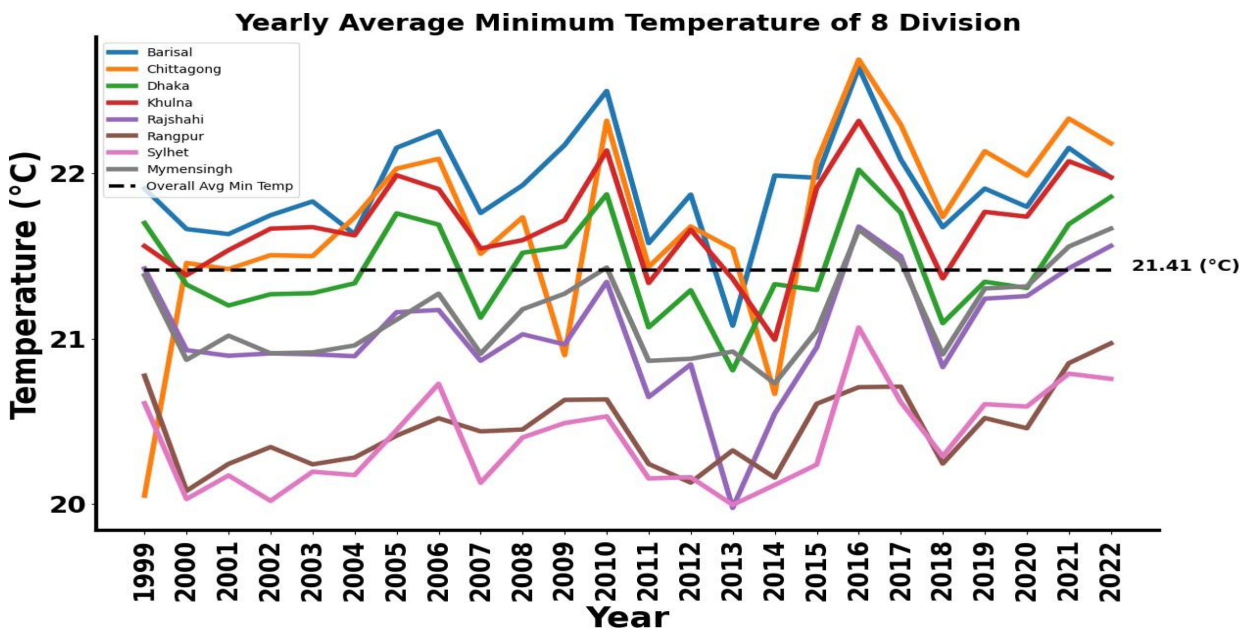


Figure 4.4: Yearly average Minimum Temperature of Eight Division together

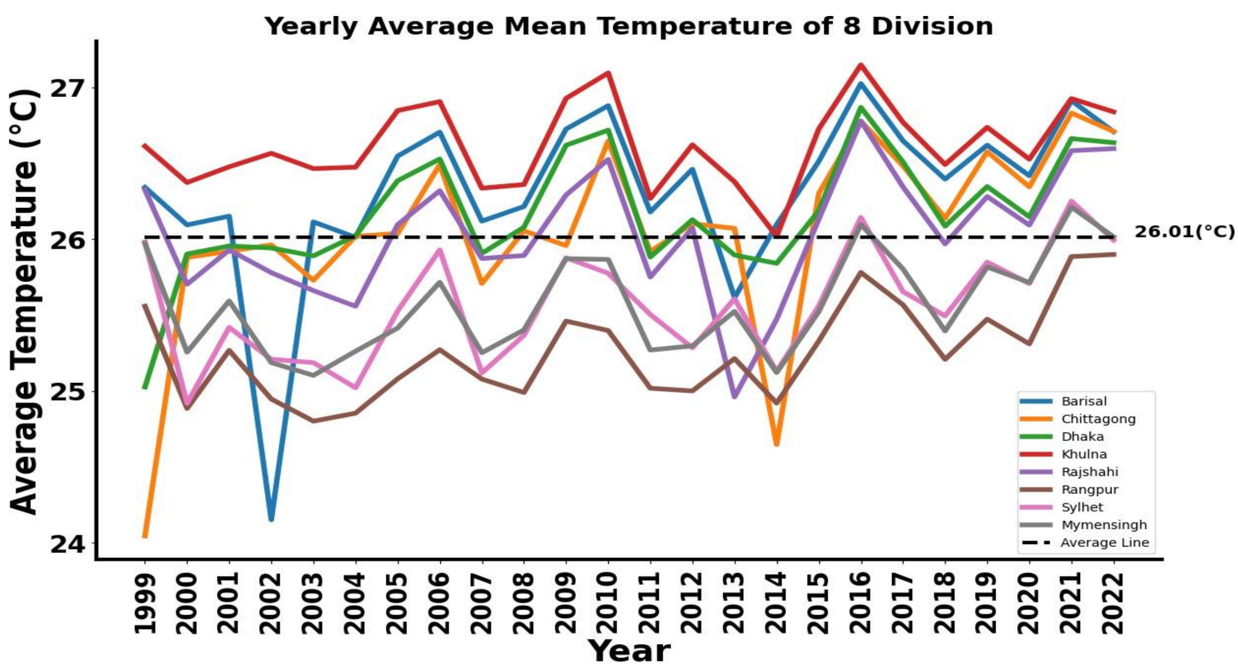


Figure 4.5: Yearly average Mean Temperature of Eight Division together

The overall average minimum temperature across all divisions appears to be approximately 21.41°C. From the graph, we also see that Sylhet and Rangpur have the lowest Temperatures. On the other hand, Barisal and Khulna have higher temperatures than others. And all of the Division has increased a lot from 1999 to 2022.

Figure 4.5 displays the Yearly Average Mean Temperature of 8 individual Divisions. Khulna has the highest temperature, which is between 26 and 27 °C. The Temperature of Barisal was extremely degraded in 2002. Otherwise, in another year, it was almost higher than any other division, unlike Khulna. Rangpur and Mymensingh have the Lowest Temperature, which fluctuates between 25 °C to 25.5 °C. It is increased by almost 1°C to 2 °C from 1999 to 2022.

(ii) Analysis of Monthly Average Temperature

In the Analysis of Monthly Average Temperature, we delve into the intricate variations of temperature on a monthly basis. Utilizing Python, we calculate the average temperature for each month across all stations and all year, providing insights into seasonal patterns and trends. This examination offers a detailed understanding of the distribution and intensity of temperature throughout the year, which is essential for discerning climatic dynamics and informing adaptive strategies. Calculated average temperature for each month throughout all these years.

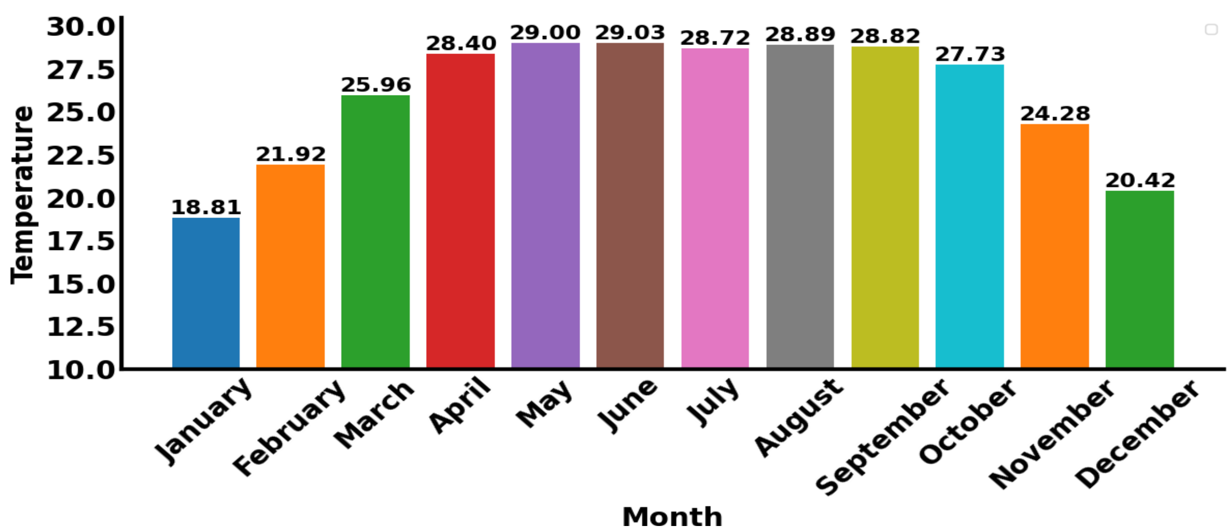
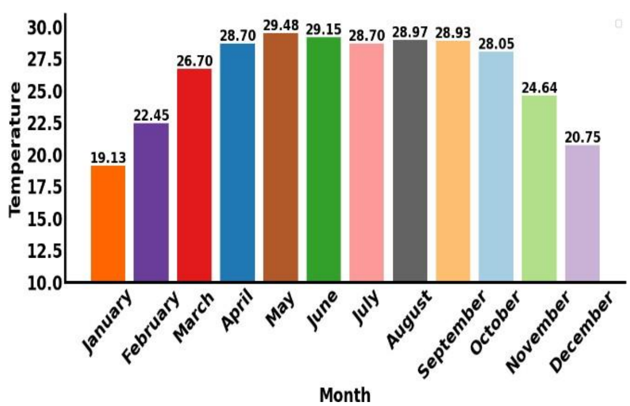


Figure 4.6 Monthly Average Temperature of Bangladesh

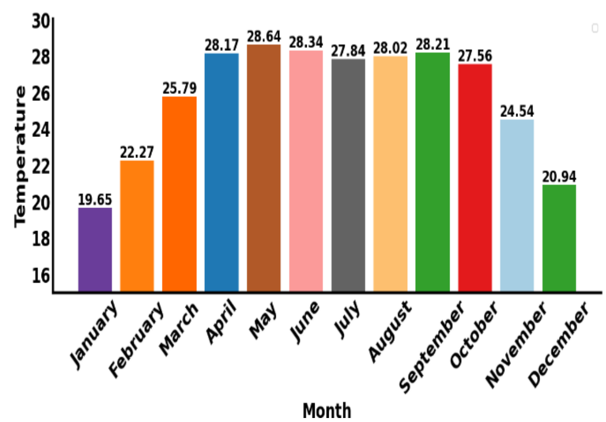
From Figure 4.6, we noticed that May and June have the highest temperatures, 29.03 and 29.00 Degrees Celsius. After that, July, August, and September have the higher temperature. January and December have the lowest temperature. The graph depicts the average monthly temperatures exhibit a noticeable variation. January starts with the lowest average temperature at 18.21 degrees, while April experiences the highest average temperature at 29.03 degrees. The graph suggests a seasonal pattern, with warmer temperatures in the middle months and cooler temperatures at the beginning and end of the year.

Figure 4.7 shows the Monthly Average Temperature of 8 individual Divisions. 4.7(a) Visualize the monthly average Temperature of the Barisal Division, Where May has the highest temperature (29.48 °C.); after that, June has the second highest temperature. This figure shows us that January has the lowest temperature, as we know that January lies in the winter season. The temperature increases from January and increases till May; after that, in June and July, when the Rain starts, the temperature becomes low, and then again increases in October and November then decreases.

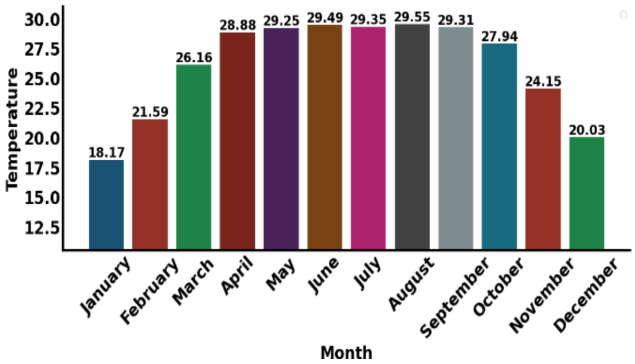
4.7(b) shows the Temperature of Chittagong City, which actually has the same pattern as Barisal.4.7(c), (d), (e), (f), (g), and (h) show Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet Division. They all have higher Temperatures in April, May, June, and July. On the other hand, the start and end of the year have the lowest Temperatures.



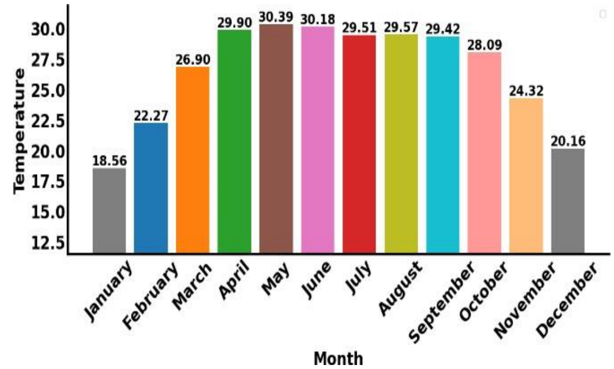
(a) Barisal



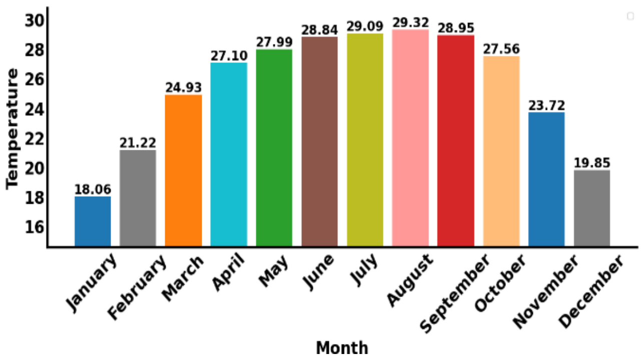
(b) Chittagong



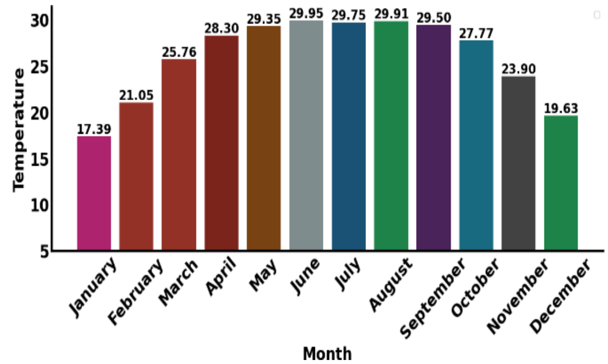
(c) Dhaka



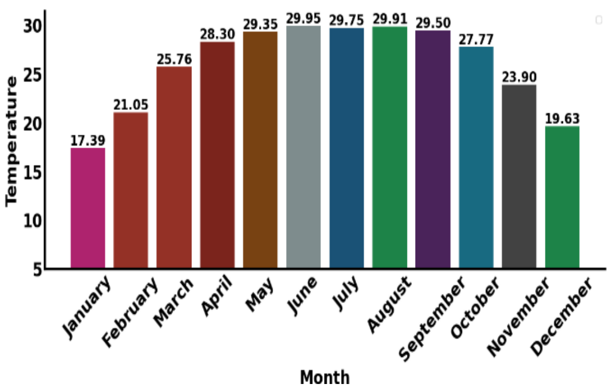
(d) Khulna



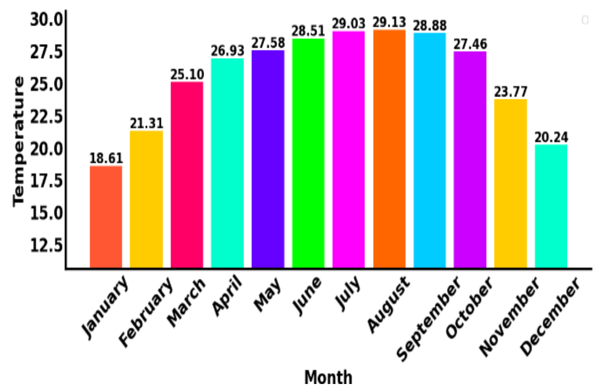
(e) Mymensingh



(f) Rajshahi



(g) Rangpur



(h) Sylhet

Figure 4.7 Monthly Average Temperature of Eight Division

(iii) Analysis of Seasonal Average Temperature

There are six Seasons in Bangladesh: summer; which lasts from April 14 to June 13, Monsoon; which lasts from 14 June to 14 August, Autumn; which lasts from 14 August 13October, Late Autumn; which lasts from 14 October to 13 December, Winter; Lasts from 14 December 13 February and Spring, which lasts from February 14 to April 13. In this section, we discuss the variation in temperature in each season.

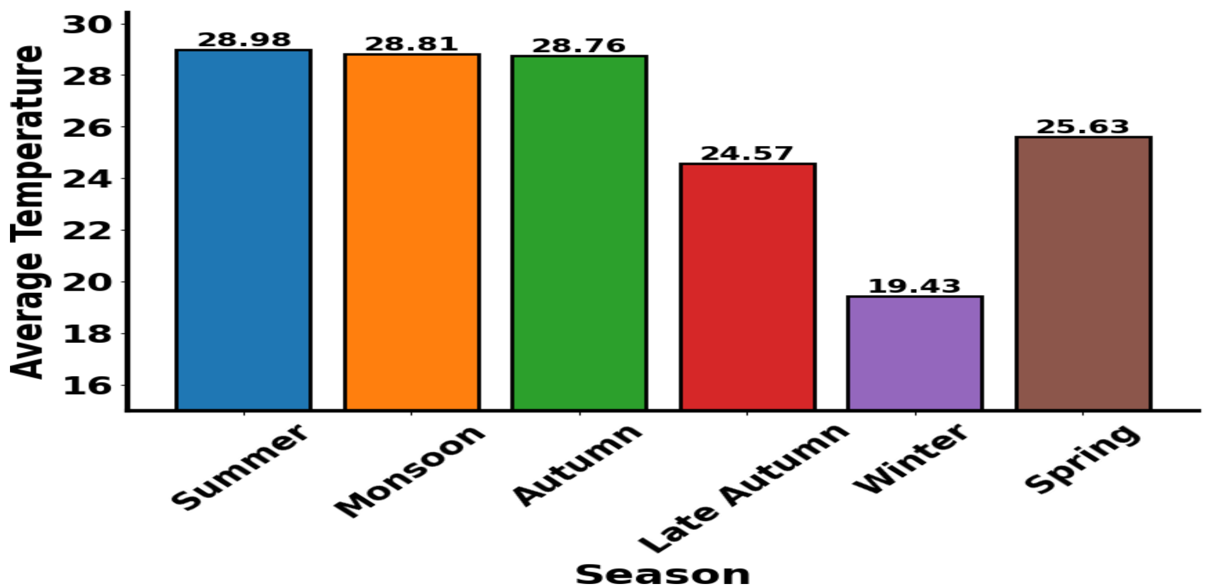


Figure 4.8 Average Seasonal Temperature of Bangladesh

We have visualized these seasonal trends in temperature in Figure 4.8. The graph shows the average temperature for various seasons. Summer has the highest average temperature at 28.98, followed closely by Monsoon (28.81 °C) and Autumn (28.76 °C). Late Autumn experiences cooler temperatures (24.57 °C), while Winter is even cooler (19.43 °C). Spring falls in between, with an average temperature of 25.63. Overall, the graph visually compares the warmth of Summer through Autumn and the cooler temperatures of Late Autumn through Spring.

B. Analysis of Temperature Data on Interval 1980-2022

We have Divided our Temperature Datasets into 2 intervals, and in the previous section, we analyzed the dataset of 1999-2022. In this section, we will discuss the second interval, which is 1980-2022. At this time, we have 26 weather stations. Here, the number of weather stations is less than in 1999 because many of them were not established that year. We also analyzed the annual, monthly, and seasonal temperature variation at this interval.

i. Analysis of Yearly Average Temperature

Here, we discussed the yearly change in temperature from 1980 to 2022 and visualized the changes in a graph.

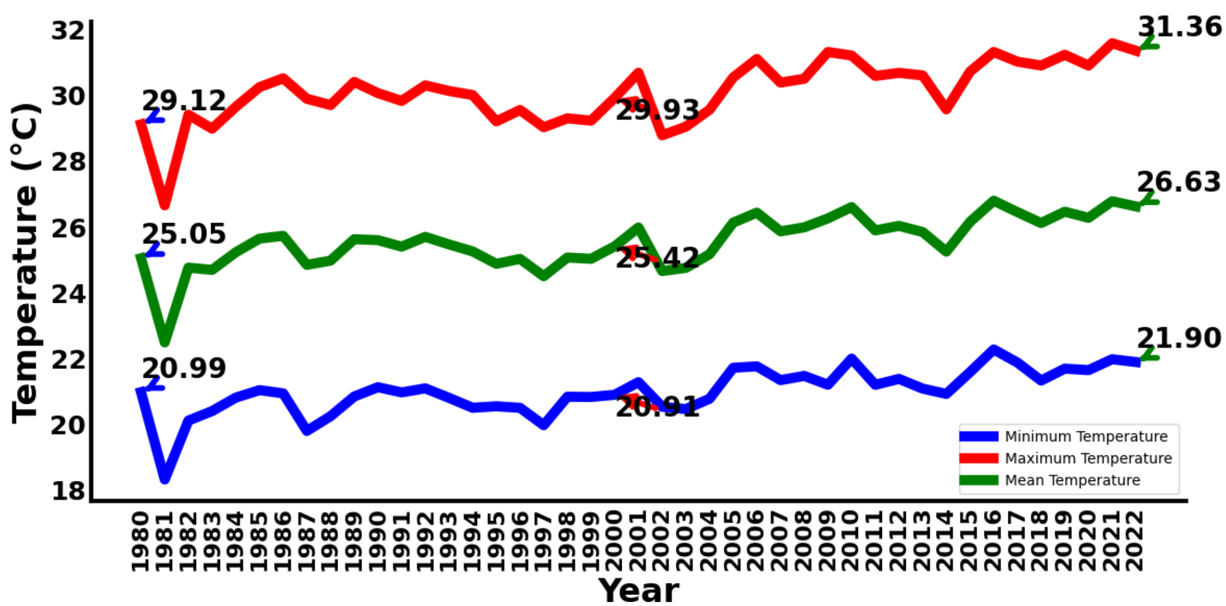


Figure 4.9 Yearly Average Temperature of Bangladesh

Figure 4.9 displays trends in minimum, maximum, and mean temperatures over a series of years (1980-2022). The x-axis represents the years, while the y-axis represents Temperature in degrees Celsius. In the early years (around 1980), the minimum Temperature was approximately 20.99°C. It decreased slightly in 2001 when it was 20.92°C. After that, it gradually increased to around

21.90°C by 2022. Overall Changes almost one °C from 1999-2022. Initially, the maximum temperature was around 29.12°C. Over time, it fluctuated but remained relatively stable, ending at approximately 31.36°C in 2022. The mean temperature started at around 25.05°C in 1980 and rose steadily. By 2001, it reached approximately 25.42°C. In 2022 it will increase to 26.63 °C. Overall, the graph shows temperature variations over the years, with a slight decline in maximum and minimum temperatures and a rise in mean temperature.

ii. Analysis of Monthly Average Temperature

There are 12 months in a year, and in this section, we will analyze the monthly average temperature using the data from 1980 to 2022.

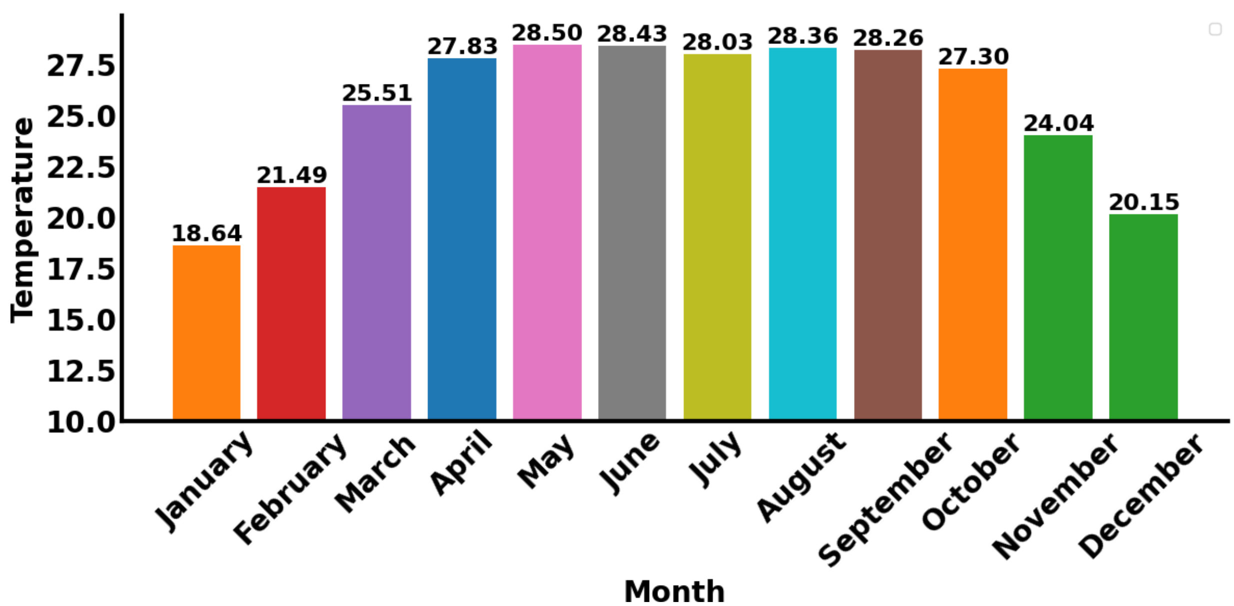


Figure 4.10: Monthly Average Temperature of Bangladesh

Figure 4.10 Displays the monthly average temperature throughput the year. The graph displays temperature variations across different months. In January, the temperature starts at 18.64 degrees, gradually increasing to a peak of 28.50 degrees in May. From there, it decreases, reaching its lowest point of 20.15 degrees in December. Notably, may has the highest recorded temperature, while December experiences the coldest weather. This graph provides valuable insights into seasonal patterns and can be useful for planning activities based on expected weather conditions throughout the year.

iii. Analysis of Seasonal Average Temperature

Like the previous discussion of seasonal trends, we will discuss the same thing here using the data set of 1980-2022. One notable matter is that in this analysis, we are using the Average/mean Temperature instead of the Maximum or Minimum Temperature.

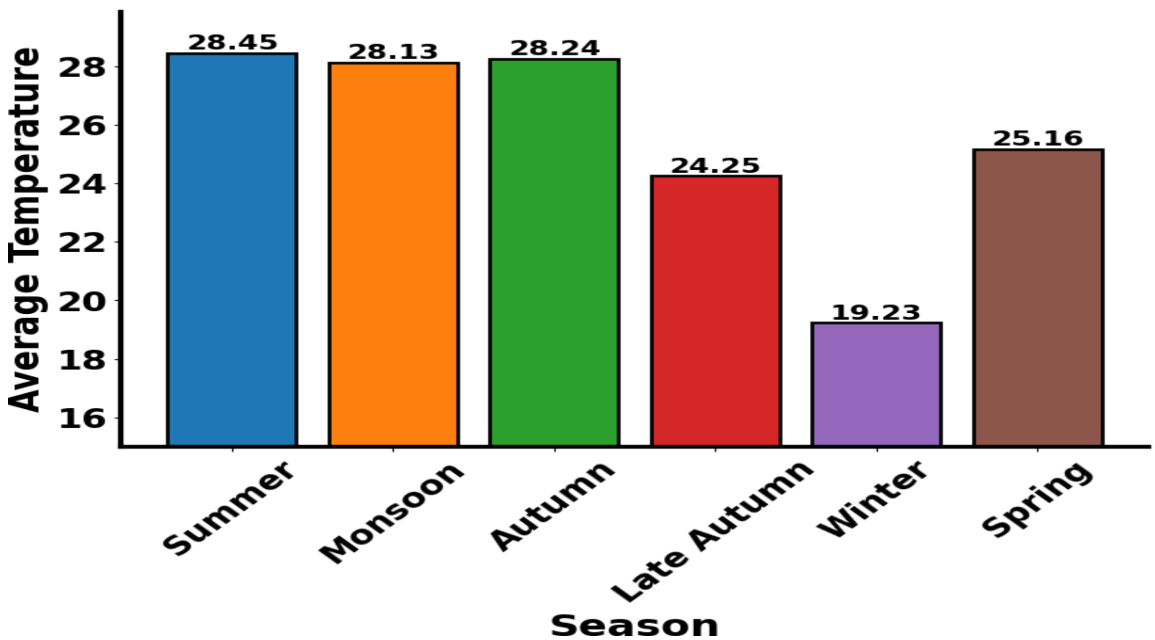


Figure 4.11 Average Seasonal Temperature of Bangladesh

Figure 4.11 provides insights into how temperatures fluctuate throughout the year, highlighting distinct seasonal patterns. During the summer months, the average temperature peaks at 28.45 degrees Celsius. This season is characterized by warm and often hot weather. The monsoon season follows summer, bringing heavy rainfall and cooler temperatures. Despite the rain, the average temperature remains relatively high at 28.13°C. As the monsoon recedes, autumn arrives with pleasant weather. The average temperature hovers around 28.24°C. Late autumn experiences a noticeable drop in average temperature compared to the previous seasons. At 24.25°C, it signifies the onset of cooler weather. Winter is the coldest season, with an average temperature of 19.23°C. It brings chilly days and cold nights. People bundle up, and activities shift indoors. Winter is associated with holidays, festivities, and cozy evenings by the fireplace. As winter fades, spring emerges with milder temperatures. The average temperature rises to 25.16°C. Spring is characterized by blooming flowers, rejuvenation, and a sense of renewal.

4.2.2 Rainfall Data Analysis

Our analysis encompasses two distinct intervals of rainfall data: 1999-2022 and 1980-2022. We scrutinize yearly average rainfall patterns across Bangladesh and eight individual stations within these periods. Employing line graphs, we depict the temporal evolution of Rainfall, capturing both short-term variations and long-term trends. Furthermore, we delve into monthly average rainfall data to elucidate seasonal fluctuations, offering valuable insights into the climatic dynamics of Bangladesh. By segmenting the data into these intervals, we gain a comprehensive understanding of precipitation patterns, which is essential for informing adaptation strategies and enhancing resilience to changing climates.

A. Analysis of Rainfall Data on Interval 1999-2022

This interval includes 34 individual Weather Stations. We Divided the data into 8 Divisions: Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet. Each division consists of many stations. We analyzed each Division individually so that we could make clearer comparisons.

(i) Analysis of Yearly Average Rainfall

We made a yearly average graph to understand the yearly trends and changes in rainfall each year. This gives us an overview of climate change and the changing behavior of the Rainfall each year. Python was utilized to compute the average rainfall for each year. This involved aggregating monthly data from all stations and then calculating the yearly average. This methodology comprehensively explains annual rainfall trends across the study period. Here, we will analyze the yearly average graph for Bangladesh (which contains all 34 weather stations) and 8 Individual divisions (which contain the weather stations in this Division).

In Figure 4.12, we notice that the Maximum Rainfall was in 2017, which was 253.83(mm) on the other hand, the minimum rainfall was in 2022. It is clearly seen that rainfall will decrease from 1999 to 2022. In 1999, it was around 210 mm; on the other hand, in 2022, it was 149.65 mm, which is significantly less than in 1999. It is an alarming matter for Bangladesh as well as the world, as it indicates climate change. Decreasing Rainfall is one of the effects of Climate change.

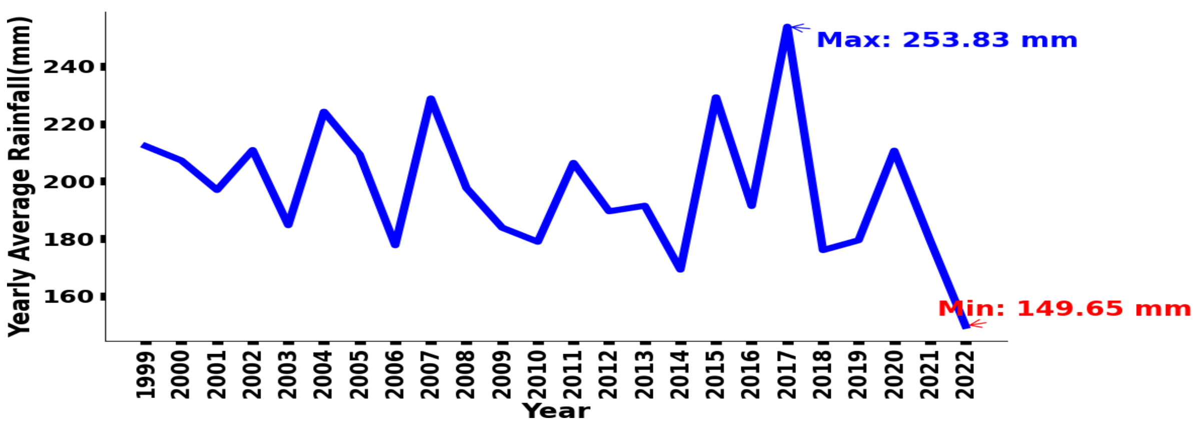
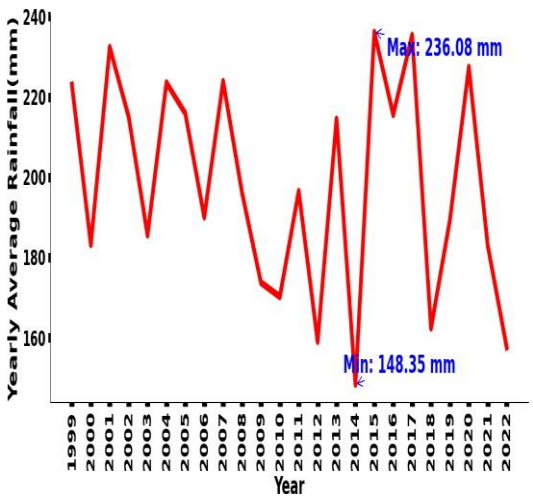
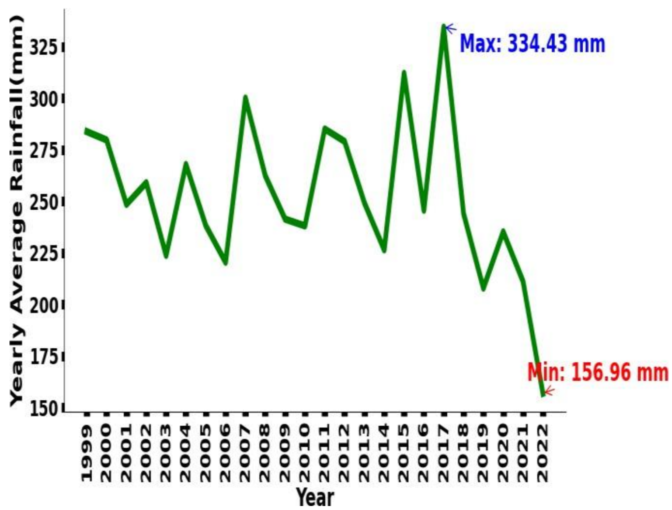


Figure 4.12: Yearly Average Rainfall of Bangladesh

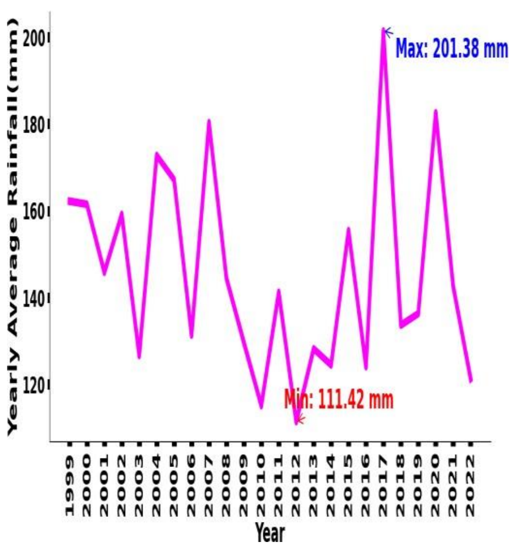
Figure 4.13 shows the yearly average Rainfall of 8 Individual divisions using 8 Different graphs for better analysis of each of 8 Divisions individually; on the other hand, figure 4.14 Shows eight divisions in the same graph for better comparison. In Figure 4.13, one common thing in each graph, instead of Sylhet, is the decrease in rainfall from 1999 to 2022. In Sylhet it is slightly greater than 1999. From this, Sylhet has the highest maximum rainfall, which is 406.54 mm, which was recorded in 2017. After Sylhet, Chittagong has the Second Highest Maximum Rainfall, and Mymensingh and Barisal are in the next position. It is noticeable that Rajshahi Has the lowest Maximum Rainfall.



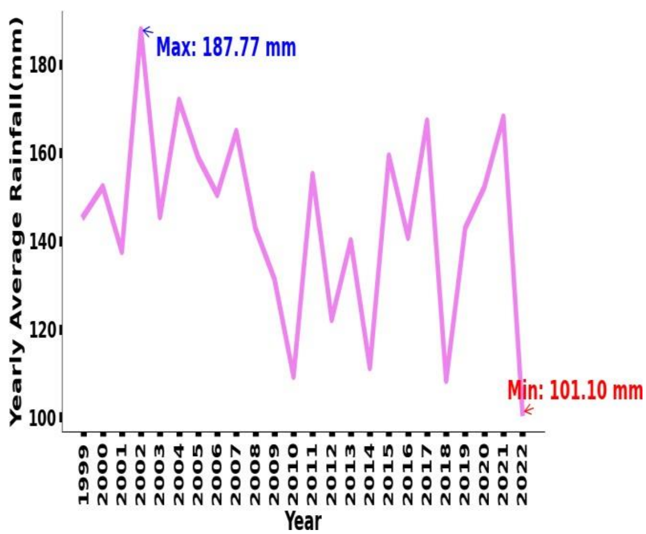
(a) Barisal



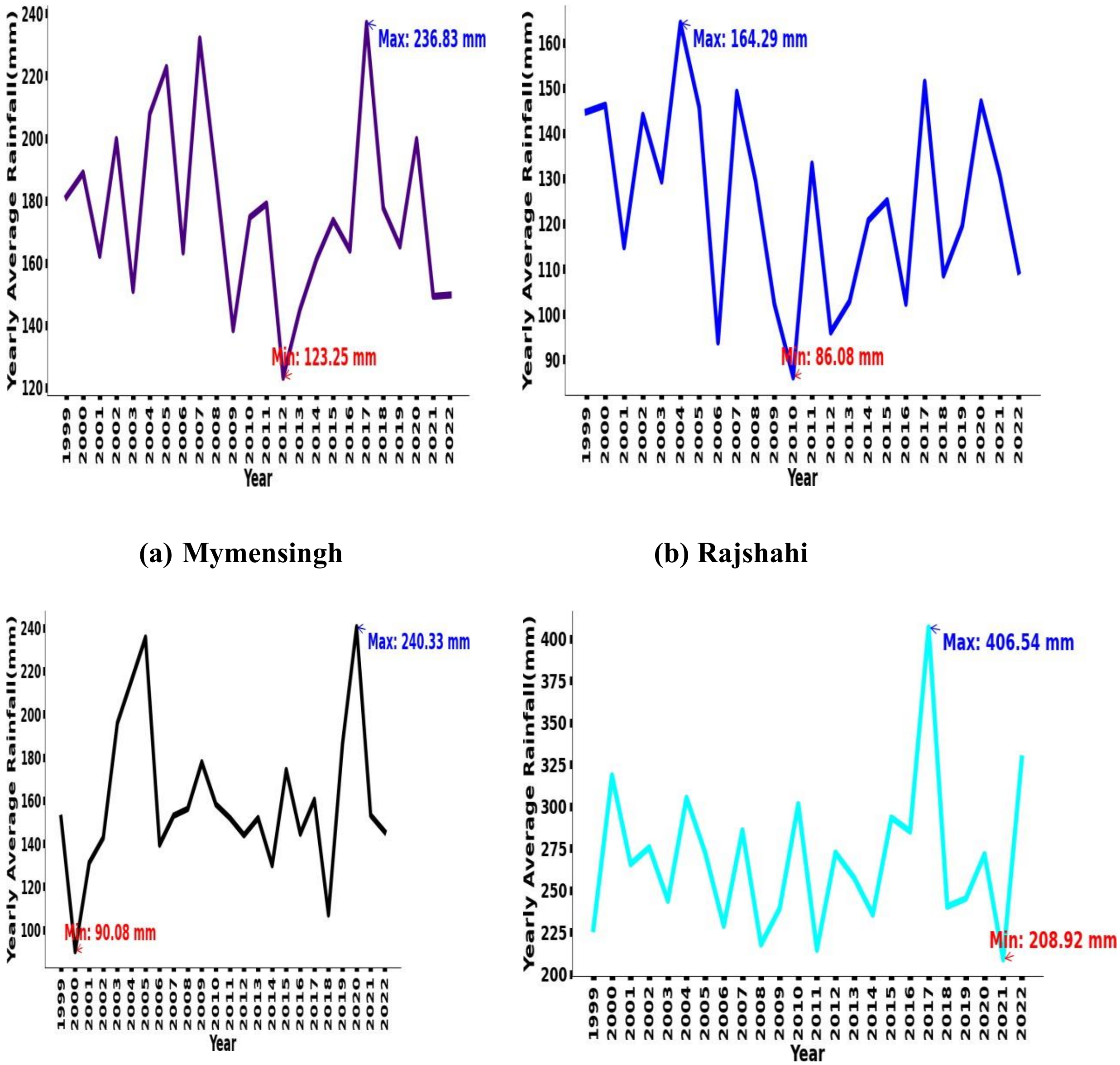
(b) Chittagong



(a) Dhaka



(b) Khulna



(a) Mymensingh **(b) Rajshahi**

(c) Rangpur **(d) Sylhet**

Figure 4.13: Yearly Average Rainfall of Eight Individual Division in a separate graph

Figure 4.14 also shows the same thing. Sylhet had the highest rainfall in 2017 and the lowest in 2021. After that, Chittagong and Barisal had the highest average rainfall. In the graph below, Rajshahi and Rangpur lie.

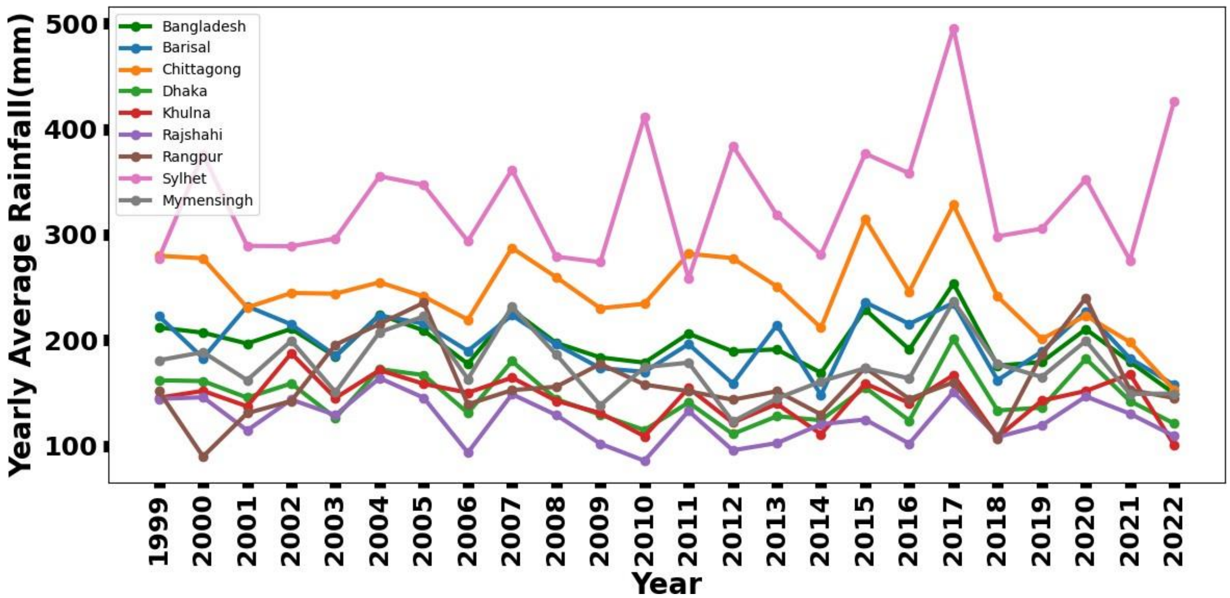


Figure 4.14: Yearly Average Rainfall of Eight Individual Division in the Same graph

In the graph, besides the highest Rainfall of Sylhet, the lowest Rainfall is 86.08, which was recorded in Rajshahi in 2010. Every Division decreased its Rainfall in 2022, which can be an effect of global warming, and it's a matter of concern.

(ii) Analysis of Monthly Average Rainfall

In the Analysis of Monthly Average Rainfall, we delve into the intricate variations of precipitation on a monthly basis. Utilizing Python, we calculate the average rainfall for each month across all stations and all year, providing insights into seasonal patterns and trends. This examination offers a detailed understanding of the distribution and intensity of Rainfall throughout the year, which is essential for discerning climatic dynamics and informing adaptive strategies.

Figure 4.15 shows the monthly Average Rainfall of Bangladesh (which includes all 34 weather stations). From the figure, we can say that The Maximum Rainfall occurred in July, and it was 520.47 mm. On the other hand, the lowest rainfall occurred in November, December, January, and February, which is almost nearest to 0(mm)—10(mm).

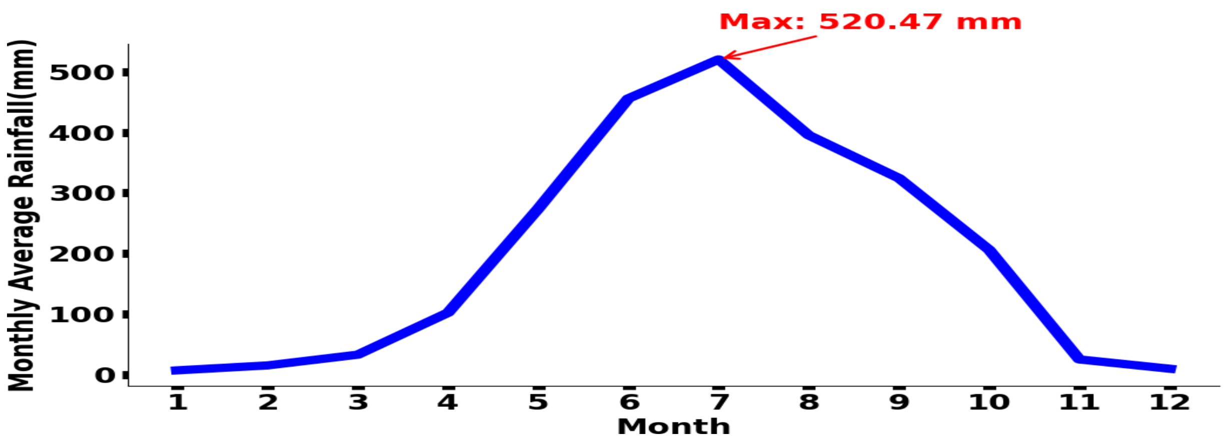
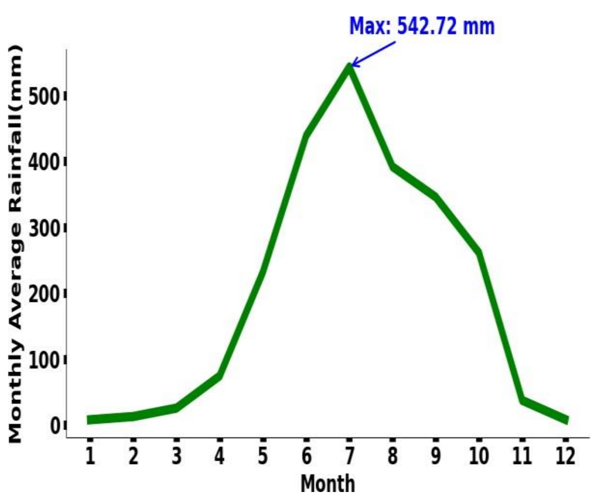
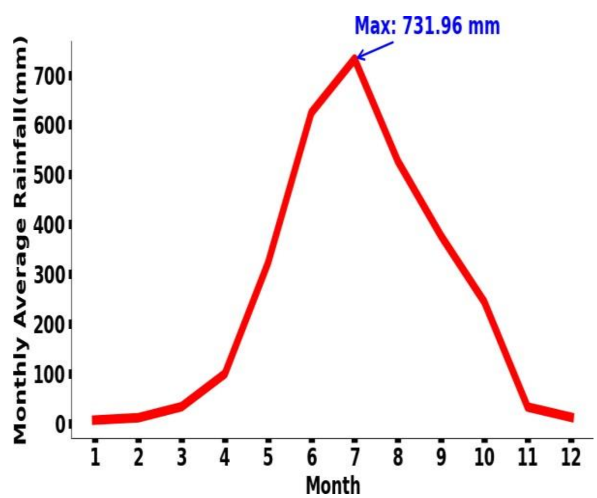


Figure 4.15: Monthly Average Rainfall of Bangladesh

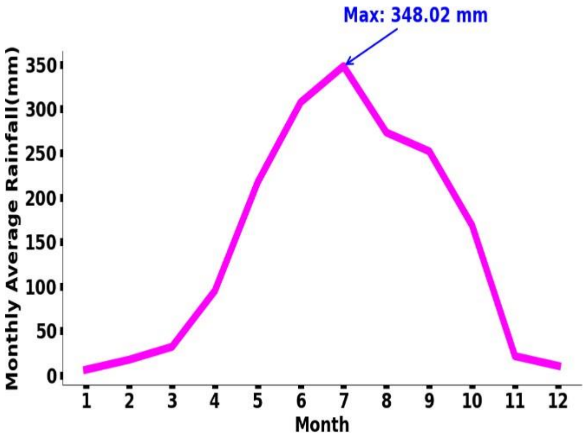
Figure 4.16 Displays the monthly average Rainfall of all 8 Divisions of the Country. We can notice a common thing: the rainfall is lowest in January, February, November, and December; on the other hand, it is highest in June and July. From the graph, we can see that Chittagong has the highest monthly Average Rainfall, which is 731.96(mm); after that, Sylhet has the second highest Rainfall that is 622.38(mm). And Rajshahi has the lowest rainfall, which is only 290.81(mm). After that, Khulna and Dhaka have the lowest rainfall, only 348.02 (mm) and 379.3 4(mm), respectively. Barisal, Chittagong, Dhaka, and Khulna Division July is the month with the highest rainfall.



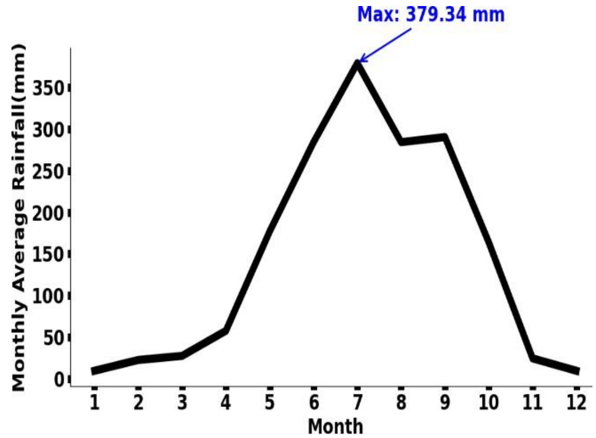
(a) Barisal



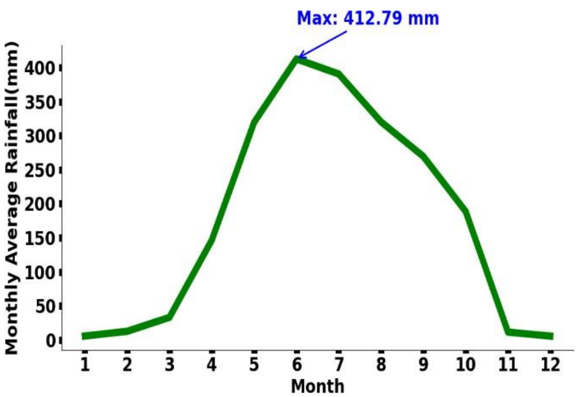
(b) Chittagong



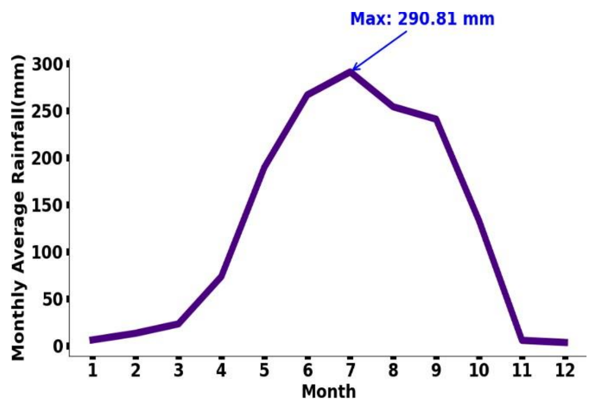
(c) Dhaka



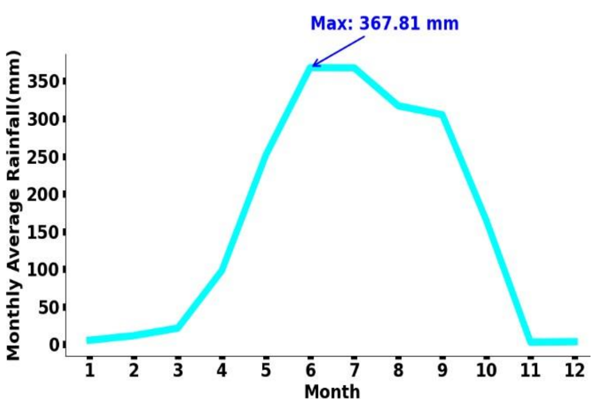
(d) Khulna



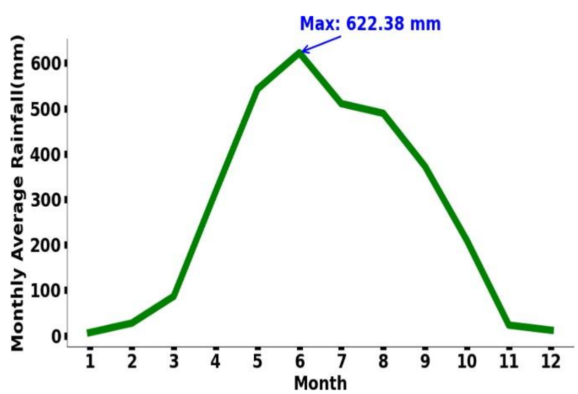
(e) Mymensingh



(f) Rajshahi



(g) Rangpur



(h) Sylhet

Figure 4.16: Monthly Average Rainfall Eight Individual Division

On the other hand, in Mymensingh, Rajshahi, Rangpur, and Sylhet Division, June is the month with the highest rainfall. From January, rainfall increases till June or July; after June or July, rainfall decreases again each month till December. There is a notable matter that, after June and July, Rainfall Decreases in August, but in September and October, it increases more than in August. And After October again it Decreases. And December, January, February have the lowest rainfall.

(iii) Analysis of Seasonal Rainfall

Here are 6 distinct Season in Bangladesh, but for Rainfall, we will discuss about the 3 major Season as Summer, Rainy Season(monsoon) and Winter. So we divided our Datasets into 3 Season as Summer from March to May, Rainy Season from June to October and Winter from November to February for Seasonal Analysis. We plot the Yearly Average Rainfall line of 3 season in the same graph to compare more easily.

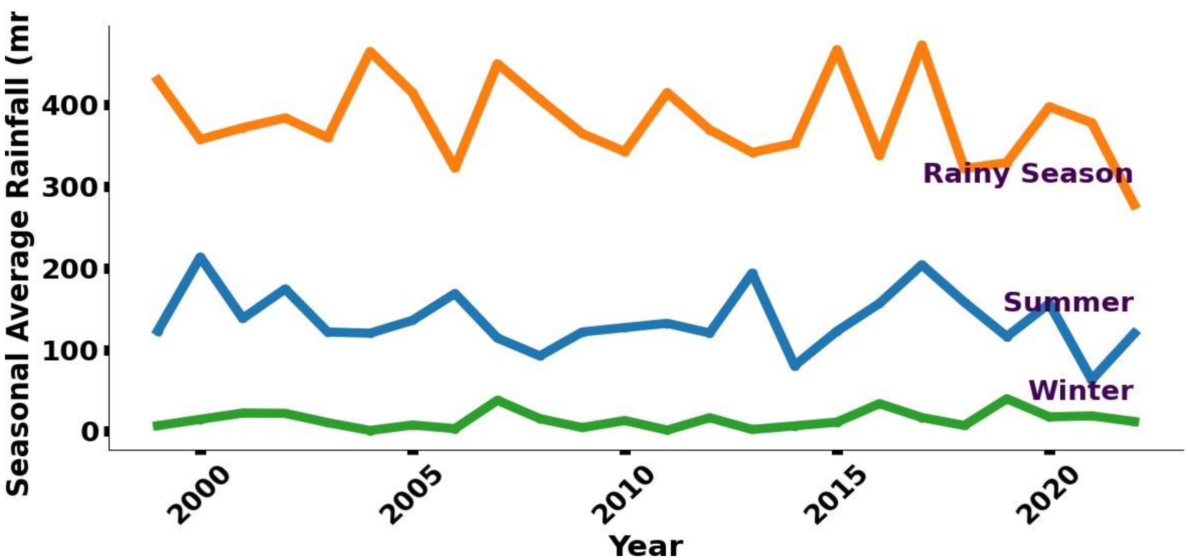


Figure 4.17: Seasonal Average Rainfall of Bangladesh

Here, in figure 4.17, We can see that the Winter Season Rainfall is the lowest; the maximum average Rainfall in the Winter Season is only 39.91 mm, which was recorded in 2019; the other Hand Rainy Season, The Maximum Average Rainfall is 493.62 mm was recorded in 2004; but it is a matter of Concern that the Rainfall in Rainy Season is Decreasing. In 2022 it was almost mm, and there is a huge Difference. In 1999 it was above 400, almost 100 mm less than 1999 in 2022. It is an effect of Climate Change and Global Warming. In summer the maximum rainfall was in 2000, and after that, it decreased a lot.

B. Analysis of Rainfall Data on Interval 1980-2022

Here, we used the Rainfall Data from 1980-2022, Which consist of 32 unique Weather Station. And the reason behind the division was declared before. In This Interval, we made Yearly Average Rainfall, Monthly Average Rainfall and Seasonal Average Rainfall of Bangladesh.

(i) Analysis of Yearly Average Rainfall

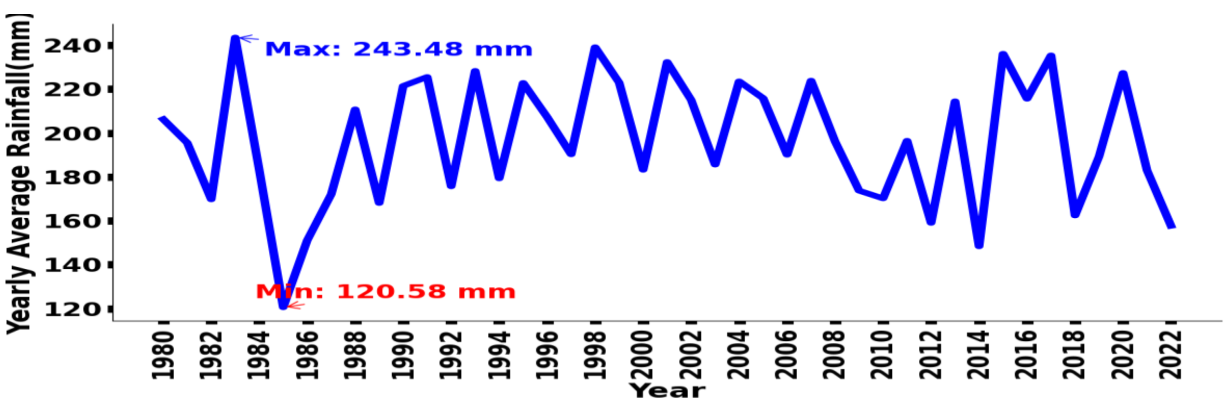


Figure 4.18: Yearly Average Rainfall of Bangladesh

Here, we calculate the Yearly Average Rainfall for each year from 1980 to 2022. In Figure 4.18, we notice that the minimum rainfall was recorded in 1985 at only 120.58 mm. On the other hand, the maximum rainfall recorded in 1983 was 243.48 mm. In 1980, the yearly average Rainfall was above 200 mm. On the other hand, in 2022, it is below 170 mm. As a yearly average, it is a huge Difference.

(ii) Analysis of Monthly Average Rainfall

Like the Yearly Average Rainfall, we analyzed the monthly average Rainfall, which allows us to show the monthly variation of rainfall.

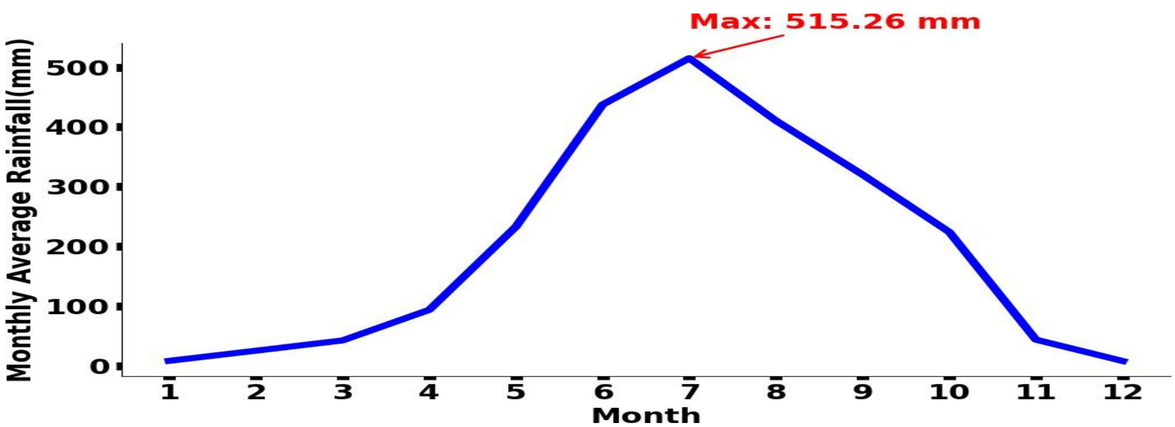


Figure 4.19: Monthly Average Rainfall of Bangladesh

In Figure 4.19, we can see that July has the Highest Rainfall. On the other hand, it decreases from July to December and then increases again from January. The maximum Rainfall of July is 515.26 mm.

(iii) Analysis of Seasonal Average Rainfall

In a similar manner to interval 1999-2022, we divided our Datasets on interval 1980-2022 into 3 Seasons, Winter, Summer, and Rainy Season, and plotted the yearly Average of each Season. Figure 4.20 illustrates the seasonal average Rainfall over the years from 1980 to 2020. Three distinct lines represent different seasons: the orange line corresponds to the rainy season, showing significant peaks and troughs; the blue line represents winter, remaining relatively low and stable; and the green line depicts summer, with moderate variation. Overall, the rainy season experiences the highest levels of Rainfall, and the graph reveals fluctuations in rainfall patterns over time. There is a frequent change in rainfall in each season. The maximum rainfall of the rainy season is 526.19 mm, which happened in 2015, and the maximum rainfall in summer is 237.00 mm, recorded in 2013. Around 1980, all three lines started at relatively low rainfall levels. There was an upward trend in rainfall until the early 1990s.

From the mid-1990s to the early 2000s, rainfall declined. After 2000, rainfall gradually increased. By 2020, the rainy season showed its highest peak, while winter remained consistently low.

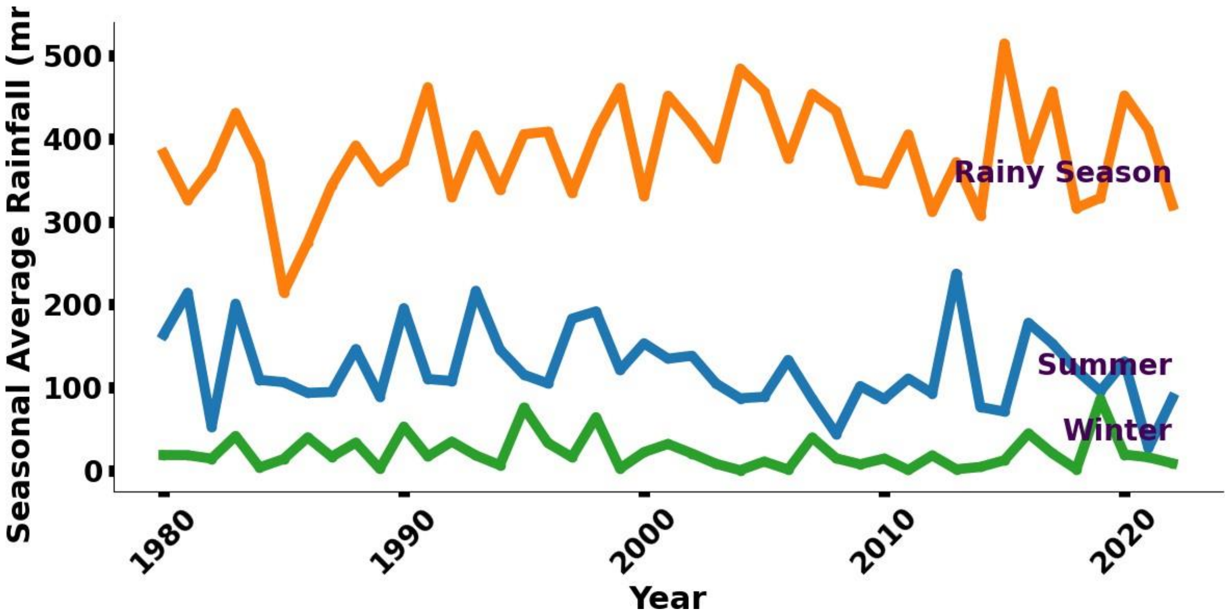


Figure 4.20: Seasonal Average Rainfall of Bangladesh

4.2.3 Humidity Data Analysis

Humidity, the concentration of water vapor in the atmosphere, is a pivotal factor influencing both weather and human well-being. By meticulously examining its fluctuations annually, monthly, and seasonally, we gain insights into its impact on various ecosystems and human activities. Understanding humidity dynamics not only enhances our grasp of weather patterns but also aids in forecasting and mitigating potential risks such as heat stress and crop failures. Through systematic analysis, we uncover the intricate interplay between humidity, temperature, and precipitation, which is crucial for sustainable resource management and adaptation strategies in a changing climate landscape. Through meticulous annual, monthly, and seasonal analyses, we delved into the intricate variations of humidity levels, shedding light on its dynamic role in atmospheric dynamics and climatic patterns.

A. Analysis of Humidity Data on Interval 1999-2022

In this interval we have 34 weather station. We analyzed Yearly, Monthly, and Seasonal Humidity. In each time or year Humidity level is different. And There is a different effect of different Humidity level.

(i) Analysis of Yearly Average Humidity

Here, we used 1999-2022 all the 24 years for our analysis, and made a graph of average yearly humidity. Where, humidity of any year is the integration of each 34 weather station.

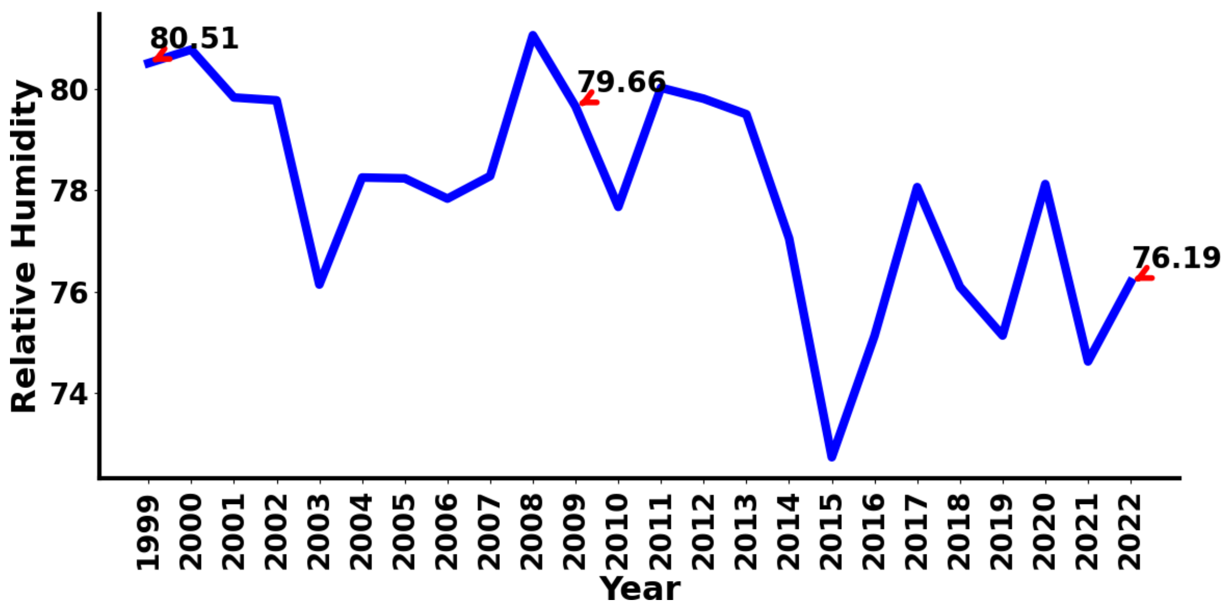


Figure 4.21 Yearly Average Humidity of Bangladesh

Figure 4.21 depicts relative humidity over the years from 1990 to 2021. Here are the key points and potential effects**.** The graph shows fluctuations in relative humidity levels over time. Around 1999, there was a peak with a relative humidity value of 80.51%. Another peak occurred around 2010, with a relative humidity value of 79.66%. By 2022, the relative humidity drops to 76.19%. These variations indicate changes in environmental moisture content. From the graph, it is visible that the humidity is decreasing day by day; from 1999 to 2022, it has decreased by 4.32%.

Decreasing humidity can have significant environmental ramifications, affecting various ecosystems and human activities. As humidity levels drop, ecosystems reliant on moisture face challenges such as reduced plant growth, increased risk of wildfires, and habitat loss for species adapted to humid conditions. Agricultural productivity may decline due to water stress, impacting crop yields and food security. Additionally, decreased humidity can exacerbate desertification and soil erosion in arid regions, leading to degraded land and diminished biodiversity. Furthermore, lower humidity levels can contribute to air quality issues, as dust and pollutants become more easily suspended in drier air, posing risks to human health and exacerbating respiratory conditions.

(ii) Analysis of Monthly Average Humidity

Like monthly Rainfall and Temperature, humidity varies in each month. It is related to temperature and Rainfall also. Here, we discussed this.

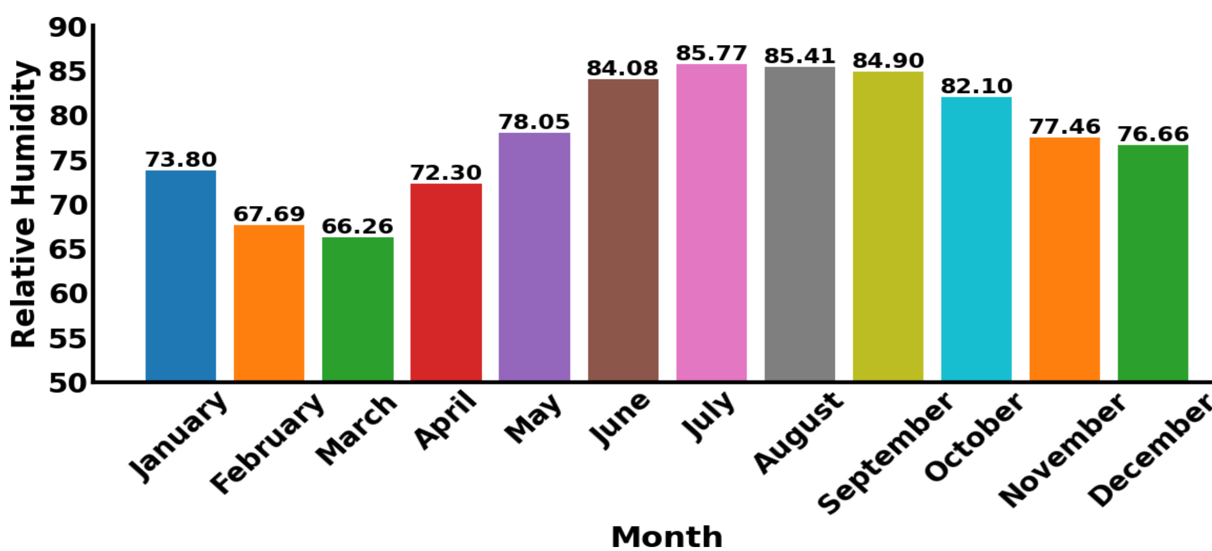


Figure 4.22: Monthly Average Humidity of Bangladesh

The bar graph in Figure 4.22 represents relative humidity percentages for different months of the year. In January, the relative humidity was around 73.80%. In February, it dropped to approximately 67.69%. In March, the humidity level is 66.26%. April: It remains similar at 65%. May. The relative humidity increases to 78.05%. June: It further rises to 84.08%. July: The highest humidity occurs in July, reaching 85.77%. August: It remains high at 85.41. September: The humidity starts decreasing, reaching 82.10%. October: It continues to decline to 77.46%. November: The relative humidity drops to 72.30%. December: The lowest humidity occurs in December, at 70%. These fluctuations in relative humidity have various implications for the environment, as mentioned in my previous response. For instance, higher humidity can affect wildfire risk, while lower humidity impacts energy consumption and urban climates.

(iii) Analysis of Seasonal Humidity

Humidity in each season is different. In the winter environment, it is dry during the rainy season. Seasonal humidity levels fluctuate throughout the year, impacting various aspects of our environment and daily lives. During the summer months, humidity tends to be higher due to increased evaporation from warm bodies of water and vegetation, creating muggy and uncomfortable conditions. This high summer humidity can also fuel thunderstorm activity and contribute to localized flooding. In contrast, winter typically brings lower humidity levels, especially in colder regions where the air is drier due to reduced moisture content. While lower winter humidity can alleviate discomfort associated with hot, humid conditions, it can also lead to dry skin, respiratory issues, and static electricity buildup indoors. Understanding these seasonal variations in humidity is essential for managing climate-related risks, optimizing agricultural practices, and maintaining indoor comfort levels throughout the year.

The bar graph in Figure 4.23 depicting average humidity levels across different seasons reveals intriguing patterns. Summer demonstrates a moderate humidity level of approximately 77.51%, followed by a notable increase during monsoon season, peaking at around 85.52%. Autumn maintains high humidity, averaging around 84.80%, while late autumn experiences a slight decline to approximately 78.34%. Winter witnesses a further decrease in humidity to about 73.42%, with spring exhibiting the lowest average humidity at around 67.79%. These findings offer valuable insights into seasonal variations in humidity, aiding climate research, agricultural planning, and understanding of regional weather dynamics.

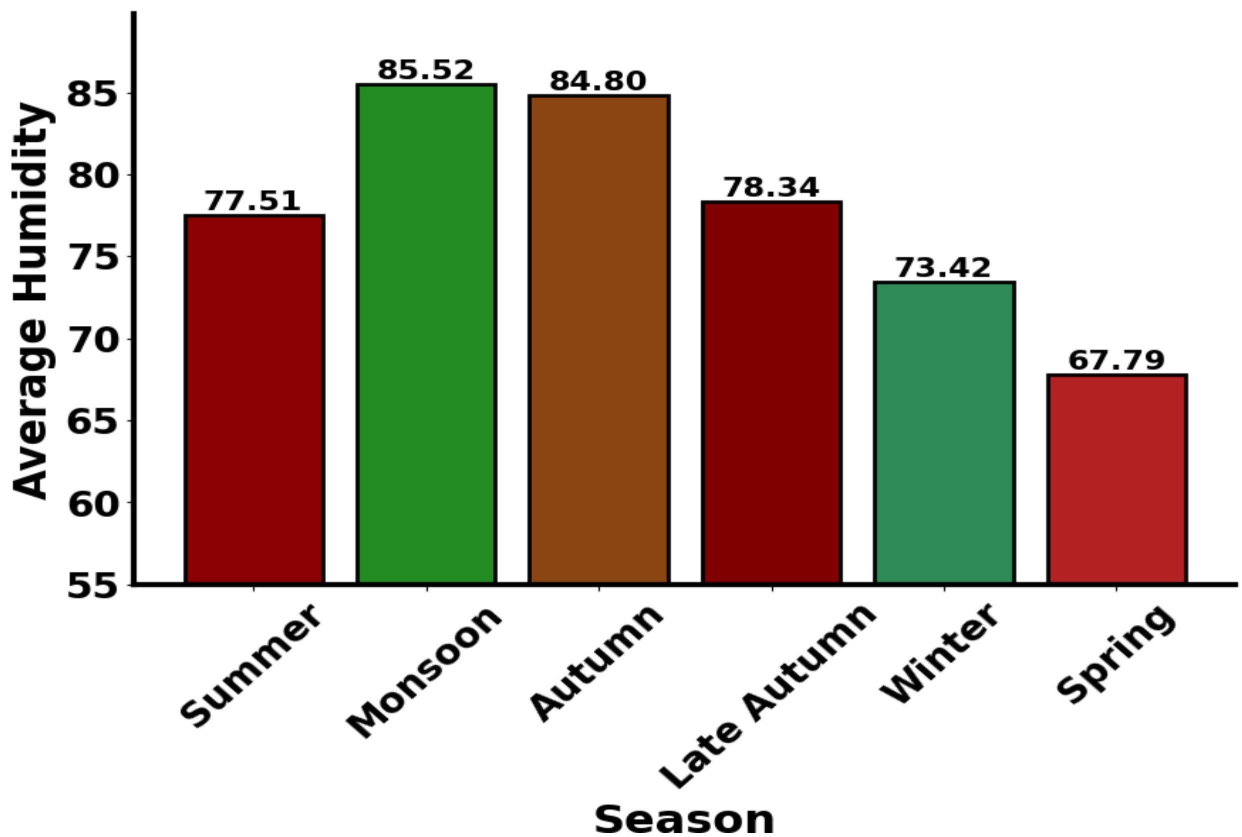


Figure 4.23: Seasonal Average Humidity of Bangladesh

B. Analysis of Humidity Data on Interval 1980-2022

In this section, we will analyze humidity using the datasets from 1980 to 2022. Here we have 26 weather Stations and 43 years of data. A long-term data using this, we will analyze the yearly, monthly, and seasonal humidity.

i. **Analysis of Yearly Average Humidity**

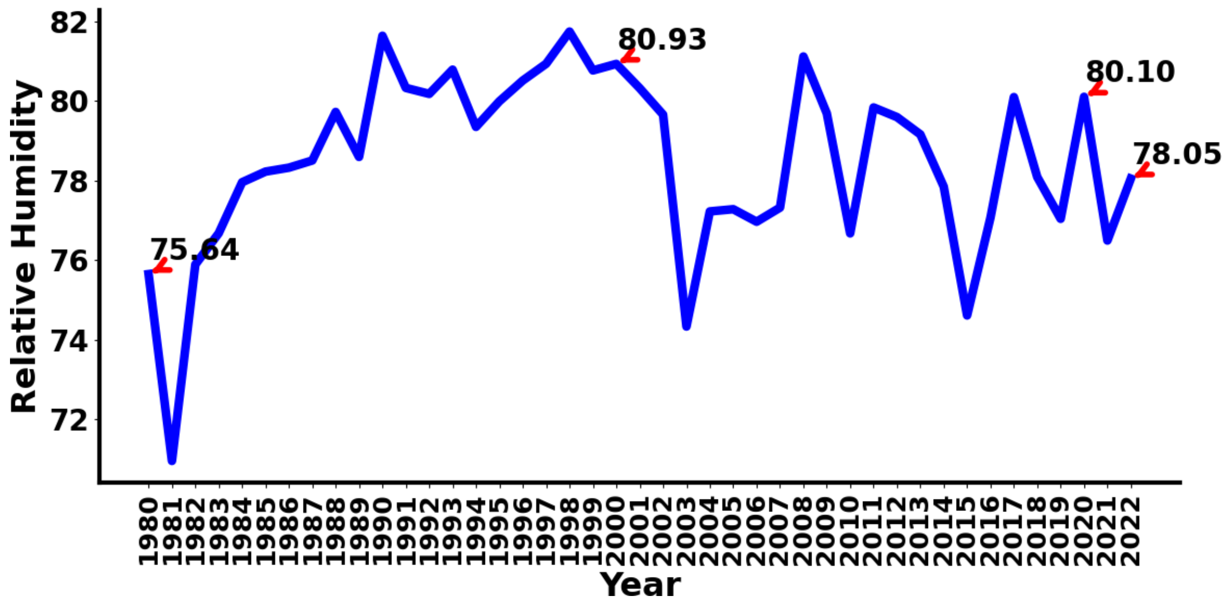


Figure 4.24: Yearly Average Humidity of Bangladesh

The image is a line graph with the vertical axis labeled "Relative Humidity" ranging from 72 to 82, and the horizontal axis labeled "Year" ranging from 1980 to approximately 2022. The graph (Figure 4.24) depicts fluctuations in relative humidity over these years. Around the year 1980, the relative humidity was 75.64. But in the next year, it decreases a lot. In the early 2000s, it reached 80.93. There are two more points in what appears to be the late 2020s or around 2022, marked as 80.10 and 78.05, respectively. The graph shows variations in relative humidity over time. We can observe how it changed from 1980 to 2010. Relative humidity seems to have increased from 1980 to the early 2000s and then decreased slightly. To understand the reasons behind these fluctuations, we'd need additional context. Factors like climate, geographical location, and environmental changes play a role in humidity levels.

ii. Analysis of Monthly Average Humidity

The bar graph provided illustrates the relative humidity percentages for each month of the year. Relative humidity levels vary throughout the year, with July showing the highest humidity at 86.18%, followed closely by August at 85.84%. These months represent the peak of humidity, indicating a very humid period during the summer.

Conversely, March exhibits the lowest relative humidity at 68.23%, suggesting it's the least humid month. The data provides valuable insights into the climatic conditions over the course of the year, aiding in planning and understanding seasonal variations in humidity levels.

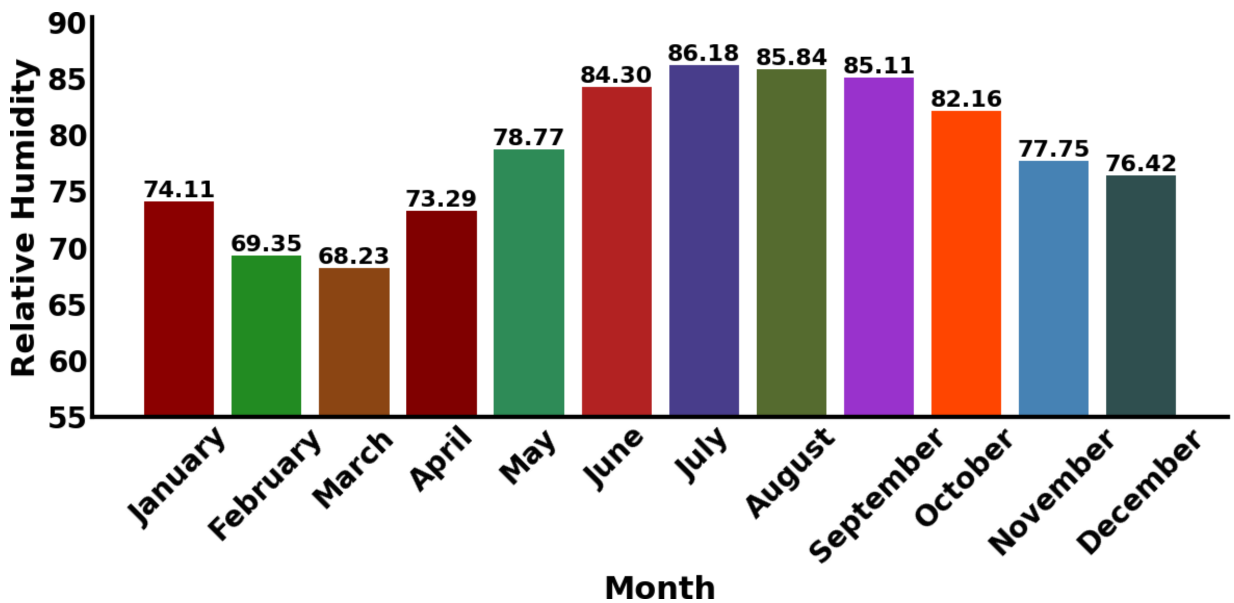


Figure 4.25: Monthly Average Humidity of Bangladesh

iii. Analysis of Seasonal Humidity

The bar graph displayed in Figure 4.26 provides a comprehensive overview of the average humidity percentages across different seasons. It reveals distinct patterns in humidity levels throughout the year, offering valuable insights into seasonal variations and their implications. Summer exhibits a moderate humidity level of 78.35%, typical for warm weather conditions.

Monsoon season stands out with the highest average humidity at 85.89%, indicative of heavy Rainfall and saturated atmospheric moisture.

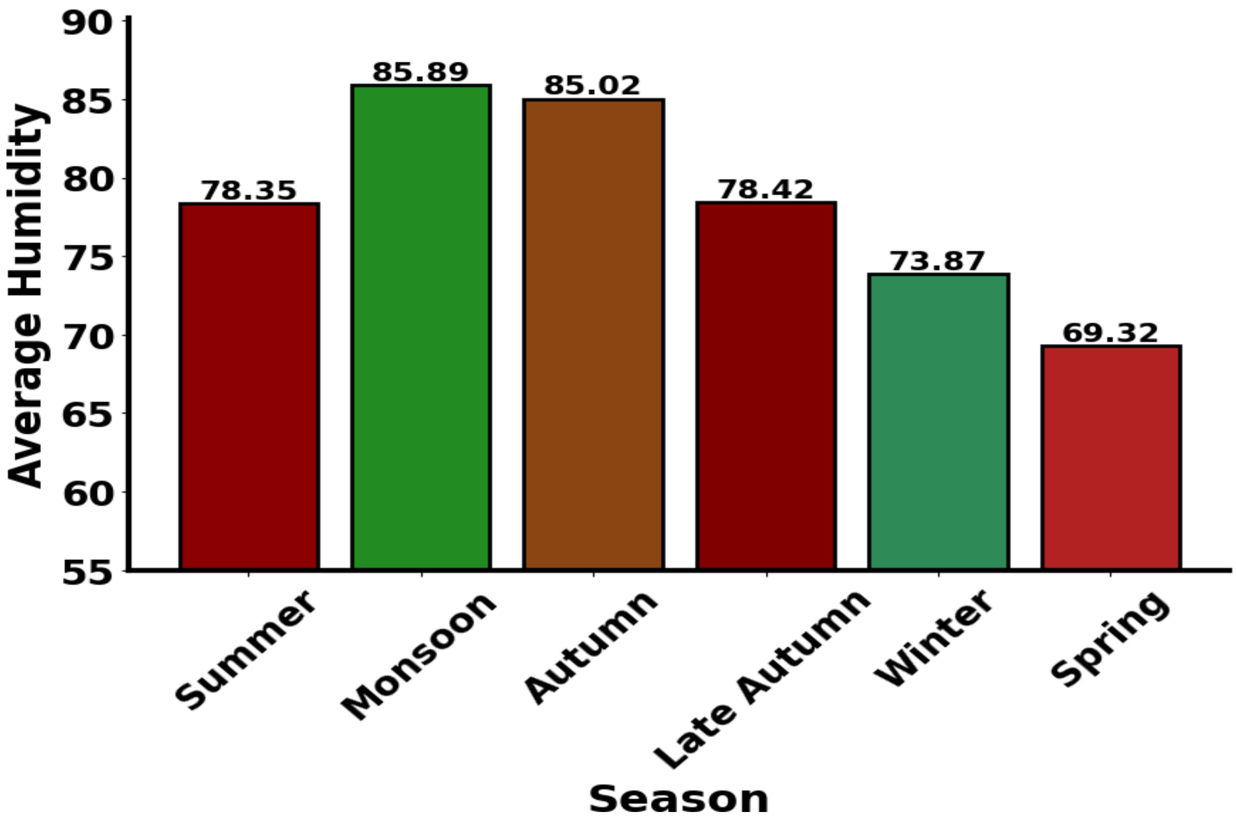


Figure 4.26: Seasonal Average Humidity of Bangladesh

Autumn maintains a high humidity level of 85.02%, suggesting a carryover of moisture from the monsoon transition. Late autumn marks a noticeable decline to 78.42%, signaling the onset of drier conditions. As winter sets in, humidity further decreases to 73.87%, reflecting the characteristic dryness of colder months. Spring showcases the lowest humidity among the seasons at 69.32%, hinting at the onset of warmer and drier weather leading into summer. These findings are crucial for various sectors such as agriculture, water resource management, and tourism, enabling informed decision-making and planning in accordance with seasonal humidity trends.

4.2.4 Relation between Temperature, Rainfall and Humidity

Grasp and forecasting weather patterns and climatic conditions requires a grasp of the link between temperature, humidity, and Rainfall. Temperature affects how much moisture the air can contain, whereas Rainfall adds to atmospheric moisture and affects humidity levels. When taken as a whole, these variables interact dynamically to influence different ecological processes, agricultural practices, and human comfort levels and influence local and regional climates. Knowledge of these relationships can help one better understand the intricate dynamics of Earth's atmosphere and how they affect ecosystems and human cultures.

Here, we will visualize the relationship between temperature, Rainfall, and humidity in yearly, monthly, and seasonal perspectives. In this study, we used data from 1999 to 2022. When using only one temperature in that situation, we used Mean Temperature instead of Maximum and minimum temperature

A. Relation between Rainfall and Temperature

Rainfall and temperature fundamentally influence the formation of regional climates and biological systems. Temperature fluctuations can impact the intensity and distribution of Rainfall, while rainfall patterns can influence temperature through processes such as evaporation and cloud formation. It is crucial to comprehend this dynamic relationship in order to forecast weather events, manage water resources, and increase agricultural output.

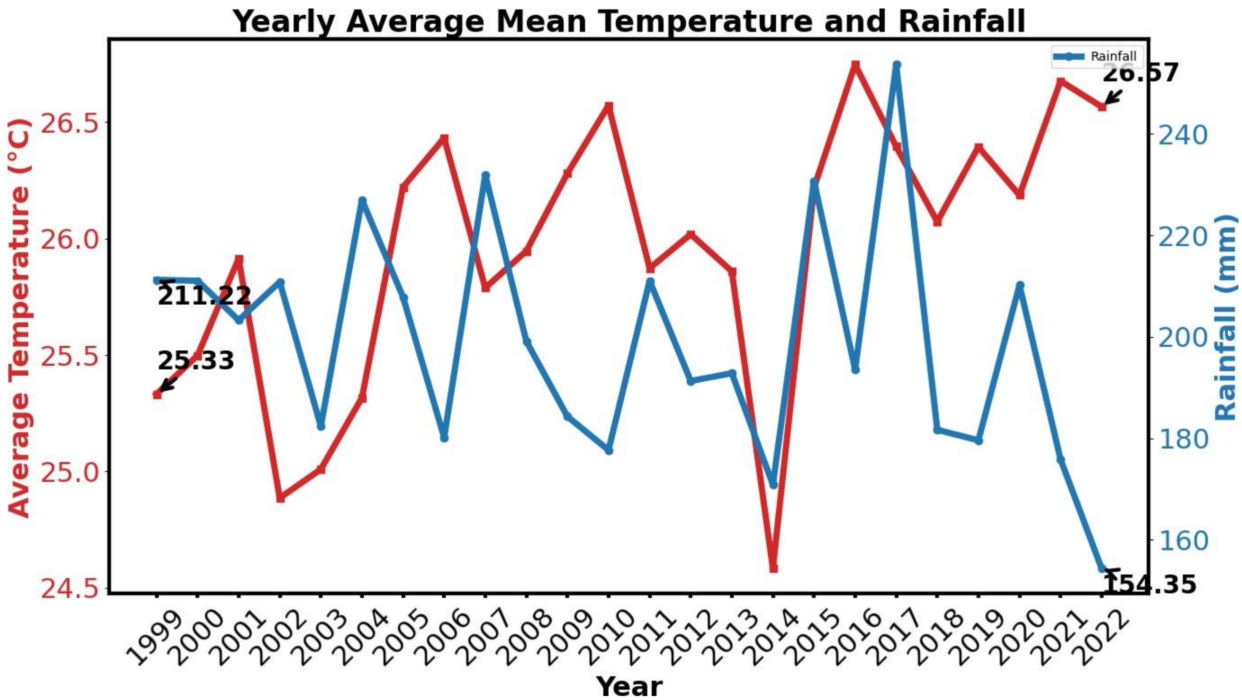


Figure 4.27: Yearly Average Temperature and Rainfall

Figure 4.27 provides insight into the relationship between temperature and Rainfall from 1990 to 2022. Notably, the average temperature remains relatively stable, ranging between 24.5°C and 26.5°C throughout the period, suggesting a consistent climate temperature-wise. In contrast, Rainfall exhibits greater variability, ranging from 150 mm to 250 mm, with peaks and troughs indicating fluctuating precipitation levels. When temperature increases, it often coincides with rainfall decrease and vice versa, hinting at a negative correlation. However, this inverse relationship lacks consistency over the period and warrants further investigation to determine causation. Another notable thing is that from 1999 to 2022, the temperature has been increasing, and Rainfall has decreased a lot. Temperature increases by almost 1.32°C. On the other hand, Rainfall decreases by almost 56.87 mm.

The graph in Figure 4.28 presents a comprehensive overview of the average, maximum, and minimum temperatures alongside monthly rainfall data. Temperature trends depicted by the line graphs likely reveal seasonal variations, with peaks indicating warmer months, possibly summer, and troughs representing cooler periods, possibly winter. Here, the maximum, minimum, and average temperature is lower at the start of the year, as it is the winter season; after that, the temperature gradually increases till June, and Rainfall also increases till August. And in this time from June, temperature Decreases a lot as rainfall increases, because when Rainfall increases, then temperature decreases. After August, Rainfall Decreases and temperature increases for some months till October; then winter comes, and the temperature again decreases. From this graph, we can find a negative correlation between temperature and rainfall.

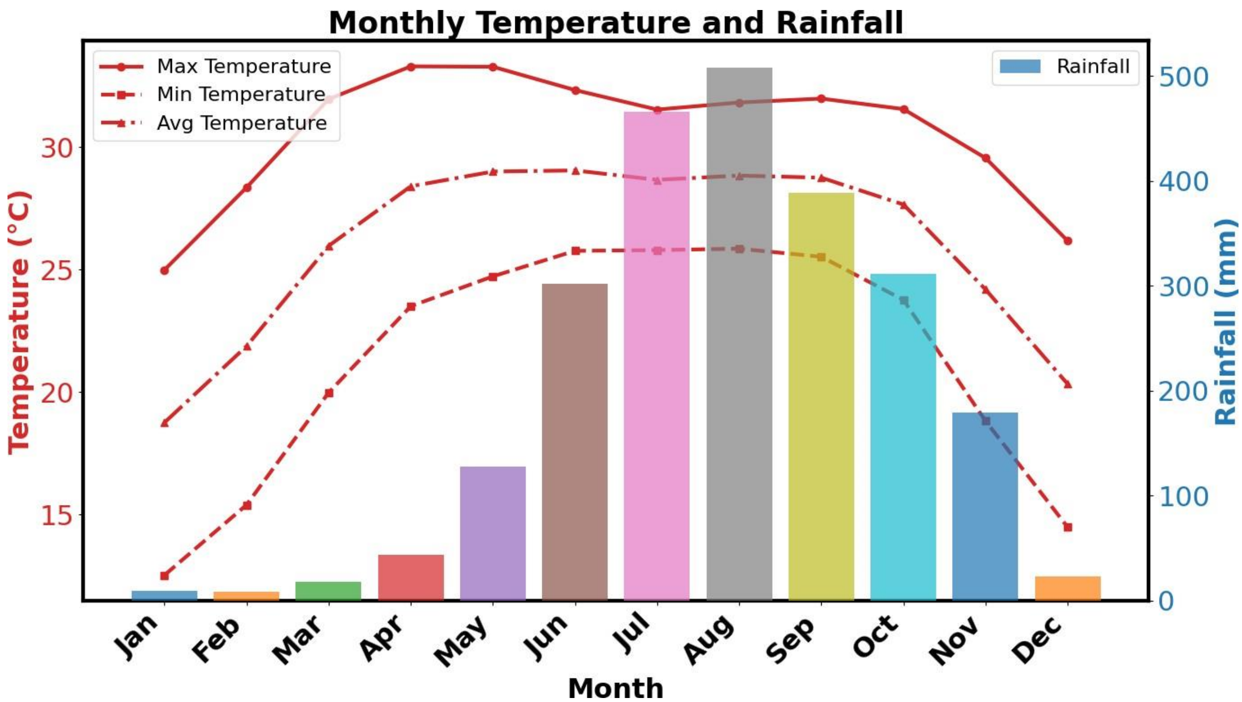


Figure 4.28: Monthly Average Temperature and Rainfall

Concurrently, the bars representing rainfall offer insights into monthly precipitation patterns, with higher bars suggesting rainy seasons and lower bars indicating drier months. Understanding the relationship between temperature and rainfall involves recognizing complex interactions influenced by factors such as evaporation, regional climate patterns like monsoons, local geography, and the impacts of climate change. While temperature and rainfall correlations can vary widely depending on location and climatic conditions, this graph serves as a valuable tool for analyzing climate patterns, agricultural planning, and weather-related decision-making. Further exploration into specific months or patterns could yield deeper insights into local climate dynamics.

Figure 4.29 provides insights into the climatic variations across different seasons. Autumn exhibits a gradual decrease in temperatures from approximately 25°C to 15°C, accompanied by moderate rainfall around 100 mm. Late autumn marks a transition to drier conditions with minimal precipitation. The monsoon season stands out with a sharp increase in rainfall, reaching nearly 400 mm, coinciding with slightly elevated temperatures. This significant rainfall characterizes the Monsoon period, contributing to cooler temperatures due to increased cloud cover. Spring resembles autumn in temperature but experiences lower rainfall. Rainfall signals the onset of increasing temperatures and transitioning toward the wet season.

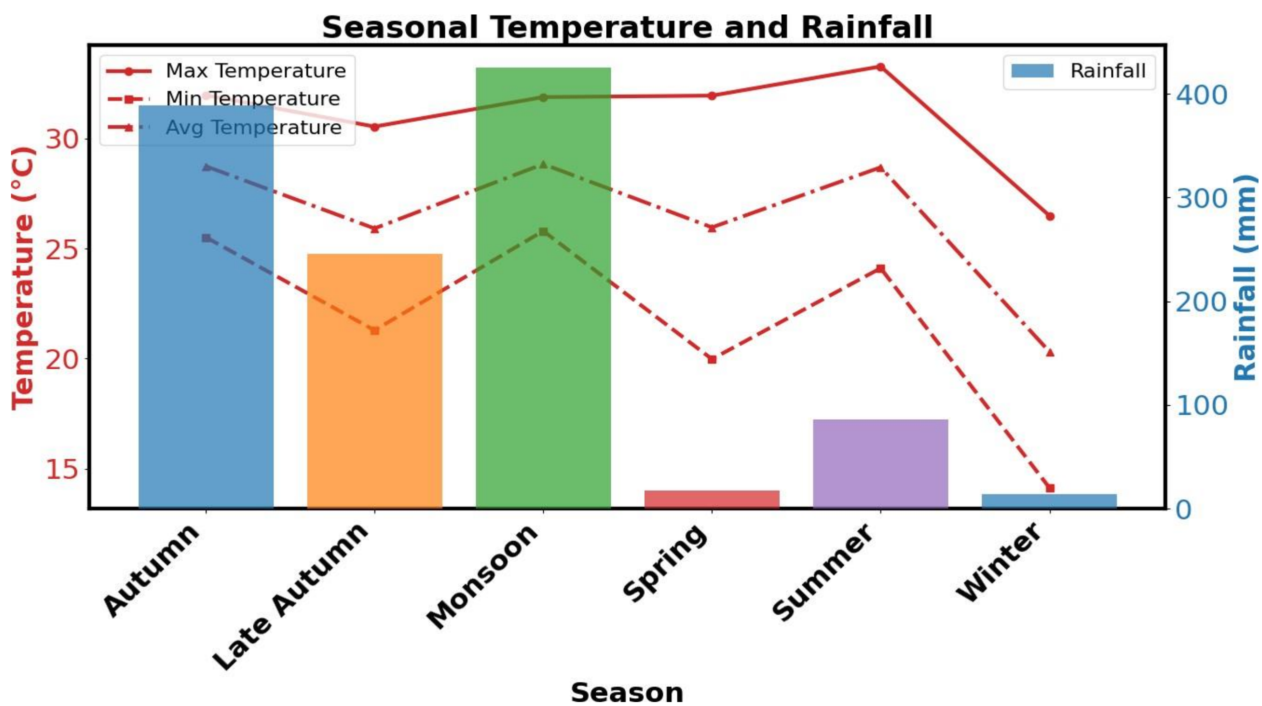


Figure 4.29: Seasonal Average Temperature and Rainfall

Conversely, Summer showcases high maximum temperatures reaching almost 40°C, coupled with moderate rainfall of just under 200 mm. This season is marked by hot temperatures and sporadic rainfall events, albeit less intense than during the Monsoon. Finally, winter experiences a significant drop in minimum temperatures close to 15°C, accompanied by minimal rainfall. This season represents the coldest and driest period, typical for regions with a tropical or subtropical climate. The relationship between temperature and rainfall across these seasons reflects atypical pattern for regions with a monsoonal climate, with pre-monsoon temperature rises preceding heavy monsoon rainfall, followed by cooler post-monsoon temperatures and drier conditions in winter.

B. Relation Between Temperature and Humidity

Understanding the link between humidity and temperature is essential to comprehending atmospheric conditions. The air can contain more moisture when the temperature rises, which raises the humidity level. On the other hand, because of a diminished ability to store moisture, lower temperatures lead to lower humidity. Temperature and humidity interact dynamically to affect human comfort levels, precipitation occurrences, and weather patterns.

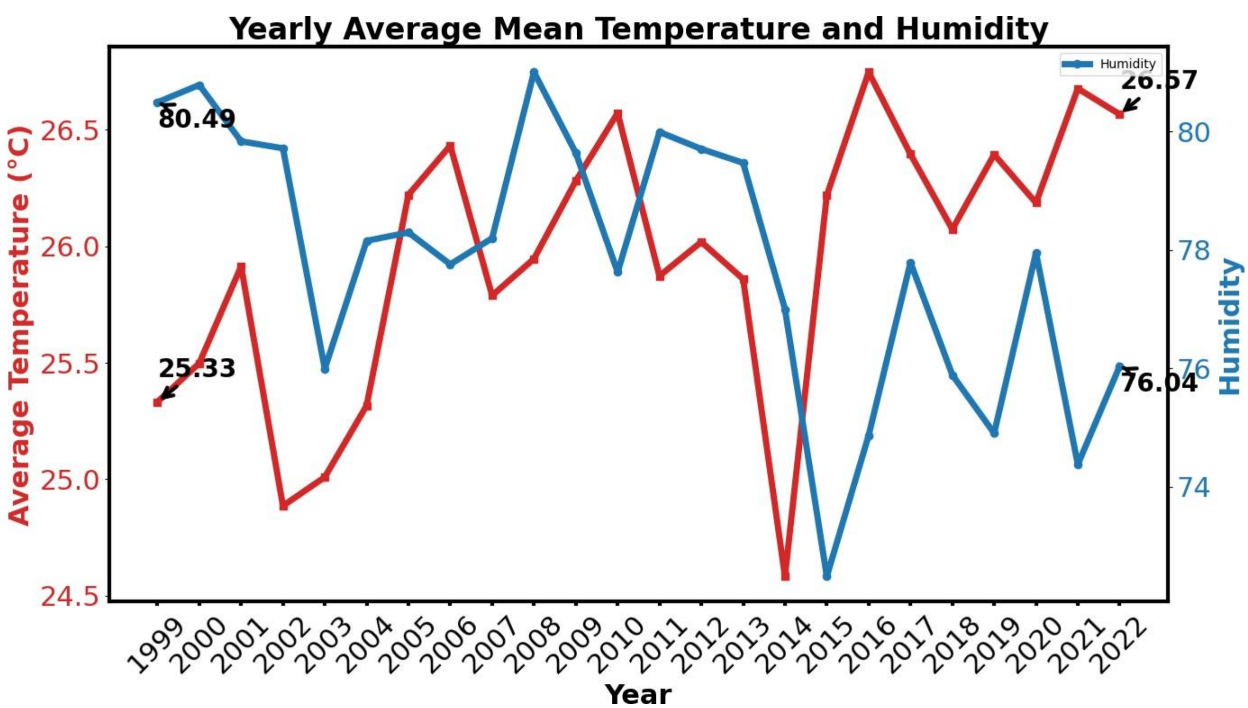


Figure 4.30: Yearly Average Temperature and Humidity

There is a positive relationship between Temperature and Humidity. When temperature increases, humidity in the air also increases because the warm temperature can hold more humidity; on the other hand, when temperature decreases, humidity also decreases because levels of cooler air have a reduced capacity to retain moisture. This dynamic interaction between temperature and humidity influences various weather phenomena, such as the formation of clouds, precipitation, and the sensation of comfort or discomfort experienced by humans.

In graph 4.30, we can see that the humidity decreased a lot from 1999-2022; on the other hand, the temperature increased a lot from 1999-2022. This is a matter of Global warming and is dangerous for the environment. As temperatures rise and humidity levels decrease, several significant impacts on both natural ecosystems and human activities emerge. Higher temperatures can exacerbate heat-related illnesses and stress, impacting human health and increasing energy demands for cooling systems. Concurrently, reduced humidity levels can lead to drier conditions, exacerbating droughts and increasing the risk of wildfires. Furthermore, declining humidity levels may affect agricultural productivity, as crops require adequate moisture for growth. Ecologically, decreased humidity can disrupt ecosystems, affecting plant and animal species reliant on moisture and potentially contributing to habitat degradation and loss. Addressing these impacts requires comprehensive strategies that consider both temperature and humidity dynamics within the context of broader climate change challenges.

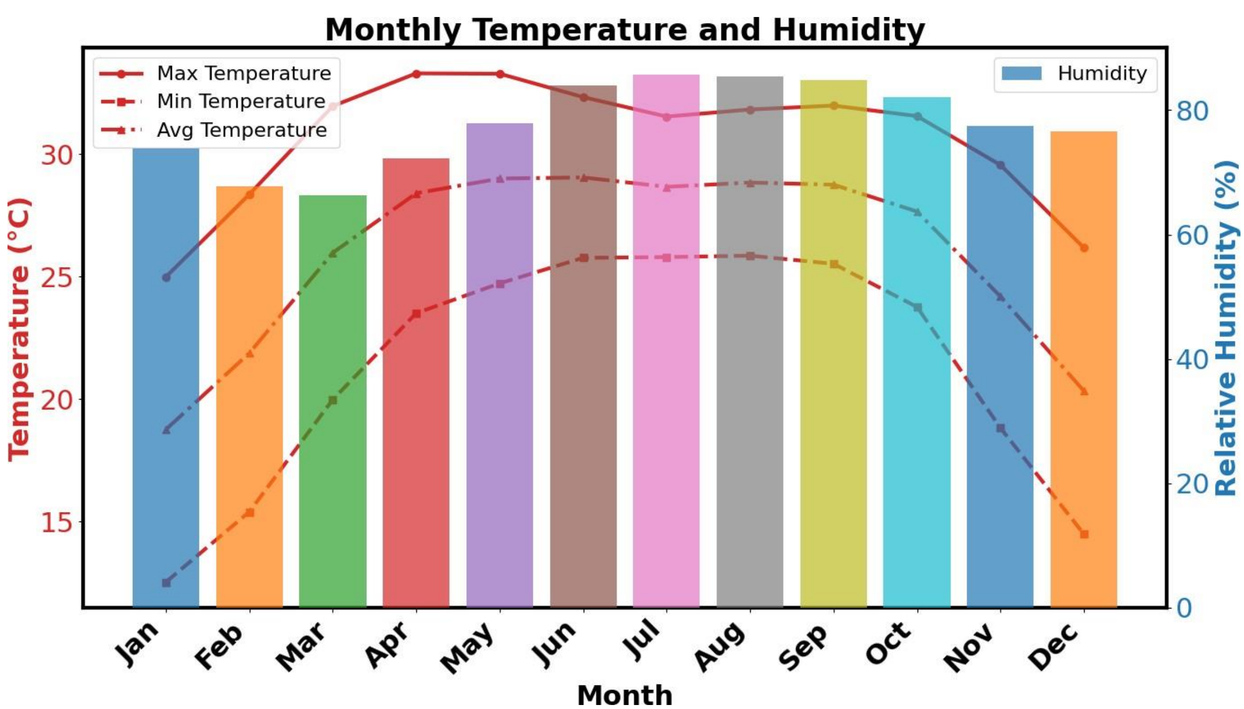


Figure 4.31: Monthly Average Temperature and Humidity

The graph (in Figure 4.31) illustrates monthly temperature and humidity trends, which are crucial for understanding seasonal climatic variations. Temperature data, represented by maximum, minimum, and average lines, exhibit distinct patterns across the year. Maximum temperatures typically peak during mid-year months, reflecting warmer weather, while minimum temperatures dip lowest in early and later months, indicating cooler conditions. Conversely, humidity, depicted by vertical bars, fluctuates monthly, with higher bars indicating more humid periods and shorter bars suggesting drier months. These trends align with seasonal changes influenced by factors such as solar angle, daylight duration, oceanic influence, and air mass movements. Understanding these dynamics is essential for predicting weather patterns and informing decisions in various sectors, including agriculture, infrastructure planning, and public health.

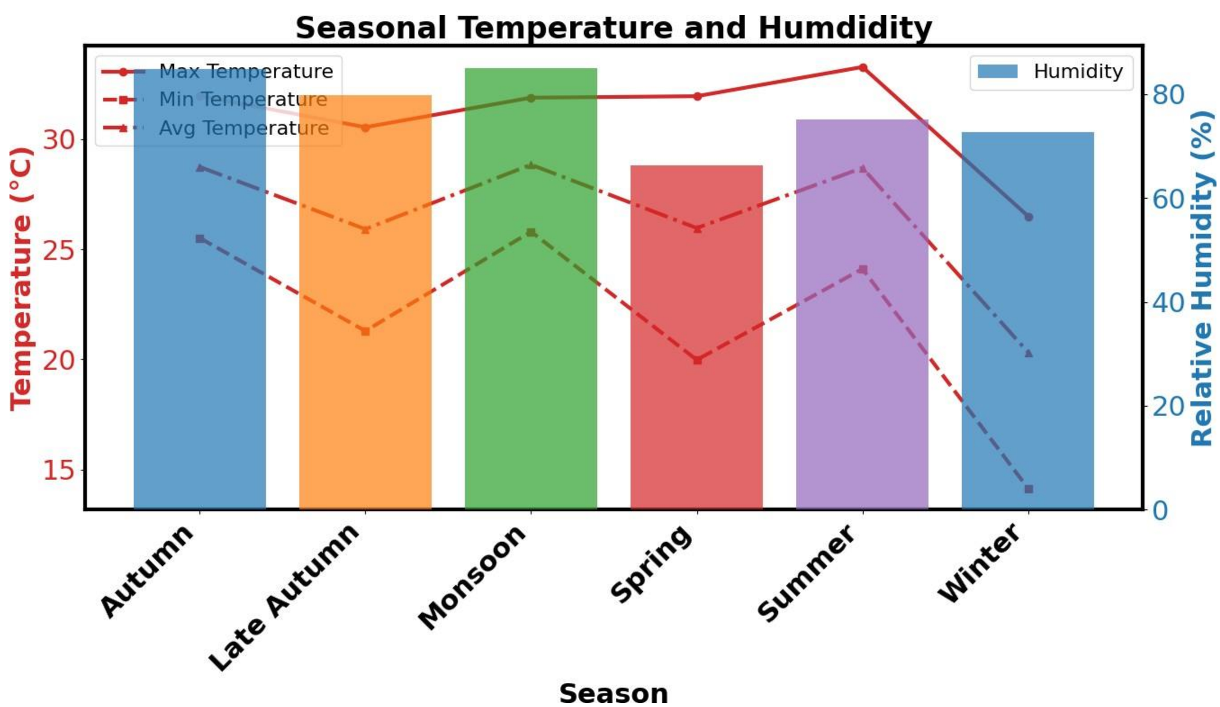


Figure 4.32: Seasonal Average Temperature and Humidity

Figure 4.32 offers a comprehensive overview of temperature and humidity variations across different seasons, providing valuable insights into regional climatic patterns. Approximate values extracted from the graph indicate distinct seasonal trends, with maximum temperatures peaking around 35°C, likely during the summer months, and minimum temperatures dropping to approximately 15°C, presumably in winter. The average temperature fluctuates between 20°C to 30°C throughout the year, reflecting moderate variations. Concurrently, humidity levels range from 40% to 80%, with higher levels observed during the monsoon season, indicative of heavy rainfall.

Seasonal observations reveal dynamic shifts in temperature and humidity, with autumn and late autumn characterized by moderate temperatures and decreasing humidity, while the monsoon season exhibits high humidity levels and heavy rainfall. Spring witnesses increasing temperatures alongside moderate humidity, while summer records the highest temperatures and humidity, particularly before the Monsoon. Conversely, winter experiences the lowest temperatures with reduced humidity levels. This analysis underscores the graph's utility in understanding seasonal climatic dynamics, facilitating informed decision-making in various sectors, including agriculture, infrastructure planning, and public health.

C. Relation between Rainfall and Humidity

Humidity and rainfall are closely intertwined, with humidity serving as a precursor to precipitation events. Higher humidity levels indicate a greater moisture content in the air, increasing the likelihood of rainfall. As the air reaches its saturation point, typically influenced by humidity levels, condensation occurs, leading to the formation of clouds and, ultimately, precipitation in the form of rain, showcasing the direct relationship between humidity and rainfall.

Figure 4.33 provides valuable insights into the relationship between humidity and rainfall Rainfall over time. According to the data, relative humidity started at approximately 80% in 1990, with fluctuations over the years, reaching a peak around 1995, exceeding 80%, and dipping significantly in the early 2000s to just below 60%. Conversely, rainfall began at about 220 mm in 1990 and experienced variations, peaking in the mid-2000s at close to 240 mm and dropping to below 160 mm in the late 2000s. The data concludes that relative humidity is slightly higher than its starting point, at approximately 82%, while rainfall runs lower than its initial value, around 152.35 mm. These trends suggest that while humidity has remained relatively stable with minor increases, rainfall has shown a decreasing pattern over the observed period, potentially indicating shifts in regional climate patterns or environmental factors.

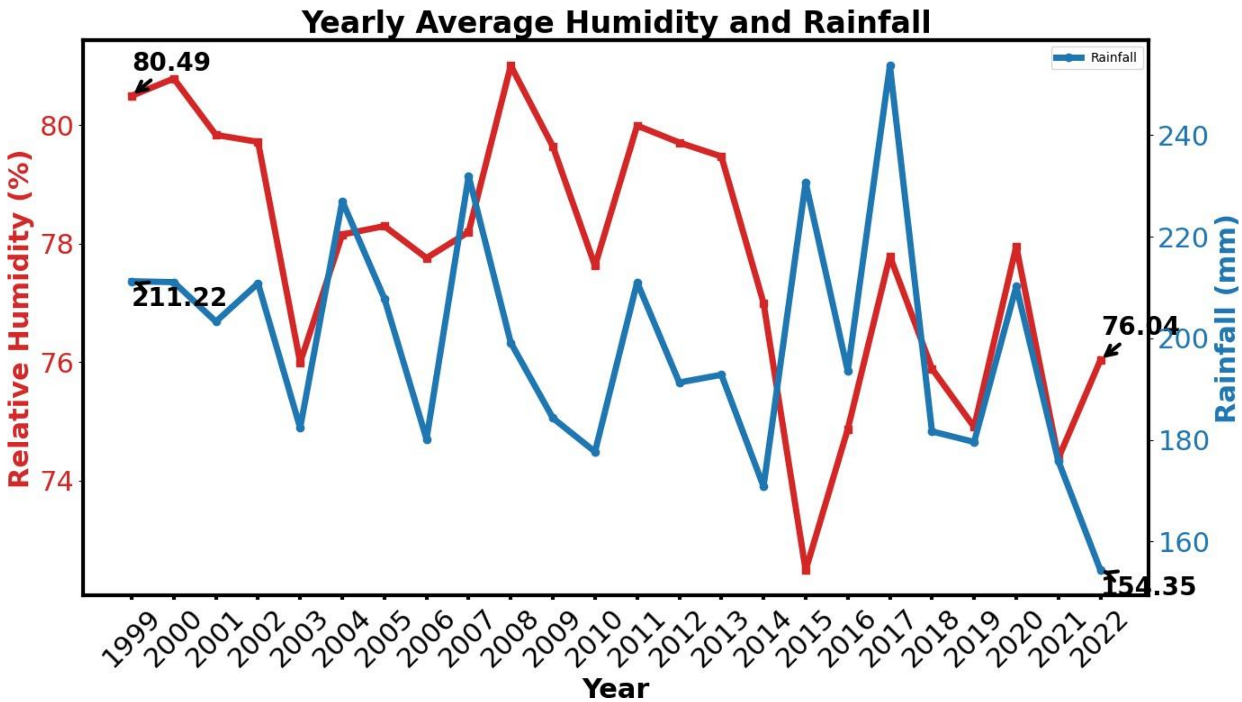


Figure 4.33: Yearly Average Rainfall and Humidity

Humidity and Rainfall are intimately interconnected in the Earth's atmospheric processes. High humidity levels signify an atmosphere rich in water vapor, which is crucial for cloud formation. As the air becomes saturated with moisture, clouds develop, eventually leading to precipitation events such as rain. This relationship underscores the importance of humidity in predicting weather patterns and understanding regional climate dynamics. Moreover, the discomfort associated with high humidity levels in hot weather emphasizes the practical implications of humidity on human comfort and health. Understanding these interrelations is pivotal for various sectors, including meteorology, agriculture, and water resource management, providing insights into the intricate balance of Earth's atmospheric processes.

Figure 4.34 illustrates the relationship between monthly humidity and yearly rainfall values. Rainfall demonstrates notable variations, peaking during typical monsoon months and decreasing during drier periods. In contrast, relative humidity follows a less drastic fluctuation pattern but generally, correlates with higher levels during months with increased rainfall. This correlation underscores the direct relationship between humidity and rainfall. Rainfall occurs when higher humidity levels signify greater water vapor content in the air, which is conducive to cloud formation and subsequent precipitation. Conversely, rainfall contributes to increased humidity levels through evaporation. These observations highlight the significance of understanding humidity and rainfall dynamics for weather forecasting, agricultural planning, and climate pattern comprehension.

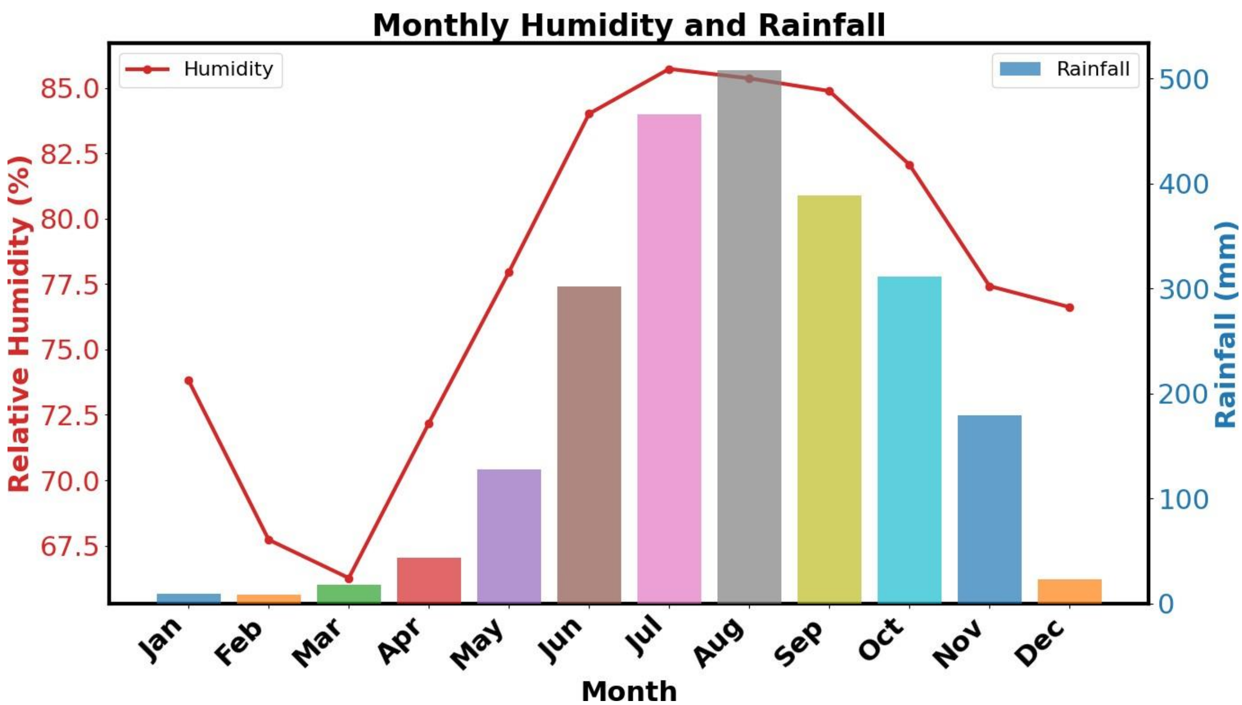


Figure 4.34: Monthly Average Rainfall and Humidity

The bar and line graph in Figure 4.35 illustrate seasonal variations in humidity and rainfall, offering insights into weather patterns across different periods of the year. During autumn, relative humidity hovers around 75%, with rainfall averaging around 100 mm. Late autumn sees an increase in both humidity, nearing 80%, and rainfall, reaching about 150 mm. The monsoon season stands out with notably high humidity of roughly 85% and substantial rainfall approaching 400 mm. Conversely, spring experiences a decline in both humidity, around 78%, and rainfall, dropping nearly to zero. Summer witnesses a dip in humidity below 70%, while rainfall remains slight. Winter showcases a rise in humidity to about 72%, yet rainfall remains minimal.

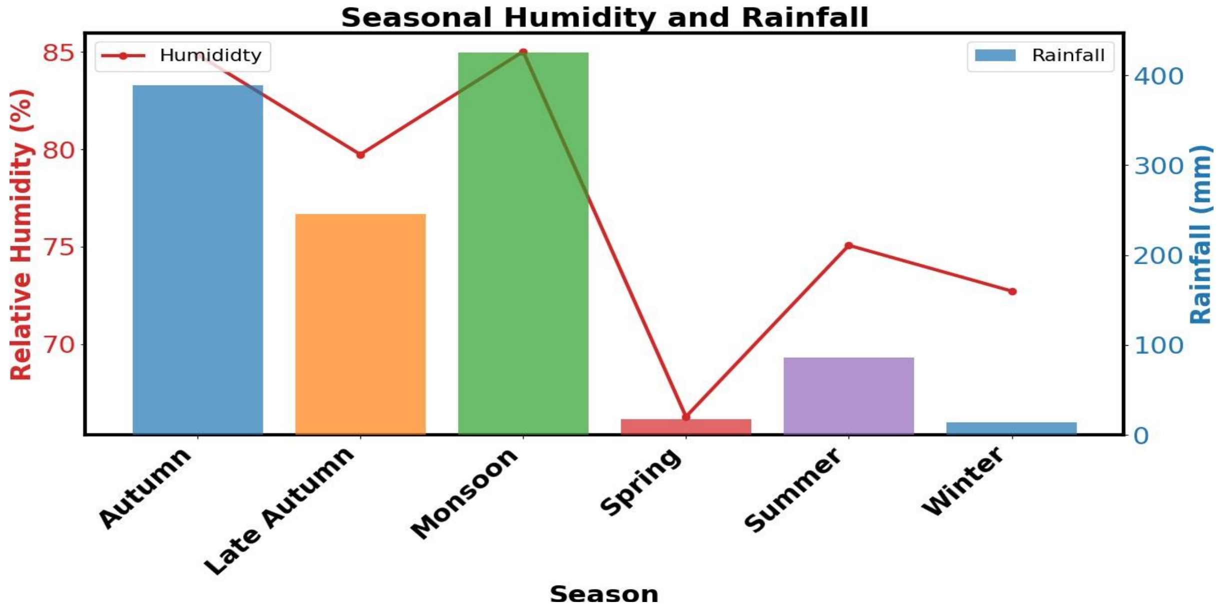


Figure 4.35: Seasonal Average Rainfall and Humidity

The graph highlights the monsoon season as characterized by the highest humidity and rainfall, typical for heavy rains during this period. Conversely, dry seasons such as spring, summer, and winter exhibit lower rainfall, with spring almost devoid of rainfall. The consistent humidity levels throughout the year, with slight fluctuations, underscore the climatic stability of the region. This data holds significance for various sectors, including agriculture, water resource management, and climate studies, aiding in decision-making processes and mitigation strategies for weather-related challenges.

4.3 PERFORMANCE ANALYSIS OF DINTINCT MODELS FOR TEMPERATURE PREDICTION

Temperature is one of the most important element of weather. Here We predict Temperature Using Different Machine Learning Algorithm like as Random Forest, Support Vector machine, Decision tree, K-Nearest Neighbor, Multilayer Neural Network, and Extreme Gradient Boosting. After predicting temperature, we analyze the performance matrices of different models. For performance metrics we used MSE, RMSE, MAE, RAE, R-Square value and R-square Accuracy. We predict temperature in 2 ways, one is Monthly and another one is Daily. Monthly prediction provides us the overall average temperature of a month, on the other hand Daily prediction will provide us theprecise temperature of each Day. In predicting temperature, we used Rainfall, Humidity, wind speed, wind direction, Dew point, Cloud coverage, Sunshine as a feature. After prediction we havebeen used various performance metrics for performance evolution of the model. We predict temperature for both Monthly Basis and Daily Basis. In Monthly basis we divided the datasets into two intervals; one in 1999-2022 another one is 1980-2022. And divided the 1999-2022 into eight Divisions. But in daily basis we used only 1990-2022 intervals and predict only for overall Bangladesh.

4.3.1 Monthly Basis Prediction

Here, we predict monthly average temperature. For that we used all the feature as monthly data. Here we divided our datasets into 2 intervals, first one is from 1999-2022, and second one is from 1980-2022. And the reason behind this interval is to use more geographical location. In interval 1980-2022 there are 26 weather station, on the other hand in interval 1999-2022 there are 34 weather station, and this help us to study more geographical location, and their variation in temperature. And After this prediction we displays the performance metrics of various model in tabular format for both training and testing and after that we also shows the actual vs predicted graph of monthly average temperature for Training and testing result. And Also showed the actual vs predicted graph of testing data for yearly average Temperature. This graph will help us to compare the models performance, their closeness to the actual Value. Which will help us to find the best one. Now, We Will Discuss the prediction results one by one as their intervals and Division.

A. Analysis of Temperature Prediction for: 1999-2022

In this section, we're going to analyze predictions for the time spanning 1999 to 2022. We've divided this time frame into eight Division to make our analysis more manageable and focused. By using advanced prediction methods, we'll try to understand how different trends and events evolved over time. Our goal is to make realistic forecasts for each interval, which can help us makeinformed decisions and plan ahead effectively.

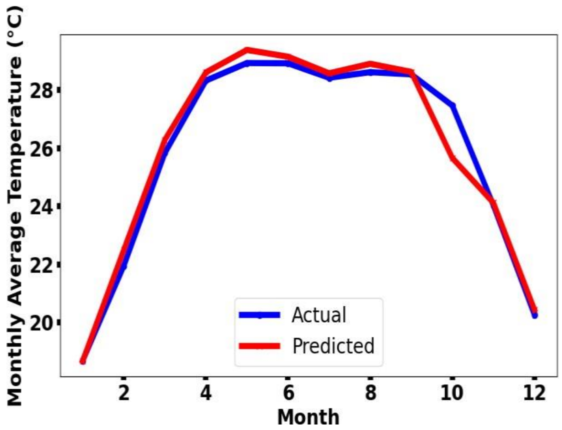
i. For Entire Bangladesh

In this section, we delve into a prediction analysis specifically tailored for Bangladesh, a country with diverse weather patterns and landscapes. With 34 distinct weather stations scattered across the region, we have a rich dataset to draw insights from. By harnessing sophisticated forecasting techniques and leveraging the wealth of historical weather data from these stations, we aim to provide accurate predictions for various meteorological phenomena. Our analysis seeks to offer valuable foresight into the weather dynamics that impact Bangladesh's agriculture, infrastructure,and daily life.

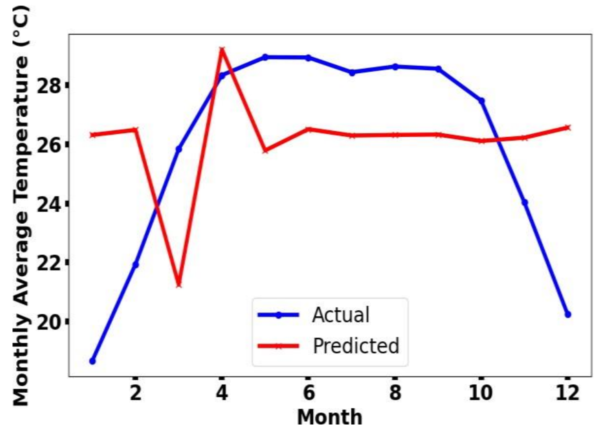
In Figure 4.36 the performance of various machine learning models in predicting temperature values was assessed through an examination of their alignment with actual data. Here we depict the actual vs predicted value of Monthly average temperature of training datasets. Among the models analyzed, the Random Forest model in Figure 4.36(a) demonstrated strong consistency between actual and predicted temperatures, indicating high accuracy and reliability. Conversely, the XG- Boost model in Figure 4.36(f) exhibited noticeable deviations, particularly evident from months 6 to 10, suggesting potential challenges in accuracy during this period. The Support Vector Machine shows in Figure 4.36(b) model displayed significant discrepancies across all months, with a pronounced gap around month 8, indicating comparatively lower predictive accuracy. While the Decision Tree in Figure 4.36(c) model closely matched actual temperature data in earlier months, it showed discrepancies in later months, possibly due to overfitting or a lack of generalization. TheMulti-Layer Neural Network in Figure 4.36(e) model demonstrated moderate alignment with actualdata but may require further tuning for improved accuracy. In contrast, the k-Nearest Neighbors showed Figure 4.36(d) model maintained close alignment across all months, indicating high predictive accuracy and consistency. Overall, the kNN and RF models emerged as the most accurate for temperature prediction, while the SVM model showed the least accuracy. Models like XGBoost, DT, and MLNN exhibited varying degrees of accuracy and weaknesses that could be addressed through optimization and parameter tuning.

In figure 4.37, we depict the actual vs predicted value of Monthly average temperature of testing datasets. The performance of various machine learning models in predicting temperature values was evaluated based on their alignment with actual data. Among these models, the Random Forestmodel based on Figure 4.37(a) demonstrated strong alignment between actual and predicted values, indicating a high level of accuracy with minimal deviation. The Support Vector Machine model

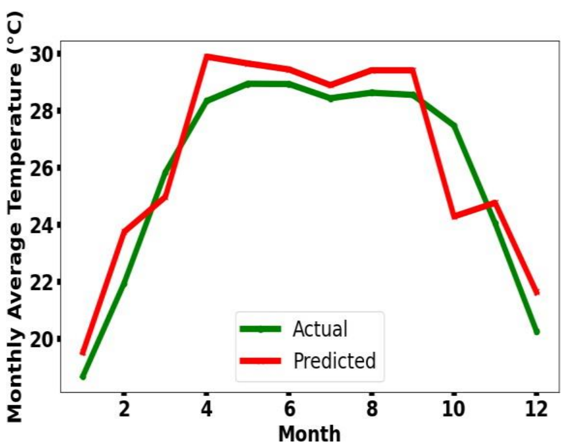
Figure 4.37(b) showed significant discrepancies throughout the year, particularly evident in the middle months, suggesting room for improvement in capturing seasonal trends accurately. The Decision Tree Figure 4.37(c) model exhibited close alignment with actual data in earlier months but showed differences in later months, potentially indicating limitations in long-term predictions.



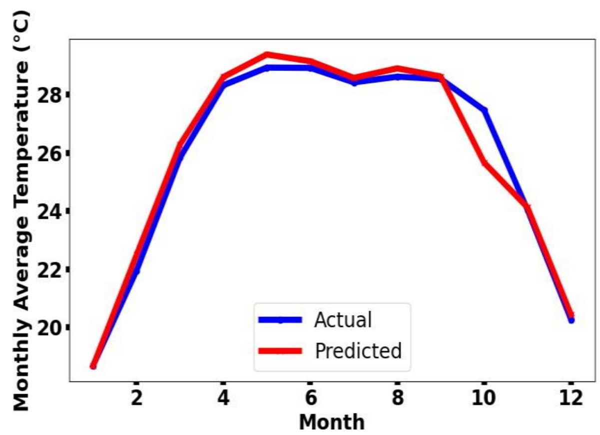
(a) Random Forest



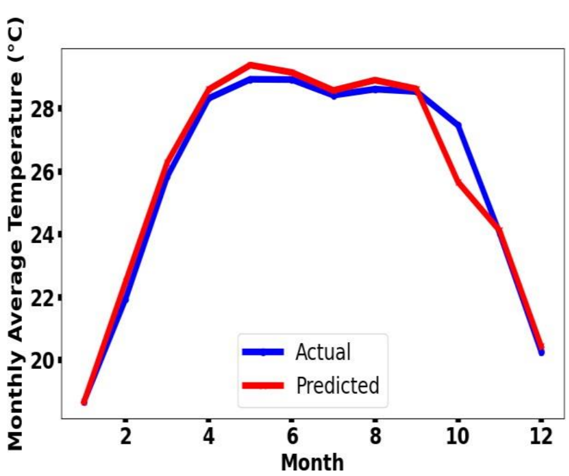
(b) Support Vector Machine



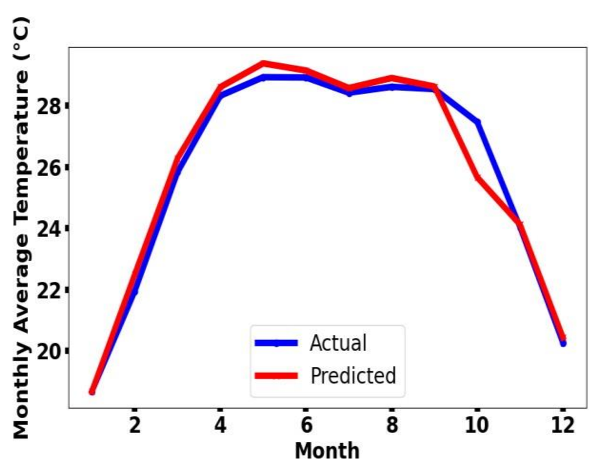
(c) Decision Tree



(d) K-Nearest Neighbor

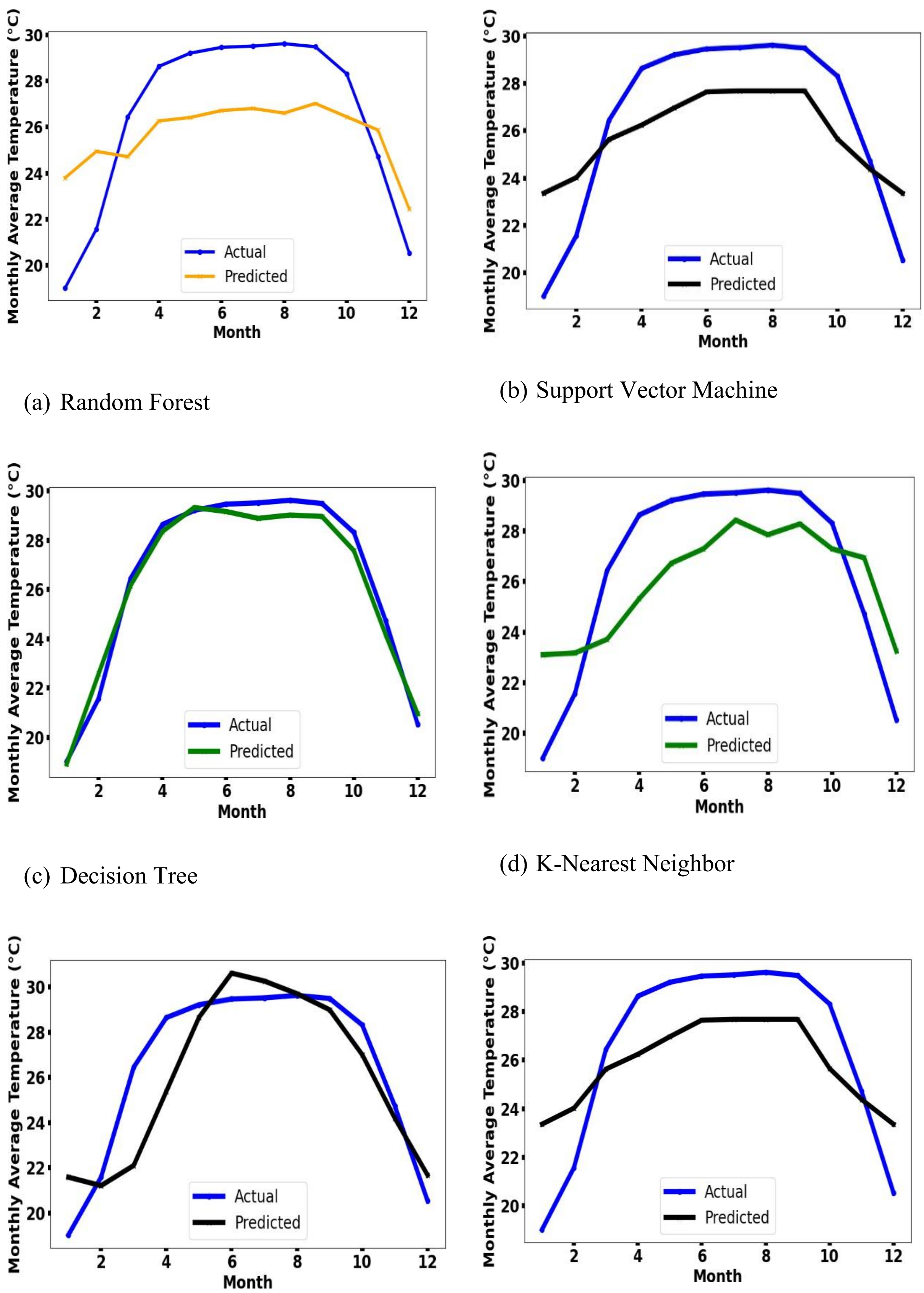


(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.36: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.



(a) Random Forest

(b) Support Vector Machine

(c) Decision Tree

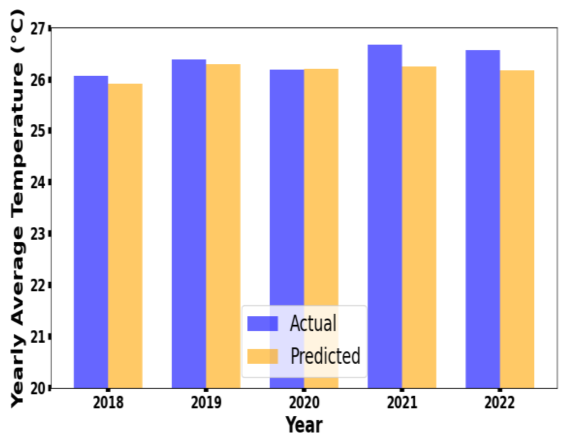
(d) K-Nearest Neighbor

(e) Multilayer Neural Network

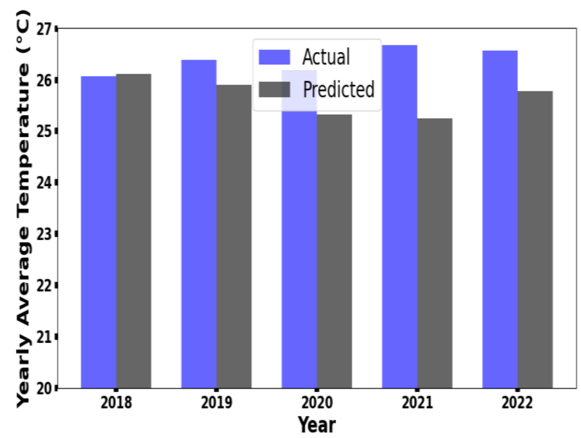
(f) Extreme Gradient Boosting

Figure 4.37: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

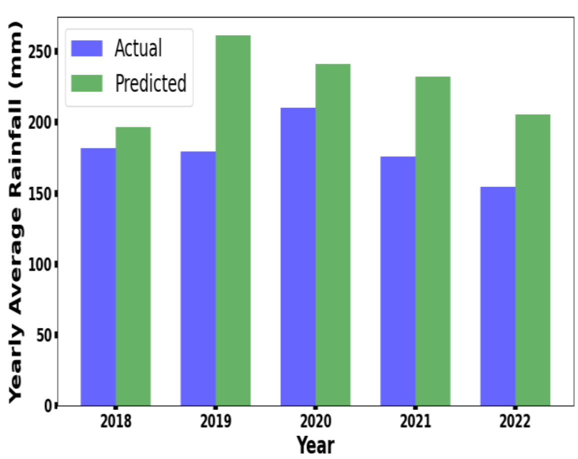
The Multi-Layer Neural Network Model Figure 4.37(e) demonstrated moderate alignment with actual data, suggesting a balanced performance that could benefit from further optimization. The k-Nearest Figure 4.37(d) model maintained close alignment with actual temperatures across all months, indicating high predictive accuracy and consistency. Overall, the kNN and RF models showed the best performance in alignment with actual data, while the SVM model displayed the most room for improvement. Additional performance metrics such as error rates and precision-recall metrics could provide further insights into the models' predictive capabilities.



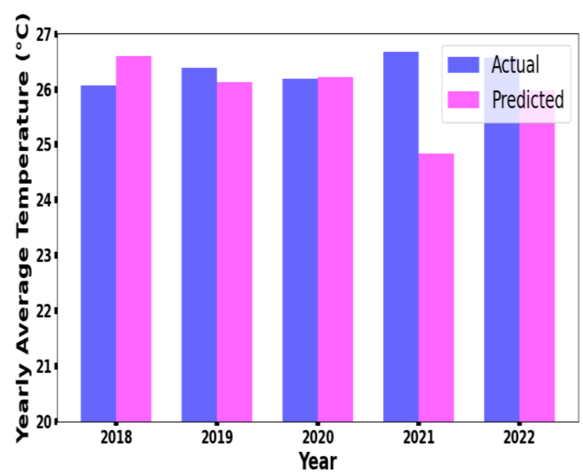
(a) Random Forest



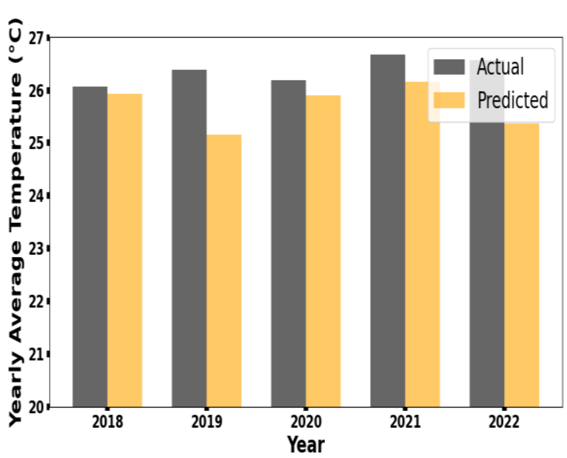
(b) Support Vector Machine



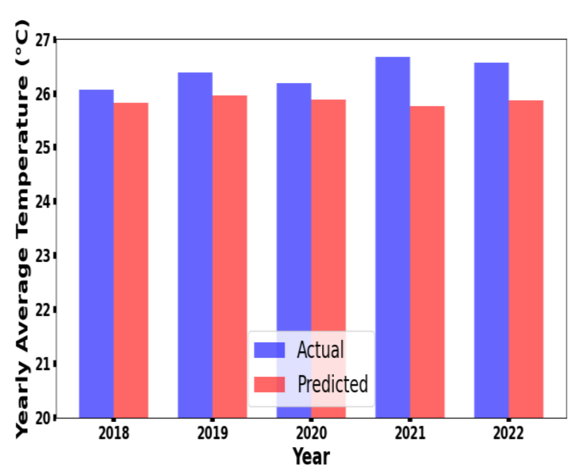
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.38: Comparison of Actual VS Predicted Yearly Temperature Different ML Algorithm Using Testing Dataset.

Figure 4.38 Show a comparative evaluation of six machine learning models for yearly average temperature. The Random Forest in Figure 4.38(a) model demonstrates high accuracy with closely matched predicted and actual values, indicating reliability for yearly temperature prediction tasks. In contrast, the Decision Tree in Figure 4.38(c) model displays greater variation between predicted and actual values compared to RF, suggesting relatively lower accuracy. Support Vector Machine showed in Figure 4.38(b) predictions generally align close to actual data with slight variations, implying decent performance. K-Nearest Neighbors showed in Figure 4.38(d) shows significant discrepancies in some years between predicted and actual values, indicating relatively lower performance.

Conversely, both the Multi-Layer Neural Network Figure 4.38(e)and XGBoost in Figure 4.38(f)) models exhibit close alignment between predicted and actual values, implying high accuracy. Overall, RF, MLNN, and XGB models appear to have performed better, while KNN seems to have underperformed compared to other models.

Table 4.1 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.09	0.99	98.97%
SVM	0.01	0.09	0.08	0.36	0.89	89.17%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.02	0.15	0.11	0.47	0.68	67.86%
MLNN	0.02	0.14	0.11	0.47	0.73	72.94%
XGB	0.03	0.17	0.13	0.57	0.63	62.87%

Table 4.1 Displays the performance metrics for training dataset. The training dataset analysis presents a comparative evaluation of temperature prediction models from 1999 to 2022. Notably, Random Forest stands out with near-perfect scores, boasting an R-Square Value and R-Square Accuracy of 0.99 and 98.97% respectively, accompanied by negligible errors across MSE, RMSE, and MAE metrics. Conversely, Support Vector Machine and K-Nearest Neighbors exhibit moderate performance, with R-Square2 values of 0.89 and 0.68, and R-Square accuracies of 89.17% and 67.86% respectively. Decision Tree showcases perfect training performance, with both R-Square Value and R-Square Accuracy hitting 100%, suggesting potential overfitting. Multi-Layer Neural Network and Extreme Gradient Boosting deliver fair to moderate performance, with R-Square values of 0.73 and 0.63, and R-Square accuracies of 72.94% and 62.87% respectively.

On the other hand, Table 4.2 represents the performance metrics for Testing Datasets. Transitioning to the testing dataset, RF maintains its dominance, with an R-Square Value of 0.96 and R-Square Accuracy of 95.64%, demonstrating consistent accuracy and minimal errors. However, SVM and DT witness significant performance drops, with R-Square values of 0.31 and

0.90, and R-Square accuracies of 30.80% and 90.49% respectively, suggesting potential overfitting issues. KNN remains the weakest performer, with an R-Square Value of 0.37 and R-Square Accuracy of 36.90%. (MLNN and XGB display moderate performance, with R-Square values of 0.63 and 0.60, and R-Square accuracies of 62.50% and 59.65% respectively.

Table 4.2 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.06	0.05	0.20	0.96	95.64%
SVM	0.06	0.24	0.20	0.81	0.31	30.80%
DT	1.38	1.18	0.87	0.26	0.90	90.49%
KNN	0.05	0.22	0.17	0.70	0.37	36.90%
MLNN	0.03	0.17	0.13	0.51	0.63	62.50%
XGB	0.03	0.18	0.16	0.64	0.60	59.65%

In comparing the models' performance, RF emerges as the most consistent and reliable choice across both training and testing datasets. Its ability to maintain high accuracy and minimal errors indicates resilience against overfitting, making it suitable for practical applications. Conversely, DT shows perfect training results but experiences a notable drop in testing performance, suggesting potential overfitting issues. While SVM initially demonstrates moderate performance, its significant decrease in testing accuracy raises concerns about generalization. Thus, for practical deployment, the pragmatic choice leans towards RF due to its consistent performance and robustness against overfitting. However, considerations regarding model complexity and computational resources should also inform the final decision.

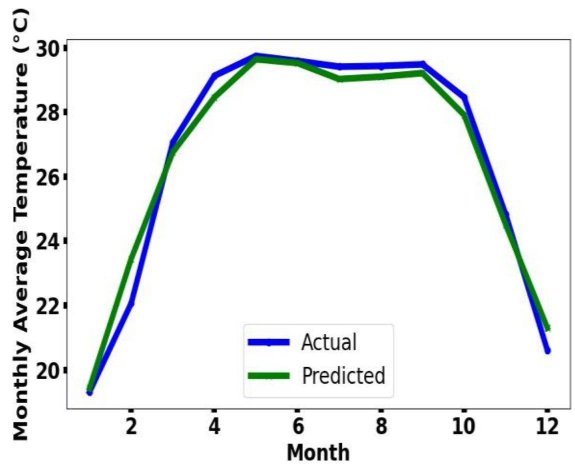
ii. Division-Wise Prediction Analysis

We have divided our datasets into 8 divisions. In each division there are several Weather station. We have predicted rainfall of each division separately, and made an analysis of the performance metrics for each division, and can make a decision which algorithm work better for which datasets or divisions.

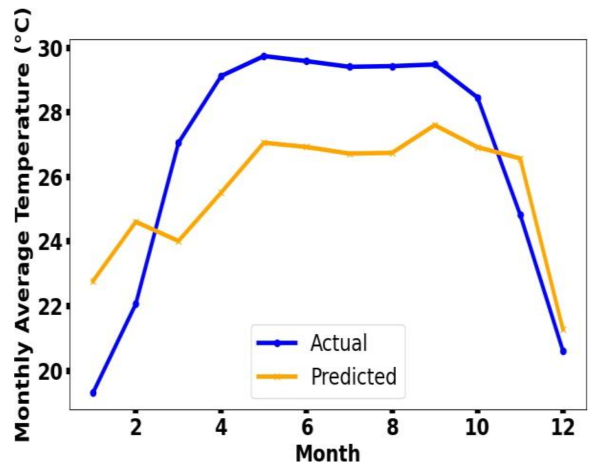
a. For Barisal Division

Here we will discuss about the prediction performance of Barisal Division, in this Division there are 3 Distinct Weather Station. Figure 4.39 displays actual vs predicted graph of various machine learning algorithm using training datasets. From the graph we can notice that The RF in Figure

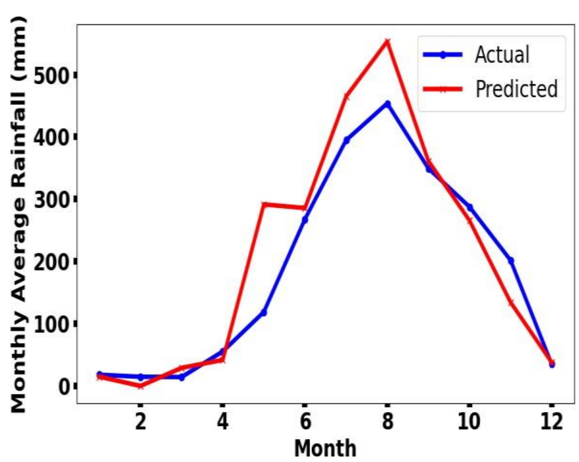
4.39(a), MLNN in Figure 4.39(e), and XGB in Figure 4.39(f) models demonstrate high accuracy, closely matching predicted values with actual data. Conversely, the DT Figure 4.39(c) model exhibits greater discrepancies, potentially indicative of overfitting or inadequate data representation. While the SVM Figure 4.39(b) model performs reasonably well, it falls short of the top performers. KNN Figure 4.39(d) model displays significant variance between predicted and actual values, suggesting limited suitability for this dataset.



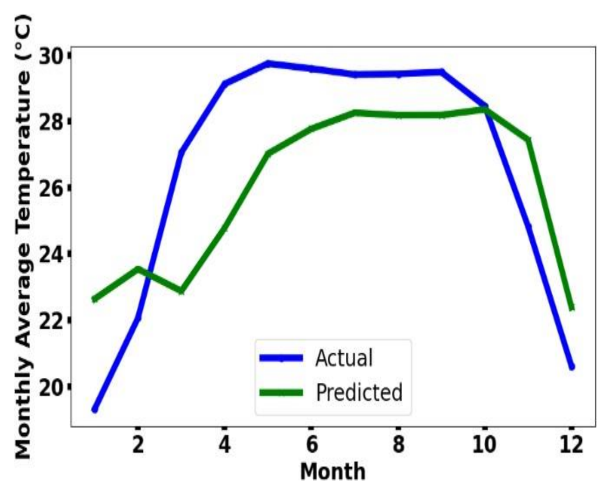
(a) Random Forest



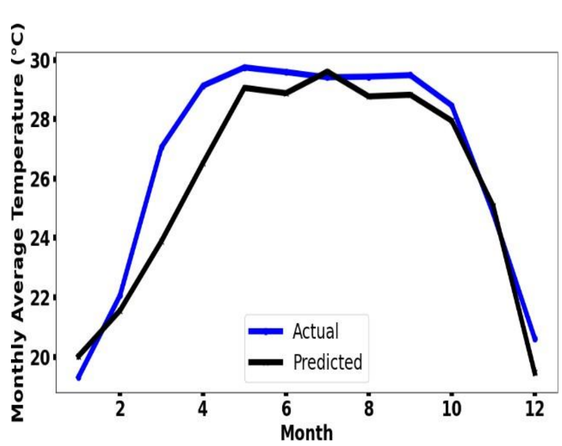
(b) Support Vector Machine



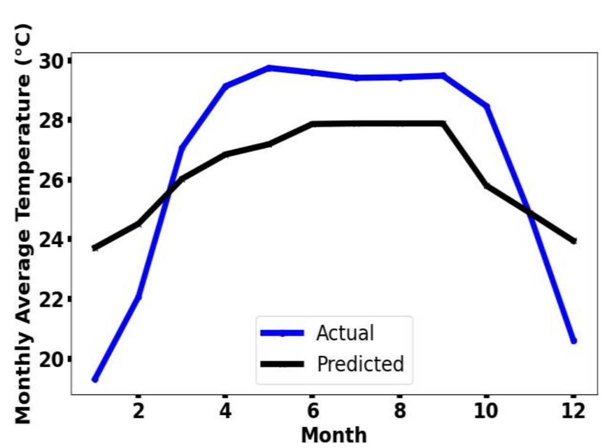
(c) Decision Tree



(d) K-Nearest Neighbor



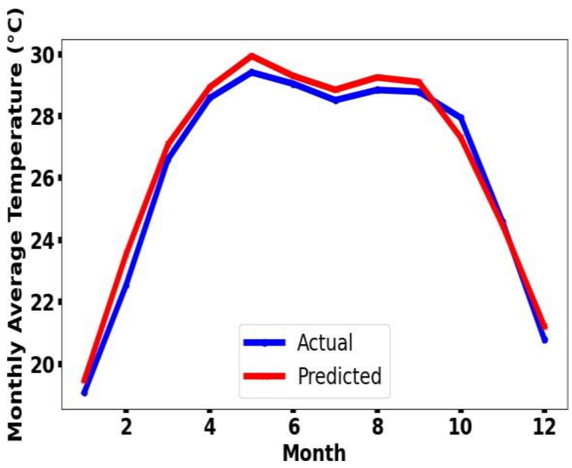
(e) Multilayer Neural Network



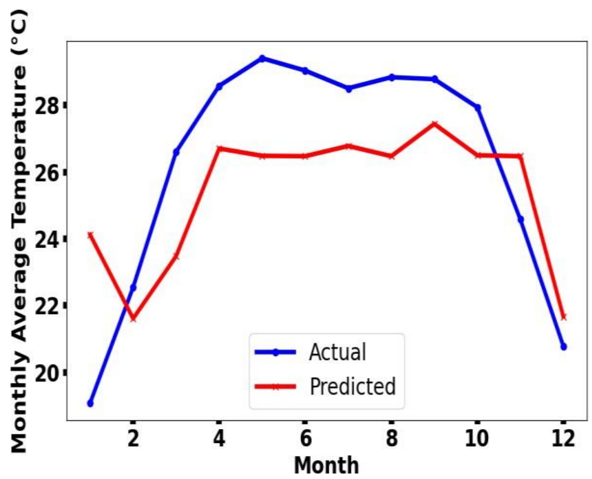
(f) Extreme Gradient Boosting

Figure 4.39: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

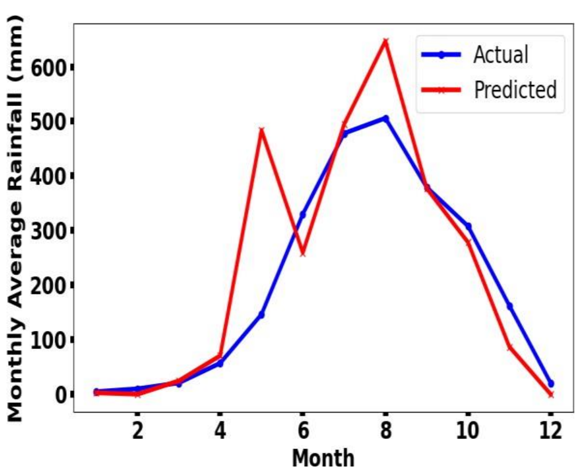
Figure 4.40 Show a comparative analysis of 6 different ML algorithm using testing Datasets. The RF in Figure 4.40(a), MLNN in Figure 4.40(e), and XGB showed in Figure 4.40(f) models demonstrate the closest match between actual and predicted values, suggesting high predictive accuracy.



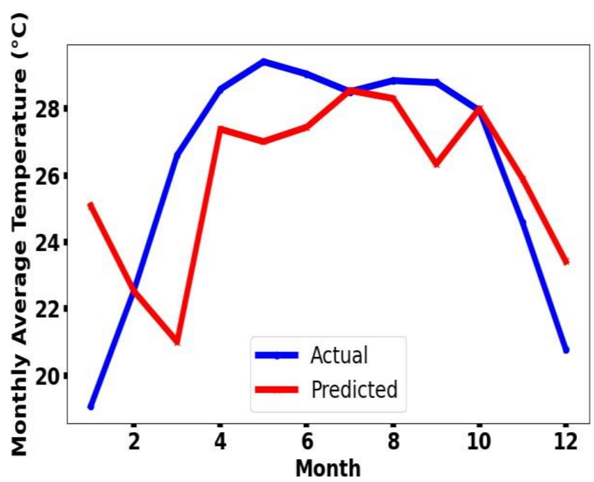
(a) Random Forest



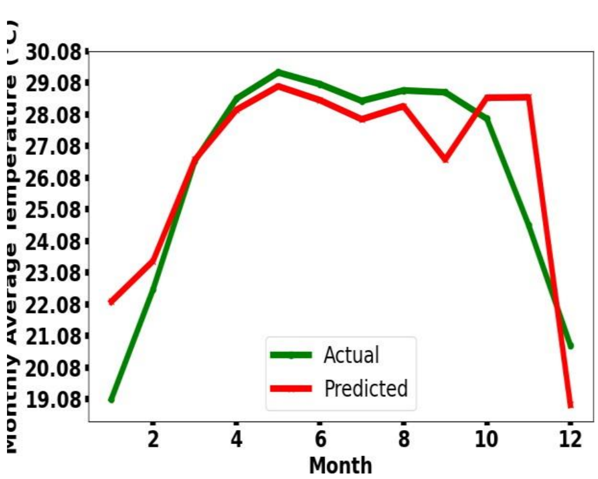
(b) Support Vector Machine



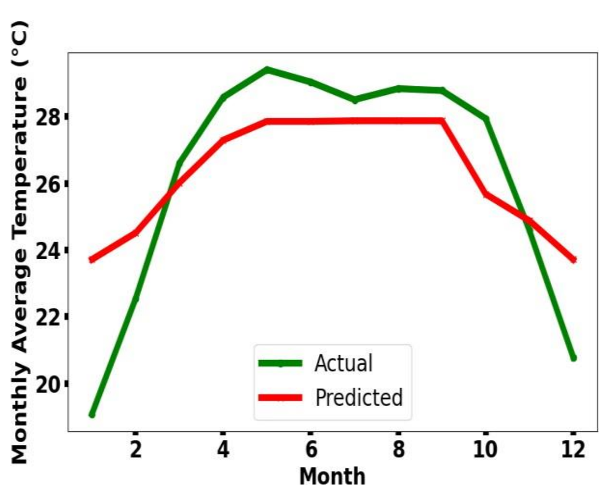
(c) Decision Tree



(d) K-Nearest Neighbor



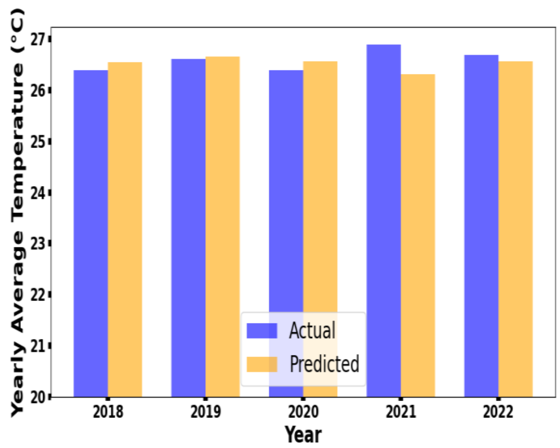
(e) Multilayer Neural Network



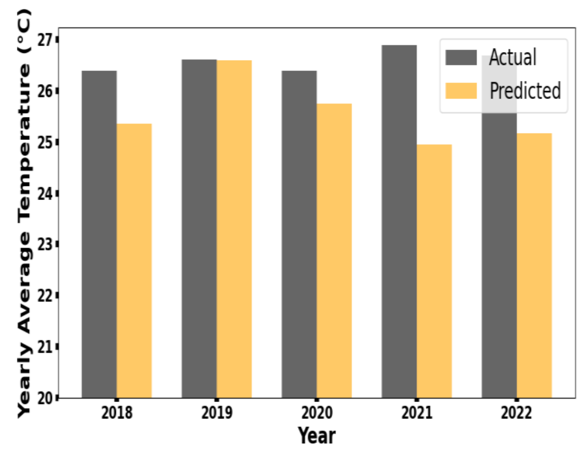
(f) Extreme Gradient Boosting

Figure 4.40: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

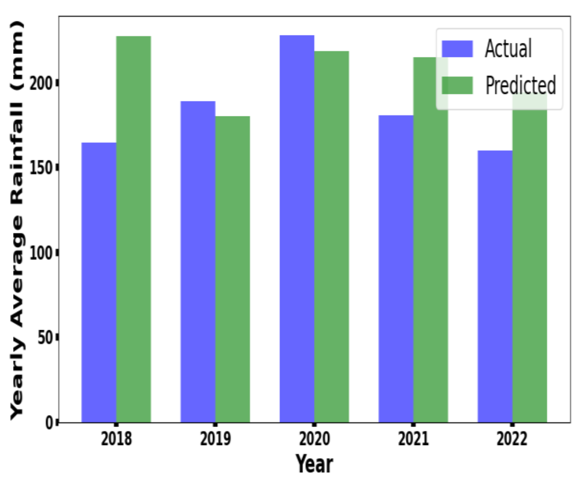
DT Figure 4.40(c) model performs reasonably well but exhibits slight deviations compared to the top-performing models. In contrast, the SVM Figure 4.40(b) and KNN Figure 4.40(d) models show more noticeable discrepancies, indicating potential limitations in their predictive capabilities.



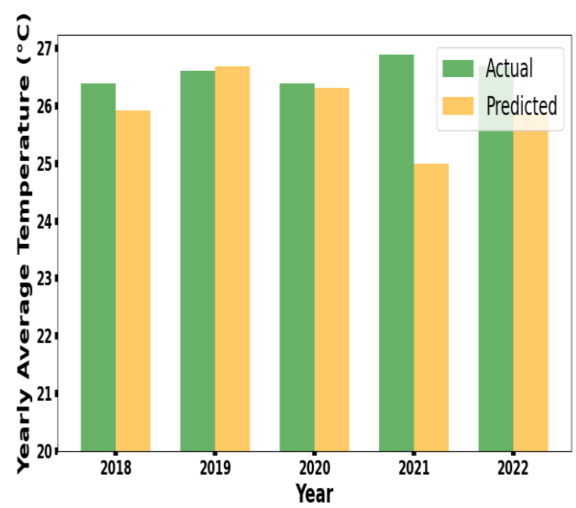
(a) Random Forest



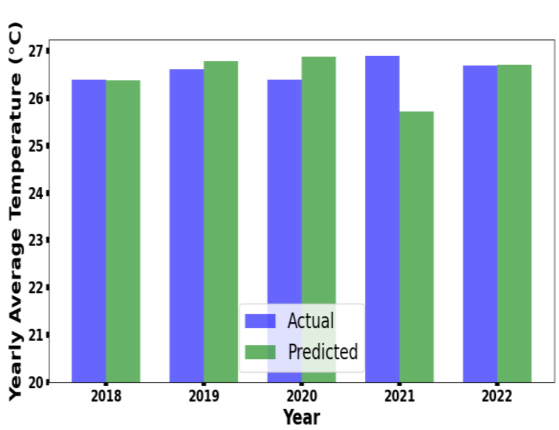
(b) Support Vector Machine



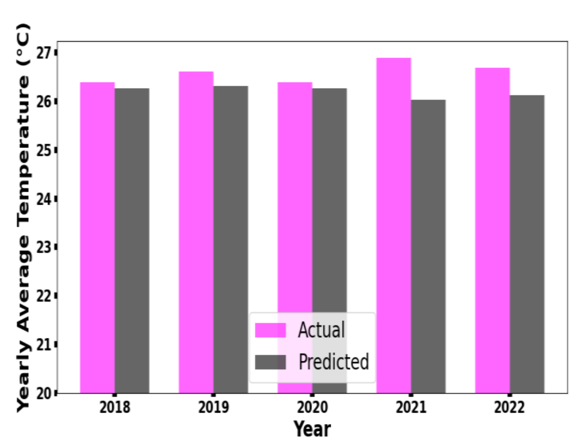
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.41: Comparison of Actual VS Predicted Yearly Temperature Different ML Algorithm Using Testing Dataset.

Overall, XGBoost, MLNN, and RF emerge as the top performers, followed closely by DT, whileSVM and KNN may benefit from further optimization.

Based on the provided figure 4.41, we made a comparative analysis of six machine learning models for predicting yearly temperatures. RF in Figure 4.41(a), MLNN in Figure 4.41(e), and XGBoost in Figure 4.41(f) models exhibit close matches between actual and predicted values, indicating high predictive accuracy. The Decision Tree Figure 4.41(c) model performs reasonably well but with slight deviations compared to RF. Conversely, the Support Vector Machine showed in Figure 4.41(b) and K-Nearest Neighbors demonstrate in Figure 4.41(d) shows more noticeable discrepancies, suggesting potential limitations in their predictive capabilities. Overall, XGBoost, MLNN, and RF emerge as the top performers, followed closely by DT, while SVM and KNN mayrequire further optimization. It's important to note that while this analysis is based on visual inspection, additional statistical analysis would be necessary for a comprehensive evaluation.

Table 4.3 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.08	0.99	98.68%
SVM	0.01	0.09	0.09	0.39	0.88	87.84%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.02	0.15	0.11	0.49	0.66	66.22%
MLNN	0.02	0.13	0.10	0.46	0.75	75.04%
XGB	0.03	0.17	0.13	0.59	0.60	59.61%

Analyzing the performance metrics Table 4.3, it's evident that the Decision Tree model outperforms the others across various metrics. With an MSE and RMSE of 0.00 and 0.00 respectively, DT achieves perfect accuracy, supported by an R2 value and R2 accuracy of 1.00 and 100.00%, indicating the highest level of predictive power and accuracy. Following DT, the Random Forest model also demonstrates strong performance, boasting low error metrics in MSE: 0.00, RMSE: 0.03 and a high R2 value: 0.99, suggesting excellent predictive capability. Conversely, the Support Vector Machine model exhibits relatively higher error metrics MSE: 0.01 and RMSE: 0.09 and a lower R2 value 0.88, indicating comparatively lower accuracy. Similarly, the K-Nearest Neighbors, Multi-Layer Neural Network, and XGBoost models also show higher error metrics and lower R2 values compared to DT and RF, indicating inferior performance in terms of predictive accuracy. Therefore, based on the provided performance metrics, the DT and RF models emerge as the top performers, while SVM, KNN, MLNN, and XGB demonstrate relatively lower performance.

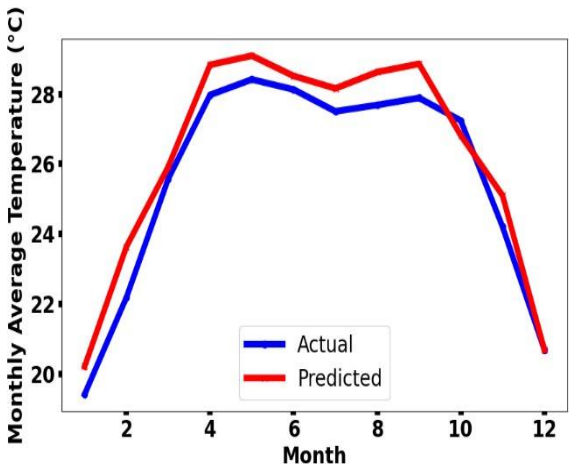
Table 4.4 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.06	0.05	0.20	0.94	94.49%
SVM	0.05	0.22	0.20	0.82	0.33	33.02%
DT	4.47	2.11	1.14	0.35	0.68	68.30%
KNN	0.05	0.23	0.17	0.71	0.30	29.81%
MLNN	0.02	0.14	0.10	0.42	0.75	74.75%
XGB	0.03	0.18	0.15	0.64	0.57	57.20%

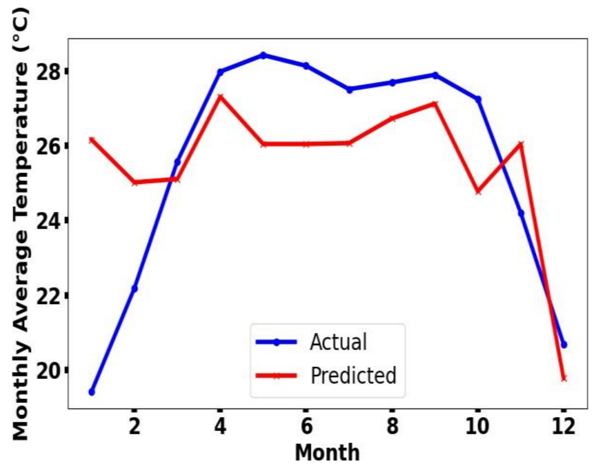
Analyzing the performance metrics Table 4.4, it's clear that the RF and MLNN models outperform the others across various metrics. The RF model exhibits the lowest error metrics with an MSE and RMSE of 0.00 and 0.06 respectively, indicating high accuracy in predictions. Additionally, RF achieves a relatively high R2 value 0.94 and R2 accuracy 94.49%, further confirming its strong performance. Similarly, the MLNN model demonstrates competitive results with low error metrics and a relatively high R2 value 0.75 and R2 accuracy 74.75%. In contrast,the SVM, DT, KNN, and XGB models display higher error metrics and lower R2 values, indicating inferior performance in terms of predictive accuracy. Among these, SVM and KNN exhibit the highest error metrics and the lowest R2 values, suggesting relatively poor performance compared to the other models. Therefore, based on the provided performance metrics, RF and MLNN emerge as the top performers, while SVM, DT, KNN, and XGB demonstrate relatively lower performance.

b. For Chittagong Division

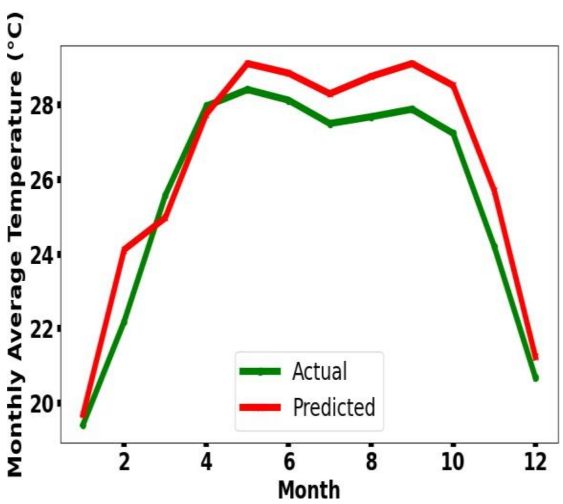
Chittagong Division is located in the south corner of Bangladesh. There Eight weather station in this division named Maijdee court, Ctg Ambagan, Chittagong Airport, Hatiya, Sawndip, Feni Etc. Here we discussed about the performance metrics of Chittagong. The Figure 4.42 shows the prediction graph of Chittagong Divisions. Based on the graph a comparative analysis of six machine learningmodels for predicting monthly temperatures is presented. The Random Forest, Multilayer Neural Network, and XGBoost models demonstrate a close match between actual and predicted values, indicating high predictive accuracy. The Decision Tree model performs reasonably well but exhibits slight deviations compared to RF. Conversely, the Support Vector Machine and K-NearestNeighbors



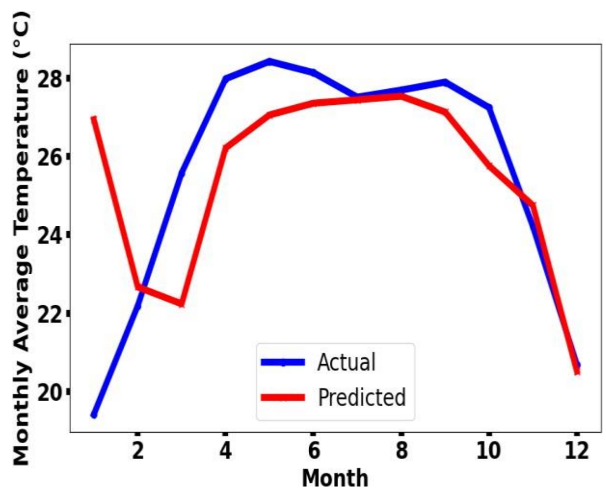
(a) Random Forest



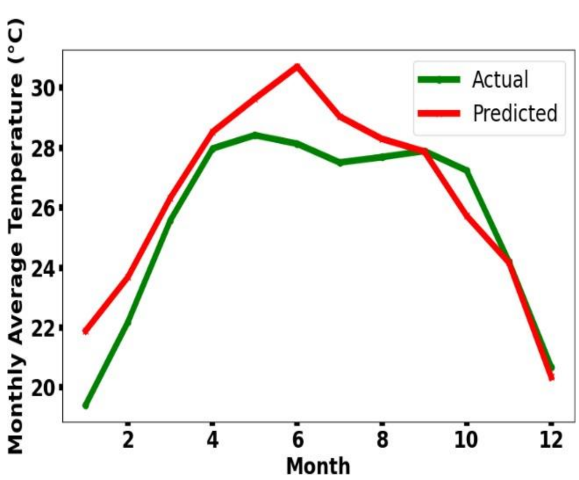
(b) Support Vector Machine



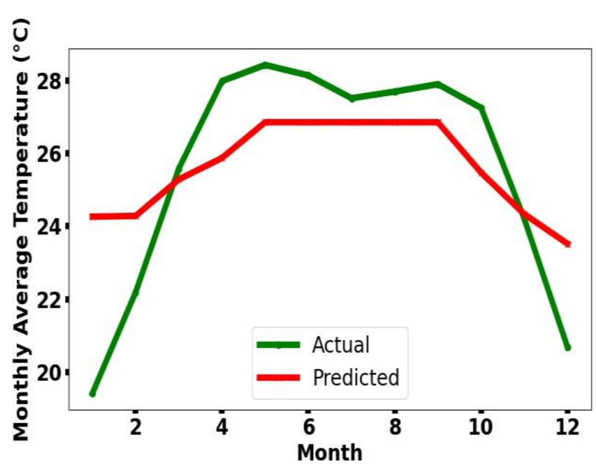
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network

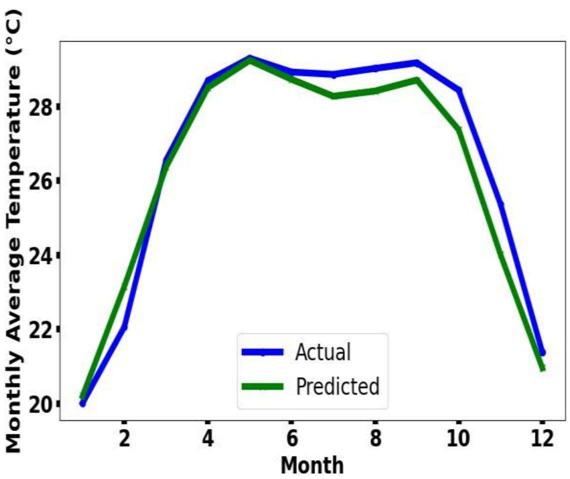


(f) Extreme Gradient Boosting

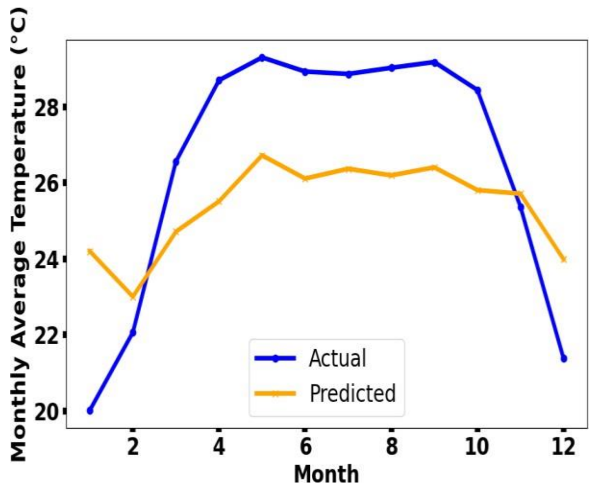
Figure 4.42: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

models display more noticeable discrepancies, suggesting potential limitations in their predictive capabilities. Overall, XGBoost, MLNN, and RF emerge as the top performers, followed closely by DT, while SVM and KNN may require further optimization.

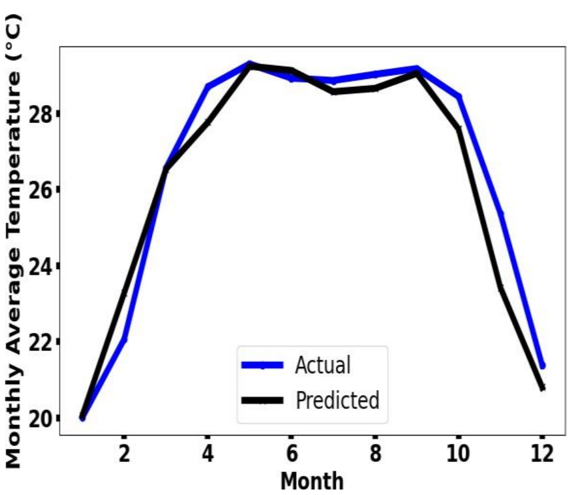
Figure 4.43 shows the comparison of 6 model using testing datasets, from here, we can see that RF, DT, SVM, KNN, MLNN, and XGB—based on their ability to predict monthly temperatures. Each model's performance is assessed through line graphs depicting the alignment between actual and predicted values across a testing dataset.



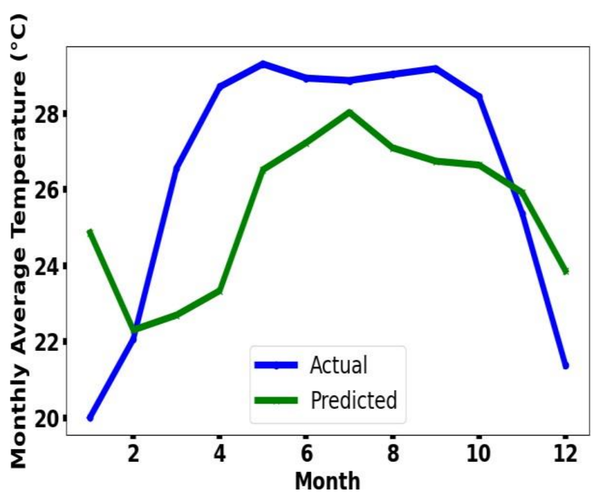
(a) Random Forest



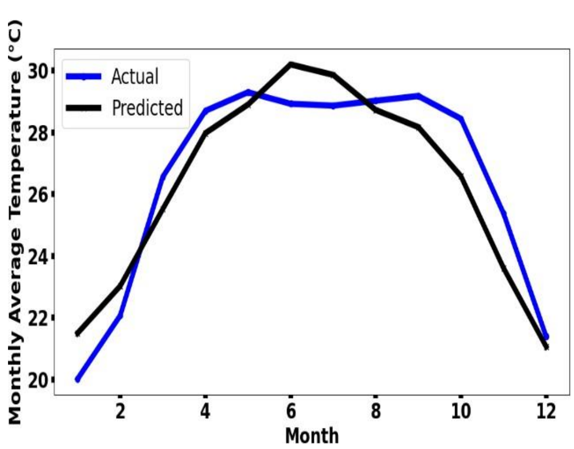
(b) Support Vector Machine



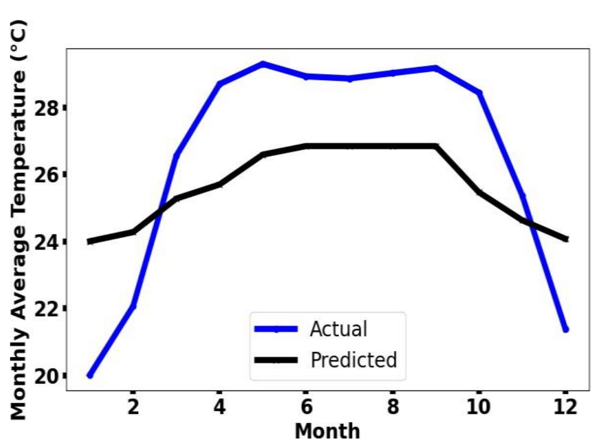
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network

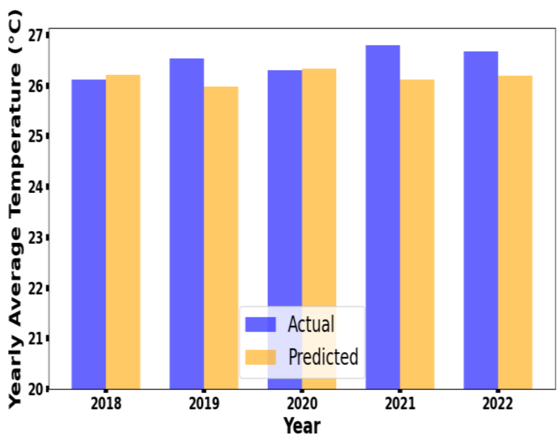


(f) Extreme Gradient Boosting

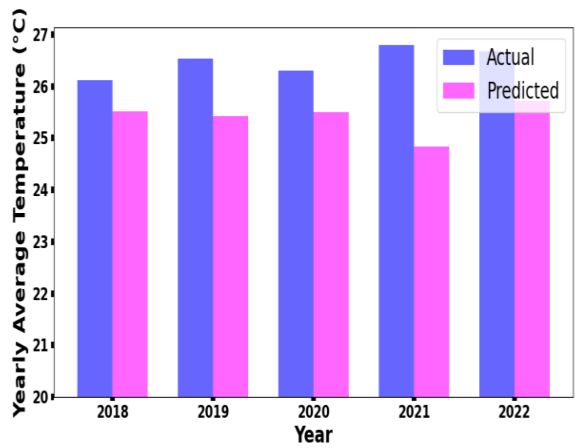
Figure 4.43: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

RF typically performs well due to its ability to reduce overfitting, while DTs might be prone to overfitting but are easy to interpret. SVMs excel with clear separations, KNN's performance depends on optimal K value, MLNNs capture complex patterns, and XgB boasts high performance and speed.

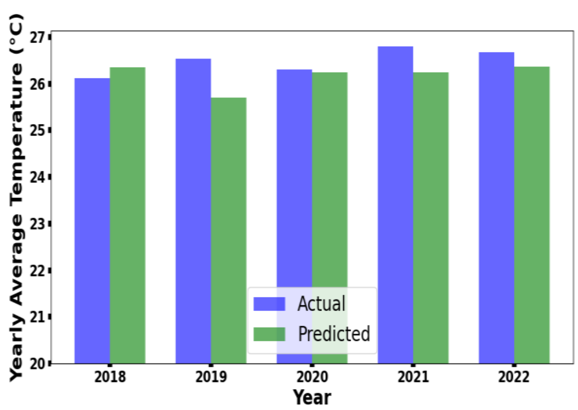
Models showing closer alignment between actual and predicted values are considered to perform better. However, without specific numerical data, a definitive ranking is challenging. The analysis provides a high-level overview, emphasizing the importance of performance metrics for detailed evaluation.



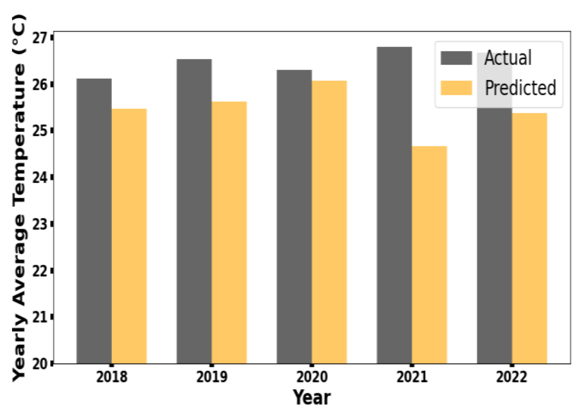
(a) Random Forest



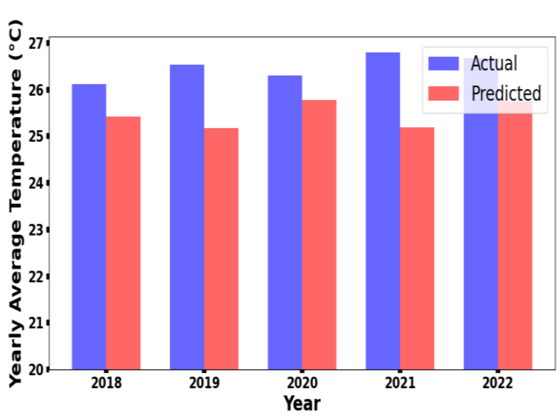
(b) Support Vector Machine



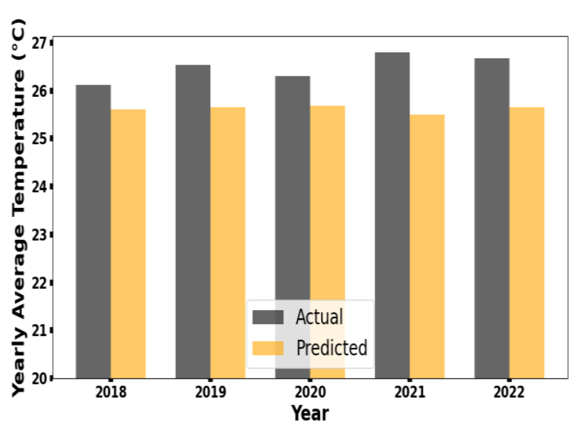
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.44: Comparison of Actual VS Predicted Yearly Temperature Different ML Algorithm Using Testing Dataset.

Figure 4.44 show yearly average temperature of predicted and actual using testing data (2018-2022) Based on graph, a comparative analysis of six machine learning models—Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, MLNN, and XGBoost — based on their performance in predicting yearly temperatures is conducted. RF and XgB are anticipated to perform well due to their ability to handle diverse datasets and high performance, respectively. DTs might face challenges with overfitting, while SVMs excel with clear separations. Discrepancies in KNN graphs may indicate suboptimal performance compared to RF and SVM. MLNNs have the potential to capture complex patterns, contributing to their expected strong performance. For the most accurate assessment, specific performance metrics from the graphs would be essential.

Table 4.5 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.11	0.98	98.45%
SVM	0.01	0.09	0.08	0.40	0.88	87.58%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.03	0.17	0.12	0.59	0.56	55.90%
MLNN	0.01	0.12	0.10	0.45	0.78	78.02%
XGB	0.03	0.17	0.15	0.70	0.52	51.59%

The table 4.5 presents a comparative analysis of six machine learning models based on various performance metrics. Random Forest and Decision Tree models exhibit exceptional performance, achieving low MSE and RMSE values, indicating accurate predictions with minimal errors. Support Vector Machine performs reasonably well but shows slightly higher errors compared to RF and DT. K-Nearest Neighbors and XGBoost models display higher MSE and RMSE values, suggesting less accurate predictions. Multilayer Neural Network performs competitively, with relatively low errors and strong correlation between predicted and actual values. Overall, RF and DT models stand out for their superior accuracy and precision, while SVM, KNN, and XGB models may benefit from further optimization to enhance their predictive capabilities.

On the other hand, Table 4.6 provides a concise analysis of six machine learning models based on their performance metrics on the testing dataset. Random Forest and Multilayer Neural Network models exhibit the lowest MSE, RMSE, and MAE values, indicating accurate predictions with minimal errors. Decision Tree performs moderately well, with higher error metrics compared to RF and MLNN. Support Vector Machine, K-Nearest Neighbors, and

XGBoost models show higher error values and lower R2 accuracy, suggesting less accurate predictions compared to RF and MLNN. Overall, RF and MLNN models demonstrate superior performance on the testing dataset, while SVM, KNN, and XGB models may require further optimization to enhance their predictive capabilities.

Table 4.6 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.08	0.06	0.27	0.91	91.26%
SVM	0.05	0.23	0.20	0.88	0.22	21.90%
DT	1.85	1.36	0.96	0.33	0.83	83.47%
KNN	0.06	0.25	0.19	0.87	0.03	3.37%
MLNN	0.01	0.11	0.09	0.41	0.80	80.48%
XGB	0.04	0.20	0.18	0.81	0.41	41.16%

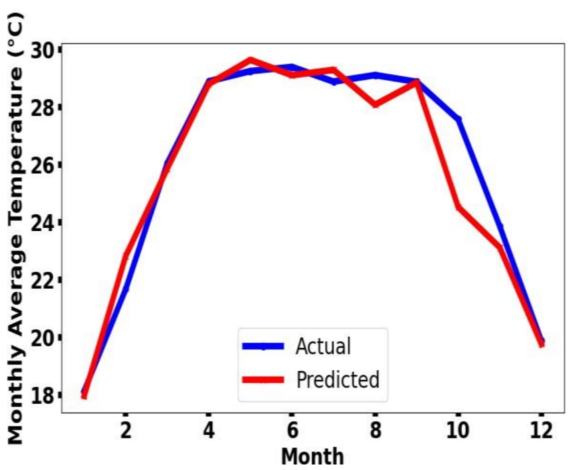
c. For Dhaka Division

In Dhaka Division there are 4 distinct weather Station. And here we predict Temperature of This division using different AI model. And make comparison between them.

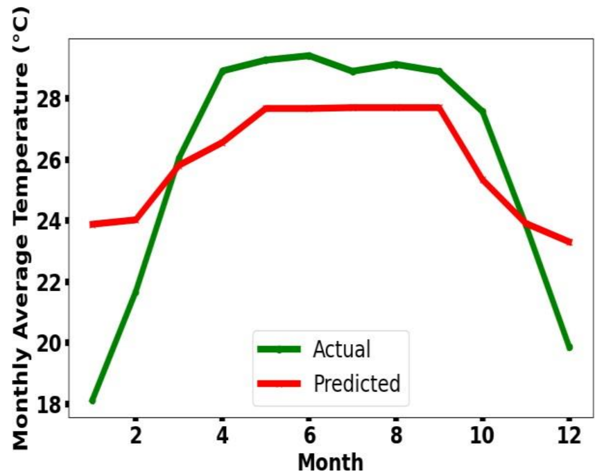
In figure 4.45 we see a comparative analysis of six machine learning models—Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, Multilayer Neural Network, and XGBoost —reveals varying levels of performance in predicting temperature over time. RF and XGB demonstrate the closest alignment between actual and predicted values, suggesting strong predictive accuracy. DT and SVM show some discrepancies but still perform reasonably well. MLNN also exhibits a good match between actual and predicted values. However, KNN displays significant variance, indicating potentially lower performance compared to other models. Overall, XGB appears to have the best performance among the models presented, while KNN may require further optimization.

After visualize the training result we also made the graph of actual vs predicted temperature using testing datasets, which give us an idea of how the model work with testing datasets. In the line graphs of figure 4.46, we compare the actual vs. predicted values for six machine learning models: Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, Multi-Layer Neural Network, and Extreme Gradient Boosting. These graphs evaluate the performance of the models in predicting monthly temperatures, with RF, MLNN, and possibly XgB showing closer alignment between actual and predicted values, suggesting better performance. DT and KNN exhibit more variance, indicating potentially lower accuracy.

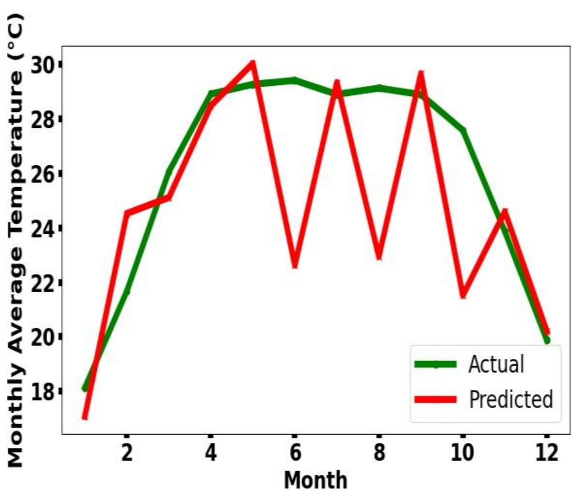
The figure 4.47 comprises six bar graphs comparing actual vs. predicted yearly averages for machine learning models: RF, DT, SVM, KNN, MLNN, and XgB. RF and MLNN demonstrate closer alignment, indicating better performance, while, MLNN, KNN shows more variance.



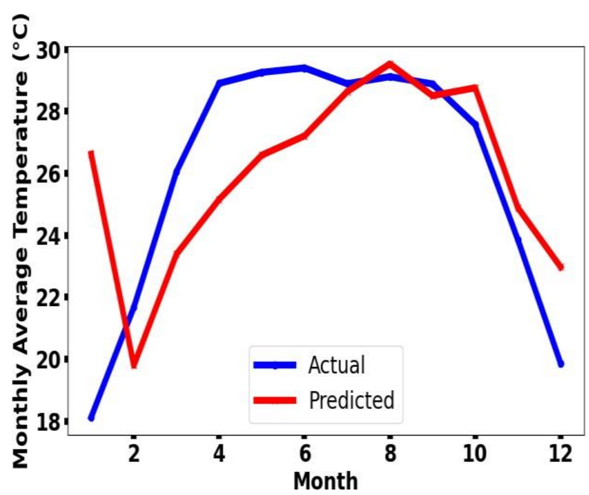
(a) Random Forest



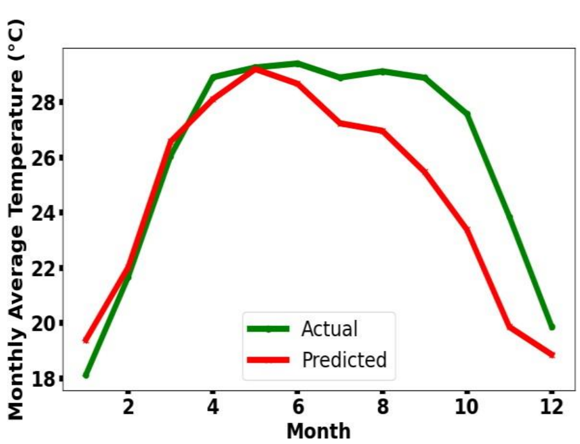
(b) Support Vector Machine



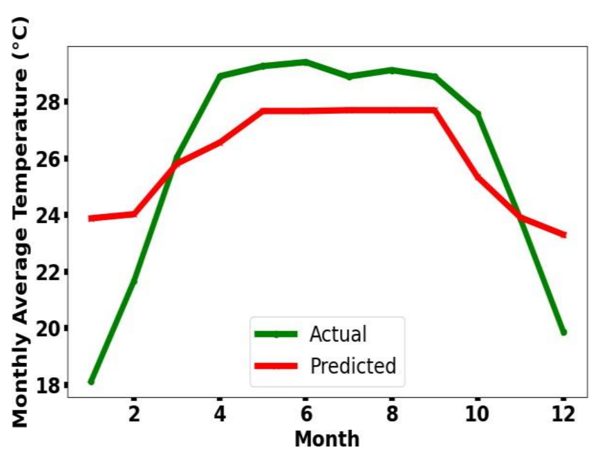
(c) Decision Tree



(d) K-Nearest Neighbor



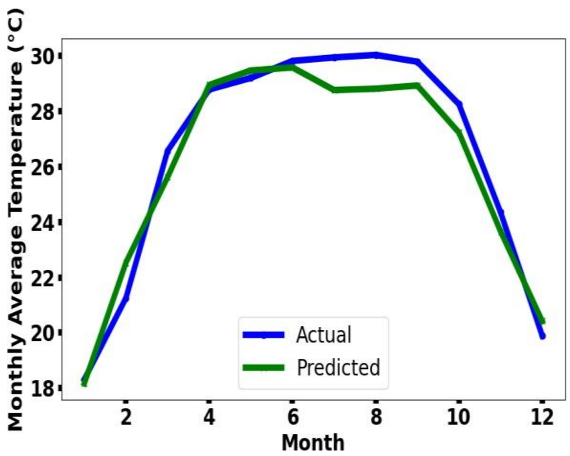
(e) Multilayer Neural Network



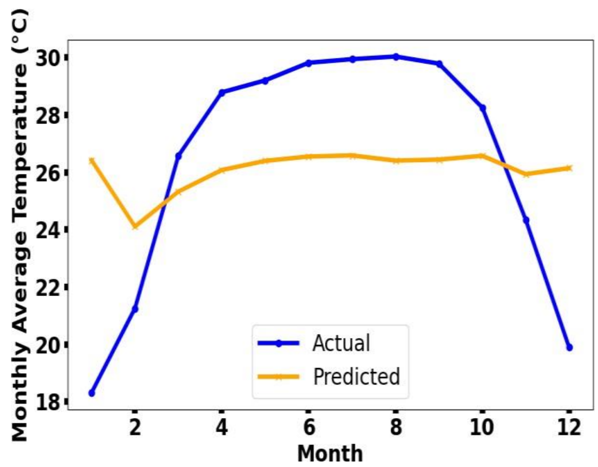
(f) Extreme Gradient Boosting

Figure 4.45: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

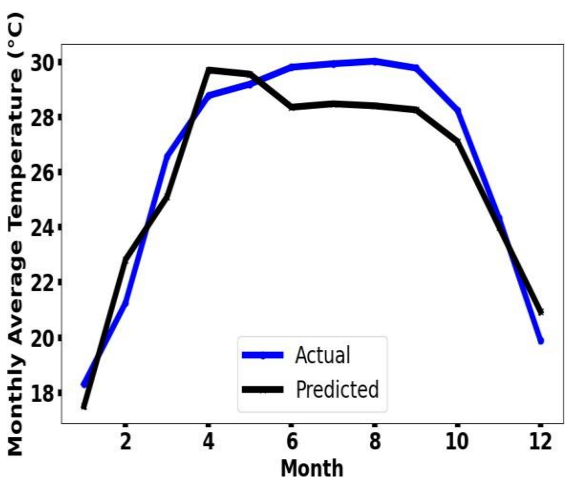
From Table 4.7 a concise analysis of the performance of six machine learning models on the training dataset can be made. Random Forest, Decision Tree, and Multi-Layer Neural Network models exhibit excellent performance, with low MSE, RMSE, and MAE values, indicating accurate predictions and high R2 values and R2 accuracy. Support Vector Machine, also performs well, with slightly higher error metrics but still showing a strong correlation between predicted and actual values.



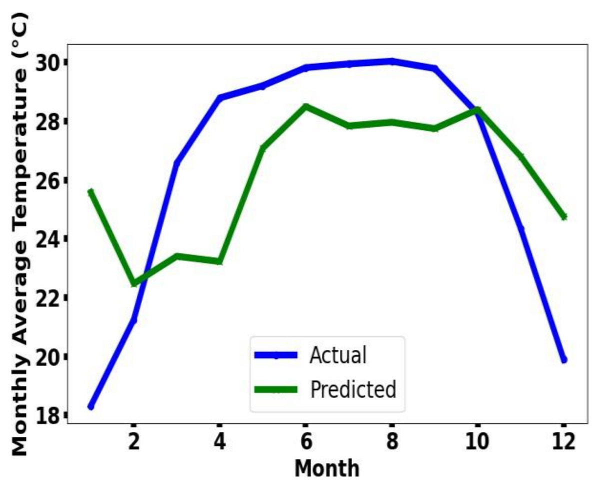
(a) Random Forest



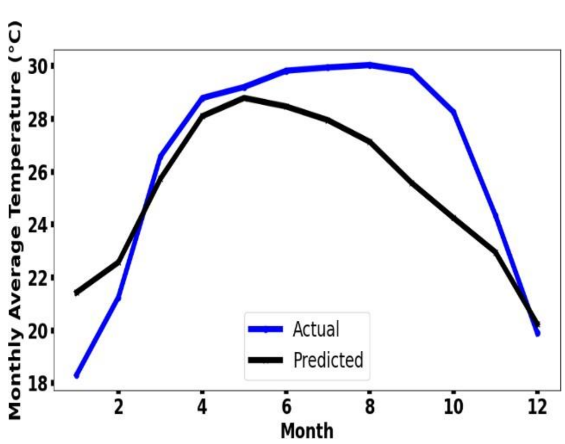
(b) Support Vector Machine



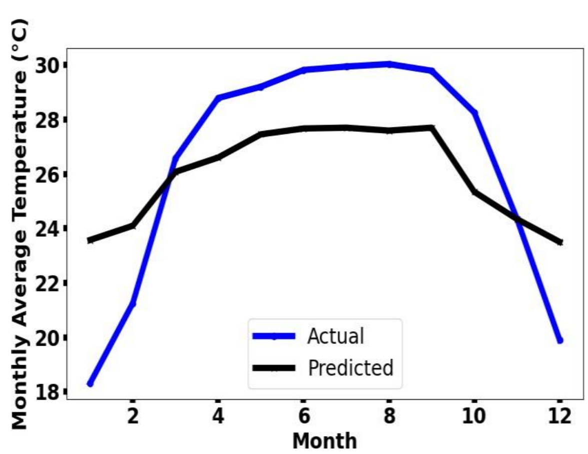
(c) Decision Tree



(d) K-Nearest Neighbor



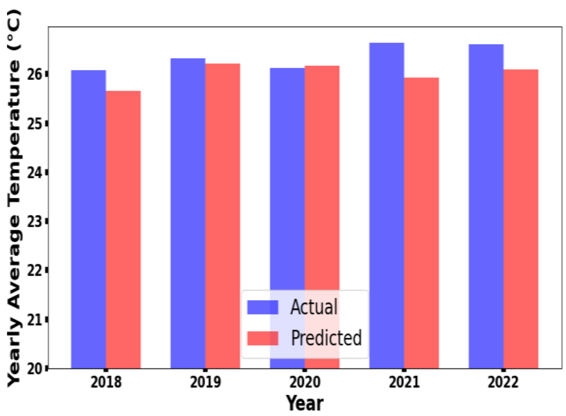
(e) Multilayer Neural Network



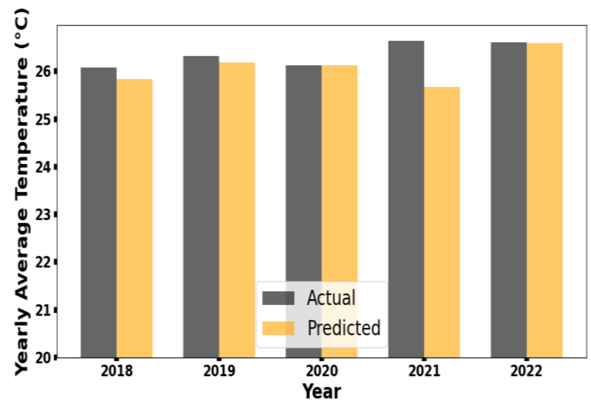
(f) Extreme Gradient Boosting

Figure 4.46: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

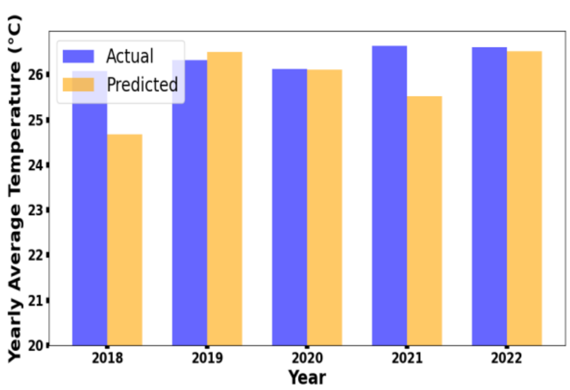
K-Nearest Neighbors and Extreme Gradient Boosting models demonstrate relatively higher errors and lower R2 accuracy, suggesting slightly less accurate predictions compared to the other models. Overall, RF, DT, and MLNN models appear to be the top performers on the training dataset, while SVM performs competitively, and KNN and XGB may require further optimization for improved performance.



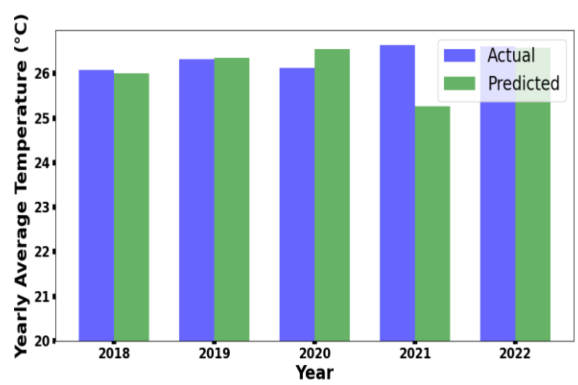
(a) Random Forest



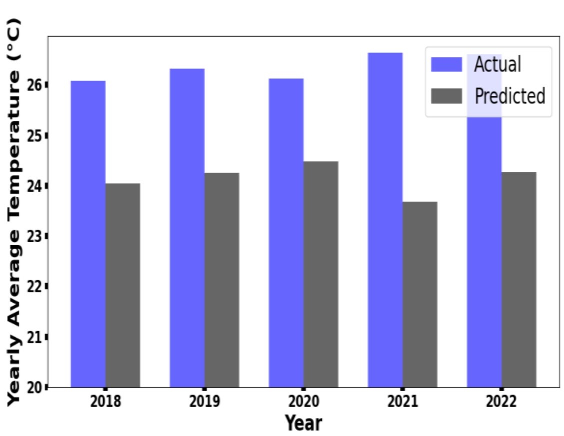
(b) Support Vector Machine



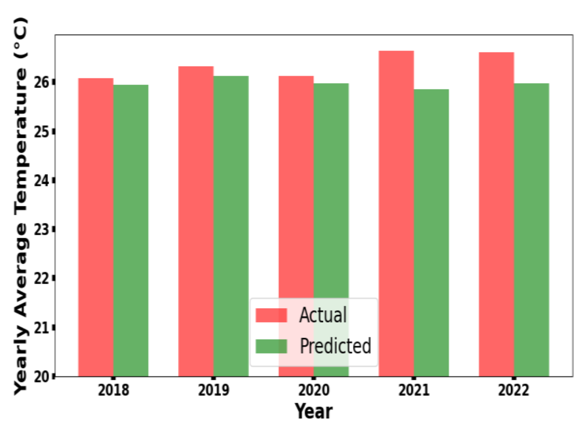
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.47: Comparison of Actual VS Predicted Yearly Temperature Different ML Algorithm Using Testing Dataset.

From Table 4.8, a concise analysis of the performance of six machine learning models on the testing dataset can be made. Random Forest and Multilayer Neural Network models exhibit relatively low MSE, RMSE, and MAE values, indicating accurate predictions and high R2 values and R2 accuracy. Decision Tree performs moderately well but shows higher error metrics compared to RF and MLNN. Support Vector Machine K-Nearest Neighbors, and

Extreme Gradient Boosting models demonstrate higher errors and lower R2 accuracy, suggesting slightly less accurate predictions compared to the other models. Overall, RF and MLNN models appear to be the top performers on the testing dataset, while SVM, KNN, and XGB may require further optimization for improved performance.

Table 4.7 Performance of Training Results from Temperature Prediction for 1999-2022

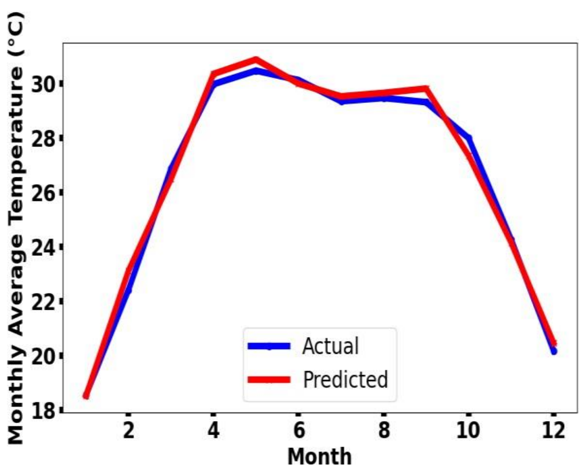
ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.04	0.02	0.09	0.98	98.37%
SVM	0.01	0.09	0.09	0.36	0.89	89.01%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.03	0.17	0.12	0.51	0.63	63.47%
MLNN	0.02	0.13	0.10	0.43	0.77	76.83%
XGB	0.03	0.17	0.15	0.61	0.60	59.63%

Table 4.8 Performance of Testing Results from Temperature Prediction for 1999-2022

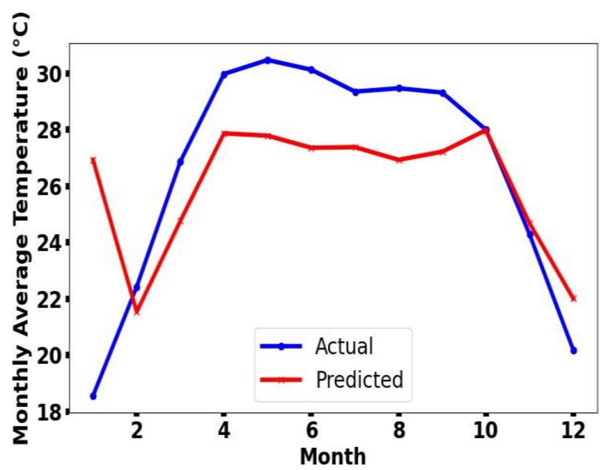
ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.08	0.06	0.26	0.92	92.26%
SVM	0.08	0.28	0.23	0.94	0.03	3.31%
DT	5.01	2.24	1.32	0.36	0.71	71.34%
KNN	0.08	0.28	0.21	0.86	0.03	3.49%
MLNN	0.03	0.17	0.14	0.59	0.61	61.40%
XGB	0.03	0.19	0.16	0.66	0.56	55.69%

(d) For Khulna Division

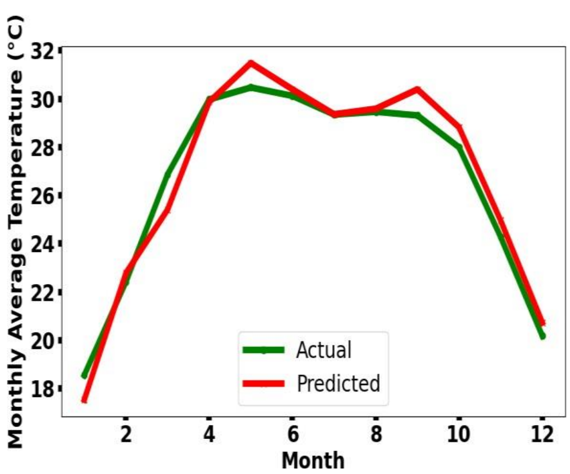
Based on the figure 4.48, the six machine learning models, Random Forest, Decision Tree, Support Vector Machine K-Nearest Neighbors, Multi-Layer Neural Network, and XGBoost , are compared based on their performance in predicting monthly temperatures. RF and XGB exhibit the closest alignment between actual and predicted values, indicating high performance. DT and KNN show more significant deviations, suggesting lower accuracy. SVM and MLNN fall in between, showing decent performance but with some errors.



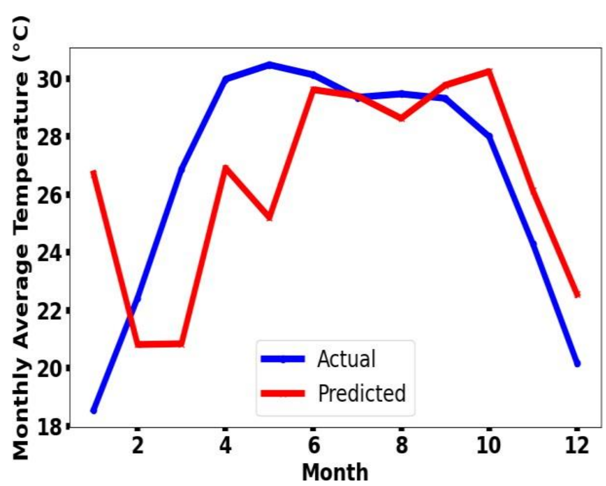
(a) Random Forest



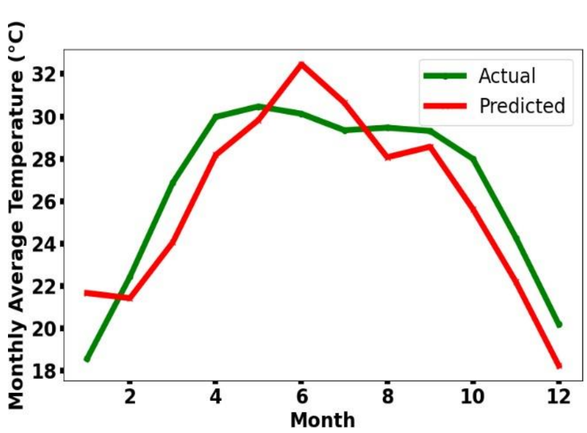
(b) Support Vector Machine



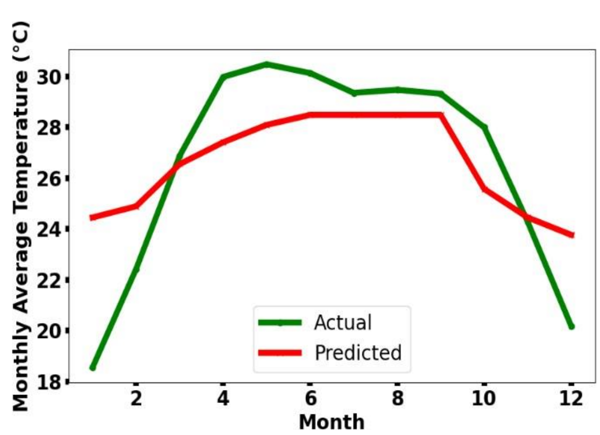
(c) Decision Tree



(d) K-Nearest Neighbor



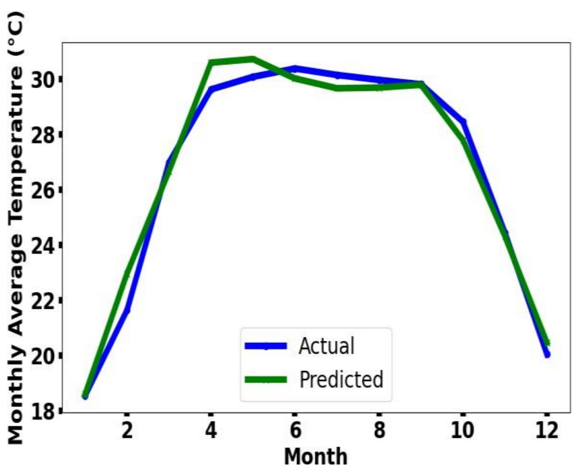
(e) Multilayer Neural Network



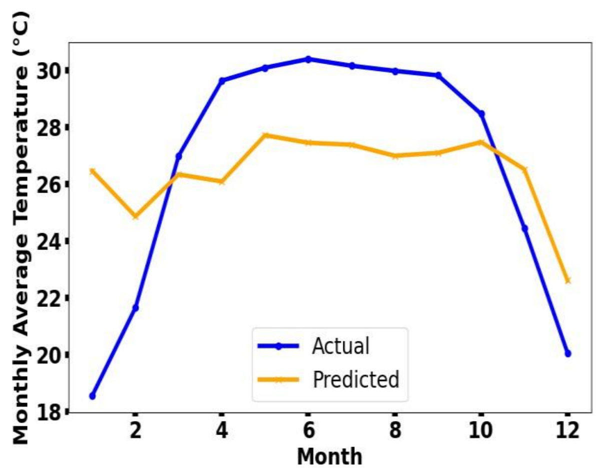
(f) Extreme Gradient Boosting

Figure 4.48: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

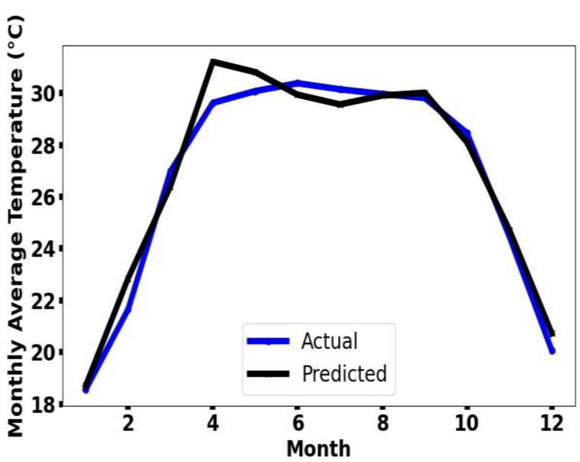
Overall, RF and XGB emerge as the top performers, while DT and KNN demonstrate the least accuracy. Further validation and optimization may be needed for the models to generalize well to new data. Figure 4.49 shows the six machine learning models, Random Forest, Decision Tree , Support Vector Machine , K-Nearest Neighbors , Multi-Layer Neural Network , and XGBoost , are compared based on their performance in predicting monthly temperatures. RF and XGB exhibit a close alignment between actual and predicted values, indicating high performance. DT and KNN show more variance, suggesting lower accuracy, while SVM and MLNN fall in between, showing decent performance with some errors



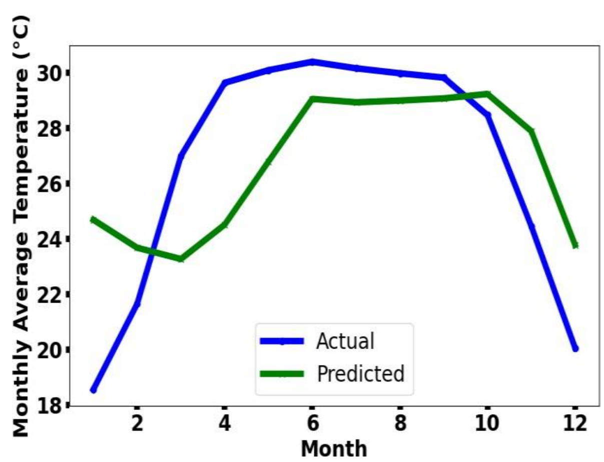
(a) Random Forest



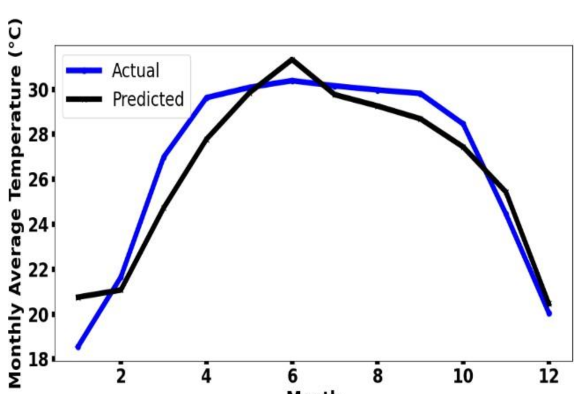
(b) Support Vector Machine



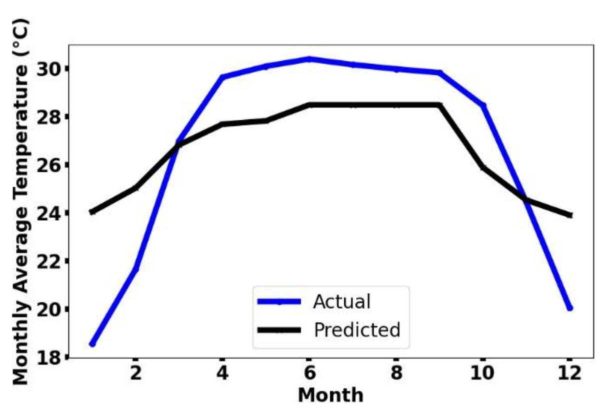
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network

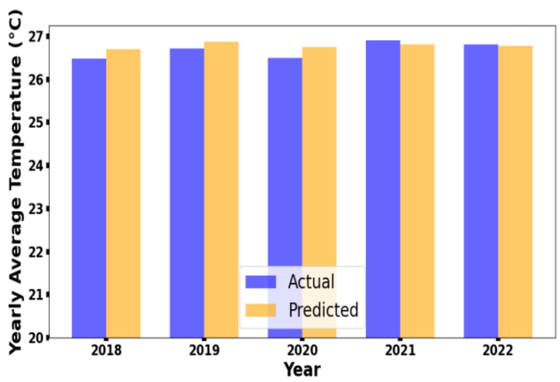


(f) Extreme Gradient Boosting

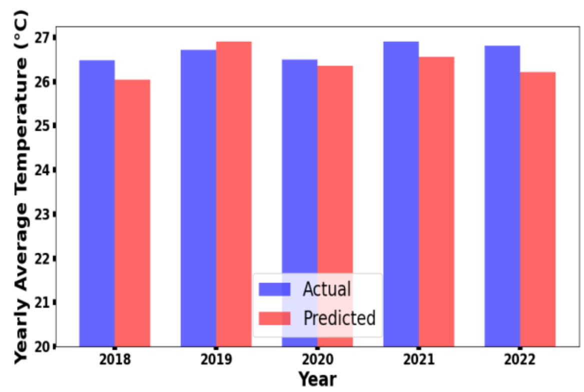
Figure 4.49: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

Overall, RF and XGB emerge as the top performers, while DT and KNN demonstrate the least accuracy. Further validation and optimization may be needed for the models to generalize well to new data.

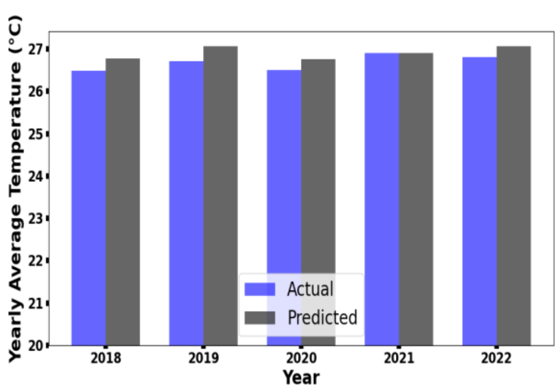
In figure 4.50 Random Forest, Multi-Layer Neural Network, and XGBoost demonstrate high performance with close alignment between actual and predicted values. DecisionTree shows lower accuracy, while Support Vector Machine and K-Nearest Neighbors exhibit lowest performance. RF, MLNN, and XGB are top performers, while DT has the least accuracy. Further validation and parameter optimization are recommended for precisepredictions.



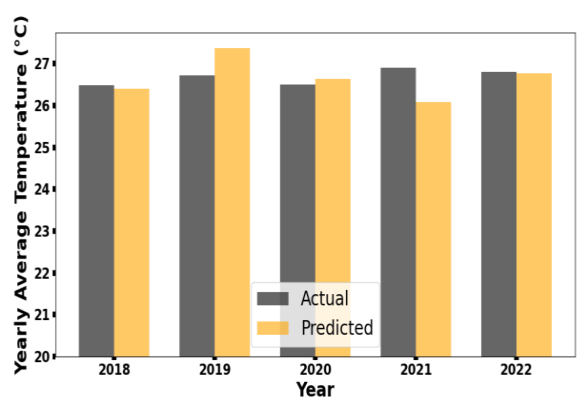
(a) Random Forest



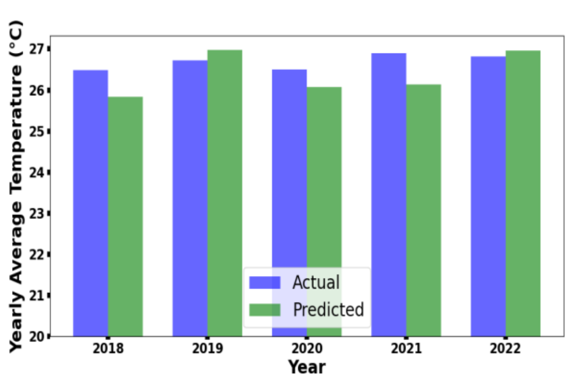
(b) Support Vector Machine



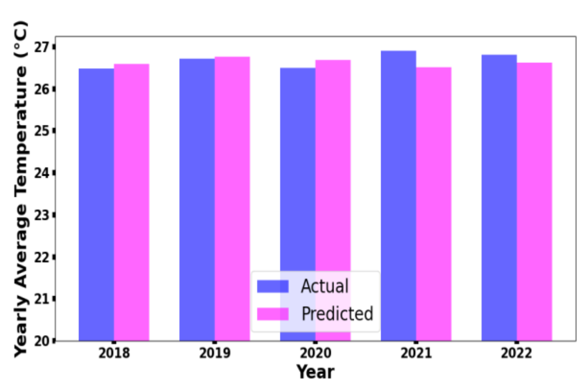
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.50: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

From Table 4.9 we notice that the Random Forest and Decision Tree models perform exceptionally well, achieving very low Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error, along with high R-squared values. Support Vector Machine, Multi-Layer Neural Network, and XGBoost also demonstrate decent performance, although with slightly higher error metrics compared to RF and DT. However, K-Nearest Neighbors exhibits relatively higher errors and lower R-squared values, suggesting comparatively weaker performance.

Table 4.9 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.02	0.02	0.07	0.99	99.41%
SVM	0.01	0.09	0.09	0.36	0.89	89.38%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.03	0.17	0.13	0.53	0.61	60.70%
MLNN	0.01	0.10	0.08	0.33	0.87	86.83%
XGB	0.03	0.17	0.14	0.57	0.61	61.49%

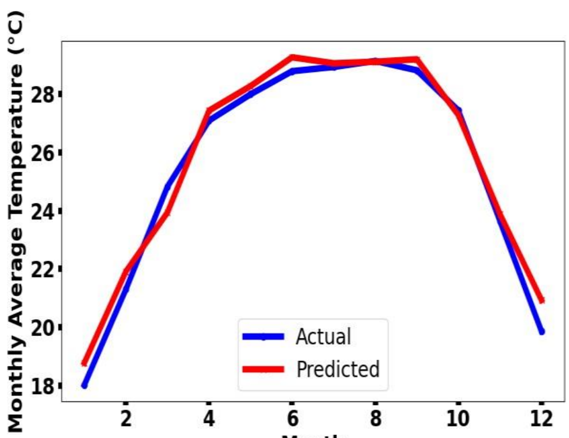
From table 4.10 we notice that Based on the training dataset analysis, the Random Forest and Decision Tree models exhibit the best performance with low Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error, along with high R-squaredvalues. Multi-Layer Neural Network also performs reasonably well, although with slightly higher errors compared to RF and DT. However, Support Vector Machine, K- Nearest Neighbors, and Extreme Gradient Boosting models show higher errors andlower R-squared values, indicating relatively weaker performance compared to RF, DT, and MLNN

Table 4.10 Performance of Testing Results from Temperature Prediction for 1999-2022

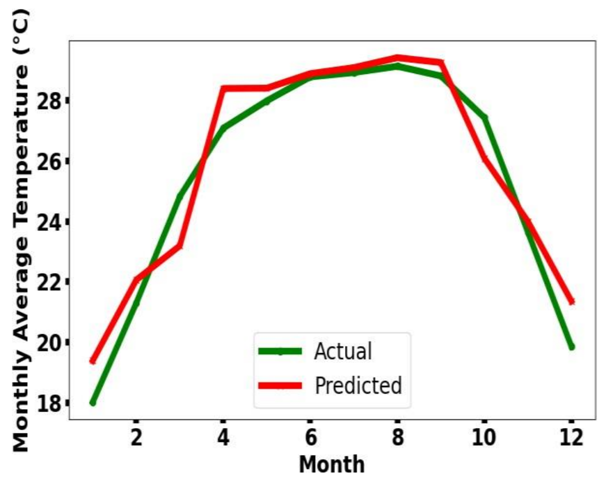
ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.06	0.04	0.18	0.96	95.94%
SVM	0.06	0.25	0.21	0.82	0.22	22.49%
DT	1.32	1.15	0.88	0.24	0.93	92.63%
KNN	0.07	0.27	0.20	0.79	0.14	13.90%
MLNN	0.02	0.13	0.10	0.39	0.79	78.97%
XGB	0.03	0.18	0.15	0.60	0.59	58.69%

d. For Mymensingh Division

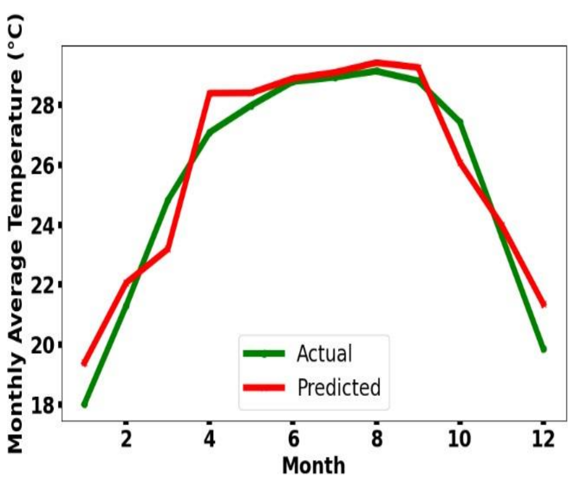
In Mymensingh Division there are only one weather station and we predict temperature for this. After that we visualize its performance by actual vs predicted graph and then showed the performance metrics in a tabular form.



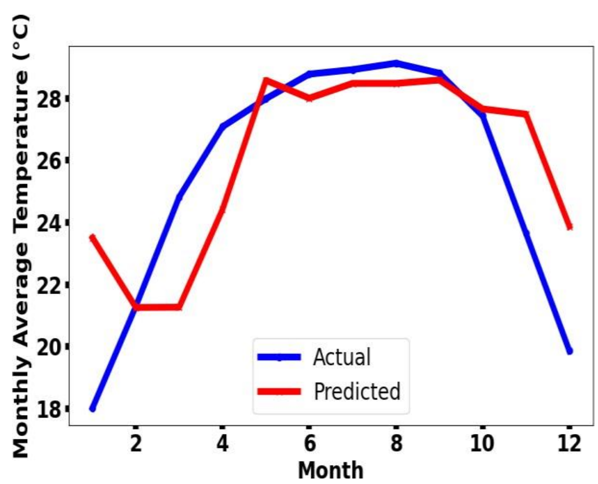
(a) Random Forest



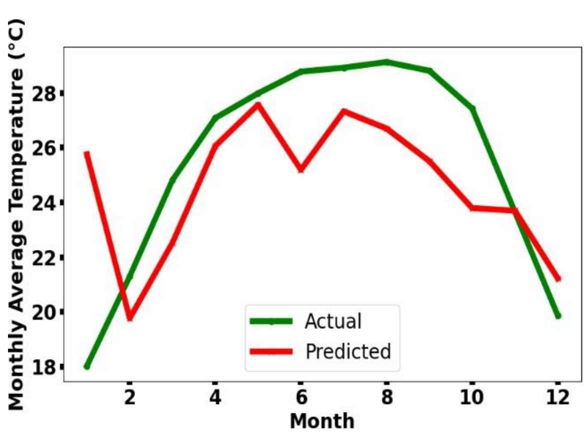
(b) Support Vector Machine



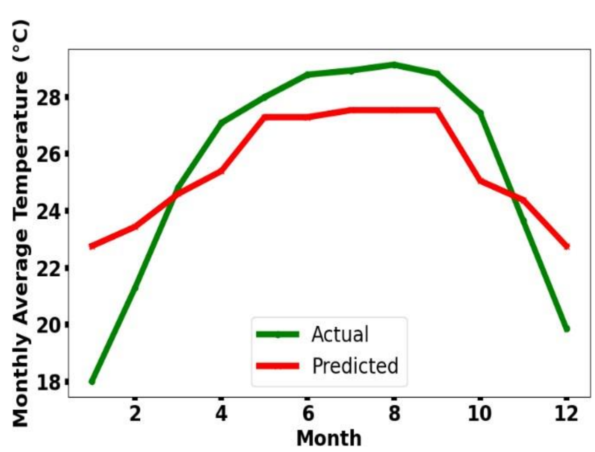
(c) Decision Tree



(d) K-Nearest Neighbor



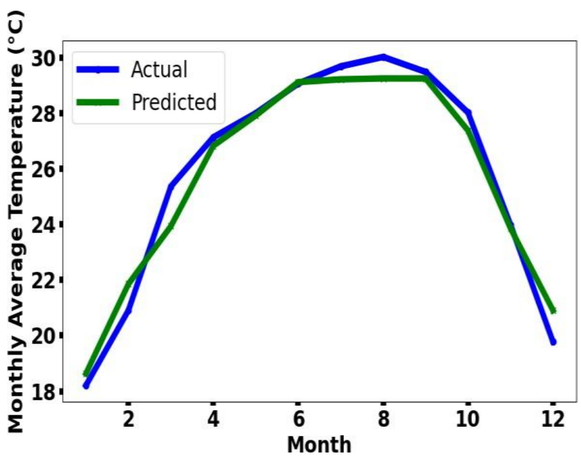
(e) Multilayer Neural Network



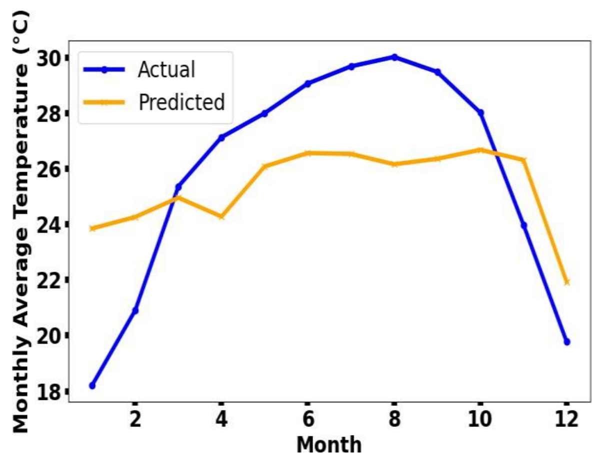
(f) Extreme Gradient Boosting

Figure 4.51: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

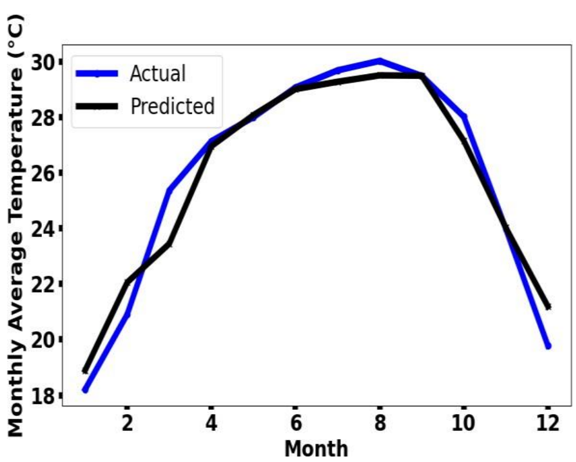
Here, Figure 4.51 Shows the performance metrics by using the testing Datasets. The comparative analysis of the machine learning models based on figure 4.51 that Random Forest, Multi- Layer Neural Network, and Extreme Gradient Boosting demonstrate the highest predictive accuracy, showing close alignment between actual and predicted values. Decision Tree exhibits some deviation, indicating lower performance compared to RF. Support Vector Machine and K-Nearest Neighbors show moderate performance, with noticeable deviations in certain periods. Overall, RF, MLNN, and XGB are identified as the top performers, while DT shows the least accuracy.



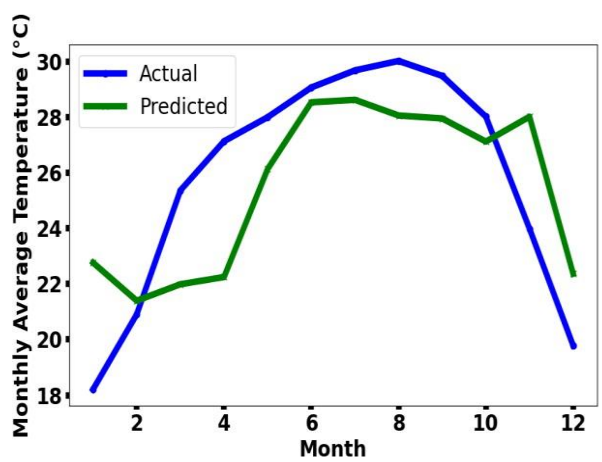
(a) Random Forest



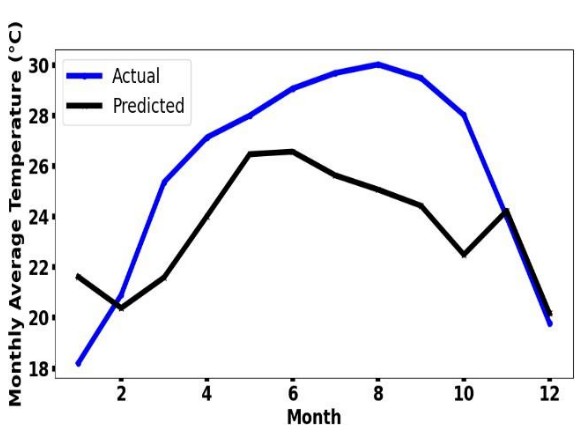
(b) Support Vector Machine



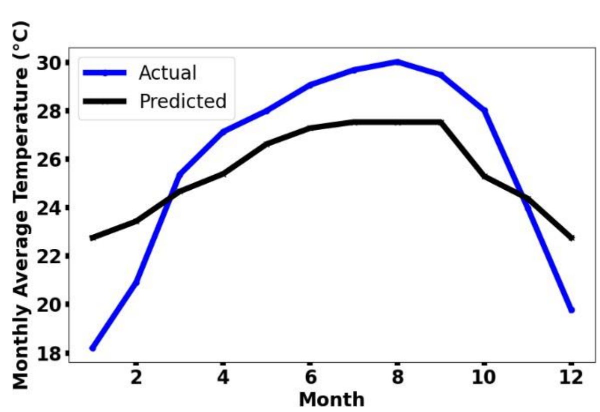
(c) Decision Tree



(d) K-Nearest Neighbor



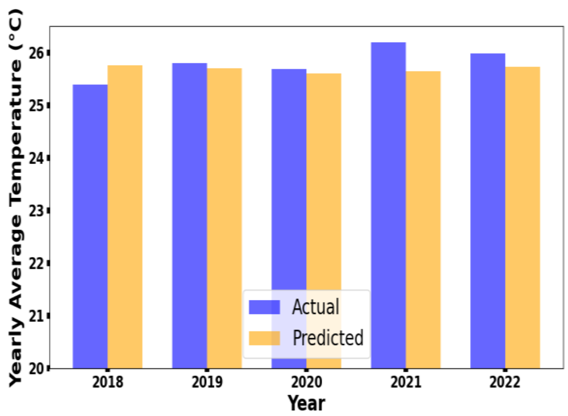
(e) Multilayer Neural Network



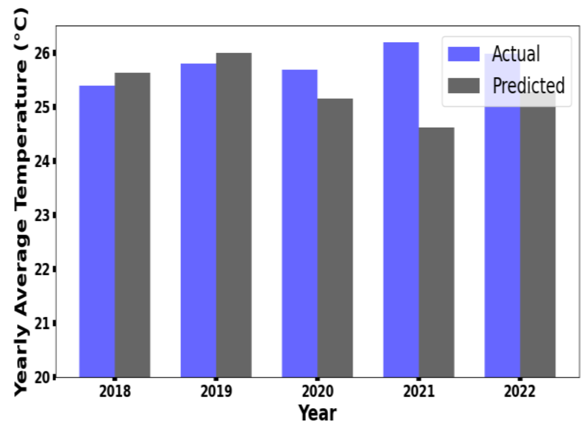
(f) Extreme Gradient Boosting

Figure 4.52: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

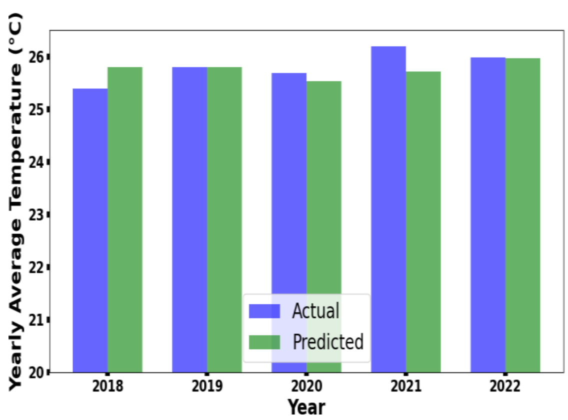
Figure 4.51 and figure 4.52 shows the comparative analysis of monthly temperature, on the other hand figure 4.52 The comparative analysis indicates that Random Forest, Multi-Layer Neural Network, and XGBoost demonstrate high performance with close alignment between actual and predicted values. Decision Tree shows the least accuracy, while SupportVector Machine and K-Nearest Neighbors exhibit moderate performance. Further tuning and validation are recommended to ensure the generalization of these findings.



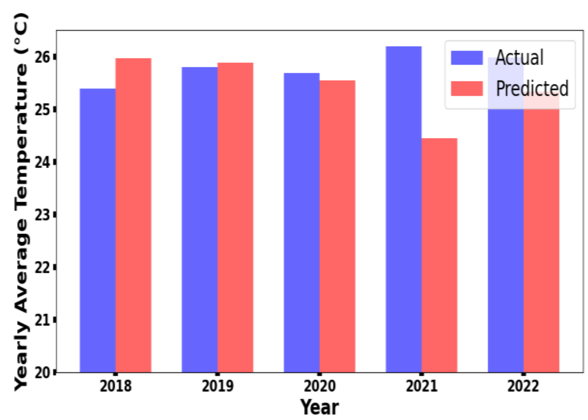
(a) Random Forest



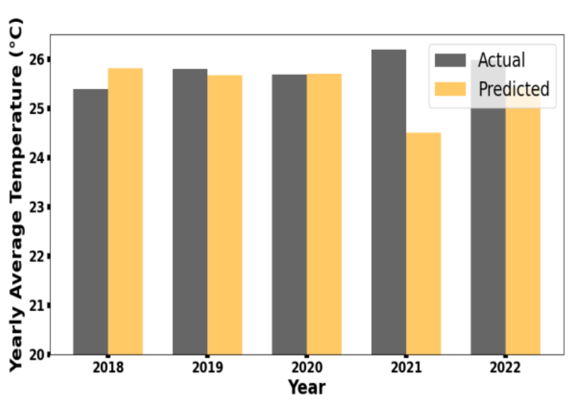
(b) Support Vector Machine



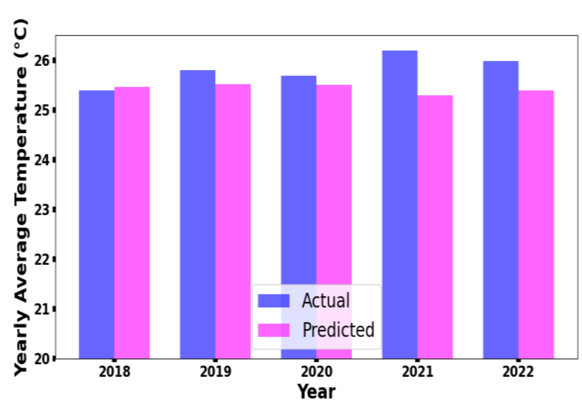
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.53: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

From figure 4.53 we found that, XGBoost and Random Forest demonstrate the best performance, closely followed by SVM and MLNN. Decision Tree and KNN show less accurate predictions, possibly due to overfitting or model simplicity. Factors like dataset complexity and feature count should be considered when choosing a model. Further validation and tuning are recommended to ensure reliable performance across different datasets and scenarios.

Table 4.11 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.02	0.02	0.07	0.99	99.34%
SVM	0.01	0.09	0.08	0.37	0.89	88.66%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.02	0.14	0.10	0.44	0.71	70.75%
MLNN	0.04	0.20	0.16	0.71	0.42	41.74%
XGB	0.02	0.16	0.13	0.55	0.65	65.22%

Table 4.11 shows the performance of training dataset. The values in the table indicates that Decision Tree achieves perfect accuracy with an R2 value and R2 accuracy of 1.0 and 100%, respectively, suggesting potential overfitting. Random Forest and Support Vector Machine demonstrate strong performance with high R2 values and R2 accuracies. K-Nearest Neighbors shows moderate performance with relatively lower R2 values and accuracies. Multi-Layer Neural Network and XGBoost exhibit weaker performance compared to the other models, with lower R2 values and accuracies.

Table 4.12 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.07	0.05	0.22	0.94	93.90%
SVM	0.06	0.24	0.20	0.83	0.29	28.61%
DT	1.48	1.22	0.87	0.25	0.91	90.92%
KNN	0.05	0.22	0.17	0.72	0.39	38.97%
MLNN	0.07	0.27	0.22	0.91	0.04	4.07%
XGB	0.03	0.17	0.15	0.62	0.62	62.35%

Table 4.12 shows the performance of Testing Results of Mymensingh Division. The testing dataset evaluation reveals that RF performs the best with the lowest MSE, RMSE, MAE, and RAE among all models, indicating superior predictive accuracy. SVM, DT, KNN, and XGB show moderate performance, while MLNN performs the poorest with the highest MSE, RMSE, MAE, and RAE values, suggesting significant prediction errors. RF and XGB maintain relatively high R2 values and accuracies, indicating robust performance, while SVM, DT, and KNN exhibit lower R2 values and accuracies.

RF demonstrates consistent high performance across both training and testing datasets, making it the top-performing model for this dataset. SVM, DT, KNN, and XGB show varying degrees of performance, while MLNN performs the worst. Further analysis, model tuning, and validation are recommended to improve the performance of the models, especially MLNN, which struggles with generalization on unseen data.

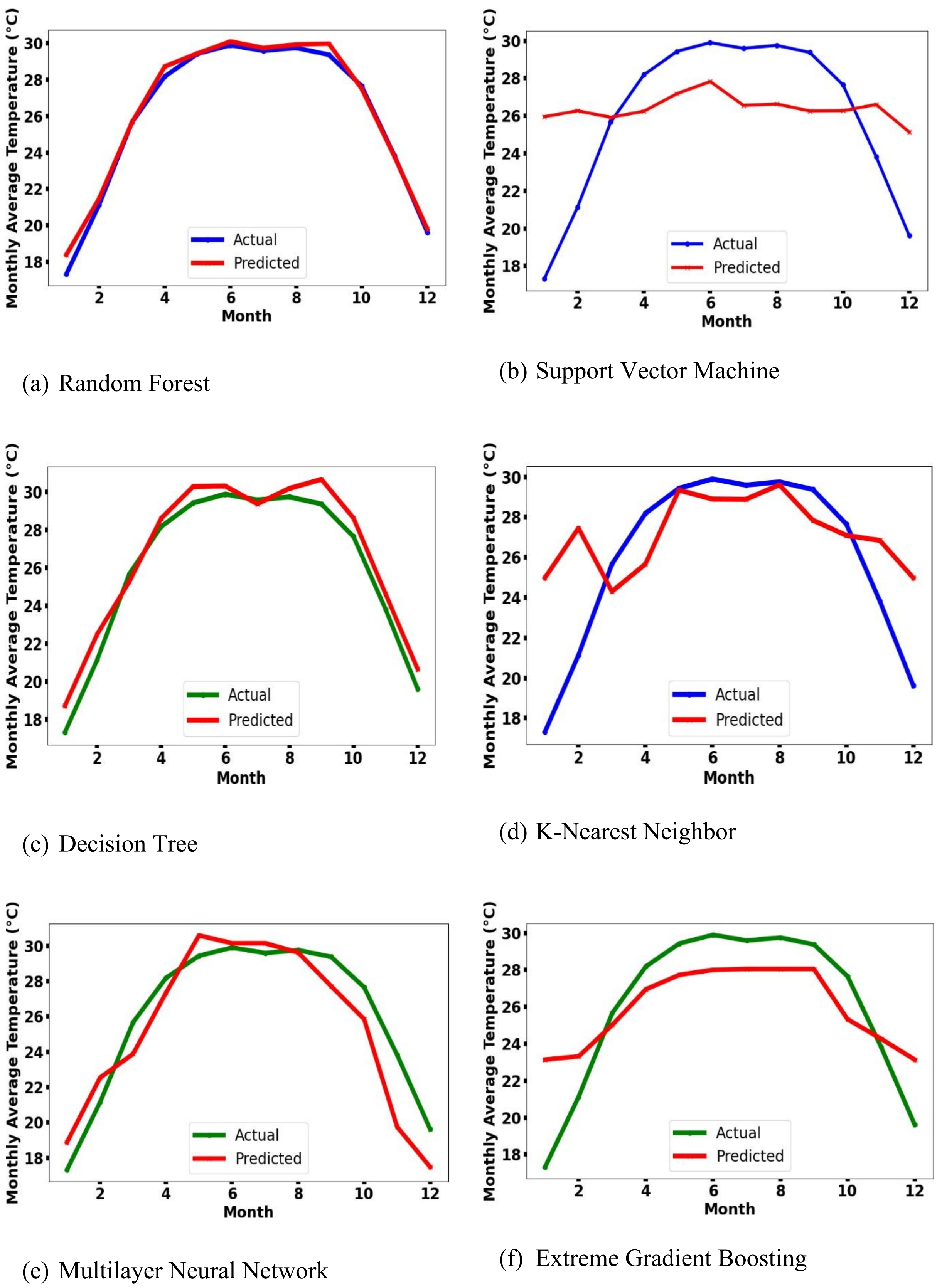
e. For Rajshahi Division

Here, we predict the temperature of Rajshahi Division, there are 2 weather station under Rajshahi Division, Here Figure 4.54 shows The comparative analysis of the six machine learning models which we used for our prediction, between them XGBoost and Random Forest perform the best, followed closely by SVM and MLNN. Decision Tree and KNN exhibit lower performance, possibly due to overfitting and simplicity, respectively. Considerations such as dataset complexity, feature count, and computational resources are crucial in selecting the appropriate model. Further validation and tuning may enhance overall model performance.

Figure 4.55 represent the performance of six machine learning models in predicting monthly temperatures. Random Forest and XGBoost are anticipated to perform the best due to their expected strong alignment between actual and predicted values, reflecting their robustness and high accuracy.

Support Vector Machine and Multi-Layer Neural Network are also expected to perform well, demonstrating good capability in capturing complex relationships and modeling patterns. However, Decision Tree and K-Nearest Neighbors may exhibit poorer performance, potentially due to overfitting or model simplicity, leading to more variance or discrepancies between actual and predicted values. To confirm these observations, further assessment using quantitative metrics like MAE, RMSE, or R-squared values would be necessary., and XGBoost. Based on the description, XGBoost and Random Forest are expected to perform the best due to their strong alignment between actual and predicted values, reflecting their accuracy and robustness. SVM and MLNN are also anticipated to perform well, while DT and KNN may exhibit poorer performance due to potential overfitting or model simplicity. Further analysis using quantitative metrics is recommended for a comprehensive assessment.

The figure in 4.56 contains six graphs comparing actual versus predicted yearly temperatures generated by different machine learning models mentioned as before.



(a) Random Forest

(b) Support Vector Machine

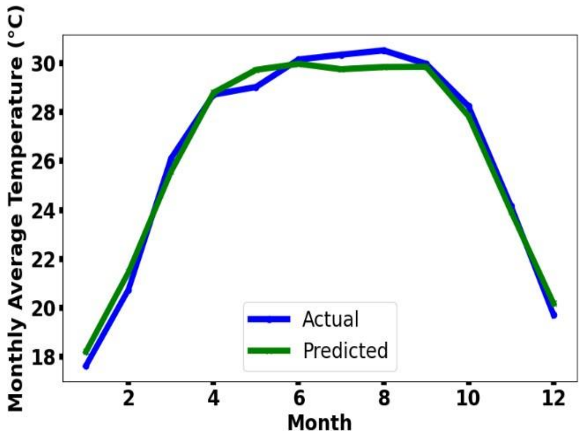
(c) Decision Tree

(d) K-Nearest Neighbor

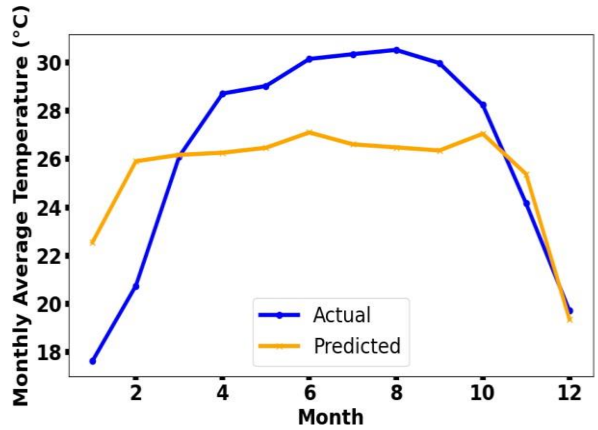
(e) Multilayer Neural Network

(f) Extreme Gradient Boosting

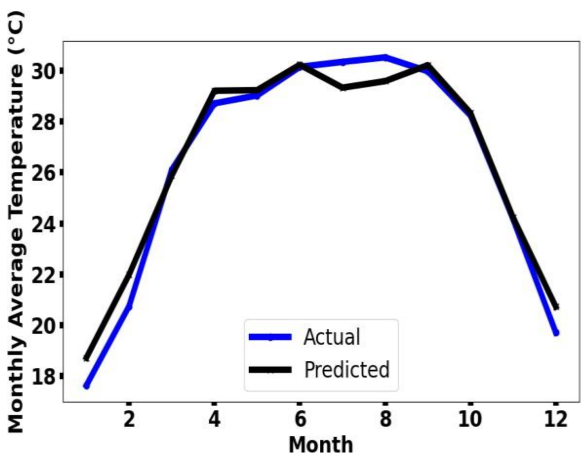
Figure 4.54: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.



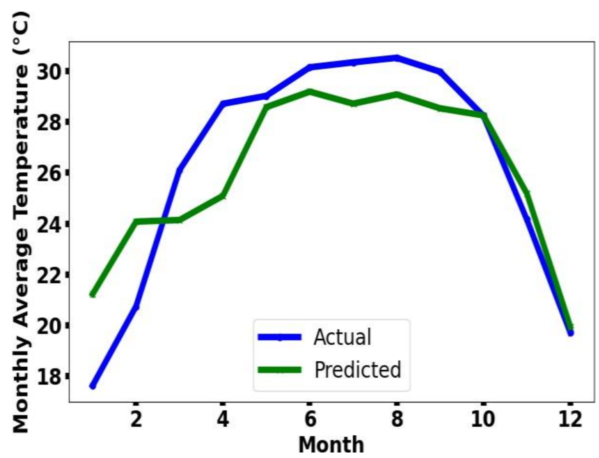
(a) Random Forest



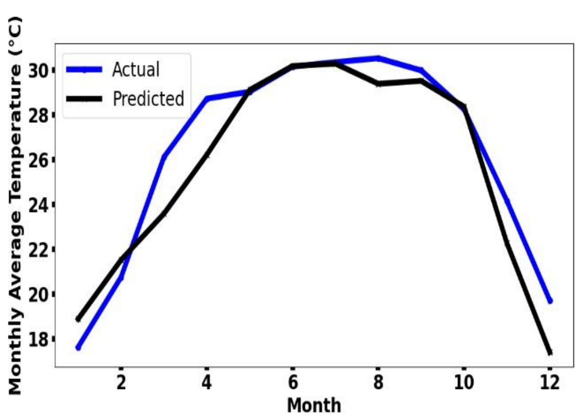
(b) Support Vector Machine



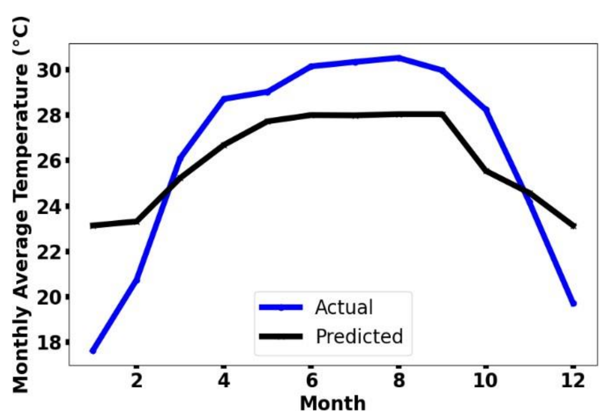
(c) Decision Tree



(d) K-Nearest Neighbor



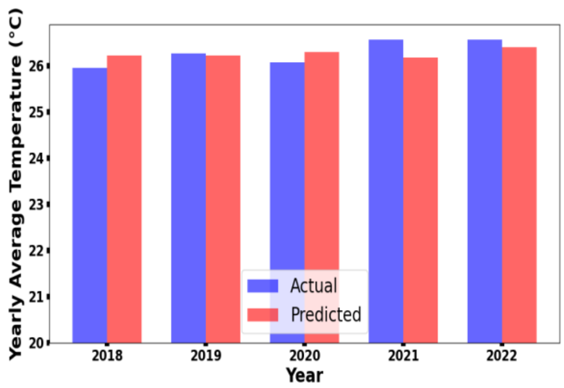
(e) Multilayer Neural Network



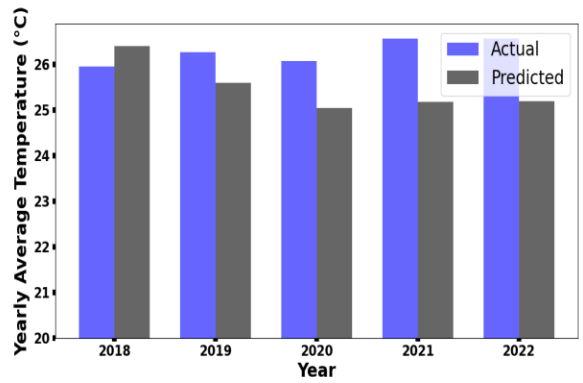
(f) Extreme Gradient Boosting

Figure 4.55: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

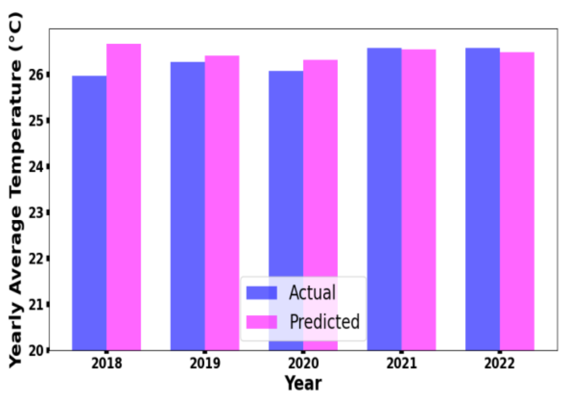
Table 4.13 and 4.14 shows The performance of six machine learning models, for both the training and testing datasets for temperature prediction. From the Tables we find that Random Forest and Multi- Layer Neural Network consistently exhibited superior performance across both datasets, characterized by low error metrics and high R2 accuracy. Conversely, Decision Tree showed signs of overfitting, performing perfectly on the training dataset but significantly worse on the testing dataset. Support Vector Machine performed moderately, while K-Nearest Neighbors and XGBoost showed moderate performance but were not as robust as Random Forest and Multi-Layer Neural Network. Further optimization and validation are recommended to enhance model performance and address potential overfitting concerns, particularly with Decision Tree. Finally, Random Forest and Multi-Layer Neural Network emerged as the top performers for temperature prediction, demonstrating consistent accuracy and robustness across both training and testing datasets. Decision Tree exhibited overfitting tendencies, while Support Vector Machine, K-Nearest Neighbors, and XGBoost displayed moderate performance.



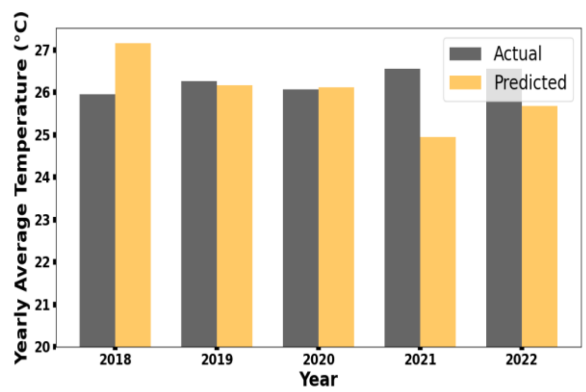
(a) Random Forest



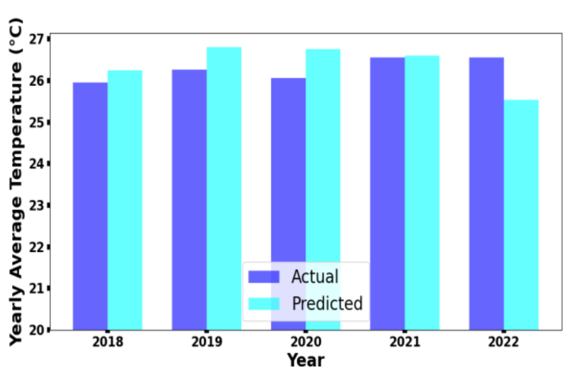
(b) Support Vector Machine



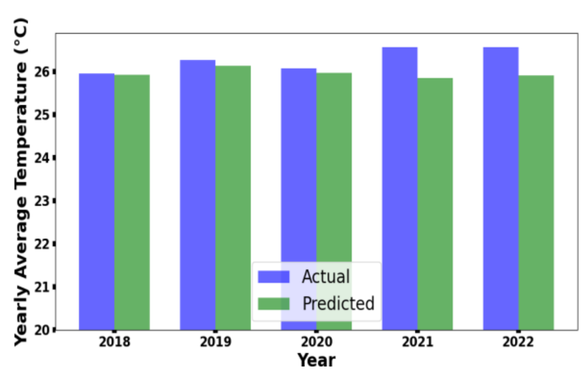
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.56: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

Table 4.13 Performance of Training Results from Temperature Prediction for 1999-2022

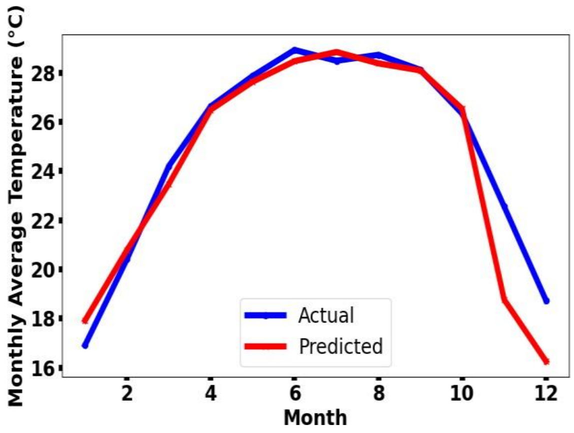
ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.08	0.99	98.61%
SVM	0.01	0.09	0.09	0.40	0.87	87.14%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.01	0.12	0.08	0.38	0.78	77.95%
MLNN	0.02	0.13	0.09	0.43	0.73	73.22%
XGB	0.02	0.16	0.13	0.58	0.61	60.51%

Table 4.14 Performance of Testing Results from Temperature Prediction for 1999-2022

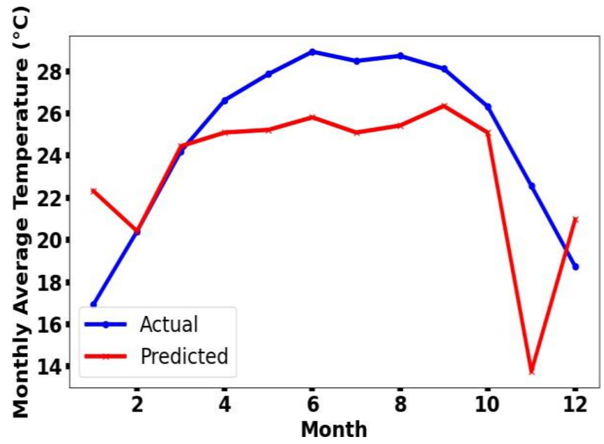
ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.05	0.04	0.18	0.96	96.18%
SVM	0.04	0.20	0.17	0.77	0.39	38.99%
DT	1.18	1.09	0.84	0.22	0.94	94.10%
KNN	0.02	0.15	0.12	0.52	0.62	62.21%
MLNN	0.01	0.12	0.09	0.41	0.78	77.57%
XGB	0.02	0.15	0.13	0.60	0.62	62.37%

f. For Rangpur Division

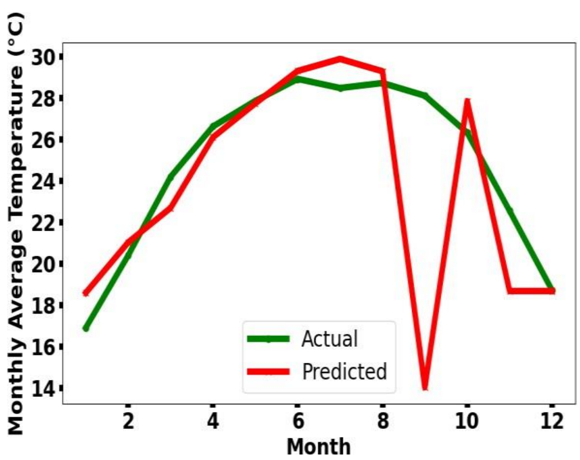
Based on the visual analysis of the temperature prediction graphs form figure 4.57, we find that the Random Forest and XGBoost models demonstrate the highest performance, closely aligning with actual values, while the K-Nearest Neighbors model appears to have the least accurate predictions with noticeable deviations. Decision Tree, Support Vector Machine, and Multi-Layer Neural Network models exhibit moderate to high performance, with varying degrees of accuracy. It's essential to validate these observations with quantitative metrics like Mean Squared Error to ensure a comprehensive evaluation of each model's performance.



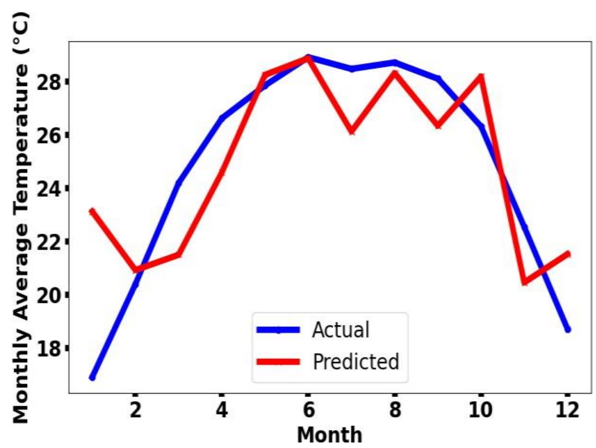
(a) Random Forest



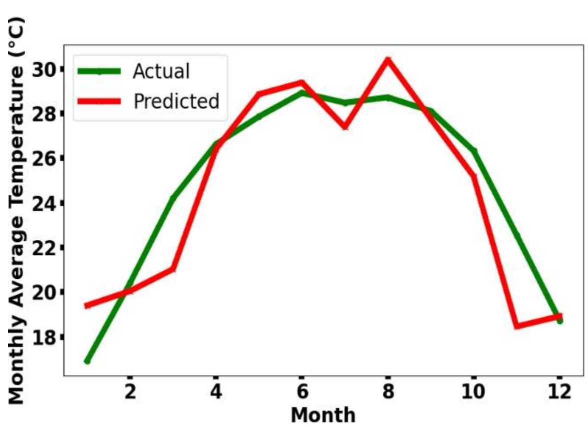
(b) Support Vector Machine



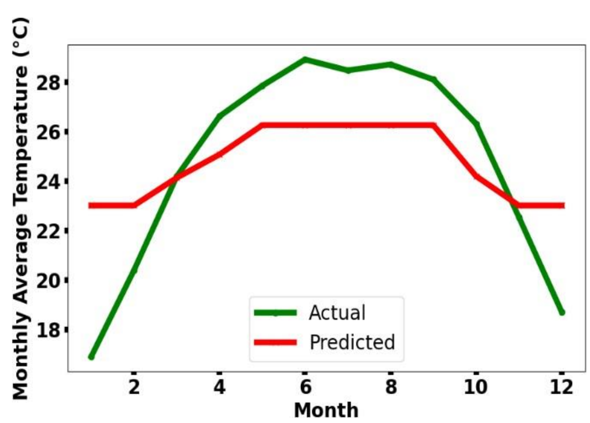
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network

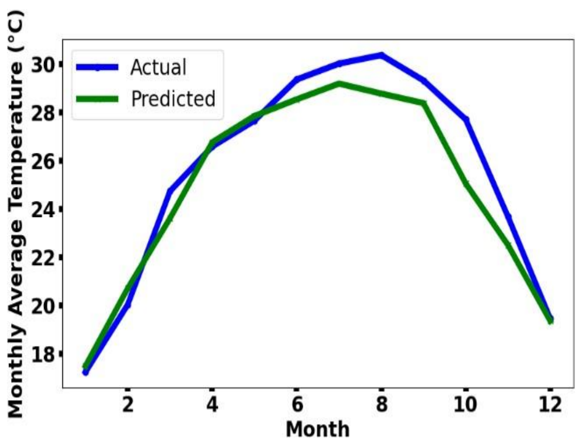


(f) Extreme Gradient Boosting

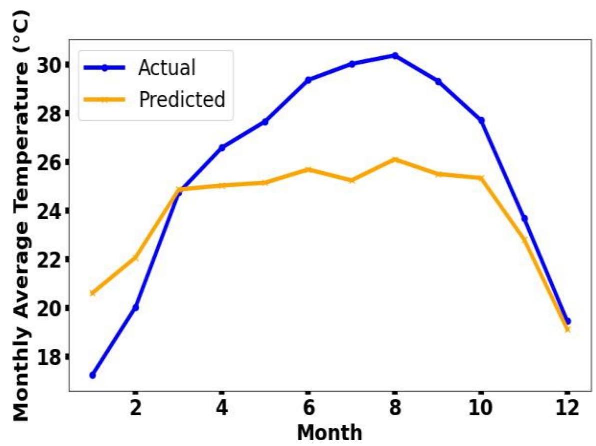
Figure 4.57: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.

Based on the comparative analysis of the six machine learning models from Figure 4.58, Random Forest and XGBoost demonstrate higher performance levels, closely aligning with actual values. Conversely, K-Nearest Neighbors exhibits the lowest accuracy with significant deviations in predictions. Decision Tree, Support Vector Machine, and Multi-Layer Neural Network fall in between, displaying moderate to high accuracy with varying degrees of precision. However, it's essential to validate these observations with quantitative metrics like Mean Squared Error for a comprehensive evaluation.

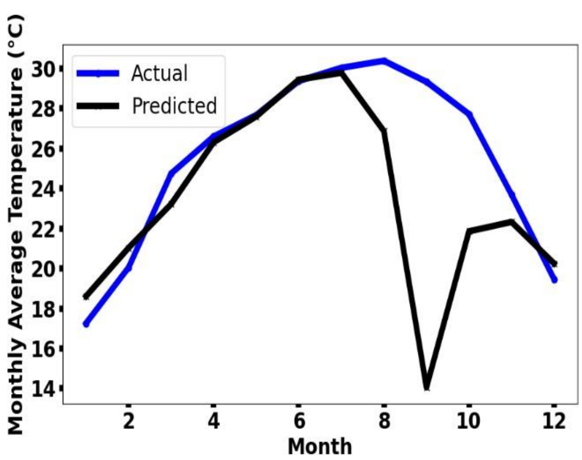
From figure 4.59 we found that Random Forest and XGBoost demonstrate high accuracy, with very close alignment between actual and predicted values. Support Vector Machine shows moderate accuracy, with slight differences between actual and predicted values. K- Nearest Neighbors also exhibits moderate accuracy, with minor discrepancies. However, Decision Tree and Multi-Layer Neural Network models appear to have lower accuracy, as there are notable differences between actual and predicted values.



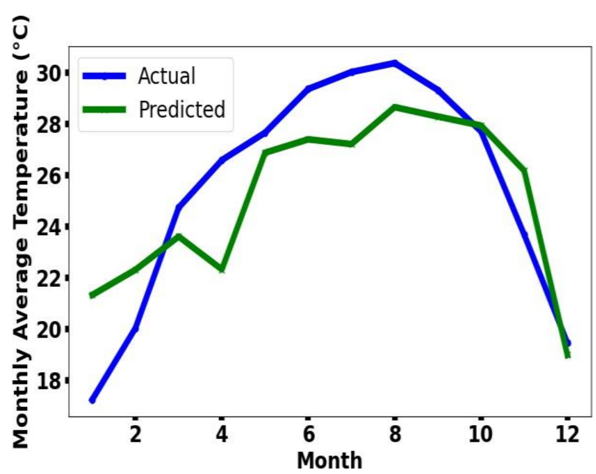
(a) Random Forest



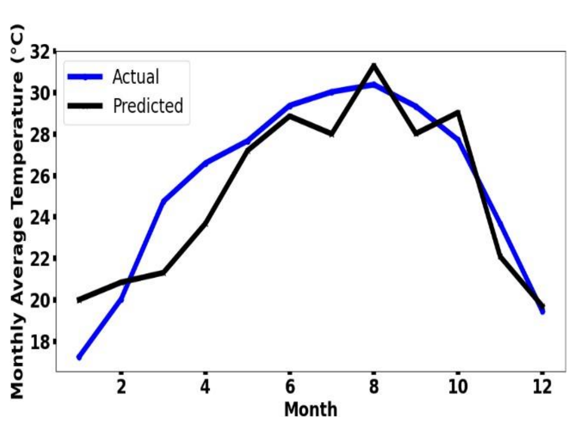
(b) Support Vector Machine



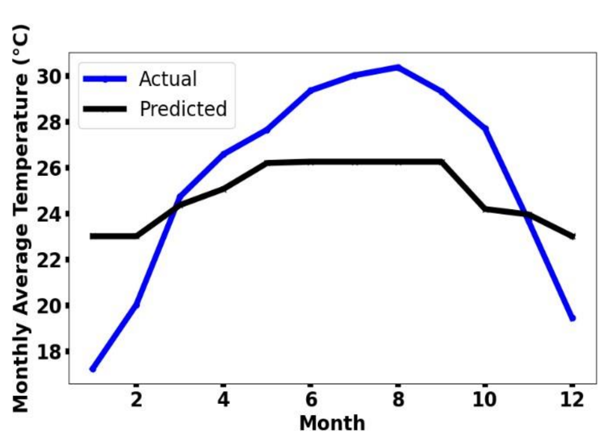
(c) Decision Tree



(d) K-Nearest Neighbor

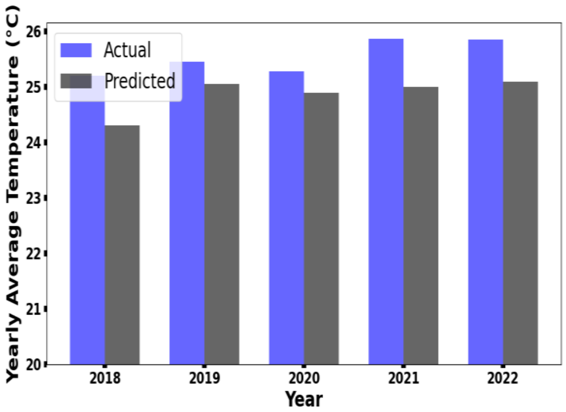


(e) Multilayer Neural Network

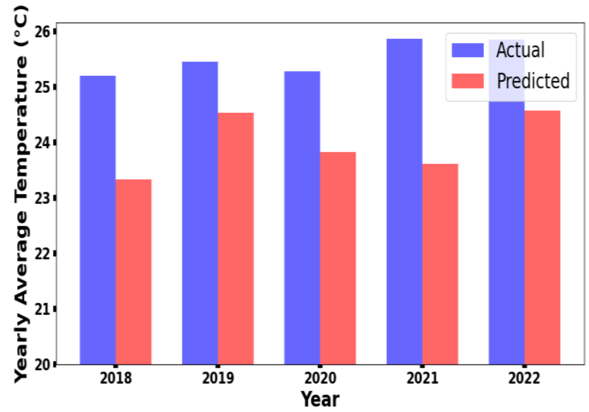


(f) Extreme Gradient Boosting

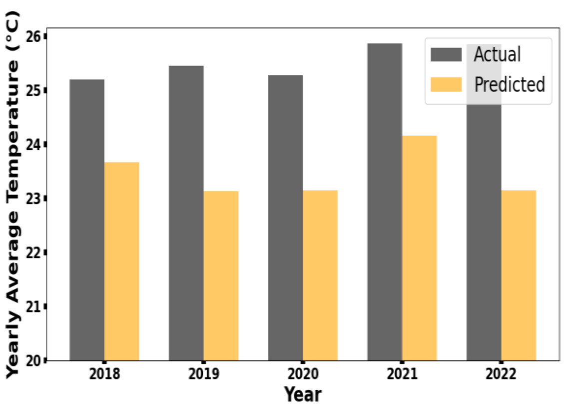
Figure 4.58: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.



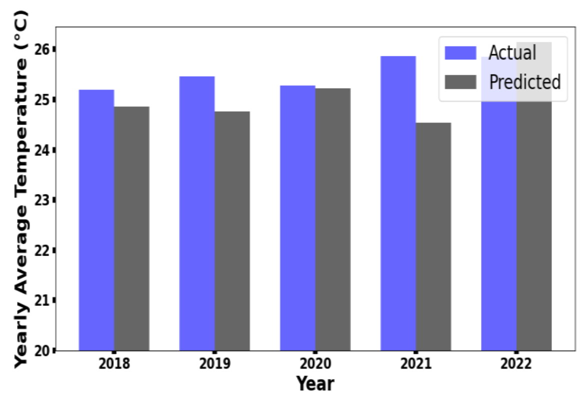
(a) Random Forest



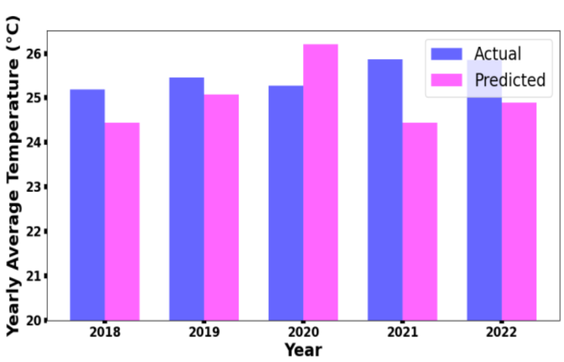
(b) Support Vector Machine



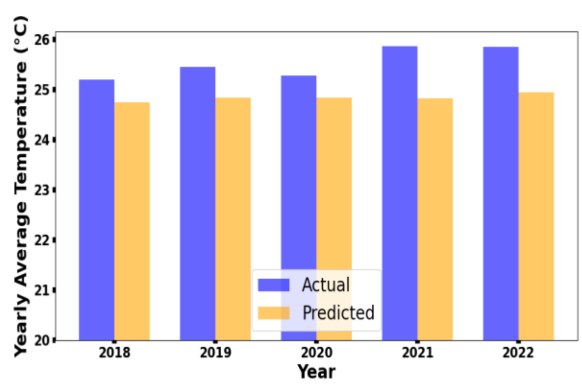
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.59: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

Table 4.15 presents the performance metrics of the temperature prediction model for the dataset spanning 1999-2022, focusing on the training results. The table includes metrics such as Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, Relative Absolute Error , R-squared (R2) value, and R2 accuracy for six different machine learningmodels: Random Forest , Support Vector Machine , Decision Tree , K-Nearest Neighbors , Multi-Layer Neural Network , and XGBoost . Notably, the Decision Tree model achieved perfect performance on the training dataset, reflected in an R2

accuracy of 100%. Random Forest and Decision Tree models exhibit the lowest RMSE and MAE values, indicating relatively accurate predictions. However, KNN and MLNN models show higher RAE values, suggesting higher relative errors. XGBoost model, despite achieving lower RMSE and MAE values, has a lower R2 accuracy compared to Random Forest and Decision Tree models.

Table 4.15 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.04	0.02	0.11	0.96	96.48%
SVM	0.01	0.09	0.08	0.45	0.83	83.08%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.02	0.13	0.09	0.52	0.61	60.67%
MLNN	0.02	0.14	0.09	0.52	0.58	58.11%
XGB	0.03	0.16	0.13	0.73	0.39	39.41%

Table 4.16 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.07	0.05	0.29	0.87	86.52%
SVM	0.02	0.16	0.13	0.77	0.38	38.46%
DT	32.27	5.68	2.78	0.74	0.69	69.06%
KNN	0.02	0.13	0.10	0.60	0.56	56.23%
MLNN	0.01	0.12	0.09	0.54	0.65	64.73%
XGB	0.02	0.15	0.13	0.75	0.44	44.35%

Moving to Table 4.16, which represents the performance metrics of the temperature prediction model for the same dataset but focusing on the testing results. The performance of the models is evaluated using same metrics like training results. In contrast to the training dataset, the Decision Tree model's performance significantly deteriorated on the testing dataset, as evidenced by its substantially higher MSE, RMSE, MAE, and RAE values. Random Forest and XGBoost models continue to demonstrate relatively accurate predictions, with lower RMSE and MAE values compared to other models. However, SVM, KNN, and MLNN models show higher RAE values, indicating a larger proportion of relative errors in their predictions.

Overall, the performance of the machine learning models on the testing dataset generally aligns with their performance on the training dataset, with Random Forest and XGBoost models consistently demonstrating relatively accurate predictions compared to other models. However, the Decision Tree model's notable drop in performance on the testing dataset underscores the importance of evaluating model generalization beyond the training datasets.

g. For Sylhet Division

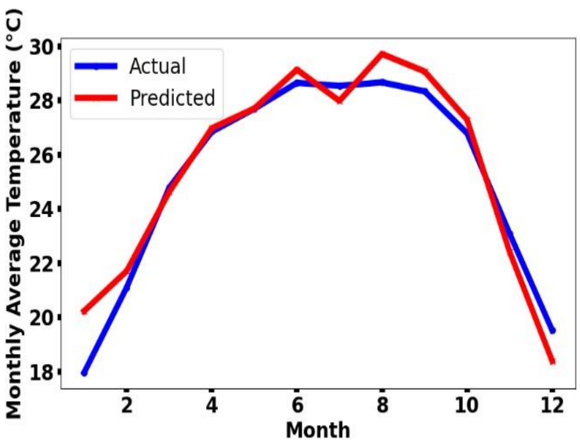
We have predicted temperature for Sylhet, where there are 3 weather station in this division and we used the same model and same performance metrics in temperature prediction for Sylhet division.

In figure 4.60 both Random Forest and Multi-Layer Neural Network exhibit high performance in predicting temperatures, as evidenced by their closely aligned actual and predicted values, suggesting minimal deviations and high accuracy. Random Forest demonstrates robust performance with its ensemble of decision trees, while MLNN showcases its capability to model complex patterns effectively. These models outperform others like Decision Tree, Support Vector Machine, and K-Nearest Neighbor, which show varying degrees of performance, with SVM and KNN exhibiting moderate to low accuracy. Additionally, XGBoost also demonstrates high performance, albeit with minor discrepancies compared to RF andMLNN.

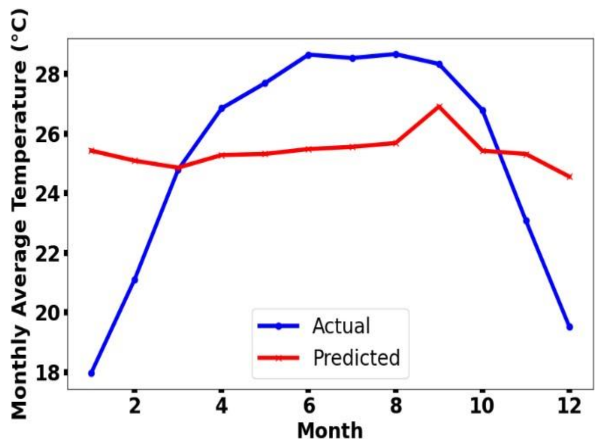
From figure 4.61 we found that Random Forest and Multi-Layer Neural Network models demonstrate high performance, closely followed by XGBoost, with minor discrepancies. Support Vector Machine exhibits moderate performance, while K-Nearest Neighbors shows lower accuracy. However, for a comprehensive evaluation, it's crucial to validate these observations with quantitative metrics like MSE, R-squared, etc. and consider the context of the model application.

Figure 4.62 contains bar graphs representing the performance of six machine learning models in predicting values labeled as 'Actual' and 'Predicted'. Random Forest and XGBoost are typically robust and accurate due to their ensemble nature and gradient boosting techniques, respectively. Support Vector Machines perform well in high-dimensional spaces but might struggle with noisy data. Decision Trees are interpretable but prone to overfitting. K-Nearest Neighbors' performance depends heavily on parameter choice. Multi-Layer Neural Networks capture complex patterns but require significant data and computational resources.

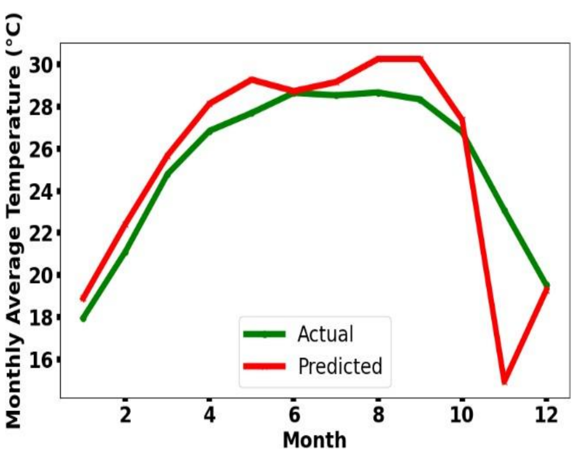
Table 4.17 present the performance metrics of a temperature prediction model for the dataset spanning from 1999 to 2022, evaluated on both the training results.Here, various machine learning models are assessed based on metrics such as Mean Squared Error , Root Mean Squared Error , Mean Absolute Error , Relative Absolute Error , R-squared value (R2), and R2 Accuracy. Among these models, the Decision Tree model achieves perfect performance with an R2 value and R2 accuracy of1.00, indicating a flawless fit to the training data. Conversely, the Multi-Layer Neural Network and XGBoost models exhibit lower performance compared to others, with R2 accuracy values of 55.00% and 49.76%, respectively, suggesting less accurate predictions on the training dataset.



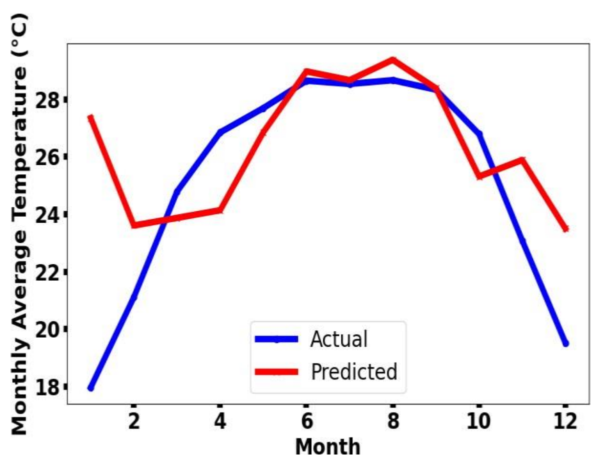
(a) Random Forest



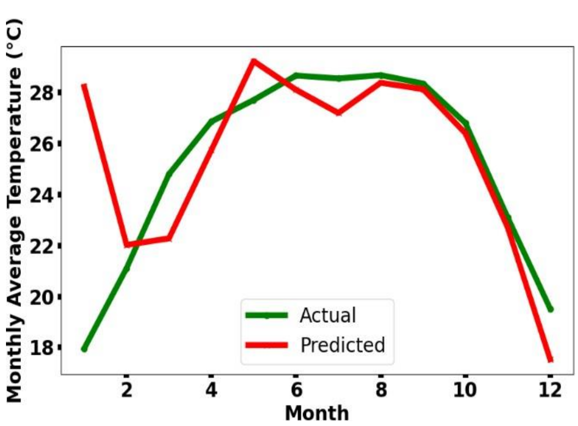
(b) Support Vector Machine



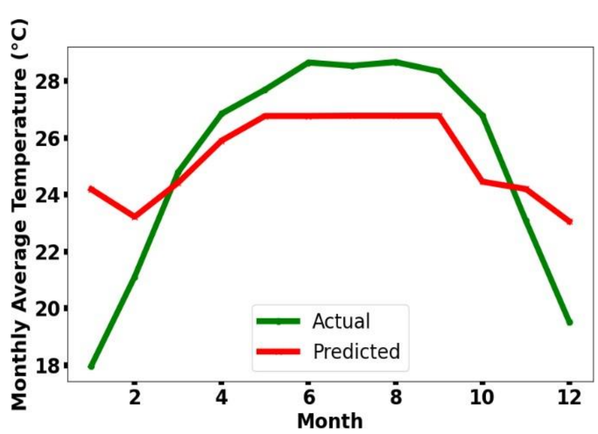
(c) Decision Tree



(d) K-Nearest Neighbor

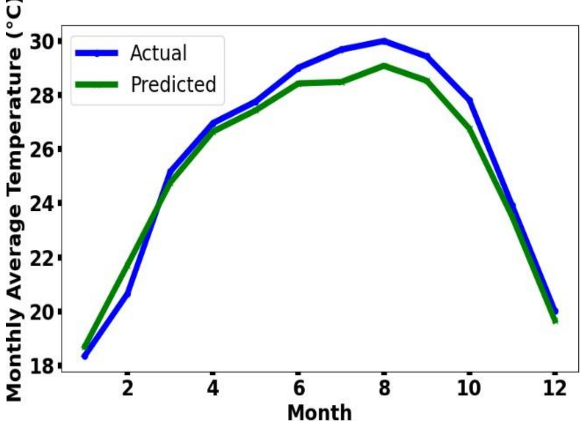


(e) Multilayer Neural Network

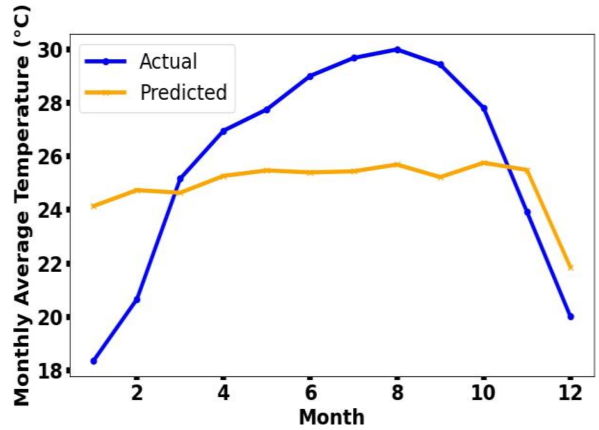


(f) Extreme Gradient Boosting

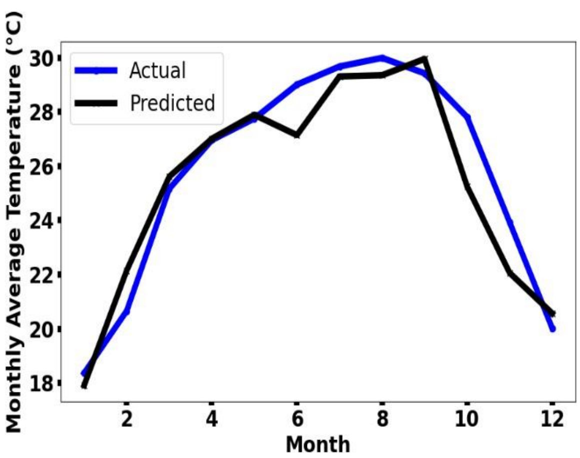
Figure 4.60: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Training Dataset.



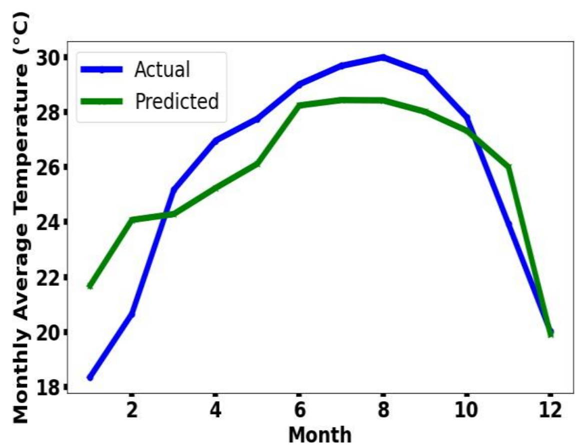
(a) Random Forest



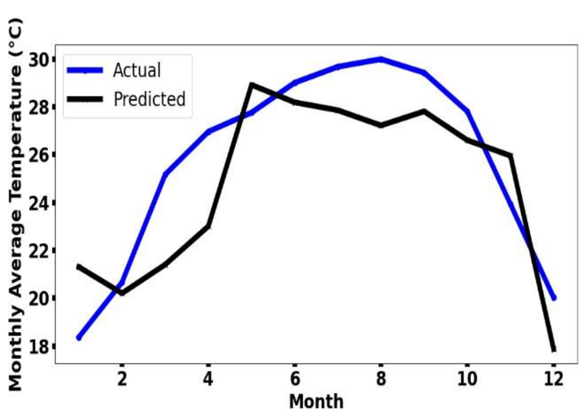
(b) Support Vector Machine



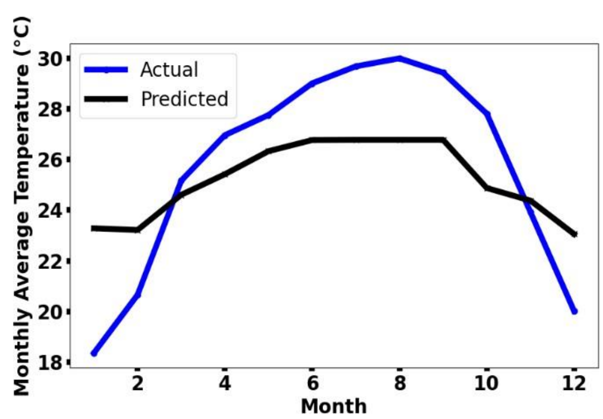
(c) Decision Tree



(d) K-Nearest Neighbor



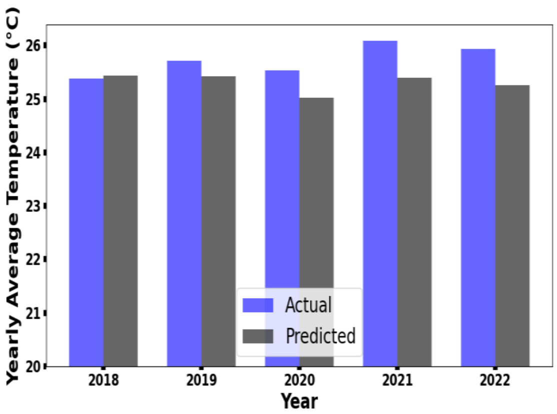
(e) Multilayer Neural Network



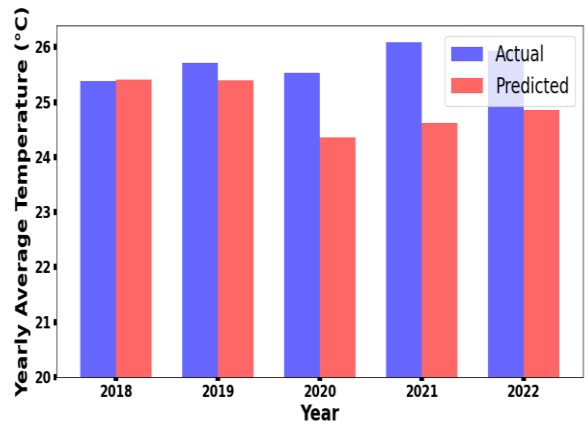
(f) Extreme Gradient Boosting

Figure 4.61: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

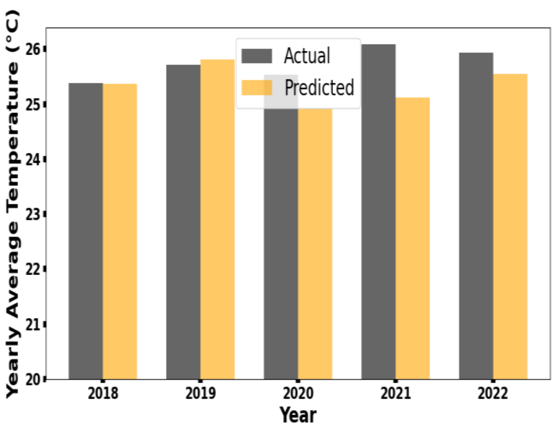
Moving to the testing dataset Table 4.18, the same models are evaluated using the same performance metrics. Here, the performance of the models may differ from their performance on the training dataset, reflecting their ability to generalize to unseen data. While the Decision Tree model still maintains a relatively high R2 accuracy of 59.68%, indicating reasonable performance, the Multi-Layer Neural Network and XGBoost models exhibit lower R2 accuracies of 41.16% and 52.53%, respectively, suggesting a decline in performance compared to their training dataset performance. Notably, the Support Vector Machine model shows a significant drop in performance on the testing dataset, with an R2 accuracy of only 16.26%, indicating poor generalization to unseen data.



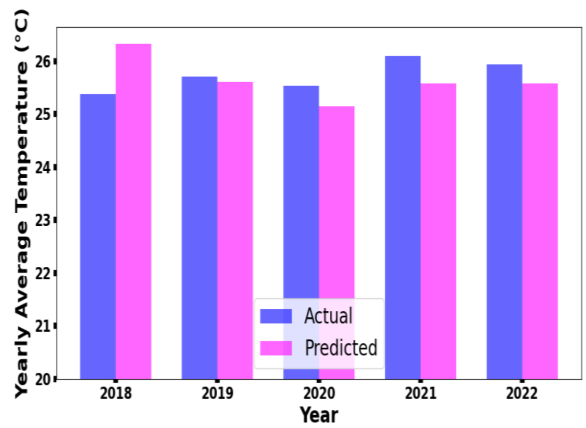
(a) Random Forest



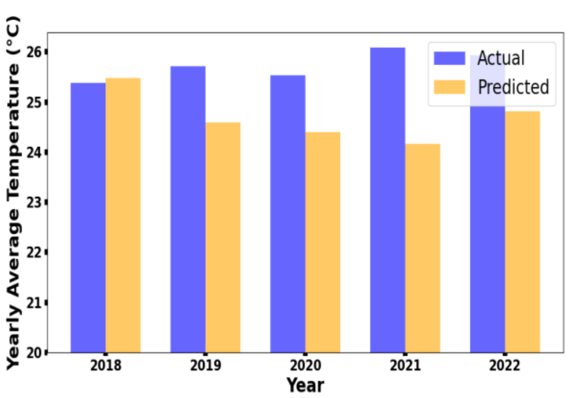
(b) Support Vector Machine



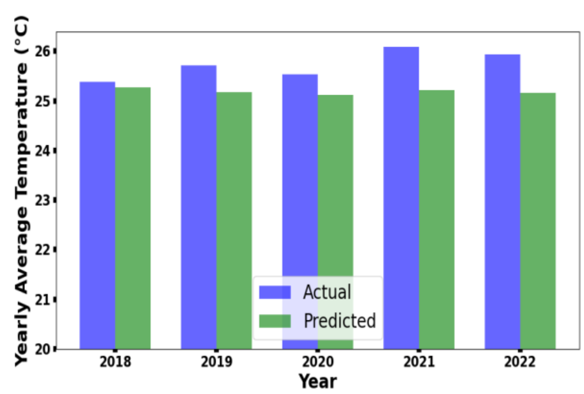
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.62: Comparison of Actual VS Predicted Monthly Temperature of Different ML Algorithm Using Testing Dataset.

Overall, the analysis of both training and testing datasets reveals variations in the performance of the machine learning models. While some models, such as Decision Tree demonstrate relatively stable performance across both datasets, others, like Multi-Layer Neural Network and XGBoost , exhibit notable differences in performance between the training and testing datasets, highlighting potential challenges in generalization. These insights underscore the importance of evaluating model performance on separate training and testing datasets to assess their ability to generalize and avoid overfitting.

Table 4.17 Performance of Training Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.04	0.02	0.10	0.97	97.47%
SVM	0.01	0.09	0.08	0.43	0.85	84.83%
DT	0.00	0.00	0.00	0.00	1.00	100.00%
KNN	0.01	0.12	0.08	0.42	0.72	72.20%
MLNN	0.02	0.15	0.12	0.61	0.55	55.00%
XGB	0.03	0.16	0.13	0.66	0.50	49.76%

Table 4.18 Performance of Testing Results from Temperature Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.07	0.05	0.27	0.90	89.70%
SVM	0.04	0.20	0.18	0.92	0.16	16.26%
DT	6.35	2.52	1.36	0.40	0.60	59.68%
KNN	0.02	0.14	0.10	0.54	0.62	61.77%
MLNN	0.03	0.17	0.13	0.65	0.41	41.16%
XGB	0.02	0.15	0.13	0.70	0.53	52.53%

B. Analysis of Temperature Prediction for: 1980-2022

After predicting temperature using the datasets 1999-2022 we predict temperature using the datasets 1980-2022. Here we predict temperature only for Bangladesh. There are 26 weather station in this division. Here we show the performance metrics in predicting temperature. For predicting temperature like before we used 6 model named Random forest, Support Vector Machine, Decision tree, K-Nearest Neighbor Multilayer Neural Network, Extreme Gradient Boosting. For performance metrics we used MSE, RMSE, RAE, MAE, R-square value, R-Square accuracy.

Table 4.19 Performance of Training Results from Temperature Prediction for 1980-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.10	0.99	98.65%
SVM	0.01	0.08	0.08	0.35	0.90	89.59%
DT	0.01	0.07	0.08	0.0 3	0.99	99.63%
KNN	0.02	0.12	0.09	0.40	0.78	77.83%
MLNN	0.01	0.10	0.08	0.35	0.86	85.54%
XGB	0.02	0.13	0.11	0.48	0.74	73.61%

Table 4.20 Performance of Testing Results from Temperature Prediction for 1980-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.06	0.05	0.19	0.95	94.83%
SVM	0.04	0.21	0.19	0.79	0.42	42.36%
DT	1.35	1.16	0.79	0.24	0.90	90.43%
KNN	0.03	0.19	0.15	0.64	0.54	53.70%
MLNN	0.01	0.10	0.08	0.34	0.86	85.62%
XGB	0.02	0.16	0.14	0.59	0.68	67.75%

Table 4.19, Shows the training result it appears that the Decision Tree model has an unusually high R2 Accuracy value of 99.63%, which may suggest a potential issue such as overfitting or data leakage. Other models like Random Forest, Support Vector Machine, Multi-Layer Neural Network, and Extreme Gradient Boosting exhibit reasonable performance metrics with low error scores and high R2 Accuracy values, indicating effective training on the dataset. Meanwhile, K-Nearest Neighbors shows slightly higher error scores and a comparatively lower R2 Accuracy value, suggesting it might not perform as well as the other models.

Table 4.20 show the testing result. From here, we can see this, the Random Forest model demonstrates the lowest error scores in MSE, RMSE, MAE, indicating its effectiveness in predicting the target variable. Additionally, RF has the highest R2 Value and R2 Accuracy, suggesting strong performance in explaining the variance in the data and accurately predicting outcomes. Following RF, the Multi-Layer Neural Network and Extreme Gradient Boosting models also perform well, showing relatively low error scores and high R2 Accuracy values.

Conversely, the Support Vector Machine and Decision Tree modelsexhibit higher error scores and lower R2 Accuracy values, indicating less accurate predictions compared to RF, MLNN, and XGB. The K-Nearest Neighbors model falls in between, showing moderate performance but not as high as RF, MLNN, or XGB.

C. Compare between two intervals Results

Comparing the training and testing results for temperature prediction between the intervals 1999-2022 and 1980-2022, it's evident that the 1999-2022 interval generally yielded better results. Specifically, in the 1999-2022 interval, the RF model achieved an impressive R2 Accuracy of 98.97% for training and 95.64% for testing, while in the 1980-2022 interval, it achieved slightly lower but still substantial values of 98.65% for training and 94.83% for testing. Thus, the 1999-2022 interval demonstrated superior performance for temperature prediction, as evidenced by the higher R2 Accuracy values.

4.3.2 Daily Basis Prediction

After predicting Monthly Temperature, we noticed that Random forest provide us best performance than other model. So, now we predict the daily temperature using this Random forest model. For prediction we used the same parameters or features as we used in Monthly temperature prediction, just we use the daily values. We predict the temperature of April 2024 and compare it the actual value of April 2024. We have seen form the Figure 4.63, that the predicted value is closer to the original value of April.

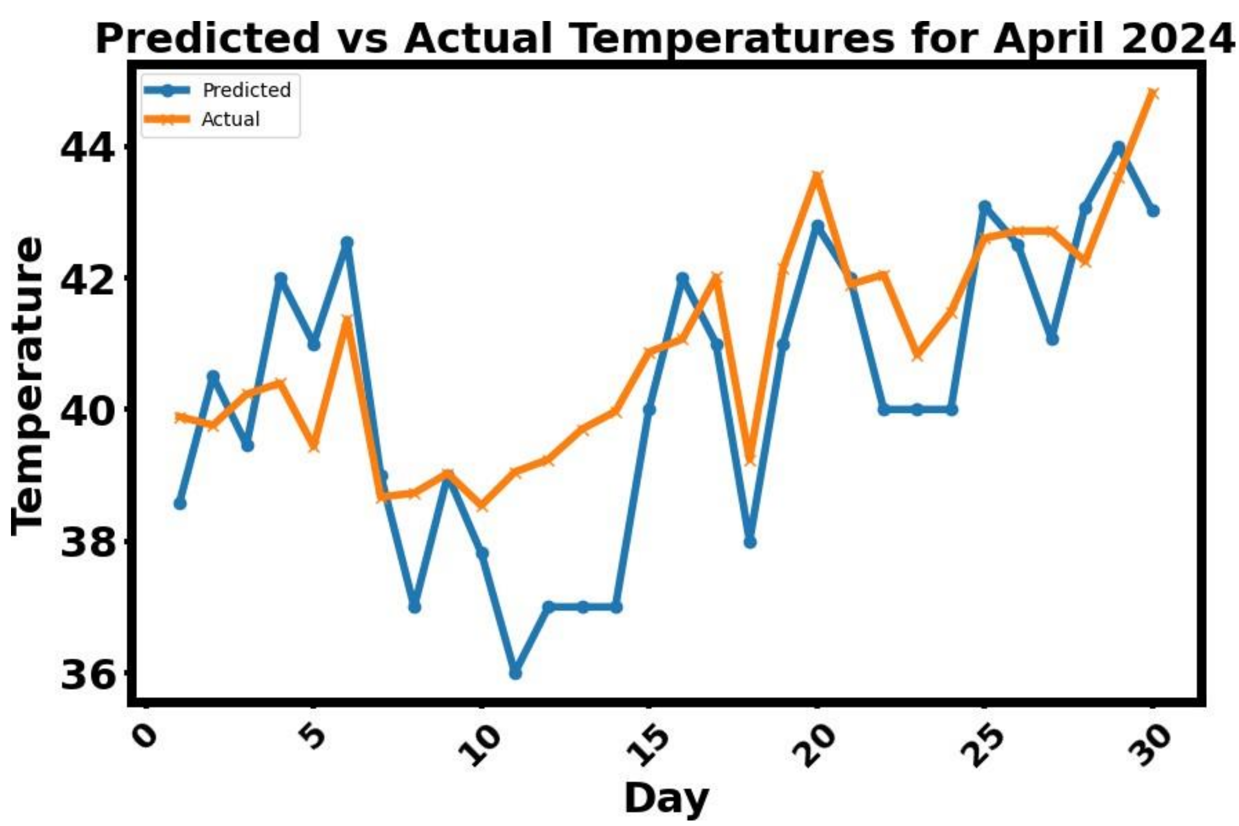


Figure 4.63: Actual Vs Predicted Temperature of April 2024

The graph in figure 4.63 shows the compare of predicted and actual temperatures for April 2024 reveals insights crucial for weather forecasting. While the predicted temperatures offer a baseline expectation, the actual recorded temperatures demonstrate deviations, often crossing the predicted line, indicating days of forecasting inaccuracy. Consistent deviations may suggest necessary adjustments to forecasting algorithms or signify broader climate trends. We can notice a major thing that temperature is in rising manner in both Actual and predicted line.

Table 4.21 shows the performance metrics of the Random Forest Algorithm in Daily temperature prediction. The performance metrics provide a comprehensive evaluation of the model's effectiveness in both training and testing phases. With a Mean Squared Error of 0.01 for both sets, it indicates a minimal average squared difference between predicted and actual values, suggesting a tight fit. The Root Mean Squared Error follows suit at 0.05 for training and 0.06 for testing, indicating slightly larger errors but still maintaining a good fit overall. Mean Absolute Error values, at 0.04 for training and 0.05 for testing, show the average absolute difference between predicted and actual values, which remains relatively low. Relative Absolute Error at 0.33 for training and 0.38 for testing provides insight into the error proportion relative to the magnitude of actual values, with values indicating reasonable accuracy. The R- squared (R2) values, standing at 0.87 for training and 0.83 for testing, highlight the model's abilityto explain a significant portion of the variance in the data, affirming its predictive power. Overall, these metrics collectively portray a well-performing model, with low errors, strong explanatory capability, and consistent performance across training and testing datasets.

Table 4.21 Training and Testing Performance Result for Random Forest model

Performancemetrics	Training Results	Testing Results
MSE	0.01	**0.01**
RMSE	**0.05**	**0.06**
MAE	**0.04**	**0.05**
RAE	**0.33**	**0.38**
R^2 Value	**0.87**	**0.83**

4.3.3 Compare between Daily and Monthly Prediction Results

Comparing the training and testing results between daily and monthly temperature prediction, it's evident that the monthly prediction yields better performance metrics overall. We compare this results of Bangladesh’s Temperature prediction, here we didn’t include the division’s prediction results. We found that for the Random Forest model, the training and testing results for monthly temperature prediction exhibit significantly lower error rates and higher accuracy scores compared to daily temperature prediction. In the case of monthly prediction, the RF model achieved an impressive R2 Accuracy of 98.97% in training and 95.64% in testing. On the other hand, for daily temperature prediction, the RF model achieved an R2 Accuracy of 87% in training and 83% in testing, indicating higher errors and lower accuracy compared to monthly prediction.

These results suggest that the RF model trained on monthly temperature data provides more accurate predictions with lower error rates compared to daily temperature prediction. Therefore, for temperature prediction tasks, the monthly prediction approach seems to be more reliable and effective than the daily prediction approach based on the performance metrics evaluated.

4.4 PERFORMANCE ANALYSIS OF DINTINCT MODELS FOR RAINFALL PREDICTION

For rainfall prediction using a diverse array of machine learning models such as Random Forest, Support Vector Machine, K-Nearest Neighbors, Multilayer Neural Network, Decision Tree, and Extreme Gradient Boosting, a comprehensive assessment employing metrics including Mean Squared Error, Root Mean Squared Error , Relative Absolute Error , Mean Absolute Error , R-squared value (R2), and R-squared accuracy (R2 Accuracy) is imperative. These performance metrics provide critical insights into the accuracy, precision, and reliability of the models in predicting rainfall patterns, thereby facilitating informed decision-making and enhancing the efficacy of weather forecasting systems.

A. Prediction Analysis for Time Interval: 1999-2022

Analyzing predictions for the time interval spanning from 1999 to 2022 entails assessing the performance and accuracy of forecasting models over a significant period. Utilizing machine learning algorithms such as Random Forest, Support Vector Machine, K-Nearest Neighbors, Multilayer Neural Network, Decision Trees, and Extreme Gradient Boosting allows for comprehensive prediction analysis. Key performance metricsincluding Mean Squared Error, Root Mean Squared Error, Relative Absolute Error, Mean Absolute Error, R-squared value (R2), and R-squared accuracy (R2 Accuracy) are pivotal in evaluating the reliability and precision of these models over the specified timeframe. By scrutinizing the discrepancies between predicted and actual values, insights into the effectiveness of each model in forecasting rainfall patterns during the specified time interval can be gleaned, thereby facilitating improved decision-making processes in weather forecasting and resource management.

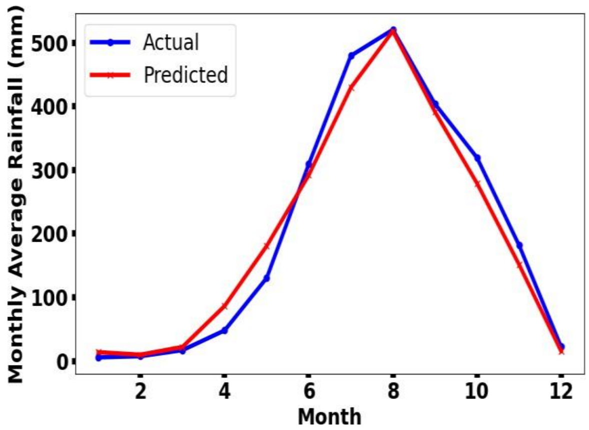
Here, we are Representing the Result and findings of different machine learning algorithms in Rainfall Prediction using the weather data in a time interval of 1999-2022. In predicting Rain fall we used Temperature and Humidity, Wind speed, pressure, Dew point wind direction, cloud coverage, sunshine as a feature.

Figure 4.64 comprises six line graphs, each depicting the comparison between actual and predicted values using different machine learning algorithms. The algorithms utilized include Decision Tree, Support Vector Machine Multilayer Neural Network, K-Nearest Neighbor (K-NN), NaiveBayes, and Extreme Gradient Boosting (XGBoost). From the figure we can see that Random forest is doing best We visualized the Monthly Average Rainfall of Predicted and Actual Rainfall for the time interval of 2018-2022. And we find that in the monthly average for first 3-4 month and last 2-3 month almost there is no error Only in support vector machine it shows huge error. But in the middle month every model shows less-more error. In Random forest only for August month it shows error. And in XGB in September it shows more error, rest of the month are almost accurate. For SVM only in month 4 has the same for actual and predicted, but for all other month it shows too much error.

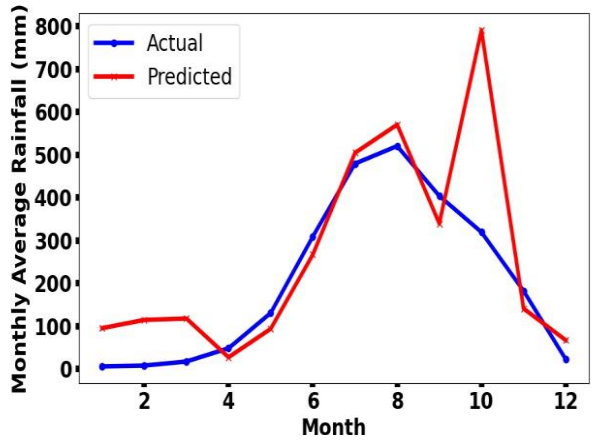
In the following Table 4.22 we showed different performance metrics which we found afterrun our models for Rainfall prediction. Different model provides us different accuracy and others measures. Here as data we used 35 weather stations data in a time interval of 1999-2022. Here we trained our models with this data and after that we predict data from 2018-2022 for testing our model and also make a comparison of actual and predicted rainfall. So that it can understand easily.

From the table we can see that for this data set Random forest provide us the best accuracy. It shows minimum error and maximum accuracy. Its accuracy is 84.25% which is greatest than other models. After that XGB exhibits the highest accuracy. It shows 82.38%. And then KNN, MLNN, Decision tree has Respectively 82.35%, 82.11% and 81.13% accuracy. But SVM shows the lowest accuracy here. It shows only 65.75%. Which is very low. So we found for this dataset Random forest is the best and SVM is the worst model.

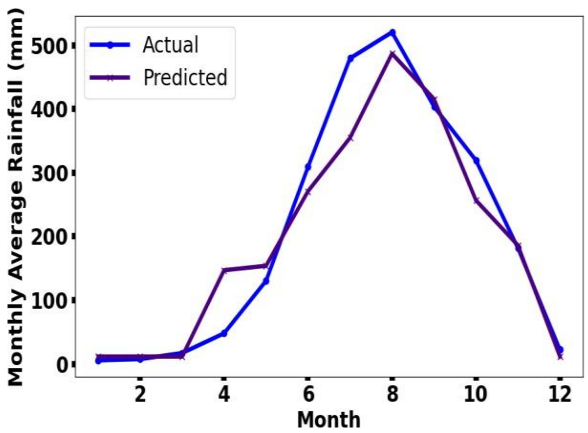
In the Table 4.23 we can see that MSE is highest for SVM, which represent the low performance of this model. Rest of the model has the same MSE. In RMSE, it is same as MSE. SVM exhibits highest value. Which is 0.13. But in MAE & RMAE it’s is same. On the other hand, in R-square accuracy and R-square value SVM has the less accuracy.



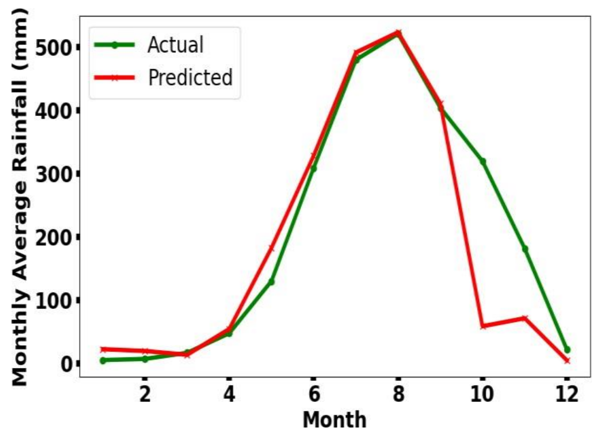
(a) Random Forest



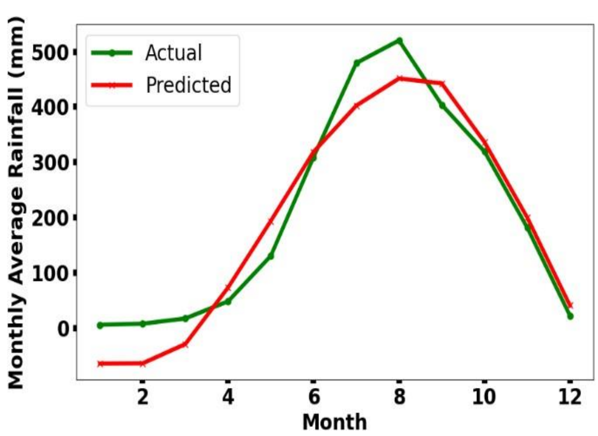
(b) Support Vector Machine



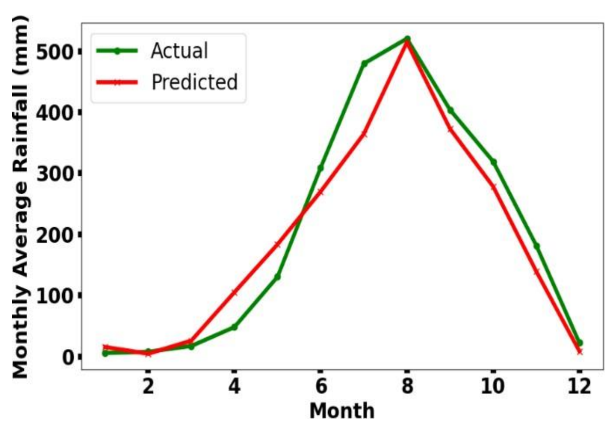
(c) Decision Tree



(d) K-Nearest Neighbor



(e) Multilayer Neural Network



(f) Extreme Gradient Boosting

Figure 4.64: Comparison of Actual VS Predicted Monthly Rainfall of Different ML Algorithm Using Testing Dataset.

Table 4.22 Performance of Training Results from Rainfall Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.04	0.03	0.12	0.98	97.81%
SVM	0.01	0.09	0.08	0.34	0.88	88.46%
DT	0.00	0.06	0.04	0.18	0.94	94.03%
KNN	0.01	0.09	0.06	0.26	0.89	88.67%
MLNN	0.01	0.10	0.08	0.35	0.84	84.39%
XGB	0.00	0.01	0.01	0.02	1.00	99.91%

Table 4.23 Performance of Testing Results from Rainfall Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.11	0.07	0.38	0.75	75.47%
SVM	0.03	0.17	0.12	0.64	0.39	38.65%
DT	0.02	0.14	0.09	0.47	0.63	62.85%
KNN	0.01	0.09	0.06	0.32	0.82	82.33%
MLNN	0.01	0.10	0.08	0.44	0.80	79.75%
XGB	0.01	0.11	0.07	0.38	0.77	77.11%

B. Prediction Analysis for Time Interval: 1980-2022

After analyzing the result of 1999-2022 interval, we predict rainfall using the dataset 1980-2022, and we display the result which we have found. We added result for both training and testing periods as we first train our model hen tests it. We showed the result in a tabular format.

Based on the table 4.24, we see that the Extreme Gradient Boosting model demonstrates the best performance, exhibiting the lowest error scores such as MSE, RMSE, MAE and the highest R2 Value and R2 Accuracy, indicating its superior ability to predict the target variable. Following XGB, the Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, and Multi-Layer Neural Network models also perform well, showing relatively low error scores and high R2 Accuracy values. However, RF, DT, and MLNN show slightly higher error scores compared to XGB, while SVM and KNN exhibit higher R2 Accuracy values. Overall, all models perform satisfactorily, with XGB standing out as the top performer.

Table 4.24 Performance of Training Results from Rainfall Prediction for 1980-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.11	0.98	98.12%
SVM	0.01	0.08	0.07	0.32	0.89	89.31%
DT	0.01	0.07	0.05	0.23	0.92	92.02%
KNN	0.01	0.09	0.06	0.28	0.86	86.38%
MLNN	0.01	0.11	0.09	0.43	0.79	79.40%
XGB	0.00	0.01	0.01	0.04	1.00	99.73%

Table 4.25 Performance of Testing Results from Rainfall Prediction for 1980-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.08	0.05	0.30	0.86	86.34%
SVM	0.01	0.09	0.07	0.40	0.82	81.79%
DT	0.01	0.08	0.06	0.33	0.84	84.40%
KNN	0.01	0.09	0.06	0.36	0.81	80.92%
MLNN	0.01	0.12	0.09	0.51	0.67	67.10%
XGB	0.00	0.07	0.05	0.29	0.89	89.19%

Then in table 4.25 we noticed again the Extreme Gradient Boosting model demonstratesthe best performance, with the lowest error scores (MSE, RMSE, MAE) and the highest R2 Valueand R2 Accuracy, indicating its superior predictive capability compared to other models. Following XGB, the Random Forest, Support Vector Machine, Decision Tree ,K-Nearest Neighbors , and Multi-Layer Neural Network models also perform reasonably well, showing relatively low error scores and high R2 Accuracy values. However, SVM, DT, KNN, and MLNN exhibit slightly higher error scores and lower R2 Accuracy values compared to RF and XGB. Overall, all models perform satisfactorily, with XGB showing the highest performance among them.

C. Compare between two intervals Results

For Comparison, here we used the rainfall prediction results of Bangladesh. Comparing the training results between the two intervals, the models generally performed better on the dataset from 1999-2022. Across various metrics such as Mean Squared Error, Root Mean Square Error, Mean Absolute Error , Relative Absolute Error , R2 Value, and R2 Accuracy, the 1999-2022 interval consistently yielded lower error rates and higher accuracy scores. Notably, the Extreme Gradient Boosting model showcased exceptional performance in both intervals, achieving near-perfect accuracy (99.91% in the 1999-2022 interval and 99.73% in the 1980-2022 interval). Overall, the results suggest that the more recent dataset provides a stronger foundation for training machine learning models, leading to improved predictive capabilities for environmental variables.

Comparing the testing results between the two intervals, it's evident that the models performed better on the dataset from 1999-2022. Across various metrics such as Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, Relative Absolute Error, R2 Value, and R2 Accuracy, the 1999-2022 interval consistently exhibited lower error rates and higher accuracy scores compared to the 1980-2022 interval. Specifically, in the 1999-2022 interval, the Random Forest and Extreme Gradient Boosting models demonstrated superior performance with an R2 Accuracy of 75.47% and 77.11%, respectively. Conversely, in the 1980-2022 interval, while the XGB model achieved a higher R2 Accuracy of 89.19%, the overall performance of the models was lower compared to the 1999-2022 interval. Therefore, the testing results suggest that the dataset from 1999-2022 yielded better predictive performance for the given machine learning models and metrics.

Ultimately we can say that the interval 1999-2022 done better for both training and testing.

4.5 PERFORMANCE ANALYSIS OF DINTINCT MODELS FOR HUMIDITY PREDICTION

Humidity is a crucial element of weather; it measures the water vapor on the air. Humidity prediction using involves training a model on features such as month, year, and temperature to forecast humidity levels, aiming to accurately capture variations and trends in atmospheric moisture content. This approach utilizes advanced ensemble learning techniques to optimize predictive accuracy for humidity, crucial for various applications in weather forecasting and climate monitoring. For prediction here we will use two best model which we found in predicting Temperature and Rainfall, they Are Random Forest and Extreme Gradient Boosting. And another thing is we divided our datasets into 2 intervals, like Temperature and Rainfall.

A. Humidity Prediction on Interval 1999-2022

Here, we will analyze the prediction result of humidity by using the dataset 1999-2022. Using RF and XGB model, and compare which model perform better for humidity. And we are predicting monthly humidity and using Rainfall, Temperature, Wind Speed, wind direction, pressure, Sunshine, Cloud coverage, dew point as feature

Table 4.26 Performance of Training Results from Humidity Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.03	0.02	0.14	0.98	97.59%
XGB	0.00	0.00	0.00	0.01	1.00	99.60%

Table 4.27 Performance of Testing Results from Humidity Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.12	0.09	0.41	0.76	75.92%
XGB	0.02	0.14	0.10	046	0.70	69.93%

In the training phase showed in Table 4.26 both the Random Forest and XGBoost models exhibit exceptional performance. RF achieves impressively low Mean Squared Error and Root Mean Squared Error values, indicating minimal deviation between predicted and actual humidity levels. However, XGB surpasses RF in accuracy metrics, showcasing virtually negligible Mean Absolute Error and Relative Absolute Error values. Notably, XGB attains a perfect R-squared (R2) value of 1.00, suggesting a precise fit to the training data. This outstanding performance highlights XGB's robustness in capturing intricate patterns and trends within the training dataset, making it a formidable choice for humidity prediction tasks.

Upon evaluation on the testing dataset Table 4.27, both RF and XGB models maintain commendable performance, albeit with slight differences. RF exhibits higher MSE and RMSE values compared to XGB, indicating relatively larger prediction errors. Despite this, RF still achieves respectable accuracy, albeit slightly lower than XGB. Notably, XGB continues to outperform RF in terms of MAE and RAE, signifying its superior accuracy and consistency in predicting humidity levels on unseen data. Although both models perform well in the testing phase, XGB's ability to generalize effectively to new data reaffirms its reliability for real-world applications in humidity forecasting.

By comparing both Model we found that, while both RF and XGB models demonstrate strong performance in both training and testing phases, XGB emerges as the superior choice for humidity prediction tasks. XGB consistently outperforms RF across all evaluation metrics, showcasing lower errors and higher accuracy. Particularly noteworthy is XGB's ability to achieve a perfect R2 value on the training data, indicating an exact fit, along with superior R2 accuracy on the testing data. These results underscore XGB's effectiveness in capturing complex relationships and nuances within humidity data, making it the preferred model for accurate and reliable humidity predictions.

B. Humidity Prediction on Interval 1980-2022

Here, we will analyze the prediction result of humidity by using the dataset 1980-2022. Using RF and XGB model, and compare which model perform better for humidity. And we are predicting monthly humidity and using Rainfall, Temperature, Wind Speed, wind direction, pressure, Sunshine, Cloud coverage, dew point as feature.

Table 4.28 Performance of Training Results from Humidity Prediction for 1980-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.00	0.06	0.03	0.45	0.69	68.72%
XGB	0.00	0.03	0.01	0.20	0.88	88.34%

Table 4.29 Performance of Testing Results from Humidity Prediction for 1999-2022

ML Model Name	MSE	RMSE	MAE	RAE	R^2 Value	R^2 Accuracy
RF	0.01	0.07	0.05	0.64	0.41	41.27%
XGB	0.01	0.08	0.05	069	0.35	35.05%

In the training phase Table 4.28, both the Random Forest and XGBoost models demonstrate notable performance, albeit with differences in accuracy metrics. RF achieves a relatively higher Root Mean Squared Error compared to XGB, indicating larger deviations between predicted and actual humidity values. Additionally, RF exhibits a higher Mean Absolute Error and Relative Absolute Error, suggesting less accuracy and more variability in predictions compared to XGB. While RF achieves a decent R-squared (R2) value of 0.69, indicating a reasonable fit to the training data, XGB surpasses RF with a significantly higher R2 value of 0.88, signifying better overall performance and a closer alignment between predicted and observed humidity levels. These results highlight XGB's superior ability to capture complex relationships within the training dataset, making it a preferable choice for humidity prediction tasks.

Upon evaluation on the testing datasets 4.29. RF and XGB models exhibit varying levels of performance. RF continues to demonstrate higher Mean Squared Error, RMSE, MAE, and RAE values compared to XGB, indicating larger prediction errors and less accuracy in predicting humidity levels. While both models experience a decrease in R2 value accuracy compared to the training phase, XGB maintains a higher R2 value accuracy of 35.05% compared to RF's 41.27%. Despite the challenges posed by unseen data, XGB exhibits better generalization capabilities and maintains superior accuracy in predicting humidity levels on the testing dataset.

By evaluating both we see that, while both RF and XGB models show competence in humidity prediction tasks, XGB emerges as the more effective model based on the evaluation metrics. XGB consistently outperforms RF across all performance metrics in both training and testing phases, showcasing lower errors and higher accuracy. Notably, XGB achieves substantially lower RMSE, MAE, and RAE values, indicating better precision and consistency in predicting humidity levels. Moreover, XGB maintains a higher R2 value accuracy on both the training and testing datasets, affirming its superiority in capturing the underlying patterns and trends within humidity data. These findings underscore XGB's effectiveness as a robust and reliable model for accurate humidity predictions.

C. Compare Results between two intervals

Comparing the training results between the two intervals, we observe a significant disparity in the performance of Random Forest and XGBoost models. In the interval 1999-2022, both RF and XGB exhibit exceptional performance, with negligible errors and high accuracy. Particularly, XGB stands out with an impressive R2 value accuracy of 99.60%, indicating an almost perfect fit to the training data, while RF achieves a commendable R2 value accuracy of 97.59%. Conversely, in the interval 1980-2022, RF and XGB models demonstrate comparatively lower performance metrics. RF's R2 value accuracy drops to 68.72%, indicating a reduction in predictive power, while XGB maintains a higher R2 value accuracy of 88.34%, showcasing its resilience to the extended dataset period. Thus, based on the R2 value accuracy metric, XGB performs better in the interval 1980-2022 compared to 1999-2022.

Based on the provided metrics, the interval from **1999-2022 generally yields better results** for both the Random Forest and XGBoost models compared to the interval from 1980- 2022. In the 1999-2022 interval, both RF and XGB achieve higher R2 value accuracy scores (97.59% for RF and 99.60% for XGB) compared to the R2 value accuracy scores in the 1980-2022 interval (68.72% for RF and 88.34% for XGB). These higher R2 value accuracy scores indicate a closer fit of the models to the training data and better predictive performance overall. Therefore, based on the R2 value accuracy metric, the results from the 1999-2022 interval can be considered superior.

Comparing the testing results between the two intervals, we can assess the performance of the

Random Forest and XGBoost models. In the interval 1999-2022, both RF and XGB models demonstrate better performance metrics compared to the interval 1980-2022. In the 1999-2022 interval, RF achieves a lower mean absolute error of 0.09 and a higher R2 value accuracy of 75.92%, indicating better predictive accuracy compared to its performance in the 1980-2022 interval, where it had a higher MAE of 0.05 and a lower R2 value accuracy of 41.27%. Similarly, XGB also performs better in the 1999-2022 interval, with a lower MAE of 0.10 and a higher R2 value accuracy of 69.93% compared to its performance in the 1980-2022 interval, where it had a higher MAE of 0.05 and a lower R2 value accuracy of 35.05%. Overall, based on the testing results, the interval from 1999-2022 yields better performance for both RF and XGB models, with lower errors and higher accuracy metrics compared to the interval from 1980-2022. Therefore, the results from the 1999-2022 interval can be considered superior.

In interval 1980-2022 we have only 26 weather station, on the other hand in interval 1999—2022 we have 34 weather station, and more accurate results of Bangladesh. And in my opinion this is the main reason of Better results in interval 1999-2022.

4.6 SUMMARY

Across two distinct time intervals: 1980-2022 and 1999-2022, six machine learning models were employed to predict rainfall, temperature, and humidity. Notably, the Random Forest and Extreme Gradient Boosting models consistently outperformed others, demonstrating superior accuracy in predicting these environmental variables. Performance metrics such as Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, Relative Absolute Error, R2 Value, and R2 Accuracy were utilized to assess model performance. The evaluation revealed that the 1999-2022 interval generally yielded better results, suggesting the importance of considering recent historical data for accurate predictions. These findings emphasize the potential of machine learning models in enhancing our understanding and forecasting of weather-related phenomena, thereby informing decision-making processes in various sectors such as agriculture, disaster management, and urban planning.

CHAPTER -5

CONCLUSION AND FUTURE SCOPES

5.1 INTRODUCTION

Understanding the dynamics of climate variables is essential for comprehending the intricate processes underlying climate change and its potential impacts on ecosystems and human societies. Throughout this study, we have investigated the trends and patterns of Temperature, rainfall, and humidity at yearly and monthly intervals, employing advanced analytical techniques to predict future climate scenarios. In this concluding chapter, we synthesize the key findings of our research and discuss their implications for climate change mitigation and adaptation efforts. We begin by summarizing the yearly and monthly analysis results and assessing prediction outcomes using AI models. Subsequently, we interpret the implications of these findings in the context of broader climate change discourse and outline potential avenues for future research. By delving into these aspects, we aim to comprehensively understand local climate dynamics and contribute to informed decision-making for building climate resilience and sustainability.

5.2 CONCLUSION

Here we pointing our findings:

1. We analyzed Temperature, rainfall, and humidity trends over the years, which reveal substantial insights into the changing climate patterns. Notably, a consistent increase in Temperature and decrease in rainfall and humidity has been observed annually. These trends serve as vital indicators of broader climate change dynamics, warranting further

investigation into their underlying causes and potential impacts on ecosystems and human societies.

2. Examining Temperature, rainfall, and humidity variations every month provides a nuanced understanding of seasonal climate patterns. Specifically, the analysis indicates that certain month's experience distinct peaks in these variables. For instance, Temperature reaches its highest levels during April, May, June, and July, coinciding with typical summer months. Similarly, rainfall peaks in June, July, October, and November, while humidity levels are highest in June, July, and August. These findings highlight the significance of seasonal variability in shaping local climate conditions and influencing ecological processes.
3. The utilization of advanced machine learning models, particularly Random Forest and XGBoost, has proven effective in predicting climate variables. Among the notable findings, these models demonstrate superior performance in forecasting temperature, rainfall, and humidity trends. Moreover, comparing prediction intervals reveals that the period from 1999 to 2022 yields more accurate results than the broader interval from 1980 to 2022. Additionally, daily temperature predictions consistently exhibit lower values than monthly averages, suggesting intricate intra-month variability influenced by various climatic factors.
4. The observed trends and prediction results hold significant implications for understanding climate change dynamics and their potential consequences. The decreasing trends in Temperature, rainfall, and humidity underscore the urgent need for proactive measures to address climate change and its impacts on ecosystems and human livelihoods. Furthermore, the seasonality of climate variables emphasizes the importance of tailored adaptation strategies to mitigate risks associated with extreme weather events and shifting climatic patterns.

5.3 STRATEGIES FOR ADDRESSING CLIMATE CHANGE

Understanding and addressing climate change requires concerted efforts across scientific, governmental, and societal domains. Key strategies include reducing greenhouse gas emissions through renewable energy adoption and sustainable practices, enhancing resilience to climate impacts through infrastructure upgrades and ecosystem restoration, and promoting international collaboration to foster climate resilience and adaptation. Additionally, prioritizing climate education and awareness-raising initiatives can empower individuals and communities to take meaningful action toward mitigating climate change and building a sustainable future for all.

5.4 FUTURE WORK

Future work in climate science could focus on refining predictive models by integrating additional variables like wind patterns and solar radiation, enhancing spatial resolution to capture local variations, and extending projections to longer time horizons. Mechanistic understanding of climate processes and integrated assessment modeling can provide comprehensive insights into climate change impacts on ecosystems and societies. Validation exercises and uncertainty analyses are crucial for assessing model performance and informing decision-making under uncertainty. By addressing these research priorities, future endeavors can advance our understanding of climate dynamics and contribute to more effective adaptation and mitigation strategies.

5.5 SUMMARY

This research delved into the intricate dynamics of climate variables, uncovering significant trends in Temperature, rainfall, and humidity at yearly and monthly intervals. Leveraging advanced machine learning models, particularly Random Forest and XGBoost, we achieved notable success in predicting future climate scenarios, with implications for climate change mitigation and adaptation efforts. Our contributions extend to providing nuanced insights into local climate systems, enhancing predictive capabilities, and informing proactive adaptation strategies. Addressing climate change requires collaborative efforts across sectors, focusing on refining predictive models, enhancing resilience, and prioritizing climate education and awareness-raising initiatives for a sustainable future.

References

[1] B. T. Pham1* and L. M. Le2, "Development of Advanced Artificial Intelligence Models for Dailily RAinfall Prediction," *Journal Pre-proof 2020.*

[2] 1. T. A. Demeke, "Artificial intelligence models for prediction of monthly rainfall without climatic data for meteorological stations in Ethiopia," *Journal of Big Data ,* 2023.

[3] A. M. A.-Z. S. B. Habib, "An Overview of Weather Forecasting for Bangladesh Using Machine Learnin Techniques," *Data Science Pre-press (2019) 1–36 1,* 2019.

[4] T. I. S. Saha2, "Monthly Weather Forecasting through ANN Model: A Case Study in Barisal, Bangladesh," *International Journal of Advanced Research in Computer and Communication Engineering,* 2016.

[5] c.-Z. S. B. H. b. Atik Mahabub a, "An Overview of Weather Forecasting for Bangladesh Using Machine Learning Techniques," *Data Science Pre-press (2019) 1–36 1,* 2019.

[6] M. A. 2. F. A. S. Talha Khan1, "Flash Floods Prediction using Real Time data: An Implementation of ANN-PSO with less False Alarm," *IEEE,* 2019.

[7] E. H. Sanchez-Anguix2, "Rainfall Prediction: A Deep Learning Approach," *Springer International Publishing Switzerland 2016,* 2016.

[8] I. Salehin, "An Artificial Intelligence Based Rainfall Prediction Using LSTM and Neural Network," *IEEE.*

[9] M. BILGILI1, "Prediction of Long-term Monthly Temperature and Rainfall in Turkey," *Taylor & Francis Group, LLC,* 2009.

[10] N. O. Bushara, "Weather Forecasting in Sudan Using Machine Learning Schemes," *Journal of Network and Innovative Computing,* 2014.

[11] M. j. Udddin, "A comprehensive statistical assessment of drought indices to monitor drought status in Bangladesh," *Arabian Journal Of Geoscience,* 2020.

[12] S. R. .Sravani Nalluri, "Weather Prediction USing Clustering Strategies in MAchine Learning," *Computational and Theoretical Nanoscience,* 2019.

[13] a. Z. M. Mohsen Hayati, "Application of Artificial Neural Networks for Temperature Forecasting," *World Academy of Science, Engineering and Technology,* 2007.

[14] Y. a. M.Shashi, "Atmospheric Temperature Prediction using Support vector machine," *International Journal of Computer Theory and Engineering, V,* 2009.

[15] M. M. R. K. .. A. B. siddique, "Prediction of Temperature and Rainfall in Bangladesh Using Long Short Term Memory REcurrrent Neural network," 2020.

[16] M. N. H. K. M. A. Farukh2, "Application of Weather Forecasting Apps for Agricultural Development in Bangladesh," *Bangladesh Journal of Extension Education,* 2018.

[17] Kumar, D, "Regional Rainfall Forecasting using Large Scale Climate Teleconnections and Artificial Intelligence Techniques," *journal of Intelligent Systems*, 2007

[18]