DENGUE FEVER PREDICTION USING MACHINE LEARNING APPROACHES

BY

Ishrat Suchita
ID: 201-15-13709

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

Supervised By

Tapasy Rabeya
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2024

APPROVAL

This Project/internship titled **"Dengue Fever Prediction Using Machine Learning Approaches"**, submitted by Ishrat Suchita, ID No: 201-15-13709 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *23 January 2024*.

BOARD OF EXAMINERS

Chairman

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Nazmun Nessa Moon (NNM)
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

Dewan Mamun Raza(DMR)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Md. Zulfiker Mahmud (ZM)
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We, therefore, declare that this undertaking has been finished by us under the supervision of **Tapasy Rabeya**, Lecturer (Senior Scale), Department of CSE, Daffodil International University. We further declare that neither an application or an educational grant has been made anywhere for this project or any part of it.

Supervised by:

22-01-24

Tapasy Rabeya
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Ishrat Suchita
23-01-24

Ishrat Suchita
ID: 201-15-13709
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are really grateful and wish our profound indebtedness to **Tapasy Rabeya**, 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. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

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

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

ABSTRACT

Dengue fever, also a viral infection spread by mosquitoes, is still a major global health concern, impacting millions of people each year. By applying a carefully controlled dataset of 521 entries and 23 variables, this study analyzes the predictive efficacy of various machine learning methods for Dengue Fever. Among the methods tested, SVM outperforms the others, obtaining an excellent accuracy of 98.88%. This remarkable accuracy highlights the algorithm's ability to capture complex patterns within the multidimensional dataset, establishing it as a strong choice for Dengue Fever detection. MLPclassifier comes in second with an impressive accuracy of 97.78%, indicating its ability to handle the dataset's constant characteristics. The success rate of Logistic Regression and GaussianNBis 96.95% and 93.64%, respectively, illustrating how they adjust to the dataset's complexities. BernoulliNB, on the other hand, lags behind with a lower accuracy of 67.85%, indicating limits in dealing with the dataset's peculiarities, particularly given its affinity for binary features. SVM exceptional accuracy highlights its promise as a significant tool for effective Dengue Fever detection. The study provides essential knowledge for health professionals and academics, guiding the selection of the most successful modeling algorithms in the context of infectious diseases.

Keywords: *Dengue, BernoulliNB, GaussianNB, SVM, LogisticRegression, MLPclassifier.*

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Dengue fever, also a viral infection spread by mosquitoes, is still a major global health concern, impacting millions of people each year. Because dengue epidemics are unforgiving, creative and proactive measures to avoid or contain them are required. In this regard, machine learning (ML) approaches appear to be a viable route for improving our ability to forecast the presence and spread of dengue fever.

The huge number of elements impacting dengue exchange, including climatic circumstances and mosquito populations to human demography and immunological responses, highlights the need for sophisticated analytical techniques. Standard methods frequently fall short of capturing the complex relationships found in this multidimensional dataset. Machine learning, with its ability to recognize patterns, learn from prior data, and adapt to changing situations, is a valuable tool for predicting dengue illness.

The current study focuses on developing and testing a machine learning-based dengue fever forecasting model. We want to capitalize on the power of ML algorithms to give accurate and timely forecasts by using a precisely collected dataset customized to the complexities of transmitting dengue. The dataset contains a variety of variables, such as demographic information, blood-related measurements, and environmental factors, which are critical when analyzing the dynamics of dengue disease.

The importance of this undertaking stems from its capacity to change the way public health officials approach preventing dengue. If the model is successful, it might be used as an early warning system, allowing for timely measures, resource allocation, and community awareness efforts. As the world struggles with rising infectious illnesses, the integration of public health expertise and cutting-edge technology becomes critical. We hope to contribute to current efforts to reduce the burden of dengue fever and create the road for a more resilient and responsive healthcare system through this work.

1.2 Motivation

The inspiration for this initiative stems from the critical need for preventive measures to tackle the significant public health threat posed by dengue disease. Dengue fever, with its risk of rapid infection and severe medical effects, need novel ways that go beyond established epidemiological methodologies. The growing frequency of dengue outbreaks around the world highlights the absence of current methods for forecasting and minimizing the disease's impact on populations.

The use of machine learning provides a game-changing answer by utilizing the power of data to uncover complex connections and patterns that conventional analysis misses. The reason for using machine learning to forecast dengue fever originates from the notion that predictive modeling can provide a critical advantage in the fight against this transmissible disease. We hope to establish a thorough and dynamic system capable of predicting the commencement and progress of dengue epidemics by developing a model that integrates varied datasets such as data on demographics, blood-related metrics, and environmental variables.

The project is inspired by the opportunity for public health approaches to be revolutionized, allowing for prompt and specific action. A successful model for prediction could enable medical organizations to better allocate resources, adopt proactive strategies, and educate communities, decreasing the burden of dengue on both people and healthcare systems. The motivation for navigating the complexity of developing viral illnesses is anchored in the notion that the integration of modern technology and public health knowledge can generate revolutionary approaches for a happier and more resilient global population.

1.3 Rationale of the Study

The research team's reasoning is based on the urgent need to address the major public health threat posed by dengue fever. Dengue fever, a mosquito-borne viral virus that affects millions of people each year, reveals complicated and changing trends that necessitate advanced forecasting methods. Classical epidemiological methodologies frequently fail to provide a timely and accurate understanding of the factors that influence the spread of dengue. The incorporation of machine learning technologies, which are known for their capacity to discover patterns from disparate datasets, provides a unique and exciting path for increasing the knowledge and prediction of cases of dengue. The present research intends to add to the arsenal of tools available to public health officials for preventive and successful disease management by building a model of prediction suited to the intricacies of the spread of dengue. The extensive dataset, which includes data on demographics, blood-related metrics, and surroundings, ensures a complete approach to modeling the complex interaction of variables impacting dengue recurrence. The argument is ultimately anchored in the quest for a transformational solution that integrates technical breakthroughs with public health goals, enabling a more flexible and adaptable strategy for treating infectious illnesses on a global level.

1.4 Research Question

i. How does age influence the likelihood of contracting dengue fever?

ii. What impact does gender have on the occurrence of dengue fever?

iii. Is there a correlation between mosquito populations and the incidence of dengue?

iv. Are there distinct patterns in platelet count that coincide with dengue fever occurrences?

v. Does the presence of specific symptoms, such as headaches or joint pain, indicate a higher probability of dengue infection?

1.5 Expected output

The design and testing of a strong machine learning-based predictive model for Dengue Fever is the expected outcome of this project. We anticipate that the model will be highly accurate in anticipating the development and spread of Dengue Fever epidemics. We hope to create a model that accurately reflects the complicated interactions between population parameters, blood-related metrics, and environmental variables by rigorous data selection, preparation, and feature creation.

In addition, the anticipated result includes helpful knowledge into the basic causes of Dengue Fever, giving light on how age, gender, and certain blood-related variables impact the likelihood of disease. The model is expected to be a significant tool for public health officials, allowing them to assign resources in advance, execute targeted preventative measures, and improve response methods based on precise forecasts.

We believe that the output will help to a better knowledge of the clinical patterns of Dengue Fever, and also to precise predictions. The model's findings could help shape future research, public health policy, and community education efforts. Finally, the research aims to empower health professionals and communities in the battle against Dengue Fever, encouraging a more resistant and informed method of managing the illness.

1.6 Project Management and Finance

Useful managing projects and budget planning are critical parts of this research initiative's success. A well-defined project management plan will be used to guarantee that each phase, from data collection to model delivery, runs smoothly. Establishing specific goals, a detailed timeline and a methodical allocation of roles between team members are critical tasks. To address any difficulties as soon as possible, regular progress evaluations and communication channels will be maintained.

Financial preparation will be carefully organized to address a wide range of issues, including data gathering costs, computing resources, and possible partnerships with subject experts. Budget transparency will be preserved, and resources will be used to maximize research outputs. Continuous financial plan monitoring will allow for reactive revisions to match with changing project requirements. This research attempts to ensure effectiveness, honesty, and responsibility through excellent project management and cautious financial planning, assuring the successful development and implementation of a solid Dengue Fever prediction model. In table 1.1, I shown the project management table.

Table 1.1: Project Management Table

Work	**Time**
Dataset	1 month
Literature Review	3 month
Experiment Setup	1 month
Implementation	2 month
Report	2 month
Total	9 month

1.7 Report Layout

- **Introduction**
  The introduction provides a clear overview of the research's context, objectives, and significance, attracting the reader's attention and laying the groundwork for further exploration.
- **Background**
  This section provides a comprehensive understanding of the research's historical context, existing literature, and relevant theories, providing a comprehensive understanding of the subject matter and the gaps it aims to address.
- **Data Collection & Data Preprocessing**
  The research employs rigorous methods for data collection, ensuring high-quality and relevant information. The data preprocessing phase cleans, organizes, and transforms the data for analysis, enhancing its credibility and integrity.
- **Research Methodology**
  This section provides a comprehensive overview of the research methodology, including design, data analysis, statistical methods, and computational techniques, thereby establishing a clear framework for the research process.
- **Experimental Result and Discussion**
  The research presents experimental results, backed by thorough analysis, and then discusses the results within the study's context, utilizing existing literature to clarify their implications and significance.
- **Impact on Society and Environment**
  The study assesses the societal and environmental implications of its findings, highlighting the potential impact on society and the environment, and the benefits for various stakeholders.
- **Summary, Conclusion, Future Research**
  This section summarizes key research findings, provides insights, and suggests future research areas for further exploration, contributing to the ongoing development of knowledge in the field.
- **References**
  The research acknowledges sources, providing a scholarly trail for readers to access cited literature, fostering transparency and recognizing contributions from others.

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

The planning stage of this research consists of fundamental tasks required for the successful implementation of the Dengue Fever prediction project. The first stage is to put together a team of researchers from different fields with experience in epidemiology, data science, and public health. This concerted effort guarantees that the complexity of Dengue Fever transmission patterns is addressed thoroughly. The ensuing steps of data collecting and model construction are guided by the definition of particular research goals and the formulation of appropriate research questions. To ensure commitment to ethical standards throughout the study process, ethical issues such as data privacy and informed permission protocols are crucial to early planning. In addition, developing a thorough project schedule, allocation of resources plan, and danger evaluation framework is critical in moving the project's progress. Early collaboration with participants, such as regional health departments and members of the community, will increase involvement and support for the initiative, contributing to its success. The early phase lays the framework for an organized and well-coordinated research effort targeted at yielding significant knowledge into Dengue Fever prediction.

2.2 Related Works

Previous Dengue Fever prediction research has mostly focused on clinical and statistical methods, which are frequently hampered by their dependency on traditional methodologies. Some research has looked into the usage of machine learning, with a focus on particular algorithms such as MLPClassifier Algorithm and Support Vector Machines. However, there is a void in the literature about the thorough integration of varied datasets into models of prediction, including demographic, blood-related, and environmental aspects.The present research aims to lose the gap by taking a comprehensive approach to Dengue Fever estimate, utilizing machine learning techniques to detect detailed patterns within a thoroughly collected dataset. The research being conducted wants to add fresh viewpoints to the current collection of information and move the field closer to more reliable and timely predictions of Dengue Fever breakouts.

Gupta, Gaurav, et al. [5] emphasized how vital it is to correctly and promptly differentiate between Dengue and its subclasses in the initial phases of sickness development. The ability to seek quick diagnosis and treatment thanks to the ability to predict Dengue fever can save lives. Data from RNA-Seq and microarrays have been crucial in the development of Dengue predictive models. When assessing sentiment from text at the sentence or phrase level, conventional techniques like Bayesian conclusions and support vector machines may not have the necessary semantic heft. The study suggests using machine learning to predict dengue illness, with the random forest classifier showing the greatest mean score of 8.72.

Majeed, Mokhalad A., et al. [6] focused on applying machine learning to forecast Dengue fever cases in Malaysia using a dataset from 2010 to 2016 that includes demographic, geographic, and climatic factors. To predict Dengue, six different LSTM models were created and evaluated. With an average RMSE of 3.17 over various lookback times, the SSA-LSTM model, which combines stacked LSTM layers with spatial attention, produced the most encouraging results of these models. This SSA-LSTM model proved effective in various Malaysian states and outperformed three benchmark models (SVM, DT, and ANN) in terms of RMSE. The findings highlight the effectiveness of the SSA-LSTM model in forecasting Dengue cases in Malaysia, with the lowest average error when compared to other models.

Zargari Marandi, Ramtin, et al. [7] presented a machine-learning model made to find markers for plasma leakage in the initial stages of dengue illness. Because it provides decision support for low- to middle-income countries like Sri Lanka, it is particularly noteworthy for its application in resource-constrained contexts. The study reveals five easily accessible variables that can help in the choice of whether to admit probable dengue cases to the hospital. Through the use of fairness considerations, Shapley additive explanations, and multi-metric evaluations, the work emphasizes model transparency and interpretability. With significant implications for monitoring and averting serious effects in dengue patients, the final model performs well, achieving an AUC of 0.80, a PPV of 76.9%, an NPV of 72.5%, specificity of 87.9%, and sensitivity of 54.8% on the test dataset.

Panja, Madhurima, et al. [8] proposed the XEWNet model, an ensemble wavelet neural network incorporating exogenous parameters, which is capable of accurately predicting dengue outbreaks across three different geographical areas. The model is flexible and can easily integrate exogenous climate variables that have been proven by statistical causality testing. The XEWNet model generates longer-term forecasts that are more reliable because it incorporates wavelet transformation into a neural network structure. While staying mathematically comprehensible, computationally effective, and simply understood, it effectively captures complex non-linear correlations between dengue incidence and rainfall. Comparative studies reveal that for both short-term and long-term dengue incidence predictions, the proposed XEWNet beats statistical, machine learning, and deep learning methods in 75% of the instances.

Yang, Hsiu, et al. [9] Utilized techniques from Extreme Gradient Boosting and Shapley Additive Explanation (SHAP), an explainable artificial intelligence methodology. The study's objective was to evaluate the spatial distribution of dengue cases homes in Kaohsiung City, Taiwan, between 2014 and 2015. The percentage of general roadways and residential areas had strong, nonlinear relationships with dengue case homes, according to key findings. On the other hand, the incidence of dengue was adversely linked with agricultural characteristics. Additionally, the Shannon diversity index showed a U-shaped relationship between dengue infections and infections, and the SHAPE-dependent plots revealed significant relationships between different types of land use and the incidence of dengue. As a result, precise landscape-based prediction maps that may be used to direct resource allocation and control strategy modifications for dengue prevention were produced.

Indhumathi, K. et al.[10] utilized electronic patient health information, it is possible to forecast and predict the spread of infectious diseases and identify those with the highest risk of spreading quickly. The study's three main goals are to forecast illness incidence using machine learning and deep learning algorithms, detect lethal symptoms depending on patient age, and predict disease occurrence before it occurs. The suggested model adopts a hybrid methodology, fusing feature selection with classification strategies. The Antlion Optimization Algorithm is utilized to pick features, whereas Random Forest combined with XG-Boost is employed to classify diseases. The study also examines disease forecasting on a daily and annual scale, making forecasts based on historical epidemic occurrences as well as climate variables like temperature and rainfall.

Hadi, Z. A., et al. [11] addressed the difficulties of combining sociodemographic and meteorological data sources to comprehend the causes of dengue occurrence. In order to conduct epidemiological and spatial investigations, it introduces the use of Geographic Information Systems (GIS) and Machine Learning (ML) methods. Without explicit programming, the machine learning technique allows for the analysis and prediction of outcomes based on data. The results of the study are intended to support efforts to manage dengue in Malaysia and maybe broaden the framework to include other mosquito-borne illnesses like malaria and chikungunya as well as ailments that are not caused by mosquitoes.

Rustom Al Nasar, Mohammad, et al. [12] focused on predicting dengue illness, this study integrates machine learning methods, notably Artificial Neural Networks (ANN) and Support Vector Machine (SVM). The statistics, which comprise records for diagnosing dengue illness (positive or negative), were collected from a government hospital in Lahore, Pakistan. 30% of the data is used for testing, while 70% is used for training. To calculate the results of the dengue diagnostic, the fused model uses membership functions. The model is saved using a cloud storage system and includes current patient data for later use. Notably, the suggested model outperforms earlier research efforts with an amazing accuracy rate of 96.19%.

Sarma, Dhiman, et al. [13] addressed the urgent need for accurate dengue epidemic forecasting given the dearth of vaccinations and antiviral medications. The study presents a unique machine-learning method for predicting dengue fever utilizing patient information, such as diagnosis reports, medical histories, and symptoms. Real-time data samples from dengue patients in Bangladesh were used to create the dataset. A 70:30 split of the data is made between training and testing, with 70% going to training. The classification model uses machine learning methods, notably decision tree (DT) and random forest (RF). In the end, the decision tree model outperforms the random forest model with an average accuracy of 79%.

Salim, Nurul Azam Mohd, et al. [14] addressed the five Selangor, Malaysia, districts with the highest dengue fever incidence from 2013 and 2017 to determine the most effective machine learning model for predicting dengue outbreaks. Each model takes into account climate factors such as temperature, wind speed, humidity, and rainfall. With an accuracy of 70%, specificity of 95%, and precision of 56%, the SVM (linear kernel) has the best prediction performance. Notably, the testing sample's SVM (linear) sensitivity rises to 63.54%, a significant improvement over the initial imbalanced data's 14.4%. The SVM model's most significant predictor turns out to be the week-of-the-year variable, demonstrating the promise of machine learning for forecasting dengue outbreaks.

Nguyen, Van-Hau, et al. [15] addressed the significance of creating a Dengue Fever (DF) early warning system in Vietnam as a climate change adaptation mechanism. Deep learning algorithms and a variety of climatic input variables will be used to construct a prediction model for DF rates in Vietnam. The best model among those put to the test is LSTM-ATT, which received an average ranking of 1.60 based on RMSE and 1.90 based on MAE. LSTM-ATT outperforms LSTM in predicting DF incidence, especially for RMSE and MAE in 12 of the 14 provinces. Deep learning models are valuable for climate-based DF forecasting in Vietnam, as shown by the ability of LSTM-ATT to reliably predict DF incidence and outbreaks up to 3 months in advance.

Iqbal, Naiyar, et al. [16] addressed on predicting dengue epidemics as having the potential to reduce mortality and boost diagnostic efficiency. In this study, a machine learning-based workflow for predicting Dengue cases—whether they are positive or negative—is established. Utilizing the WEKA method to forecast dengue outbreaks is the main research emphasis. Eight parameters are used to evaluate the performance of seven well-known machine learning algorithms. The LogitBoost ensemble model performs best, with sensitivity and specificity of 90% and 94%, respectively, and an amazing classification accuracy of 92%.

Sajana, T., et al. [17] focused on machine learning as essential for using various algorithms to provide quick and accurate results. Since they can be used effectively and practically on personal electronics like computers, cellphones, or tablets, these computational methods are very intriguing. Predictions can be made quickly using trained statistics and machine learning models. To help clinicians evaluate the risk in dengue patients, this study will use non-invasive machine learning approaches. In a comparison investigation of the Simple Classification and Regression Tree (CART), Multi-layer Perceptron (MLP), and C4.5 algorithms, it was shown that the Simple CART algorithm had an astounding 100% accuracy in identifying afflicted and unaffected patients.

Rahman, Md Siddikur, et al. [18] used machine learning to map the socioeconomic, climatic change, dengue knowledge, attitude, and practices (KAP) and landscape characteristics in northeastern Thailand in order to estimate the quantity of female adult Aedes aegypti mosquitoes.

Using geographic information systems (GIS), it is possible to figure out which homes have high and low abundance. The random forest (RF) model demonstrated the greatest prediction performance across various parameters when five supervised learning models were tested.The RF model effectively predicted the abundance of female adult Ae. aegypti in northeastern Thailand, with an AUC of 0.93 to 0.96, classification accuracy of 0.86 to 92, and F1 score of 0.85 to 0.90. The model's most significant predictor was dengue prevention efforts, resulting in an AUC of 0.99, CA of 0.96, and F1 of 0.95.

Sanjudevi, D., et al. [19] focused on using classification methods to increase prediction accuracy using a dataset on Dengue disease. Using WEKA as a data mining tool, the proposed method is contrasted with the decision tree. Data collection from the UCI repository, feature selection using regression techniques, application of SVM and decision tree classification algorithms, and evaluation using sensitivity (SE), specificity (SP), and area under the curve (AUC) metrics are the four steps of the study. The outcomes show that the suggested SVM classification method performs better than the decision tree, reaching an astounding 99% accuracy, a quicker processing time, and a reduced error rate. The study recommends using more sophisticated technologies to analyze the consequences of dengue more thoroughly.

Macedo Hair, Gleicy, et al. [20] proposed employing machine learning to find features for a risk criterion for the severity of dengue patients. Self-organizing maps (SOM) and random forest algorithms were used to examine the clinical characteristics of 523 confirmed dengue cases, resulting in the discovery of four natural clusters. One cluster had severe dengue patients with a high frequency of warning indicators, two clusters represented dengue without warning signs or mild sickness, and another displayed intermediate characteristic. Age was found to be the main factor in clustering, but other warning indicators, such as stomach pain, clinical fluid buildup, mucosal bleeding, and others, should also be taken into account when determining the severity of a dengue patient. These findings highlight the significance of age in dengue-endemic areas and the necessity of careful observation of warning symptoms, particularly in youngsters. They also point to the urgency of more longitudinal research to fully comprehend the clinical manifestations of dengue.

Ming, Damien K., et al. [21] Utilized combined data from adult and pediatric dengue hospital patients, a study in Ho Chi Minh City, Vietnam, produced supervised machine learning prediction models. Hyperparameter optimization was carried out using ten-fold cross-validation after the data was split into training and hold-out sets. Utilizing the hold-out set, the models were assessed. 5.4% of the total 4,131 patients—477 adults and 3,654 children—had dengue shock syndrome, according to the final dataset. Age, sex, weight, the day the patient was hospitalized, the hematocrit, and the platelet indices over the first 48 hours were all taken into account as predictors. The most effective model was an artificial neural network (ANN), which achieved an AUROC of 0.83 when measured against a separate hold-out set and an AUROC of 0.82 overall.

Abdualgalil, Bilal, et al. [22] proposed ensemble machine learning techniques (EMLT) prediction models for dengue sickness. K-Nearest Neighbor (KNN), Gradient Boosting Classifier (GBC), Extra Tree Classifier (ETC), eXtreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LightGBM), five effective machine learning models, were used. Using the dataset and the 10-fold Cross-Validation and Holdout Cross-Validation procedures, all classifiers were trained and tested. Various measures, including accuracy, F1-score, recall, precision, AUC, and operation time, were evaluated on a test set. By achieving the maximum accuracy in both Holdout (99.12%) and 10-fold cross-validation (99.03%) techniques, the Extra Tree Classifier (ETC) emerged as the best performer. The paper emphasizes how the suggested dengue prediction system might help doctors forecast diseases accurately.

Kaur, Supreet, et al.[23] proposed a framework for data collecting and analysis to predict the existence of dengue infection and evaluate risk levels by categorizing patients into several dengue infection groups. Real-time diagnosis based on symptomatic and clinical investigations is used, along with a tertiary classification technique based on machine learning. Patients are notified by the system of the potential for internal bleeding through warning indicators, pushing them to seek emergency medical attention. According to the World Health Organization's classifications of dengue fever, dengue hemorrhagic fever, and dengue shock syndrome, the model demonstrates high sensitivity and specificity in predicting infection levels, achieves high accuracy of over 90%, and is accurate in predicting dengue infection levels in all three of these forms. The model's performance effectiveness and usefulness are highlighted by the experimental evaluation, which is supported by statistical methods.

Gomes, Ana Lisa V., et al. [24] presented a unique use of the Support Vector Machines (SVM) method to examine the patterns of 12 genes during the acute viral infection in peripheral blood mononuclear cells (PBMCs) of 28 dengue patients (13 DHF and 15 DF). Utilizing leave-one-out cross-validation, the SVM model attained an accuracy of over 85%. The study pinpoints seven genes, including MYD88 and TLR7, as crucial in discriminating between DF and DHF patients through the selective removal of gene expression data. The overall accuracy rises to nearly 96% once the SVM model is retrained using the two key genes and five additional genes. The classification of DF and DHF patients may be aided by these findings, which also highlight the importance of numerous genes but still need further experimental confirmation.

Since Gaussian Naive Bayes and Logistic Regression produce the same outcomes, their simplicity, and understanding may be useful in cases that need transparent model outputs. Although suited for binary features, Bernoulli Naive Bayes has difficulties when dealing with the complicated Dengue Fever dataset.

In conclusion, the comprehensive examination indicates the success of machine learning in predicting Dengue Fever. The method chosen should meet the needs of the project, taking into account elements such as understanding, computational effectiveness, and the necessity to handle nonlinear interactions. The findings provide helpful information for public health professionals looking for ideal models for proactive Dengue Fever care, providing detailed knowledge of the strengths and trade-offs associated with various machine learning approaches.

2.3 Comparative Analysis and Summary:

Table 2.1: Accuracy Comparison between Papers

Author	Dataset Collection	Algorithm	Accuracy Result
Zargari Marandi, Ramtin, et al. [7]	own	Machine Learning	87.9%
Rustom Al Nasar, Mohammad, et al. [12]	own	Machine Learning	96.19%
Sarma, Dhiman, et al. [13]	own	Machine Learning	79%
Iqbal, Naiyar, et al. [16]	no	Machine Learning	92%
Salim, N A Mohd,et al. [14]	own	Machine Learning	70%
My Paper	own	Machine Learning	98.88%

A brief overview of dengue-related research was presented in Table 2.1 observed that most of the researches focused on publicly available secondary data. Therefore, achieving a satisfactory accuracy of a good model has no rationale where the origin of data and the area of implementation are both different geographic locations. Moreover, most of the datasets had limited features, whereas, in this study, we added relevant fever-specific attributes through the patient's personal, diagnostic, and symptom data.

2.4 Scope of the Problem

The nature of the Dengue Fever problem is broad and goes past the acute health consequences. Dengue fever is a big strain on worldwide public health systems, with periodic outbreaks affecting millions of people each year. The disease's geographical spread hampers control even more, needing thorough prevention and control techniques. Furthermore, Dengue Fever has a significant financial impact, especially in areas with minimal resources.

The development of global interconnection increases the risk of Dengue delivery, underlining the importance of prediction algorithms that can adjust to changing situations. The breadth of the problem grows as global warming continues to impact vector behavior and disease spread, requiring flexible and adaptable responses.

2.5 Challenges

The creation of a Dengue Fever predicted model is filled with difficulties due to the complexities of the disease dynamics and the data involved. Data quality is a key challenge since differences or mistakes in demographic, medical, or environmental data could compromise the model's trustworthiness. The dynamic nature of Dengue Fever spread, which is impacted by environmental elements and human activity, adds another layer of complexity to the model, leading it to adjust to changing situations.

In addition, the lack of sufficient and high-quality Dengue Fever predicted datasets may limit the model's training and universality. Accessing and using real-world datasets is complicated by moral issues, particularly data protection and informed consent. Furthermore, striking a balance between the level of complexity and clearness is critical, as too complex models can hinder comprehension and acceptability by physicians and the general public. Addressing these issues is critical for the effective development and implementation of a reliable Dengue Fever prediction model.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The research project's topic is Dengue Fever projection, with a particular emphasis on using machine learning technologies for accurate and overriding planning of Dengue outbreaks. The study is to look at the intricate relationship of demographic parameters, blood-related measurements, and environmental variables to develop an in-depth prediction model.

The current research's equipment includes the use of advanced machine learning techniques such as Random Forest, Support Vector Machine (SVM), Gaussian Naive Bayes, Logistic Regression, and Bernoulli Naive Bayes. These algorithms act as tools for analysis, detecting patterns in a well-curated dataset specific to the subtleties of Dengue infection. The instrumentation includes data preparation techniques, selecting features methodologies, and model assessment measures, all of which are intended to aid in the development of a strong and accurate prediction model. The research topic and equipment constitute the framework for a comprehensive investigation of Dengue Fever prediction, employing cutting-edge technology to improve our understanding and effective control of this public health concern.

3.2 Data Collection Procedure

The information-gathering approach for this Dengue Fever predicting research began with the meticulous compilation of a dataset of 521 items and 23 attributes. Primary information on demographics, such as age and gender, was gathered with key blood-related parameters, such as hemoglobin (HB%), ESR, total WBC count, neutrophils, lymphocytes, monocytes, eosinophils, and basophils. Environmental parameters such as MCV, MCH, MCHC, RDW-CV, RDW-SD, Platelet Count, MPV, PDW, and PCT were also included in the dataset.

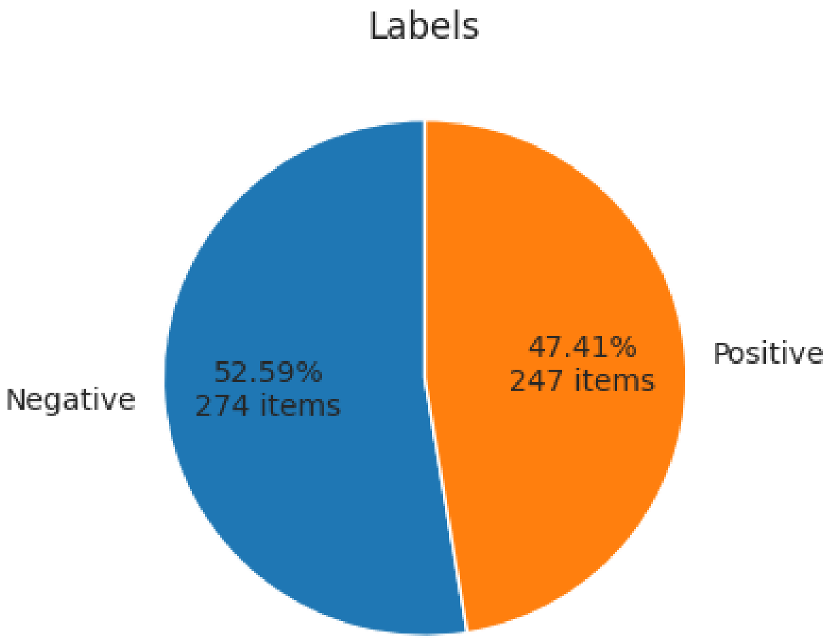

Figure 3.1: Data Ratio

Every record was carefully classified based on the existence or absence of Dengue Fever, with two classes in the "Labels" attribute: Negative 274 entries and Positive 247 entries. Data gathering followed ethical guidelines, assuring privacy and informed permission. Partnerships with healthcare providers and specialists facilitated access to important medical records, assuring the accuracy and completeness of the dataset. This dataset's rigorous management serves as the foundation for strong and relevant analyses, propelling the future phases of preprocessing, feature selection, and machine learning model construction in the pursuit of effective Dengue Fever detection.

3.3 Statistical Analysis

The mathematical investigation of the Dengue Fever predictor dataset entails a thorough evaluation of the correlations between numerous parameters in order to discover patterns and discoveries. Descriptive metrics provide an overview of the dataset's major tendencies and variability, providing knowledge about the patterns of age, gender, and blood-related parameters. Correlation analysis evaluates the degree and direction of correlations between variables, identifying probable connections that are important for modeling predictions.

In addition, statistical methods such as t-tests or chi-square tests can be used to establish the importance of differences across groups, particularly between Dengue-positive and Dengue-negative cases. Statistical techniques such as ANOVA can be used to investigate variations across many categories. These investigations help to identify critical characteristics that influence Dengue Fever incidence, leading to further stages of feature selection and model construction. The incorporation of strong statistical approaches improves the research's stability, guaranteeing that the prediction model is founded on good data analysis.

3.4 Proposed Methodology

The recommended approach for predicting Dengue Fever using machine learning algorithms entails an ordered and thorough set of stages.

Flow chart:

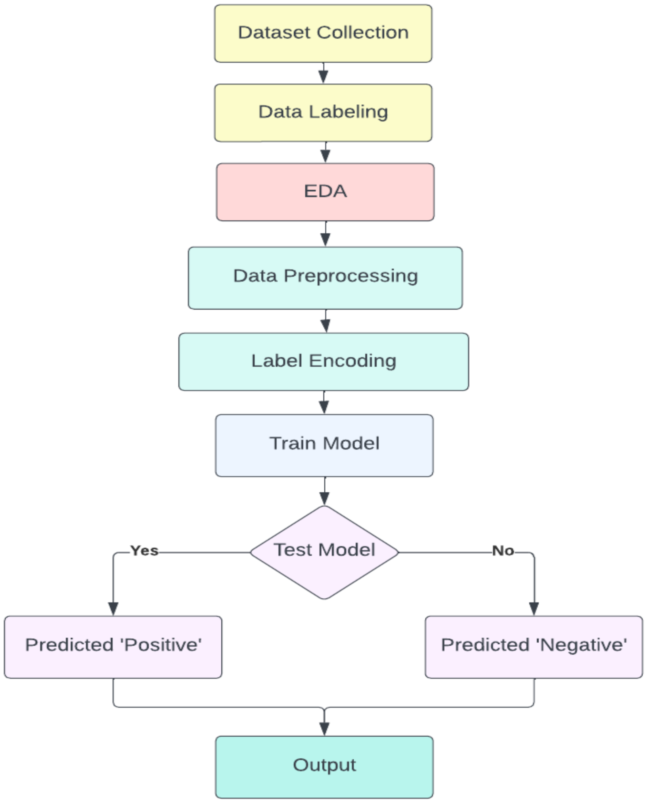

Figure 3.2: Methodology Flowchart

3.4.1 Data Collection

The method begins with the collection of a carefully selected dataset of 521 entries and 23 variables that include demographic information, blood-related metrics, and environmental variables. The dataset was created primarily to capture the complexities of Dengue Fever dynamics.

3.4.2 Data Preprocessing

The following stage is to address problems with data quality by dealing with missing values, resulting in a durable collection for research. The use of labels is also used to transform categorical variables to a numerical representation. Scaling changes, such as Min-Max growing, are used to normalize numerical features to improve model performance.

3.4.3 Feature Selection

Finding and choosing essential features to minimize complexity and improve understanding of models is a critical step. To select attributes important for Dengue Fever prediction, techniques such as correlational analysis and removal of recursive features are used.

3.4.4 Encoding

Individual variables having more than two categories are stored at once, making them easier to include in machine learning models.

3.4.5 Exploratory Data Analysis (EDA)

EDA involves presenting and analyzing a dataset using statistics and graphic methodologies. EDA findings guide the following steps in model building.

3.4.6 Model Training

Applying the previously collected dataset, machine learning methods such as 'BernoulliNB', 'GaussianNB', 'MLPClassifier', 'LogisticRegression','SVM 'are trained.

Support Vector Machine (SVM):

Support Vector Machine (SVM) is a strong supervised learning technique for classification and regression tasks. SVM works by identifying a hyperplane that best divides the data into various classes while maximizing the margin between them. It performs particularly well in high-dimensional spaces and can handle both linear and non-linear relationships using kernel methods. SVM attempts to discover the ideal decision boundary that increases margin while minimizing classification mistakes, making it flexible and applicable to a wide range of applications, including image recognition and text classification.

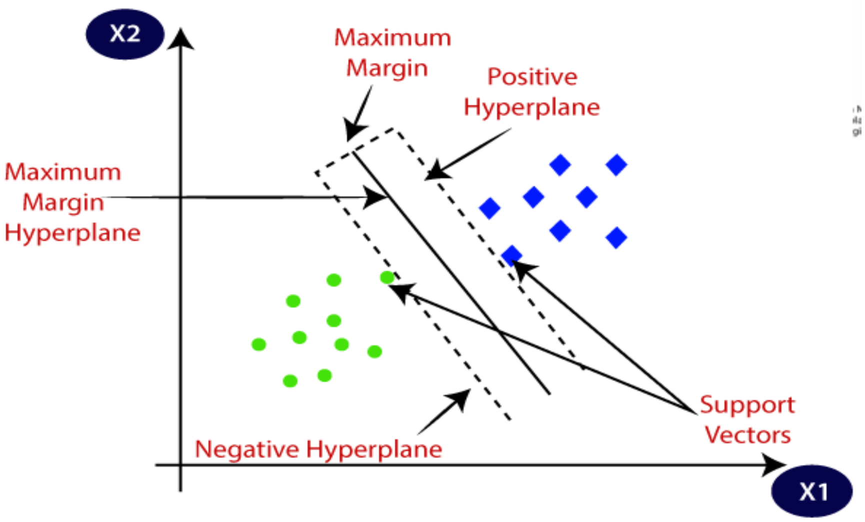

Figure 3.3: SVM model architecture

MLPClassifier (Multilayer Perceptron):

The Multilayer Perceptron (MLP) is an artificial neural network used for classification problems. It is made up of many layers of nodes (neurons) connected by weighted edges. MLPClassifier employs a feedforward design, in which information travels from the input layer through hidden levels to the output layer. MLPClassifier, which can capture complicated associations in data, is useful for handling complex problems like picture recognition and natural language processing. The model's parameters, such as the number of hidden layers and brain cells, can be adjusted to improve performance.

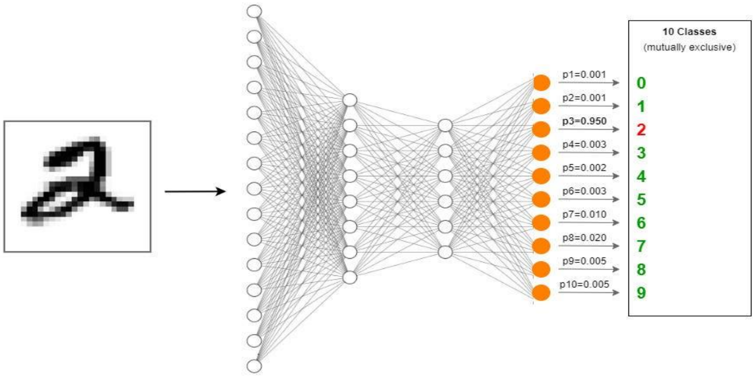

Figure 3.4: MLPClassifier model architecture

Logistic Regression:

Logistic regression is a linear model that solves binary classification problems. It uses the logistic (sigmoid) function to estimate the probability of an instance belonging to a specific class based on a linear combination of input data. Logistic regression is frequently used in many domains due to its simplicity, clarity, and efficiency. It is a fundamental algorithm in machine learning and is frequently used as a baseline model to compare more advanced approaches. Logistic Regression is best suited for instances where the connection between features and the target variable is nearly linear.

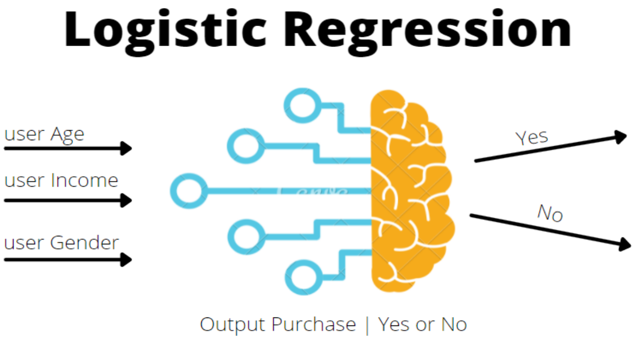

Figure 3.5: LR model architecture

Gaussian Naive Bayes (GaussianNB):

Gaussian Naive Bayes is a statistical classification algorithm that relies on Bayes' theorem and the concept of feature independence. It is especially beneficial for datasets that have continuous features. GaussianNB models the distribution of each class using Gaussian distributions and determines the probability of a data point belonging to a specific class. Despite its "naive" assumption of feature independence, GaussianNB is typically effective in practice and computationally economical. It is widely used in text classification and spam filtering, where it performs well with a minimal implementation.

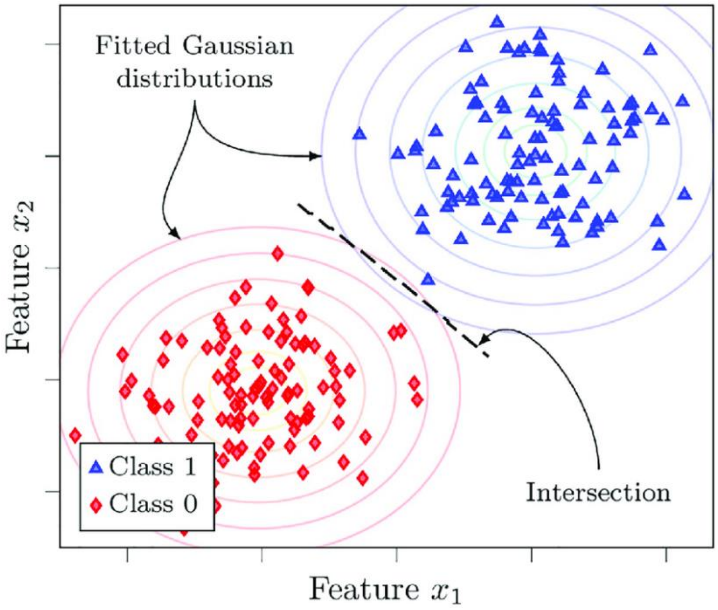

Figure 3.6: GaussianNB model architecture

Bernoulli Naive Bayes (BernoulliNB):

Bernoulli Naive Bayes is a Naive Bayes variation that is optimized for binary feature data in which features are either present or missing. It simulates the distribution of each class using Bernoulli distributions and determines the likelihood of an instance belonging to a specific class depending on the presence or absence of characteristics. BernoulliNB is often used in document classification tasks such as sentiment analysis, where the emphasis is on binary feature representations, making it appropriate for instances where the presence or absence of specific features is critical for classification.

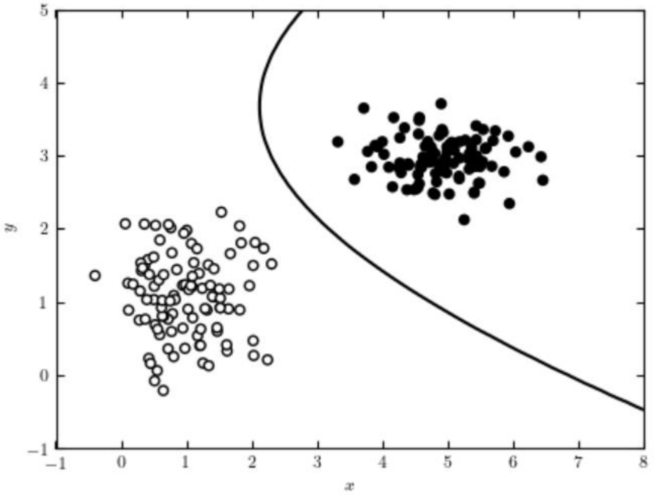

Figure 3.7: BernoulliNB model architecture

Model Evaluation

This model's accuracy is carefully evaluated using several classification measures, confirming the selection of the most successful Dengue Fever prediction method.

Test Model

The model's accuracy on a different set of tests is then evaluated, offering a true evaluation of its prediction skills on unknown information. The suggested method ensures a methodical and based-on-data method for Dengue Fever forecasting, which is consistent with best practices in machine learning model construction.

3.5 Implementation Requirements

The particular necessary are required for the effective execution of the Dengue Fever predicted model. A strong computer system with enough processing capacity is required to handle the complex structures of machine-learning algorithms. For model building and research, using programming languages such as Python and necessary libraries (such as scikit-learn, pandas, and matplotlib) is required.

The set of files, which has 521 entries and 23 attributes, is a prerequisite, underlining the need for precise and accurate information for training and testing. Communication with hospitals and experts also provides access to important topical knowledge, which enhances the model's knowledge of the environment.

A set of tools for single work and recording, such as Google Collaboratory, is suggested to promote single work and control of versions. Keeping the model's correctness over time requires constant access to current data for model adjustment and adjustment. The method of execution requires respect to ethical standards such as data privacy and informed consent. These conditions compose the facilities and resources required for the successful development and implementation of an effective Dengue Fever predicted model.

CHAPTER 4

Experimental results and discussion

4.1 Experimental Setup

An ordered collection of instruments and resources is required for the experimental setup for Dengue Fever projection. The technology system consists of a powerful system with the ability to successfully run machine learning algorithms. The use of computer languages like Python, as well as libraries like scikit-learn, pandas, and matplotlib, allows for smooth model building, training, and evaluation.

The dataset, which contains 521 entries and 23 attributes, is at the heart of the experimental setup. Access to relevant healthcare expertise, which may be obtained through relationships with people and places, improves knowledge about Dengue dynamics. Git version control systems enable working together, whereas Jupyter Notebooks, on the other hand, enable visible description and analysis.

The laboratory setting requires constant access to current data for model adjustment and modification. The quality of the research is maintained by keeping to ethical criteria, including those concerning data protection and thoughtful permission. This extensive setup is intended to verify the Dengue Fever prediction experiment's dependability, validity, and ethical behavior.

4.2 Experimental Results & Analysis

The research results show that the selected machine learning methods for Dengue Fever detection using the available dataset have varied degrees of accuracy.

The recall, accuracy, and F1 Score measures confirm the system's efficacy by displaying high levels of accuracy, sensitivity, and overall model performance. The experimental results and following analysis lead to a deeper recognition of every model's strengths and limits, allowing for more educated system optimization and model selection decisions based on unique requirements and concerns.

Accuracy: Accuracy measures the overall correctness of the model's predictions by comparing the number of correctly classified samples to the total number of samples. When classes are unbalanced, it gives a broad indication of the model's effectiveness but might not give a whole picture.

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

Precision: Out of all positive predictions generated by the model, precision focuses on the percentage of true positive forecasts.

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

Recall: Also known as sensitivity or true positive rate, recall is the percentage of true positive predictions made out of all truly positive samples.

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

F1 rating: The F1 score is the harmonic mean of recall and precision. It provides a reasonable evaluation metric that considers recall and precision. The F1 score is useful when classes are uneven since it accounts for both false positives and false negatives. A high F1 score denotes a well-balanced precision to recall ratio.

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

The result of the deep learning model is compared based on Accuracy, Precision, Recall, and F1 Score in Table 4.1:

Table 4.1. Performance Evaluation

Model Name	Accuracy	Precision	Recall	F1-Score
SVM	98.88%	0.989076	0.989011	0.989014
MLPClassifier	97.78%	0.978092	0.978022	0.978027
LogisticRegression	96.95%	0.971003	0.969780	0.969799
GaussianNB	93.64 %	0.937580	0.936813	0.936850
BernoulliNB	67.85%	0.678222	0.678571	0.678145

Support Vector Machine's excellent accuracy underscores its potential for this predicting job, possibly due to its capacity to manage non-linear correlations and capture feature importance. The success of MLPClassifier emphasizes the importance of studying the shape of features in the dataset. The accuracy ratings of MLPClassifier and SVM were 96.95% and 98.89%, respectively, suggesting their competence in Dengue Fever detection. BernoulliNB, on the other hand, had the lowest accuracy at 67.85%, indicating limits in dealing with the dataset's peculiarities. LogisticRegression and GaussianNB perform admirably, demonstrating how they adapt to the features of the dataset. BernoulliNB's lesser accuracy emphasizes the need for more complicated models capable of addressing the dataset's complexity. These findings highlight the significance of algorithm selection and offer suggestions for improving the Dengue Fever prediction model.

Performance Analysis

SVM:

Table 4.2. Performance Evaluation(SVM)

	Precision	Recall	F1-Score	Support
0	0.98	0.99	0.99	173
1		0.98	0.99	191
Accuracy	0.98		0.99	364
Macro avg	0.99	0.99	0.99	364
Weighted avg	0.99	0.99	0.99	364

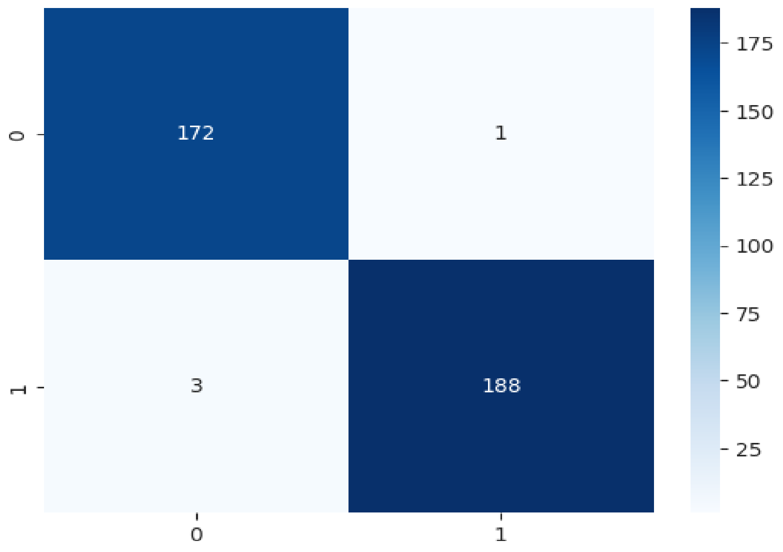

Figure 4.1: Confusion Matrix (SVM)

In Figure 4.1 the confusion matrix of SVM er shows It correctly identified 172 positive cases (true positives) and 188 negative cases (true negatives). It incorrectly classified only 1 positive case as negative (false positive) and 3 negative cases as positive (false negative). This suggests that the SVC model performs well in accurately distinguishing between positive and negative examples.

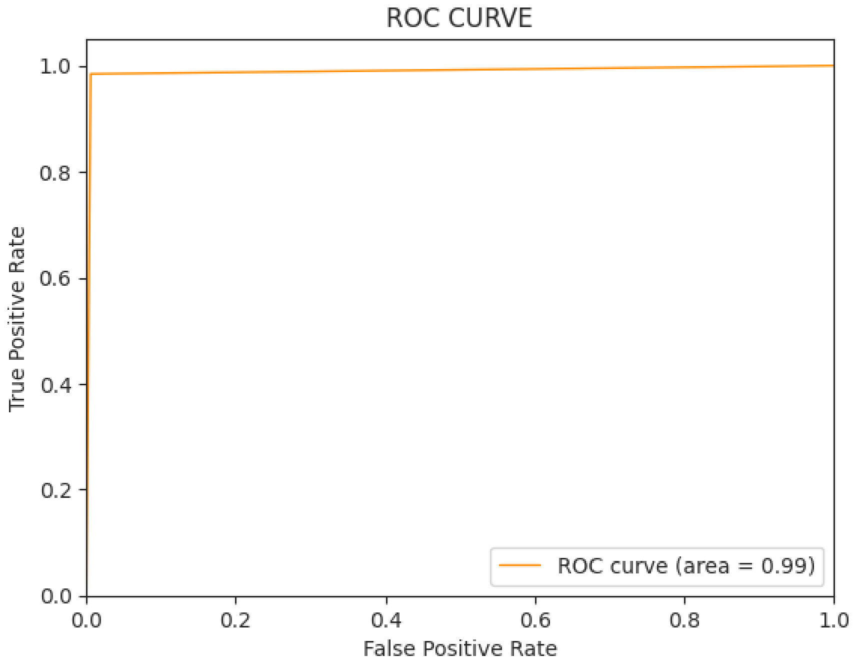

Figure 4.2: AUC ROC Curve (SVM)

Figure 4.2 presents an AUC ROC curve for an SVM model, employed for binary classification. This curve, tracing the trade-off between true and false positive rates, suggests a relatively strong model performance with an AUC value of 0.99. While this indicates a commendable ability to distinguish classes, the non-perfect curve implies cases of both false positives and negatives. To optimize performance, comparative analysis with other models or SVM hyperparameter tuning could be explored.

MLPClassifier:

Table 4.3. Performance Evaluation (MLPClassifier)

	Precision	Recall	F1-Score	Support
0	0.97	0.98	0.98	173
1	0.98	0.97	0.98	191
Accuracy	0.98	0.98	0.98	364
Macro avg	0.98	0.98	0.98	364
Weighted avg	0.98	0.98	0.98	364

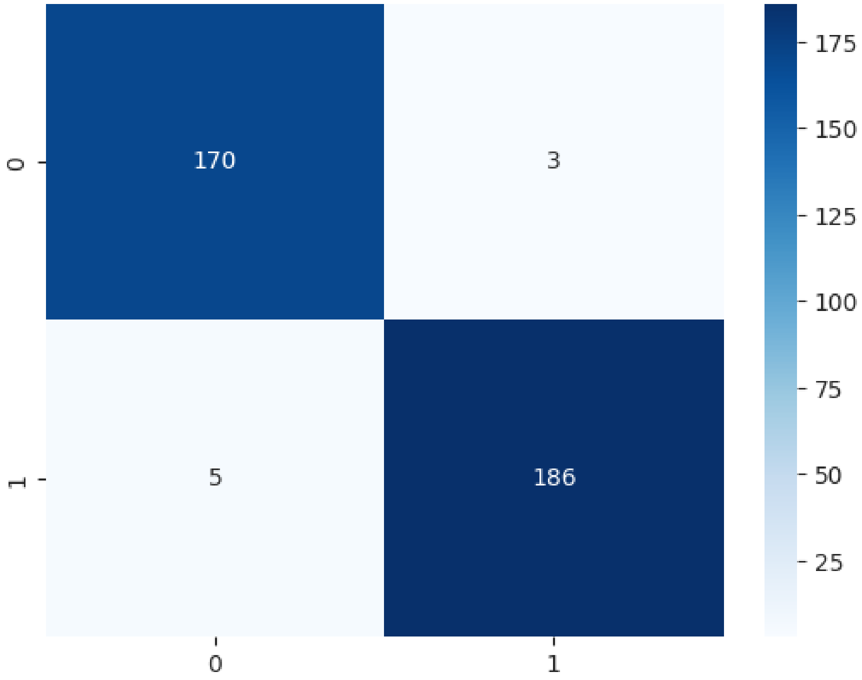

Figure 4.3: Confusion Matrix (MLPClassifier)

In Figure 4.3 the confusion matrix of MLPClassifier shows 170 true positives, meaning 170 cases were correctly identified as positive. There are only 3 false positives, indicating very few incorrect positive predictions. The matrix also shows 186 true negatives, meaning 186 cases were correctly identified as negative. This demonstrates the model's ability to accurately distinguish between positive and negative examples.

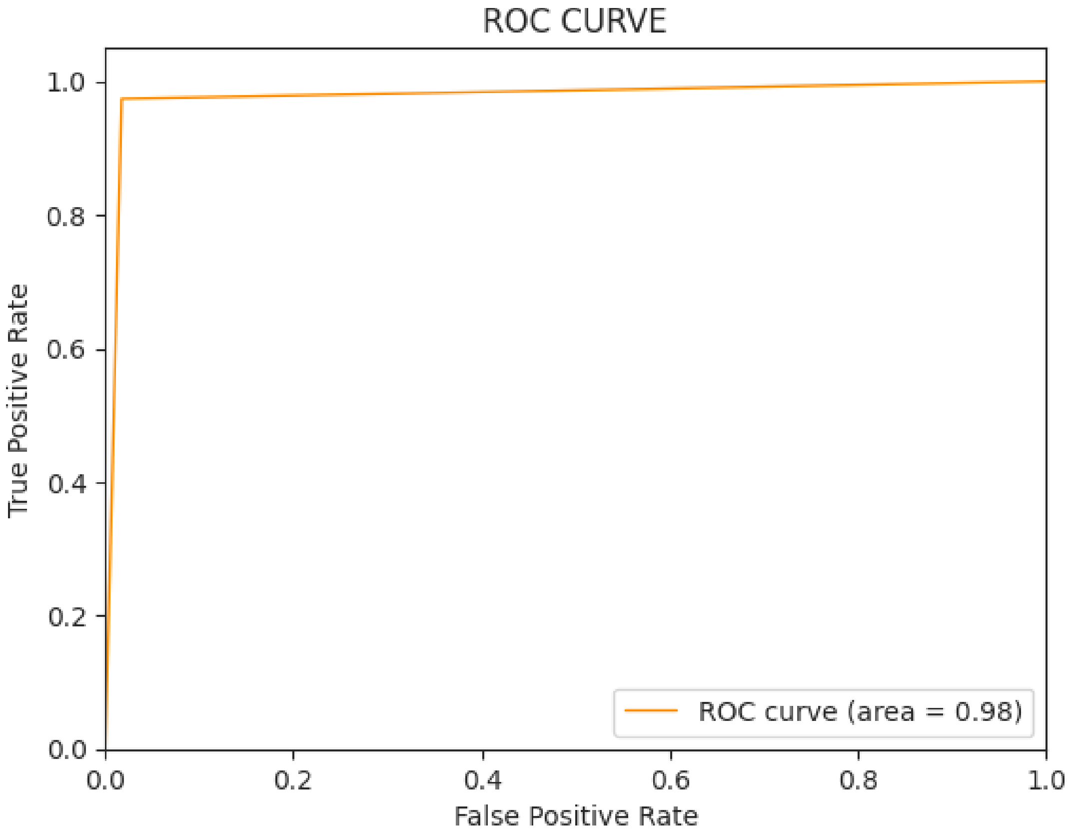

Figure 4.4: AUC ROC Curve (MLPClassifier)

The ROC curve for the MLPClassifier has an area of 0.98, indicating excellent performance. It demonstrates a high true positive rate, correctly identifying most positive cases, while maintaining a low false positive rate, minimizing incorrect positive identifications. This balance reflects the classifier's ability to effectively distinguish between true and false positives.

LogisticRegression:

Table 4.4. Performance Evaluation (LogisticRegression)

	Precision	Recall	F1-Score	Support
0	0.95	0.99	0.97	173
1	0.99	0.95	0.97	191
Accuracy			0.97	364
Macro avg	0.97	0.97	0.97	364
Weighted avg	0.97	0.97	0.97	364

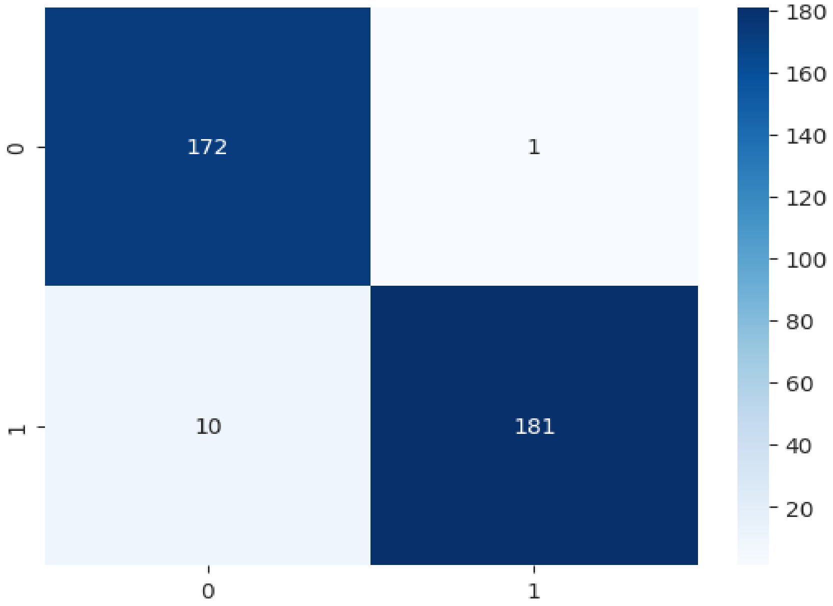

Figure 4.5: Confusion Matrix (LogisticRegression)

In Figure 4.5 the confusion matrix of Logistic Regression shows the model correctly classified most cases (353 out of 364). It has a high accuracy of approximately 97% (353 correct predictions out of 364 total cases). There's a slight tendency to overpredict positive cases, as evidenced by the 10 false positives. However, it rarely misses positive cases, as indicated by the low number of false negatives (1).

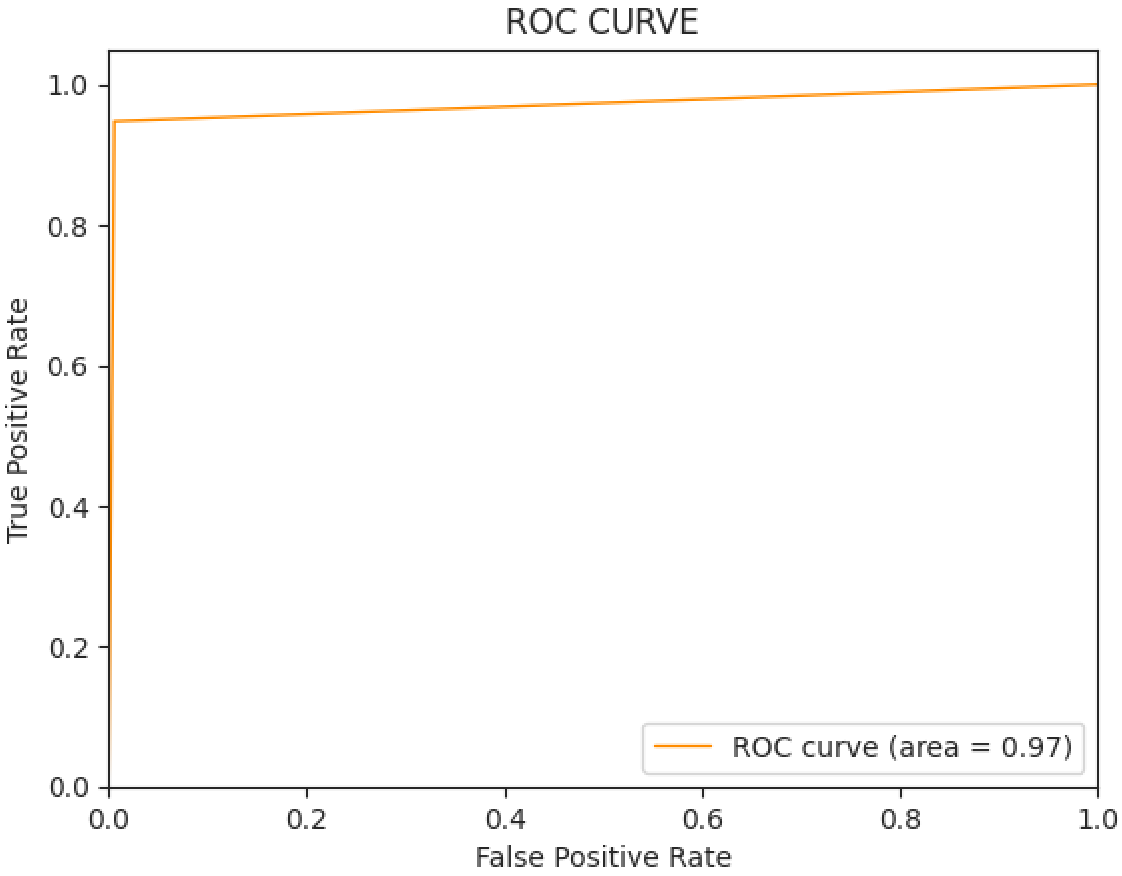

Figure 4.6: AUC ROC Curve (LogisticRegression)

The ROC curve for the Logistic Regression model has an area of 0.97, indicating excellent performance. It demonstrates a high true positive rate, correctly identifying most positive cases, while maintaining a low false positive rate, minimizing incorrect positive identifications. This balance reflects the classifier’s ability to effectively distinguish between true and false positives.

GaussianNB:

Table 4.5. Performance Evaluation (GaussianNB)

	Precision	Recall	F1-Score	Support
0	0.92	0.95	0.93	173
1	0.96	0.92	0.94	191
Accuracy				364
Macro avg	0.94	0.94	0.94	364
Weighted avg	0.94	0.94	0.94	364

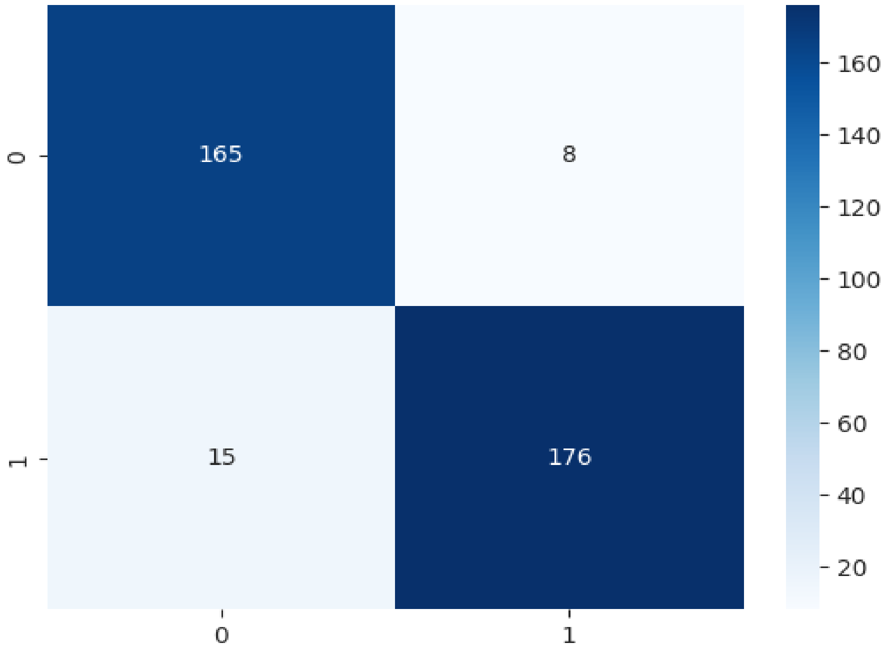

Figure 4.7: Confusion Matrix (GaussianNB)

In Figure 4.7 the confusion matrix of GNB shows,165 cases were correctly identified as positive. 8 cases were incorrectly classified as positive. 176 cases were correctly identified as negative.15 cases were incorrectly classified as negative. It shows a slight tendency to underpredict positive cases, as there are more false negatives than false positives. Overall, the model performs well but could be improved in terms of identifying positive cases more accurately.

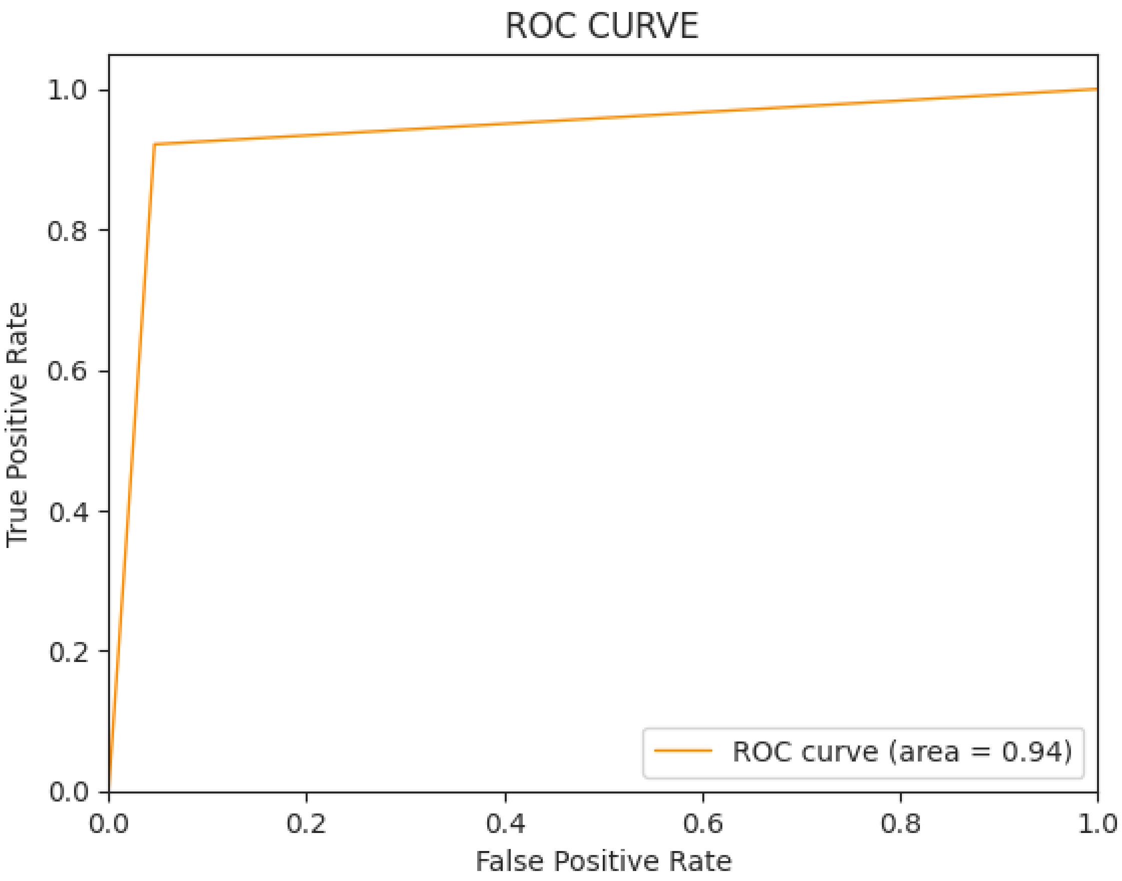

Figure 4.8: AUC ROC Curve (GaussianNB)

The ROC curve of GNB in the image has an area under the curve (AUC) of 0.94, which is considered good performance. This means that the model is able to correctly distinguish between positive and negative cases 94% of the time.

BernoulliNB:

Table 4.6. Performance Evaluation (BernoulliNB)

	Precision	Recall	F1-Score	Support
0	0.67	0.64	0.65	173
1	0.69	0.71	0.70	191
Accuracy			0.68	364
Macro avg	0.68	0.68	0.68	364
Weighted avg	0.68	0.68	0.68	364

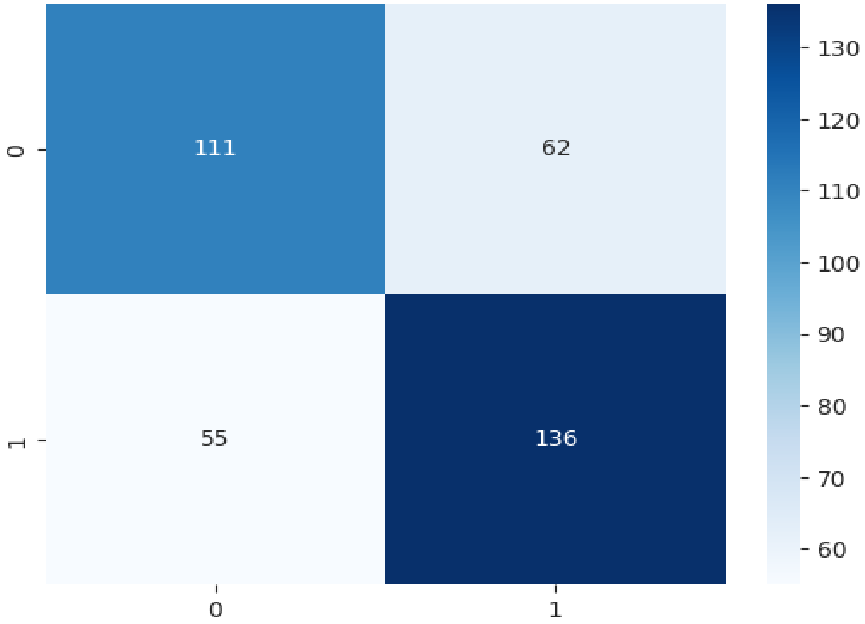

Figure 4.9: Confusion Matrix (BernoulliNB)

Figure 4.9 shows the confusion matrix of BernoulliNB It shows True positives (111) and true negatives (136), indicating good overall accuracy. False positives (62) are higher than false negatives (55), suggesting a slight tendency to over-predict. The model correctly classified 247 out of 354 cases, achieving an accuracy of about 68%.

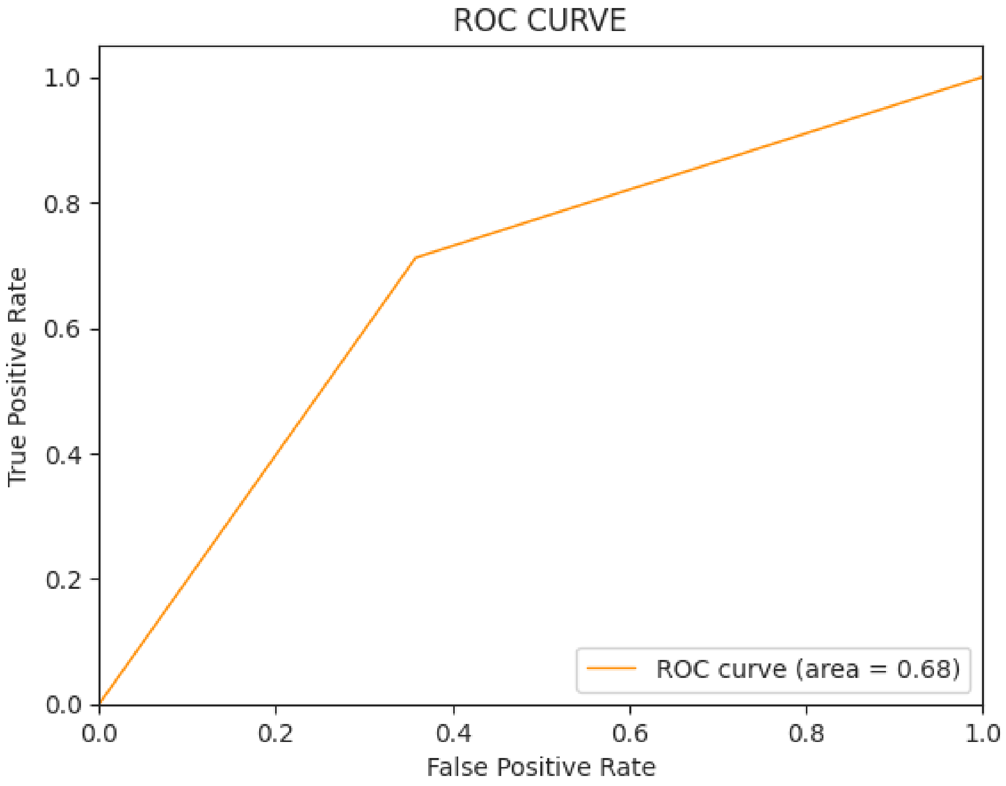

Figure 4.10: AUC ROC Curve (BernoulliNB)

Figure 4.10 shows the ROC curve for BernoulliNB has an area under the curve (AUC) of 0.68, which is considered to be moderate performance. This means the model is able to correctly distinguish between positive and negative cases 68% of the time.The curve starts off in the bottom left corner and rises steadily, which is good. However, it doesn't reach the top left corner and plateaus instead. This indicates that the model struggles to correctly identify positive cases as the threshold increases.Overall, the ROC curve suggests that BernoulliNB might not be the best choice for this classification task, as other models in the previous bar chart achieved higher AUC scores.

4.3 Accuracy

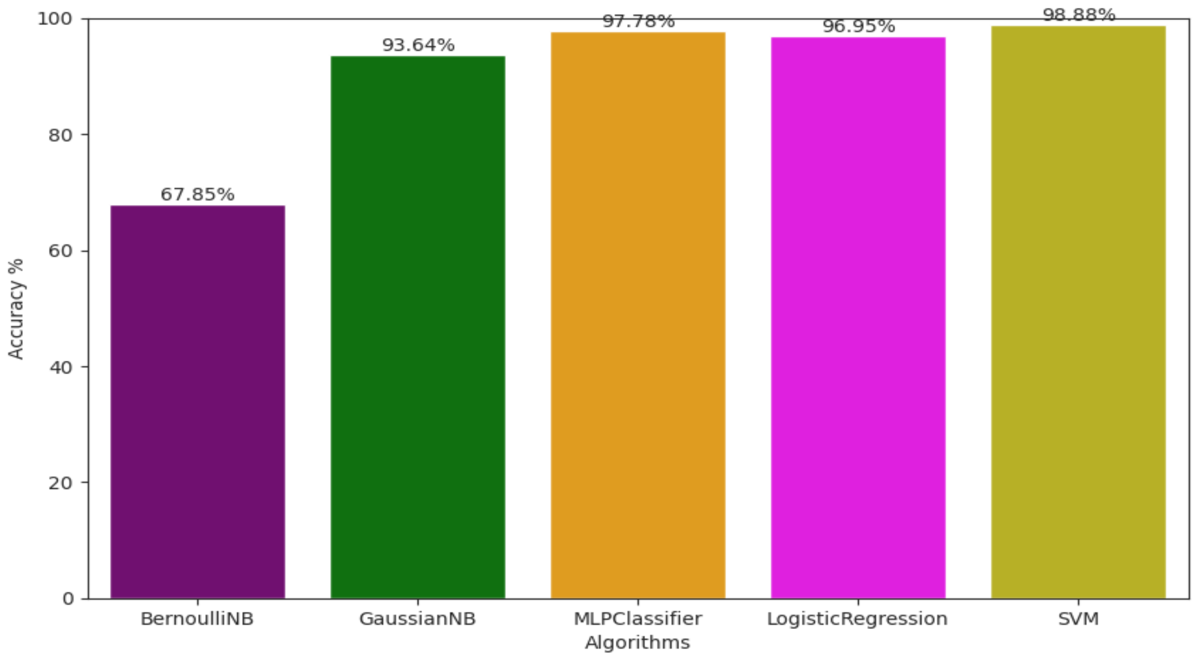

Figure 4.11: Comparative Model Accuracy Bar Plot

The bar chart shows how accurate five different machine learning models are at predicting dengue fever. The best model is SVM, with an accuracy of 98.88%. This means that it correctly predicted almost 99 out of every 100 dengue fever cases. The next best model is MLPClassifier, with an accuracy of 97.78%. LogisticRegression is close behind at 96.95%. GaussianNB is less accurate, at 93.64%, and BernoulliNB is the least accurate, at only 67.85%. Overall, the bar chart shows that SVM and MLPClassifier are the most accurate models for predicting dengue fever.

4.4 Discussion

The results of a study of the machine learning algorithms used to detect Dengue Fever show different strengths and areas for development. SVM outperforms the competition with a stunning accuracy of 98.89%, demonstrating its ability to recognize complex patterns within the dataset. MLPClassifier follows suit with an accuracy of 97.8 %, displaying ability in dealing with continuous features. In addition to accuracy levels 96.96% of and 93.64%, respectively, LogisticRegression and GaussianNB perform admirably, demonstrating their adaptability to the dataset's intricacies. BernoulliNB, on the other hand, falls behind with an accuracy of 67.85%, indicating problems in dealing with the dataset's peculiarities, possibly due to its suitability for linear features.The range of accuracy levels highlights the necessity of algorithm selection that is suited to the complexities of the dataset. SVM emerges as a strong contender, highlighting the importance of teamwork in capturing complicated interactions. These performance measures give useful information for improving the Dengue Fever prediction model, directing additional tuning and potential collective techniques for improved accuracy and universality.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

This breakthrough has a dramatic influence on society, giving an important tool for public health management in the face of Dengue Fever episodes. The proposed prediction model, powered by machine learning algorithms, has the potential to transform how health authorities plan for and react to disease spread. The model's accuracy allows for the overriding allocation of resources, specific actions, and public education campaigns, thereby minimizing the burden of Dengue Fever on suffering communities.

In addition to direct health outcomes, the societal impact influences policy formation and allocation of resources for public health efforts. Improved response and readiness measures help communities cope with the potential of Dengue epidemics. This study shows the good connection between advanced technologies and public health, encouraging a more educated and responsive societal method for disease control.

5.2 Impact on Environment

The environmental effect of this research is indirect yet considerable, as it contributes to the ethical handling of Dengue Fever, influencing broader environmental variables. The approach supports aggressive actions for public health by employing machine learning for precise prediction, ultimately minimizing the need for reactive responses that may have environmental consequences. Rapid action can reduce the need for chemical vector control measures while encouraging environmentally acceptable alternatives.

Meanwhile, the model's capacity to advise specific actions can help preserve environments by reducing uncontrolled consumption of resources. Because Dengue Fever is so closely linked to climatic and environmental conditions, the study highlights the necessity of studying and minimizing the influence of viral illnesses on the natural environment. This research therefore contributes to establishing a healthier and more balanced interaction between human activities and the environment through the lens of avoiding illnesses and environmentally friendly methods.

5.3 Ethical Aspects

According to research and deployment of the Dengue Fever predicted model, moral decisions are crucial. Data privacy and security are critical, requiring stringent processes for handling sensitive health information. Transparency and respect for those who created the dataset require informed consent processes.

Data collection and representation must be equal to avoid disadvantage and ensure fair representation of varied groups. The appropriate application of machine learning algorithms entails avoiding unfairness and reducing unforeseen repercussions, as well as encouraging trust in the model's conclusions.

Transparency interaction between consumers, such as physicians and the general public, respects moral standards and allows for informed choice-making. Regular evaluation and revision of moral issues during the research process is critical for matching the project with moral values and fostering responsible and ethical use of the Dengue Fever modeling system.

5.4 Sustainability Plan

The environmental sustainability strategy for this Dengue Fever prediction research includes several components to ensure the project's durability and beneficial impact. First and foremost, the prediction model will be continuously monitored and updated, including new data and improving algorithms to improve accuracy and usefulness.

In progress interaction between medical organizations and community partners will be encouraged, boosting information exchange and ensuring that the model remains relevant to changing public health requirements. Additionally, attempts will be made to link the model with current public health facilities, allowing professionals to use it easily.

This dedication to open-access publication of study results and model outcomes would increase understanding and practicality, helping worldwide efforts to eliminate Dengue Fever. Throughout the project's lifecycle, moral principles and appropriate data practices will be respected, assuring the continued trust of both producers and end-users. This overall conservation plan strives to maximize the long-term positive effects on society and the environment of the Dengue Fever projected model.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In general, the current research focuses on predicting Dengue Fever utilizing machine learning methods using a dataset of 521 entries with 23 variables. The efficiency of five methods was looked at BernoulliNB, GaussianNB,MLPClassifier, LogisticRegression, and SVM. The best model is SVM, with an accuracy of 98.88%, demonstrating its applicability to continuous information. MLPClassifier performed well, with an accuracy of 97.78%, In addition to accuracy levels 96.96% of and 93.64%, respectively, LogisticRegression and GaussianNB performed admirably, demonstrating their adaptability to the dataset's intricacies. BernoulliNB, on the other hand, had a lower accuracy of 67.85%, demonstrating difficulties in dealing with the dataset's features. These results highlight the significance of algorithm choice in Dengue Fever detection and offer suggestions for model enhancement. SVM, in particular, emerges as a powerful tool, leading to further improvements and underlining the value of ensemble approaches in improving forecast accuracy for illness management.

6.2 Conclusions

At last, the research on Dengue Fever prediction using machine learning algorithms showed helpful findings regarding the efficacy of various models. SVM achieved remarkable accuracy, demonstrating its ability to detect complicated patterns in the dataset. MLPClassifier excelled at handling continuous features, while LogisticRegression and GaussianNB showed excellent versatility. However, BernoulliNB had limits in effectively predicting Dengue Fever, highlighting potential issues with binary feature appropriateness.

The results highlight the importance of algorithm selection and highlight the complex nature of the dengue fever data. This research adds to the field by directing both professionals and scholars to the most appropriate models for reliable and exact forecasts. Further development and investigation of aggregation approaches, notably utilizing the characteristics of SVM, offer interesting paths for improving Dengue Fever model prediction capabilities. Overall, the findings of this study lay the groundwork for future research aiming at increasing the precision and durability of machine learning applications in illness detection and management.

6.3 Implication for Further Study

This research on Dengue Fever detection using machine learning algorithms opens up new areas for investigation and improvement. The outstanding results of SVM show the capacity to capture subtle patterns in Dengue Fever data. Future research could focus on improving ensemble techniques and analyzing how they apply across different datasets.

Studying switch machine learning architectures and advanced algorithms may overcome the constraints discovered in BernoulliNB, allowing for a more complete understanding of model applicability for Dengue Fever prediction. Incorporating real-time data streams and dynamic features could also improve the model's flexibility to shift illness dynamics.

In addition, the effects of model projections on public health measures and resource allocation merit further examination. Studying the financial and environmental elements that influence the spread of dengue and combining them into predictive models could improve our understanding and future forecasts.

In general, the investigation establishes the groundwork for future research in the field, promoting the use of sophisticated methodology, diversified datasets, and cross-disciplinary partnerships to improve the efficacy and usability of machine learning in Dengue Fever predictions and control of public health.

REFERENCE

[1] Rana, Sandeep Kumar, et al. "Dengue fever prediction using machine learning analytics." 2022 International conference on machine learning, big data, cloud and parallel computing (COM-IT-CON). Vol. 1. IEEE, 2022.

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

[3] Anitha, A., and DC Joy Winnie Wise. "Forecasting dengue fever using classification techniques in data mining." 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2018.

[4] Al Nasar, Mohammad Rustom, et al. "Detection of Dengue Disease Empowered with Fused Machine Learning." 2022 International Conference on Cyber Resilience (ICCR). IEEE, 2022.

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

[6] Majeed, Mokhalad A., et al. "A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention." International journal of environmental research and public health 20.5 (2023): 4130.

[7] Zargari Marandi, Ramtin, et al. "Development of a machine learning model for early prediction of plasma leakage in suspected dengue patients." PLOS Neglected Tropical Diseases 17.3 (2023): e0010758.

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

[9] Yang, Hsiu, et al. "An Integrative Explainable Artificial Intelligence Approach to Analyze Fine-Scale Land-Cover and Land-Use Factors Associated with Spatial Distributions of Place of Residence of Reported Dengue Cases." Tropical Medicine and Infectious Disease 8.4 (2023): 238.

[10] Indhumathi, K. "An Efficient Framework For Predicting And Forecasting Of Infectious Diseases Using Machine Learning Techniques In Big Data Environment." (2023).

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

[12] Rustom Al Nasar, Mohammad, et al. "Detection of Dengue Disease Empowered with Fused Machine Learning." 2022 International Conference on Cyber Resilience (ICCR). 2023.

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

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

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

[16] Iqbal, Naiyar, et al. "Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers." Informatica 43.3 (2019).

[17] Sajana, T., et al. "Classification of dengue using machine learning techniques." Int J Eng Technol 7.2.32 (2018): 212-218.

[18] Rahman, Md Siddikur, et al. "Mapping the spatial distribution of the dengue vector Aedes aegypti and predicting its abundance in northeastern Thailand using machine-learning approach." One Health 13 (2021): 100358.

[19] Sanjudevi, D., et al. "Dengue fever prediction using classification techniques." Int Res J Eng Technol (IRJET) 6.02 (2019): 558-563.

[20] Macedo Hair, Gleicy, et al. "Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach." BMC infectious diseases 19 (2019): 1-11.

[21] Ming, Damien K., et al. "Applied machine learning for the risk-stratification and clinical decision support of hospitalised patients with dengue in Vietnam." PLOS digital health 1.1 (2022): e0000005.

[22] Abdualgalil, Bilal, et al. "Early Diagnosis for Dengue Disease Prediction Using Efficient Machine Learning Techniques Based on Clinical Data." Journal of Robotics and Control (JRC) 3.3 (2022): 257-268.

[23] Kaur, Supreet, et al. "Predicting Infection Positivity, Risk Estimation, and Disease Prognosis in Dengue Infected Patients by ML Expert System." Sustainability 14.20 (2022): 13490.

[24] Gomes, Ana Lisa V., et al. "Classification of dengue fever patients based on gene expression data using support vector machines." PloS one 5.6 (2010): e11267.

DENGUE FEVER PREDICTION USING MACHINE LEARNING APPROACHES

ORIGINALITY REPORT

14%	10%	8%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	sd.blackball.lv Internet Source	<1%
4	Submitted to Grambling State University Student Paper	<1%
5	Submitted to Sheffield Hallam University Student Paper	<1%
6	github.com Internet Source	<1%
7	libweb.kpfu.ru Internet Source	<1%
8	repository.cuilahore.edu.pk Internet Source	<1%
9	Archana Thirugnanam, Faritha Banu Jahir Hussain. "Exploring Machine Learning	<1%