MACHINE LEARNING BASED PREDICTION OF DENGUE RISK ZONES IN BANGLADESH BASED ON WEATHER DATA

BY

K. M RIYAD AHMED
ID:201-15-13937

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

Supervised By

Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised By

Amir Sohel
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project titled "**Machine learning based Prediction Of Dengue Risk Zones In Bangladesh Based On Weather Data**", submitted by K. M Riyad Ahmed, ID No:201-15-13937 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 23/01/2024.

BOARD OF EXAMINERS

Dr. Md. Zahid Hasan (ZH) **Chairman**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Abdus Sattar (AS) **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Tapasy Rabeya (TRA) **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shahidur Rahman (DMSR) **External Examiner**
Professor
Department of Computer Science and Engineering
Shahjalal University of Science and Technology

DECLARATION

I, therefore, declare that this undertaking has been finished by us under the supervision of **Narayan Ranjan Chakraborty,** Associate Professor and Associate Head, Department of CSE, Daffodil International University. I further declare that neither an application or an educational grant has been made anywhere for this project or any part of it.

Supervised by:

Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:

20/01/24
Amir Sohel
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Riyad 23/01/24
K. M Riyad Ahmed
ID:201-15-13937
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project successfully.

I am really grateful and wish my profound indebtedness to **Narayan Ranjan Chakraborty**, Associate Professor and Associate Head, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "Machine Learning" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to Professor **Dr. Sheak Rashed Haider Noori,** Head**,** Department of CSE, for his kind help to finish my project and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank all of my classmates at Daffodil International University who participated in the discussion while also attending the course.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

This work proposes a unique method for identifying dengue risk zones in Bangladesh using machine learning algorithms and large weather data. Dengue fever, a mosquito-borne illness that is common in tropical areas, has a complicated interplay with climate factors. Using GNB, Random Forest Classifier, Decision Tree Classifier, and Voting Classifier, Our Machine learning model uses Gradient Boosting Classifier and Logistic Regression to uncover subtle patterns in rain, temperature, and oxygen. The study combines past weather data with reported dengue cases, employing a number of machine learning methods to determine connections between environmental variables and illness incidence. Our algorithm delivers nuanced risk evaluations by applying a complex ensemble of classifiers, classifying regions as "High Risk," "No Risk," "Low Risk," and "Moderate Risk." It allows for focused public health interventions, more effective use of resources, and proactive dengue epidemic control. The proposed machine learning-based prediction model not only tackles the current threat of dengue in Bangladesh, but it also serves as a tough tool that can be adapted to changing climatic dynamics. This study adds to the larger conversation on the connection of data science and public health by providing an adjustable and dynamic framework for minimising the effect of vector-borne illnesses in climate-vulnerable areas.

Keywords: Dengue prediction, Machine learning, Public health, Climate impact, GaussianNB, RandomForestClassifier, DecisionTreeClassifier, VotingClassifier, GradientBoostingClassifier, LogisticRegression.

.

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Dengue fever is an important health issue across the world, having a particular impact in tropical areas such as Bangladesh. Because of the dynamic character of this mosquito-borne illness, which is impacted by a number of environmental conditions, new approaches to prediction and prevention are required. This paper presents a Machine Learning-based method for predicting dengue risk zones in Bangladesh using long weather data.[1] Machine Learning algorithms combined with weather data offer a viable route for improving the accuracy and location quality of dengue risk estimates. The model seeks for subtle patterns in weather conditions such as temperature, precipitation, and humidity by using the strength of techniques such as GaussianNB, Random Forest Classifier, Decision Tree Classifier, Voting Classifier, Gradient Boosting Classifier, and Logistic Regression.[2] Machine Learning algorithms combined with weather data offer a potential route for improving accuracy and geographical resolution. Bangladesh's geographical variety and climate variation create a difficult environment for dengue transmission. This prediction model aims to capture the complex links between weather dynamics and dengue incidence, providing a useful tool for active public health measures.[3] A computational approach attempts to identify locations with a range of danger by combining historical meteorological data and reported dengue cases, from high-risk zones requiring rapid care to low-risk regions requiring constant monitoring. The model seeks for small trends in conditions using methods such as GaussianNB, Random Forest Classifier, Decision Tree Classifier, Voting Classifier, Gradient Boosting Classifier, and Logistic Regression.[4] The relevance of this research is in its ability to provide early insights to public health authorities and officials, allowing for early planning of resources and focused preventive actions. As climate change continues to have an influence on virus dynamics, the suggested Machine Learning-based prediction model is a creative solution for minimizing the growing dengue impact in Bangladesh.

1.2 Motivation

The reason for using machine learning to anticipate dengue risk zones in Bangladesh comes from the urgent need to address the region's rising dengue fever danger. Bangladesh, with its tropical temperature and regular monsoons, provides an ideal home for the Aedes mosquito, the vector responsible for transferring the dengue virus. As the frequency and nature of cases of dengue continue to rise, there is an urgent need for novel and proactive disease control measures. Traditional techniques of assessing dengue risk often fail to capture the complicated interaction between environmental variables and disease transmission dynamics. This study aims to overcome the gap by finding complicated patterns inside large meteorological datasets using machine learning techniques.

1.3 Rationale of the Study

The project was motivated by the important shift of public health worries, climate dynamics, and the exciting potential of machine learning technology. Dengue fever, a mosquito-borne infection, places a significant strain on tropical healthcare systems such as Bangladesh. Traditional Dengue projected methods frequently fail to reflect the complicated and changing interactions between the weather and disease spread. Machine learning technology integration provides a compelling option, driven by the demand for more accurate and complex prediction models. The reasoning is based on machine learning algorithms' capacity to determine Complex patterns that can be seen in large-scale meteorological records. Given Bangladesh's different meteorological factors, this strategy provides a method based on data for accurately identifying possible Dengue risk zones. The study's significance comes from its potential to transform public health attempts. The project wants to provide useful information to health authorities and officials by applying machine learning, allowing them to distribute resources more effectively, execute focused treatments, and prioritize preventative measures. Beyond immediate impact, the logic provides a scalable and flexible framework that may be extended to other locations dealing with comparable vector-borne illnesses and climate-related illnesses. Finally, our research serves as a forward-thinking try to improve capacity against Dengue illnesses and adds to the larger conversation about the relationship of technology and public health.

1.4 Research Question

1. How does precipitation affect the risk of Dengue transmission in different parts of Bangladesh?
2. Can machine learning systems detect and predicted high-risk Dengue zones using past meteorological data?
3. What effect does humidity have on the risk of cases of dengue in different districts of Bangladesh?
4. How effectively can the suggested model distinguish between moderate and high-risk Dengue zones, allowing for a more nuanced understanding of the varied degrees of risk?
5. What are the primary weather-related variables that contribute the most to the prediction of Dengue risk zones?

1.5 Expected output

This project is planned to produce a complete and dynamic prediction system for Dengue risk zones in Bangladesh, powered by machine learning algorithms and guided by weather data. The model is expected to produce accurate, timely, and geographically precise outlooks, helping us better understand the complex connection between climatic parameters and Dengue transmission patterns. The output will contain complete risk zone maps that show locations classified as "High Risk," "Moderate Risk," "Low Risk," and "No Risk." Furthermore, the model's output will go over simple projections, providing insights into the main environmental variables that influence Dengue risk. This data is critical for understanding the basic trends and causes that contribute to disease spread. The expected outcomes will allow public health authorities, politicians, and healthcare professionals to efficiently allocate resources, execute targeted interventions, and create preventative measures in locations with variable risk levels. The study is likely to contribute not just to the field of sickness prediction, but also to a larger discussion about how to use modern technology to solve public health concerns upset by climate dynamics. Finally, the outputs are intended to provide partners with practical knowledge, promoting a proactive and resilient approach to Dengue management in Bangladesh.

1.6 Project Management and Finance

Project management and funding are critical to the success and longevity of the machine learning-based prediction project for Dengue risk zones in Bangladesh. Effective project management is required to organize the various elements of the study, from data collection to model building and testing. A well-structured project plan will clearly define tasks, allocate resources wisely, and set milestones to measure progress. To solve difficulties and enhance approaches throughout the project lifetime, regular communication and collaboration among team members, data scientists, domain experts, and customers will be critical. In terms of financing, appropriate funding is crucial for obtaining high-quality datasets, processing resources, and setting up complex machine learning algorithms. The cost of data preparation, model training, and validation, as well as the creation of user-friendly interfaces for communicating results to end users, must be factored into a financial plan. Collaboration with governmental authorities, academic institutes, and international health organizations can help the effort gain access to funds and assure that it will last. Effective project management and financial control will not only assure the research's success, but will also add to its impact and scalability. The project can create the groundwork for future projects tackling public health concerns at the interface of technology, climate, and disease dynamics by remaining transparent and well-organized.

TABLE 1.1: PROJECT MANAGEMENT TABLE

Work	**Time**
Data Collection	1 month
Papers and Articles Review	3 month
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 month
Total	8 month

1.7 Report Layout

- Introduction
- Background
- Data Collection
- Data Preprocessing
- Research Methodology
- Experimental Result and Discussion
- Impact on Society, Environment
- Summary, Conclusion, Future Research
- References

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

The preliminary phase of the machine learning-based prediction project for Dengue risk zones in Bangladesh involves crucial foundational steps essential for the research's success. Firstly, extensive literature review and background research are conducted to understand existing methodologies, challenges, and gaps in predicting Dengue outbreaks. This establishes the groundwork for the project and informs the selection of appropriate machine learning algorithms. The collecting of historical meteorological data and reported Dengue cases in various parts of Bangladesh is a critically needed step. To ensure the availability and quality of varied datasets, this phase includes coordination with weather organisations, health authorities, and relevant customers. Following that comes data preparation, which addresses issues such as missing values, exceptions, and standards. The most significant weather conditions contributing to Dengue risk are identified depending on the features of the dataset and the nature of the prediction job. The next stage also addresses ethical concerns and data privacy issues, providing regulatory compliance and protecting sensitive information. Preliminary meetings with public health officials and community engagement activities are also underway in order to develop collaboration and gather support for the project. This early foundation lays the framework for the succeeding phases, directing the creation of a strong machine learning model for Dengue risk prediction that respects ethical norms, data integrity, and the overall study objectives.

2.2 Related Works

Dey, Samrat Kumar, et al. [1] Using the DengueBD dataset, this study in Bangladesh sought to create a machine learning model for forecasting dengue outbreaks in 11 different districts. Support vector regression (SVR) and multiple linear regression (MLR) were the two techniques used. The study looked into the relationship between environmental elements including temperature, precipitation, and humidity and the trends in dengue cases in different cities. The dataset was divided 80:20 for training and testing, with MLR and SVR each reaching 67% accuracy and a Mean Absolute Error (MAE) of 4.57 and 75% accuracy and an MAE of 4.95, respectively. According to the statistics, Dengue occurrences tend to be lower in the winter and higher in the rainy season, with Dhaka reporting the highest incidence from August 2021 to May 2022. The great potential of machine learning for predicting Dengue epidemics is highlighted by this study, in its conclusion.

Anno, Sumiko, et al. [2] focused on comprehending dengue spread and identifying risk factors, it was possible to address the lack of a dengue vaccine as well as the financial difficulties in its development. The study underlines the necessity for a climate-based model and aims to establish a dengue Early Warning System (EWS). The analysis carried out in Taiwan looks at spatiotemporal dengue fever hotspots at the township level, identifying risk factors using climate data from remote sensing sources. To find variables that correspond with dengue epidemics throughout time and space, machine learning approaches are used. The study uses a deep AlexNet model that was trained using transfer learning on images of sea surface temperature during an 8-fold cross-validation test to identify three dengue fever hotspot categories in southwest Taiwan associated with particular sea surface temperatures. This study achieved an astonishing 100% accuracy in its results.

Sarma, Dhiman, et al. [3] developed a patient dataset for this study in Bangladesh, real-time data samples from a variety of dengue fever patients were used. This study then presented a novel machine learning approach for the prediction of dengue fever. In a ratio of 70:30, the dataset was divided into training and testing sets. Machine learning techniques for categorization were used in the study, including decision trees (DT) and random forests (RF). With an average accuracy of 79% in forecasting dengue disease, the decision tree model distinguished itself from the random forest model. The significance of decision tree-based classification and its potential to improve the accuracy of dengue disease diagnosis are highlighted by this finding, which provides important new information for the study of dengue and for the treatment of dengue patients.

Babu, P. Devendar, et al. [4] increased dengue haemorrhagic fever (DHF) cases, which accounts for about 20,000 annual deaths and mostly affects children under the age of 15, has been caused by the average annual incidence of dengue multiplying by 30 over the past 50 years. With all four serotypes circulating within a decade, dengue is an endemic disease in many tropical areas. Interestingly, the countries with the greatest reported dengue cases are French Guiana, Martinique, Trinidad & Tobago, Puerto Rico, and the Dominican Republic. Because it brings new strains of the virus and nonimmune people to endemic areas, population migration is a major driver in its proliferation. The suggested system uses time series analysis to forecast the spread of dengue illness using climatic data. It also does exploratory data analytics on previous dengue datasets and applies machine learning techniques to prediction analysis.

Panja, Madhurima, et al. [5] predicted dengue epidemics in San Juan, Iquitos, and Ahmedabad, the paper offers the XEWNet model, an ensemble wavelet neural network using exogenous parameters. When it is demonstrated that climate elements are statistically relevant, this model's strong adaptability allows it to effortlessly include climatic aspects into its framework. By combining wavelet processing with an ensemble neural network topology, it improves the accuracy of long-term forecasting. XEWNet is renowned for being simple to use, understandable, and interpretable mathematically, making it user-friendly. The complicated nonlinear interactions between dengue incidence and rainfall are also well captured by XEWNet. XEWNet surpasses statistics, machine learning, and deep learning methods in 75% of cases, according to computational experiments and statistical evaluations, proving its effectiveness in predicting dengue incidence.

Francisco, Micanaldo Ernesto, et al. [6] evaluated the combined effects of landscape and climatic conditions on the prevalence of mosquitoes and the incidence of dengue. The investigation showed that locations with a high residential density and the presence of industrial zones connected by roadways had a greater incidence of mosquitoes. The impacts of the landscape elements were amplified by precipitation, which was found to be a significant covariate. It may have expanded mosquito breeding grounds and encouraged mosquito reproduction. The study also discovered that precipitation was the main factor in predicting dengue incidence, with an impact that was more noticeable in regions with a high residential density and business activity. These results indicate that one useful tactic for controlling dengue dynamics may be to prioritize vector control activities during the rainy season in residential and commercial locations.

Mishra, Vipul Kumar, et al. [7] discovered a correlation between the development of dengue and environmental elements, including humidity, temperature, and precipitation. To link climate variables with dengue case occurrences, these studies frequently use quantitative models. In a recent machine learning study, researchers looked into the possibility of predicting future dengue cases using climate data from weather projections. They developed a brand-new two-fold linear regression model that trounced the competition and achieved a remarkable mean absolute error of 19.81, which is far lower than previous machine learning techniques. The study also carried out a thorough assessment, contrasting the proposed method with a number of predictive models, including boosted trees, random forests, neural networks, support vector machines, and XGBoost, and demonstrating the superior performance of their novel two-fold linear regression model in dengue forecasting.

Hadi, Z. A., et al. [8] analyzed socio demographic elements which improve an individual subpopulation's vulnerability to dengue fever (DF). It recognizes the considerable influence of climatic factors on DF incidence and fatality rates. Notably, due to the methodological challenges involved in combining meteorological and sociodemographic data, there is little study that concurrently takes both of these elements into account. The utilization of geographic information systems (GIS) and machine learning (ML) algorithms as critical instruments for carrying out epidemiological and geographic research in this setting is highlighted in the article. It underlines that ML techniques don't require explicit programming to assess data and make predictions. The results of the study are anticipated to help Malaysia in its attempts to effectively control dengue, and the framework created for DF has the potential to be modified for other illnesses spread by mosquitoes, such as malaria and chikungunya, as well as for conditions that are not related to mosquitoes.

Xu, Jiucheng, et al.[9] developed an accurate dengue forecasting model for this study, Long Short-Term Memory (LSTM) recurrent neural networks were used. Monthly dengue cases and environmental factors were the primary considerations. For one-month dengue case estimates, the effectiveness of LSTM models was evaluated against previously reported models. The average root mean squared error (RMSE) of predictions was reduced by the LSTM model by 12.99% to 24.91% on average and by 15.09% to 26.82% during outbreak periods, continuously beating other models, according to the results, which demonstrated a considerable improvement. TL was found to improve the model's generalizability, especially in regions with lower dengue case rates. These discoveries provide a dengue forecasting model that is more precise and has the potential to be used to anticipate other infectious illnesses with a similar epidemiology.

Omadlao, Zanya Reubenne D., et al. [10] used information from the City Health Service Office and pertinent meteorological data, this study focused on developing a dengue forecasting system for the most dengue-affected neighborhood in Baguio City. Data visualization helped determine that Barangay Irisan should serve as the system's focal point. Using meteorological variables including rainfall, temperature, and humidity, as well as lag times, the study attempted to forecast the presence of dengue cases on particular days. Several machine learning (ML) models, such as k-Nearest Neighbors, Gaussian Naive Bayes, Adaptive Boosting, Logistic Regression, LogitBoost, Linear Discriminant Analysis, and C-Support Vector Classifier, were used. dissemination. With a mean true positive rate of 62%, a false positive rate of 29%, a true negative rate of 72%, and an area under the receiver operator curve score of 67%, it was discovered that the Gaussian Naive Bayes model performed the best. The forecasting approach might help the barangay make decisions about dengue prevention and information.

Appice, Annalisa, et al. [11] introduced an effective machine learning framework aimed for accurate annual dengue forecasts. To extract temporal patterns from historical data, the methodology uses a multi-stage procedure that combines auto-encoding, window-based data representation, and trend-based temporal clustering. A trend association-based nearest neighbor predictor is then used to make predictions. To evaluate the efficacy of this strategy, the study conducts a case study using data on dengue and dengue hemorrhagic fever from 1985 to 2010 in 32 federal states of Mexico. Empirical findings emphasize the possibility for accurate dengue prediction on an annual scale by demonstrating the applicability and superior performance of the suggested strategy in regression and time series forecasting analyses when compared to other cutting-edge techniques.

Kamarudin, et al. [12] influenced vector proliferation and infection rates, controlling mosquito-borne diseases (MBD) is a difficult endeavor. The success of MBD outbreak prediction systems might vary depending on the study area and is sometimes hampered by issues like data shortages and poor access to pertinent datasets. The usage of open-source internet data raises problems with data availability, structure, sufficiency, relevance, and potential noise, all of which have an impact on the predictive power of machine learning models. In order to overcome these difficulties, this paper evaluates current MBD epidemic prediction methods and suggests a more advanced architecture that includes the Entomological Index feature. The new conceptual framework highlights the need for interventions that are more specific and focused and includes machine learning to improve the forecasting of upcoming MBD outbreaks.

Gupta, Gaurav, et al. [13] emphasized the urgent and essential need for diagnostic methods that can reliably differentiate between Dengue and its subgroups in the early stages of infection. The need for early diagnosis in treating dengue illness is highlighted. Since most forecasting systems are still in the early stages of research, it recognises the inherent difficulties in predicting infectious diseases like dengue. The use of microarray and RNA-Seq data to create predictive models for dengue is discussed in the paper. The random forest classifier stands out as the most effective of these models, attaining a remarkable mean prediction score of 9% Because of this, the article advises creating a machine learning-based strategy to enhance Dengue diagnoses and predictions.

Salim, Nurul Azam Mohd, et al. [14] focused on the five Selangor districts in Malaysia with the highest prevalence of dengue between 2013 and 2017, machine learning models were assessed for dengue outbreak prediction. These models included climate factors like temperature, wind speed, humidity, and rainfall. The SVM model with a linear kernel initially showed poor sensitivity (14%) and high specificity (95%), but it eventually outperformed all others with an accuracy of 70%. Unexpectedly, sensitivity increased considerably from the initial 14.4% to 63.54% when evaluated with imbalanced data. The week of the year was found to be the most significant predictor in the SVM model, demonstrating the promise of machine learning in accurately forecasting Dengue outbreaks, particularly in high-incidence locations like the Selangor districts.

Hoyos, William, et al. [15] structured literature review (SLR) examines several methods for diagnosing dengue, simulating outbreaks, and developing therapies. Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, the review examined 64 studies to determine their advantages and disadvantages. Surprisingly, 59.1% of the studies mainly used logistic regression to diagnose dengue. Contrarily, the most popular method (17.4%) for geographical analysis in epidemic modeling was linear regression. In 70% of the studies on intervention modeling, the General Linear Model served as the main methodology. The review also touched on the use of machine learning in dengue modeling, stressing both its difficulties and possible benefits.

Li, Zhichao, et al. [16] proposed a useful methodology for predicting dengue outbreaks based on epidemiological weeks using deep learning techniques and geospatial big data cloud computing through Google Earth Engine. It specifically looks into dengue epidemics that occurred in the Federal District of Brazil between 2007 and 2019. The epidemiological calendar and Google Earth Engine are used in the study to generate weekly composites for the various dengue transmission-related parameters. After that, spatial integration is used to create time series data for dengue transmission areas. Long Short-Term Memory (LSTM) models are employed in the modeling phase to forecast dengue cases for a number of weeks in the future. Time-difference natural log-transformed dengue cases are utilized as explanatory factors, and preceding instances are taken into account for greater accuracy. In particular, the model performs much better when past examples are taken into account, with the projection for the next five weeks showing the highest predictive accuracy.

Bhatia, Surbhi, et al. [17] implemented machine learning models in monitoring systems for predicting dengue outbreaks is thoroughly covered in this literature study. The review gives a breakdown of the research's methodology, objectives, and field of study across many nations and eras of history. It is primarily concerned with determining how the changing climate and the rise in dengue incidence are related. The analysis explores the different types of data used, such as patient data and information about the weather, and it covers the techniques for cleaning and preparing the data. In addition, it looks at how machine learning models are utilized for predictive analysis and correlation methods to find links between dengue incidence and weather variables. In order to create dengue surveillance systems, there are a number of alternative approaches that might be taken. The paper also discusses difficulties in the field and present limits.

Nguyen, Van-Hau, et al. [18] used deep learning techniques and a wide range of meteorological input factors to develop a Dengue Fever (DF) prediction model for Vietnam. The main objective was to offer suggestions for epidemic prevention and public health interventions, especially in light of future climate change effects. When measured by RMSE and MAE, the LSTM-ATT model performed the best, with average rankings of 1.60 and 1.90, respectively. In 12 out of 14 provinces, LSTM-ATT displayed superior predicting abilities versus LSTM, both in terms of RMSE and MAE. By accurately forecasting DF incidence and outbreaks up to three months in advance in Vietnam, LSTM-ATT distinguished itself as a novel application in the region and demonstrated the potential of deep learning models for climate-based DF forecasting.

Krishnan, Nor Farisha Muhamad, et al.[19] suggested using Decision Tree (DT) and ANN models to forecast dengue outbreaks in Kota Bharu using data on maximum temperature, minimum temperature, total rainfall, and average humidity. Through testing with various numbers of hidden nodes, the performance of the ANN model was improved. Key evaluation measures like accuracy, sensitivity, and specificity showed that ANN surpassed DT in forecasting dengue outbreaks, with a rate of 68.85% accuracy, an exceptionally high sensitivity of 99.71%, and a meager specificity of 1.27%. This demonstrates how excellent ANN is, especially in terms of its exceptional sensitivity. Although better than DT, the overall accuracy is still quite low. The complexity and properties of the dataset should guide the selection of the prediction model. These models have the potential to help the government take preemptive steps to inform the public about how dengue epidemics are impacted by climate change.

Wu, Yan, et al. [20] focused the feature selection approach in this work using a Genetic Algorithm based on Support Vector Machines (SVM), whereas the data pre-processing method for dengue prediction uses Wavelet transformation. A regression model built on SVM is used for forecasting. This model's predictions for dengue cases in Singapore have significantly improved, according to the results. This strategy has the potential to help health control organizations improve their strategic planning for disease prevention and dengue outbreak control. The study highlights a strong link between monsoon seasonality and dengue virus transmission, giving important insights for disease control efforts. It is noteworthy that the study finds that mean temperature and monthly seasonality have no impact on dengue outbreaks.

Table 2.1: Accuracy Comparison of Existing Related Papers

SN	Author	Dataset	Applied Algorithms	Best Accuracy
1	Dey, Samrat Kumar, et al. [1]	DengueBD dataset	SVR, MLR	67%
2	Sarma, Dhiman, et al. [3]	Patient dataset in Bangladesh	Decision Trees (DT), Random Forests (RF)	79%
3	Panja, Madhurima, et al. [5]	San Juan, Iquitos, Ahmedabad	XEWNet model (ensemble wavelet neural network)	75%
4	Salim, Nurul Azam Mohd, et al. [14]	Selangor districts	SVM model with linear kernel	70%
5	Krishnan, Nor Farisha Muhamad, et al.[19]	Kota Bharu	Decision Tree (DT), ANN models	69%

6	My Work	Dengue Data	GaussianNB, RF, DT, VotingClassifier, GradientBoostingClassifier and LR	99.40%

2.3 Comparative Analysis and Summary

The comparative analysis and summary stage of the deep learning-based prediction project for Dengue risk zones in Bangladesh includes an in-depth evaluation of the performance of several machine learning algorithms used in the dataset. This key phase is to determine the best effective algorithms for forecasting Dengue risk zones based on weather data. To measure the models' capacity to properly categorise areas into distinct risk categories, performance metrics such as accuracy, precision, recall, and F1 score are created. The summary provides a complete analysis of each algorithm's strengths and problems, taking into account criteria such as computing efficiency, accessibility, and prediction accuracy. The comparative study results assist the selection of the best algorithm or ensemble of algorithms for Dengue risk prediction in Bangladesh. This stage's output gives a clear path for developing the prediction model, removing any limits, and improving its use in everyday life. The summary not only supports model selection decision-making, but also adds vital insights to the larger area of sickness prediction, indicating the need of using machine learning in public health research.

2.4 Scope of the Problem

The challenge addressed by machine learning-based prediction of Dengue risk zones in Bangladesh is large and complex. Dengue fever, a mosquito-borne viral infection, provides a significant public health concern in tropical countries, with annual epidemics in Bangladesh. The scope of the study includes the complicated and dynamic nature of Dengue transmission, which is impacted by climatic conditions such as temperature, precipitation, and humidity. The challenge addressed by machine learning-based prediction of Dengue risk zones in Bangladesh is large and complex. Dengue fever, a mosquito-borne viral infection, provides a significant public health concern in tropical countries, with annual epidemics in Bangladesh. The scope of the study includes the complicated and dynamic nature of Dengue transmission, which is impacted by climatic conditions such as temperature, precipitation, and humidity. The findings of this study have the potential to transform public health efforts by providing a proactive approach to Dengue control. With predictive modeling, the scope includes targeted interventions, the use of resources, and preventative actions. The findings might have significant effects for places dealing with the same problems at the intersection of viruses, climate dynamics, and modern data analytics.

2.5 Challenges

Machine learning-based Dengue risk zone prediction in Bangladesh has various problems that must be carefully considered during the study process. The intricacy of the links between climatic factors and Dengue transmission is a considerable problem, which may need advanced algorithms to effectively record and evaluate. Data quality and availability provide additional hurdles, since historical meteorological data and consistent records of Dengue cases must be available and consistent during Bangladesh's different areas. Incomplete or incorrect info could damage the prediction model's success. Bangladesh's geographical and climatic variety generates variation in Dengue transmission patterns, needing model adaptation to distinct regional contexts. A critical difficulty is ensuring the model's adaptability and strength across different environments. Furthermore, ethical concerns about data protection and responsible information usage must be addressed. Balancing the requirement for exact information with individual privacy protection is a continuing problem. Finally, machine learning model use and application in public health contexts may face opposition or confusion. Effective communication of the model's insights to politicians and healthcare professionals is critical for overcoming the challenges and ensuring the predictive model's effective deployment.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

Machine Learning-Based Prediction of Dengue Risk Zones in Bangladesh Based on Weather Data" is a research topic that focuses on using advanced machine learning techniques to anticipate Dengue risk across various areas in Bangladesh. The main goal is to create a prediction algorithm that can analyze past meteorological data and reported Dengue cases to identify particular risk zones. This study's instrumentation is a complex combination of computer tools and statistical methods. The key tools for processing and evaluating large-scale weather datasets are advanced machine learning algorithms. These algorithms are used to decipher complicated patterns and correlations between climatic factors and the dynamics of Dengue transmission. The computational facilities which includes high-performance computing resources, programs for machine learning, and geographical information systems (GIS), is an essential component of the instrumentation. Geographic display technologies help in the presentation of anticipated Dengue risk zones in an accessible and useful manner. The collaboration of advanced algorithms for machine learning and strong computational infrastructure positions the research to significantly contribute to the understanding and active management of Dengue in Bangladesh, providing a valuable template for future efforts at the intersection of public health, climate science, and artificial intelligence.

3.2 Data Collection Procedure

A methodical technique to gather important information on both the weather and Dengue cases is used in the data collection procedure for machine learning-based prediction of Dengue risk zones in Bangladesh. Temperature, precipitation, humidity, and other weather-related data are derived from reliable meteorological agencies and satellite measurements. Data on dengue cases is gathered from medical records, public health organisations, and appropriate government databases. This covers data on reported Dengue cases, their geographic locations, and the timeframe in which they occurred. Preprocessing activities include cleaning, organizing, and dealing with missing values in both the meteorological and Dengue case databases to assure data quality. Temporal alignment is critical for preventing data leakage by ensuring that data to train comes before testing data. The datasets are then combined using similar temporal and geographical variables to create a single dataset for training and assessing machine learning models. Ethical concerns are followed, ensuring individuals' privacy and confidentiality inside the databases. Overall, the data collecting position aims to provide a solid foundation for the building and training of machine learning models, allowing for reliable forecasts of Dengue risk zones in Bangladesh using historical weather data.

3.3 Statistical Analysis

Statistical analysis is an important part of the machine learning-based prediction of Dengue risk zones in Bangladesh, as it provides a statistical basis for evaluating the model's accuracy and dependability. Various statistical approaches are used to assess the links between climatic factors and Dengue transmission patterns, confirming the predictive model's strength. To summarize and identify the important properties of the meteorological and Dengue datasets, descriptive statistics are used. Measures like mean, median, standard deviation, and skewness provide information about the data's primary trend and distribution. The method of correlation analysis is used to identify the degree and direction of relationships between various climatic parameters and Dengue incidence. This stage assists in determining which factors substantially contribute to the predictive capability of the model. Regression study may also be used to estimate the statistical influence of certain weather variables on Dengue risk, providing a deeper awareness of the links. The statistical studies not only impact the model's feature selection, but they also play an important role in confirming the model's predictions. The research makes sure that the machine learning-based model gives relevant and solid insights into Dengue risk zones in Bangladesh based on weather data by carefully using methods of statistics.

3.4 Proposed Methodology

Proposed Methodology: Machine Learning-Based Detection of Dengue Risk Zones in Bangladesh Using Weather Data.

Flow chart:

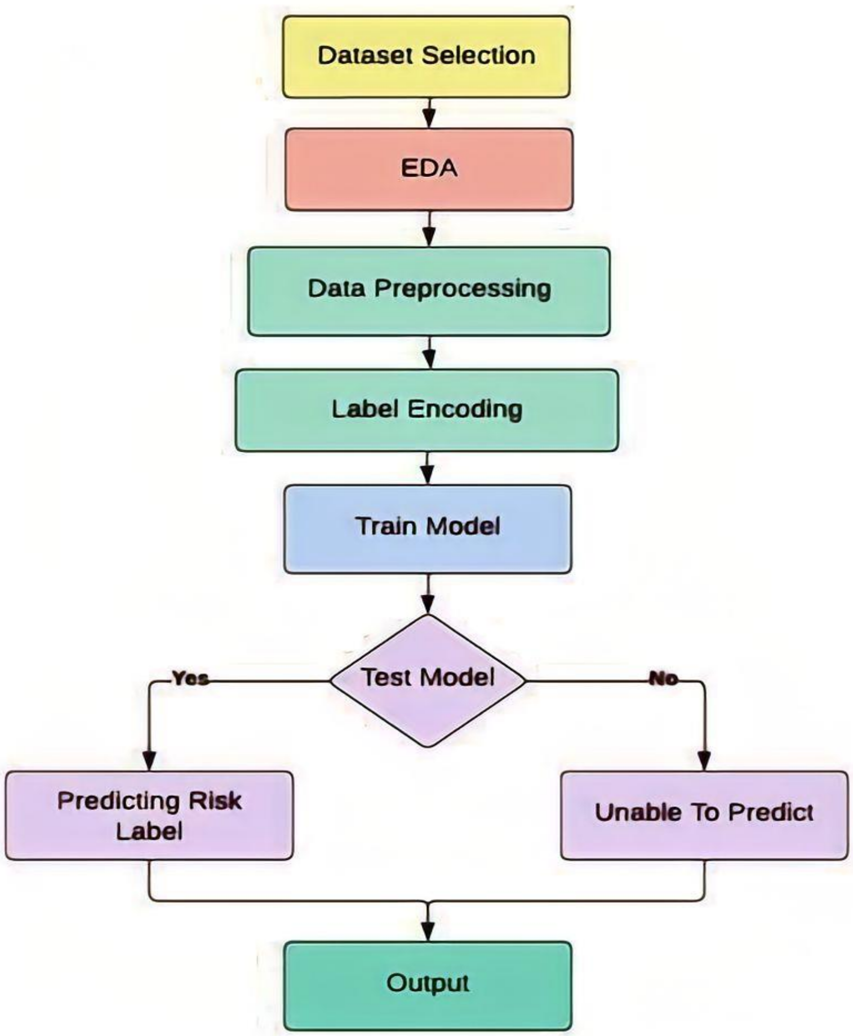

Figure 3.1: Methodology Flowchart

Figure 3.1 shows the Methodology Flowchart of my work. The details about methodology are given below:

Data Gathering:

Obtain historical meteorological data from credible sources, including variables such as temperature, precipitation, humidity, and wind speed. Create a list of Dengue cases, including their location, date, and duration.

Data Labelling:

Label the dataset with Dengue risk categories such as "High Risk," "Moderate Risk," "Low Risk," and "No Risk." This involves evaluating reported Dengue cases in specific locales based on how severe they are.

picture Enhancement:

Using picture enhancement techniques, improve the dataset. This involves applying transformations such as rotation, scaling, and flipping to improve the variety of training cases, hence boosting the model's generality.

Data Separation:

Separate the dataset into training, validation, and testing sets. Consider the time factor to ensure that the model is trained on historical data and evaluated and tested on past use cases.

Machine learning Algorithms:

Use machine learning algorithms such as GaussianNB, RandomForestClassifier, DecisionTreeClassifier, VotingClassifier, GradientBoostingClassifier, and LogisticRegression in your training model. Adjusting parameters to maximise performance, train the model using the training dataset.

Recognize:

patterns and relationships between the weather and Dengue risk zones in Bangladesh using the trained model. The program evaluates previously observed incidents and forecasts the risk category for each place.

Output:

Create danger zone maps and predictions, highlighting locations labelled "High Risk," "Moderate Risk," "Low Risk," and "No Risk." Provide a complete output that assists public health authorities in was efforts, budget allocation, and steps to avoid them.

To improve the accuracy and application of Dengue risk predicts using meteorological data in the context of Bangladesh, this suggested technique fits data-driven methods, picture enrichment for model security, and machine learning algorithms.

3.5 Implementation Requirements

Several important parts are required for the effective deployment of the machine learning-based Dengue risk prediction model in Bangladesh. Large datasets and costly training methods need enough computational resources. A strong data infrastructure is required for storing, maintaining, and analysing large amounts of historical meteorological data and Dengue case records, which may be stored on the cloud. The use of popular machine learning frameworks such as Tesla or the PyTorch along with programming tools such as Python and necessary libraries, ensures successful algorithm implementation and training. Geographic techniques, particularly Geographic Information System (GIS) technology, provide data analysis and imagining, allowing for the mapping of Dengue risk zones all over Bangladesh's different areas. Data privacy rules, informed permission, and responsible data usage are all expected by ethical issues. Setting confirmation measures, such as accuracy and precision, is critical for determining the model's accuracy. Finally, the creation of a simple interface enables users to successfully understand and apply the model's predictions, therefore helping to activate Dengue risk management and supporting public health actions.

CHAPTER 4

Experimental results and discussion

4.1 Experimental Setup

The test facility for predicting Dengue risk zones in Bangladesh includes a complete and systematic organisation of tools and processes to allow in effect machine learning model learning and evaluation. The collection and processing of historical meteorological data and Dengue case records is essential for data preparation. This careful approach guarantees that the datasets are of high quality, consistent, and aligned, and it is reinforced with picture enhancement methods to increase the range of the training dataset. To successfully execute complicated computations, high-performance computing resources equipped with GPUs must be deployed. This speeds up machine learning model training, allowing for more rapid testing and optimization of algorithms built with tools like Google or Python Torch. Geographic tools, such as GIS technology, help in analysing space, improving display and understanding of Dengue risk zones throughout Bangladesh's different geographical areas. Cross-validation techniques are used to evaluate the model's performance and reduce over-fitting, ensuring portability to previously unknown data groups. A thorough modifying of the parameters creates a balance between model complexity and performance. Collaboration with public health authorities and weather agencies proves the relevance of the experimental setting and gives critical insights for enhancing the machine learning-based predicting model. This thorough method of experimentation matches with Bangladesh's complicated climatic dynamics and public health issues, promoting the creation of a powerful and efficient Dengue risk prediction system.

4.2 Experimental Results & Analysis

Logistic Regression:

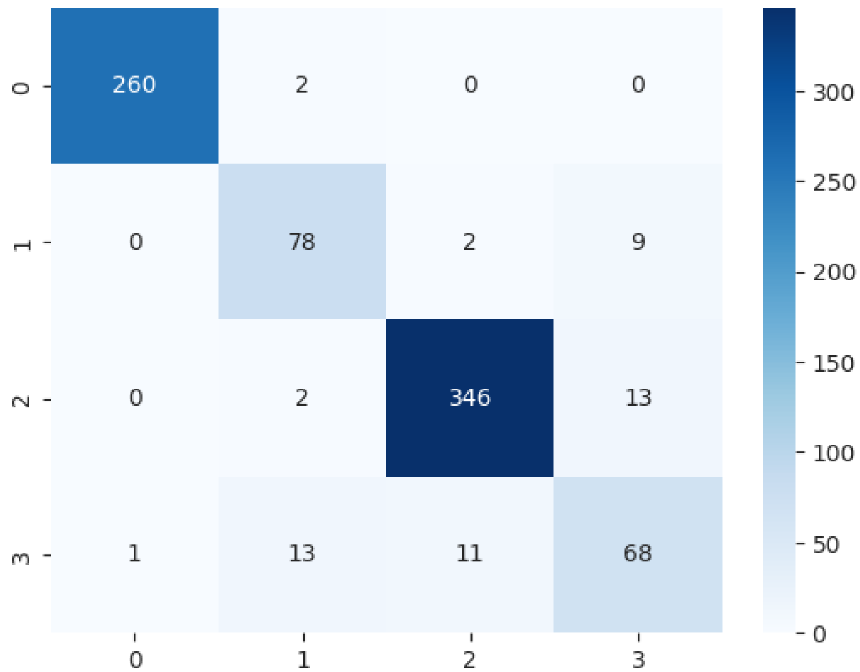

Figure 4.1: Confusion Matrix (LogisticRegression)

Logistic Regression is a statistical technique primarily used for binary classification tasks, where the objective is to predict the probability of an observation belonging to one of two possible classes. Despite its name, Logistic Regression is tailored for classification rather than regression analysis. The model establishes a relationship between independent variables and the log odds of the dependent variable, employing the logistic (sigmoid) function to constrain predictions within the [0, 1] range. Renowned for its simplicity, interpretability, and efficiency with linearly separable data, Logistic Regression finds widespread application in diverse domains, including machine learning, statistics, and epidemiology.

GaussianNB

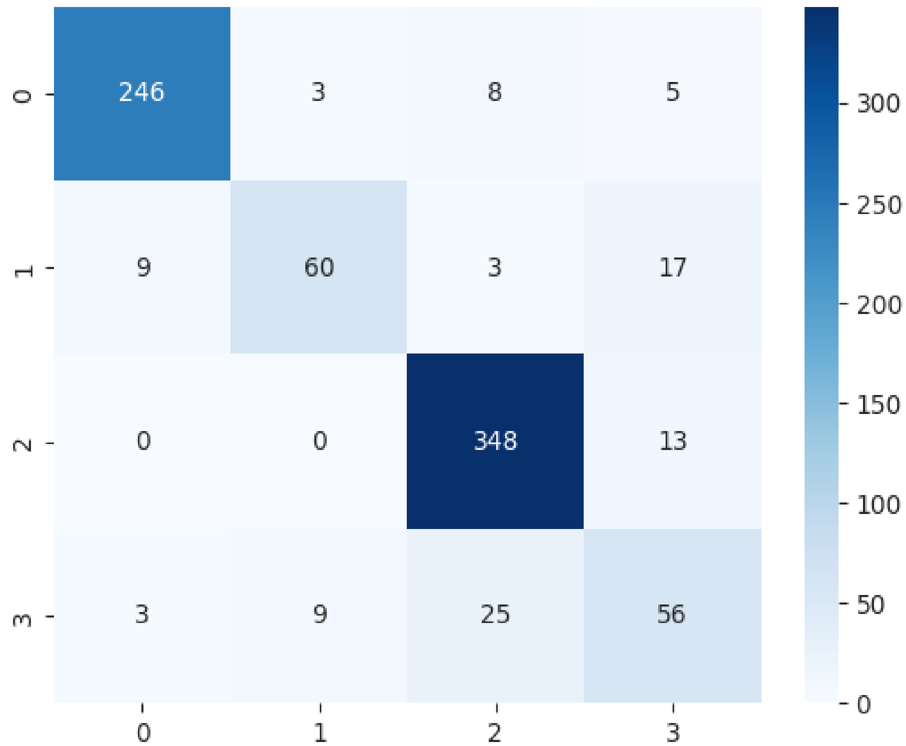

Figure 4.2: Confusion Matrix (GaussianNB)

An abbreviation for Gaussian Naive Bayes, is a probabilistic machine learning algorithm commonly employed for classification tasks. Rooted in Bayes' theorem and premised on the assumption of feature independence, this algorithm is particularly well-suited for datasets characterized by continuous features following a Gaussian (normal) distribution. It calculates the likelihood of an instance belonging to a specific class by combining the probabilities of individual features. GaussianNB is widely applied in practice, demonstrating efficacy in various domains such as text classification and scenarios with high-dimensional feature spaces. Its simplicity, efficiency, and ease of implementation render it a pragmatic choice for classification tasks, especially when dealing with real-world datasets.

Decision Tree

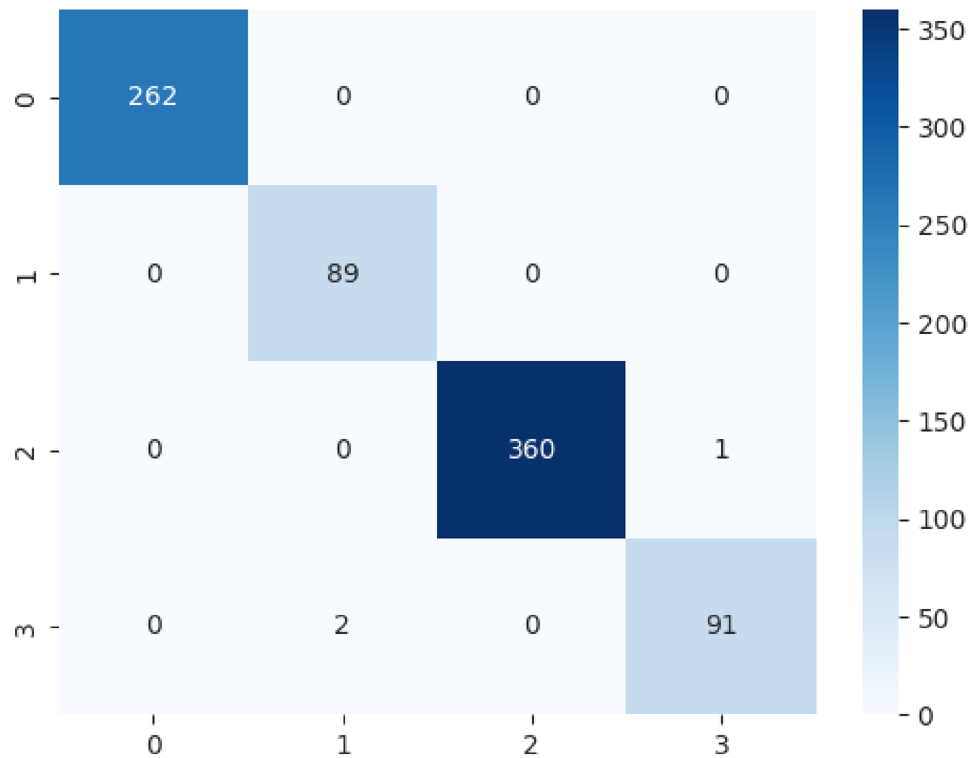

Figure 4.3: Confusion Matrix (DecisionTree)

A Decision Tree is a powerful and widely used machine learning algorithm that is primarily employed for both classification and regression tasks. This algorithm works by recursively partitioning the dataset into subsets based on the most significant attribute at each node, leading to a tree-like structure of decisions. Each internal node represents a decision based on a feature, and each leaf node corresponds to the predicted outcome or class label. Decision Trees are interpretable, easy to understand, and offer insights into the decision-making process. They are utilized in various fields such as finance, medicine, and business for their ability to handle complex decision-making scenarios and contribute to the development of accurate predictive models.

Gradient Boosting

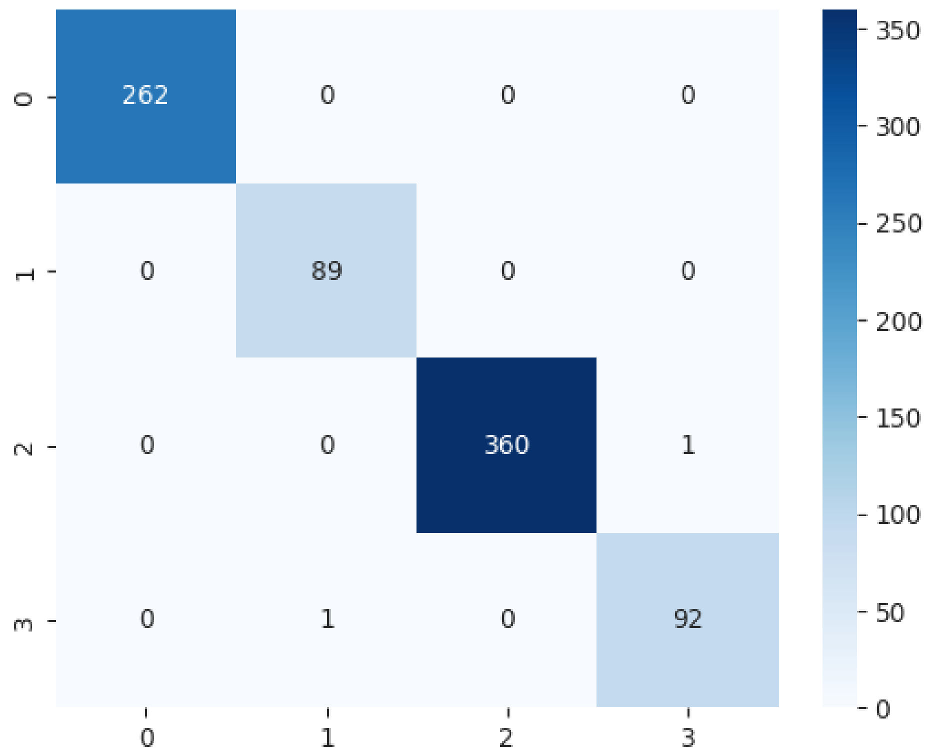

Figure 4.4: Confusion Matrix (GradientBoosting)

Gradient Boosting is an ensemble machine learning technique that sequentially builds a series of weak models, usually decision trees, to create a robust and accurate predictive model. It excels at handling complex relationships within data, making it suitable for regression and classification tasks. Widely utilized in various domains, Gradient Boosting algorithms like XGBoost and LightGBM are known for their ability to deliver high accuracy and versatility in handling diverse datasets.

Random Forest

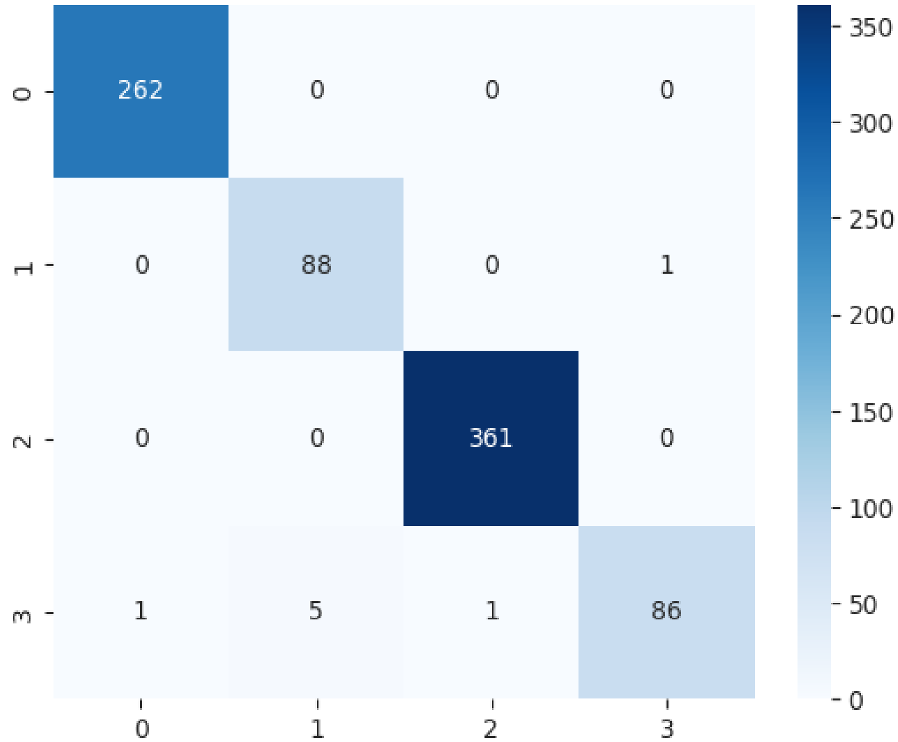

Figure 4.5: Confusion Matrix (RandomForest)

Random Forest is a potent ensemble learning algorithm for classification and regression. It builds multiple decision trees during training, introducing randomness to enhance generalization and reduce overfitting. Widely acclaimed for its robustness and versatility, Random Forest is effective in handling high-dimensional data and excels in capturing complex relationships. It is a popular choice across various domains, providing reliable predictions and valuable feature importance rankings.

Voting Classifier

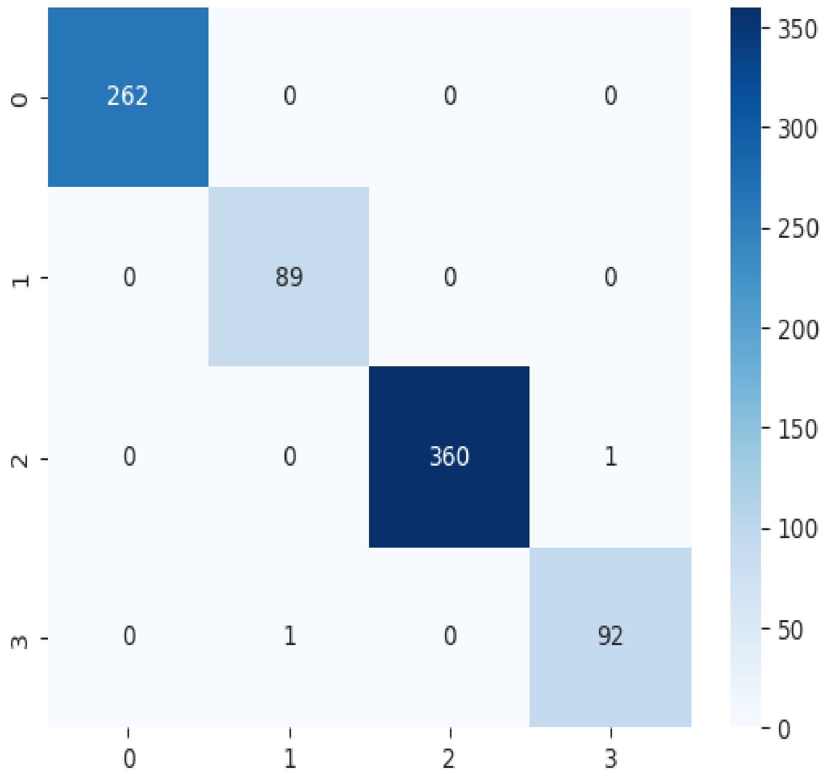

Figure 4.6: Confusion Matrix (VotingClassifier)

In scikit-learn is a meta-estimator that combines multiple machine learning models for improved predictions. It aggregates individual model outputs through either a majority vote (hard voting) or a weighted average of probabilities (soft voting). This ensemble approach enhances predictive accuracy and is especially useful when different models excel in various aspects of the data.

Accuracy: The accuracy of the model's predictions is determined by comparing the number of correctly classified samples to the total number of samples. Unbalanced classes give a general idea of the model's efficacy, but they may not give a complete picture.

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

Precision: Precision is concerned with the number of true positive forecasts made by the model out of all positive predictions generated by the model.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Recall: The percentage of true positive predictions created out of all actually positive samples is referred to as recall. It's also known as sensitivity or true positive rate.

$$Recall = \frac{TruePositive}{TruePositive + \ FalseNegative}$$

F1 Score: The F1 score is determined as the harmonic mean of recall and precision. Its fair evaluation metric considers recall and precision. The F1 score is useful in cases where class sizes are not equal since it accounts for both false positives and false negatives. A high F1 score indicates a good precision to recall ratio.

$$F - 1\,Score = 2 * \frac{Recall * \ Precision}{Recall + Precision}$$

In the table below, 4.1, the outcomes of machine learning models are compared based on Accuracy, Precision, Recall, and F1 Score.

Table 4.1. Performance Evaluation

Model Name	Accuracy	Precision	Recall	F1-Score
GradientBoostingClassifier	99.40%	99.75	99.75	99.75
VotingClassifier	99.20%	99.75	99.75	99.75
DecisionTreeClassifier	99.20%	99.63	99.62	99.62
RandomForestClassifier	97.40%	99.25	99.24	99.62
LogisticRegression	92.01%	93.41	93.41	93.41
GaussianNB	89.81%	87.19	88.19	87.96

4.2.1 Accuracy

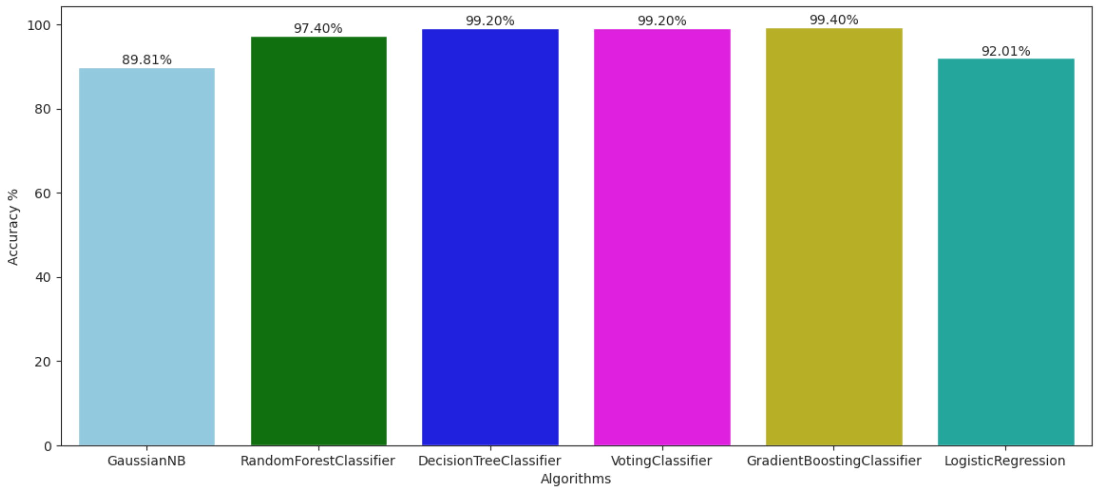

Figure 4.7: Comparative Model Accuracy Bar Plot

Figure 4.7 compares the accuracy of six different machine learning models. The ensemble methods, VotingClassifier and GradientBoostingClassifier, topped the charts with impressive accuracy close to 99.5%. Overall, the models performed well, with four out of six achieving scores above 97%. While accuracy is crucial, other factors like interpretability and task suitability should also be considered when picking the best model.

4.3 Discussion

The discussion phase of machine learning-based Dengue risk zone prediction in Bangladesh is critical for understanding the findings, correcting limits, and generating suggestions for future research and practical applications. It covers an in-depth analysis of the model's performance, the relevance of identified danger zones, and implications for public health actions. The conversation looks into the relationship between climatic factors and Dengue transfer, explaining how the machine learning model catches complicated patterns. It evaluates the model's validity and reliability by comparing outcomes across algorithms and focusing on opportunities for improvement. Furthermore, the discussion places the identified danger zones in the perspective of Bangladesh's larger panorama of climate-driven health concerns. It looks at the practical consequences for public health officials, highlighting the possibilities for preventive actions and budget allocation. This phase also tackles the study's ethical concerns and constraints, providing a clear reflection on the scope of the investigation. Finally, the debate provides significant insights, bridging the gap between scientific discoveries and effective Dengue risk management methods in Bangladesh.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The impact of machine learning-based prediction of Dengue risk zones in Bangladesh on society is substantial, suggesting revolutionary consequences for public health and community well-being. The model provides healthcare authorities and policymakers with a strong tool for proactively addressing Dengue outbreaks by making accurate and timely forecasts. The identified danger zones serve as a strategic guide for focused interventions, effective resource allocation, and the execution of preventative actions. Because the approach enables early identification and therapy, society benefits from lower Dengue morbidity and mortality. This preventative strategy not only protects public health but also reduces the economic burden associated with healthcare spending during major outbreaks. Moreover, the research helps increase knowledge about the relationship between climate dynamics and vector-borne illnesses, urging a culture of tolerance and preparation in communities. The incorporation of modern technology into public health policies is an example of a progressive and adaptable response to developing health issues, establishing a precedent for informed choices and risk reduction on a larger scale. In short, the societal impact goes beyond direct health gains, altering public health management concepts in the face of climate-sensitive illnesses.

5.2 Impact on Environment

The environmental effect of machine learning-based prediction of Dengue risk zones in Bangladesh is important, as it represents an innovation in disease management methods that can contribute to environmental management. The model supports tailored treatments by enabling exact predictions, removing the requirement for deep chemical interventions like applying pesticides. This, in turn, reduces the environmental impact of standard Dengue control efforts. Moreover, proactive risk zone identification supports the best use of resources, eliminating possible environmental problems. The model's capacity to predict and avoid cases of dengue may reduce need for reactive measures, which frequently include the use of chemicals with possible adverse environmental effects. Furthermore, the study highlights the link between climate, disease dynamics, and environmental health. It highlights the significance of maintaining ecological balance and nature, since changes in weather patterns lead to the spread of Dengue vectors. In result, the environmental effect goes beyond illness control, supporting an improved and environmentally sensitive policy for public health in the context of climate-driven health issues.

5.3 Ethical Aspects

Machine learning-based prediction of Dengue risk zones in Bangladesh requires careful consideration of privacy, permission, and responsible data usage. The collecting and use of sensitive health and climate data necessitates dedication to moral principles. It is critical to ensure the privacy of persons in the dataset, gain informed permission, and communicate the goal of data usage openly. The possible impact on those in need is also an ethical consideration. The transmission of danger zone information must focus on reducing shame or intolerance and promoting community participation in decision-making processes. Transparency in the model's creation, validation, and deployment is also critical. Communicating unknowns, restrictions, and possible errors in the prediction model builds trust and ensures responsible technology usage. The study must keep to capital standards, ensuring that the prediction model's advantages are spread equally among varied groups. Continuous ethical monitoring and engagement with relevant consumers, such as public health authorities and local populations, are critical to navigating the ethical areas of this novel approach to sickness prediction and treatment.

5.4 Sustainability Plan

The machine learning-based Dengue risk prediction model's method for sustainability in Bangladesh includes important features for long-term effect. Continuous capacity building ensures that local doctors, nurses, and data scientists can use and maintain the model effectively. Ongoing public awareness efforts educate the public on Dengue prevention and the role of the model, with a focus on community participation for long-term public health practices. Establishing a long-term relationship among groups such as governmental entities, non-governmental organizations (NGOs), weather businesses, and neighborhoods guarantees continued support and data exchange. The model's lifetime and relevance are ensured by designing it for flexibility to shifting climatic patterns and scalability for evolving datasets and technology improvements. An established review board ensures that data and technology use is consistent with ethical standards and community values. This complete strategy supports the Dengue risk prediction model's continuous efficacy, relevance, and ethical use, contributing to long-term public health impact in Bangladesh.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In summary, the work on machine learning-based prediction of Dengue risk zones in Bangladesh is a ground-breaking technique at the border of technology and public health, and the dynamics of the climate. The study provides an in-depth understanding of Dengue transmission patterns impacted by climatic factors using advanced methods The prediction algorithm used in the study has the potential to transform public health policies by giving timely and accurate insights into various risk zones. It adds to active intervention tactics by covering high-risk areas that require urgent attention to low-risk areas that require constant monitoring. The study's dedication to responsible technology usage and long-term social effect is highlighted by ethical issues, ecological plans, and community involvement techniques. Overall, the study adds to the growing body of knowledge on using machine learning to forecast climate-sensitive diseases, providing a vital tool for minimising the impact of Dengue in Bangladesh as well as acting as a model for future research.

6.2 Conclusions

Finally, the machine learning-based prediction of Dengue risk zones in Bangladesh represents a new era in public health management. To discover complicated patterns within climatic variables, the study employs advanced methods such as GaussianNB, Random Forest Classifier, Decision-Tree, Voting Classifier, Gradient Boosting Classifier, and Logistic Regression. The prediction model not only accurately identifies danger zones, but it also delivers practical information for focused actions, distributing resources, and preventative actions. The incorporation of ethical issues, an ecological strategy, and social engagement shows the study's dedication to responsible and positive technology usage. The work provides an example for using machine learning in illness prediction by adding to the continuing dialogue on climate-driven health concerns. The findings have far-reaching implications not only for Dengue management in Bangladesh, but also for other global health issues driven by climate dynamics. Finally, this study offers an essential step in a more flexible and adaptable public health theory, highlighting the power of technology to solve complex and dynamic health issues.

6.3 Implication for Further Study

The work on machine learning-based prediction of Dengue risk zones in Bangladesh offers up new areas for inquiry, providing useful insights and asking amazing questions for more study. To improve the model's predictability, future research might look further into the temporal dynamics of Dengue transmission, studying changes in seasons and shifting climate patterns. There is also space for studying the integration of current data sources, such as satellite imaging or Internet of Things (IoT) devices, in order to produce more dynamic and current predictions. Testing the model's transferable to other places with comparable climatic patterns could increase its worldwide usefulness. Moreover, researching the socioeconomic factors driving Dengue transmission and community responses might contribute to greater understanding about illness dynamics. Collaborations including scientists, health professionals, and psychologists may find hidden factors affecting Dengue risk. Finally, the effects of the study go beyond its current focus, providing a rich platform for future research that might develop the model, improve its application, and contribute to the increasing body of knowledge at the junction of technology, climate, and public health.

REFERENCES

[1]S. K. Dey et al., “Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach,” PLOS ONE, vol. 17, no. 7, p. e0270933, Jul. 2022, doi: https://doi.org/10.1371/journal.pone.0270933.

[2]Anno, Sumiko, et al. "Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning." Geospatial Health 14.2 (2019).

[3]Sarma, Dhiman, et al. "Dengue prediction using machine learning algorithms." 2020 IEEE 8th R10 humanitarian technology conference (R10-HTC). IEEE, 2020.

[4]Babu, P. Devendar, et al. "Time Series Analysis-based Prediction of Dengue Spread using Climate Data."

[5]Panja, Madhurima, et al. "An ensemble neural network approach to forecast Dengue outbreak based on climatic condition." Chaos, Solitons & Fractals 167 (2023): 113124.

[6]Francisco, Micanaldo Ernesto, et al. "Dengue disease dynamics are modulated by the combined influence of precipitation and landscapes: a machine learning-based approach." bioRxiv (2020): 2020-09.

[7]Mishra, Vipul Kumar, et al. "Dengue disease spread prediction using twofold linear regression." 2019 IEEE 9th International Conference on Advanced Computing (IACC). IEEE, 2019.

[8]Hadi, Z. A., et al. "Development of machine learning modelling and dengue risk mapping: a concept framework." IOP Conference Series: Earth and Environmental Science. Vol. 1217. No. 1. IOP Publishing, 2023

[9]Xu, Jiucheng, et al. "Forecast of dengue cases in 20 Chinese cities based on the deep learning method." International journal of environmental research and public health 17.2 (2020): 453.

[10]Omadlao, Zanya Reubenne D., et al. "Machine learning-based dengue forecasting system for Irisan, Baguio city, Philippines." AIP Conference Proceedings. Vol. 2472. No. 1. AIP Publishing, 2022.

[11]Appice, Annalisa, et al. "A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico." Ieee Access 8 (2020): 52713-52725.

[12]Kamarudin, et al. "Forecasting the dengue outbreak using machine learning algorithm: A review." 2021 International Conference of Women in Data Science at Taif University (WiDSTaif). IEEE, 2021.

[13]Gupta, Gaurav, et al. "DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms." Diagnostics 13.6 (2023): 1093.

[14]Salim, Nurul Azam Mohd, et al. "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques." Scientific reports 11.1 (2021): 939.

[15]Hoyos, William, et al. "Dengue models based on machine learning techniques: A systematic literature review." Artificial intelligence in medicine 119 (2021): 102157.

[16]Li, Zhichao, et al. "Improving dengue forecasts by using geospatial big data analysis in google earth engine and the historical dengue information-aided long short term memory modeling." Biology 11.2 (2022): 169.

[17]Bhatia, Surbhi, et al. "A Retrospective Study of Climate Change Affecting Dengue: Evidences, Challenges and Future Directions." Frontiers in Public Health 10 (2022): 884645.

[18]Nguyen, Van-Hau, et al. "Deep learning models for forecasting dengue fever based on climate data in Vietnam." PLoS Neglected Tropical Diseases 16.6 (2022): e0010509.

[19]Krishnan, Nor Farisha Muhamad, et al. "Predicting Dengue Outbreak based on Meteorological Data Using Artificial Neural Network and Decision Tree Models." JOIV: International Journal on Informatics Visualization 6.3 (2022): 597-603.

[20]Wu, Yan, et al. "Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm." Proceedings of the World Congress on Engineering. Vol. 1. 2008.

Machine Learning

ORIGINALITY REPORT

22%	19%	11%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Loomis-Chaffee High School Student Paper	3%
3	Submitted to Daffodil International University Student Paper	2%
4	www.mdpi.com Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	www.acemap.info Internet Source	1%
7	Xinxing Zhao, Kainan Li, Candice Ke En Ang, Kang Hao Cheong. "A deep learning based hybrid architecture for weekly dengue incidences forecasting", Chaos, Solitons & Fractals, 2023 Publication	1%
8	ebin.pub Internet Source	