# A MACHINE LEARNING APPROACH FOR PREDICTIVE-PROBABILITY ANALYSIS AND PREVENTIVE STRATEGIES IN DENGUE FEVER


**BY**

**Abdul Fattah Amid**
**ID: 202-15-3808**
**AND**

**Ashraful Islam Jahid**
**ID: 202-15-3814**


## 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

**Israt Jahan**
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

## Co-Supervised By

**Nushrat Jahan Ria**
Lecturer
Department of Computer Science and Engineering
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**July 2024**

## APPROVAL

This Project titled "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever", submitted by **Abdul Fattah Amid** and **Ashraful Islam Jahid** 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 **15-07-2024.**

## BOARD OF EXAMINERS

**Dr. Md. Ismail Jabiullah(MIJ)** — **Board Chairman**
**Professor**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Mr. Md Assaduzzaman (MA)** — **Internal Examiner 1**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Ms. Sharun Akter Khushbu (SAK)** — **Internal Examiner 2**
**Sr. Lecturer**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

**Dr. Ahmed Wasif Reza (DWR)** — **External Examiner**
**Professor**
Department of CSE
East West University

# DECLARATION

We hereby declare that this project has been done by us under the supervision of **Israt Jahan, Lecturer (Senior Scale), Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

**Supervised by:**

14.7.24

**Israt Jahan**
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

**Co-Supervised by:**

for, 14.7.24

**Nushrat Jahan Ria**
Lecturer
Department of CSE
Daffodil International University

**Submitted by:**

Amid 14-07-24

**Abdul Fattah Amid**
ID: 202-15-3808
Department of CSE
Daffodil International University

Jahid 14-07-2024

**Ashraful Islam Jahid**
ID: 202-15-3814
Department of CSE
Daffodil International University



*©Daffodil International University, Bangladesh* iii

# ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Israt Jahan**, **Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever" 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 **Dr. Sheak Rashed Haider Noori, Professor & Head, Department of Computer Science and Engineering, Daffodil International University**, 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

Dengue fever, a common mosquito-borne viral infection, often goes undiagnosed until severe symptoms appear. Early detection is important for effective management and prevention of complications. This study explores the potential of machine learning (ML) and artificial intelligence (AI) to develop predictive models for dengue fever and severity assessment. This study used secondary datasets including patient data including medical history, environmental factors, symptoms and possible genetic predisposition to train various machine learning algorithms such as CNN, Linear Regression, Random Forest and K-Nearest Neighbors (KNN). This method involves data collection, preprocessing, feature selection, model training, cross-validation, and hyperparameter tuning to optimize performance and avoid overfitting. Models are evaluated based on metrics such as accuracy and AUC-ROC, including attribute importance scores and shape quality analysis. The dataset consisted of 531 patient records with 20 attributes each, which ensured data quality through preprocessing steps such as handling missing values and encoding categorical variables. A high-performance computer, cloud resources, powerful hardware and software including Python, TensorFlow, PyTorch and Skit-Learn are essential for implementation. Effective project management and financial analysis ensure research success and sustainability. Studies have shown that linear regression and random forest models achieve the highest accuracy (99%), followed by CNN (97%) and KNN (54%). These findings underscore the potential of machine learning to enhance dengue prediction, improve patient outcomes, and inform public health strategies.

# TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

## 1.1 Overview

Dengue fever, a mosquito-borne viral infection, poses a significant public health challenge, affecting millions of individuals worldwide and leading to serious health complications and death. Traditional methods of diagnosing and predicting dengue fever rely heavily on symptom assessment and laboratory tests, which can be time-consuming and often result in delayed intervention. With advances in technology, machine learning offers a promising alternative by providing accurate, non-invasive prediction models based on large datasets comprising patient demographics, medical history, environmental factors and other relevant parameters. This paper explores the development and application of machine learning algorithms to predict the probability of dengue fever, with the aim of improving early detection and personalized treatment plans. Healthcare providers can improve patient outcomes, optimize resource allocation, and reduce the overall burden of this widespread disease. This study addresses the challenges associated with dengue fever.

## 1.2 Background and Present State

To facilitate a clear understanding of the key concepts and methods employed in this paper on “A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever”, it is influential to lay down a terminology and foundation. It is essential to establish a basic understanding of key concepts. First, dengue fever, a mosquito-borne viral infection, is interpreted as an important public health challenge, leading to serious health complications and death. Next, the symptoms and risk factors associated with dengue fever are outlined, highlighting the importance of identifying at-risk individuals. The section introduces central machine learning concepts, including predictive modeling, feature selection, and evaluation metrics. Understanding these machine learning values is important for developing accurate prediction models Moreover, the concept of cross-validation is introduced as a method to measure model performance and ensure generalizability. By clarifying these conditions, the foundation is laid for exploring the application of machine learning to predict dengue fever, play an important role in healthcare and preventive medicine. A thorough analysis of related work and previous research is provided in the background chapter. Abstract This section provides a general understanding of keystone, technological developments and methods related to dengue fever detection and machine learning. It provides background information for the present study and identifies data gaps.

### 1.3 Problem Statement

Nowadays, dengue fever is a very common disease. Lately not only adults but children and teenagers are also affected by this disease. The spread of dengue is exacerbated by urbanization, increased travel and climate change, which create favorable conditions for the Aedes mosquito that transmits the virus. Clinical practice currently lacks robust predictive tools to identify patients with dengue fever. The lack of a systematic approach to identifying factors and patterns predicting reactive outcomes rather than proactive management places a significant burden on patients and healthcare systems. Furthermore, it is difficult to develop a reliable predictive framework due to the multifactorial nature of dengue fever, including environmental factors, human dynamics, and genetic variation of the virus. Emphasizes the urgent need for a comprehensive predictive model to enable personalized preventive strategies. Closing this gap in predictive analytics for dengue fever is critical to transforming healthcare delivery, optimally allocating resources, and dramatically improving patient outcomes. Thus, this study aims to develop a reliable prediction model that integrates environmental and demographic data to help reduce the risk of dengue fever.

### 1.4 Objectives

In recent years, the incidence of dengue fever has increased worldwide, with alarming outbreaks occurring in various regions, including our country. The increasing frequency of dengue cases emphasizes the urgent need for proactive measures to address this public health challenge. Communities are witnessing the true effects of dengue fever, with increasing numbers of people succumbing to the debilitating disease as hospitals and health care facilities are flooded with dengue patients, many of whom are in critical condition, highlighting the severity of the situation. In this context, increasing public awareness and undertaking massive efforts to prevent the spread of dengue is a critical imperative. By focusing on the latest trends and using a data-driven approach, our initiative seeks to mitigate the impact of dengue outbreaks and empower communities with the knowledge and resources needed to protect public health. Through targeted interventions and education campaigns, we will aim to stem the tide of dengue transmission and build a culture of active prevention. Presented below are a few illustrations of potential for our paper:

- Analysis of Data and Identification of Patterns.
- Decision Support Systems Focused on Patients.
- Insights Derived from Data for Healthcare Policies.
- Increasing Awareness of Public Health.
- Strengthening Strategies for Preventive Healthcare.
- To explore the people to know about dengue disease.
- To know the predictive probability rate of people getting dengue disease.
- To understand what preventive measures can be taken in future.

Each of these goals seeks to maximize the potential of machine learning to provide further insight into dengue disease management to improve treatment outcomes, optimize preventive measures. The overall goal of this research is to use machine learning for predictive analytics in healthcare with a special focus on dengue fever and how to prevent or reduce the impact of this problem.

### 1.5 Scope and Limitations

This research aims to develop and validate machine learning models to predict dengue fever, using comprehensive datasets that include patient demographics, medical history, environmental factors and other contextual variables. The primary objective is to develop a predictive tool. Scope of study includes selection of appropriate machine learning algorithms, pre-processing of data, model training and validation, and evaluation of model performance against established benchmarks.

Several limitations must be acknowledged. The accuracy and reliability of predictive models is linked to the quality an;d completeness of the data used internally. Incomplete, biased, or unrepresentative datasets can lead to flawed predictions and reduced generalizations. Additionally, the applicability of these models to different patient populations. Ethical considerations, particularly regarding patient privacy and data security, are important challenges that must be addressed to maintain trust and compliance with regulatory standards. Furthermore, successful implementation of these models in real-world settings requires widespread validation, acceptance, and adoption by the medical community, which can be a time-consuming and complex process. Despite these challenges, the potential for machine learning to transform dengue prognosis and improve patient outcomes underscores the importance of continued research and development in this field.

## 1.6 Report Organization

Through methodology, conclusions, and study results, the proposed report has a thorough layout, taking readers through. Each chapter has a distinct function and adds overall unity and depth to the document.

**Chapter 1: Introduction**

The study is paced by an introductory chapter that provides background information on the importance of dengue detection and the need for a predictive-probability analysis and the use of machine learning. It provides an overview of the paper's area, research question and general framework.

**Chapter 2: Literature Review**

A thorough analysis of related work and previous research is provided in the background chapter. Abstract A general understanding of keystones, technological developments, and methods related to dengue prediction and machine learning is provided in this section. It provides background information for the present study and finds information gaps.

Chapter 3: Methodology/ Requirement Analysis & Design Specification

The methodology used to conduct the research is described in the Research Methods chapter. It focuses on the rationale behind the design of models, data collection systems, study design and specific collection methods. Ethical issues and potential limitations of the study are also covered in this section.

**Chapter 4: Implementation**

Implementation section details our machine learning model for dengue forecasting. We evaluated four algorithms: CNN, Linear Regression, Random Forest, and K-Nearest Neighbors (KNN). Data were pre-processed to ensure its quality for analysis, including strategies to handle missing values and generate potentially new informative features. The selected algorithms were then trained on the prepared data, essentially allowing them to learn patterns associated with dengue. Finally, we employed evaluation metrics to evaluate the performance of each model in predicting dengue presence. This section provides an overview of the technical process adopted to develop machine learning models for dengue forecasting.

**Chapter 5: Result and Analysis**

The paper begins by summarizing performance metrics, including accuracy, precision, memory, F1-score, and area under the receiver operating characteristic curve for each machine learning algorithm considered. Then relative analysis is performed to compare the performance of different algorithms and classify the most effective methods for dengue prediction. Overall, the new findings and analyzes provide valuable insights into the performance of machine learning models for dengue prediction using secondary datasets. Directs future research and informs clinical decision-making processes.

**Chapter 6: Impact on Society, Environment and Sustainability**

"A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever," claims the study, beyond the realm of academia, promising significant impact on society and healthcare. Predictive models can help healthcare providers identify at-risk individuals Dengue develops earlier, allowing timely interpolation and treatment, and interpolation may reduce the need for expensive medical interventions associated with advanced dengue.

**Chapter 7: Conclusion and Further Work**

In conclusion, this paper reveals the potential of machine learning techniques in predicting dengue rates. By studying various lifestyle topics, we developed predictive models with capable accuracy rates. These models also provide valuable insights into preventive health care strategies that are only at risk in emerging dengue.

### 1.7 Summary

In summary, explores the development and application of machine learning algorithms to predict dengue fever probability, aiming to improve early detection and personalized treatment plans. The research aims to develop and validate predictive models using comprehensive datasets including patient demographics, medical history, environmental factors, and other contextual variables. Closing the gap in predictive analytics is critical for transforming healthcare delivery, optimizing resource allocation, and improving patient outcomes. However, limitations include accuracy, reliability, applicability to different patient populations, ethical considerations, and widespread medical community acceptance.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Overview

The literature review section provides a comprehensive overview of existing research and methods related to dengue prediction and management, particularly focusing on the application of machine learning techniques. It examines traditional diagnostic methods or others, their limitations, and research investigating the growing interest in using data-driven methods for early detection and intervention. Key topics include dengue epidemiology, clinical manifestations, and current medical approaches to diagnosis and treatment. The review also explores recent advances in machine learning algorithms, such as decision trees, logistic regression, support vector machines, KNNs and deep learning techniques such as neural networks, highlighting their potential for predictive accuracy and clinical decision making. Furthermore, this section discusses the challenges and gaps in the existing literature, setting the stage for the present study's contribution to filling these gaps through the innovative application of machine learning to dengue prediction.

## 2.2 Related Works

Samrat Kumar Dey et al [1], says The study of this paper is to develop a machine learning model that can use relevant dengue causality information within a geographic region. Epidemiological classification and forecasting techniques for dengue outbreaks, this study scenarios the impact of dengue disease in 11 different districts. MLR, SVR algorithm is used in this paper. The results show that the number of dengue police decreases during the winter season of the designated country and increases during the rainy season. The results of this data-driven analysis show that machine learning algorithms are a must-see for dengue epidemics Diseases of the country and work accordingly currently lacks the availability of these studies to set standard data if these studies do not reflect the true spectrum of dengue.

Dhiman Sarma et al [2], Authors have done this research on dengue forecasting using machine learning algorithm. Dengue is an arbovirus disease caused by Aedes species and is a mosquito-borne disease. Which is basically DENV. A medical status has been set in this regard. Two algorithms are used in this regard, which are Random Forest and Decision Tree. Which gives precision with stability. This model will allow doctors to quickly diagnose fever without the need to follow patients.

Gaurav Gupta et al [3], discuss The dengue strain Aedes strain of the study is an arbovirus disease caused by mosquito-borne dengue viruses (DENVs). Dengue hemorrhagic fever can progress to dengue shock syndrome which can be fatal. It is known that dengue fever can be predicted in advance. In this paper, SVR, Decision Tree, Random Tree, ANN algorithms are used. proposed to develop a machine learning method to predict dengue fever. All of the algorithms require significantly more time to compute because they yield higher results These paper models can help analyze and interpret dengue disease data.

Naizhuo Zhao et al [4], states In this paper discuss on dengue in Colombia uses machine learning to predict dengue virus. Random Forest, ANN and AMIRA models have been used. Long-term forecasting horizons This study demonstrates the feasibility of dengue fever casting RF with a probabilistic approach using a national floral model for fine-spatial forecasting.

T. Marimuthu et al [5], says In this paper the authors present a novel approach for sequence miners. The evolution of the dengue virus is known to have a major impact on the cause of dengue worldwide. Computational models have become a tedious problem due to the scarcity of dengue virus. It is also said that the hierarchical association rules between substances are visualized through interactive tools. Therefore, it is desirable that the structure of a protein can be determined from its sequence through computational models. The accuracy of the proposed model is 96.74% calculated with 10,735 sequences of different lengths as input, 10,198 sequences are correctly classified. The proposed biocomputational model will be helpful in identifying toxic proteins. Presence of recently discovered virus serotypes. Above all, the relationship helps predict dengue serotypes with the proposed tool.

Rao, K et al [6], The authors in the paper discuss the use of decision trees for classification rules in spatial data mining for dengue disease. This highlights the importance of spatial patterns in GIS applications and the potential of decision tree learning algorithms for rule discovery. The paper also explores future prediction models using decision trees to predict the probability of dengue in tribal areas. Two models are developed using real-life problems.

Sarah F et al [7], discuss The paper presents a novel machine learning approach for forecasting dengue fever epidemics in Brazil using weather and population susceptibility cycles. The approach achieved an accuracy of 72%, surpassing naive predictors. Time-series feature extraction is used for supervised learning models. The paper provides a comprehensive analysis of dengue fever dynamics.

Shaukat, K et al [8], In this paper author said that Dengue fever, a disease caused by the Aedes aegypti mosquito, is a global health concern, particularly in tropical or subtropical regions. Researchers used data mining algorithms to analyze dengue fever incidence in Jhelum, Pakistan, in 2011. The study used the DBSCAN algorithm and other clustering techniques, including k-means, DBSCAN, and OPTICS, for dengue clustering. These techniques can be used in disaster prevention planning and comparing different clustering algorithms. However, the DBSCAN algorithm has drawbacks.

Bhatia, R et al [9], says The study compares Neural Network and Nonlinear Regression Models for dengue outbreak prediction in Malaysia. The Neural Network Model outperforms the Nonlinear Regression Model, with Architecture IV being most effective. The findings can help government and health departments plan early intervention programs to prevent dengue fever spread. However, the study has limitations and could benefit from a more detailed explanation of the models used.

P. Manivannan et al [10], In this paper the authors discuss dengue fever, a virus infection that is carried by mosquitoes in tropical and subtropical climates. This highlights the importance of predicting dengue serotypes to aid biotechnologists and bioinformatics in antibiotic discovery. The paper focuses on four stages: prepressing, feature selection, clustering and dengue prediction using clustering algorithms.

Sudsom, N et al [11], The paper discusses the use of insecticide spray-ultra low volume (ULV) to reduce mosquito transmission of dengue virus, but argues for further research on its effectiveness. It also addresses public concerns about human health and environmental impacts. The paper suggests spatial clustering studies can help identify high-risk areas for dengue virus transmission and improve control strategies. It also suggests the use of spatial and temporal risk factor analysis for local dengue prevention and control.

Sharef, Nurfadhlina Mohd et al [12], discuss This paper presents a new temporal proposition approach for trend analysis, focusing on temporal correlations for predicting dengue outbreaks and cyber intrusions. The method uses back propagation and nonlinear regression, with an accuracy of 95%. The method optimizes predictions using temporal abstraction and logic accuracy, with applications in medicine and cybersecurity.

Pi Guo et al [13], states The author discusses the use of machine learning algorithms to predict dengue outbreaks in China, highlighting the lack of accurate forecasting. The study uses an SVR algorithm, which outperforms other forecasting techniques, to track epidemic trajectories and predict dengue outbreaks. The findings could help the government identify necessary control initiatives.

R. Sanjudevi et al [14], says The paper compares data mining and classification techniques for predicting dengue fever, focusing on the SVM algorithm. The SMO technique is used for training and data segmentation, resulting in a more accurate method. The framework effectively uses network classification and data mining techniques, outperforming other relational learning methods in dengue prediction data.

Dewa Ayu Putri Wulandari et al [15], The author discusses the use of machine learning methods, specifically the K-Nearest Neighbor (K-NN) method, to predict the length of hospital stay for patients with dengue fever. It compares different algorithms, including diabetic patients, and uses laboratory blood test results. The K-NN method achieved an accuracy of 65.67%, with the best value of k=13.

### 2.3 Comparison Between Existing Works

This study compares the performance of machine learning models like decision trees, random forests, support vector machines, and neural networks in predicting dengue fever outbreaks. It evaluates their accuracy, precision, recall, and F1-score to determine the most suitable model for predictive analysis. The study also examines the importance of different features used in machine learning models, such as temperature, humidity, precipitation, and population density, in predicting dengue fever outbreaks. It also compares the effectiveness of preventive strategies like vector control measures, public awareness campaigns, and community-based interventions in reducing dengue fever incidence. The study's summary highlights the importance of using machine learning for predictive analysis and preventive strategies in combating dengue fever. Key findings include the predictive accuracy of machine learning models, the identification of important contributing factors, and the effectiveness of preventive strategies in reducing disease transmission. The study's implications for public health policy and practice emphasize the importance of early detection and proactive preventive measures in controlling dengue fever spread and reducing its burden on affected communities. Future directions include refining machine learning algorithms, integrating additional data sources for improved prediction accuracy, and evaluating the long-term impact of preventive strategies on dengue fever transmission dynamics.

Table 2.3.1: Comparative Analysis Table

| **Author** | **Model** | **Dataset** | **Result** | **Contribution** | **Limitation** |
|---|---|---|---|---|---|
| Samrat Kumar Dey.et al. [1] | SVR, MLR | The DengueBD dataset | 67% | Predicts dengue incidence in Bangladesh using machine learning algorithms and data sources, focusing on environmental factors and providing insights for policymakers and health authorities. | Dengue prevalence data in Bangladesh using the DengueBD dataset have limitations in quality, completeness and representativeness, and its predictions may not be applicable to other regions. |
| Dhiman Sarma. et al. [2] | Random Forest, Decision Tree | The study collected real-time patient information, including diagnosis reports, medical histories, and symptoms, from various dengue fever patients. | A machine learning method was developed to predict dengue fever using patient datasets, enhancing accuracy through decision trees and random forest algorithms, demonstrating potential for timely | The paper presents a machine learning model for predicting dengue fever in Bangladesh, utilizing patient data and association rules, achieving 99% accuracy and reducing doctor workload. | Presents a machine learning-based dengue prediction method, but its limitations include a small dataset, lack of demographic information, and inability to |

| | | | | | |
|---|---|---|---|---|---|
| | | | outbreak prediction. | | address real-world healthcare challenges. |
| Gaurav Gupta.et al. [3] | SVM; Decision Tree; Random Forest; ANN | Consists of 1456 data points, while the testing dataset had 416 data points. The total dataset used in the study was 1872. | Proposes a machine learning-based diagnostic model for early dengue disease detection and prediction, using KNN, decision tree, random forest, and Gaussian neighborhood boundaries. | Develops an initial dengue disease detection model using various machine learning algorithms including KNN classifier, decision tree, random forest, Gaussian naive Bayes, support vector classifier and deep convolutional neural network. | Emphasizes the need for improved epidemiological data quality for dengue forecasting, highlighting potential challenges in early stage and implementation of forecasting systems. |
| Naizhuo Zhao et al. [4] | RF, ANN, and ARIMA models | The SIVIGILA portal, a national surveillance program in Colombia that is accessible at the department level, provided the dengue case data. | A 2015 dengue outbreak in Colombia demonstrated the effectiveness of the national random forests model in predicting dengue cases, highlighting its potential for finer spatial scales. | Random Forest (RF) and Artificial Neural Network (ANN) models were compared to predict dengue burden in Colombia, finding RF models to be more accurate and reduce prediction errors. | Focuses on dengue prognosis in Colombia and results may not be directly applicable to other regions or countries. |

| | | | | | |
|---|---|---|---|---|---|
| T. Marimuthu et al. [5] | Association rules, Support vector machine (SVM) | Four dengue virus serotypes' gene sequences are included in the collection, which was gathered from the National Center for Biotechnology Information's (NCBI) online composite database. | Introduces a novel 'Sequence Miner' bio-computational model for dengue gene sequence mining, which achieves 96.74% accuracy and features a protein structure visualization module. | Proposed a novel bio-computational model called 'Sequence Miner' for dengue gene sequence mining, which addresses the lack of computational models in analyzing and interpreting the relationships between dengue viruses. | Introduced a novel bio-computational model for dengue gene sequence mining, but lacks details on datasets, computational resources, hardware requirements, validation, scalability, and visualization module biases**.** |
| Rao, K. et al. [6] | Decision tree (DT) | The paper does not explicitly mention the specific dataset used for the analysis | Presents a decision tree algorithm-based model for predicting dengue disease prevalence in a tribal area, using parameters such as education, income and location but lacks specific numerical results. | Presents a decision tree algorithm-based model for predicting dengue disease outbreaks in a tribal area, emphasizing the importance of spatial data mining and decision tree algorithms. | Specific numerical results, dataset size, accuracy, performance, potential bias, generalizability, and comparison with other disease prediction algorithms, including decision |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | tree algorithms, are lacking. |
| Sarah F. et al. [7] | Support Vector Machine (SVM) | Annual dengue prevalence in 20 Brazilian municipalities, using meteorological data including temperature and rainfall frequency to assess population susceptibility to infection. | A new machine learning method in Brazil predicts dengue fever epidemics using ensemble learning, combining top 11 models and population susceptibility information, showing high performance in specific municipalities. | Presents a machine learning approach for accurate dengue fever epidemic forecasting in Brazil, highlighting the importance of population susceptibility and immunity in improving weather-based predictions. | Presents a machine learning approach for dengue fever prediction in Brazil, highlighting its limitations, challenges, and the need for further research and validation. |
| Shaukat, K et al. [8] | K-Mean, DBSCAN | The dataset was obtained from the office of the Executive District Officer EDO (Health), District Jhelum | A data mining algorithm was used to analyze dengue fever outbreaks in Jhelum, Pakistan using a dataset from the EDO office and comparing different clustering algorithms. | Using data mining algorithms to analyze dengue fever outbreaks in Jhelum, Pakistan, provides insights into disease behavior, geographic patterns, and clustering patterns. | Discusses the sensitivity of the DBSCAN algorithm, its inability to handle outliers and noisy data, and its inability to provide a detailed dataset overview. |

| | | | | | |
|---|---|---|---|---|---|
| Sharef, Nurfadhlina Mohd, et al. [12] | Backpropagation and nonlinear regression | Dengue dataset was used, originally prepared by the Malaysian Ministry of Health and published on the open data portal. The dataset contains four attributes: year, week (Wk), number of dengue-affected areas in that district (NL), and number of cases (NC). | Presents a temporal-based pattern mining method for dengue outbreak and penetration prediction, demonstrating improved accuracy compared to multilayer perceptron and SMOReg algorithms. | Present a temporal mining relation-based method for predicting dengue prevalence and penetration using pattern mining and machine learning algorithms and compares it with a baseline model. | Discusses a proposed temporal mining relationship-based approach for predicting dengue outbreaks and penetration, but lacks specific details on performance improvements, limitations, and comparisons with existing models. |
| Dewa Ayu Putri Wulandari et al. [15] | K-NN | The dataset was the result of performing complete blood laboratory tests on patients with dengue fever as a parameter to predict the length of hospital stay. | Researchers used K-NN method to predict hospital days for dengue fever patients, achieving 65.67% accuracy, but may be lower due to laboratory checks and other factors. | The K-NN method accurately predicts hospital days and costs for dengue fever patients, with an optimal k value of 13, and could be expanded to include additional parameters. | Mainly due to insufficient consideration of other factors and dataset limitations. |

### 2.4 Open Issues

Current research on "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue fever" faces several open problems. Data availability and quality remain significant challenges, as large, diverse and high-quality datasets are difficult to obtain due to privacy concerns and source variability. Feature selection and data preprocessing require careful handling to ensure relevant and consistent inputs. Algorithm-specific challenges also arise: KNNs struggle with computational inefficiency, random forests with interpretability, and linear regression with overfitting. Furthermore, model evaluation lacks standardization, making comparisons difficult. Ensuring model interpretability and integration into clinical practice is crucial for clinician trust and usability. Ethical and legal considerations, including data privacy and bias, must be addressed. It is essential to ensure the generalizability and robustness of the model across different populations and settings. Finally, models need continuous monitoring and updating to maintain their relevance and accuracy over time. Addressing these issues will improve the development and application of machine learning models for dengue prediction, ultimately improving patient care.

**Data Availability and Quality:**

- Lack of large, diverse, and high-quality datasets specific to dengue prediction.
- Difficulty in acquiring comprehensive medical records due to privacy concerns and ethical considerations.
- Variability in data sources that can introduce inconsistencies.

**Feature Selection and Data Preprocessing:**

- Identification of relevant features that significantly impact dengue prediction.
- Challenges in handling missing, incomplete, or noisy data.
- Need for standardization in data preprocessing techniques across different studies.

**Model Evaluation and Validation:**

- Lack of standard evaluation metrics across different studies.
- Need for comprehensive validation strategies, including cross-validation and external validation.

- Difficulty in comparing results due to variations in study designs and datasets.

**Ethical and Legal Considerations:**

- Ensuring patient data privacy and security during model development and deployment.
- Addressing potential biases in models that could lead to unfair or inaccurate predictions.
- Compliance with legal regulations and standards in healthcare data usage.

**Long-Term Model Maintenance:**

- Establishing protocols for continuous monitoring and updating of models to maintain accuracy over time.
- Adapting models to evolving medical knowledge and changes in clinical practices.
- Ensuring models remain relevant and effective as new data becomes available.

## 2.5 Summary

In summary, the literature review provided a comprehensive overview of the current state of knowledge regarding the role of machine learning in enhancing dengue prediction and diagnostic capabilities. It highlighted dengue outbreaks, emphasizing the challenges associated with traditional diagnostic methods and the need for more accurate. Existing research has explored various predictive models and algorithms, demonstrating advances in machine learning techniques such as supervised learning and deep learning. The review also underscored the importance of robust data collection and preprocessing methods as well as model validation and performance evaluation in medical applications. However, gaps remain in understanding the optimal integration of machine learning for dengue prediction, including issues of data quality, model interpretability, and ethical considerations. Moving forward, this study aims to contribute to this evolving field by developing and validating predictive models designed to improve dengue management through innovative machine learning approaches.

# CHAPTER 3

# METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

## 3.1 Overview

Research overview of “A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue fever” targeting prevention and prevention. Dengue fever, a common mosquito-borne viral infection that affects millions of people worldwide, often goes undiagnosed until symptoms are severe. Early detection and intervention are crucial for effective management and prevention of complications. Traditional diagnostic methods rely heavily on clinical examination and laboratory tests, which cannot always accurately predict dengue fever or assess its severity. By leveraging advances in machine learning (ML) and artificial intelligence (AI), this research aims to develop predictive models to aid early detection and severity assessment of dengue fever.

A brief description of this approach involves the use of machine learning techniques to develop predictive models to identify individuals at risk of dengue fever or those who may already have the condition but have not yet been diagnosed. The paper explores the feasibility, accuracy, and effectiveness of various machine learning algorithms that can analyze relevant data such as medical history, environmental factors, symptoms, and possibly even genetic predisposition to predict the likelihood or severity of dengue fever. The comparative aspect evaluates the performance of different algorithms.

This research using machine learning techniques to develop models of dengue fever has the potential to assist healthcare professionals and public health officials in making informed decisions, resulting in timely interventions, improved patient outcomes and reduced healthcare costs. The integration of machine learning algorithms into dengue fever diagnosis and public health strategies holds great promise for enhancing healthcare delivery and patient care. Using the power of data-driven approaches, this study seeks to address existing challenges and pave the way for more accurate, efficient and personalized management of dengue fever.

### 3.2 Proposed Methodology/ System Design

The proposed method for predicting dengue using machine learning techniques involves several key steps. Secondary datasets, which consist of patient data and dengue status, are first collected and pre-processed to guarantee data quality and suitability for training models. It handles numeric feature scaling, categorical variable encoding, and missing values. Feature selection techniques are then used to determine whether relevant predictors of dengue incidence are present. After feature selection, the pre-processed dataset is used to train various machine learning algorithms including CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN), etc. Cross-validation and hyperparameter tuning are used to maximize and avoid model performance. Overfitting Initially, datasets consisting of patient medical records with characteristics such as gender, region, region type, medical history and blood tests are collected and preprocessed. This involves cleaning the data, normalizing features and encoding categorical variables. Feature selection is then performed using statistical methods and dimension reduction techniques to identify the most relevant predictors. Various machine learning algorithms including CNN, linear regression, random forest, KNN are trained and validated using k-fold cross-validation and hyperparameter tuning. Models are evaluated based on metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, along with a confusion matrix to visualize performance. To interpret the model, attribute importance scores and SHAP values are analyzed. The flowchart are given below:

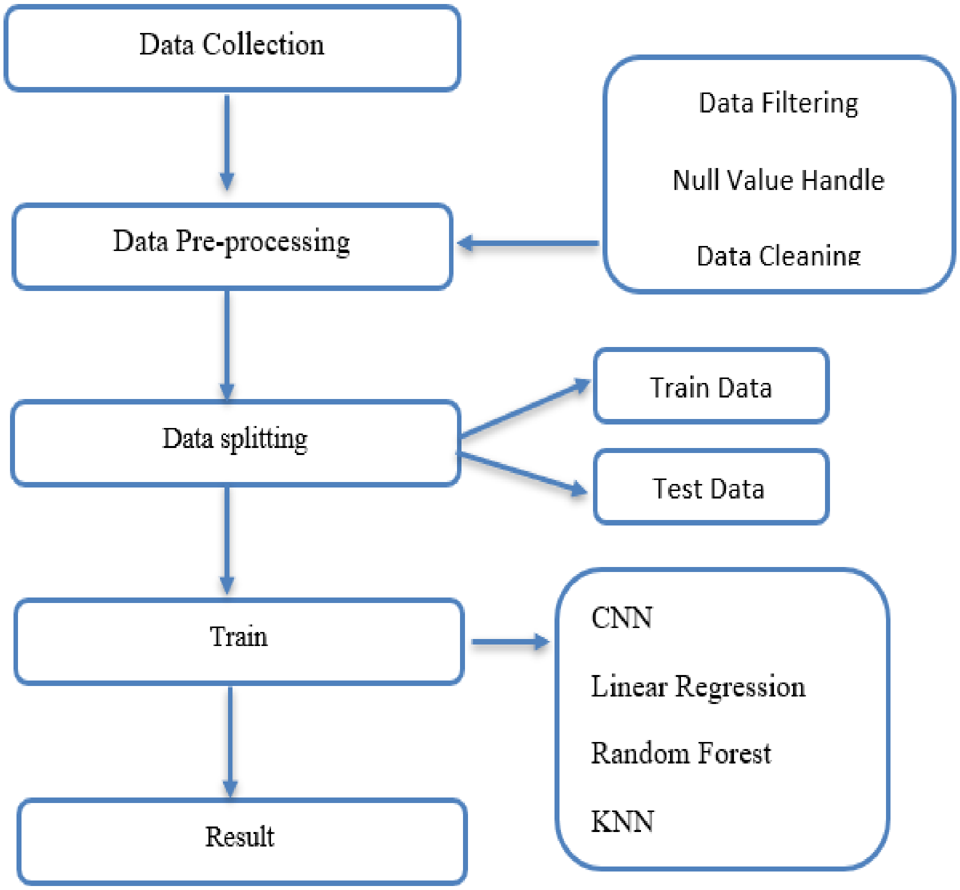


Figure 3.2.1: Workflow of Methodology

**Dataset:**

The secondary dataset utilized in this study comprises patient information collected from medical hospital and Kaggle. The dataset includes a comprehensive range of demographic, clinical, and diagnostic variables relevant to dengue prediction. Key variables collected for each patient include age, gender, area, area type, district, some blood test information.

| 1 | Gender | Age | Area | AreaType | HouseType | District | Humidity | Rainfall | Headche | Pain_behir | Joint_mus | Taste_in_r | Appetite_l | Abdominal | Nausea_v | Diarrhoea | NS1 | IgG | IgM | Outcome |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Female | 45 | Mirpur | Undevelop | Building | Dhaka | 78.82488 | 342.33 | yes | no | no | yes | yes | yes | no | yes | 0 | 0 | 0 | 0 |
| 3 | Male | 17 | Chawkbaz | Developed | Building | Dhaka | 72.60296 | 498 | no | yes | yes | no | no | no | no | no | 0 | 0 | 1 | 0 |
| 4 | Female | 29 | Paltan | Undevelop | Other | Dhaka | 76.90323 | 356.45 | | no | no | yes | yes | yes | yes | no | 0 | 0 | 0 | 0 |
| 5 | Female | 63 | Motijheel | Developed | Other | Dhaka | 73.94381 | 368 | yes | yes | no | no | no | no | no | no | 1 | 1 | 0 | 1 |
| 6 | Male | 22 | Gendaria | Undevelop | Building | Dhaka | 77.36037 | 423.5 | yes | no | no | no | no | no | yes | no | 0 | 0 | 0 | 0 |
| 7 | Female | 36 | Dhanmond | Developed | Other | Dhaka | 85.68286 | 525 | yes | no | no | no | no | yes | yes | no | 0 | 0 | 1 | 0 |
| 8 | Female | 15 | New Mark | Undevelop | Building | Dhaka | 88.05714 | 460 | yes | no | no | yes | no | no | no | no | 0 | 0 | 1 | 0 |
| 9 | Male | 26 | New Mark | Developed | Other | Dhaka | 88.05714 | 460 | yes | yes | no | yes | no | no | no | no | 0 | 0 | 0 | 0 |
| 10 | Female | 31 | Dhanmond | Undevelop | Tinshed | Dhaka | 85.68286 | 525 | yes | yes | yes | no | no | no | no | no | 0 | 0 | 1 | 0 |
| 11 | Female | 10 | Sher-e-Bar | Developed | Tinshed | Dhaka | 85.15524 | 375.32 | yes | no | | no | no | yes | no | no | 0 | 0 | 1 | 0 |
| 12 | Female | 31 | Kafrul | Undevelop | Building | Dhaka | 71.65457 | 183.33 | | no | yes | no | yes | no | yes | no | 1 | 1 | 0 | 1 |
| 13 | Male | 10 | Dhanmond | Developed | Tinshed | Dhaka | 85.68286 | 525 | yes | no | yes | no | no | no | yes | no | 0 | 0 | 0 | 0 |
| 14 | Female | 13 | Pallabi | Undevelop | Building | Dhaka | 70.27755 | 471 | yes | no | no | yes | yes | yes | no | yes | 1 | 1 | 0 | 1 |
| 15 | Female | 43 | Mohamma | Developed | Building | Dhaka | 82.44332 | 473 | no | yes | yes | no | no | no | no | no | 1 | 1 | 0 | 1 |
| 16 | Male | 52 | Shahbagh | Undevelop | Other | Dhaka | 67.54839 | 463.5 | | no | no | yes | yes | yes | yes | no | 0 | 0 | 0 | 0 |
| 17 | Female | 12 | Shyampur | Developed | Tinshed | Dhaka | 84.28762 | 368 | yes | yes | no | no | no | no | no | no | 1 | 1 | 1 | 1 |
| 18 | Male | 18 | Kalabagan | Undevelop | Tinshed | Dhaka | 80.05524 | 380 | yes | no | no | no | no | no | yes | no | 0 | 0 | 1 | 0 |
| 19 | Male | 56 | Bosila | Developed | Building | Dhaka | 80.35023 | 418 | yes | no | no | no | no | yes | yes | no | 0 | 0 | 1 | 0 |
| 20 | Male | 9 | Mirpur | Undevelop | Tinshed | Dhaka | 78.82488 | 342.33 | yes | no | no | yes | no | no | no | no | 1 | 1 | 1 | 1 |
| 21 | Male | 27 | Mirpur | Developed | Tinshed | Dhaka | 78.82488 | 342.33 | yes | yes | no | yes | no | no | no | no | 1 | 1 | 0 | 1 |
| 22 | Female | 31 | Jatrabari | Undevelop | Tinshed | Dhaka | 74.15012 | 181.95 | yes | yes | yes | no | no | no | no | no | 0 | 1 | 0 | 1 |
| 23 | Female | 23 | Adabor | Developed | Other | Dhaka | 70.42166 | 586 | yes | no | | no | no | yes | no | no | 0 | 0 | 0 | 0 |
| 24 | Female | 37 | Kamrangir | Undevelop | Tinshed | Dhaka | 66.45872 | 285 | | no | yes | no | yes | no | yes | no | 0 | 0 | 1 | 0 |

Figure 3.2.2: Datasets

The dataset contains 531 patient records, each represented as a row, and 20 attributes, corresponding to various patient characteristics and characteristics. Data collection procedures adhere to ethical guidelines and patient privacy regulations, ensuring confidentiality and anonymity of patient information. Preprocessing steps, such as handling missing values, encoding categorical variables, and scaling numerical features, were performed to ensure data quality and consistency for machine learning analysis.

Table 3.2.1: Dataset Used in the Train, Test and Validation Sets

| | **Normal** | **Dengue** |
|---|---|---|
| Raw Dataset | 469 | 531 |

### 3.3 Hardware/ Software Requirement

For To implement a machine learning system for dengue prediction, robust hardware and software are essential. On the hardware side, a high-performance computer with a multi-core CPU, at least 16 GB of RAM, SSD storage, and a dedicated GPU are required to handle intensive calculations. Additional cloud resources such as AWS or Google Cloud may be required for scalability. In terms of software, a compatible operating system (Windows, MacOS, or Linux) and programming languages like Python are crucial. Essential libraries and frameworks include TensorFlow or PyTorch for model building, Skit-Learn for machine learning algorithms, and Pandas for data manipulation. Tools for data management (SQL/NoSQL databases), version control (Git), and deployment (Docker, Flask) are also important to ensure smooth development, collaboration, and deployment processes. These requirements collectively ensure efficient and effective execution of a machine learning-based dengue prediction system.

**Hardware**

- High-Performance Computer
- Server Infrastructure
- Peripheral Devices

**Software**

- Operating System
- Programming Languages and Development Environments
- Machine Learning Frameworks and Libraries
- Data Management and Processing Tools
- Version Control and Collaboration
- Deployment and Monitoring
- Security and Compliance Tools

## 3.4 Project Management and Financial Analysis

Effective project management and financial oversight are integral to ensuring the success and sustainability of any endeavor, and they hold particular significance in the context of the proposed research on "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever". The project management aspect encompasses meticulous planning, coordination and execution, from data collection to model training and evaluation. A well-defined project plan will outline key milestones, judiciously allocate resources, and establish timelines to encourage a structured and efficient workflow. At the same time, robust financial management will be paramount, including budget allocations for project performance, equipment, computational resources and potential research collaborations. A transparent and equitable financial strategy will be implemented ensuring optimal utilization of resources while maintaining fiscal responsibility. These integrated approaches to project management and funding are essential pillars that will empower research teams to effectively navigate research complexities, ensure that allocated resources are aligned with research objectives, and facilitate continuous progress toward project goals.

TABLE 3.4.1: Financial Analysis

| SN | Components | Estimated Cost (BDT) |
|---|---|---|
| 01 | Hardware/Infrastructure | 4500-6000 |
| 02 | Visiting Stakeholders | 1000-1500 |
| 03 | Software and Tools | 2000-3000 |
| 04 | Data Collection and Processing | 3000-3500 |
| 05 | Documentation and Report Writing | 500-1000 |
| 06 | Miscellaneous (e.g., licenses, tools) | 1000-1500 |
| 07 | Contingency (10% of total) | 1500-2000 |
| Total Estimated Cost | | 13,500-18,500 |

**3.5 Summary**

In this study on "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever"., the approach includes a systematic approach including data collection, preprocessing, model selection and evaluation. Data will be collected from electronic health records, patient surveys and clinical reports, ensuring a comprehensive dataset. Preprocessing steps include handling missing values, normalizing data, and selecting relevant features to increase model accuracy. Four machine learning algorithms—CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN)—have been selected for their varied strengths in the classification task. Each model will be trained and validated using a cross-validation technique to ensure robustness and prevent overfitting. Performance will be evaluated based on metrics such as accuracy, precision, recall and F1-score. Ethical considerations, including patient privacy and data protection, are integral to the design. This comprehensive approach aims to develop a reliable, interpretable and clinically useful model to predict dengue, ultimately improving patient outcomes and efficiency of care.

# CHAPTER 4
# IMPLEMENTATION

## 4.1 Overview

The implementation section discusses the practical steps taken to develop machine learning models for dengue forecasting. Here, we will uncover specific machine learning algorithms evaluated for this task. We will also focus on the data pre-processing techniques used to ensure that the data is clean, structured and suitable for analysis by the chosen algorithm. This section will further detail the training process employed to build the model, essentially how the model outlines patterns associated with dengue. Finally. We will discuss selected evaluation metrics to evaluate the performance of the model in predicting dengue occurrence. This overview provides a roadmap for understanding the technical complexities involved in building machine learning models for dengue forecasting.

## 4.2 Train Model/ Prototype Design

The implementation of "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever" involves several key steps to develop and validate predictive models. Initially, the dataset, consisting of 1000 patient records with 20 features, undergoes thorough preprocessing. These include handling missing data, normalization, and feature selection to enhance model performance. Four machine learning algorithms—CNN, linear regression, random forest, k-nearest neighbors (KNN)—were selected based on their suitability for the classification task and initial evaluation metrics.

**Train Model:**

Four machine learning algorithms—CNN, linear regression, random forest, k-nearest neighbors (KNN)—are chosen based on their suitability for the classification task and initial evaluation metrics. The dataset is divided into training and validation sets, typically using techniques such as stratified sampling to maintain the distribution of classes. Cross-validation methods such as cross-validation are employed to train each algorithm on different subsets of the training data and evaluate their performance across multiple iterations.

During training, the algorithms learn to create predictive models from the training data that can distinguish between patients with dengue and those without. Model performance is evaluated using standard metrics such as accuracy to assess how well each algorithm generalizes to new, unseen data.

The training phase involves fine-tuning hyperparameters specific to each algorithm to further optimize performance. This iterative process ensures that the models achieve high accuracy and robustness while avoiding overfitting. The results of this training phase lay the foundation for subsequent validation and deployment of machine learning models in clinical practice, with the goal of increasing diagnostic accuracy and improving patient care outcomes related to dengue prognosis.

**Prototype Design:**

The prototype design for the implementation of "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever" focuses on developing a user-friendly interface that integrates the trained models into clinical practice. The design ensures accessibility for healthcare professionals and aims to streamline the prediction process based on patient data.

**User Interface:** The prototype has a graphical user interface (GUI) that allows clinicians to easily input patient information. The interface is designed to be intuitive with clear prompts for entering relevant medical data such as symptoms, medical history and demographic information.

**Data Integration:** The prototype integrates seamlessly with existing electronic health record (EHR) systems or standalone data input mechanisms. This ensures that clinicians can efficiently and securely retrieve patient data for predictive purposes.

**Prediction Engine:** Behind the interface, the prototype has trained machine learning models—CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN). These models are able to analyze the inputted data to predict the likelihood of a patient having dengue.

**Output and Interpretability:** The prototype provides clear and interpretable output for clinicians, indicating the predicted probability or classification of dengue. Visual aids such as charts or risk scores can accompany predictions for understanding and decision making.

**Ethical considerations:** Throughout the design process, ethical considerations regarding patient data privacy and model transparency are paramount. The prototype ensures compliance with data protection regulations and includes features to explain model predictions to clinicians.

**Testing and Validation:** Before deployment, the prototype undergoes rigorous testing to ensure accuracy, reliability and usability in real-world clinical situations. Validation against independent datasets or through clinical trials further confirms the efficacy and generalizability of predictive models.

**Iterative Improvement:** Iterative updates based on continuous feedback from healthcare professionals and improvements to prototypes of performance metrics. This iterative approach aims to improve prediction accuracy, user satisfaction, and overall clinical utility over time.

Focusing on these design principles, the prototype aims to bridge the gap between machine learning research and practical healthcare applications and contribute to patient care outcomes in dengue prediction.

### 4.3 System Testing/ Model Evaluation

The implementation of "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever" involves rigorous system testing and model evaluation to ensure accuracy, reliability, and usability in clinical practice.

**Testing Framework:** A comprehensive testing framework is established to evaluate the prototype's performance. This includes unit testing of individual components such as data preprocessing modules, machine learning algorithms, and the user interface. Integration testing ensures seamless interaction between different system modules, including data input, prediction engine, and output visualization.

**Performance Metrics:** Standard performance metrics such as accuracy employed to assess the predictive models' effectiveness. These metrics provide quantitative insights into how well the models classify patients into dengue and non-dengue categories based on input data.

**Cross-Validation:** Cross-validation techniques, such as cross-validation, are utilized to validate the models' generalizability and robustness. By splitting the dataset into multiple subsets, training and testing are performed iteratively to mitigate overfitting and ensure that the models perform well on unseen data.

**Validation Against Baselines:** The performance of the machine learning models—CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN)—is compared against baseline methods or existing clinical practices. This comparative analysis helps validate the models' superiority in terms of accuracy and predictive power.

**Real-World Simulation:** Simulated real-world scenarios and datasets may be used to further evaluate the prototype's performance under varied conditions. This testing phase assesses how well the predictive models adapt to different patient demographics, symptom profiles, and clinical settings.

**Clinical Validation:** Collaboration with healthcare professionals facilitates clinical validation of the prototype. Clinicians evaluate the prototypes' predictions against actual patient outcomes, providing qualitative feedback on usability, interpretability, and clinical relevance.

**Ethical Considerations:** Throughout the testing and evaluation process, ethical considerations regarding patient privacy, data security, and model transparency are paramount. Measures are taken to ensure compliance with regulatory requirements and ethical standards in healthcare data usage.

By conducting thorough system testing and model evaluation, this implementation phase aims to validate the effectiveness and applicability of machine learning-based dengue prediction models in real-world clinical settings. The insights gained from these evaluations contribute to refining the models, improving accuracy, and ultimately enhancing patient care outcomes.

## 4.4 Summary

This section outlines the implementation process for developing a machine learning model to predict dengue. We detail the machine learning algorithm selected, the data pre-processing techniques employed to clean and prepare the data for analysis, and the training process used to build the model. In addition, we describe the evaluation metrics used to evaluate the performance of the model in predicting dengue. This section provides a broad overview of the technical aspects involved in developing machine learning models for dengue forecasting.

# CHAPTER 5
# RESULT AND ANALYSIS

## 5.1 Overview

This study on "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever" provided insightful results on the performance of four algorithms: CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN). The dataset is then split into training and testing sets. Model performance is evaluated using appropriate evaluation metrics, such as accuracy and area under the receiver operating characteristic curve (AUC-ROC). Comparative analysis is performed to compare the performance of different algorithms and identify the most effective method for dengue prediction. Validation techniques including sensitivity analysis and external validation are employed to ensure the reliability and generalizability of the results. Following this experimental setup, the study aims to develop accurate and reliable predictive models for dengue prognosis, which contribute to improved healthcare decision-making and patient outcomes. The study found that Linear Regression and Random Forest achieved the highest accuracy among the algorithms, with an impressive rate of 99%. Linear regression and random forest ability demonstrate its robustness and suitability for medical prediction tasks in effectively classifying patients into normal and affected categories. CNN also performed strongly, achieving an accuracy of 97%, demonstrating its ability to handle complex data structures and variables. KNN demonstrated an accuracy of 54%, indicating their competitive performance in predicting dengue outcomes. These results suggest that machine learning algorithms can significantly contribute to the accuracy and efficiency of dengue prediction, providing healthcare practitioners with valuable tools to improve diagnostic processes and patient care strategies. Further analysis highlights the need for continuous refinement and validation of models for different patients.

## 5.2 Experimental/ Simulation Result

This section compares four main machine learning algorithms. Classification performance comes first. Then, the overall metrics of those models are reviewed to collect, describe, identify potential causes and areas for improvement in results. The section begins by summarizing performance metrics including accuracy and area under the receiver operating characteristic curve (AUC-ROC) for each machine learning algorithm considered. Comparative analysis is then performed to compare the performance of different algorithms and identify the most effective method for dengue prediction. Additionally, visualizations such as confusion matrix, ROC curve can be provided to further illustrate the performance of the models. The strengths and weaknesses of each algorithm are discussed, highlighting factors such as predictive accuracy, interpretability, and computational efficiency. Sensitivity analysis is performed to assess the stability of model predictions under different conditions, external validation techniques can be employed to assess model performance on independent datasets if available. Overall, the experimental results and analyzes provide valuable insights into the performance of machine learning models for dengue prediction using secondary datasets, informing future research directions and clinical decision-making processes.

Table 5.2.1: Precision, Recall, F1-Score

| **Architecture** | **Training Accuracy** | **Model Accuracy** |
|---|---|---|
| **CNN** | **97.17%** | **97%** |
| **Linear Regression** | **99.29%** | **99%** |
| **Random forest** | **99.35%** | **99%** |
| **KNN** | **53.52%** | **54%** |

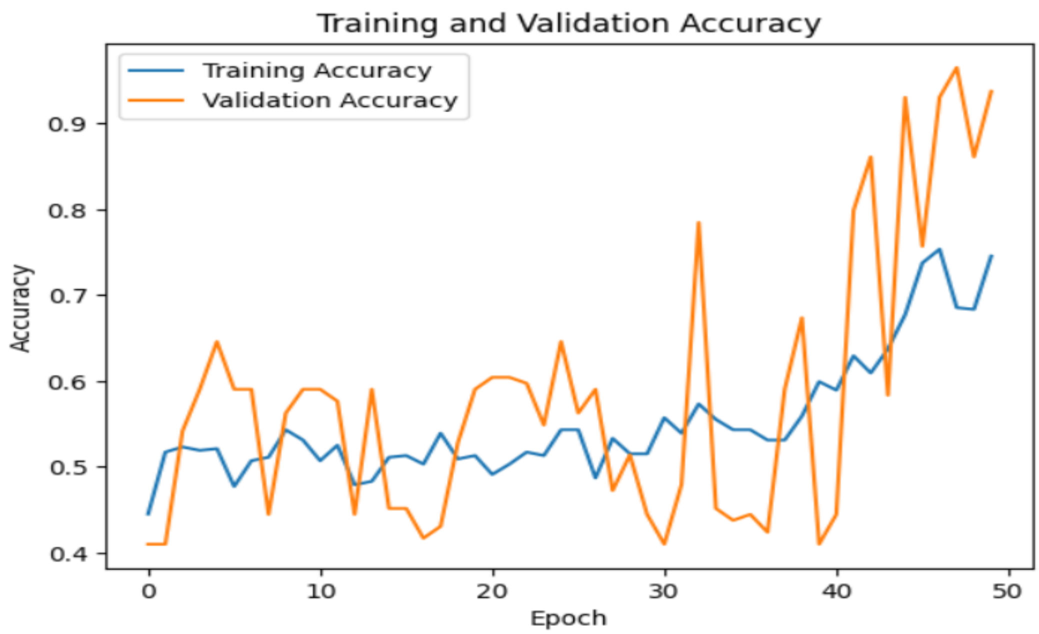


Figure 5.2.1: Training and Validation Accuracy.

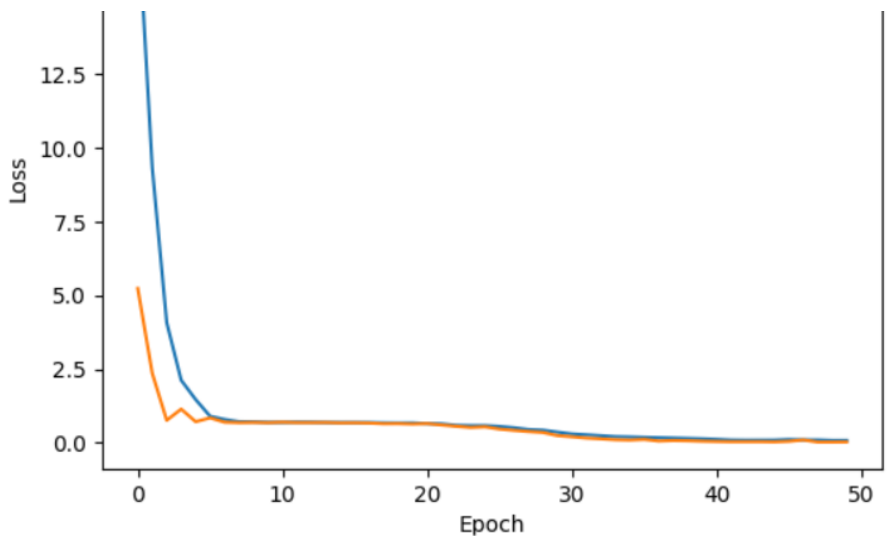


Figure 5.2.2: Loss curve for the CNN model.

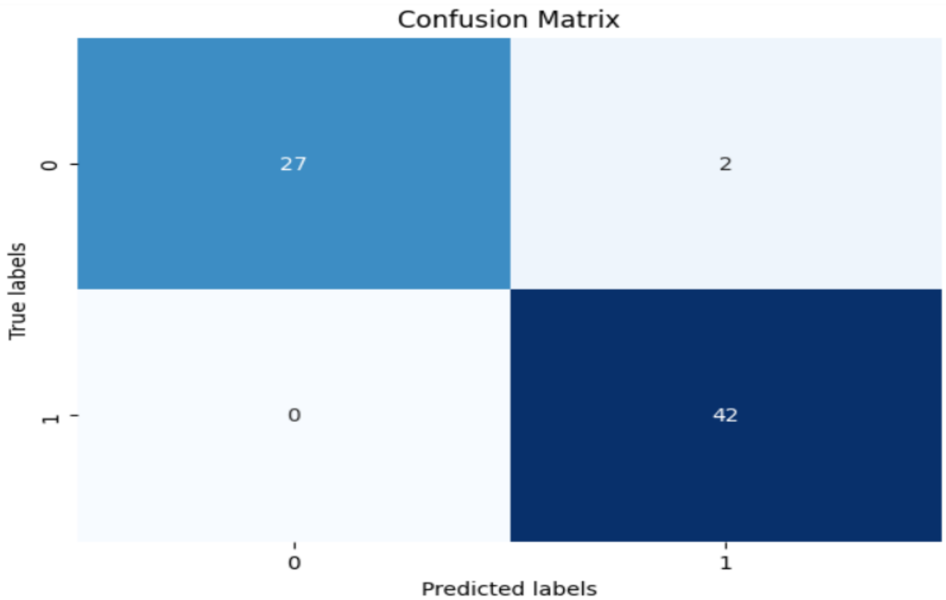


Figure 5.2.3: Confusion Matrix

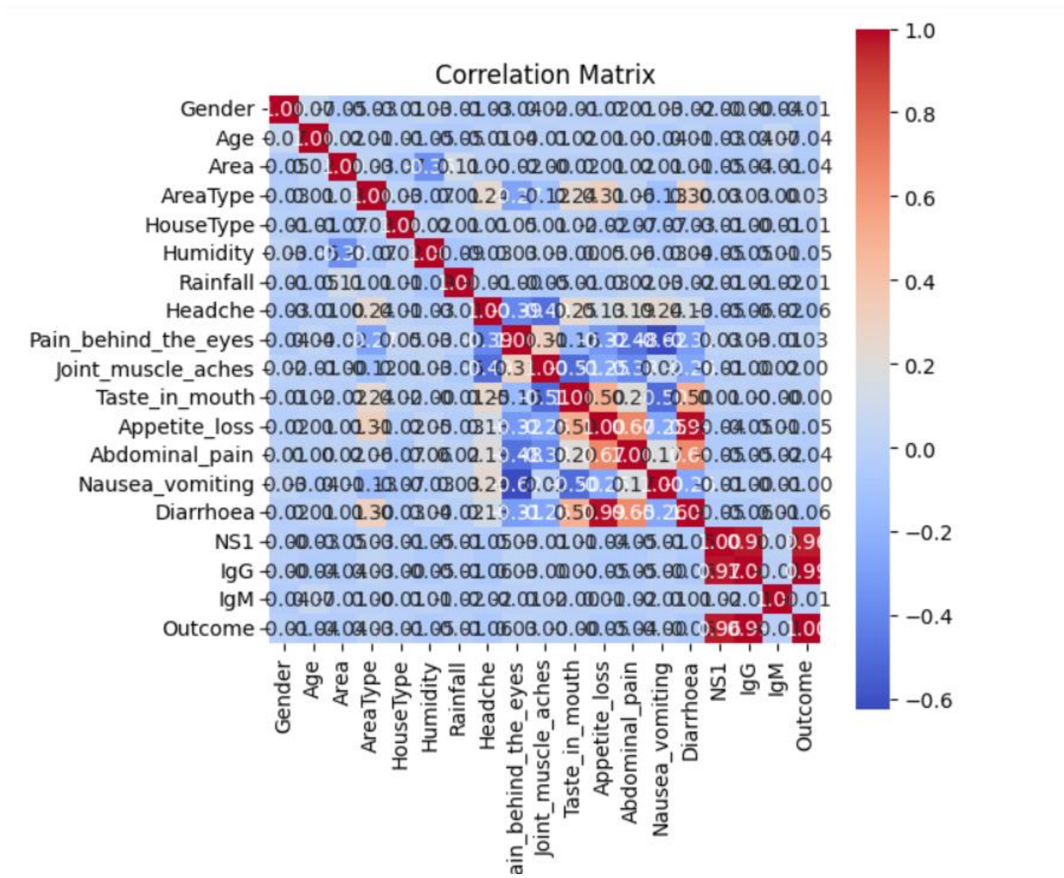


Figure 5.2.4: Correlation Matrix.

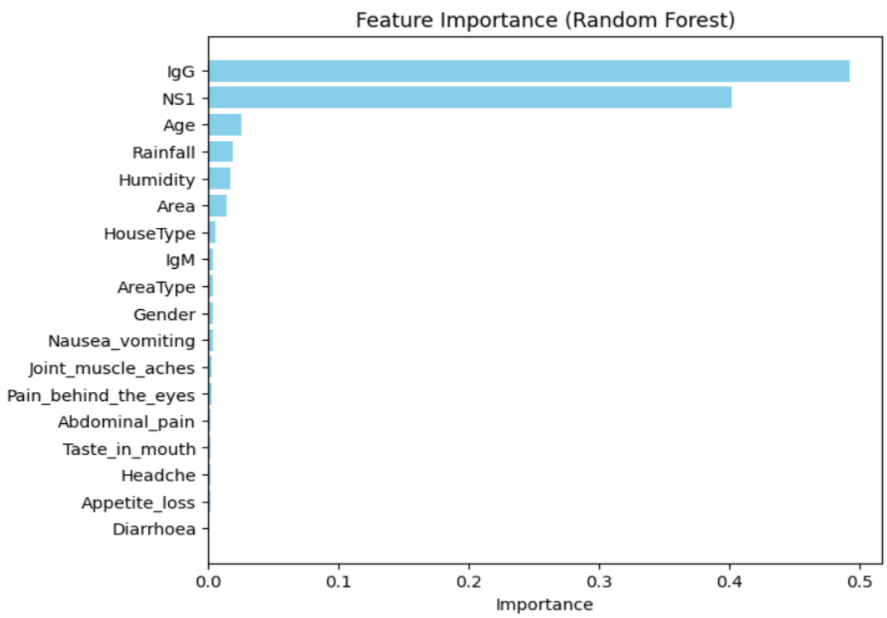


Figure 5.2.5: Bar Plot for Random Forest.

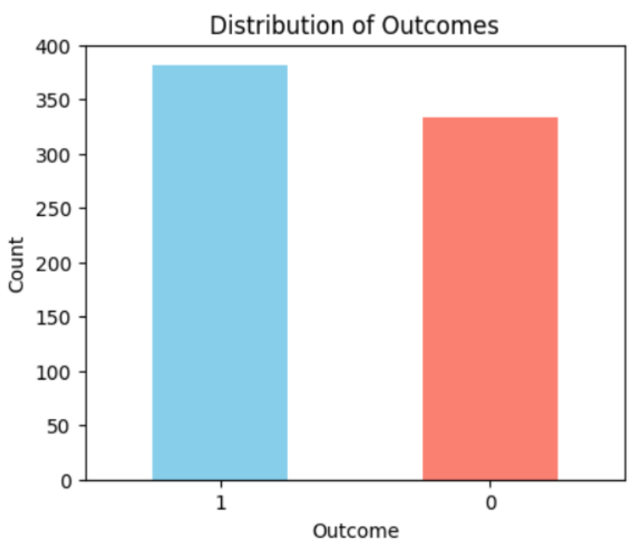


Figure 5.2.6: Bar plot for Outcome.

The model appears to perform well overall, as evidenced by its high ratio of true positives to true negatives. Nevertheless, a small percentage of false positives and false negatives occur. False negative results can cause patients to wait longer for necessary treatment, while false positive results can lead to unnecessary care for patients who do not actually have dengue. It is important to note that this is only a small sample of data, and the performance of the model may vary based on the dataset used for training. Furthermore, the confusion matrix only provides part of the picture. Additional metrics that can reveal more about the model's performance.

### 5.3 Performance/ Comparative Analysis

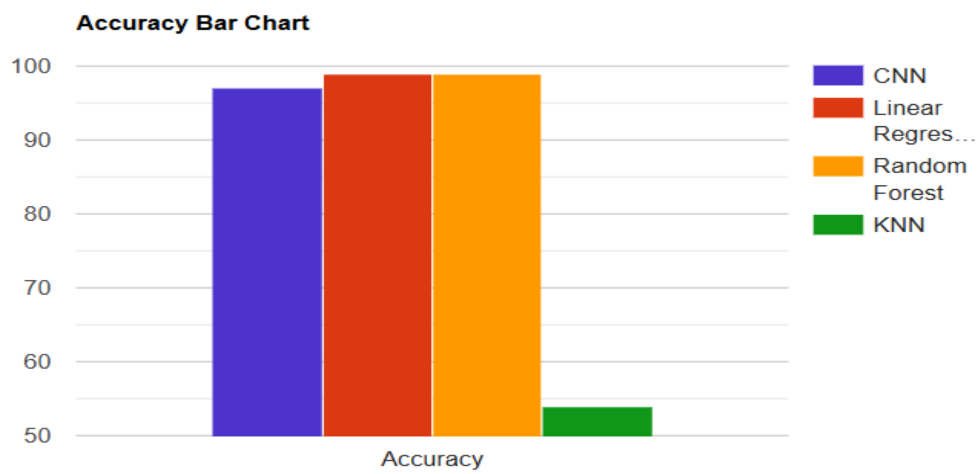


Figure 5.3.1: Accuracy Bar Chart

In this section of the thesis paper "A machine learning method for predictive-probability analysis and preventive strategies in dengue fever" where multiple machine learning algorithms achieve high accuracy, such as linear regression and random forest with 99%, CNN 97%, K-Nearest Neighbors (KNN) with 54% several points can be highlighted to contextualize and analyze these results. First, the discussion can focus on comparing the performance of each algorithm, emphasizing their strengths and weaknesses in terms of dengue prediction. For example, the high accuracy of linear regression and random forest can be attributed to their ability to effectively define nonlinear decision boundaries, while the performance of random forest can be attributed to their ability to handle high-dimensional data and capture complex interactions between features. Similarly, the discussion may explore the trade-off between interpretability and accuracy, noting that while CNN achieves higher accuracy, KNN may offer more interpretability due to their simpler decision-making processes. Additionally, the discussion is the potential utility of machine learning models to assist healthcare professionals in informing targeted interventions or preventive measures. Furthermore, the discussion may consider potential sources of bias or confounding that may affect model performance, as well as the generalizability of findings across different patient populations or healthcare settings. Overall, the discussion provides critical insight into the significance of the high accuracy achieved by multiple machine learning algorithms in dengue forecasting and provides research implications.

### 5.4 Summary

The Results and Analysis section presents the results obtained from the machine learning model developed for dengue forecasting. Here, we discuss the performance of the model on selected evaluation metrics. This involves presenting accuracy metrics used to assess the model's ability to accurately predict the presence of dengue. We will further analyze the results to gain insight into the strengths and weaknesses of the model. This may include exploring factors affecting model performance, identifying potential areas for improvement, or visualizing model predictions to understand the decision-making process. The Results and Analysis section explores the performance of machine learning models developed for dengue forecasting. We evaluated four algorithms: CNN, Linear Regression, Random Forest, K-Nearest Neighbors (KNN). Among these algorithms, Linear Regression and Random Forest achieved the highest accuracy of 99%, followed by CNN at 97%, KNN at 54%. This indicates that linear regression and random forest demonstrated the strongest ability to accurately predict dengue presence in our dataset. Further analysis will delve deeper into the performance of each model. We will explore factors that may affect the results, such as the complexity of the data or specific characteristics of each algorithm. Additionally, we can consider visualization techniques to understand how each model predicts and identify potential areas for improvement. This section aims to provide a comprehensive understanding of the performance of the models in predicting dengue. By comparing their performance and exploring underlying factors, we can gain valuable insight into the suitability of different machine learning methods for these specific tasks.

# CHAPTER 6

# IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

## 6.1 Impact on Life

The impact of accurately predicting dengue using machine learning extends significantly into various aspects of life, both for individuals and the healthcare system. For patients, early and precise prediction can lead to timely interventions, reducing the discomfort and complications associated with advanced stages of the disease. This proactive approach can improve quality of life by minimizing symptoms such as high fever, severe headaches, and joint and muscle pain, and can prevent the progression to severe dengue, which can be life-threatening.

From a healthcare perspective, machine learning models can optimize resource allocation by identifying high-risk individuals and areas, enabling more efficient management of healthcare resources and reducing the burden on medical professionals. This can be particularly crucial in areas where dengue is endemic, as it allows for targeted vector control measures, such as mosquito eradication programs, and focused public health campaigns to reduce transmission.

Moreover, the integration of predictive analytics can enhance patient education and self-care, empowering individuals to take preventive measures such as using mosquito repellents, installing screens on windows, and eliminating standing water where mosquitoes breed. This can significantly reduce the incidence of dengue and its spread within communities.

Overall, the adoption of machine learning for dengue prediction promises to streamline healthcare processes, reduce costs, and improve patient outcomes, marking a significant advancement in the medical field. By leveraging data-driven insights, healthcare systems can become more proactive and responsive in managing dengue, ultimately saving lives and enhancing public health.

## 6.2 Impact on Society and Environment

The social and environmental implications of using machine learning for dengue forecasting are multifaceted and profound. At the societal level, these technological advances democratize access to preventive health care, particularly benefiting disadvantaged and rural communities that may lack specialized medical services. By facilitating early detection and management, machine learning can reduce the incidence of severe dengue disease, thus reducing overall healthcare costs and reducing the economic burden on individuals and healthcare systems. In addition, it promotes a culture of preventive health, encouraging individuals to seek early advice and take measures to avoid mosquito bites and eliminate breeding sites.

From an environmental perspective, medical interventions and reduced hospital visits associated with improved dengue treatment lead to reduced medical waste and resource consumption. Fewer hospitalizations and intensive care requirements mean less reliance on disposable medical supplies, thereby contributing to environmental sustainability. Furthermore, the shift towards digital health solutions reduces the need for physical infrastructure and travel, further reducing the carbon footprint associated with healthcare delivery.

Furthermore, predictive analytics can help target vector control, reducing the need for widespread use of pesticides and thus reducing their environmental impact. By identifying high-risk areas, authorities can implement more efficient and localized mosquito control measures, preserving ecosystems while still fighting the spread of dengue.

In summary, the application of machine learning to dengue prediction not only enhances public health and economic efficiency but also supports environmental conservation efforts. By using technology to improve early detection and prevention, society can reduce the effects of dengue and build a healthier, more sustainable future.

### 6.3 Ethical Aspects

The research on "A Machine Learning Approach for Predictive-Probability Analysis and Preventive Strategies in Dengue Fever" is underpinned by a commitment to ethical considerations throughout its methodology. The ethical considerations for a research project include informed consent, data privacy and security, bias and fairness, transparency and accountability, beneficence and non-maleficence, equitable access to benefits, regulatory compliance, public engagement and communication, long-term implications, and collaboration and interdisciplinary approach. Informed consent ensures participants understand the purpose of data collection and how it will be used. Data privacy and security protect sensitive information and anonymize data to prevent re-identification. Be vigilant about biases in data or algorithms used in machine learning approaches to prevent discriminatory outcomes. Transparency in research methodology ensures accountability and allows for scrutiny and reproducibility of findings. Maximize the benefits of predictive models for preventive strategies in dengue fever while minimizing potential harm. Ensure equitable access to benefits for all populations, including marginalized or underprivileged communities. Adhere to relevant regulations and guidelines governing research involving human subjects and healthcare data. Public engagement and communication facilitate informed decision-making by communicating findings clearly and understandably. Consider the long-term implications of deploying machine learning models for predictive analysis and preventive strategies in dengue fever and collaborate with experts from diverse fields to develop holistic preventive strategies. By addressing these ethical considerations, the research can contribute to the advancement of knowledge and the development of ethically sound approaches to combating dengue fever.

By addressing these ethical considerations, researchers can ensure that the development and deployment of predictive models for dengue prediction is conducted in a responsible and ethically sound manner, prioritizing patient well-being, autonomy, and justice.

### 6.4 Sustainability Plan

The sustainability plan for the study "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever" is designed to minimize environmental impact while ensuring ongoing societal contributions. This includes continuous data collection, algorithm refinement, and increasing model performance for better prediction accuracy. Collaboration with healthcare institutions, academic researchers, and industry stakeholders is critical to accessing the necessary data, resources, and expertise. Educating healthcare professionals, patients and the wider community through workshops, educational materials and online platforms is essential to promote understanding and effective use of technology. Evaluating impact involves conducting prospective studies, monitoring patient outcomes, and making iterative improvements. Scaling implementation involves expanding access to technology, addressing resource constraints, and adapting models to different settings and populations. This comprehensive approach ensures that predictive models for predicting dengue have a lasting and meaningful impact on healthcare delivery, patient outcomes and public health.

### 6.5 Summary

Implementing machine learning for dengue predicting has significant implications for society, environment and sustainability. Socially, this technology increases the accessibility and equity of healthcare and benefits remote communities. It reduces the incidence of severe dengue, reduces the economic burden on individuals and healthcare systems, and fosters a proactive health culture. Environmentally, reducing critical care and hospitalizations reduces medical waste and resource consumption, contributing to sustainability. Fewer hospitalizations mean less use of disposable medical supplies in line with environmental conservation efforts. Additionally, adoption of digital health solutions reduces the need for extensive physical infrastructure and reduces travel, thereby reducing the carbon footprint of the healthcare sector. Overall, integrating machine learning into dengue prediction fosters a sustainable, efficient, and equitable healthcare system, driving positive social and environmental outcomes.

# CHAPTER 7
# CONCLUSION AND FURTHER WORK

## 7.1 Conclusions

This study examines the feasibility of using machine learning algorithms to predict dengue using a primary dataset of patient information. Through extensive experiments and analysis, several key conclusions can be drawn. First, the results demonstrate the effectiveness of machine learning techniques such as CNN, Linear Regression, Random Forest K-Nearest Neighbors (KNN) to accurately predict dengue probability based on patient demographics, medical history and symptoms. These findings highlight the potential of machine learning as a valuable tool in medical decision-making processes supporting dengue diagnosis and management. In addition, comparative analysis of different algorithms revealed insights into their respective strengths and weaknesses. Random forests offer a balance between accuracy and interpretability, making it a practical choice for clinical settings where clarity and ease of interpretation are paramount.

However, it is essential to acknowledge the limitations of this study. The secondary datasets used may have inherent biases or limitations, and further investigation is needed to generalize the results to larger patient populations or healthcare settings. Additionally, the performance of machine learning models can be affected by factors such as data quality, feature selection, and model hyperparameters, which require careful consideration and optimization. Furthermore, the development of interpretable machine learning models can facilitate the adoption of predictive analytics in real-world healthcare settings. In conclusion, although this study represents an important step toward leveraging machine learning to predict dengue, more research and collaboration between healthcare professionals and data scientists is needed to fully realize the potential of predictive modeling in improving patient outcomes and healthcare delivery.

## 7.2 Further Suggested Works

Future research on "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever" can build upon this study by addressing several key areas. Firstly, expanding the dataset to include a larger and more diverse population would enhance the model's generalizability and robustness. Incorporating additional features, such as climatic data, socio-economic factors, and individual health history, could improve prediction accuracy. Additionally, developing real-time prediction systems integrated into public health surveillance and electronic health records would facilitate timely and effective interventions. Longitudinal studies to monitor the model's performance over time and its impact on patient outcomes and disease spread would provide valuable feedback for continuous improvement. Collaboration with public health officials and clinicians to refine and validate the models in practical settings is crucial for ensuring their applicability and effectiveness. Finally, addressing ethical considerations, including bias mitigation and ensuring data privacy, will be essential as these models are further developed and deployed in public health environments.

### 7.3 Limitations/ Conflict of Interests

This study on "A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever", while promising, has several limitations. Post secondary datasets may not be fully representative of the larger patient population, potentially limiting the generalizability of the model. An imbalance in the dataset with a high proportion of normal patients may affect model performance and lead to biased predictions. Additionally, reliance on self-reported data introduces the potential for inaccuracies and inconsistencies. Although the selected algorithms (CNN, Linear Regression Random Forest, KNN) are powerful, each has inherent limitations, such as the computational inefficiency of KNN with large datasets. The study also did not include external validation, which is essential to assess the applicability of the model in real-world clinical settings. Furthermore, there may be potential conflicts of interest if researchers have affiliations or financial relationships with organizations that may benefit from certain results. It will be important to address these limitations in future studies to develop more accurate, reliable and clinically useful predictive models for dengue.

## REFERENCES

[1] Dey, S. K., Rahman, M. M., Howlader, A., Siddiqi, U. R., Uddin, K. M. M., Borhan, R., & Rahman, E. U. (2022). Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach. PLoS One, 17(7), e0270933.

[2] Sarma, D., Hossain, S., Mittra, T., Bhuiya, M. A. M., Saha, I., & Chakma, R. (2020, December). Dengue prediction using machine learning algorithms. In 2020 IEEE 8th R10 humanitarian technology conference (R10-HTC) (pp. 1-6). IEEE..

[3] Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B. I., Siddiqui, T., ... & Al-Subaie, M. (2023). DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. Diagnostics, 13(6), 1093.

[4] Zhao, N., Charland, K., Carabali, M., Nsoesie, E. O., Maheu-Giroux, M., Rees, E., ... & Zinszer, K. (2020). Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. PLOS Neglected Tropical Diseases, 14(9), e0008056.

[5] Marimuthu, T., & Balamurugan, V. (2015). A novel bio-computational model for mining the dengue gene sequences. International Journal of Computer Engineering & Technology, 6(10), 17-33.

[6] Rao, N. K., Varma, G. S., Rao, D., & Cse, P. (2014). Classification rules using decision tree for dengue disease. International Journal of Research in Computer and Communication Technology, 3(3), 340-343.

[7] McGough, S. F., Clemente, L., Kutz, J. N., & Santillana, M. (2021). A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. Journal of the Royal Society Interface, 18(179), 20201006.

[8] Shaukat, K., Masood, N., Shafaat, A. B., Jabbar, K., Shabbir, H., & Shabbir, S. (2015). Dengue fever in perspective of clustering algorithms. arXiv preprint arXiv:1511.07353.

[9] Husin, N. A., & Salim, N. (2008, August). Modeling of dengue outbreak prediction in Malaysia: a comparison of neural network and nonlinear regression model. In 2008 international symposium on information technology (Vol. 3, pp. 1-4). IEEE.

[10] Manivannan, P. I. D. P., & Devi, P. I. (2017, March). Dengue fever prediction using K-means clustering algorithm. In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-5). IEEE.

[11] Sudsom, N., Thammapalo, S., Pengsakul, T., & Techato, K. J. J. T. (2016). A spatial clustering approach to identify risk areas of dengue infection after insecticide spraying. Jurnal Teknologi, (5-3).

[12] Sharef, N. M., Husin, N. A., Kasmiran, K. A., & Ninggal, M. I. H. (2019). Temporal Trends Analysis for Dengue Outbreak and Network Threats Severity Prediction Accuracy Improvement. J. Digit. Inf. Manag., 17(3), 122.

[13] Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., ... & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. PLoS neglected tropical diseases, 11(10), e0005973.

[14] Sanjudevi, D., & Savitha, D. (2019). Dengue fever prediction using classification techniques. Int Res J Eng Technol (IRJET), 6(02), 558-563.

[15] Wulandari, D. A. P., Permana, K. A. B., & Sudarma, M. (2018). Prediction of Days in Hospital Dengue Fever Patients using K-Nearest Neighbor. International Journal of Engineering and Emerging Technology, 3(1), 23-25.

# Appendix A

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

**Title:**
**A MACHINE LEARNING APPROACH FOR PREDICTIVE-PROBABILITY ANALYSIS AND PREVENTIVE STRATEGIES IN DENGUE FEVER**

**Student ID: [202-15-3808, 202-15-3814]**

## CO Description for FYDP

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

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

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

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

| SN | EP Definition | Attainment | CO | Justification (with Knowledge Profile) | References |
|---|---|---|---|---|---|
| 1. | **EP1: Depth of Knowledge required** | Yes | CO1, CO2, CO3, CO5, CO6, CO7 and CO8 | For our research, we need to collect data from online and understand how to build a machine learning-based model that addresses **K3** and **K4.** | **Page no:** [2-3] **Section:** [1.4] |
| | | | | Our paper needs to include many elements and we collected data with **K5** and **K6** | **Page no:** [22-23] **Section:** [3.4] |
| | | | | We have studied existing models with similar goals (**K8**). | **Page no:** [7-10] **Section:** [2.1, 2.2] |
| 2. | **EP2: Range of Conflicting Requirements** | Yes | CO2, and CO7 | We had trouble finding specific laws governing the collection of data from various individuals for our work. | **Page no:** [16-17] **Section:** [2.4,2.5] |

| | | | | | |
|---|---|---|---|---|---|
| **3.** | **EP3: Depth of analysis required** | Yes | CO2, and CO6 | With so much information, it's hard to come up with a definitive answer for our paper. We carefully considered several options and conducted a thorough investigation before selecting a particular algorithm. | **Page no:** [19-22] **Section:** [3.2,3.3] |
| **4.** | **EP4: Familiarity of Issues** | Yes | CO8 | For our paper, in order to understand and analyze physical health and dengue we had to explore many things that we were not very familiar with. In particular, we need to determine the best way to quantify dengue. | **Page no:** [22-23] **Sectio:** [3.4] |
| **5.** | **EP5: Extends of application codes** | No | CO5 | N/A | N/A |
| **6.** | **EP6: Extends of stakeholders involved and conflicting** | Yes | CO8 | Our project measures dengue based on individual activity. This convention will help us determine the appropriate data for analysis. | **Page no:** [2-3] **Section:** [1.4] |
| **7.** | **EP7: Interdependence** | Yes | CO5 | Our paper consists of interdependent subsystems. These responsibilities involve data collection, processing, machine learning model training and predictive analytics. | **Page no:** [22-23] **Section:** [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 paper utilizes diverse resources such as high-performance computing infrastructure, machine learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in dengue detection. | **Page no:** [24-27] **Section:** [4.1,4.2,4.3] |
| 2. | EA2: Level of interaction | No | | N/A | N/A |
| 3. | EA3: Innovation | No | | N/A | N/A |
| 4. | EA4: Consequences for society and the environment | Yes | | This paper contributes to society by improving healthcare through advanced A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy. | **Page no:** [35-38] **Section:** [6.1,6.2,6.3,6.4] |

| | | | | | |
|---|---|---|---|---|---|
| **5.** | **EA-5: Familiarity** | Yes | | This paper expands upon existing research by examining a novel approach in A Machine Learning Approach For Predictive-Probability Analysis And Preventive Strategies In Dengue Fever through machine learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field. | **Page no:** [18-21] **Section:** [3.1,3.2] |

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

| SN | COs | Attainment | Justification | References |
|---|---|---|---|---|
| **1** | CO4 | Yes | This work tackles **CO4** through the integration of efficient project management and financial supervision, guaranteeing careful planning, resource distribution, and budget estimation for the best possible resource use during the course of the research process. | **Page no:** [22-23] **Section:** [3.4] |
| **2** | CO9 | Yes | By putting patient privacy first, getting informed consent, and openly reporting the research process, the paper demonstrates adherence to ethical principles. It also ensures responsible knowledge dissemination and societal well-being through the ethical application of cutting-edge healthcare technologies that abide by **CO9**. | **Page no:** [35-37] **Section:** [6.2,6.3] |
| **3** | CO10 | No | N/A | N/A |
| **4** | CO12 | Yes | The extensive data collection, rigorous statistical analysis, careful methodology development, and thorough experimental results and analysis all demonstrate the paper's dedication to continuous learning **(CO12)** and adaptation within the ever-evolving technological landscape. This shows a commitment to staying current and improving techniques to address contemporary challenges. | **Page no:** [28-34] **Section:** [5.2,5.3] |

Plagiarism Report Checking for Final Defense v1 for Ashraful and Amid

ORIGINALITY REPORT

| 15% | 12% | 5% | 2% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 2% |
| 2 | www.fastercapital.com<br>Internet Source | 1% |
| 3 | www.frontiersin.org<br>Internet Source | 1% |
| 4 | www.ijraset.com<br>Internet Source | 1% |
| 5 | fastercapital.com<br>Internet Source | 1% |
| 6 | Bani Ahmad, Oday Ali. "Diabetic Retinopathy (DR) Prediction by the RuleFit Algorithm Using Routine Lab Results", Oklahoma State University, 2024<br>Publication | 1% |
| 7 | Submitted to Daffodil International University<br>Student Paper | 1% |
| 8 | www.science.gov<br>Internet Source | |