ANTICIPATING DENGUE: ADVANCED MACHINE LEARNING FOR TIMELY EPIDEMIC PREDICTION

BY

Jannatul Ferdous
ID: 202-15-14413
AND

Md. Arafat Rahman
ID: 201-15-14020

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

Sharmin Akter
Lecturer (Senior Scale)
Department of Computer Science & Engineering
Daffodil International University

Co Supervised By

Ms. Most. Hasna Hena
Assistant Professor
Department of Computer Science & Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled "ANTICIPATING DENGUE: ADVANCED MACHINE LEARNING FOR TIMELY EPIDEMIC PREDICTION", submitted by Jannatul Ferdous and Md. Arafat Rahman 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 July 01, 2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH) — **Board Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.07.24

Ms. Nazmun Nessa Moon (NNM) — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.7.24

Mr. Deawan Rakin Ahamed Remal (DRAR) — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.07.2024

Dr. Abu Sayed Md. Mostafizur Rahaman (ASMR) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Sharmin Akter, Lecturer (Senior Scale), Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

11.07.24

Sharmin Akter
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co Supervised by:

11/7/24

Ms. Most. Hasna Hena
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Jannat

Jannatul Ferdous
ID: 202-15-14413
Department of CSE
Daffodil International University

Arafat

Md. Arafat Rahman
ID: 201-15-14020
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Sharmin Akter, Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Data Science 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 **Professor Dr. Sheak Rashed Haider Noori, Professor & 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

This research is about finding better ways to predict and manage dengue outbreaks, a serious illness spread by mosquitoes. To enhance predictions, we are utilizing sophisticated machine learning methods and intelligent computer programs. Put simply, our focus is on understanding and forecasting potential dengue outbreak sites and times. The study highlights the significance of trustworthy data, moral data management, and machine learning proficiency. Using a dataset that includes important variables like year, month, area, affected population, and total population, our goal is to create a tool that can quickly assess risk for people traveling to new places. Based on input parameters like area and month, we hope to use advanced algorithms to classify the likelihood of dengue occurrence into four categories: Low, Medium, High, and Risky. Input output analysis prioritizes privacy protection while ensuring data organization and utility. We prioritize privacy while making sure the data is well organized and helpful for our computer programs. Our project management guarantees an organized strategy with efficient teamwork, financial analysis, budgeting, and flexibility in the face of difficulties. In summary, the goal of this research is to provide important new information for the comprehension and forecasting of dengue outbreaks. The methodical approach, which is based on state-of-the-art machine learning techniques and data-centric approaches, aims to close current gaps and offer a strong framework for managing and predicting dengue outbreaks. After a thorough analysis of the data, the Decision Tree algorithm was the best performer, with an astounding 84% accuracy, 60% F1 score, 67% precision, and 55% recall rate. Considering these results, the Decision Tree is the recommended algorithm for dengue epidemic prediction. Consequently, decision trees are the best option for this purpose since we can accurately and consistently predict dengue outbreaks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
DGHS	Directorate General of Health Services
SVM	Support Vector Machine
KNN	K-Nearest Neighbors

CHAPTER 1
INTRODUCTION

1.1 Overview

Dengue fever is becoming more commonplace in many parts of the world, it has become a serious global health concern. The public's health is seriously threatened by this virus spread by mosquitoes, especially in tropical and subtropical areas where the Aedes mosquito vector is abundant. Dengue outbreaks frequently lead to high rates of morbidity and mortality, which severely tax healthcare systems and negatively impact the local economy in the affected areas. Urbanization, population growth, climate change, and globalization are some of the reasons contributing to the rapid spread of dengue disease. These elements provide an ideal environment for mosquito reproduction and dengue virus dissemination, which results in frequent and extensive outbreaks.

Researchers and medical practitioners can reduce the effect of epidemics by identifying high risk locations, implementing targeted treatments, and gaining useful insights into the spatiotemporal patterns of dengue transmission by utilizing big data and predictive analytics. We hope to create predictive models that can precisely and timely predict dengue outbreaks by utilizing cutting edge machine learning techniques and large-scale dataset analysis. The public health authorities will be able to adopt early and efficient control measures with the use of these models, which will also improve our understanding of the dynamics of dengue transmission. Through an examination of past dengue case data, and population dynamics, the suggested model seeks to offer current insights on the probability and intensity of outbreaks in particular regions. The ultimate purpose of this research is to provide communities, governments, and healthcare professionals with the information and resources we need to lessen the effects of dengue fever and safeguard public health. This initiative employs a multidisciplinary strategy that integrates epidemiology, data science, and public health to support the worldwide effort to combat dengue and other diseases spread by vectors.

1.2 Problem Statement

Dengue fever is still a major public health problem despite advances in technology and treatment, particularly in areas like Bangladesh where the illness burden is high. The current approaches to forecast dengue epidemics may not be very accurate, which makes it difficult to allocate resources and intervene in time. Although dengue is common in Bangladesh, there is a dearth of studies concentrating on precise prediction models adapted to the nation's unique circumstances. To improve the precision and dependability of forecasts, the project will make use of sophisticated machine learning methods to create a solid dengue outbreak prediction model. the project aims to address the pressing need for timely and accurate dengue outbreak prediction in Bangladesh by leveraging advanced machine learning techniques to improve public health preparedness and response. Dengue outbreak prediction techniques now in use frequently rely on historical data and conventional epidemiological models. These methods might not be able to fully capture the intricate dynamics of dengue transmission, which would result in poor prediction accuracy and insufficient readiness. The impact of dengue outbreaks must be reduced, and this requires prompt action and early detection. Increasing prediction accuracy can support preemptive actions including public awareness campaigns, healthcare resource allocation, and vector control activities, which can result in more efficient outbreak management. By increasing prediction accuracy, the project seeks to support healthcare authorities in effective resource allocation, early intervention, and effective public health planning to mitigate the impact of dengue outbreaks.

1.3 Motivation and Objectives

This project is our first challenge because dengue is one of the serious diseases which afflicts a lot of people either in a particular region or worldwide. Public health continues to suffer greatly from dengue fever, especially in areas with inadequate infrastructure and resources for disease control. Implementing efficient control measures, such as vector control, community education, and healthcare supply, requires early epidemic identification. Unfortunately, current dengue surveillance techniques frequently fail to provide precise and timely forecasts, which causes interventions to be delayed and the disease burden to rise. The necessity of creating fresh methods for epidemic prediction is highlighted by the appearance of new dengue strains and the shifting dynamics of disease transmission. The goal of this project is to improve the ability for proactive dengue management by addressing the shortcomings of conventional monitoring strategies and utilizing a variety of data sources in conjunction with cutting edge machine learning techniques. Create a cutting-edge machine learning model that can accurately and timely forecast dengue epidemics. To improve forecast granularity and reliability, incorporate a variety of data sources into the predictive model, such as past dengue cases, population demographics, environmental variables, and migration patterns. Assess the developed model's performance with real world data sets and strict validation methods. Give stakeholders healthcare officials, legislators, and community organizations actionable insights and suggestions to help direct preventative dengue control activities. our determination is to solve a notable health matter. The target is to make an instrument that can assist in this endeavor. Moreover, it is expected that this activity will help to improve public health. To continue further we now will consider our vision how we can realize these goals in particular.

1.4 Project Scope

The prime purpose is to take advantage of ML's technology to forecast dengue local outbreaks through model building. The goal is to build a system that will allow for early warnings on disease outbreaks, gathering of historical information on population demographics, movement patterns, and dengue cases. Preprocessing will be done on the data to guarantee its quality and suitability for analysis. The device is meant to be reliable with the sole purpose of contributing to public health. Not only we develop the product, but we also test it to ensure that it works well. This means testing the product to establish whether it is indeed accurate in identifying dengue disease outbreaks. If it is successful, it will be a true revolution in the way we combat epidemics. On one hand, we are possibly moving, but on the other side, we do not pay attention to other issues. We will not study the treatment of patients with dengue like doing research on it or designing a vaccine. These are the undoubtedly the main points. Hence, the protagonist of the novel is given the role of the prophet. Despite the down to earth dengue area of health, we concentrate on forecasting and preparation for the dengue outbreak. We will be narrowing down to the forecasting of the outbreaks and, we hope that our efforts will be meaningful. The fact that we will know the time and area of the dengue virus spread will help us implement measures before the epidemic breaks out. We intend to carry out ML by using it to predict dengue outbreaks and testing for the accuracy.

1.5 Project Outcome

The main target here is to make a predicting device with machine learning algorithms that are efficient and precise in forecasting dengue outbreaks using the cutting-edge tech. Supported by artificial intelligence, it functions more or less the same as a specialist assigned to notify places and times of a dengue epidemic. The tool is believed to be efficient and thus can be revolutionary in detection methods of dengue. Our main goal under interpretation is to grant healthcare officials and decision makers with sufficient authority. Having reliable dengue forecasts about certain places will let them take the smart decisions aimed at preventing and controlling the spread of the disease. Therefore, it constitutes proper budgeting and planning, utilizing preventive mechanisms, and ultimately reducing the population vulnerability to dengue. It has nothing to do with technology. It is the reflection of the health cause. When the timing of the dengue breakdown is predicted, it becomes even more possible to combat this epidemic. Therefore, we are likely to see a positive effect on the Dengue area leading to the reduction of the number of patients infected with it to lower levels. Specifically, our desires lie in the establishment of a well-functioning system that can help predict outbreaks of dengue in a more advanced manner. In conclusion, the most important curative function that can result in enhanced health condition is the final benefit of the public health. Also, the ways of implementation of the solution for this disease will be mentioned, and the significance of it will be brought out for the people who suffer from the dengue outbreaks.

1.6 Report Organization

The structure of the proposed report is logical and offers clear guidance to readers in terms of the progression of the research process, data analysis, and implications. Every chapter is important as it contributes to making the document as coherent as possible and making it contain as much information as possible.

Chapter 1: Introduction

The necessity of dengue prediction is discussed in the beginning of the work to explain why this research is necessary, and how machine learning has been used in this area, but no extensive information is given on these topics although, the authors did a good job in presenting the overview. It presents a brief overview of the objectives and the conceptual work for the study.

Chapter 2: Literature Review

A comprehensive analysis of the knowledge base and research background of the project will be conducted. This analysis will highlight and discuss the gaps present in existing works, which our study aims to address.

Chapter 3: Methodology

The chapter on research methodology describes the method that has been used to carry out the study on dengue prediction using the machine learning approach. It gives an understanding of the approaches used to design the study, gather data, build the model, and the reasons for selecting certain methodologies. The chapter also covers ethical concerns and any constraints that may affect the research.

Chapter 4: Implementation

This chapter discusses the training and testing of machine learning techniques to forecast dengue fever incidences. It involves the establishment of the models, preparations of the models and the assessment of the models.

Chapter 5: Experimental Result and Discussion

This chapter gives the result from the experiment conducted using machine learning in dengue fever prediction. Discussions on how various types of prediction models have performed are presented in the paper, accompanied by graphical illustrations and summarization of the results.

Chapter 6: Impact on Society, Environment and Sustainability

The chapter of the dissertation focuses on the societal implications of the study results in healthcare and disease prevention. This section explains how enhanced dengue prediction approaches can have a beneficial impact on patient management, treatment outcomes, and public health. The possibility of an advancement in technology and its effect in the society is also taken into account.

Chapter 7: Summary, Conclusion, Recommendation, and Implications for Future Research

This is the final chapter: it provides a summary of the contents of the entire report. It restates the findings in form of a concise conclusion, underscores some research findings and recommendations on its feasibility for practical use and for the propagation of further research. Future research directions for the study are highlighted, which will give other researchers an outline on how to advance the present work. This structured format promotes a systematic presentation of information from the introduction, through the specifics of the research study, to the broader implications and recommendations section. It made it easier to fully appreciate the research and the implications that it could ever have on the academic and applied sides of dengue prediction using machine learning.

1.7 Summary

In conclusion, our work seeks to address the urgent problem of dengue outbreak prediction by the application of cutting-edge Machine Learning (ML) techniques. Our goals are simple and motivated by the urgency of the health issue: we want to create and evaluate a machine learning tool that can provide healthcare practitioners accurate forecasts and useful insights. Our study is entirely focused on using machine learning (ML) to anticipate outbreaks, with the ultimate goal of developing a powerful tool that has a major impact on public health outcomes. We'll go into great depth about our research journey in the next chapters as we go along, going over the literature review, methods, experimental results, and wider implications of our work.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

In recent years, more and more people around the world are getting sick with dengue fever. Every year, there are over 390 million cases, and a lot of them have serious symptoms. The main reasons for this increase are the widespread presence of mosquitoes that spread the disease, fast and uncontrolled growth of cities, more people traveling internationally, and not enough effective ways to stop the sickness. The World Health Organization says that places like the Americas, Southeast Asia (especially Thailand, Indonesia, and the Philippines), and the Western Pacific are most affected by dengue. Now, even in Bangladesh, including its capital city Dhaka and different places, there are more cases of dengue. Dengue fever has become a big health problem in Bangladesh. To understand it better, a study was done from various related paper on this topic. The goal was to figure out when dengue happens the most and how it's connected to the weather, places and many more attributes. Researchers used fancy techniques like SVM, DF, RF, Naive Bayes, and many more ways to the implement problem more deeply. The hope is that this research will help us better understand how dengue spreads, especially in Bangladesh.

2.2 Related Works

The author [1] used machine literacy to read Dengue; this is helps ease the issues traditional styles have when it comes to projecting complaint circumstances. Their study focuses on the significance of applying state-of-the-art styles for diving Dengue by adding rigor and real-time prognostications in complaint circumstances, which is veritably important in executing preventative measures in the affected communities. This paper [2] is organized in a manner that reveals an evolutionary model for the prevalence of Dengue through the use of machine literacy ways. The study focuses on the use of sophisticated algorithms The description of the development of contemporary styles provides information on the progress of line identification for enhanced Dengue vaticination. According to the specified environment of this paper, the author [3] calculates the dengue outbreaks in Bangladesh especially, the outbreak of Dhaka megacity. They're evolving different Californian algorithms in order to achieve a high position of delicacy. From the tuned SVM algorithm, they achieved the loftiest delicacy of 97 percent. This paper focuses on [4] a review of different machine literacy styles for further enhancement of the discovery of the dengue complaint with special attention paid to the significance of the early prognostic to reduce the burden of these transmitted conditions throughout the world. therefore, the author of the study establishes that the Random Forest Algorithm shows bettered effectiveness compared to other mechanisms relating to dengue bracket while italicizing that timely discovery of dengue is possible. In this study, thus, five sections [5] and their climate variables from 2013 to 2017 are assessed grounded on the Dengue prevalence rate. The work also determines the Support Vector Machine with a direct kernel as the most applicable bone while stressing that the week of the time be viewed as an influential variable. The study adds to the literature on Dengue vaticination in the environment of exercising machine literacy ways for public health operation operations. The exploration correlates [6] with previous studies concerning the influence of climatic factors in the circumstance of contagious conditions especially Dengue. This study joins this order by exploring precise climatic characteristics and enforcing the machine learning algorithms to boost the Dengue outbreak vaticination. The author [7] participates in epidemiological soothsaying using the Multilayer Perceptron (MLP) algorithm. Specifically, the three stages used in the work include data formatting and pre-processing, selection of SMO features, and final point reduction from MLP, which increases delicacy in the prognostic of Dengue conditions. The SMO algorithm is reported to be a crucial determinant and helpful in producing individual measures to inform on the actuality of the Dengue complaint. Regarding the global epidemic of dengue, the paper author [8], has concentrated on forestallment to check the prevailing situation by changing out its symptoms at an early stage. Comparing different data mining styles in Weka toolbox the study is little more inclined towards the identification and webbing of symptoms and conditions. The findings presented then reveal Random Forest to have performed better than others and thus good at the early opinion of dengue. This paper focuses on [9] the proposal of a system that would diagnose early dengue fever using machine learning algorithm, with a dataset of 400 records. There is potential of the PDS for effectiveness in resource-constrained health care environment because of the ability to arrive at right diagnosis and for learning. The objective of this research is, [10] among other things, to estimate the Dengue fever cases in Jhelum district utilizing several classification algorithms by employing the WEKA tool. The research uses classification algorithms namely Naive Bayes, REP Tree, Random Tree, J48 and SMO. The analysis of these methods shows that Naive Bayes has the highest accuracy of 92% and has the qualities of both efficiency and better estimation of probability. It further focuses on these classification techniques for the prediction of Dengue fever and suggests that among all the techniques Naive Bayes and J48 have been found most effective. This paper focuses on [11] the use of classification algorithms in the analysis of data to forecast the incidence of Dengue in the Surat district and its surroundings. The study obtains the results in the Weka tool based on different classification algorithms like Naïve Bayes, J48, Random Tree, Logistic Model Tree (LMT), Sequential Minimal Optimization (SMO0, and so on. Analyzing all these techniques the J48 method attains accurate results that are significantly appropriate with an accuracy rate of 96 percent. 58%. The research focuses on the importance of data mining in disease prediction and determines that as per the parameters of accuracy, efficiency, and probability of estimation, J48 is superior. The focus area of discussion in this paper is a severe health threat that is, dengue fever; and why it is vital to forecast it. The authors apply a type of technology commonly known as machine learning that helps in enhancing predictions. Focused on the issues of current models and introduced a new model based on the microarray and RNA Seq data sets. The paper's literature review examines people's feelings,' referred to as 'sentiment analysis,' as well as why artificial neural networks (ANN) and support vector machines (SVM) are significant. As indicated by the results, the classifier of the selected type, the Random Forest, functions effectively. The exploration [14] extends former work on arbitrary timber and provides perceptivity into opting variables in the environment of the bracket. This helps offer guidelines when choosing good styles depending on the characteristics of data available or needed. This paper presents [15] a fashion to identify excrescences in the images of mammary glands. In other words, it begins with perfecting the images and also pressing the areas containing excrescence. From there, it excerpts features for describing these areas and classifies them using a kind of machine literacy called as SVM classifier. Random timbers [16] are intelligent in the sense that they're suitable to make prognostications. A decision tree aimlessly has numerous decision trees and chooses a different set of features to make a conjecture. Indeed, when the data isn't clear, they aren't veritably bad off at prognosticating effects. The paper also goes to state that if we mix up the features still further, we should be suitable to make indeed better suppositions. The paper [17] discusses in detail on how the use of exemplifications can enable computers to suppose wisely as we do. It presents a fashion appertained to as ID3 thus making decision tree from exemplifications. It also describes how we can modify this approach to address noisy or missing data. The paper also discusses some possible unborn developments of this system. The paper also presents a training system [18] aimed at enhancing computer recognition by minimizing the distance that separates different groups of data. It can be used for several purposes similar as relating shapes, letters and its training is veritably presto; lower than an hour for complicated problems like handwriting recognition. The paper focuses on creating a decision tree, [19] when data is resolved between two individuals and their information isn't disclosed. They explain how this might be done using a third-party garçon and bandy how important information might be given away. Further, in the future, they want to corroborate how important of information gets exposed through decision trees and test other ways of concealing data. This paper demonstrates [20] that using color histogram, the support vector machines (SVMs) are suitable to rightly classify the images. In similar ways, they were suitable to work at the high delicacy of image recognition and didn't bear the use of complex data. In this paper [21], the idea of exercising the K Nearest Neighbor classifier to classify normal and protrusive program actions using system call frequency differences was introduced. This is done by treating 'system calls' as words and program prosecutions as 'documents' hence this system registers low false cons in this intrusion discovery as proved by the preliminary tests done on 1998 DARPA BSM inspection data. There's a recommendation of an approach [22] that was developed to alter the support vector machines (SVMs) through center discovery and keeping implicit support vectors. As suggested from the trials, the choice of kernel and the handle of Lagrange multipliers for this option were set up precious. Classifier performance stability was attained by optimizing kernel parameters and Lagrange multipliers and that of radial base functions for the studies conducted displayed hopeful results. The paper [23] concentrated on the use of the decision trees and arbitrary timber algorithms in the determination of the performance of the first time University scholars. Grounded on the scholar's records, data, and checks, the experimenters concluded that several attributes have a major influence on literacy. According to them, out of all the models bandied, the arbitrary timber models have an advanced delicacy compared to decision trees. therefore, this increased delicacy provides information that can be salutary for designing new effective styles for the supporting of the scholars and effective academic recovery. Hence, this paper considers the K- Nearest Neighbor (KNN) bracket [24] since it's easy to understand and apply. Two, KNN is veritably useful for explicatory tasks because of its nice explicability. nonetheless, it's resource ferocious particularly when applied in large data sets and also can be affected by noisy attributes. Despite its plumpness, KNN would conceivably be supplanted by styles that are more sophisticated similar as Support Vector Machines or Neural Networks on more complicated problems. KNN algorithms known as k-Tree [25] and k*Tree are efficiently designed in this paper to describe the phrasings of stoutly opting k values for each of the test samples to enhance the bracket perfection as well as speed. Still, as the authors of similar approaches also agree that over this introductory form of KNN major advancements have been made by training to learn these optimal ‘k’ and by using decision trees for brisk prognostications, these methodologies are way better in terms of delicacy compared to the time spent by the instructional KNN and calculation time. In addition, to enhances the performance and stores farther information into the flake bumps hence dwindling the runtime. The effectiveness of these styles has been caught on by experimental results attained in the course of this exploration of 20 datasets.

2.3 Comparison between existing works

Here, we looked at past studies about dengue fever. It will help us to understand data over time and predict future trends, like how fast dengue fever spreads and how severe it is. We compared my findings with other studies to see how they match up or differ. This helps us understand dengue fever better and find ways to deal with it.

Table 2.3.1: Comparative analysis with previous work

Author Name	Used Algorithm	Best Accuracy with Algorithm
D. Sarma et al. [1]	Decision Tree, Random Forest	Decision Tree = 79%

N. V. Ganapathi Raju et al. [2]	Linear Regression, Support Vector Regressor, Kernel Ridge	Support Vector Machine = 90.63%
M. T. Sarwar et al. [3]	Support Vector Machine, XGBOOST, Logistic Regression	Support Vector Machine = 97%
Latha Jothi et al. [4]	Decision Tree, Random Forest, KNN, Support Vector Machine	Random Forest = 9.13%
Salim, N.A.M. et al. [5]	Support Vector Machine, Naive Bayes, ANN	Support Vector Machine = 70%
Yavari Nejad et al. [6]	Support Vector Machine, Naive Bayes, RBF tree, TempeRain Factor (TRF)	TempeRain Factor (TRF) = 92.35%
Janani, V. et al. [7]	Multilayer Perceptron, Decision Tree	Multilayer Perceptron = 99.874%
Rajathi, N. et al. [8]	Naive Bayes, J48, Random Forest, REP tree, SMO, LWL, AdaboostM1, ZeroR	Random Forest = 83.3%
Khan, J. R. et al. [9]	Decision Tree, Random Forest, Naïve Bayes	Random Forest = 94.3%
Shaukat, K. [10]	Naive Bayes, Rep tree, Random Tree, J48, SMO	Naive Bayes = 92%
Patel, Arvindbhai et al. [11]	Naive Bayes, J48, Random tree, Logistic Model Tree {LMT), SMO	J48 = 96.58%
Gupta G et al. [12]	Random Forest, Decision Tree, Gaussian NB, Support Vector Classifier, KNN	Random Forest = 8.72%

2.4 Gap Analysis

After going through many research papers, I noticed something important. it has become clear that there are a number of important gaps in the knowledge about dengue fever prediction in Bangladesh. Bangladesh has escaped significant epidemics of dengue thus far, whereas neighboring nations are dealing with severe outbreaks. Nonetheless, the concerning expansion of dengue in adjacent areas points to an impending health emergency for Bangladesh. Proactive steps to stop dengue spread are required due to the possible burden on the nation's healthcare system and the possibility of future epidemics. The majority of the examined research papers concentrate on particular dengue fever outbreaks or years, frequently offering case studies or analyses specific to the involved nations. Nonetheless, a noteworthy dearth of thorough studies addressing Bangladesh's dengue problem exists. In spite of the country's rising dengue case incidence in recent years, systematic research employing cutting edge machine learning approaches to forecast and control outbreaks in various Bangladeshi cities is lacking. There is a chance for research intervention because of this vacuum in the literature. This research gap may be filled by utilizing cutting edge machine learning techniques like K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process, Naïve Bayes, Logistic Regression and Support Vector Machine (SVM) to create predictive models tailored to Bangladesh's dengue dynamics. In order to promote public health activities and prevent future epidemics, the objective is to predict and efficiently manage dengue outbreaks in different parts of the nation.

2.5 Summary

In summary, the chapter on literature review offers a thorough analysis of the body of research on dengue fever prediction, with a special emphasis on studies that are pertinent to Bangladesh. The summary emphasizes the growing incidence of dengue fever worldwide and the pressing need for preventative actions to lessen its effects. Despite the rising incidence of dengue in nearby regions, we have found a research deficit in Bangladesh specifically connected to dengue prediction through a thorough examination of related publications. A comparison of previous research indicates differences in predicted accuracy between various machine learning algorithms, highlighting the significance of method selection and model optimization. This analysis provides context for our study and directs how we will approach filling the identified gap in the literature. In the future, our research hopes to create predictive models specific to Bangladesh's dengue dynamics by utilizing cutting edge machine learning techniques. We hope to significantly improve public health outcomes and prevent future epidemics.

CHAPTER 3
METHODOLOGY

3.1 Overview

In this thesis we would like to find out how we can improve dengue outbreak predictions using smart computer algorithms. Dengue is a significant health issue, and instead of using the old methods we are trying the advanced solutions which largely rely on data. This part provides a broad view as to why implementing these technology programs is vital for tracking when the dengue may occur. We also had a look at what others have done, and we identified some areas where new and improved suggestions are required. The plan we put forth seeks to address these gaps by integrating sophisticated methodologies grounded on difficult but invaluable data. In order to improve our research, we are including several other modern techniques, which will create a much more accurate and robust predictive model for the dengue outbreaks. These techniques bridge the deficiencies in the literature by adding to general understanding of dengue modeling. We cover a number of important areas in our technique, such as requirement analysis, data collecting, model development, evaluation measures, and project management. Through a methodical approach to each of these elements, our goal is to guarantee the strength and efficiency of our prediction models. In order to accurately anticipate dengue fever outbreaks, we stress the significance of utilizing cutting edge machine learning methods including K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process, Naïve Bayes, Logistic Regression and Support Vector Machine (SVM). These algorithms are capable of many different tasks and are ideally adapted to tackle the intricacies of the dengue prediction challenge.

We draw attention to the importance of the dataset from the Bangladeshi Directorate General of Health Services (DGHS), which offers insightful information about the temporal and spatial dynamics of dengue transmission in the nation. The data includes daily report on dengue cases, which are publicly accessible at DGHS website. The goal is to generate the recommendations that are so straight forward and comprehensive that armed with them one could manage dengue epidemics with ease.

3.2 Requirement Analysis

The need assessment phase is also an important activity of our project because it outlines the components that are necessary for predicting dengue fever. Develop accurate prediction models for the occurrences of dengue fever in Bangladesh. Predict the severity of epidemics in different regions; people can take preventive measures against dengue fever.

Historical Dengue Data: Gathered from the Directorate General of Health Services (DGHS), which gives information about the temporal and spatial distribution of dengue fever.

Technical Requirements:

High Performance Computing: We have found high-performance computing capabilities as extremely essential. There is need for large data handling and the computations which are sometimes higher level and necessary for the predictions in the predictive modelling. In turn, this will ensure that the data processing time is reduced and the models are trained within the shortest time possible hence improving the prediction accuracy.

Machine Learning Expertise: Knowledge of many different strategies in the field of Machine Learning. Development and enhancement of the hypothetical models which are essential for precise outbreak prediction require comprehensiveness of various fields of Machine Learning.

Stakeholder Collaboration: The company to partner with some with local public health organizations, in order to improve its understanding of local health needs. These organizations can give essential data and make sure the correctness of the data contained in our models. This way the data that is used is well validated and the information gathered is exhaustive thereby improving on the model.

Public Health Organizations: Sharing in the collection of information and developing means of verifying information.

Financial Resources: Equipment, software, or fares for stakeholders' meetings require acquisition of financial funds for operating expenditures. Adequate funding entails that all the project areas are well funded and reduces any major hindrance in the implementation of the predictive system.

3.3 Proposed Methodology

We have developed a systematic technique that combines data preprocessing, feature engineering, model selection, training, evaluation, and deployment to predict dengue fever outbreaks in Bangladesh.

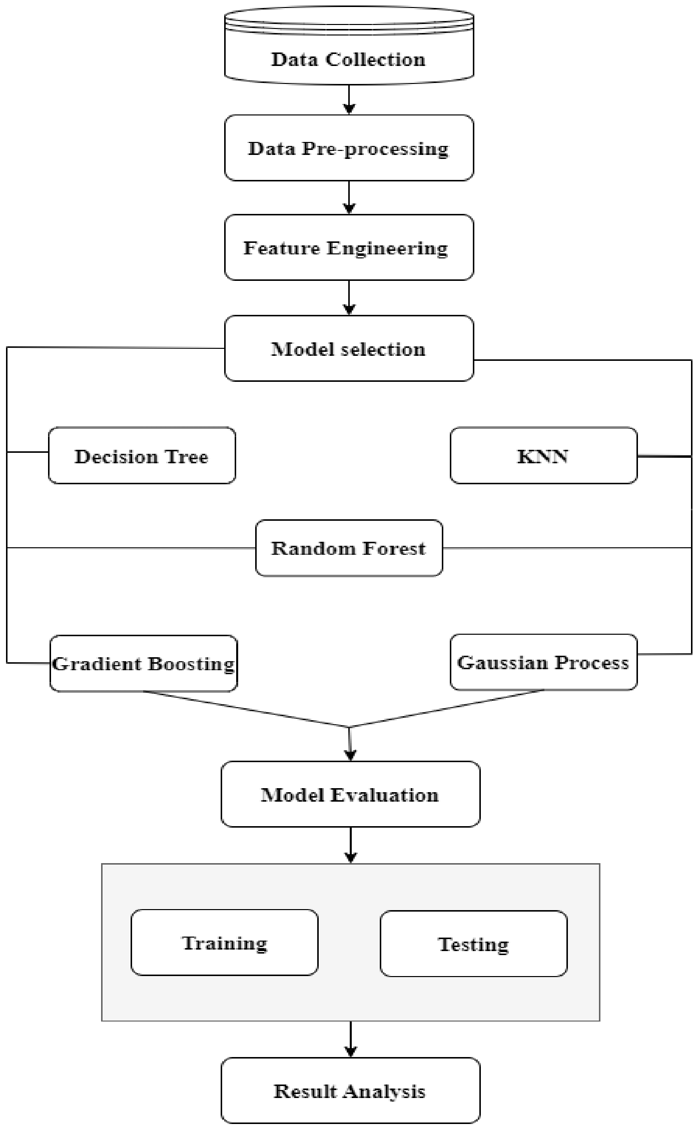

Figure 3.3.1: Proposed Methodology of Dengue Prediction

3.3.1. Preprocessing data:

Cleaning: Deal with outliers, inconsistent data, and missing values in the dataset that was retrieved from the daily dengue bulletin from the DGHS in Bangladesh.

Normalization/Standardization: To improve model performance, make sure that all features scale uniformly. Conver string value data o numeric format data.

Splitting: To make it easier to validate the model, split the dataset into training and testing sets.

3.3.2. Model Choice:

Evaluation of methods: Evaluate the effectiveness of various machine learning methods, such as K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process, Naïve Bayes, Logistic Regression and Support Vector Machine (SVM).

Hyperparameter tuning: To enhance model performance, optimize algorithm parameters through methods such as grid search or random search.

3.3.3. Training and Evaluation:

Division of the dataset into training and testing subsets, initially using an 80:20 split and subsequently a 70:30 split for further validation. Training of models using the training data and evaluation of their performance on the testing data.

3.3.4. Model Assessment:

Metrics: Use metrics like F1 score, accuracy, precision, recall, and area under the receiver operating characteristic curve to assess the performance of the model.

Validation: To guarantee robustness and generalizability to previously untested data, validate models on the testing set.

3.3.5. Model Implementation:

Deployment: To forecast dengue outbreaks in real time, deploy the model or ensemble of models with the greatest performance.

Integration: Provide healthcare professionals and policymakers with simple access to the predictive model by including it into an application or user interface that is straight forward to use.

Constant Monitoring: To keep the model accurate and relevant, keep an eye on its performance over time and adjust it as needed.

3.4 Data Collection

In order to build and analyze models, gathering data is an essential first step in our dengue fever prediction effort. Our dataset is derived from the daily dengue bulletin available from the Directorate General of Health Services (DGHS) Bangladesh [13]. This dataset includes detailed statistics on dengue cases that have been recorded over a number of years in different parts of Bangladesh. The dataset contains variables that indicate the intensity of the outbreak, along with accompanying labels, and includes information on year, month, area, total number of affected individuals, and percentage of affected population. The records are categorized by time intervals (e.g., months) and geographic locations, offering both temporal and spatial insights into the dynamics of dengue transmission. We do pre-processing processes after acquiring the raw dataset to guarantee data quality and analytic applicability. In order to address missing values, outliers, and inconsistencies that could impair the accuracy of our predictive models, this entails cleaning the data. Additionally, we normalize the features to preserve scale consistency and make model training easier. We reference the dataset obtained from DGHS Bangladesh for all analyses and model development stages. Dengue severity is labeled using a predetermined method that determines the percentage of the affected population in a specific location and applies labels such as low, medium, high, or risk to each record. To maintain quality and consistency, we preprocess the daily dengue bulletin data that we acquire from DGHS Bangladesh before using it in our data collection process. The dataset includes all of the temporal and spatial characteristics required for model construction, as well as detailed information on dengue cases in Bangladesh. With the use of this dataset, we hope to create precise forecasting models that support the nation's efficient control and containment of dengue fever outbreaks.

Year	Month	Area	Total Affected	Percentage	Labelling
2023	January	Dhaka	279	0.08686744775	2
2023	January	Faridpur	0	0	1
2023	January	Gazipur	17	0.005292998608	1
2023	January	Gopalganj	0	0	1
2023	January	Kishorganj	0	0	1
2023	January	Madaripur	9	0.002802175734	1

Figure 3.4.1: Dengue Dataset

3.5 Project Management and Financial Analysis

Project Management:

Project Objectives:

- Develop a machine learning model capable of predicting dengue outbreaks.
- Implement a user-friendly interface for accessing and interpreting dengue prediction data.
- Collaborate with public health authorities to integrate the predictive model into their decision-making processes.

Project Timeline:

A. Phase 1: Literature Review (4-8 weeks)
B. Phase 2: Data Collection and Preprocessing (4-10 weeks)
C. Phase 3: Interface Design and Development (6-8 weeks)
D. Phase 4: Integration with Public Health Systems (2-4 weeks)
E. Phase 5: Training and Testing (3-5 weeks)
F. Phase 6: Deployment (2-3 weeks)

Resource Planning:

Equipment and Tools: High performance computing resources for model training, software development tools, collaboration tools, data visualization tools.

Software and Technologies: Machine learning frameworks (e.g., scikit learn).

Data and Content: Historical dengue outbreak data, geographical information system (GIS) data, relevant research papers and publications.

Training and Skill Development: Provide training in machine learning, data analysis, software development, and project management to ensure the team has the necessary expertise.

Risk Management:

SWOT Analysis:

Strengths: Advanced machine learning capabilities, potential to significantly impact public health outcomes.

Weaknesses: Reliance on historical data for prediction, potential ethical concerns regarding data privacy.

Opportunities: Collaboration with public health authorities, potential for expansion into other disease prediction models.

Threats: Technological limitations, regulatory hurdles, competition from other predictive models.

Cause and Effect Analysis (Fishbone Diagram): Identify potential causes of delays or failures in the project, such as data quality issues, algorithmic complexities, resource constraints, and stakeholder misalignment.

Communication Plan:

A. Frequency of Communication: Weekly progress meetings with the project team, bi weekly updates to stakeholders, monthly reports on project milestones and outcomes.

B. Interview with Expert: Conduct interviews with epidemiologists, public health officials, and machine learning experts to gather insights, validate assumptions, and refine project objectives and methodologies. Schedule expert consultations at key milestones to ensure alignment with industry standards and best practices.

By effectively managing resources, identifying and mitigating risks, and maintaining open communication with stakeholders and experts, the project aims to develop a robust predictive model for dengue outbreaks, ultimately contributing to more timely and effective public health interventions.

Financial Analysis:

Table 3.5.1: Measure the cost for Project

SN	Components	Cost (BDT)
01.	Hardware/Infrastructure	1000-2000
02.	Visiting Stakeholders	2000-3000
03.	Software and Tools	500-1000
04.	Miscellaneous Expenses	1500-2000
	Total Cost	5,000-8,000

In our project we need a lot of resources, a specialized infrastructure might be necessary, such as high-performance computing. Collaborative work with the public health organizations or other stakeholders could incur costs of travel, excursions, and meeting arrangements. Machine learning frameworks, data visualization tools, and other software required for the development and analysis of models will introduce costs. Unplanned costs like software licenses, data subscriptions, or project specific expenditures may come up.

3.6 Summary

This research aims to enhance dengue outbreak prediction using smart computer programs and advanced machine learning techniques. The overview emphasizes the shift to data centric solutions, addressing gaps in existing approaches. In Requirement Analysis, key elements include obtaining reliable raw data, ethical data handling, ensuring computational power, and expertise in machine learning techniques. The Proposed Methodology outlines a comprehensive plan involving ensemble learning, feature engineering, data splitting, labeling the dataset. Data Collection focuses on gathering information from DGHS, utilizing input output analysis. Project Management and Financial Analysis ensure structured execution, team collaboration, effective budgeting, financial scrutiny, and adaptability. These efforts collectively contribute valuable insights to understanding and predicting dengue outbreaks.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The purpose of this chapter is to provide an elaboration on the working of the proposed dengue prediction model. In this process, the data is prepared for analysis, a model is built, the system is tested, and evaluated for accuracy and reliability. In this case, the main finding is to design a reliable system that can be used in accurately predicting the occurrences of dengue.

4.2 Train Model/ Prototype Design

In the training of the models, we use a number of machine learning algorithms with the include K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process, Naïve Bayes, Logistic Regression and Support Vector Machine (SVM). The implementation steps are as follows:

Data Preprocessing: First of all, the continuity and quality of the data are examined, missing values are treated, and string data are converted into numerical variables for training. This is done in order to facilitate in passing clean data into the model to enhance its performance.

Model Training: In this regard, we have further divided the collected data in ratio of 70:30 to assign it in the training as well as testing data set. It is customary to employ hyperparameter tuning for every machine learning model used in the subsequent stages. The models are then trained on the given training set, and then tested on the testing set with the help of parameters like accuracy, F1 score, precision and recall. We sacrificed 70% of training and 30% of testing data. And we also examine 80% training dataset and 30% testing dataset to see the difference between them.

Environment Setup: The environment setup is important when it comes to training and testing the model and selecting the appropriate environment plays a very crucial role. For our experiments, we choose Google Colab, and we decided to add a block in which we optimize the training process with the help of using GPUs. In this environment it is convenient and secure to implement/improve our models and then test their stability.

Hardware: Make use of GPUs that are usually provided in Google Colab for computation of models.

Software: Programming language of choice is Python and the machine learning libraries are scikit-learn.

Version Control: Use Google Colab tools to manage code as well as check on changes that are made within the code.

4.3 System Testing/ Model Evaluation

Once the models have been trained, a comprehensive set of tests are done to assess their accuracy.

Accuracy Measurement: To evaluate the accuracy of each model, we use the formula percent accurate number of correctly classified instances over the total number of instances.

F1 Score: This metric takes into account both the precision and the recall and it offers just one value of the given model. It is particularly useful when there is data imbalance since it is designed to work with such datasets in mind.

Precision and Recall: Precision is calculated as the ratio of correctly predicted positive instances to all instances predicted as positive while precision is the ratio of correctly predicted actual positive instances to all instances that are actually positive. These metrics help in understanding how accurate the model is in giving out alerts on dengue outbreaks.

4.4 Summary

Consequently, this chapter outlined the deployment of our dengue prediction model. The data preprocessing steps to be followed were explained, then followed by the model training process and the techniques used for model evaluation. Employing and comparing various machine learning algorithms and testing helps develop our model that is capable of predicting the dengue outbreak with high accuracy and minimum time.

CHAPTER 5
EXPERIMENTAL RESULT AND DISCUSSION

5.1 Experimental Setup

We are utilizing really powerful hardware like GPUs on Google Colab. It helps us train machine learning models quickly. We're implementing different types of classifiers. These include K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process and many more algorithms. First, we preprocess the dataset. We deal with missing values and encode categorical variables. Next, we train the models by tuning their hyperparameters where using 30% data for testing and 70% data for train the model. This improves their performance. Then, we evaluate the models using metrics like accuracy and F1 score. We also validate them through cross validation techniques. This ensures the models work well in real situations. Finally, Google Colab gives us an organized place to do all these experiments. It's a reliable environment for our project. For understanding the models, a details analysis is discussed below.

K Nearest Neighbors (KNN): K Nearest Neighbors, termed KNN, classifies data points. It examines new points and assigns labels based on nearby neighbors. This algorithm is versatile, handling tasks like classification or regression. Despite being simple, KNN excels with modest datasets. However, choosing appropriate values like distance metrics and k values is crucial. Also, extensive computations may slow things down. In summary, KNN offers straightforward functionality with some caveats.

Random Forest: Random Forest is a method that combines many decision trees. Each tree learns from random parts of the data. The trees predict separately. Then their predictions are combined. This makes Random Forest good at classification and regression tasks. It improves accuracy by combining many different trees' predictions. It also reduces overfitting, a common issue with decision trees.

Decision Tree: A Decision Tree is a type of machine learning algorithm. It sorts information into groups using feature values. It creates a structure that looks like a tree. Each question is a decision point. Each final answer is a leaf on the tree.

Gradient Boosting: Ensemble learning method (boosting). It puts together a decision tree and builds an ensemble of the decision trees also sequentially. Every new tree is trained to rectify the errors made by the other trees. The form of the final model is a sum of the trees weighted by some coefficient. From a single tree and continues to build more trees to reduce the loss function (for regression, it is a mean squared error, and for classification, log loss).

Gaussian process: Probabilistic model. It specifies a distribution over functions and Bayesian learning is used to update the distribution based on the observed data. All functions must be incorporated, to make predictions. Requires computing the covariance matrix with the training data and then applying it to transform the prior distribution of the functions unto the posterior distribution.

5.2 Experimental Results & Analysis:

In this analysis, we evaluated the performance of many machine learning algorithms but in our opinion these five models have performed the best.: K Nearest Neighbors (KNN), Random Forest, Decision Tree, Gradient Boosting, Gaussian process. The experiments were conducted using the "Google Colab" environment. Using plots, we visualized the performance evaluation of each algorithm. These graphs offer a clear representation of the models' performance during both training and validation phases. By comparing the accuracy, F1 score, precision, and recall of each algorithm, we gain insights into their effectiveness in predicting dengue outbreaks. These visualizations aid in understanding how each algorithm performs in different aspects, helping us make informed decisions about which model to deploy for future predictions.

F1 Score: The F1 score balances precision and recall, evaluating classification models. It is a useful metric for binary tasks. A high F1 score shows good performance for both precision and recall. Precision measures the correctly classified positive instances out of all positive predictions. Recall measures the correctly classified positive instances out of all actual positive instances. The F1 score combines these measures into one value. It provides a balanced way to assess a model's overall performance.

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

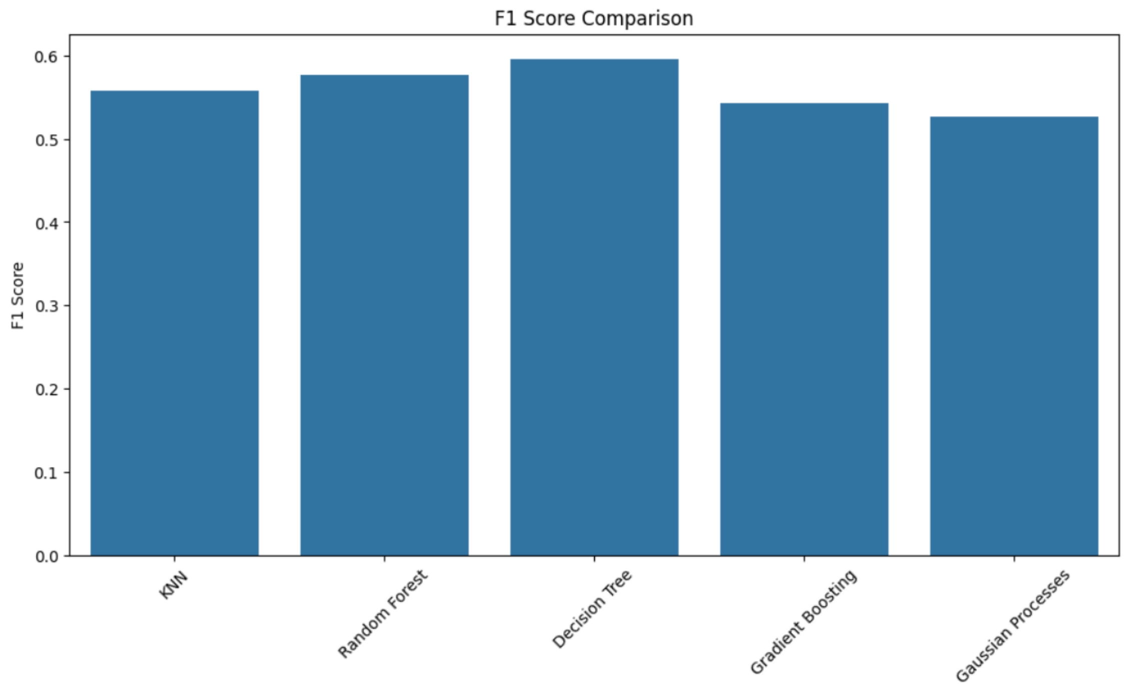

Figure 5.2.1: F1 Score Comparison

Figure 5.2.1 shows the F1 score for five algorithms. Decision Tree had the highest at 0.60, so it balanced precision and recall very well. Random Forest came second with 0.59. The scores for KNN were 0.56, which is also good. But Gradient Boosting and Gaussian processes had the lowest at sequentially 0.54 and 0.53.

Precision: Precision calculates how good a model is at correctly identifying the things it says are positive. It looks at all the model's positive predictions and checks what portion were actually true positives. Higher precision means the model made fewer incorrect positive calls, being more precise when labeling things as positive.

$$\text{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}}$$

Five algorithms' precision scores are compared in Figure 5.2.2. K Nearest Neighbors (KNN) and Gaussian processes attained 0.70, the highest precision, indicating the most true positives among all positives. Next, Decision Tree scored 0.67 precision, accurate positive predictions effectiveness. Gradient Boosting performed admirably, with 0.66 precision. Random Forest has the lowest, 0.61 precision, suggesting more false positives than others.

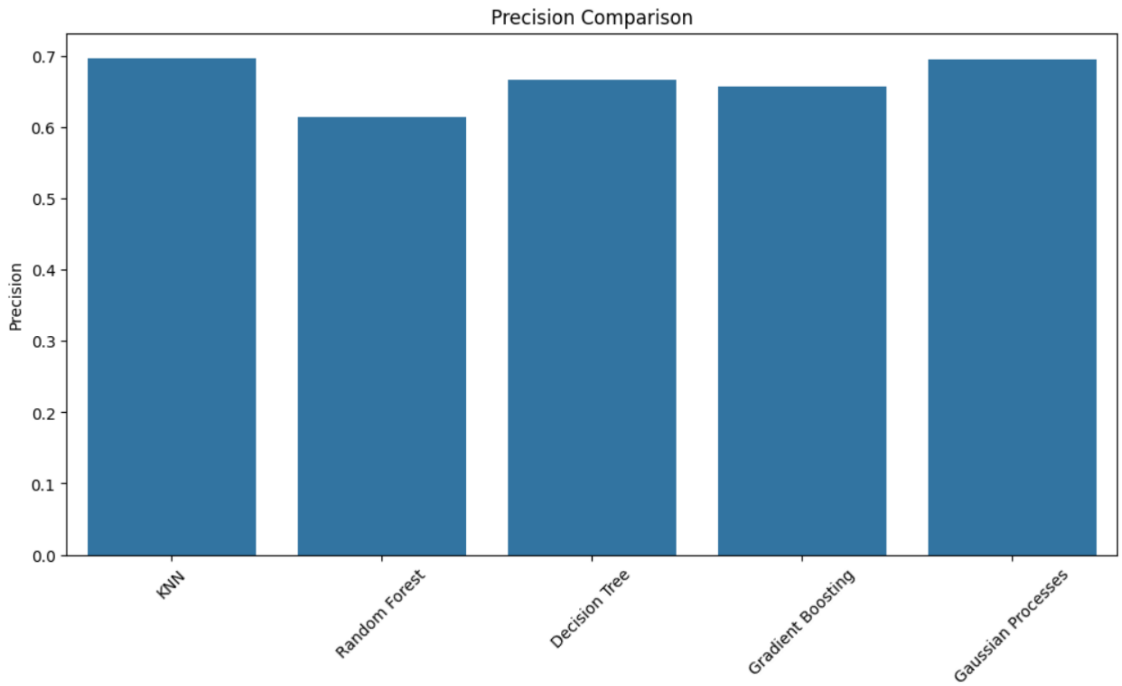

Figure 5.2.2: Precision Comparison

Recall: Recall, also known as sensitivity, is a metric used in binary classification tasks to evaluate the model's ability to correctly identify all positive instances in the dataset. It measures the proportion of true positive predictions (correctly predicted positive instances) out of all actual positive instances in the dataset.

$$\text{Sensitivity/Recall} = \frac{TP}{TP+FN}$$

A high recall indicates that the model is effectively capturing most of the positive instances in the dataset, minimizing the number of false negative predictions.

Decision Tree and Random Forest had the best recall rate, 0.55 from Figure 5.2.3, catching many right answers. KNN is the next at 0.50, also finding lots of positives. Nearby Gradient Boosting scored 0.49, but, Gaussian processing trailed, only reaching 0.46 recall more often missing real positives.

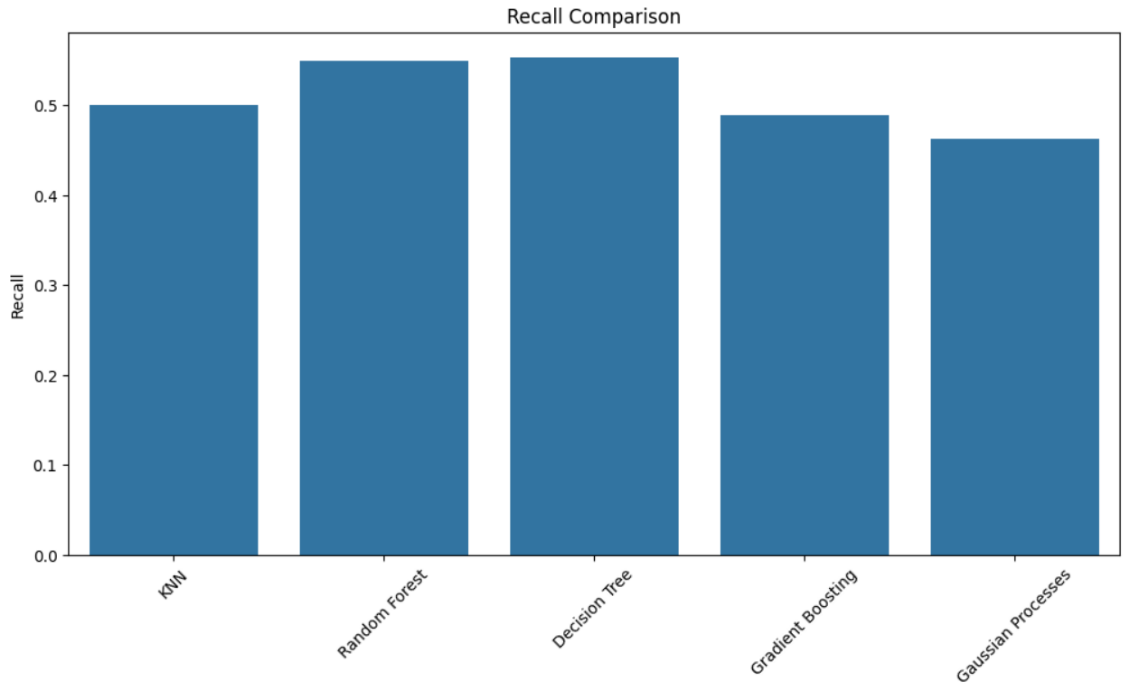

Figure 5.2.3: Recall Comparison

Accuracy: Accuracy is a metric used to evaluate the performance of a classification model. It measures the proportion of correctly predicted instances out of the total instances in the dataset. Mathematically, accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions. In machine learning, accuracy shows how frequently the model's predictions match actual dataset labels. A higher accuracy indicates that the model is making more correct predictions, while a lower accuracy suggests that the model is making more errors in its predictions.

$$\text{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FN+FP}}$$

Figure 5.2.4 shows, Decision Tree demonstrated the highest accuracy at 84%, closely trailed by K Nearest Neighbors (KNN) and Gradient Boosting at 83%. Then Random Forest and Gaussian process at 82%.

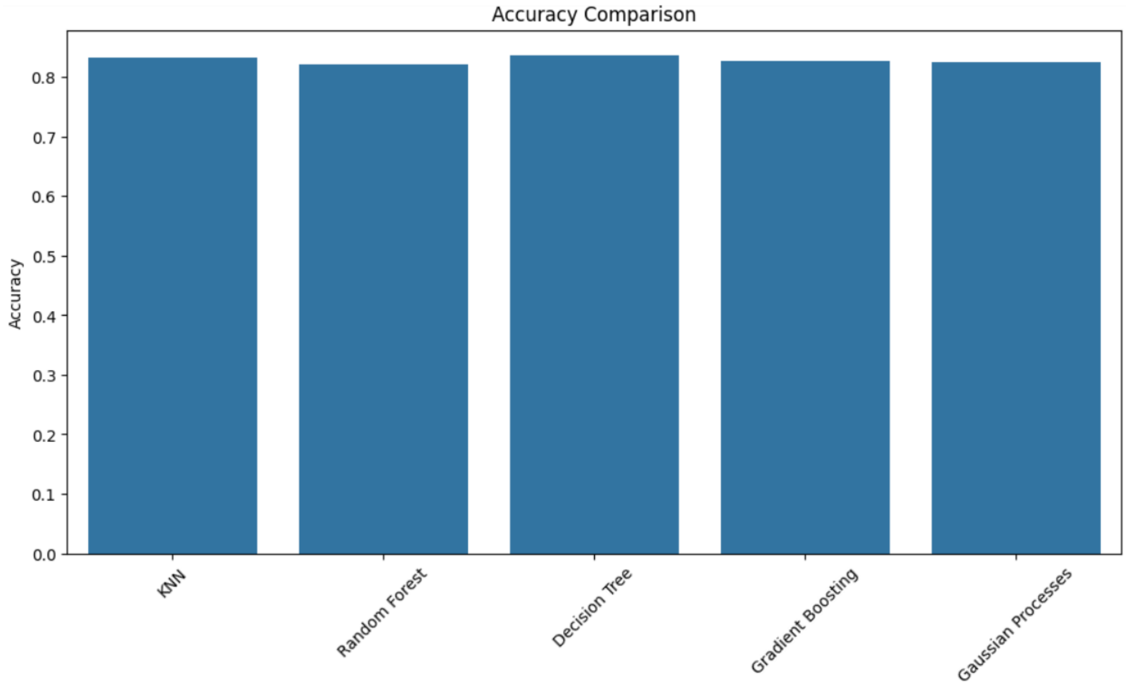

Figure 5.2.4: Accuracy Comparison

5.3 Discussion

After reviewing the results in Table 5.3.1 Decision Tree emerged as the top performing algorithm. It achieved a remarkable 84% accuracy rate Considering these results, Decision Tree emerges as the recommended algorithm for predicting dengue outbreaks. Its robust performance across various evaluation metrics suggests that it can provide reliable predictions while maintaining a good balance between precision and recall. Therefore, Decision Tree is the preferred choice for this task, offering both accuracy and reliability in forecasting dengue outbreaks.

Table 5.3.1: Algorithm Comparison

Algorithm	Accuracy	F1 score	Precision	Recall
Decision Tree	84%	60%	67%	55%
KNN	83%	56%	70%	50%
Gradient Boosting	83%	54%	66%	49%
Random Forest	82%	58%	61%	55%
Gaussian Process	82%	53%	70%	46%

Naïve Bayes, Logistic Regression, Support Vector Machine (SVM) these were also used and the accuracy was 77%, F1 score equal to 0.22, Precision is 0.19, Recall is 0.25. It was found that accuracy and f1 scores of all the models are intuitive, including precision and recall scores. We also completed, 80% of the training data and 20% of the testing data. There we found KNN and Decision Tree as the highest accuracy as they have 84%. However, the accuracy of both models is the same, we do not want to have complexity between, which for that reason we choose the result of 70% of training data and 30% of testing data. Otherwise, KNN and Gradient Boosting have provided 83% accuracy and it stands at second position. Random Forest and Gaussian process have predicted 82% of total datasets and Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) have 77% accurately.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Society

Our dengue fever prediction project's execution has a significant impact on society, affecting many facets of public health, neighborhood well-being, and hospital infrastructure. Our approach gives communities and healthcare authorities early warning signs about dengue epidemics, allowing for proactive response actions. In order to slow the disease's spread and lessen its effects on society, timely information is made possible, enabling the mobilization of resources, the adoption of preventive measures, and the coordination of healthcare services. Predicting dengue epidemics facilitates the most efficient use of healthcare resources, such as doctors, hospital beds, and pharmaceuticals. This lessens the burden on healthcare institutions and enhances the quality of care for dengue patients by guaranteeing that impacted populations receive prompt and adequate care. Our prediction models aid in lowering dengue related morbidity and fatality rates by facilitating early detection and proactive measures. Healthcare authorities can prevent the spread of the disease and lessen its impact on society by identifying high risk locations and vulnerable groups and implementing targeted interventions such vector control measures, public health campaigns, and community-based education initiatives. this project gives communities the tools and information we need to take preventative measures against dengue disease. We empower people to take proactive steps to protect their health and wellbeing by increasing illness awareness, encouraging preventative practices, and encouraging community engagement. This builds community resilience and cohesion. Our dengue fever prediction project has a wide range of social effects, including better public health outcomes, better healthcare delivery, stable economic conditions, and empowered communities. To build a healthier and more resilient society that is better able to handle the problems presented by dengue fever and other infectious diseases, we aim to make use of cutting-edge predictive analytics and proactive intervention techniques.

6.2 Impact on Environment

Although public health is our main priority, the environment will be significantly impacted by the way our dengue fever prediction study is implemented. Many environmental factors affect the spread of dengue disease, and our project indirectly supports environmental sustainability in a number of ways, The transmission of dengue fever is directly correlated with meteorological factors like temperature, humidity, and rainfall. Our approach offers preemptive public health initiatives that help reduce the spread of dengue by properly forecasting outbreaks. This ultimately helps with efforts to mitigate climate change by lessening the strain on healthcare systems and the need for reactive treatments. Insecticides and habitat alteration are two major methods used to stop the spread of dengue. Nevertheless, these actions may have unforeseen effects on nearby ecosystems, such as the destruction of natural habitats and the decline in biodiversity. Our approach lessens the need for vector control efforts, reducing their negative effects on the environment and protecting ecosystems, by offering early warning signs for dengue outbreaks. Dengue transmission is mostly caused by rapid urbanization because urban areas offer Aedes mosquitoes suitable breeding grounds. Our study helps sustainable urban planning and development by identifying high risk regions and providing information to decision makers on infrastructure, sanitation, and land use. Accurate dengue prediction makes this possible. This encourages the development of urban areas that are healthier, more resilient, and less prone to dengue outbreaks.

6.3 Ethical Aspects

Ethical issues are essential to the design and execution of our dengue fever prediction study. To guarantee the ethical conduct of research and the defense of people's rights and welfare, we place a high priority on ethical principles and practices. Important moral components of our project consist of, we are dedicated to protecting the security and privacy of personal health information that is utilized in our prediction algorithms. We make sure that personally identifiable information is anonymized and follow stringent data protection procedures to avoid unwanted access or exposure. We get people's informed consent, when necessary, before using their data for research purposes. This entails making certain that participants are aware of the study's objectives, its possible advantages and disadvantages, and their legal rights surrounding the use of their data. preserve openness in all facets of our project, including the gathering of data, the creation of models, and the decision-making procedures. In order to foster responsibility and confidence, we make information easily understandable and available to all relevant parties, such as legislators, healthcare professionals, and members of the community. We work hard to make sure that the advantages and disadvantages of our initiative are distributed equally and fairly. This entails addressing inequalities in access to resources and healthcare, taking into account the needs and viewpoints of varied communities, and advancing social justice in our research endeavors. To make sure that their perspectives are heard and their concerns are taken seriously, we actively engage with the communities impacted by dengue fever. We create partnerships based on respect and cooperation by involving community members in the planning, execution, and assessment of our project.

6.4 Sustainability Plan

To maximize the impact of our dengue fever prediction initiative on public health and societal well-being, we must ensure its long-term viability. Our sustainability plan includes a range of tactics designed to sustain the project's efficacy, scalability, and viability over time. Among the essential elements of our sustainability plan are, putting money into training and capacity building programs will provide community residents, researchers, and healthcare workers in the area the information and abilities they need to maintain and advance the prediction models. This comprises public health, epidemiology, and data science related workshops, seminars, and instructional initiatives. forming strategic alliances with governmental bodies, nonprofits, educational institutions, and business partners to promote cooperation, resource sharing, and group initiatives. By utilizing the knowledge, connections, and assets of many partners, we can increase the project's impact and more skillfully handle challenging issues. constructing and maintaining a strong technical foundation to enable ongoing prediction model deployment, upgrading, and monitoring. This entails making investments in data management programs, cloud computing resources, and software platforms that facilitate real-time data analysis, model improvement, and decision assistance. To maintain financial sustainability beyond the original project phases, it is important to diversify funding sources and investigate alternative financing options. This entails obtaining funding, contributions, and sponsorships from governmental bodies, charitable institutions, business associates, and other parties devoted to furthering public health and illness preventive initiatives. putting in place reliable monitoring and assessment systems to gauge the success, impact, and efficiency of our project interventions. This entails monitoring key performance indicators, evaluating project results on a regular basis, and getting input from stakeholders to guide ongoing adaptation and development. By incorporating these components into our sustainability plan, we hope to improve the quality of life for dengue fever affected communities and advance the overall objectives of sustainable development by having a long-lasting effect on public health, environmental resilience, and societal wellbeing.

CHAPTER 7

SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FUTURE RESEARCH

7.1 Summary of the Study

The goal of our research was to create a solid framework for forecasting dengue fever epidemics in Bangladesh by applying cutting-edge machine learning techniques. We performed a thorough examination of the several parameters impacting dengue transmission, including temporal trends, regional patterns, and meteorological conditions, using data from the DGHS Bangladesh daily dengue bulletin. By carefully preprocessing the data, creating features, and training the model, we were able to create predictive models that were able to predict dengue outbreaks quite well. The importance of proactive disease surveillance and intervention measures in reducing the impact of dengue fever on public health is highlighted by our findings. We showed that it is possible to predict outbreaks before they reach critical levels, allowing for resource allocation and prompt reaction measures, by utilizing machine learning techniques. Furthermore, in order to stop the spread of dengue fever, communities, researchers, governments, and healthcare professionals must work together, as demonstrated by our study. our study advances the field of infectious illness prediction by demonstrating the effectiveness of machine learning techniques in improving the surveillance and management of dengue fever. Our prediction models show potential for assisting in decision-making and directing focused measures to reduce the impact of dengue fever on society, provided they are further improved and validated.

7.2 Conclusions

A powerful toolkit for forecasting dengue virus outbreaks is provided by sophisticated machine learning algorithms. Through the examination of several datasets that comprise geographical, climatic, and temporal characteristics, we have proven our ability to create precise predictive models. The introduction of predictive models provides proactive disease surveillance, allowing for early detection and timely response to potential dengue epidemics. By taking a proactive stance, public health is better equipped to fend off infectious disease threats. The effectiveness of policies, data scientists, healthcare professionals, and community stakeholders in collaboration is critical to the prediction and control of dengue disease. We can effectively solve complex public health concerns by leveraging varied knowledge and resources via the development of interdisciplinary collaborations. The quality of input data determines the prediction models' accuracy and dependability. Maximizing the predictive accuracy and practicality of machine learning models requires maintaining the timeliness, completeness, and integrity of dengue related datasets. Because of changing disease dynamics, environmental factors, and technological improvements, the area of infectious disease prediction is dynamic. Predictive models must be continuously improved and validated in order to adjust to shifting circumstances and increase forecasting accuracy over time. Our findings highlight the promise of machine learning based strategies for improving the detection and management of dengue fever. We can improve community wellbeing and lessen the negative effects of dengue fever on public health by utilizing interdisciplinary collaboration and adopting data driven approaches.

7.3 Implication for Further Study

It is necessary to improve the predictive models' accuracy and dependability further. To get optimal performance, future research can investigate how to integrate new data sources, adjust model parameters, and assess innovative machine learning methods. Longitudinal studies that examine dengue fever patterns over longer periods of time might yield important information about the seasonality and dynamics of the disease. In order to guide focused intervention methods, longitudinal research makes it easier to identify recurrent patterns, temporal shifts, and developing trends in dengue transmission. Identification of localized dengue hotspots and the characterization of spatiotemporal patterns of disease transmission may be made possible by the development of spatial temporal modeling approaches. Using geographic information systems (GIS) and geospatial analysis tools can help identify high risk locations and direct the distribution of resources for focused interventions. Evaluating the ability and resilience of healthcare systems to handle dengue epidemics is essential to guaranteeing efficient monitoring and management of the disease. Subsequent investigations may concentrate on assessing healthcare infrastructure, resource accessibility, and healthcare seeking behavior in order to pinpoint areas in need of enhancement and provide guidance for policy recommendations. It is crucial to interact with dengue fever affected communities in order to raise awareness, encourage behavior modification, and organize group action. To empower people and communities in dengue preventive and control activities, future research can examine creative approaches to community participation, such as participatory techniques, social mobilization campaigns, and digital health treatments. Future research can improve our knowledge of the dynamics of dengue fever transmission and help develop evidence-based strategies to lessen the disease's negative effects on society and public health by tackling these research goals. Translating research findings into practicable interventions and promoting long-term progress in dengue fever prevention and control require cooperation between researchers, legislators, healthcare providers, and community stakeholders.

REFERENCES

[1] D. Sarma, S. Hossain, T. Mittra, M. A. M. Bhuiya, I. Saha and R. Chakma, "Dengue Prediction using Machine Learning Algorithms," 2020 IEEE 8th R10 Humanitarian Technology Conference (R10HTC), Kuching, Malaysia, 2020, pp. 16, doi: 10.1109/R10HTC49770.2020.9357035.

[2] N. V. Ganapathi Raju, P. G. Krishna, K. Manognya, G. S. S. Raj Kiran, P. Rohit and K. Likhith, "Evolution of predictive model for Dengue incidence by using machine learning algorithms," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 5159, doi: 10.1109/ICCES45898.2019.9002147.

[3] M. T. Sarwar and M. Al Mamun, "Prediction of Dengue using Machine Learning Algorithms: Case Study Dhaka," 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, Bangladesh, 2022, pp. 16, doi: 10.1109/ICECTE57896.2022.10114535.

[4] Latha Jothi, V., S. Rajasri, and M. S. R. Ramya Vipashinhi. "DENGUE PREDICTION USING MACHINE LEARNING."

[5] Salim, N.A.M., Wah, Y.B., Reeves, C. et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. Sci Rep 11, 939 (2021). https://doi.org/10.1038/s41598020791932

[6] Yavari Nejad, F., Varathan, K.D. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. BMC Med Inform Decis Mak 21, 141 (2021). https://doi.org/10.1186/s1291102101493y

[7] Janani, V., et al. "Dengue prediction using (MLP) multilayer perceptron—A machine learning approach." Int. J. Res. Eng. Sci. Manag 3 (2020): 226231.

[8] Rajathi, N., Kanagaraj, S., Brahmanambika, R., & Manjubarkavi, K. (2018). Early detection of dengue using machine learning algorithms. International Journal of Pure and Applied Mathematics, 118(18), 38813887.

[9] Khan, J. R., Farooqui, S. A., Raza, S. K., & Siddiqui, F. A. (2023). Development and Evaluation of a Predictive Diagnostic System for Dengue Fever using Machine Learning Techniques.

[10] Shaukat, K., Masood, N., Mehreen, S., & Azmeen, U. (2015). Dengue fever prediction: A data mining problem. Journal of Data Mining in Genomics & Proteomics, 2015, 15.

[11] Patel, Arvindbhai and DR. Mahammadidrish I. Sandhi. “DENGUE DISEASE PREDICTION USING DATA MINING TECHNIQUES.” (2020). Patel, Arvindbhai and DR. Mahammadidrish I. Sandhi. “DENGUE DISEASE PREDICTION USING DATA MINING TECHNIQUES.” (2020).

[12] Gupta G, Khan S, Guleria V, Almjally A, Alabduallah BI, Siddiqui T, Albahlal BM, Alajlan SA, ALsubaie M. DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. Diagnostics. 2023; 13(6):1093. https://doi.org/10.3390/diagnostics13061093

[13] DGHS Bangladesh daily dengue bulletin accessible at https://old.dghs.gov.bd/index.php/bd/home/5200dailydenguestatusreport

[14] J. L. Speiser et al., "A comparison of random forest variable selection methods for classification prediction modeling," Expert Systems with Applications, vol. 136, pp. 112122, 2019.

[15] Y.I. Anna Rejani et al., "Early detection of breast cancer using SVM classifier technique," International Journal on Computer Science and Engineering, vol. 1, no. 3, pp. 127130, 2009.

[16] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 532, 2001.

[17] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81106, 1986.

[18] B. E. Boser et al., "A training algorithm for optimal margin classifiers," in Proceedings of the fifth annual ACM workshop on Computational learning theory, pp. 144152, 1992.

[19] W. Du and Z. Zhan, "Building Decision Tree Classifier on Private Data," in Proceedings of the Australasian Computer Science Conference (ACSC 2002), pp. 3542, Australian Computer Society, Inc., 2002.

[20] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support Vector Machines for HistogramBased Image Classification," IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 10551064, 1999.

[21] Y. Liao and V. R. Vemuri, "Use of KNearest Neighbor Classifier for Intrusion Detection," Computers & Security, vol. 21, no. 5, pp. 439448, 2002.

[22] Y. Yang, J. Li, and Y. Yang, "The Research of the Fast SVM Classifier Method," in 2015 International Conference on Computational Science and Computational Engineering (CSCE), pp. 146151, IEEE, 2015.

[23] HuynhCam, T.T, Chen L.S. and Le H, "Using Decision Trees and Random Forest Algorithms to Predict and Determine Factors Contributing to FirstYear University Students'Learning Performance," Algorithms, 14(11), 318. https://doi.org/10.3390/a14110318.

[24] P.Cunningham and S.J.Delany "kNearest neighbor classifiers" ACM Computing Surveys, vol. 54, no. 6, pp. 125, 2021.

[25] S.Zhang et al., "Efficient kNN Classification With Different Numbers of Nearest Neighbors," ACM Transactions on Intelligent Systems and Technology, vol. 8, no. 3, pp. 119, IEEE, 2017.

Appendix A

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

Title:
ANTICIPATING DENGUE: ADVANCED MACHINE LEARNING FOR TIMELY EPIDEMIC PREDICTION

Student ID: [202-15-14413], [201-15-14020]

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **dengue prediction** 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 these 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 (K3)** principles by employing machine learning techniques. We are discussing about what need to do for our future progress. In this section, we describe we need good efficient of data, high power level of computer. The project demonstrates **specialist knowledge (K4)** by conducting labeling the dataset, use machine learning approach for predicting dengue.	**Page no:** [16] **Section:** [3.2] **Page no:** [1] **Section:** [1.1]

				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing our system analysis and design approach in the methodology section. Make a diagram for future work, which helps to what we do next to next.	**Page no:** [16] **Section:** [3.2] **Page no:** [19] **Section:** [3.4]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from review many conference and journal papers for betterment of work.	**Page no:** [7-12] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in dengue fever prediction, including limitations of solving the problem of this work, data collection process is not so easy. Because every hospital's data is very confidential, so that's why face some problems in collecting data.	**Page no:** [13-14] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by discuss about proposed methodology, where describe using modern machine learning approaches for getting a better accuracy in work.	**Page no:** [26-30] **Section:** [5.2]

4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting medical diagnostics in respiratory diseases like dengue, contributing to advancements in public health and healthcare practices which indicates **EP-4**.	**Page no:** [7-13] **Section:** [2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in medical diagnostics which ensures **EP-7**.	**Page no:** [16-19] **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 project utilizes diverse resources such as high-performance computing infrastructure, raw dataset and ethical considerations to ensure systematic research and contribute to advancements in dengue prediction using advanced machine learning model.	**Page no:** [20-22] **Section:** [3.5]
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 project contributes to society by improving healthcare through advanced dengue prediction methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [25-31] **Section:** [5.1, 5.2, 5.3]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in dengue prediction through advanced machine learning techniques, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [7-12] **Section:** [2.1, 2.3]

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

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [4-6] **Section:** [1.6]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical	**Page no:** [30-31] **Section:** [5.3]

			application of advanced healthcare technologies which comply with **CO9**.	
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [16-22, 26-30] **Section:** [3.2, 3.3, 3.4, 3.5, 5.2]

report dengue

ORIGINALITY REPORT

17%	13%	8%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Tianyu Zhang, Fethi Rabhi, Ali Behnaz, Xin Chen, Hye-young Paik, Lina Yao, Chandini Raina MacIntyre. "Use of automated machine learning for an outbreak risk prediction tool", Informatics in Medicine Unlocked, 2022 Publication	1%
5	libweb.kpfu.ru Internet Source	1%
6	ouci.dntb.gov.ua Internet Source	1%
7	Muhammad Danish Waseem, Ali Nawaz, Uzair Rasheed, Abir Raza, Mubarak Omar Albarka. "A Time Series Regression-based Model for Predicting the Spread of Dengue	1%