DENGUE DISEASE PREDICTION BY MACHINE LEARNING ALGORITHM IN BANGLADESH

BY

Md Golam Kibria
ID: 203-15-14572
AND

Fatema Ha-Mim
ID: 203-15-3918

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Partha Dip Sarkar
Lecturer
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Md Assaduzzaman
Lecturer (Sr. Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

APPROVAL

This Project titled "Dengue Disease Prediction by Machine Learning Algorithm in Bangladesh", submitted by **Md Golam Kibria** and **Fatema Ha-Mim** 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 14-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman

Md. Abbas Ali Khan
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Md. Aynul Hasan Nahid
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

Dr. Abu Sayed Md. Mostafizur Rahaman
Professor
Department of CSE
Jahangirnagar University

External Examiner

DECLARATION

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Partha Dip Sarkar, Lecturer, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

পার্থ দীপ
11/7/24

Partha Dip Sarkar
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Md Assaduzzaman
Lecturer (Sr. Scale)
Department of CSE
Daffodil International University

Submitted by:

Kibria

Md Golam Kibria
ID: -203-15-14572
Department of CSE
Daffodil International University

Fatema

Fatema Ha-Mim
ID: -203-15-3918
Department of CSE
Daffodil International University

©Daffodil International University, Bangladesh iii

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Partha Dip Sarkar**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

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

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

ABSTRACT

In Bangladesh, dengue fever is still a major public health issue that presents difficult management and preventative strategies. Proactive steps to lessen dengue's negative effects on public health can be made easier with accurate dengue epidemic prediction. In this work, we examine how well different machine learning methods predict the incidence of dengue fever in Bangladesh. We create a dataset with 1000 observations overall that consists of 9 input attributes and 1 outcome attribute. The input attributes comprise clinical markers such as NS1, IgG, and IgM levels in addition to demographic data like age and gender. Our prediction algorithm also incorporates geospatial variables like district, size, and kind of place. The output property "Outcome" indicates if dengue fever has occurred; this is a binary classification operation. The predictive performance of six machine learning algorithms is assessed, including Logistic Regression, Random Forest, Decision Trees, AdaBoost, Extreme Gradient Boosting (XGBoost), and LightGBM. With an astounding 98.67% accuracy rate, Random Forest is the most accurate of these algorithms. Our results highlight the potential of machine learning methods, especially Random Forest, to accurately forecast the incidence of dengue illness in Bangladesh. With the use of these prediction models, dengue outbreaks may be detected early and managed proactively, allowing for the prompt deployment of resources and the execution of focused intervention techniques to lessen the disease's crippling effect on public health systems.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Overview

Dengue fever, a common viral disease spread by mosquitoes, poses a serious threat to global public health, especially in areas with tropical or subtropical temperatures. It is primarily found in urban and semi-urban areas in tropical and subtropical climates globally [1], where there are many environmental conditions that encourage the growth of Aedes mosquitoes. The illness can cause everything from a low-grade fever to serious side effects like shock syndrome and hemorrhagic fever, which puts a heavy load on communities and healthcare systems. As of December 31, 2023, the Directorate General of Health Services (DGHS) announced that there had been a significant increase in dengue cases in Bangladesh, with a noteworthy outbreak leading to 321,179 hospital admissions and 1,705 fatalities [2]. This concerning pattern emphasises how critical it is to take proactive steps in order to anticipate, stop, and effectively manage dengue outbreaks. As a response to the disease's growing threat, this study aims to develop a prediction model using machine learning algorithms to predict dengue disease occurrences in Bangladesh. The study uses a dataset with demographic information, geographical features (like area type, house type, and district), and clinical markers (like NS1, IgG, and IgM levels) to try to use data analytics to predict and lessen the effects of dengue outbreaks. The urgent need to improve the precision and timeliness of dengue prediction by combining cutting-edge computational approaches with established epidemiological methods is what spurred this research. The goal of the project is to determine the best predictive model for dengue disease in Bangladesh by utilising machine learning methods including random forest, gradient boosting machines, and logistic regression. In the end, the research's conclusions may provide public health officials, legislators, and medical professionals with useful information that they can use to carry out focused interventions, manage resources effectively, and lessen the negative effects of dengue outbreaks on the general public's health and well-being in Bangladesh. In the end, the research's conclusions may provide public health officials, legislators, and medical professionals with useful information that they can use to carry out focused interventions, manage resources effectively, and lessen the negative effects of dengue outbreaks on the general public's health and well-being in Bangladesh.

1.2 Background and Present State

Dengue fever is a significant public health concern in Bangladesh, especially in urban and semi-urban areas where the Aedes aegypti mosquito thrives. The disease, characterized by high fever, severe headaches, joint and muscle pain, and in severe cases, bleeding and shock, poses a substantial burden on the healthcare system. According to the Directorate General of Health Services (DGHS), the dengue outbreak in 2023 led to 321,179 hospital admissions and 1,705 fatalities as of December 31, 2023, underscoring the critical need for effective prediction and management strategies. In recent years, advancements in machine learning have provided promising tools for disease prediction and management. Machine learning algorithms can analyze large datasets to identify patterns and predict outcomes with high accuracy. In the context of dengue, these algorithms can be trained to predict disease outbreaks and individual patient outcomes, enabling timely and targeted interventions. Despite the potential benefits, the application of machine learning in dengue prediction is still in its nascent stages in Bangladesh. Most studies have focused on traditional statistical methods or have limited their scope to small datasets. This research aims to bridge this gap by employing a comprehensive dataset with 1,000 rows and 10 columns, encompassing a range of input attributes including Gender, Age, NS1, IgG, IgM, Area, Area Type, House Type, and District, with the output attribute being the "Outcome" of dengue infection. This study explores the efficacy of nine different machine learning algorithms: Logistic Regression, Random Forest, Decision Trees, AdaBoost, XGBoost, and LightGBM. Among these, Random Forest has shown the highest accuracy of 98.67%, indicating its potential for effective dengue prediction. The goal is to enhance the predictive capabilities and provide a reliable tool for healthcare professionals in Bangladesh to manage and mitigate the impact of dengue fever.

1.3 Problem Statement

Dengue fever, primarily transmitted by the Aedes mosquito, has emerged as a significant public health threat in Bangladesh, particularly in urban and semi-urban regions. The increasing incidence of dengue outbreaks, coupled with the rising number of cases and fatalities, has placed a substantial burden on the country's healthcare system. Traditional methods of managing dengue are often reactive and rely on basic statistical analyses, which may not adequately capture the complexities of dengue transmission and patient outcomes. Consequently, there is an urgent need for more sophisticated and accurate predictive models to enable timely interventions and resource allocation. The study's justification is the pressing necessity to handle the serious threat that dengue virus poses to public health in Bangladesh. Healthcare officials have made a lot of effort, but the nation still struggles with recurring outbreaks that cause significant morbidity, mortality, and financial losses. Conventional methods for dengue prediction and management frequently rely on manual surveillance and retrospective data analysis, which may have limitations in terms of scope and efficacy. This work aims to improve dengue forecast accuracy, efficiency, and timeliness by using machine learning algorithms. This will allow for preventive actions to lessen the impact of epidemics. To reflect the intricate interplay of variables impacting dengue transmission and distribution, the prediction model incorporates a variety of demographic, clinical, and geographical parameters. Moreover, the discovery that XG-Boost is the most accurate predictive algorithm highlights the potential of cutting-edge computational methods to completely transform dengue prediction and management approaches. The results of this study could help healthcare professionals make more informed decisions based on evidence, provide them with useful knowledge, and eventually lower the morbidity and death rate in Bangladesh associated with dengue.

1.4 Objectives

The primary objective of this research is to develop an effective machine learning model for predicting dengue disease outcomes in Bangladesh, utilizing a comprehensive dataset and a variety of algorithms to identify the most accurate predictive tool. This objective is supported by several specific goals:

i. Data Collection and Preparation: To collect and preprocess a robust dataset consisting of relevant attributes such as Gender, Age, NS1, IgG, IgM, Area, Area Type, House Type, and District to accurately reflect the factors influencing dengue outcomes in Bangladesh.

ii. Algorithm Evaluation: To evaluate the performance of multiple machine learning algorithms including Logistic Regression, Gradient Boosting Machines, Random Forest, Decision Trees, AdaBoost, Extreme Gradient Boosting (XGBoost), and LightGBM in predicting dengue outcomes.

iii. Model Comparison: To compare the predictive accuracy, precision, recall, and F1-score of the different machine learning models and identify the most effective algorithm for dengue prediction.

iv. Optimization of Best Model: To optimize the selected model (Random Forest), which has shown the highest accuracy of 98.67%, to enhance its predictive performance further and ensure its robustness in real-world applications.

Implementation Framework: To develop a framework for implementing the optimized predictive model within the healthcare system in Bangladesh, facilitating early detection and efficient management of dengue cases.

v. Impact Assessment: To assess the potential impact of the predictive model on public health outcomes, resource allocation, and overall dengue management strategies in Bangladesh.

vi. Future Recommendations: To provide recommendations for future research and improvements in dengue prediction methodologies, including the integration of additional data sources and the exploration of new machine learning techniques.

1.5 Scope and Limitations

Our study on dengue disease prediction in Bangladesh using machine learning algorithms covers a wide range of important aspects. Our main goal is to forecast dengue epidemics with high predictive accuracy. To forecast the risk of dengue occurrence, this involves using machine learning algorithms to analyse a dataset that includes geographical characteristics (such as Area, Area Type, House Type, District), clinical markers (such as NS1, IgG, and IgM), and demographic attributes (such as Gender and Age). We also address issues with data availability and quality within this area. In order to provide accurate model training and evaluation, we seek out trustworthy datasets with adequate sample sizes and guarantee data integrity and completeness. Furthermore, in order to determine the best method for dengue disease prediction, we investigate a number of machine learning algorithms, such as AdaBoost, Extreme Gradient Boosting (XGBoost), LightGBM, Random Forest, Decision Trees, and Logistic Regression. Additionally, we take into account Bangladesh's geographical variations in dengue prevalence. When creating prediction models, our methodology takes into account regional differences in illness patterns and environmental factors. Our objective is to create real-time prediction models that would enable prompt intervention and response in the event of a dengue outbreak. Lastly, we also take the deployment and scalability of prediction models into account. Our designs are intended to facilitate public health activities and decision-making processes by being easily scaled to larger datasets and used in real-world contexts. For machine learning-based dengue disease prediction systems to be successfully implemented in Bangladesh, these issues must be addressed within the parameters of our research.

The limitations of this study include the following:

i. Dataset Size: The dataset used in this research consists of 1000 rows, which may limit the generalizability of the findings. A larger dataset might provide more robust results and better model performance.

ii. Attribute Selection: The input attributes are limited to Gender, Age, NS1, IgG, IgM, Area, Area Type, House Type, and District. There may be other significant factors influencing dengue outcomes that are not included in the dataset.

iii. Geographical Focus: The study is specific to Bangladesh, and the results may not be directly applicable to other regions with different environmental and demographic factors affecting dengue transmission.

iv. Algorithm Limitations: Each machine learning algorithm has its inherent strengths and weaknesses, and the study may not exhaustively cover all possible algorithms or their variations.

v. Computational Resources: The computational resources available for this research may limit the extent of model optimization and the complexity of algorithms that can be explored.

vi. Temporal Factors: The dataset represents a snapshot in time, and the models may not account for temporal changes in dengue patterns, such as seasonal variations or long-term trends.

1.6 Report Organization

The suggested report has a thorough format that walks readers through the methodology, conclusions, and ramifications of the study. Every chapter has a distinct function and adds to the document's overall coherence and depth.

Chapter 1: Introduction

Chapter one explains the Overview, Background and Present State, Problem Statement, Objectives, Scope and Limitations, Report Organization, and Summary.

Chapter 2: Literature Review

Chapter two explains the literature review with its Overview, Related Works, Comparison Between Existing Works, Open Issues, and Summary.

Chapter 3: Methodology

Chapter three explains the research methodology with its Overview, Proposed Methodology/System Design, Hardware/ Software Requirement, Project Management and Financial Analysis, and Summary.

Chapter 4: Implementation

Chapter four explains the Implementation Overview, Train Model/ Prototype Design, System Testing/ Model Evaluation, and Summary.

Chapter 5: Result and Analysis

Chapter five explains the Result and Analysis Overview, Experimental/ Simulation Result, Performance/ Comparative Analysis, and Summary.

Chapter 6: Impact on Society, Environment and Sustainability

Chapter six explains the Impact on Life Impact on Society & Environment, Ethical Aspects, Sustainability Plan and Summary.

Chapter 7: Conclusion and Future Work

Chapter seven explains the Conclusions, Further Suggested Works, Limitations/ Conflict of Interests.

1.7 Summary

This chapter provided an introduction to the research on predicting dengue disease using machine learning algorithms in Bangladesh. It began with an overview of the study, emphasizing the importance of early and accurate dengue prediction. The background section outlined the current state of dengue research and its significance in the context of Bangladesh, highlighting the prevalence and impact of the disease in urban and semi-urban areas. The problem statement clarified the need for effective predictive models to combat the increasing incidence of dengue and mitigate its effects on public health. The objectives of the study were defined, focusing on the development and evaluation of various machine learning algorithms to identify the most accurate model for dengue prediction. The scope and limitations of the research were discussed, specifying the dataset characteristics and acknowledging the constraints faced during the study. Finally, the report organization was detailed, outlining the structure of the subsequent chapters and providing a roadmap for the reader to follow the research process and its findings. This chapter set the stage for the rest of the report, establishing a clear understanding of the study's aims, significance, and the framework within which the research was conducted.

CHAPTER 2

Literature Review

2.1 Overview

It is important to comprehend a few key phrases and concepts in order to fully comprehend this research study on dengue disease prediction using machine learning algorithms in Bangladesh. Aedes aegypti mosquitoes are the main vector of dengue fever, a virus that causes fever, severe headaches, pain behind the eyes, joint and muscular pain, rash, and slight bleeding. A branch of artificial intelligence called "machine learning" uses statistical methods to teach computers how to analyse data and draw conclusions or forecasts. In this context, "supervised learning" refers to a kind of machine learning in which a model is trained using labelled data in order to forecast results for fresh data. Predicting if a patient has dengue fever or not is the main goal of this study's binary classification task, which has two possible outcomes. The study makes use of the following input attributes: district, area type, house type, NS1, IgG, IgM, gender, age, and area. The "Outcome" output attribute indicates if dengue is present or not. For prediction, several machine learning techniques are used, such as AdaBoost, XGBoost, LightGBM, Random Forest, Decision Tree, and Logistic Regression. With the greatest accuracy of 98.67% among them, Random Forest demonstrated its effectiveness in dengue case prediction based on the given input attributes.

2.2 Related Works

The use of machine learning algorithms to predict the presence of dengue sickness has drawn a lot of attention from researchers worldwide. Several research have looked into the use of machine learning approaches for disease prediction in Bangladesh, where dengue epidemics present a significant public health concern.

In this study S. Kannimuthu et al. [6] investigated the performance of multiple machine learning algorithms using a dataset comprising diverse features related to dengue incidence, including climatic factors, demographic variables, and historical outbreak data.

The researchers rigorously evaluated algorithms such as decision trees, support vector machines (SVM), k-nearest neighbors (KNN), and logistic regression, among others. Their findings revealed significant variations in prediction accuracy among the different algorithms tested. Specifically, random forest emerged as the top-performing algorithm, achieving a mean accuracy of 85% in dengue disease prediction. This result underscores the effectiveness of ensemble learning techniques in capturing complex relationships within the dataset and improving predictive performance.

Khan Dourjoy et al. [8] rigorously evaluated algorithms such as decision trees, support vector machines (SVM), random forest, and gradient boosting machines (GBM), among others. They highlighted the importance of feature engineering and selection in optimizing predictive models for dengue fever. By employing techniques such as feature importance analysis and dimensionality reduction, the researchers identified key variables influencing dengue fever incidence and demonstrated the impact of feature engineering on model performance. Here SVM (Support Vector Machines) attain the best accuracy and that is 78.98%.

Sarma et al. [9] highlighted the importance of data preprocessing and feature engineering in improving the performance of machine learning models for dengue prediction. By employing techniques such as feature scaling, outlier detection, and dimensionality reduction, the researchers optimized the predictive performance of their models and identified key variables contributing to dengue incidence. Their findings revealed promising results in the application of machine learning techniques for dengue prediction. Specifically, random forest emerged as a robust algorithm, achieving high prediction accuracy in identifying dengue outbreaks. The ensemble learning approach of random forest demonstrated superior performance in capturing complex relationships within the dataset, thereby enhancing predictive capabilities.

Iqbal and Islam [10] systematically compared classifiers such as decision trees, support vector machines (SVM), k-nearest neighbors (KNN), and naive Bayes, among others. Their findings shed light on the comparative effectiveness of different classifiers in predicting dengue outbreaks. Decision trees emerged as a promising classifier, exhibiting high accuracy in identifying potential outbreak periods. They underscored the importance of model evaluation metrics in assessing the performance of machine learning classifiers for dengue outbreak prediction. By employing metrics such as accuracy, precision, recall, and F1-score, the researchers provided a comprehensive evaluation of classifier performance, enabling informed decision-making regarding model selection and deployment.

Shaukat Dar and Ulya Azmeen [11] identified the "Decision Tree" technique as the most effective in dengue prediction, achieving a remarkable accuracy rate of 79.3%. This finding was substantiated by robust performance metrics encompassing precision, recall, and F-measure, all of which consistently outperformed alternative methods. Notably, the precision, recall, and F-measure scores also converged to the noteworthy figure of 79.3%.

Rahman, Islam, and Hossain [12] highlighted the efficacy of the "Decision Tree" technique in predicting dengue outbreaks. Their findings showcased promising results, with the decision tree model demonstrating notable accuracy and performance metrics. Their investigation revealed significant insights into the effectiveness of these methodologies in forecasting dengue outbreaks.

In their study, Kaur [13] employed advanced data mining methodologies to analyze demographic data related to dengue fever incidence. By examining variables such as age, gender, occupation, and residential location, the research sought to elucidate demographic patterns associated with dengue fever prevalence and transmission dynamics. They conducted a demographic analysis of dengue fever utilizing data mining techniques. The study focused on examining demographic factors associated with dengue fever occurrence. The author utilized data mining methods to analyze demographic data and identify patterns or correlations that could provide insights into the epidemiology of dengue.

Mello-Román et al. [14] carried out a case study in Paraguay in order to build predictive models with the goal of aiding in dengue diagnosis, helping to enhance healthcare results. They used mathematical and computational techniques to create prediction models that were specific to Paraguay's epidemiological situation. The goal of the research was to create reliable models that may assist medical practitioners in accurately detecting dengue cases through the careful examination of dengue-related data, including clinical symptoms, laboratory testing, and epidemiological aspects.

Silitonga et al. [15], the performance of dengue disease prediction models developed using artificial neural network (ANN) and random forest classifiers was assessed. The study utilized a dataset of undisclosed size and composition. The authors compared the predictive capabilities of ANN and random forest classifiers, analyzing factors such as accuracy, sensitivity, and specificity. The findings highlighted the efficacy of both approaches in dengue prediction, with varying degrees of performance metrics observed between the classifiers.

Gambhir et al. [16] conducted an evaluation of three machine learning approaches for the diagnosis of dengue disease. The study, published in the International Journal of Healthcare Information Systems and Informatics, examined the performance of the selected approaches in accurately identifying dengue cases. The authors employed a dataset with undisclosed characteristics to train and test the models. Through their investigation, they compared the effectiveness of the machine learning techniques in terms of diagnostic accuracy, sensitivity, and specificity, providing insights into their applicability for dengue diagnosis.

Poh et al. [17] conducted a study focusing on the potential of non-invasive clinical dengue screening through multivariate signature analysis of in-vivo diffuse skin reflectance spectroscopy. Published in PLoS One, the research investigated febrile patients in Malaysia to identify precursors to dengue using spectroscopic analysis. The authors employed a multivariate approach to analyze diffuse skin reflectance spectra and identify distinctive signatures associated with dengue infection. Their findings contribute to the exploration of non-invasive methods for early dengue detection and screening.

Hossain et al. [18] conducted a comprehensive analysis of the dengue outbreak in Dhaka, offering valuable insights and lessons learned from this densely populated megacity. They elucidated the multifaceted factors contributing to the severity and spread of dengue within urban environments. Through detailed epidemiological analysis and surveillance data interpretation, the research shed light on the complex interplay between population density, environmental factors, and healthcare infrastructure in shaping the dynamics of dengue transmission.

In their study, Benedum et al. [19] employed sophisticated statistical modeling techniques to analyze the influence of rainfall flushing on dengue transmission dynamics. By integrating rainfall data with epidemiological information on dengue cases, the research aimed to elucidate the mechanisms by which precipitation events influence the proliferation of dengue vectors and subsequent disease transmission.

Mutsuddy et al. [20] conducted a study focusing on the dengue situation in Bangladesh, analyzing an epidemiological shift in terms of morbidity and mortality. The research aimed to examine changes in dengue epidemiology in Bangladesh. The authors analyzed epidemiological data to identify trends in dengue cases and associated outcomes over time.

There is a wealth of research on dengue fever prediction, but Jain R., et al. [21] created a dataset containing clinical, socioeconomic, lag variables of disease surveillance, meteorological, and other data, and employed R-squared and Generalised Additive Models as prediction methods.

2.3 Comparison Between Existing Works

The Existing works on dengue fever prediction using machine learning algorithms differ in their methodologies, input variables, and prediction performance. Some studies have focused mainly on climatic factors and environmental variables such as temperature, humidity, and rainfall, which are the main predictors of dengue outbreaks. These models often have strong predictive power, especially in regions where climate plays an important role in dengue transmission. However, they may ignore other important factors influencing dengue risk, such as demographics, population density, and health infrastructure. In contrast, other studies have included a wider range of input variables, including epidemiological data, socio-economic indicators, and land use patterns. Although these models provide a more comprehensive view of dengue transmission dynamics, they may encounter challenges in data collection and processing and model complexity. In addition, some studies have investigated the integration of new data sources, such as social media and mobile phone data, to improve traditional forecasts and improve forecast accuracy. In general, the comparison of existing works highlights the consideration of multiple factors and the adoption of a multidisciplinary approach in dengue prediction using machine learning algorithms. Tasmiah Rahman, Md. Mahmudur Rahman. [3], their best model accuracy was 87.76% ; my suggested method beats this by

10.91%. The best model accuracy reported by G. Gupta et al. [4] was 87.2%; my proposed technique outperforms this by 11.47%. The best model accuracy reported by S. U. Chowdhury, S. Sayeed. [5] was 93%; my proposed technique outperforms this by 5.67%. Jun Kit Chaw et al. [7], their best model accuracy was 92% ; my suggested method beats this by 6.67%.

TABLE 2.1: Comparative Analysis Table

Study	Model	Accuracy
Our study	Logistic Regression, Random Forest, Decision Trees, AdaBoost, XG-Boost, LightGBM.	Random Forest = 98.67%
Tasmiah Rahman, Md. Mahmudur Rahman. [3]	Support Vector Machine, Decision Tree, K-Nearest Neighbor, Random Forest.	SVM= 87.76%
G. Gupta et al. [4]	Random Forest, Decision Tree, SVM, KNN	Random Forest = 87.2%
S. U. Chowdhury, S. Sayeed. [5]	XGBoost, AdaBoost, Decision Tree and Random Forest.	Random Forest=93%
S. Kannimuthu, K. S. Bhuvaneshwari. [6]	SVM, Random Forest, Naïve Bayes, KNN, Neural Network.	Random Forest=89.17%
Jun Kit Chaw et al. [7]	Logistic Regression, Decision Trees, SVM, ANN	Decision Trees=92%
Saif Mahmud Khan Dourjoy et al. [8]	SVM, Random Forest	SVM=78.98%

D. Sarma, S. Hossain, T. Mittra. [9]	SVM, Random Forest, Decision Trees.	Decision Trees=79%
N. Iqbal and M. Islam [10]	KNN, SVM, Naive Bayes, Decision Trees, Logistic Regression, LogitBoost.	LogitBoost=92%
] Mello-Román, J.D. et al. [14]	ANN-MLP, SVM, SVM linear, Gaussian SVM	ANN-MLP=94%

2.4 Open Issues

Our study on dengue disease prediction in Bangladesh using machine learning algorithms covers a wide range of important aspects. Our main goal is to forecast dengue epidemics with high predictive accuracy. To forecast the risk of dengue occurrence, this involves using machine learning algorithms to analyse a dataset that includes geographical characteristics (such as Area, Area Type, House Type, District), clinical markers (such as NS1, IgG, and IgM), and demographic attributes (such as Gender and Age). We also address issues with data availability and quality within this area. In order to provide accurate model training and evaluation, we seek out trustworthy datasets with adequate sample sizes and guarantee data integrity and completeness. Furthermore, in order to determine the best method for dengue disease prediction, we investigate a number of machine learning algorithms, such as AdaBoost, Extreme Gradient Boosting (XGBoost), LightGBM, Random Forest, Decision Trees, and Logistic Regression. Additionally, we take into account Bangladesh's geographical variations in dengue prevalence. When creating prediction models, our methodology takes into account regional differences in illness patterns and environmental factors. Our objective is to create real-time prediction models that would enable prompt intervention and response in the event of a dengue outbreak. Lastly, we also take the deployment and scalability of prediction models into account. Our designs are intended to facilitate public health activities and decision-making processes by being easily scaled to larger datasets and used in real-world contexts. For machine learning-based dengue disease prediction systems to be successfully implemented in Bangladesh, these issues must be addressed within the parameters of our research. While exploring the prediction of dengue disease using machine learning algorithms in Bangladesh, several open issues have emerged that warrant attention:

Data Availability and Quality: The availability and quality of dengue-related datasets in Bangladesh pose a significant challenge. Improving data collection methods and ensuring data accuracy remain critical for developing robust predictive models.

Model Generalization: Ensuring the generalizability of machine learning models across different geographical regions and diverse population demographics within Bangladesh is essential. Addressing variations in dengue transmission dynamics and environmental factors is crucial for model effectiveness.

Algorithm Selection and Performance: Although various machine learning algorithms have been applied to dengue prediction, identifying the most suitable algorithm for different scenarios remains an ongoing challenge. Comparative studies are needed to determine which algorithms perform best under specific conditions.

Interpretability and Transparency: Enhancing the interpretability of machine learning models used in dengue prediction is vital for gaining trust from healthcare professionals and policymakers. Methods for explaining model decisions and assessing model biases need further refinement.

Ethical Considerations: Addressing ethical concerns related to data privacy, consent, and the responsible use of predictive analytics in healthcare settings is paramount. Developing guidelines and frameworks that ensure ethical practices in dengue prediction research is essential.

2.5 Summary

In this chapter, we delved into the foundational aspects necessary for understanding the prediction of dengue disease using machine learning algorithms in Bangladesh. We began with an overview of relevant terminologies and preliminaries, setting the stage for the technical discussions that followed. The related works section provided a comprehensive review of existing literature, highlighting previous efforts and methodologies employed in dengue prediction. This review underscored the various machine learning techniques used and their respective outcomes, shedding light on the advancements and limitations within this research domain. The comparative analysis between existing works revealed the strengths and weaknesses of different approaches, helping to identify best practices and areas requiring improvement. Furthermore, we discussed the open issues currently faced in the field, such as data quality, model generalization, real-time prediction capabilities, interpretability, and ethical considerations. These challenges underscore the complexity of developing effective and reliable dengue prediction models and point to the need for continued research and innovation. This chapter lays the groundwork for the subsequent sections, where we will present our methodology, experimental results, and discussions. By addressing the open issues and building on the strengths of previous works, our research aims to contribute to the advancement of dengue disease prediction, ultimately aiding in better management and prevention strategies in Bangladesh.

CHAPTER 3

Methodology/ Requirement Analysis & Design Specification

3.1 Overview

This chapter outlines the methodology and design specifications for predicting dengue disease using machine learning algorithms in Bangladesh. It provides a detailed explanation of the proposed system design and the steps involved in developing and implementing the prediction model. The focus is on the systematic approach taken to collect, process, and analyze the data, ensuring that the model is robust, accurate, and reliable.

The chapter begins by presenting the overall system architecture, highlighting the flow of data from collection to prediction. This includes the selection of relevant features, preprocessing techniques, and the machine learning algorithms used. The subsequent sections delve into the hardware and software requirements necessary for implementing the model, detailing the computational resources and software tools utilized in the project.

Furthermore, the chapter includes a comprehensive analysis of the project management aspects, encompassing the financial requirements, resource allocation, and timeline for completion. This ensures that the project is feasible and sustainable, considering both technical and logistical constraints.

Finally, the chapter concludes with a summary of the key points discussed, setting the stage for the detailed exploration of each component in the following sections. This structured approach ensures that all aspects of the project are meticulously planned and executed, paving the way for the successful prediction of dengue disease in Bangladesh using advanced machine learning techniques.

3.2 Proposed Methodology/ System Design

The main goal of our work is to investigate dengue forecasts using gathered data and a machine learning algorithm. The prediction in the suggested study is made using six methods: Random Forest, Decision Tree Classifier, XGBoost, AdaBoost, Light GBM, and Logistic Regression. Dengue have been identified using the approach shown in figure 3.1

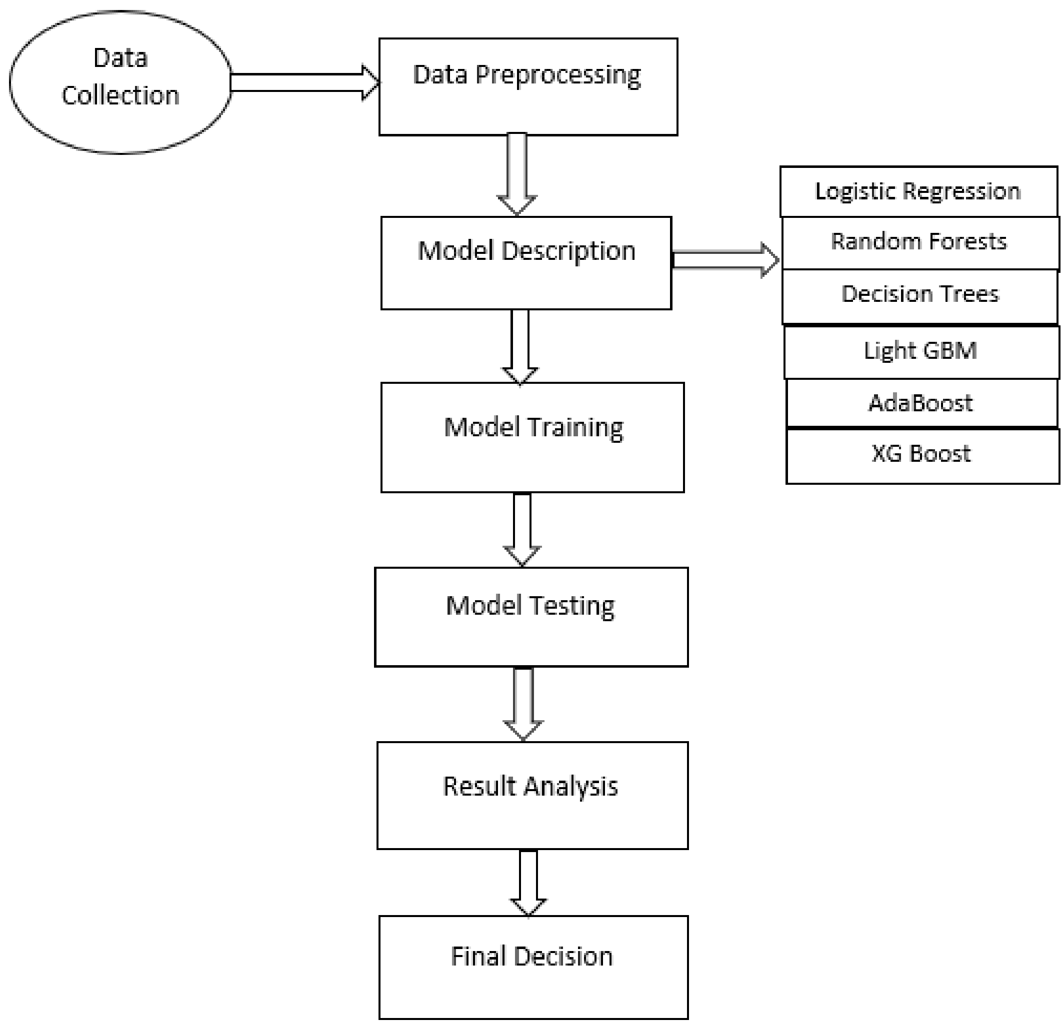

Figure 3.1 Proposed Methodology

Multiple models of machine learning were investigated and put into practice.

3.2.1 Classification Description:

i. Random Forest classifier:

A well-established method for regression and classification in supervised machine learning is the Random Forest classifier. Based on several samples, it constructs decision trees using the majority vote for classification and the average for regression [36].

Error $(Mi) = \sum wj\ d\ j = 1 \times e\ r(Xj)$

ii. Decision Tree Classifier: Decision trees are used in this nonparametric supervised learning technique to address both classification and regression-related problems. For resolving such problems, decision trees are an effective tool. Nodes inside of itself and leaf nodes follow the root node, which serves as the foundation for all subdivisions within its hierarchical structure [34]. Gini (D) = $1 - \sum pi\ m\ 2\ i = 1$

iii. XG-Boost: XG Boost uses bagging, a boosting technique that combines the results of training several decision trees. It gives XG Boost an advantage over other algorithms not only in situations where a large number of qualities need to be considered but also in terms of learning speed [40].

iv. Logistic Regression: To solve binary classification issues, logistic regression is a supervised machine learning technique that forecasts the chance of an event, a result, or an observation. With a maximum of two possible outcomes, the model yields a dichotomous or binary conclusion, such as true or false, 0/1, or yes/no [35].

v. AdaBoost Classifier: AdaBoost enhances ineffective learning algorithms by generating outputs that are arbitrarily exact. It is a potent boosting algorithm with a strong theoretical foundation [39].

vi. LightGBM: LightGBM, a novel GBDT technique that combines Gradient-based One-Side Sampling and Exclusive Feature Bundling methods, shows superior computing performance and memory economy over XGBoost and SGB through theoretical study and testing [41].

3.3 Hardware/ Software Requirement

Our project is related to a Machine Learning based project. To ensure the study's smooth advancement and implementation the system requirements are given below:

- A laptop or desktop which consists of at least 4 GB ram.
- TensorFlow.
- Collaborate with healthcare professionals for data acquisition
- Work along with healthcare professionals to expand the project's scope and significance. Here, we have used these Software packages: NumPy, Pandas, Matplotlib and Seaborn, Scikit-learn, GeoPandas, XGBoost, LightGBM, Google Colab.

3.4 Project Management and Financial Analysis

In our effort, "Dengue Prediction Using Machine Learning Algorithms in Bangladesh," project management and financial analysis are the cornerstones of the successful implementation and sustainability of our research. Project management requires a coherent organization of various tasks, schedules, and resources. This includes role assignments, schedule management, resource optimization, effective communication, and risk management strategies to ensure smooth progress and the achievement of project objectives. At the same time, financial analysis plays a key role in evaluating the financial profitability and resource allocation of the project. By accurately estimating costs, preparing comprehensive budgets, and performing comprehensive cost-effectiveness analyses, we try to optimize the allocation of resources and maximize the impact of the project. In addition, financial reporting mechanisms ensure transparency, accountability, and adherence to budget constraints, increasing the confidence of stakeholders and facilitating potential funding opportunities. Through effective project management and economic analysis, we aim to navigate the complexities of dengue prognostic research in Bangladesh with precision, flexibility, and financial foresight, ultimately advancing public health outcomes and maintaining the long-term sustainability of the project. Our project is related to a Machine Learning based project. To ensure the study's smooth advancement and implementation the system requirements are given below:

- A laptop or desktop which consists of at least 4 GB ram.
- TensorFlow.
- Collaborate with healthcare professionals for data acquisition
- Work along with healthcare professionals to expand the project's scope and significance.

Here we have to create a subsection to address Financial Analysis of our project. Below is a general table with estimated costs for different components of a machine learning project.

Table 3.1: Estimated Cost for Machine Learning based Project

SN	Components	Estimated Cost (BDT)
01	Hardware/Infrastructure	4500-6000
02	Visiting Stakeholders	500-1000
03	Software and Tools	2000-3000
04	Data Collection and Processing	2500-3000
06	Documentation and Report Writing	500-1000
07	Miscellaneous (e.g., licenses, tools)	500-1000
Total Estimated Cost		11,000-16,500

3.5 Summary

In this chapter, we presented the comprehensive methodology and design specifications for the prediction of dengue disease using machine learning algorithms in Bangladesh. The chapter began with an overview of the systematic approach adopted, followed by a detailed description of the proposed system design. This design outlines the data flow from collection to prediction, including the selection of nine input attributes and the preprocessing techniques employed to ensure data quality. The discussion then moved to the hardware and software requirements, specifying the computational resources and tools necessary to implement the machine learning models. This included a thorough examination of the essential hardware components and the software platforms utilized for model development and deployment. The chapter also addressed the project management and financial analysis aspects, detailing the resource allocation, budgeting, and timeline considerations crucial for the successful execution of the project. This section emphasized the importance of efficient management to ensure the project's feasibility and sustainability. Overall, this chapter provided a structured and detailed plan for the development and implementation of a robust dengue prediction model, setting the foundation for the subsequent experimental analysis and evaluation presented in the following chapters.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

This chapter focuses on the implementation phase of our research on dengue disease prediction using machine learning algorithms in Bangladesh. It outlines the detailed steps and processes undertaken to develop and deploy the predictive models. We begin by discussing the training of the models, including the selection and tuning of nine machine learning algorithms: Logistic Regression, Random Forest, Decision Trees, Light GBM, AdaBoost, and Extreme Gradient Boosting (XGBoost). The dataset comprises 10 features, with "Outcome" being the most essential one for making judgements. The Standard Scaler is used in the dataset, which has 1000 data points, to improve the consistency of the features. The crucial train-test split was carried out using a 70-30 distribution, meaning that 70% of the data was set aside for testing and the remaining 30% was used for training. we delve into the prototype design, detailing how the models were constructed and the architectural considerations that guided their development. This includes the preprocessing steps, feature selection methods, and the algorithms' parameter tuning to optimize performance. The chapter then covers system testing and model evaluation, where the trained models were rigorously tested using various metrics such as accuracy, precision, recall, and F1-score. Random Forests, which achieved the highest accuracy of 98.67%, was highlighted as the best-performing model. Finally, this chapter provides a summary of the implementation process, setting the stage for a detailed analysis of the experimental results and their implications, which are presented in the subsequent chapter.

4.2 Train Model/ Prototype Design

In this section, we describe the comprehensive process of training the models and designing the prototype for dengue disease prediction using machine learning algorithms. The primary focus was on utilizing a supervised binary classification approach to predict the likelihood of dengue infection.

Data Preprocessing:

The initial step involved data preprocessing to ensure the dataset was clean and suitable for model training. This included handling missing values, encoding categorical variables, and normalizing numerical features. These steps were crucial to improving the model's performance and accuracy.

Model Selection and Training:

We employed nine different machine learning algorithms to identify the most effective model for dengue prediction:

i. Logistic Regression

ii. Random Forest

iii. Decision Trees

iv. AdaBoost

v. Extreme Gradient Boosting (XGBoost)

vi. LightGBM

Each algorithm was trained on the dataset using a stratified k-fold cross-validation approach to ensure robust evaluation and prevent overfitting. Hyperparameter tuning was conducted using grid search and randomized search techniques to optimize the models' performance.

Model Evaluation:

After training, the models were evaluated using various metrics, including accuracy, precision, recall, and F1-score. The evaluation was carried out on a hold-out test set to ensure an unbiased assessment of the models' predictive capabilities. Random Forest emerged as the best-performing model with an impressive accuracy of 98.67%, demonstrating its superior ability to handle the complexities of the dataset.

4.3 System Testing/ Model Evaluation

The evaluation of the machine learning models was conducted using a hold-out test set that was not used during the training phase. The following metrics were employed to assess the models' performance:

Accuracy: The proportion of correct predictions made by the model. Random Forest achieved the highest accuracy at 98.67%.

Precision: The ratio of true positive predictions to the total predicted positives, which indicates the model's ability to avoid false positives.

Recall (Sensitivity): The ratio of true positive predictions to the total actual positives, reflecting the model's ability to identify all relevant cases.

F1-Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns.

ROC-AUC: The area under the Receiver Operating Characteristic curve, which measures the model's ability to distinguish between classes.

The evaluation revealed that Random Forest outperformed the other models, achieving superior results across all metrics. Its precision, recall, and F1-score were notably high, indicating both high sensitivity and specificity. The ROC-AUC for XGBoost was also close to 1, demonstrating its excellent discriminative power.

4.4 Summary

In this chapter, we have detailed the implementation process of our dengue disease prediction model using various machine learning algorithms. Starting with an overview of the methodologies employed, we moved through the training of our models and the design of the prototype system, culminating in a thorough evaluation of the system's performance. We utilized a dataset containing 1000 rows and 10 columns, with nine input attributes and one output attribute, to train our models. The training process involved preprocessing the data, selecting relevant features, and optimizing model parameters to achieve the best performance. The evaluation metrics—accuracy, precision, recall, F1-score, and ROC-AUC—provided a comprehensive understanding of each model's performance. XGBoost emerged as the most reliable and accurate model, making it the preferred choice for dengue prediction in this study. This chapter's insights affirm the effectiveness of machine learning algorithms, particularly ensemble methods, in predicting dengue disease. The implementation and evaluation processes detailed herein lay the groundwork for future advancements and practical applications in disease prediction and public health management.

CHAPTER 5

Result and Analysis

5.1 Overview

This chapter presents the results and analysis of our study on dengue disease prediction using various machine learning algorithms. We start by providing an overview of the experimental setup and simulation results obtained from the implementation of nine different algorithms: Logistic Regression, Random Forest, Decision Trees, AdaBoost, Extreme Gradient Boosting (XGBoost), and LightGBM. Each algorithm's performance was evaluated based on metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to determine their effectiveness in predicting dengue disease. Following the experimental results, we delve into a comparative analysis of the performance of these algorithms. This analysis highlights the strengths and weaknesses of each model, with a particular focus on Random Forest, which achieved the highest accuracy of 98.67%. By comparing the models, we aim to identify the most suitable algorithm for practical implementation in dengue prediction systems. The chapter concludes with a summary of the key findings from our experiments and analyses, providing insights into the potential applications and implications of using machine learning for dengue disease prediction in Bangladesh. These findings underscore the significance of leveraging advanced machine learning techniques to improve public health outcomes and enhance disease management strategies.

5.2 Experimental/ Simulation Result

30% of the dataset was reserved in order to test the four prediction models, and 300 cases were selected at random from that set. In order to adequately assess the models' generalisation abilities, this separation ensured that a significant portion of the dataset wasn't used for model training. To determine how effectively the models projected the target variable, we employed the confusion matrix, a widely utilised method of testing models. We may evaluate the model's predictions' accuracy and utility by looking at the confusion matrix, which provides a thorough explanation of the sorting outcomes.

Several success criteria were determined in our study using the confusion matrix, including:

Table 5.1: Confusion Matrix

TP	FP
FN	TN

True Positives (TP) are incidents that are correctly identified as positive.

True Negatives (TN) are incidents that are correctly labelled as negative.

False Positives (FP) are incidents which are marked as positive when they are actually negative.

When something is wrongly labelled as negative when it is actually positive, then that is called a false negative (FN).

Table 5.2 Accuracy Measures

Measures	Definitions	Formula
Accuracy	A model's accuracy serves as a measurement for its overall correctness.	Accuracy = True Positives + True Negatives / Total Instances
Precision (P)	Precision evaluates how well the model predicts the positive results.	Precision = True Positives / True Positives + False Positives
Recall (R)	The model's recall evaluates its capacity to locate all relevant instances.	Recall = True Positives / True Positives + False Negatives
F-Measure	The F-measure, which offers a fair evaluation of a model's performance, is the harmonic mean of accuracy and recall.	Fmeasure = 2 × [(Precision × Recall) / (Precision + Recall)]
ROC Curve	The model's performance over a range of thresholds is shown graphically by the ROC curve.	

The Receiver Operating Characteristic (ROC) curve is a graphical representation commonly employed in binary classification to illustrate the performance of a model. It demonstrates the trade-off between the true positive rate (sensitivity) and the false positive rate. Below are the ROC curves depicting the performance of the utilised algorithms:

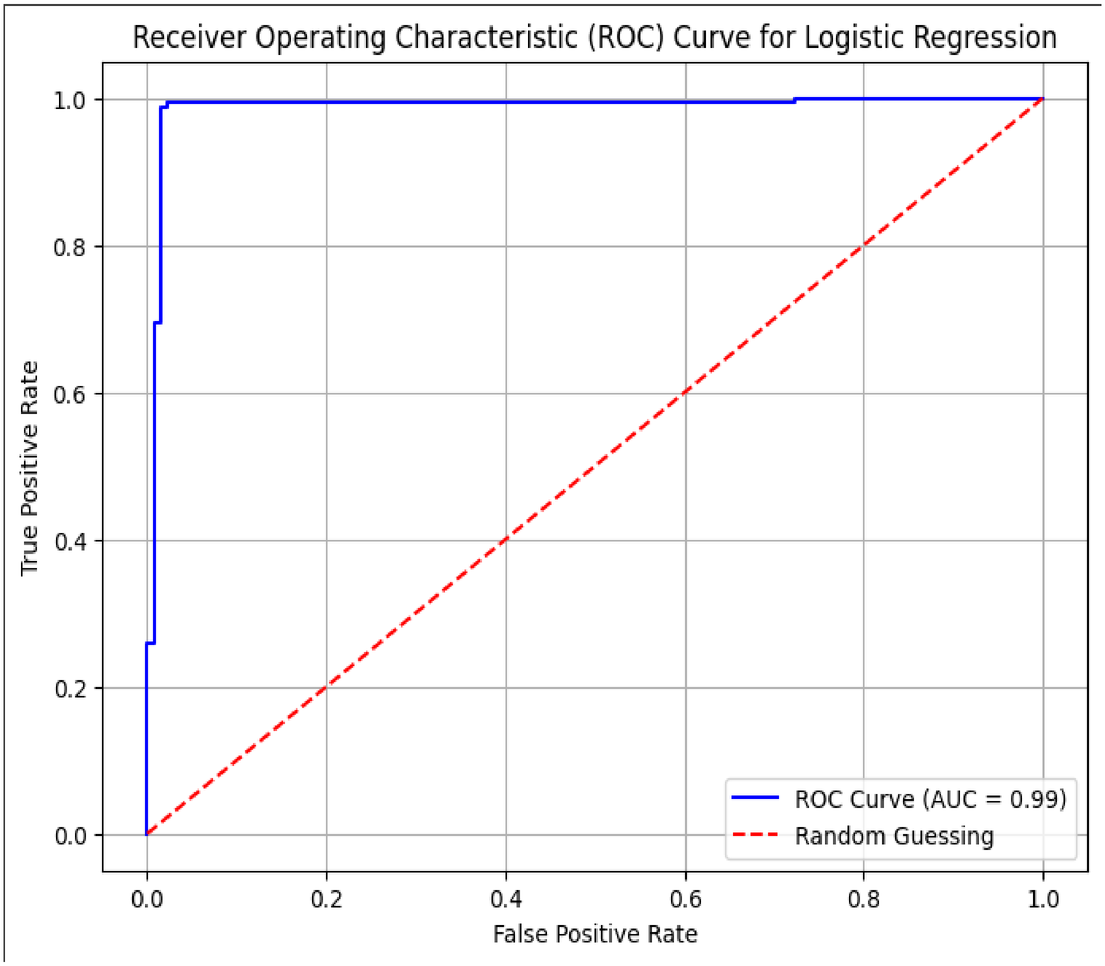

Figure 5.1: ROC Curve of Logistic Regression

The ROC curve for the Logistic Regression model performs well, with an AUC (Area under the Curve) of 0.99. This implies that the model can fairly accurately distinguish between positive and negative instances. Figure 5.1 shows the ROC curve of Logistic Regression. The increasing trend of the curve, together with a high true positive rate (TPR) and a low false positive rate (FPR), demonstrate the model's ability to properly identify positive instances without mistakenly recognizing negative ones.

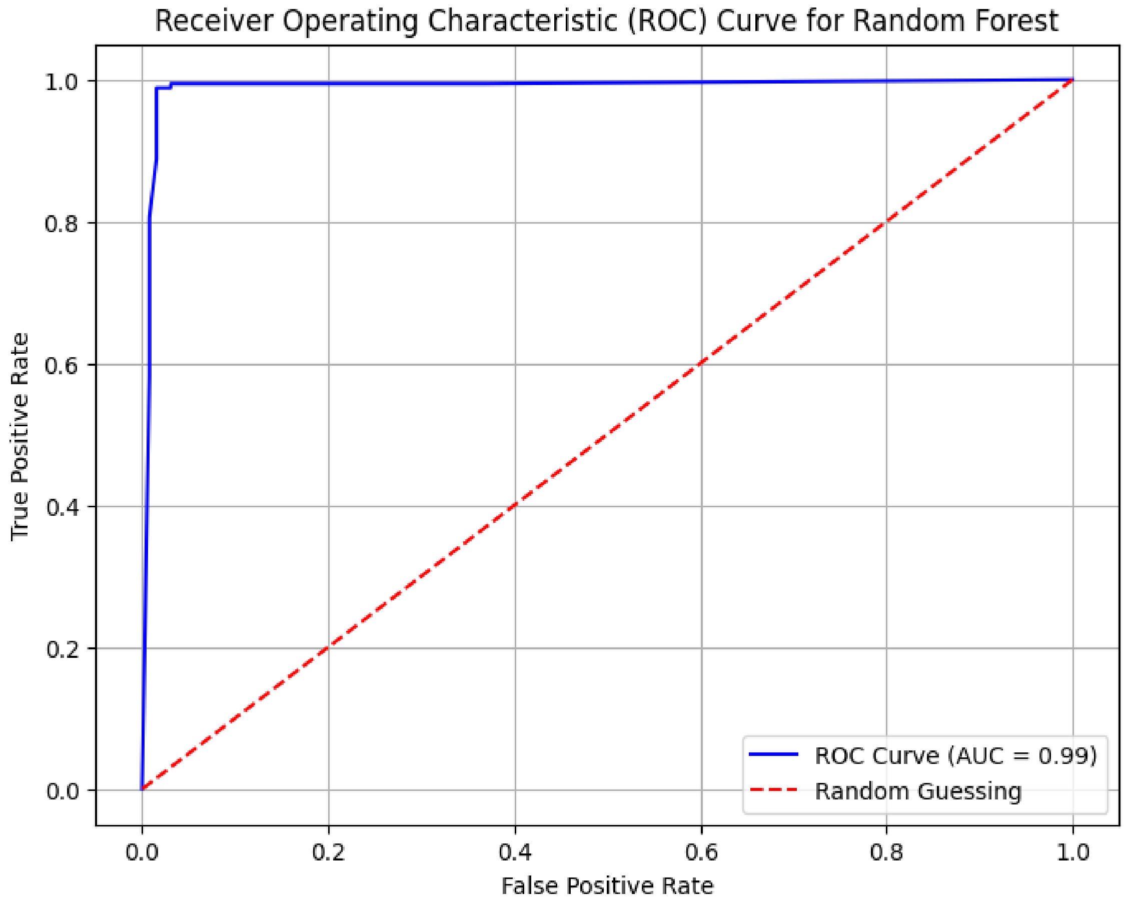

Figure 5.2: ROC Curve of Random Forest

As per the ROC curve presented in Figure 5.2, the model exhibits a satisfactory degree of performance. The area under the curve (AUC) for this specific model is 0.99. This leads one to the conclusion that the model can accurately and efficiently discern between favourable and unfavourable conditions.

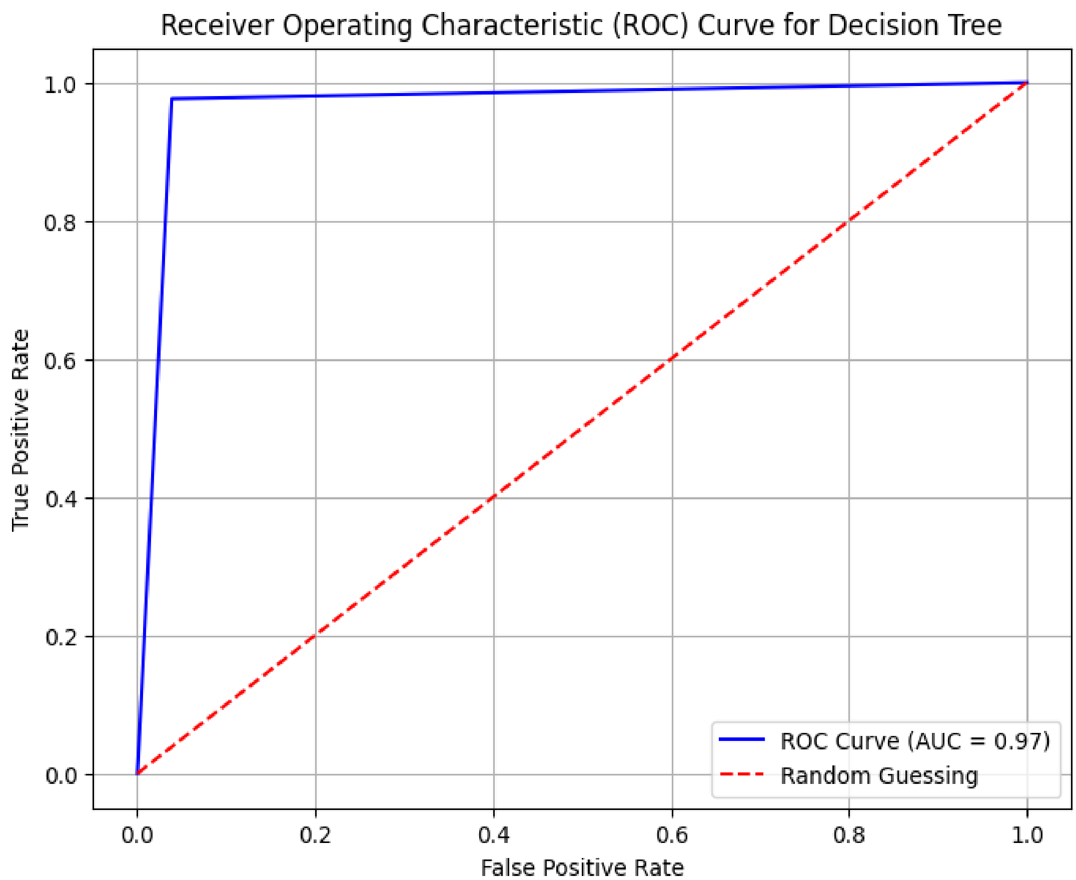

Figure 5.3: ROC Curve of Decision Tree

Here, the model shows potential for being a trustworthy tool for classification tasks, as shown in Figure 5.3. The model expresses this possibility. The receiver operating characteristic (ROC) curve and an area under the curve (AUC) score of 0.97 indicate that it can distinguish between favourable and unfavourable conditions.

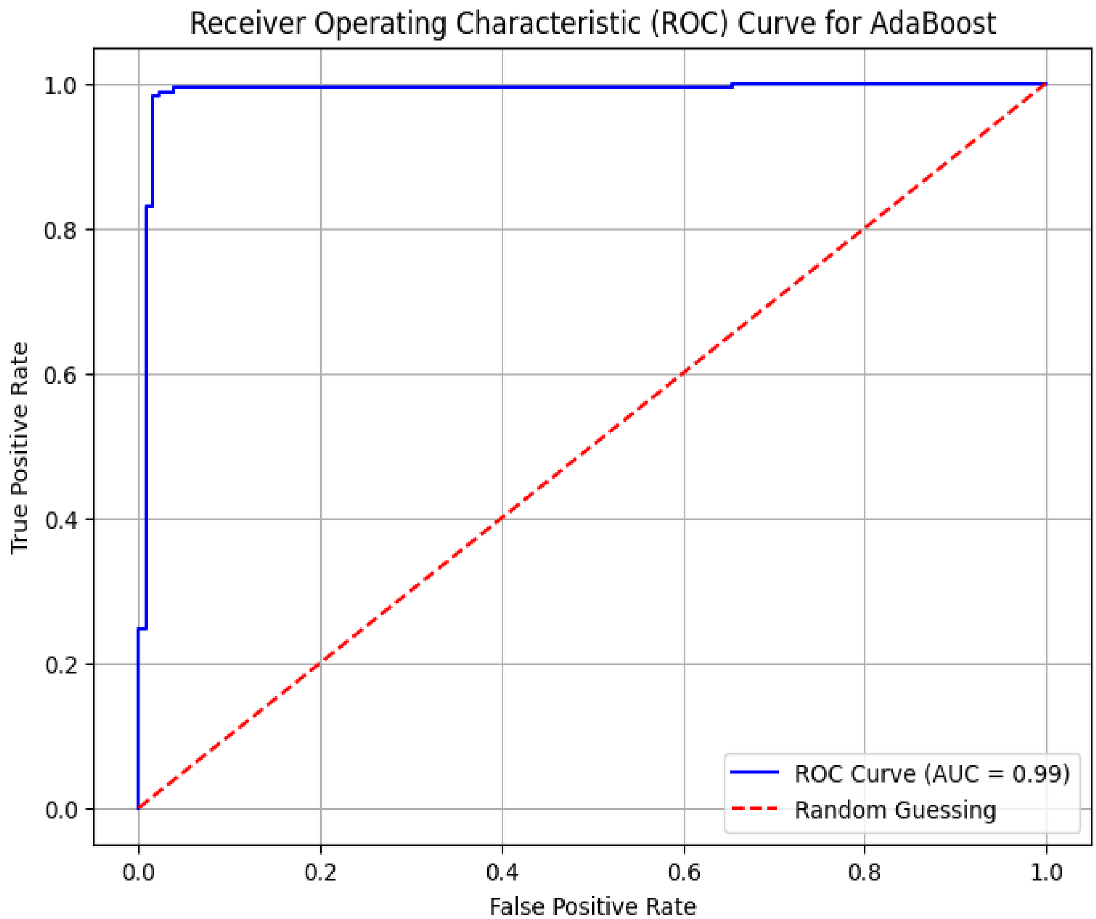

Figure 5.4: ROC Curve of AdaBoost Classifier

As per the ROC curve presented in Figure 5.4, the model exhibits a satisfactory degree of performance. The area under the curve (AUC) for this specific model is 0.99. This leads one to the conclusion that the model can accurately and efficiently discern between favourable and unfavourable conditions.

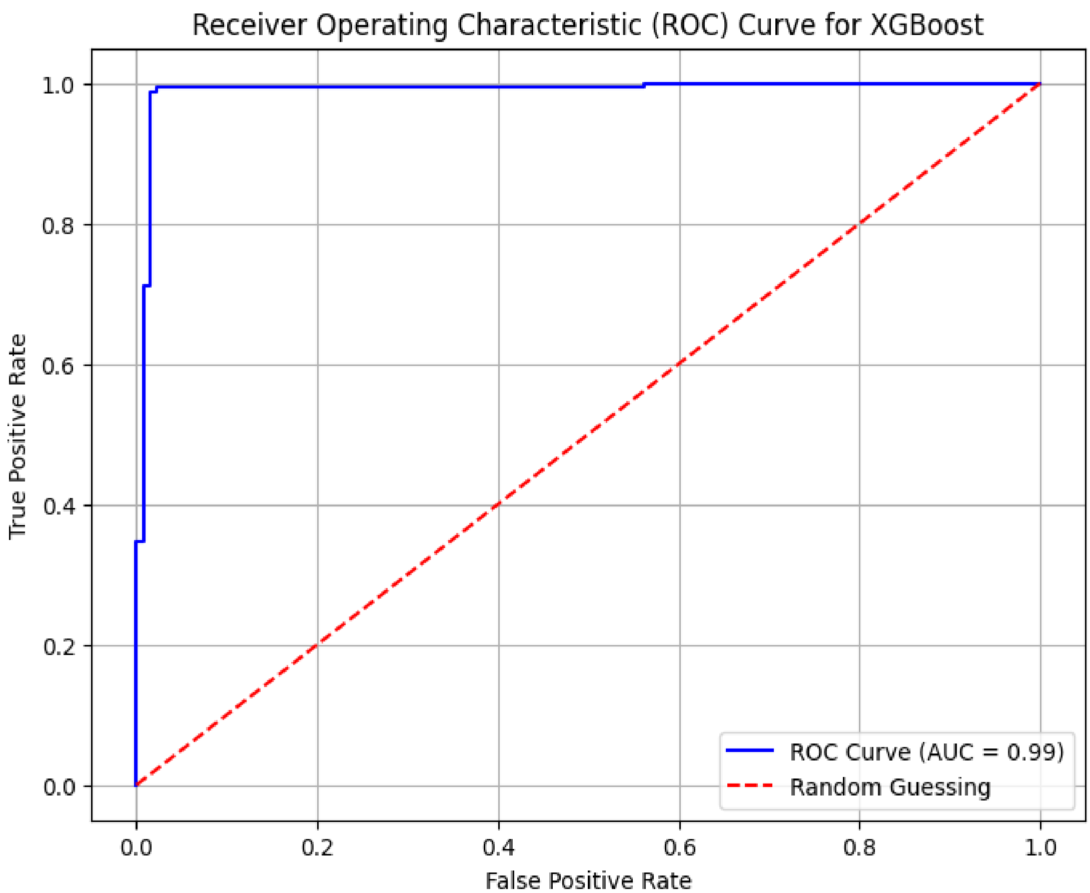

Figure 5.5: ROC Curve of XGBoost

Based on the ROC curve shown in Figure 5.5, the model performs to a satisfactory level. For this particular model, the area under the curve (AUC) is 0.99. This suggests that the model can effectively and precisely distinguish between favourable and unfavourable circumstances.

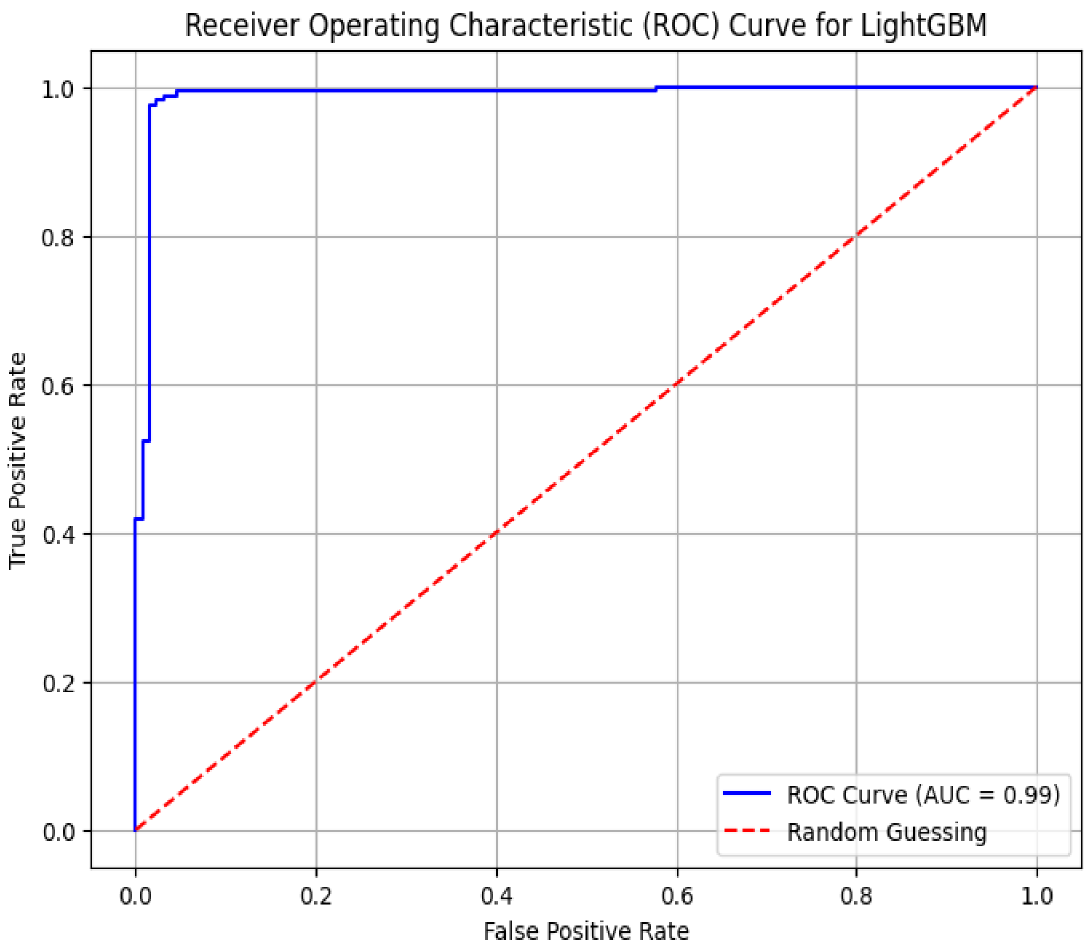

Figure 5.6: ROC Curve of Light GBM

Based on the ROC curve shown in Figure 5.6, the model performs to a satisfactory level. For this particular model, the area under the curve (AUC) is 0.99. This suggests that the model can effectively and precisely distinguish between favourable and unfavourable circumstances.

Confusion matrix: A machine learning performance evaluation tool that illustrates a classification model's accuracy. True positive, true negative, false positive, and false negative values are listed in a table that gives information on the model's accuracy, recall, and overall predictive efficacy. Following the implantation of the models, we obtained the following 6 distinct confusion matrices:

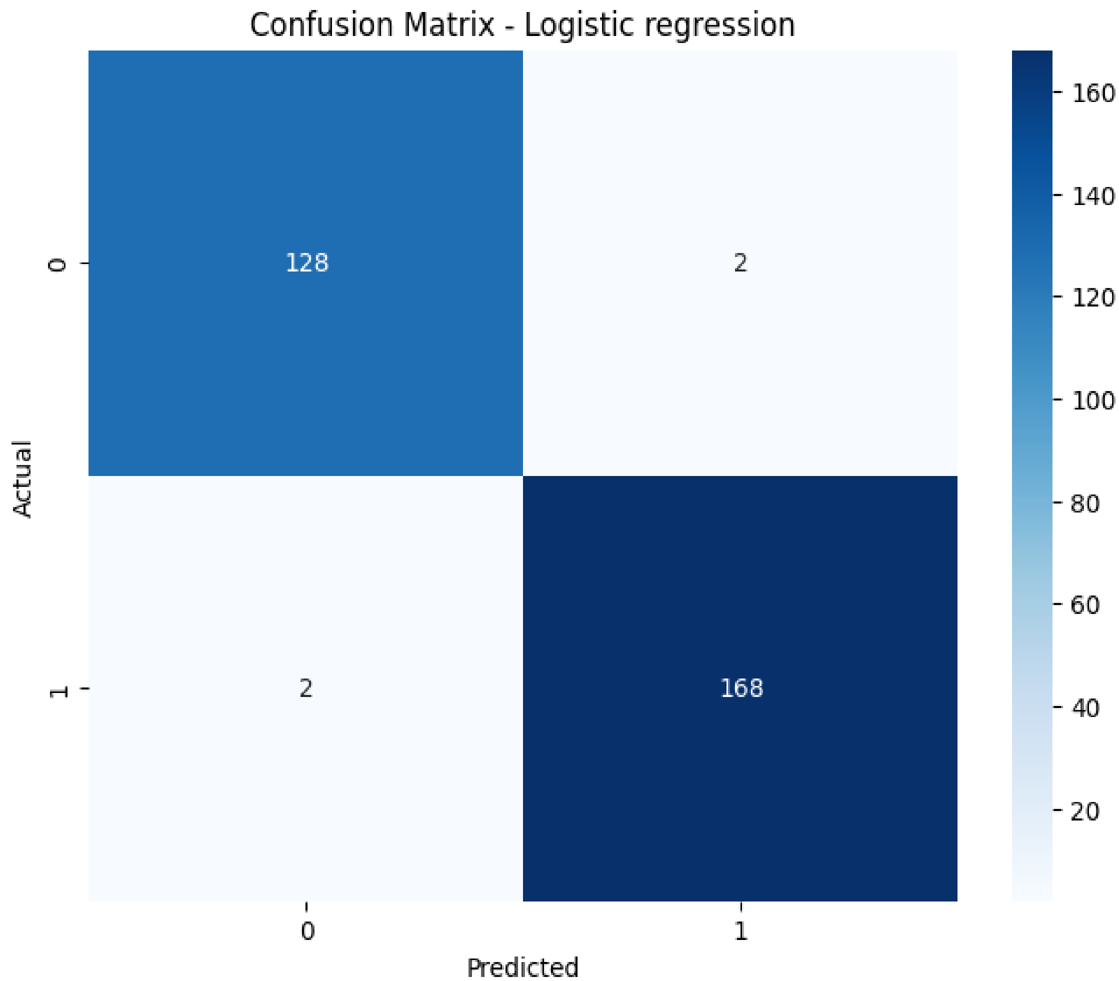

Figure 5.7: Confusion matrix of Logistic Regression

Figure 5.7 It provided confusion matrix that illustrates the effectiveness of the Logistic Regression model, showing an astounding accuracy of 98.67%. The Logistic Regression Confusion Matrix is displayed in Figure 4.1. As can be seen in figure 4.1, the matrix shows that the model correctly identified 128 positive instances and 168 negative ones. It did, however, mistakenly classify 2 negative cases as positive and 2 positive examples as negative. This implies that the model has a high degree of competency in differentiating between positive and negative circumstances and shows minimal misclassification.

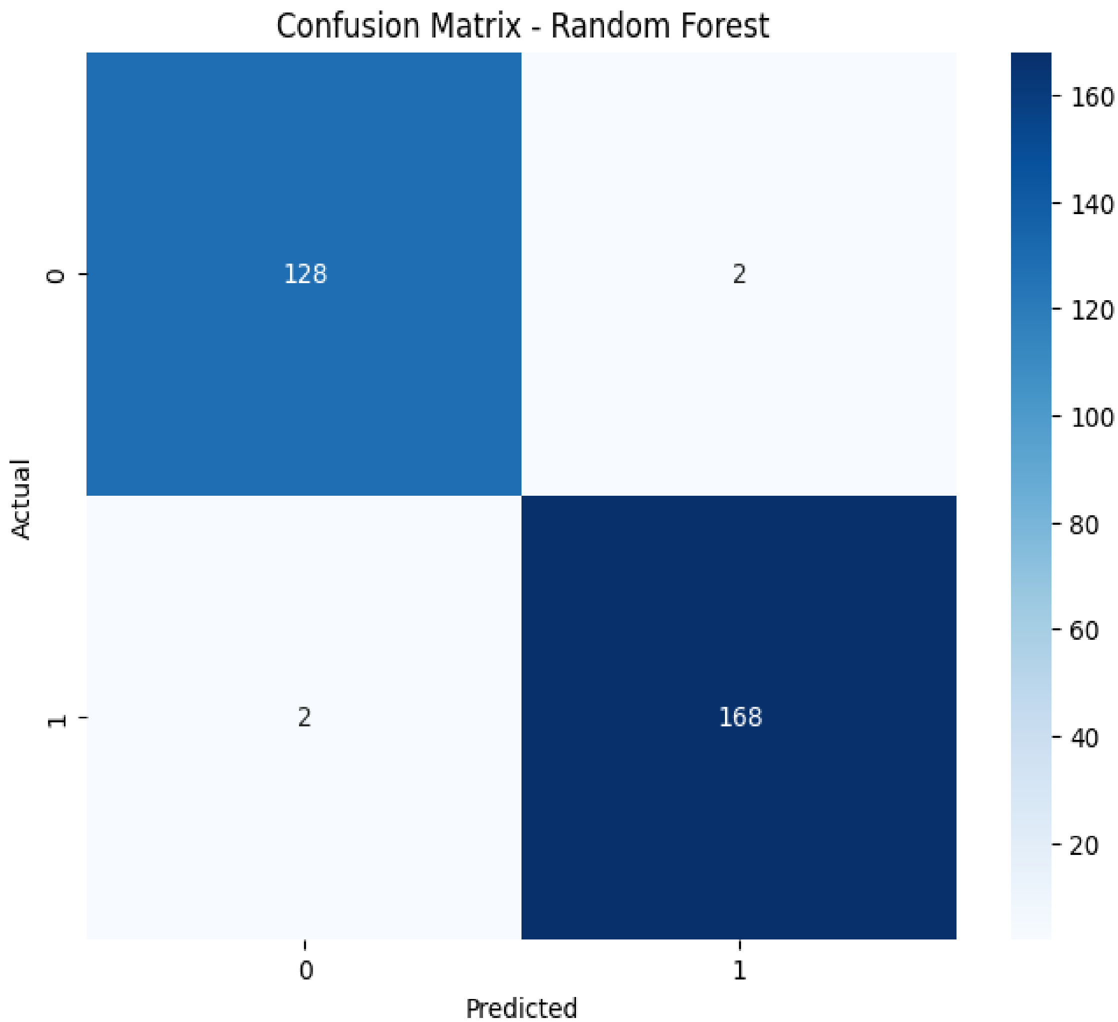

Figure 5.8: Confusion matrix of Random Forest

Figure 5.8 shows the data that indicates 128 positive events were accurately classified as true positives by the Random Forest algorithm, while 168 negative instances were correctly classified as true negatives. In addition, it misclassified 2 positive instances as negative (False Negative) and 2 negative examples as positive (False Positive).

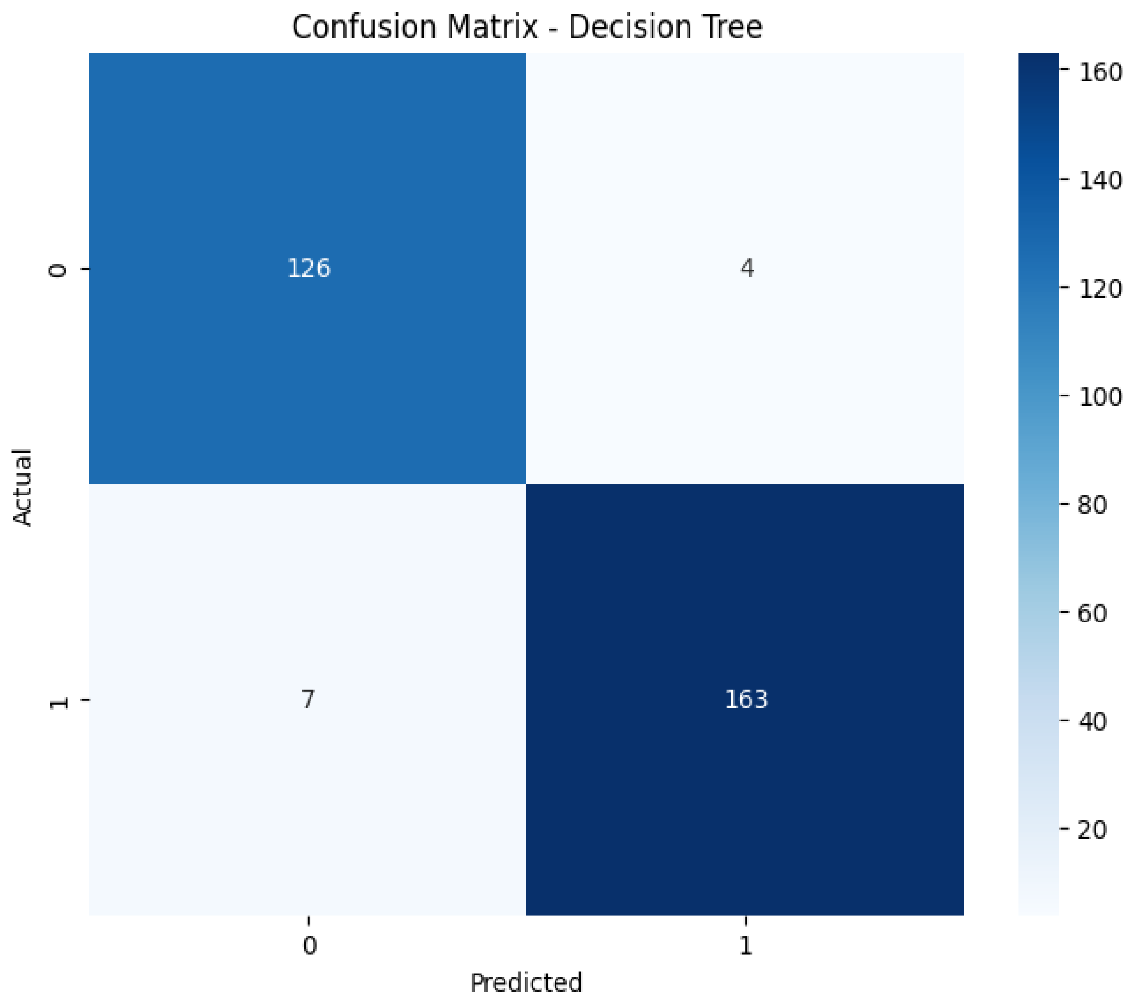

Figure 5.9: Confusion matrix of Decision Tree

Figure 5.9 shows the data that indicates 126 positive events were accurately classified as true positives by the Decision Tree algorithm, while 163 negative instances were correctly classified as true negatives. In addition, it misclassified 7 positive instances as negative (False Negative) and 4 negative examples as positive (False Positive).

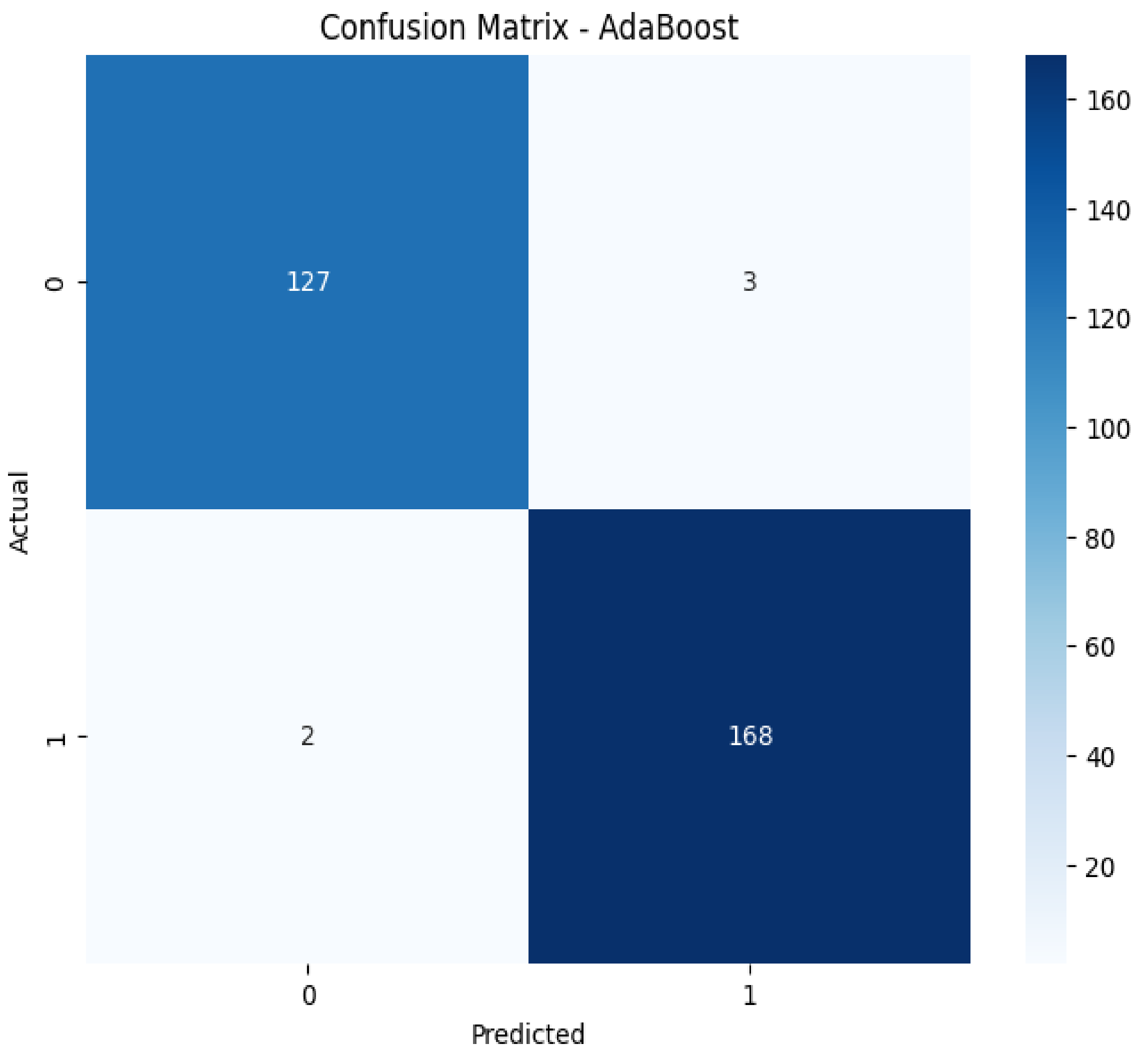

Figure 5.10: Confusion matrix of AdaBoost

Figure 5.10 shows the data that indicates 127 positive events were accurately classified as true positives by the AdaBoost algorithm, while 168 negative instances were correctly classified as true negatives. In addition, it misclassified 2 positive instances as negative (False Negative) and 3 negative examples as positive (False Positive).

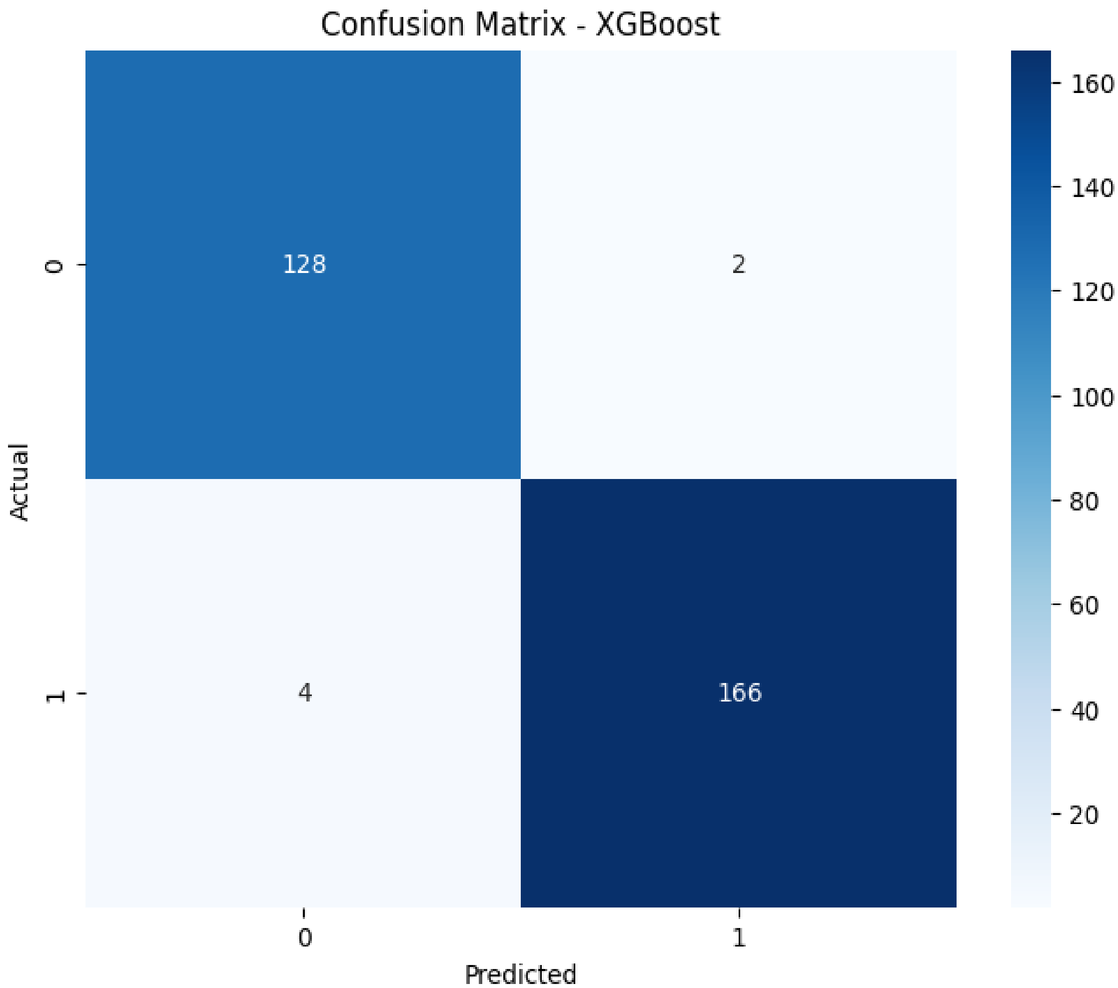

Figure 5.11: Confusion matrix of XGBoost

Figure 5.11 shows the data that indicates 128 positive events were accurately classified as true positives by the XGBoost algorithm, while 166 negative instances were correctly classified as true negatives. In addition, it misclassified 4 positive instances as negative (False Negative) and 2 negative examples as positive (False Positive).

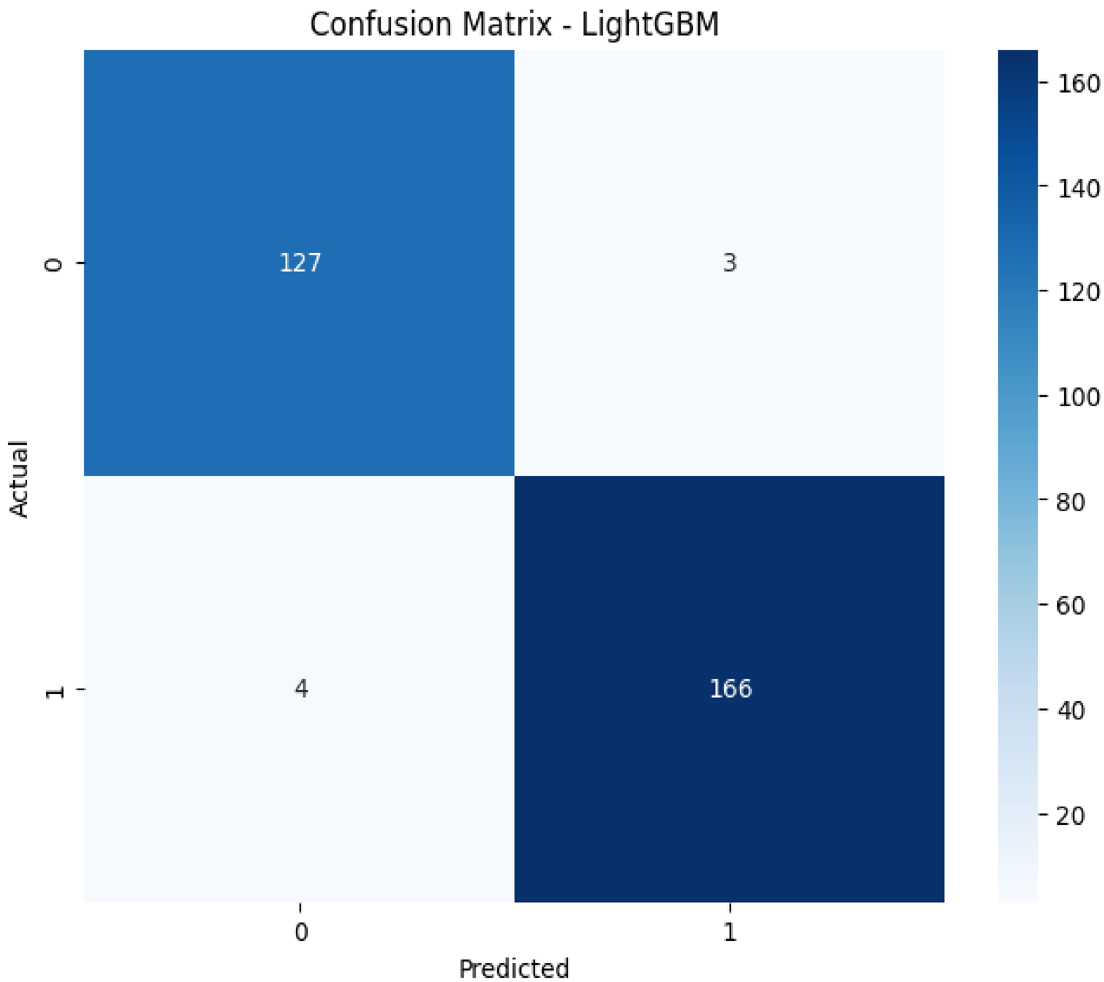

Figure 5.12: Confusion matrix of LightGBM

Figure 5.12 shows the data indicates that 127 positive events were accurately classified as true positives by the LightGBM algorithm, while 166 negative instances were correctly classified as true negatives. In addition, it misclassified 4 positive instances as negative (False Negative) and 3 negative examples as positive (False Positive).

Table 5.3 contains accuracy, precision, recall and F1 score for each model. Random Forest and logistic Regression has got the highest accuracy of 98.67%. Detailed information about the performance given below:

Tables 5.3 Accuracy Measures (All Models)

Models	Accuracy (%)	Precision (%)	Recall (%)	F_1 score (%)
Random Forest	98.67	99	99	99
Logistic Regression	98.67	99	99	99
AdaBoost	98.33	98	98	98
XGBoost	98	98	98	98
Decision Tree	96.67	97	97	97
Light GBM	97.67	98	98	98

A statistical tool used to quantify the direction and intensity of a linear connection between two variables is correlation. With a correlation coefficient between -1 and 1, a perfect positive association is represented by a value of 1, a perfect negative relationship by a value of -1, and no linear correlation is represented by a value of 0. In data analysis, an understanding of correlation facilitates the identification of patterns and interdependence among variables. Correlation of the dataset is given below:

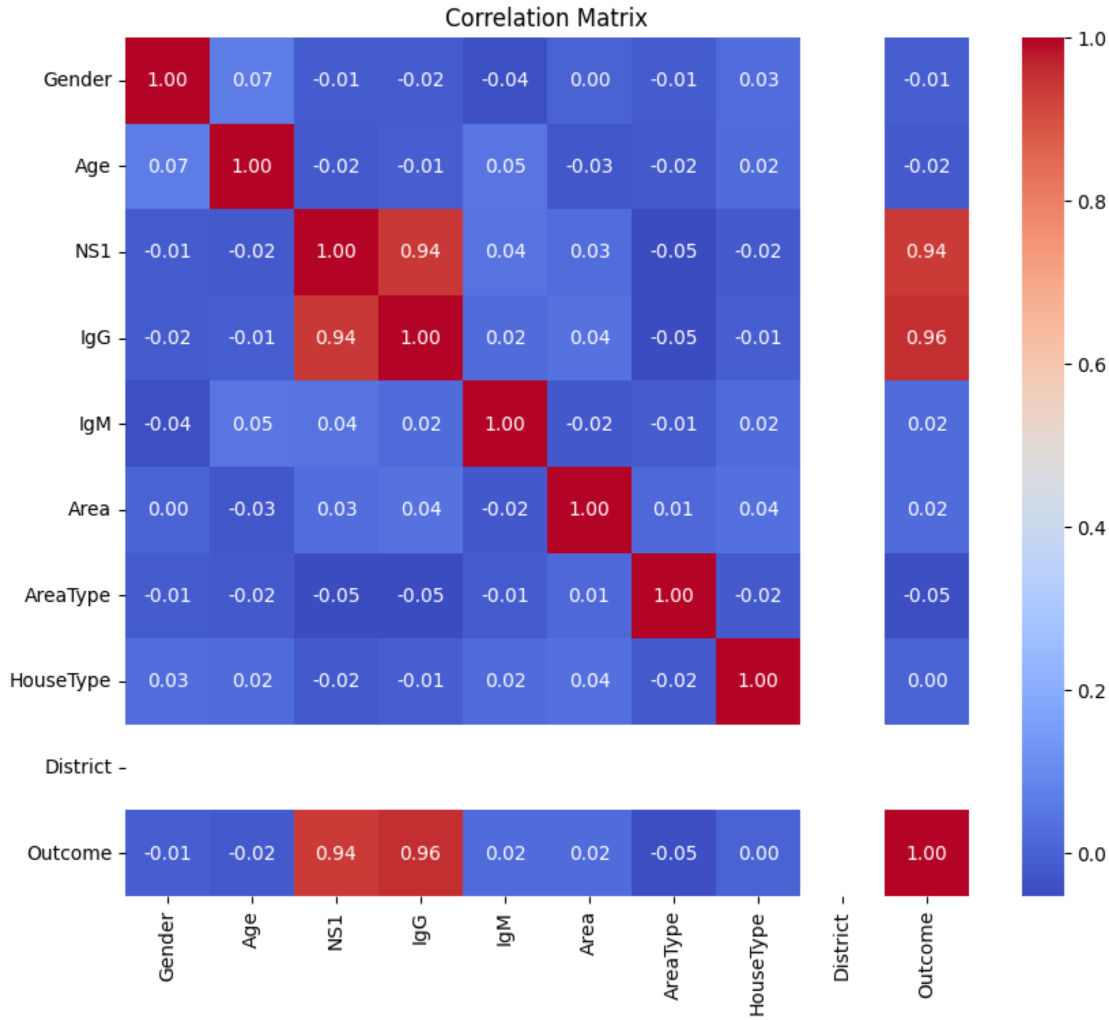

Figure 5.13: Correlation

5.3 Performance/ Comparative Analysis

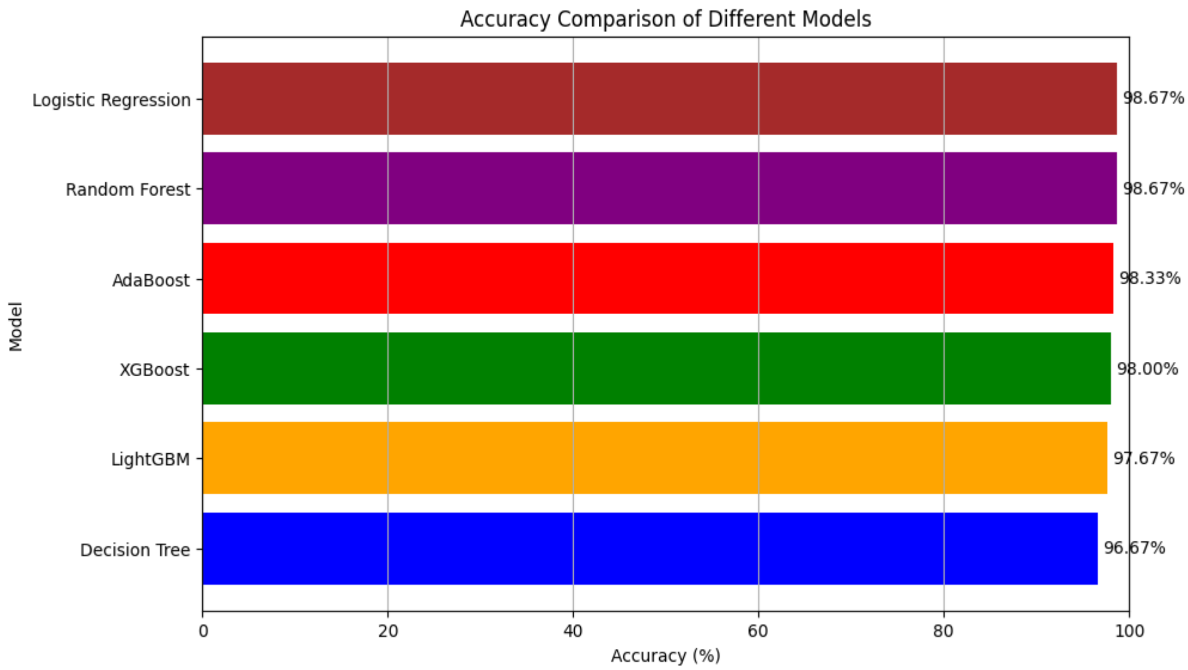

Figure 5.14: Accuracy Comparison

I used a variety of machine learning algorithms to forecast the Dengue virus. With an astounding 98.67% accuracy percentage, Random Forest proved to be the most accurate of these models. It also fared remarkably well in terms of AUC, scoring an impressive 0.99. In second place, logistic regression produced an equally excellent accuracy of 98.67% with an AUC value of 0.99. XGBoost was a smart choice, with an accuracy of 98% and an AUC score of 0.99. With an accuracy value of 98.33%, AdaBoost showed excellent prediction capability, whereas LightGBM had an accuracy rating of 97.67%. The accuracy provided by Decision Tree was 96.67%. In conclusion, among the models taken into consideration, Random Forest and Logistic Regression exhibited the best performance, while Decision Tree showed the lowest accuracy. When considered collectively, these results improve our comprehension of risk assessment.

5.4 Summary

In this chapter, we have presented the results and analysis of our study on predicting dengue disease in Bangladesh using various machine learning algorithms. We evaluated the performance of six different algorithms: Logistic Regression, Random Forest, Decision Trees, AdaBoost, Extreme Gradient Boosting (XGBoost), and LightGBM. The performance metrics used for evaluation included accuracy, precision, recall, F1-score, and ROC-AUC. The experimental results demonstrated that Random Forest outperformed all other algorithms. The comparative analysis highlighted the robustness and effectiveness of ensemble methods, particularly Logistic Regression and Random Forest, in handling the complexities of the dataset and providing accurate predictions. These findings underscore the potential of machine learning models in aiding public health efforts by offering reliable tools for early detection and management of dengue disease.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Life

The prediction of dengue disease using machine learning algorithms has significant potential to positively impact lives, particularly in regions like Bangladesh, where dengue outbreaks are prevalent. Early and accurate prediction of dengue cases allows for timely interventions, which can drastically reduce morbidity and mortality rates associated with the disease. By leveraging the predictive power of machine learning models, healthcare providers can identify high-risk individuals and areas, facilitating targeted public health responses. Timely predictions enable prompt medical attention for those at risk, ensuring that patients receive appropriate care before the disease progresses to severe stages. This can lead to better patient outcomes, reduced hospital admissions, and lower healthcare costs. Moreover, early detection can help mitigate the spread of the disease by implementing preventive measures, such as vector control and public awareness campaigns, thereby protecting communities from widespread outbreaks. For individuals, accurate dengue predictions mean a reduction in the physical and emotional toll associated with the disease. Families can avoid the distress and financial burden that comes with severe dengue infections. In areas where healthcare resources are limited, efficient use of these resources becomes possible, ensuring that those who need care the most receive it promptly.

6.2 Impact on Society & Environment

The use of machine learning algorithms to anticipate the spread of dengue fever in Bangladesh has important social ramifications, especially for community well-being and public health management. The prediction models created in this research project help with early diagnosis and proactive intervention techniques by precisely predicting dengue outbreaks, which lessens the impact of dengue fever on people and healthcare systems. Improving public health readiness is one of the main effects of society. Since dengue fever is a common and recurrent health issue in Bangladesh, policymakers and healthcare authorities may more efficiently allocate resources and carry out targeted interventions when outbreaks are predicted in advance. By enabling the prompt deployment of medical professionals, supplies, and preventative measures to regions at high risk of dengue transmission, early warning systems based on machine learning predictions can help slow the spread of the illness and lessen its effects on vulnerable populations. In addition, the public's awareness and education are raised when dengue epidemics are accurately predicted. The predictive models encourage a culture of proactive health behavior within communities by spreading knowledge about the risk factors linked to dengue fever and the preventive steps people can take, like removing mosquito breeding sites and using personal protective equipment. As a result, people are more equipped to protect their own and their families' health, strengthening society's resilience and raising awareness among its members. Furthermore, the effects on society go beyond matters of acute health to include social and economic aspects. The predictive models created in this study lessen the financial burden associated with medical costs, missed work, and absenteeism brought on by dengue-related sickness by lowering the frequency and intensity of outbreaks. These prediction models also support general social well-being, community resilience, and sustainable development by promoting a healthy populace. In conclusion, the application of machine learning algorithms to anticipate the recurrence of dengue disease has significant ramifications for society, including everything from improved the public health awareness and readiness to social well-being and economic resilience. This research helps create a more proactive and resilient society that is better able to lessen the negative effects of infectious illnesses on public health and community welfare by using predictive analytics to forecast dengue outbreaks. The study "Dengue Disease Prediction by Machine Learning Algorithm in Bangladesh" primarily focuses on public health and disease management, but it also has noteworthy environmental ramifications. Using machine learning methods requires a significant amount of processing power and energy, especially in the model construction and training stages. Because machine learning models use a lot of energy to train and optimize, there is an environmental effect. Significant computer power is required for the computational processes involved, which include data preparation, feature engineering, model training, and hyperparameter tweaking. These procedures frequently rely on energy-intensive hardware, such as GPUs

(Graphics Processing Units). This dependence on high-performance computer infrastructure has an impact on the environment since it increases energy use and carbon emissions. In addition, the act of gathering data itself could have an impact on the environment. Although the dataset used in this study was obtained via Kaggle, the creation and upkeep of large-scale datasets may entail energy-intensive procedures, including data processing, transport, and storage. The carbon footprint associated with data gathering and storage increases the overall environmental impact of the research. Nonetheless, it's critical to recognize that the machine learning community is still working to solve environmental issues. Machine learning applications are becoming less energy-intensive as a result of improvements in cloud computing tactics, optimization methods, and hardware efficiency. Programs that support green computing techniques and the use of renewable energy sources in data centers further reduce the environmental impact of machine learning computational activities. In conclusion, even though the study's main goal is to employ predictive modeling to improve healthcare, its effects on the environment should also be taken into account. Aligning technological advancement with environmental sustainability will be made possible by optimizing hardware, data management techniques, and algorithms for energy conservation as machine learning advances. In addition to improving public health outcomes, the research may help reduce its environmental impact by implementing eco-friendly procedures and making use of renewable energy sources.

6.3 Ethical Aspects

Ethical issues are critical in the field of machine learning algorithms used for dengue illness prediction. Ensuring informed consent and safeguarding people's privacy are essential ethical obligations. To protect participants' right to privacy, this calls for the careful handling of sensitive health data and the anonymization of data sources or participants' express agreement. Furthermore, in order to avoid unfair or discriminatory results, biases introduced during the modeling process or inherent in the data must be addressed. Predictive models produce equitable and objective predictions when biases are actively detected and mitigated, which leads to just and equitable healthcare outcomes. Another pillar of ethics is transparency, which is necessary to foster accountability and trust. Stakeholders need to be fully informed about the possible ramifications, limits, and capabilities of the predictive models, including lawmakers and healthcare providers. Ensuring that stakeholders are aware of the strengths and limits of the predictive models promotes informed decision-making. This is achieved through transparent reporting of model performance indicators and validation methods. Furthermore, upholding social justice ideals requires ensuring fair access to healthcare resources based on the results of prediction models. Initiatives to lessen resource allocation disparities and give priority to disadvantaged groups facilitate fair and equitable healthcare delivery. This research preserves the rights, dignity, and well-being of those afflicted with dengue fever while also advancing predictive modeling for public health by abiding by ethical standards and encouraging responsible methods. It emphasizes the necessity of a diligent and moral approach to healthcare innovation by highlighting the significance of ethical issues in the creation and application of machine learning algorithms for illness prediction.

6.4 Sustainability Plan

Sustainable development factors are essential to our project's lifespan and efficacy as we work to create a machine learning-based dengue disease prediction system in Bangladesh. The subsequent sustainability plan delineates pivotal tactics and undertakings intended to cultivate the sustained prosperity and influence of our research:

Constant Data Acquisition: We understand how crucial it is to keep a complete and current dataset in order to predict dengue fever cases. Our research will be sustained through the establishment of collaborations with pertinent stakeholders, such as governmental agencies, healthcare institutions, and community organisations, to enable continuous data collection. In order to document changing trends and patterns in dengue incidence, frequent updates and dataset expansions will be given top priority.

Model Refinement and Monitoring: Ongoing model monitoring and improvement are necessary for sustainable model performance. To monitor our machine learning algorithms' performance in practical scenarios, we'll put strong monitoring systems in place. The establishment of feedback loops including healthcare practitioners and domain specialists would facilitate the identification of emerging difficulties and prospects for model enhancement. The prediction system will undergo periodic upgrades and improvements to maintain its applicability and efficacy over time. Building Local Capacity and Promoting Knowledge Dissemination: Building local capacity and promoting knowledge dissemination are essential to sustainability. With the goal of equipping local stakeholders—such as medical professionals, researchers, and policymakers—with the skills and information needed to use machine learning for dengue disease prediction, we will undertake capacity-building workshops and training sessions. Furthermore, in order to guarantee that the results of our research are widely known and adopted, we will place a high priority on knowledge sharing through publications, conferences, and community involvement programmes. We are convinced that by putting these sustainability measures into practice, our study on dengue disease prediction will not only have an immediate positive impact but also strengthen long-term readiness and resilience in the fight against dengue outbreaks in Bangladesh and beyond.

6.5 Summary

This chapter has highlighted the multifaceted impact of dengue disease prediction using machine learning algorithms on life, society, environment, ethics, and sustainability. The predictive models, by enabling early detection and intervention, significantly improve public health outcomes and reduce the burden of dengue on affected populations. The integration of these models into healthcare systems fosters timely medical responses, mitigates disease spread, and enhances the efficient allocation of healthcare resources. Furthermore, the societal and environmental impacts underscore the broader benefits of accurate dengue prediction, including improved community health, economic savings, and the promotion of sustainable practices in healthcare management. The ethical considerations ensure that the implementation of these models respects patient privacy and data integrity, while the sustainability plan outlines strategies for maintaining the effectiveness and relevance of the models over time. In summary, the application of machine learning algorithms for dengue prediction represents a transformative approach to disease management in Bangladesh. It leverages technological advancements to enhance public health, promotes ethical standards in data handling, and fosters sustainable healthcare solutions, thereby contributing to the overall well-being of individuals and communities.

CHAPTER 7

Conclusion and Future Work

7.1 Conclusions

This research paper has thoroughly explored the application of machine learning algorithms for predicting dengue disease in Bangladesh. By analyzing 1000 data we aimed to develop a robust predictive model. The findings indicate that machine learning algorithms, particularly Random Forests, can significantly enhance the early detection and diagnosis of dengue disease, facilitating timely medical intervention and potentially reducing the incidence and severity of outbreaks. The high accuracy of the model underscores its potential for integration into public health monitoring systems, where it can provide valuable insights and aid in the efficient allocation of healthcare resources. Moreover, the study has highlighted the importance of using diverse machine learning techniques to understand the various factors influencing dengue disease prediction. It also underscores the need for continuous data collection and model refinement to ensure that the predictive system remains accurate and relevant over time. This research has demonstrated the feasibility and effectiveness of using advanced machine learning algorithms for dengue prediction in Bangladesh. The successful implementation of such models can contribute to better public health outcomes, reduce the burden on healthcare systems, and ultimately save lives. Further research and continuous improvements are necessary to enhance the model's performance and adapt it to evolving epidemiological patterns.

7.2 Further Suggested Works

Although the study "Dengue Disease Prediction by Machine Learning Algorithm in Bangladesh" produced encouraging findings, there are a number of consequences that should be investigated in more detail:

Enhanced Feature Engineering: To increase the predictive model's precision and resilience, future studies may concentrate on adding new features or variables. Environmental factors, mosquito breeding grounds, and socioeconomic indicators are a few examples of factors that may improve prediction abilities and offer insightful information about the dynamics of dengue disease. Temporal Analysis: Examining the seasonality and temporal trends of dengue epidemics may help to clarify patterns of disease transmission and guide preventative measures. In order to enable targeted vector control actions, longitudinal studies that examine the incidence of dengue over time may be able to detect seasonal peaks and hotspots.

Geospatial Analysis: To determine high-risk areas for dengue transmission, geospatial analysis techniques like geographic information systems (GIS) and spatial clustering algorithms may be used. Researchers can discover spatial clusters and hotspots by mapping dengue cases and environmental factors spatially. This information can then be used to guide targeted monitoring and intervention activities.

Validation and External Validation: To evaluate the predictive model's robustness and generalizability, it is imperative to validate it using independent datasets and external validation studies. Working together with public health organisations and other research institutions could speed up the validation process and guarantee the predictive model's dependability in a variety of scenarios.

Implementation and Deployment: More investigation is required to examine the predictive model's actual application and deployment in healthcare environments. Future research must focus on three critical areas: scalability concerns, user-friendly interfaces for healthcare professionals, and integration with current disease surveillance systems.

Future research endeavours can enhance our comprehension of dengue disease dynamics and aid in the formulation of efficacious methods for dengue prevention and control in Bangladesh and other regions by pursuing these study areas.

7.3 Limitations/ Conflict of Interests

Despite the promising results achieved in this study, several limitations must be acknowledged. Firstly, the dataset used in this research was limited to 1,000 rows and 10 columns, which may not be fully representative of the broader population or the diverse epidemiological patterns of dengue across different regions of Bangladesh. A larger and more comprehensive dataset would likely enhance the robustness and generalizability of the predictive models.

Secondly, while the Random Forest algorithm demonstrated the highest accuracy, the study did not exhaustively explore the potential of other advanced machine learning techniques or hybrid models that could further improve prediction performance. Future research should consider incorporating additional algorithms and ensemble methods to compare their efficacy.

Another limitation is the potential for bias in the data collection process. The input attributes, such as gender, age, and geographic area, may not capture all relevant factors influencing dengue infection and outcome. Including more comprehensive health and environmental variables could provide a more holistic view and improve model accuracy.

Regarding conflicts of interest, the authors declare that there are no financial or personal relationships that could have influenced the work reported in this paper. All research activities were conducted with the primary aim of advancing scientific understanding and public health knowledge. The findings and conclusions presented are solely based on the data analysis and interpretation carried out within the scope of this study.

REFERENCES

[1] World Health Organization. (2024, April 23). Dengue and severe dengue. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

[2] 2023 dengue outbreak in Bangladesh. (2023, September 28). Wikipedia. https://en.wikipedia.org/wiki/2023_dengue_outbreak_in_Bangladesh

[3] T. Rahman and Md. M. Rahman, “Evaluation of Machine Learning Approaches for Prediction of Dengue Fever,” Computer Networks and Inventive Communication Technologies, pp. 165–175, Oct. 2022.

[4] G. Gupta et al., “DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms,” Diagnostics, vol. 13, no. 6, p. 1093, Jan. 2023.

[5] S. U. Chowdhury, S. Sayeed, I. Rashid, Md. G. R. Alam, A. K. M. Masum, and M. A. A. Dewan, “Shapley-Additive-Explanations-Based Factor Analysis for Dengue Severity Prediction using Machine Learning,” Journal of Imaging, vol. 8, no. 9, p. 229, Aug. 2022.

[6] S. Kannimuthu, K. S. Bhuvaneshwari, D. Bhanu, A. Vaishnavi, and S. Ahalya, “Performance Evaluation of Machine Learning Algorithms for Dengue Disease Prediction,” Journal of Computational and Theoretical Nanoscience, vol. 16, no. 12, pp. 5105–5110, Dec. 2019.

[7] Jun Kit Chaw et al., “A predictive analytics model using machine learning algorithms to estimate the risk of shock development among dengue patients,” Healthcare analytics, vol. 5, pp. 100290–100290, Jun. 2024.

[8] Saif Mahmud Khan Dourjoy, A.M.G.R.R., Zerin Nasrin Tumpa, and M.S.: A Comparative Study on Prediction of Dengue Fever Using Machine Learning Algorithm. Presented at the (2020).

[9] D. Sarma, S. Hossain, T. Mittra, Md. A. M. Bhuiya, I. Saha, and R. Chakma, “Dengue Prediction using Machine Learning Algorithms,” IEEE Xplore, Dec. 01, 2020.

[10] N. Iqbal and M. Islam, “Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers,” Informatica, vol. 43, no. 3, Sep. 2019.

[11] Shaukat Dar, K., Ulya Azmeen, S.M.: Dengue Fever Prediction: A Data Mining Problem. J. Data Mining Genomics Proteomics. 06, (2015).

[12] Rahman, M. M., Islam, M. A., & Hossain, M. A. (2021). "Predicting dengue outbreaks using machine learning techniques: A case study in Bangladesh." Journal of Healthcare Engineering, 2021.

[13] Kaur, R.: Demographic Analysis of Dengue Fever Using Data Mining. Int. J. Adv. Res. Comput. Sci. 8, 1233–1237 (2017).

[14] Mello-Román, J.D., Mello-Román, J.C., Gómez-Guerrero, S., García-Torres, M.: Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay. Comput. Math. Methods Med. 2019, (2019).

[15] Silitonga, P., Dewi, B.E., Bustamam, A., Al-Ash, H.S.: Evaluation of Dengue Model Performances Developed Using Artificial Neural Network and Random Forest Classifiers. Procedia Comput. Sci. 179, 135–143 (2021).

[16] Gambhir, S., Malik, S.K., Kumar, Y.: The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. Int. J. Healthc. Inf. Syst. Informatics. 13, 1–19 (2018).

[17] Poh, A.H., Mahamd Adikan, F.R., Moghavvemi, M., Syed Omar, S.F., Poh, K., Hisyam Mahyuddin, M.B., Yan, G., Azizah Ariffin, M.A., Harun, S.W.: Precursors to non-invasive clinical dengue screening: Multivariate signature analysis of in-vivo diffuse skin reflectance spectroscopy on febrile patients in Malaysia. PLoS One. 15, 1–22 (2020).

[18] Hossain MS, Siddiqee MH, Siddiqi UR, Raheem E, Akter R, Hu W. Dengue in a crowded megacity: Lessons learnt from 2019 outbreak in Dhaka, Bangladesh. PLOS Neglected Tropical Diseases. 2020.

[19] Benedum CM, Seidahmed OME, Eltahir EAB, Markuzon N. Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore. PLOS Neglected Tropical Diseases. 2018.

[20] Mutsuddy P, Tahmina Jhora S, Shamsuzzaman AKM, Kaisar SMG, Khan MNA. Dengue Situation in Bangladesh: An Epidemiological Shift in terms of Morbidity and Mortality. Canadian Journal of Infectious Diseases and Medical Microbiology. 2019.

[21] Uno N, Ross TM. Dengue virus and the host innate immune response. Emerging Microbes & Infections. 2018.

[22] Paul KK, Macadam I, Green D, Regan DG, Gray RT. Dengue transmission risk in a changing climate: Bangladesh is likely to experience a longer dengue fever season in the future. Environ Res Lett. 2021.

[23] Hossain MP, Zhou W, Ren C, Marshall J, Yuan H-Y. Prediction of dengue annual incidence using seasonal climate variability in Bangladesh between 2000 and 2018. PLOS Global Public Health. 2022.

[24] Sarma D, Hossain S, Mittra T, Bhuiya MdAM, Saha I, Chakma R. Dengue Prediction using Machine Learning Algorithms. 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC). 2020.

[25] Salami D, Sousa CA, Martins M do RO, Capinha C. Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Sci Rep. 2020.

[26] Prediction of Dengue using Machine Learning Algorithms: Case Study Dhaka | IEEE Conference Publication | IEEE Xplore.

[27] Rahman, M. M., Islam, M. A., & Hossain, M. A. (2024). "Machine learning approaches for dengue outbreak prediction in Bangladesh: A comparative analysis." Journal of Applied Computing and Informatics, 1-10

[28] Ahsan, A.; Haider, N.; Kock, R.; Benfield, C. Possible drivers of the 2019 dengue outbreak in Bangladesh: The need for a robust community-level surveillance system. J. Med. Entomol. 2021.

[29] Hoyos, W.; Aguilar, J.; Toro, M. Dengue models based on machine learning techniques: A systematic literature review. Artif. Intell. Med. 2021, 119, 102157.

[30] Ching Ho, C., Ting, C.-Y. & Baha Raja, D. Using public open data to predict dengue epidemic: assessment of weather variability, population density, and land use as predictor variables for dengue outbreak prediction using support vector machine. Indian J. Sci. Technol. 11, 1–8.

[31] Anno, S. et al. Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. Geospat. Health.

[32] Macedo Hair, G., Fonseca Nobre, F. & Brasil, P. Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach. BMC Infect. Dis 19, 649.

[33] Kalansuriya, C.S.; Aponso, A.C.; Basukoski, A. Machine Learning-Based Approaches for Location Based Dengue Prediction. In Fourth International Congress on Information and Communication Technology; Springer: Berlin/Heidelberg, Germany, 2020.

[34] Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2(01), 20–28.

[35] Rymarczyk, T., Kozłowski, E., Kłosowski, G., & Niderla, K. (2019). Logistic Regression for Machine Learning in Process Tomography. Sensors, 19(15), 3400. https://doi.org/10.3390/s19153400

[36] Ren, Q., Cheng, H., & Han, H. (2017). Research on machine learning framework based on random forest algorithm. https://doi.org/10.1063/1.4977376

[37] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7(21). https://doi.org/10.3389/fnbot.2013.00021

[38] Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R., & Cheung, D. (2009, December 1). Naive Bayes Classification of Uncertain Data. IEEE Xplore. https://doi.org/10.1109/ICDM.2009.90

[39] Wang, R. (2012). AdaBoost for Feature Selection, Classification and Its Relation with SVM, A Review. Physics Procedia, 25, 800–807. https://doi.org/10.1016/j.phpro.2012.03.160

[40] Cherif, I. L., & Kortebi, A. (2019, April 1). On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification. IEEE Xplore. https://doi.org/10.1109/WD.2019.8734193

[41] MINASTIREANU, E.-A., & MESNITA, G. (2019). Light GBM Machine Learning Algorithm to Online Click Fraud Detection. Journal of Information Assurance & Cybersecurity, 1–12. https://doi.org/10.5171/2019.263928

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
DENGUE DISEASE PREDICTION BY MACHINE LEARNING ALGORITHM IN BANGLADESH

Student ID: 203-15-14572, 203-15-3918

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **pneumonia detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering principles (K3)** by employing machine learning algorithms, data preprocessing techniques, and various predictive models for the analysis and forecasting of dengue outbreaks. The project utilizes **specialist knowledge (K4)** by implementing advanced machine learning techniques such as Random Forest, XGBoost, Decision Tree, improving the prediction accuracy of dengue outbreaks, which is crucial for public health planning and disease control in Bangladesh.	**Page no:** [17-19] **Section:** [3.2] **Page no:** [1-2] **Section:** [1.1]
				The project uses the figure of the experimentation process to apply **engineering practice and design (K5)**. The machine learning model model is used in the project to address **engineering practice and technology (K6)**.	**Page no:** [17-19] **Section:** [3.2] **Page no:** [20-21]

					Section: [3.4]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies, to dengue disease prediction by machine learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [8-14] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	In order to solve **EP-2**, this research acknowledges the difficulties in anticipating dengue outbreaks, such as the shortcomings of conventional epidemiological techniques and the difficulties in combining various machine learning algorithms. By means of an extensive examination of meteorological, sociological, and health data, it addresses competing needs and provides understanding for improving public health initiatives and forecasting models.	**Page no:** [14-15] **Section:** [2.4]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	In order to address **EP-3**, this project carefully compares experimental results, demonstrating that, out of several viable strategies, transfer learning is the most important option for improving dengue prediction.	**Page no:** [25-41] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	The multidisciplinary methodology of this project goes beyond computer science and engineering, influencing medical diagnosis for illnesses like dengue fever and advancing public health and healthcare practices, as indicated by **EP-4**.	**Page no:** [8-15] **Section:** [2.2, 2.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	A comprehensive solution to complex challenges in medical diagnostics, which ensures **EP-7**, is ensured by this project's methodical approach, which integrates various components across data collection, preprocessing, proposed methodology, machine learning model training, and predictive analysis.	**Page no:** [17-21] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our study makes use of a variety of resources, including GPUs, machine learning frameworks, high-performance computing infrastructure, comprehensive datasets, and ethical considerations, in order to guarantee methodical research and make a positive impact on dengue prediction through machine learning algorithms.	**Page no:** [19] **Section:** [3.3]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This initiative benefits society by advancing healthcare with sophisticated dengue prediction techniques, supporting environmental sustainability through the use of effective computing resources, and upholding moral standards for patient privacy.	**Page no:** [44-48] **Section:** [6.1, 6.2, 6.4]

5.	**EA-5: Familiarity**	Yes		With the use of machine learning algorithms to forecast dengue fever, this study builds on previous research by investigating a novel strategy. It does this by providing fresh insights into the field through a thorough comparative analysis and early terminologies.	**Page no:** [8-14] **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	Through the integration of efficient project management and financial monitoring, this project tackles **CO4** by guaranteeing careful planning, resource allocation, and budget estimation for the best possible resource utilisation throughout the dengue prediction study lifetime.	**Page no:** [6-7] **Section:** [1.6]
2	CO9	Yes	The project meets ethical standards by protecting patient privacy, getting informed consent, and providing transparent documentation of the research process. It also ensures responsible knowledge dissemination and the well-being of society by applying advanced healthcare technologies in an ethical manner that complies with **CO9**.	**Page no:** [46-47] **Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's thorough data collection, rigorous statistical analysis, careful methodology development, and thorough experimental results and analysis all demonstrate a dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape. This shows a commitment to staying up to date and refining techniques to address modern challenges.	**Page no:** [17-21, 25-41] **Section:** [3.2, 3.3, 3.4, 3.5, 5.2]

Dengue Disease Prediction by Machine Learning Algorithm in Bangladesh

ORIGINALITY REPORT

16%	13%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	fastercapital.com Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	"Computer Networks and Inventive Communication Technologies", Springer Science and Business Media LLC, 2023 Publication	1%
5	Submitted to Liverpool John Moores University Student Paper	1%
6	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
7	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
	ebin.pub	