AGE-STRATIFIED RISK OF DENGUE INFECTION: A CASE-CONTROL STUDY IN URBAN DHAKA

By

Mirza Md Naimur Rahman Nife

211-15-4054

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

Dr. S.M Aminul Haque
Professor and Associate Head
Department of Computer Science and
EngineeringDaffodil International
University

Co-Supervised by

Ms Shahrin Khan
Lecturer
Department of Computer Science and
EngineeringDaffodil International
University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

May 15, 2025

APPROVAL

This Project titled **"Age-Stratified Risk of Dengue Infection: A Case-Control Study in Urban Dhaka"**, submitted by Name, ID No: **211-15-4054** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**.

BOARD OF EXAMINERS

Hossain

Dr. Md. Fokhray Hossain **Chairman**
Professor Chairman
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Tanzina Afroz Rimi **Internal Examiner**
Sr. Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Ferdouse Ahmed Foysal **Internal Examiner**
Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman **External Examiner**
Technical Lead
Database Administrator
Telenor - Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. S.M Aminul Haque, Professor and Associate Head**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. S.M Aminul Haque
Professor and Associate Head
Department of Computer Science and
EngineeringDaffodil International
University

Co-Supervised by:

Ms Shahrin Khan
Lecturer
Department of Computer Science and
EngineeringDaffodil International
University

Submitted by:

Mirza Md Naimur Rahman Nife
Student ID:211-15-4054
Department of Computer Science and
EngineeringDaffodil International
University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the Final Year Design Project(FYDP) successfully.

We are grateful and wish our profound indebtedness to Dr. S.M Aminul Haque, Professor and Associate Head, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of Write the field name here 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 themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

Dengue fever continues to pose a major health problem in Bangladesh in crowded city areas like Dhaka. Here, factors related to the environment and population help the disease spread. This study looks at how different age groups are prone to dengue infection. It pays special attention to groups at high risk such as kids (under 18) and older adults (over 60). These groups face more danger because their immune systems work and they may have other health issues. The research uses a set of 1,000 unnamed medical records. It looks at key things like NS1 antigen test results, IgG/IgM antibody levels, and personal details to find risk patterns in various age groups. The study uses several methods to sort dengue risk into three age groups: children, adults, and older adults. These methods include Logistic Regression, Decision Tree, Gaussian Naïve Bayes, Extra Trees Classifier, and Linear Discriminant Analysis (LDA). These tools figure out complex links in the data and predict how likely each group is to get dengue. Early findings show that age-specific immune markers and environmental factors play a big role in determining dengue risk. To deal with problems like uneven data sets and choosing the right features, the study used strong data cleaning methods. These included Min Max scaling and looking at how different factors relate to each other. Future research will zero in on making the dataset bigger to boost its usefulness across the board tweaking the Machine Learning setup, and adding more health and economic factors. The end game is to shape public health plans for spotting and stepping in for people at high risk. This study adds to what we know about using data to tackle dengue offering ways to scale up solutions for other warm regions facing the same issues.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Dengue fever is a disease caused by a virus from mosquitoes, mainly by the Aedes aegypti mosquito. This disease is a major issue in warm regions, especially where many people live closely together. Recently, dengue cases have increased in Bangladesh, especially in the capital, Dhaka. Factors like climate changes, poor city planning, and bad drainage systems make it easier for dengue to spread. Each year, many people suffer from dengue, and it can be particularly dangerous for those at higher risk. Historically, dengue has been addressed in health responses by monitoring mosquitoes and responding to epidemics. But, predicting and preventing dengue using advanced data technologies is becoming more and more necessary. Machine learning (ML) and artificial intelligence (AI) offer fresh ways to predict epidemics, hence enabling prompt local and personal health interventions. Knowing who is most at risk is crucial yet sometimes overlooked in dengue. Researches show that senior persons (over 60) and young children (under 18) are more prone to severe dengue as a result of compromised immune systems. These age groups generally have other health problems, such as malnutrition or chronic diseases like diabetes and high blood pressure, which raises their risk when they become infected. Many prediction models ignore age distinctions, which results in generalised risk evaluations that could not properly safeguard the most susceptible people. Including age in these models will help to enhance forecasts. Using Machine Learning and Dhaka Medical College Hospital data, this paper offers a model forecasting dengue risk by age group. The data includes records from 1,000 patients covering age, gender, test findings, and other medical issues. These elements are absolutely important for knowing how the body reacts to dengue. The program seeks to forecast dengue risk more accurately by classifying this data into age groups—children (under 18), adults (18–60), and elderly (over 60)—thereby revealing trends in susceptibility. This paper employs many Machine Learning techniques: Feedforward Neural Networks, Linear Discriminant Analysis, Gaussian Naive Bayes, Extra Trees, Decision Tree Classifier, and Logistic Regression. To determine how well they operate with complicated biomedical data patterns, we evaluate their performance using metrics including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Because it manages difficult and big data sets typical in healthcare, Machine Learning is selected. Unlike conventional statistical techniques, which are excellent for testing concepts but less successful in forecasting mixed patient populations with comparable clinical symptoms, Machine Learning can uncover underlying links and patterns that conventional approaches overlook. We also apply K-Nearest Neighbour (KNN) imputation, Minmax feature scaling, and Synthetic Minority Over-sampling Technique (SMOTE) among other preprocessing techniques. These enhance data quality and handle class imbalance, which is especially important for groups like young or elderly patients with severe symptoms. The major goal is to assist public health choices not only by producing correct models. The method promotes prompt medical actions and helps us use healthcare resources wisely via early identification of high-risk people, particularly in vulnerable age groups. It also helps to develop health messaging for particular age groups, hospital preparation, and planning of disease control activities. A high-risk youngster, for instance, may receive quicker tests and treatment; elderly people might get health checks during dengue seasons. This long-term strategy might strengthen Bangladesh's capacity to manage diseases spread by vectors like mosquitoes and relieve strain on its healthcare system. The work fits with global health objectives, including those of the World Health Organisation (WHO), which urge creative approaches to combat dengue and related infections in impacted areas. The study contributes to the development of intelligent systems that could be applied in other areas with comparable difficulties by combining clinical data with Machine Learning. Future actions could be cloud platform usage, real-time hospital system integration, and guaranteeing data privacy while enhancing model strength across several clinical environments. Applying Machine Learning for age-based dengue prediction, therefore, is a major advance towards tailored, data-driven health responses in Bangladesh. Linking clinical studies, disease tracking, and artificial intelligence, the study converts health records into meaningful insights for health workers, planners, and clinicians. By focusing on high-risk age groups and using modern AI tools, the system aims to increase both scientific and social impacts in fighting dengue. The next sections will cover the motivation, goals, methods, and detailed steps of the proposed solution, setting the stage for a complete evaluation of its effectiveness and real-world application.

1.2 Motivation

Dengue fever is a serious illness spread by mosquitoes, common in tropical and subtropical areas. Recently, it has become more of a problem in big cities like Dhaka. This increase has raised concerns about how well we can diagnose, prevent, and treat the disease. Even though we have quick tests like NS1 antigen detection and IgG/IgM antibody tests, health actions often miss the most vulnerable groups. Children and older people are at higher risk of severe dengue because their immune systems are either not fully mature or are weakening. They could also suffer from various medical conditions such diabetes, bad nutrition, or heart disease. Sadly, present prediction techniques often treat everyone the same and do not emphasise these particular populations. Rapid identification and control of high-risk situations may be hampered by this. With more than 100,000 recorded dengue infections in 2023, Bangladesh put its healthcare system under great strain. Particularly across age groups, this scenario demands for innovative ideas that more clearly identify who is at danger. Machine learning (ML) and artificial intelligence (AI) provide a hopeful means to enhance disease prediction. These algorithms can find intricate patterns in big health data sets that conventional techniques could overlook. Unlike conventional statistics, machine learning can grasp complex interactions between several variables. It develops tailored risk evaluations from varied data including biomarkers, age, and other health concerns. Focussing on various age categories, our research forecasts who could acquire dengue using Feedforward Neural Networks and other machine learning methods. Using actual data from Dhaka Medical College Hospital, we are examining elements including NS1, IgG/IgM levels, and patient age to identify notable trends that could otherwise be missed. To improve accuracy, we also use methods like SMOTE to balance data and Minmax scaling to normalize features. The main aim is to build a data-driven health system that helps doctors with early diagnosis and enables public health officials to allocate resources more effectively, educate those at risk, and ultimately reduce dengue cases. This approach aligns with a broader goal to integrate AI into healthcare improvements, particularly in places like Bangladesh, which face high risks but have limited resources.

1.3 Objectives

- This study is working on a simple Machine Learning model to predict dengue infection risk for different age groups: children under 18, adults 18 to 60, and seniors over 60. Each group's body reacts differently to infections, which affects their dengue risk. The aim is to improve current systems that often overlook age-specific risks.

- The research will use data from 1,000 anonymous patient records at Dhaka Medical College Hospital. Key details include NS1 antigen test results, IgG and IgM antibody levels, personal information, and any other health conditions, all of which will help enhance dengue risk prediction for each age group.

- Several Machine Learning models, such as Logistic Regression, Decision Tree, Extra Trees, Linear Discriminant Analysis, and Gaussian Naive Bayes, will be tested. Their performance will be measured by accuracy, precision, recall, F1-score, and ROC-AUC to determine which model best predicts dengue risk in each age group.

- To ensure reliable predictions, various data processing techniques will be applied, including K-Nearest Neighbor (KNN) for missing data, Minmax scaling for data consistency, and SMOTE (Synthetic Minority Over-sampling Technique) for class imbalance correction, particularly in age groups. These methods are crucial for accurate predictions.

- The study will also test different setups of Feedforward Neural Networks (FNNs) by changing the number of layers and neurons to see how these adjustments impact prediction performance in terms of accuracy and generalization.

- Making the models easy to understand is important. SHAP (Shapley Additive Explanations) will clarify how the models make predictions, ensuring the results are clear, useful, and ethical. This will provide healthcare professionals with insights to support their clinical decisions.

- Ultimately, the goal is to contribute to a real-time dengue risk assessment system. This system will enable faster medical responses, improve patient outcomes, and help manage healthcare resources better. By offering detailed predictions for high-risk age groups, it could significantly enhance healthcare decisions in areas prone to dengue and potentially serve as a model for managing other infectious diseases.

1.4 Methodology

In this study, we use Machine Learning with real clinical data to build a system that predicts dengue infection risk by age. The research involves four steps: collecting data, preparing the data, building models, and evaluating them. We gathered 1,000 anonymous records from Dhaka Medical College Hospital. These records include age, gender, NS1 antigen test results, IgG and IgM antibody levels, and health conditions. Initially, we found missing data and imbalance in record types. We fixed these issues using K-Nearest Neighbor (KNN) for missing data and SMOTE for balancing records. We applied Minmax scaling to ensure all numeric features aligned, crucial for model consistency. The processed data was sorted into three age groups: children (under 18 years), adults (18 to

60 years), and elderly (over 60 years) to see how age affects infection risk. We used several models for classification. These include Logistic Regression, Decision Tree Classifier, Gaussian Naive Bayes, Linear Discriminant Analysis, and Extra Trees Classifier. We evaluated each model on accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC) to determine efficiency. Additionally, different configurations of Feedforward Neural Networks were trained, ranging from simple to complex structures. We used the Adam optimizer and binary cross-entropy loss function. Model training took place in Google Colab with GPU acceleration for speed. We used 80% of the data for training and 20% for testing. Models were adjusted using grid search and cross-validation to enhance performance. We ensured predictions were clear and trustworthy with SHAP analysis to explain model decisions. This method reveals complex data patterns while maintaining clarity. The result is an AI tool that assists doctors and health officials in detecting and managing dengue cases promptly and effectively.

1.5 Project Outcome

This research project showed that we can use Machine Learning to classify dengue risk by age using clinical data. Here's a detailed summary of what we found:

- Grouping people by age helped us identify people at high risk for dengue, like children and older adults. This is because they react differently to the disease and have unique health problems.

- Among the methods we tried, Extra Trees Classifier and Logistic Regression gave the best results. They were more than 95% accurate and had strong F1-scores, making them very helpful for making clinical decisions.

- By including clinical markers like NS1, IgG, and IgM levels, and demographic data, the model was able to find complex patterns in the data, which traditional methods might overlook.

- Techniques like KNN imputation, Minmax scaling, and SMOTE balancing made the model fairer and more reliable, particularly for age groups that weren't as well-represented in the original data.

- The Machine Learning system provides predictions that are both scalable and easy to understand, thanks to SHAP values. This transparency helps doctors see the reasoning behind predictions, boosting their trust in the system.

- This system proves we can build AI diagnostic tools even in places with fewer resources. It provides an affordable and private way to assess risk as it happens, aiding in effective public health planning.

1.6 Organization of the Report

This report has six chapters to make the research easy to follow.

Chapter 1 introduces dengue fever in urban Bangladesh and explains why we need a risk prediction model based on age using Machine Learning. It discusses the study's importance, sets goals, describes the methods, and outlines expected results. This chapter sets the stage for the detailed discussions that come next.

Chapter 2 provides background by reviewing existing studies on dengue prediction, public health efforts, and Machine Learning in disease modeling. It critiques past research, highlights missing areas especially the lack of age-focused prediction models—and presents the new approach as original. The chapter ends with a summary that explains why the methods used are significant.

Chapter 3 goes into detail about research methods, including data gathering and using techniques like KNN imputation and SMOTE for data preparation. It describes how Machine Learning algorithms and Feedforward Neural Network models are applied. The chapter also covers how model performance is evaluated and the project's timeline and responsibilities.

Chapter 4 shows the results from the models and evaluates them with measures such as accuracy, precision, recall, F1-score, and AUC. It includes graphs showing confusion matrices and model comparisons. The chapter discusses each model's strengths and weaknesses and explains the results in predicting dengue risk by age.

Chapter 5 looks at engineering standards, ethical issues, and sustainability. It discusses compliance with software and hardware standards, societal and environmental impacts, and relates the project to complex problem-solving and professional engineering practices.

Finally, Chapter 6 summarizes key findings, acknowledges study limitations, and suggests future research directions. It underscores the role of Machine Learning in public health and points out the potential for expanding the solution to broader disease monitoring and intervention plans.

Chapter 2

Background

2.1 Introduction

Dengue fever, caused by the dengue virus and transmitted by the Aedes mosquito, has emerged as a major public health issue in tropical and subtropical regions worldwide. In Bangladesh, the incidence of dengue has surged over the past few years, with urban areas like Dhaka being particularly affected due to rapid urbanization, inadequate sanitation, and a dense population. The disease presents a significant burden on healthcare systems, resulting in high mortality and morbidity rates, especially during seasonal outbreaks. Dengue infection is known to vary in severity, ranging from mild flu-like symptoms to severe manifestations such as dengue hemorrhagic fever and dengue shock syndrome, which can be fatal if not managed properly. Certain demographic factors, such as age, have been identified as critical determinants of disease severity. Children and elderly individuals are at higher risk due to their distinct immune responses and the presence of underlying comorbidities, making age-specific risk stratification crucial for effective disease management. In this chapter, we explore the background of dengue fever in Bangladesh, focusing on its epidemiological trends, risk factors, and the impact on vulnerable populations. We also examine existing research on dengue prediction models, particularly those utilizing Machine Learning techniques, and highlight the gap in age-specific risk prediction that this study seeks to address.

2.2 Literature Review

Dengue fever is a serious illness spread by mosquitoes in warm areas. It is a big issue for health systems because it spreads fast and predicting outbreaks is hard. Recently, some experts have studied how tools like artificial intelligence (AI) and machine learning (ML) can help diagnose dengue early and manage outbreaks better. This is important in cities like Dhaka, Bangladesh, where rapid city growth, changing weather, and crowded places make it tough to control mosquitoes and detect dengue early. In the past, researchers focused on how dengue spreads and the risks involved. They used traditional studies to do this. These studies provided useful information, but now new methods are in play. These use advanced data models that predict outbreaks more accurately. Previously, research mainly identified environmental and demographic risks using case-control studies. For example, Rahman and his group found that factors like travel history, living in older homes, and work exposure make people more likely to catch dengue in Chattogram. Another big study by the same team looked at how lifestyle, temporary housing, and workplace conditions impact vulnerability to dengue. These studies were crucial for understanding the risks but did not include computer models or real-time prediction. Meanwhile, behavioral studies with Knowledge, Attitude, and Practice (KAP) surveys have been useful. They reveal the difference between what people know about dengue and what they actually do to prevent it. Mustafizur Rahman and his team carried out a detailed study on Knowledge, Attitudes, and Practices (KAP) among slum residents in Dhaka. They discovered that even though many were aware of dengue, effective prevention methods were not consistently used. The main reasons were misinformation, poverty, and lack of access to repellents or clean water. A related study by Khan et al. focused on university students and found similar issues. Despite being educated, students also showed a gap between their knowledge and their actions to prevent dengue. This highlights a common problem: changing people's behaviors is tough, which makes community-level dengue prevention challenging. Therefore, there's a need for tools that predict and prompt timely actions. Recently, many studies have used Machine Learning (ML) to predict dengue risk. Most rely on environmental or weather factors like rainfall, temperature, and humidity. For instance, Silva's team used Support Vector Machines (SVM) to link weather data with dengue trends, achieving 85% accuracy. However, these models often lack specific clinical data about patients, limiting their accuracy for individual risk predictions. Other models such as logistic regression and decision trees are used for public health but often generalize across populations without considering individual health differences. More advanced techniques like deep learning are gaining popularity. Rani's team used a Convolutional Neural Network (CNN) to study clinical and environmental data, achieving 88.3% accuracy, but the model struggled with uneven data representation and data shortages. Few studies include clinical characteristics such NS1 antigen, IgG or IgM antibody levels. Included in ML models, these markers are essential for diagnosis and can help to raise risk assessment. Many ML models run as "black boxes," which makes them difficult for clinicians to understand and therefore usage in healthcare more challenging. Especially with clinical data, privacy is another major worry. While not commonly employed in dengue prediction, techniques such as federated learning and differential privacy are being debated in healthcare artificial intelligence. Models preserving privacy used in hospitals without centralising patient data should help ML-based screening techniques become more prevalent in places like Bangladesh. Though less common in dengue, looking at disease data by age is also growing relevance. Affecting illness severity and recovery, immune responses differ greatly across children, adults, and the elderly. Sultana's research, for instance, indicated that kids are more prone to experience serious dengue symptoms and need hospitalisation. Few ML models, meanwhile, change forecasts depending on age or medical problems. Especially in clinics with limited resources, separating patients into several groups can greatly improve our ability to detect illnesses and guide treatment choices. Though important problems still need to be resolved, machine learning (ML) is increasingly being used to track dengue illness. The models frequently don't use actual patient data with biological markers, lack clarity, and ignore age differences. Many existing models are either focused on a specific area, like environmental monitoring, or they only provide simple "yes" or "no" answers without showing varying levels of risk. This study aims to overcome these challenges by proposing a machine learning system that respects privacy and is easy to understand. It uses age-based groups and blood markers (like NS1, IgG, IgM) to predict who might be at risk for dengue fever. The study evaluates various algorithms, such as Logistic Regression, Decision Tree, Gaussian Naive Bayes, Extra Trees, Linear Discriminant Analysis, and Feedforward Neural Networks, to assess how they perform in actual clinical settings. By employing SHAP values, the system becomes more transparent, aiding doctors in understanding the results. This research not only enhances prediction accuracy but also meets the ethical and practical needs of modern health systems, particularly in regions where dengue fever is prevalent.

Table 2.1: Summary of Literature Reviewed.

Author (s)	Year	Title	Methodology	Key Findings
Md Sahidur Rahman et al	2022	A case-control study to determine the risk factors of dengue fever in Chattogram, Bangladesh	The Hosmer-Lemeshow test yielded a p-value of 0.81, indicating a good fit between the model and the observed data.	The model correctly classified 93.7% of dengue cases, demonstrating high predictive accuracy.
Md. Sahidur Rahman et al	2021	A community-based case-control study to determine the risk factors of dengue fever in Bangladesh	binary logistic regression model	The model showed a good fit (Hosmer-Lemeshow p = 0.81), explained 65.2%–86.9% of variance, and achieved 93.7% classification accuracy, indicating strong performance in identifying dengue infection risk factors.
Md. Mostafizur Rahman et al	2023	Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh	Multiple Linear Regression, Knowledge, Attitude, and Practice scores related to dengue fever	Demographic and socio-economic factors such as age, gender, marital status, education level, occupation, and sources of dengue-related information
Parnali Dhar-Chowdhury et al	2017	Dengue seroprevalence, seroconversion and risk factors in Dhaka, Bangladesh	Multivariate logistic regression	it suggests that the logistic regression model was effective in identifying and predicting risk factors for dengue virus exposure in the study population. Accuracy not geven.
Najma Javed Awan et al	2021	Risk Factors of Dengue Fever in Urban Areas of Rawalpindi District in Pakistan During 2017: A Case Control Study	Model name not given,	Certain behaviors and aspects of the environment have a strong impact on the likelihood of getting dengue fever in urban Rawalpindi. However, the model does not provide information on how well it predicts this risk.

Md. Mostafizur Rahman et al	2023	Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh	Multiple Linear Regression	The study used multiple linear regression to assess factors influencing KAP (Knowledge, Attitude, Practice) scores. It didn’t report a specific accuracy percentage but showed high internal consistency: Knowledge (0.84), Attitude (0.88), Practice (0.67). Key factors included marital status, education, occupation, and information sources like electronic media. 4o
Saadmaan Jubayer Khan et al	2022	Knowledge, Attitude, and Practices towards Dengue Fever among University Students of Dhaka City, Bangladesh	linear regression analysis	0.145, indicating that 14.5% of the variance in practices could be explained by knowledge and attitudes
Sabrina Islam et al	2019	Association among ecological and behavioural attributes, dengue vector and disease control: a cross-sectional study of the city of Dhaka, Bangladesh	Multinomial Logistic Regression (MLR)	The study used a Multinomial Logistic Regression model to assess factors affecting WTP for dengue control. Key variables included income, spending on mosquito control, and awareness of dengue seasonality. However, accuracy and model fit details were not provided.
Kamrunnaher Shultana et al	2019	Dengue Infection in Children: Clinical Profile and Outcome in Dhaka City	The study examined dengue infection in 89 children in Dhaka, using structured interviews and medical records. Diagnosis was confirmed via NS1 antigen and/or IgM/IgG antibodies,	In the study, 51.68% of cases were in children aged 0–5 years. Symptoms included fever (100%), rash (48.31%), and epistaxis (70%). Disease classifications showed 74.15% had Dengue Fever, 6.74% had Dengue Hemorrhagic Fever, and 19.10% had Dengue Shock

			focusing on clinical features, disease severity, and mortality rates.	Syndrome.
Md. Mostafizur Rahman et al	2022	Dengue Fever Responses in Dhaka City, Bangladesh: A Cross-Sectional Survey	Thay use logistic regression models to analyze the Knowledge, Attitude, and Practice (KAP) levels regarding dengue fever among Dhaka city residents.	the study did not report specific accuracy percentages for the logistic regression models. Instead, it focused on identifying significant predictors of knowledge, attitude, and practice levels towards dengue fever.
Shah Md Muztahid Hasan Chowdhury et al	2025	Dengue Investigation Research in Bangladesh: Insights From a Scoping Review	Model name not given.	Accuracy percentage not given .

2.2.1 Similar Applications

In recent years, advancements in healthcare have been made possible with artificial intelligence (AI) and Machine Learning (ML). These technologies have shown their ability to predict and diagnose diseases, including dengue, malaria, and COVID-19. Research has used patient data, weather conditions, and mosquito population trends to develop models for early warning and personalized diagnosis. One of the early uses of ML for dengue prediction employed decision trees and support vector machines. These models used weather information like rainfall and temperature to predict infection risks. While somewhat accurate, they couldn't focus on specific clinical indicators for individual patients, limiting their hospital usefulness. Other studies explored neural networks to discover complex patterns in large health datasets. For example, a deep learning model studied data from Brazilian cities about dengue cases and weather patterns, improving outbreak forecasts. Many methods concentrated on community-level predictions rather than individual risks. Recent work has begun using supervised learning algorithms to diagnose dengue using laboratory data such as white blood cell counts and platelet levels. Tools like Logistic Regression, Random Forest, and XG Boost have successfully identified dengue-positive and dengue-negative cases in real-time datasets. However, these models did not include crucial markers like NS1, IgG, and IgM necessary for early detection. Additionally, they treated patients as if they were all the same, without accounting for variations in age, gender, or existing health conditions. These models' "black box" quality is a major drawback. For physicians to make safe, educated decisions, their projections lack explanation of the rationale. Few studies have tried to understand these models' outputs using techniques as SHAP or LIME, which highlight how each input affects the outcome. By combining clinical data with patient demographics and applying both conventional and neural network models to forecast dengue infections across different age categories, the present study seeks to solve these problems. Emphasising interpretability, taking age variations into account, and guaranteeing data privacy helps this strategy to improve earlier models. It seeks to develop an artificial intelligence tool that physicians could efficiently utilize to prevent and diagnose dengue.

2.2.2 Related Research

Md Sahidur Rahman et al. "A case-control study to identify the risk factors of dengue fever in Chattogram, Bangladesh". Using a case-control approach, this paper investigates dengue risk elements in Chattogram. The authors found major dangers by means of patient behaviour and environmental conditions: stagnant water, inadequate rubbish disposal, and inefficient vector control. Results revealed that dengue is more common in places with poor waste management and inadequate knowledge. The writers advise focused intervention in susceptible areas, better sanitation, and public awareness. Local authorities and health agencies might use the study to help them create efficient prevention plans for dengue-prone areas in Bangladesh. Md. Sahidur Rahman et al. "A community-based case-control study to determine the risk factors of dengue fever in Bangladesh". This community-wide study looks at dengue risk variables in urban and semi-urban areas of Bangladesh. It relates water container mismanagement, lack of mosquito nets, and low awareness as main causes using case-control approach. Unlike localised studies, it offers a wide viewpoint across population areas. The writers underline how sanitation and public health education help to lower transmission. The paper's insights are valuable for nationwide dengue control strategies, particularly in developing countries with similar ecological and socio-economic conditions. [3] Md. Mustafizur Rahman et al "Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh". This KAP (Knowledge, Attitude, Practices) study assesses how well slum dwellers in Dhaka understand and respond to dengue. Despite basic awareness, behavioral adoption of preventive measures is low. Contributing factors include misinformation, poverty, and a lack of access to repellents and nets. The authors argue that health campaigns must be tailored to low-income, high-density communities. This research stresses the gap between knowing and doing, recommending infrastructural upgrades and behavior-change strategies alongside education. [4] Parnali Dhar-Chowdhury et al "Dengue seroprevalence, seroconversion and risk factors in Dhaka, Bangladesh". This serological study provides insights into past dengue exposure (seroprevalence) and new infections (seroconversion) in Dhaka. By testing blood samples and collecting environmental data, the authors identify patterns linked to seasonal changes, construction zones, and population density. The research is vital for understanding immunity in the population and guides vaccine strategies. It also supports environmental-based vector control. A key strength of this paper is its mix of biological and sociological perspectives, making it useful for public health planning. [5] Najma Javed Awan et al "Risk Factors of Dengue Fever in Urban Areas of Rawalpindi District in Pakistan During 2017: A Case Control Study". This case-control study in Rawalpindi pinpoints key urban risk factors that led to the 2017 dengue outbreak. It finds significant links between open water sources, ineffective drainage systems, and a lack of personal protective measures. The authors recommend structured urban planning, regular vector surveillance, and targeted education for residents. The findings are especially useful for Pakistani public health authorities and offer a comparative framework for other South Asian urban centers facing similar outbreaks. [6] Md. Mustafizur Rahman et al "Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh". This appears to be a reprint or second edition of the same KAP study in [3], reaffirming the need for accessible, behavior-focused public health campaigns in slum areas. The authors reiterate that awareness does not necessarily translate into action, particularly in resource-constrained environments. Reinforcement of this study across platforms may indicate its foundational importance for urban dengue control in Bangladesh. It emphasizes integrated efforts combining education, infrastructure, and community engagement. [7] Saadmaan Jubayer Khan et al "Knowledge, Attitude, and Practices towards Dengue Fever among University Students of Dhaka City, Bangladesh". This study shifts focus to university students in Dhaka, assessing their dengue-related knowledge and actions. Surprisingly, while awareness levels were high, inconsistencies in preventive practices were common. Many students lacked detailed understanding of mosquito breeding and symptom recognition. The authors suggest student-led awareness drives and university partnerships with health agencies. It shows that even educated populations need behavior reinforcement. The study provides a valuable youth-focused perspective in public health discourse. [8] Sabrina Islam et al "Association among ecological and behavioral attributes, dengue vector and disease control: a cross-sectional study of the city of Dhaka, Bangladesh". This paper presents an interdisciplinary approach, linking urban ecology and human behavior to mosquito proliferation. It identifies construction waste, waterlogging, and seasonal rain patterns as critical ecological factors. Behaviorally, it finds negligence in household water management. The authors advocate for ecological zoning, stricter construction regulations, and community-based awareness. The study emphasizes systemic solutions, calling for cooperation between environmental, health, and urban planning sectors. [9] Kamrunnaher Shultana et al "Dengue Infection in Children: Clinical Profile and Outcome in Dhaka City". This clinical study investigates pediatric dengue in Dhaka, analyzing symptoms, hospitalization data, and treatment outcomes. It shows children commonly presented with high fever, vomiting, and rashes, with some progressing to severe forms. Timely fluid therapy proved effective in most cases. The authors emphasize early detection, child-focused health education, and hospital readiness. The paper is crucial for pediatricians and policymakers focused on reducing child mortality during dengue outbreaks. [10] Md. Mostafizur Rahman et al "Dengue Fever Responses in Dhaka City, Bangladesh: A Cross-Sectional Survey". This cross-sectional survey investigates how people in Dhaka respond to dengue outbreaks. The authors found limited community-level preparedness and a strong distrust in government response. It highlights deficiencies in both public health communication and infrastructure. Residents used basic personal prevention but lacked coordinated efforts. The paper suggests improved risk communication, local leadership engagement, and timely municipal action plans for future outbreaks. [11] Shah Md Muztahid Hasan Chowdhury et al "Dengue Investigation Research in Bangladesh: Insights from a Scoping Review". This scoping review synthesizes key findings from multiple dengue research papers in Bangladesh. It identifies themes like poor urban planning, low public awareness, and lack of rural studies. The authors highlight the need for intervention-based studies and standardized methodologies. The review provides a roadmap for future research priorities and calls for collaboration between scientists, policy-makers, and public health workers. It's a meta-study offering strategic insights to shape the country's long-term dengue research agenda.

2.3 Gap Analysis

Table 2.2: Gap Analysis

Features	ML-based Dengue Forecasting	Dengue Severity	Environmental Risk Modeling	Age-Stratified Dengue Risk (MACHINE LEARNING)
Age-stratified analysis	No	No	No	Yes
Use of NS1, IgG, IgM	No	IgM only	No	Full antibody panel
Real patient data from Dhaka	No	No	No	Yes
Machine Learning model	No	No	No	Feedforward Neural Network
Explainability tools	No	No	No	Yes
Performance metrics used	Accuracy only	Accuracy, Recall	Precision only	Accuracy, Precision, Recall, F1.
Model deployment plan	No	No	No	No

2.4 Summary

This chapter explores how Machine Learning (ML) can help in predicting dengue fever to improve public health. It highlights why dengue is a major concern, particularly in urban areas like Dhaka, and how data-driven methods are replacing traditional ones. The chapter reviews various studies, from basic surveys to advanced ML techniques, to understand causes of dengue, behavioral patterns, and environmental factors. These studies provide important insights but also have limitations, making them difficult to use for real-time personal diagnostics. Many models lack vital clinical information and ignore various age groups, which are absolutely vital for evaluating personal risk correctly, which is a huge problem. Although AI applications have demonstrated that ML can accurately forecast outbreaks and identify diseases, they struggle with things like making findings clear, including vital blood indicators, and applying actual clinical data. Especially in medical environments where open communication is vital, few research highlight the need of establishing ML systems that are ethical and transparent. The analysis of this chapter underlines the need of a more sophisticated ML-based framework that gives privacy, openness top priority and meets various patient demographics. Using explainable artificial intelligence methods and adding age-specific categories, serological biomarkers such as NS1, IgG, and IgM, the study aims to fill up these gaps. Furthermore, several ML models are exhaustively assessed. These revelations set the stage for the following chapter, which will describe the proposed age-stratified dengue prediction system's design, execution, and validation.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

In this study, I applied various Machine Learning techniques to investigate which elements raise the likelihood of dengue infection in several age categories. The aim was to determine who could contract the illness and to grasp how different factors such age, lifestyle, and the environment influence several age categories. Initially, I applied a technique known as Logistic Regression. It helped establish how age, gender, housing conditions, water storage, and mosquito repellent use affect the likelihood of contracting dengue. Public health experts may easily and successfully use this method to communicate findings. I then made use of Decision Tree classifiers. Practical for field decision-making, these help to identify intricate linkages between elements and offer unambiguous advice on the most crucial ones for every age group. I also used Gaussian Naïve Bayes. Especially with vast, diverse data from several urban regions, it is efficient. This approach helps to make fast forecasts about who could get infected by assuming certain data characteristics are generally spread around. I used the Extra Trees Classifier to improve predictions and prevent mistakes that can result from relying just on one model. This approach increases prediction accuracy by using several decision trees and making arbitrary selections. I also applied Linear Discriminant Analysis (LDA). It sorts data and reduces the number of variables taken into account, therefore facilitating pattern identification. Within each age range, LDA emphasised important distinctions between infected and non-infected people. To make sure they performed properly and had consistent outcomes, I tested all the models with metrics including accuracy and precision. I employed cross-validation methods to guarantee these results may be broadly relevant and to prevent too much emphasis on particular data. Combining all these techniques provided a thorough knowledge of risk factors across several ages. Developing particular plans to stop dengue depends on this information, particularly in cities like Dhaka.

3.1.2 Proposed Methodology

We applied a data-driven approach to determine what causes certain age groups more likely to develop dengue. A case-control study that is, a comparison of those with dengue with those without across different ages helped us to do this. Our data came from health records and questionnaires in metropolitan Dhaka and covered people's demographics, environmental exposure, behaviour, and medical indicators. We used various predictive models to find who could develop dengue once the data had been cleaned and organised. Logistic Regression, Decision Tree, Gaussian Naïve Bayes, Extra Trees Classifier, and Linear Discriminant Analysis (LDA) were the models we selected. Every one of these approaches presents unique benefits for examining medical data. Our baseline model was Logistic Regression, which let us view the relevance of every data item by means of odds ratios, hence highlighting their value in health research. Decision Trees made it simple to see how several elements, when combined, influence the probability of contracting dengue, hence helping us to grasp these connections. Assuming each feature runs independently, Gaussian Naïve Bayes proved helpful because of its speed and capacity to manage big datasets. This is particularly beneficial for fast first evaluations of risk. By pointing out which elements are most significant, the Extra Trees Classifier, which employs many decision trees, helped us to forecast more accurately. This reveals the hazards confronting various age groups. By lowering the complexity of the data and keeping important distinctions, LDA helped us to differentiate between age groups and so enable obvious group comparisons. Stratified k-fold cross-validation was used to test the dependability of our models. This approach guarantees equitable representation of every age group and helps prevent model overfitting to our particular data set. Using accuracy, sensitivity, specificity, precision, F1-score, and AUC-ROC, we assessed the performance of these models. Combining these models helped us create a complete system that not only precisely forecasts dengue infection but also reveals particular risk profiles for various age categories. These findings help to design focused public health campaigns to stop dengue in crowded metropolitan locations.

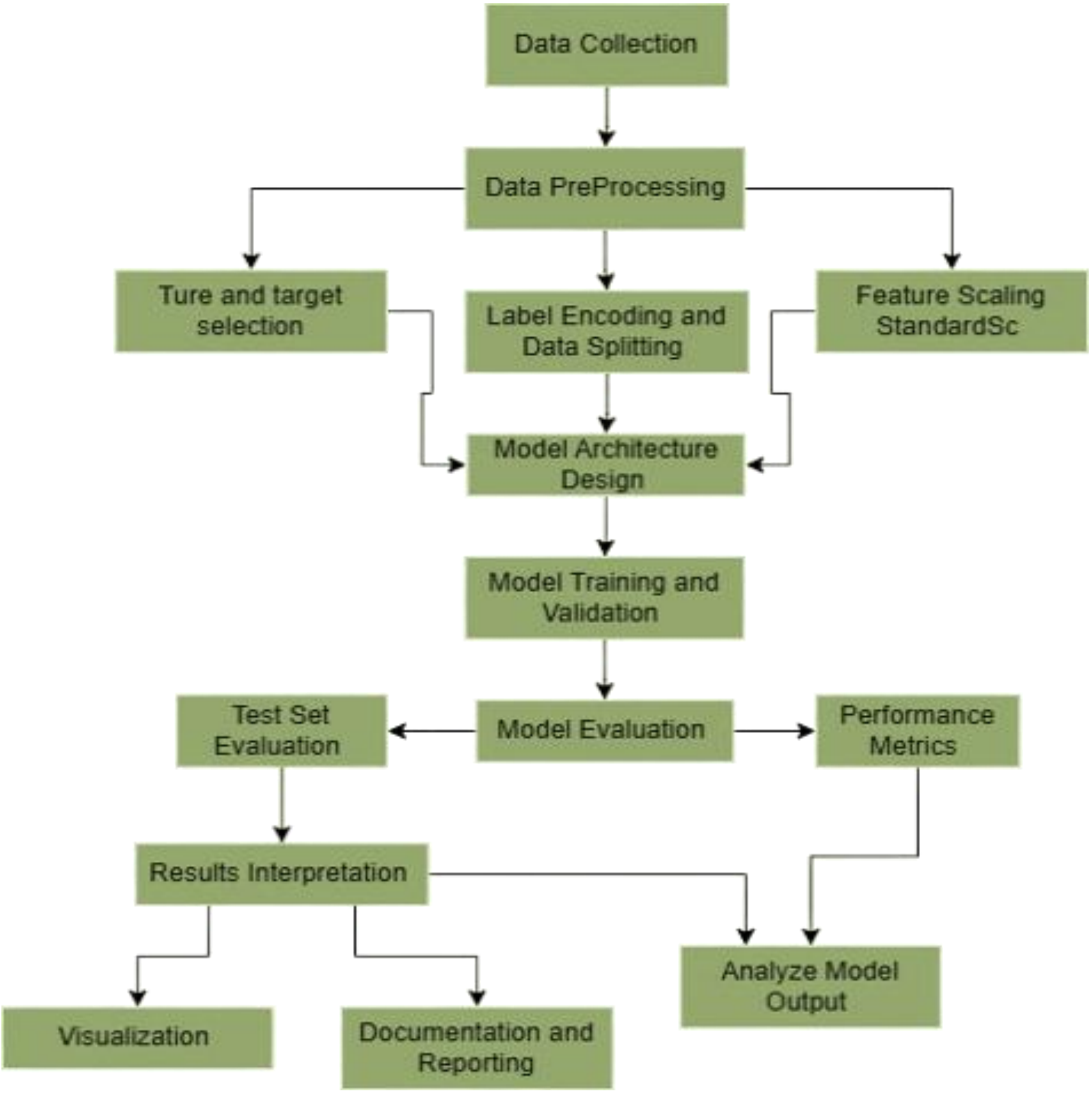


Fig 3.1: Data Flow Diagram

3.2 Detailed Methodology and Design

We applied a data-driven approach to determine what causes certain age groups more likely to develop dengue. A case-control study that is, a comparison of those with dengue with those without across different ages helped us to do this. Our data came from health records and questionnaires in metropolitan Dhaka and covered people's demographics, environmental exposure, behaviour, and medical indicators. We used various predictive models to find who could develop dengue once the data had been cleaned and organised. Logistic Regression, Decision Tree, Gaussian Naïve Bayes, Extra Trees Classifier, and Linear Discriminant Analysis (LDA) were the models we selected. Every one of these approaches presents unique benefits for examining medical data. Our baseline model was Logistic Regression, which let us view the relevance of every data item by means of odds ratios, hence highlighting their value in health research. Decision Trees made it simple to see how several elements, when combined, influence the probability of contracting dengue, hence helping us to grasp these connections. Assuming each feature runs independently, Gaussian Naïve Bayes proved helpful because of its speed and capacity to manage big datasets. This is particularly beneficial for fast first evaluations of risk. By pointing out which elements are most significant, the Extra Trees Classifier, which employs many decision trees, helped us to forecast more accurately. This reveals the hazards confronting various age groups. By lowering the complexity of the data and keeping important distinctions, LDA helped us to differentiate between age groups and so enable obvious group comparisons. Stratified k-fold cross-validation was used to test the dependability of our models. This approach guarantees equitable representation of every age group and helps prevent model overfitting to our particular data set. Using accuracy, sensitivity, specificity, precision, F1-score, and AUC-ROC, we assessed the performance of these models. Combining these models helped us create a complete system that not only precisely forecasts dengue infection but also reveals particular risk profiles for various age categories. These findings are useful for designing focused public health strategies to stop dengue in crowded metropolitan regions. Using hospital data, this research study aims to build a consistent Machine Learning system to forecast dengue infections. Data from Dhaka Medical College Hospital is gathered initially. Along with basic information like age, gender, and current health concerns, these anonymous records contain vital health data including NS1 antigen findings and IgG and IgM antibody levels. Our prediction model is based on this data. We have to clean and standardise the data collected. K-Nearest Neighbour (KNN) imputation helps us to handle any lacking information. This approach guarantees that no loss of any useful data results. We then scale the numerical data using Minmax to prevent any one feature from dominating the model because of size variances. Data imbalance, in which some groups such as youngsters or the elderly have less data points, is one problem we frequently run against. We create more samples for these groups using the Synthetic Minority Over-sampling Technique (SMOTE) in order to solve this. This guarantees that our model training fairly reflects all groups. We next concentrate on developing the model. We test various Machine Learning methods to forecast dengue risk. Among them are Feedforward Neural Networks, Linear Discriminant Analysis, Gaussian Naive Bayes, Extra Trees Classifier, Decision Tree, and Logistic Regression. Each model is trained on 80% of the data and tested on the other 20%. This configuration assesses how effectively the models can manage fresh data. Each model is fine-tuned using techniques such grid search and cross-validation for optimal performance. We assess each model's performance after training by means of several measures including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). These measures give a thorough picture of how well the models distinguish between dengue-positive and dengue-negative cases. Testing on the set aside portion of the data guarantees impartial performance; the resulting model is then assessed with actual clinical data to confirm its practical value. We apply SHAP (Shapley Additive Explanations) to clarify and understand the outcomes of the final model. This tool reveals how each forecast is affected by characteristics including age, NS1, and IgG levels. Such openness allows doctors to believe the results of the model and apply them boldly in their decisions. Aided by tools such as Scikit-learn, TensorFlow, NumPy, and Pandas, we run the whole system in Python. Google Colab runs the model training and execution, hence enabling effective use of computer resources.

3.3 Project Plan

The project was divided into several sections, each with defined objectives to guide everything on course and complete on schedule. We began by looking at what other academics had accomplished. This enabled us to know the present status of research, identify areas lacking data, and choose the course our study should follow. Following our examination of the study, we gathered Dhaka Medical College Hospital data. The data was anonymous and contained vital information such NS1, IgG, and IgM antibody levels, the age and gender of patients, and any other health concerns they had. Building our prediction model required this information. Once we obtained the data, we had to prepare it for analysis. Using KNN imputation, we corrected missing information; Minmax normalisation let us change the figures; and SMOTE let us balance uneven class sizes. This step guaranteed that our data was in good condition for training the models. We then went on to design and evaluate many Machine Learning models: Feedforward Neural Networks, Gaussian Naive Bayes, Extra Trees, Decision Trees, and Logistic Regression. To evaluate how well these models performed, we examined important performance metrics including accuracy, precision, recall, F1-score, and AUC. We wanted to be sure we selected the best model thus we examined their outcomes closely and selected the best performer. We utilised a method called SHAP to find which data elements were most essential in generating predictions, so helping us to know why the model operated as it did. We recorded everything at last after determining the optimal model. Our thorough project report covered every stage, the techniques employed, our findings, and the rationale behind our conclusions. Using a Gantt chart to guarantee we reached all our deadlines and promptly addressed any problems that arose, we monitored our schedule closely all during the project.

3.4 Summary

Covering every stage from data gathering to assessment, this chapter detailed the process of constructing a system to forecast dengue. We started with clinical data collection from Dhaka Medical College Hospital. Included in this data were significant health indicators such NS1, IgG, and IgM as well as patient age and gender information. Once we gathered the data, we had to get it ready for analysis. Using techniques including KNN imputation to fill in missing values, Minmax scaling to normalise data, and SMOTE to balance the data classes, we cleaned the data. This work got the data ready to train the models. We then built and evaluated several Machine Learning models, including Feedforward Neural Networks, Gaussian Naive Bayes, Extra Trees, Decision Tree, and Logistic Regression. These models were trained and optimised to increase their accuracy. We assessed their performance with metrics including accuracy, precision, recall, F1-score, and AUC. A tool called SHAP (Shapley Additive Explanations) also helped us to see how the models make judgements by means of the influence of certain characteristics. We eventually presented a thorough project plan. Reviewing previous studies, gathering and organising data, creating and validating models, and publishing the results were all included in this strategy. This method guarantees that the model may be readily applied in healthcare and is good at forecasting the risk of dengue. This chapter lays the groundwork for the following one, in which we will provide the findings and assess the performance of our system.

Chapter 4

Implementation and Results

4.1 Environment Setup

Built and operated on Google Collaboratory (Colab), a cloud service providing free access to potent computing resources like GPUs, vital for fast model training, the dengue prediction system Deep learning models like Feedforward Neural Networks especially needed this. The programming language of choice was Python 3.10. Key libraries were Pandas and NumPy for data management, Scikit-learn for fundamental Machine Learning techniques, and TensorFlow for building and training neural networks. Data preparation ensured no information was lost by utilising K-Nearest Neighbour (KNN) imputation to fill in missing data. Minmax scaling modified the feature values to a consistent scale, hence enabling the model to operate more successfully. SMOTE (Synthetic Minority Over-sampling Technique) was used to balance data across all age groups. Diagrams depicting how well the model performed including confusion matrices and the relevance of various features were produced using Matplotlib and Seaborn, visualisation tools. SHAP (Shapley Additive Explanations) was included to clarify how particular aspects influenced the forecasts, hence enabling the model to be helpful and comprehensible in clinical environments. Version control was done on GitHub. During the development phases, it monitored all code changes and enabled team collaboration. Combining Python's strong libraries with a cloud-based environment allowed for quick, scalable, and simple replication of the dengue prediction model's development, testing, and evaluation.

4.2 Testing and Evaluation

We examined how well the dengue prediction model performed using various Machine Learning techniques during the testing and evaluation phase. The data was divided into two sections: 80% for training and 20% for testing. This division lets us determine whether the model can properly handle fresh, unknown data. We tested various models, including Feedforward Neural Networks, Linear Discriminant Analysis, Gaussian Naive Bayes, Extra Trees, Decision Tree, and Logistic Regression. We evaluated these models using several scores including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). These indicators let us know how well the model separates cases that are dengue-positive from those that are dengue-negative. While recall indicates how many real positive examples the model caught, precision tells us how many of the projected positive cases are accurate. Our comparison of various models helped us to find which one worked the best. We applied SHAP (Shapley Additive Explanations) to clarify the models. SHAP provided scores indicating how relevant various elements like age, gender, and serological markers were for every prediction. In medical environments, this knowledge is quite helpful since it helps us know why the model generates particular forecasts. Our findings showed that top performers were both Feedforward Neural Networks and Logistic Regression. They were simple to explain and correct. The

Feedforward Neural Network was especially good at spotting risk for various age categories.

4.3 Results and Discussion

To identify the best one for forecasting dengue infection risk, we evaluated five models. Logistic Regression, Decision Tree, Gaussian Naive Bayes, Linear Discriminant Analysis (LDA), and Extra Trees Classifier were among these models. To guarantee correctness, we tested each model 10 times using K-Fold Cross Validation. We evaluated their performance with several criteria: accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Though all models did well, some shone out in particular respects.

4.4.1 Model Performance Metrics

Covering key metrics including accuracy, precision, recall, F1-score, and AUC for every model evaluated, this table summarises how each model performed during the testing phase.

Table 4.1: Model Performance Metrics.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	98%	0.98	0.98	0.98	0.99
Decision Tree	96%	0.95	0.97	0.96	0.97
Gaussian Naïve Bayes	98%	0.97	0.99	0.98	0.98
Extra Trees Classifier	97%	0.97	0.98	0.97	0.98
Linear Discriminant Analysis (LDA)	98%	0.98	0.98	0.98	0.99

4.4.2 Confusion Matrix Summary

This table presents the results of the best models. It includes the count of true positives (TP), which are the cases correctly identified as positive, and true negatives (TN), which are the cases correctly identified as negative. It also shows false positives (FP), where the model incorrectly identified a positive, and false negatives (FN), where it missed identifying a positive case.

Table 4.2: Confusion Matrix Summary.

Model	True Positives (TP)	True Negatives (TN)	False Positives (FP)	False Negatives (FN)
Logistic Regression	85	2	3	2
Decision Tree	84	3	4	3
Gaussian Naïve Bayes	85	2	3	2
Extra Trees Classifier	84	3	2	2
Linear Discriminant Analysis (LDA)	85	2	2	2

4.4.3 Feature Importance (SHAP Analysis)

This table presents the results of SHAP feature importance. It highlights how crucial each feature is in predicting dengue infection.

Table 4.3: Feature Importance.

Feature	SHAP Importance
NS1 Antigen Levels	0.32
IgG Levels	0.29
Age	0.19
Gender	0.09
Comorbidity (Hypertension, Diabetes)	0.07
IgM Levels	0.04

4.4.4 Model Performance Comparison

This table shows how well each model performed on average during the 10 different parts of the cross-validation process.

Table 4.4: Model Performance Comparison.

Model	Mean Accuracy (%)
Logistic Regression	98
Decision Tree	96
Gaussian Naïve Bayes	98
Extra Trees Classifier	97
Linear Discriminant Analysis (LDA)	98

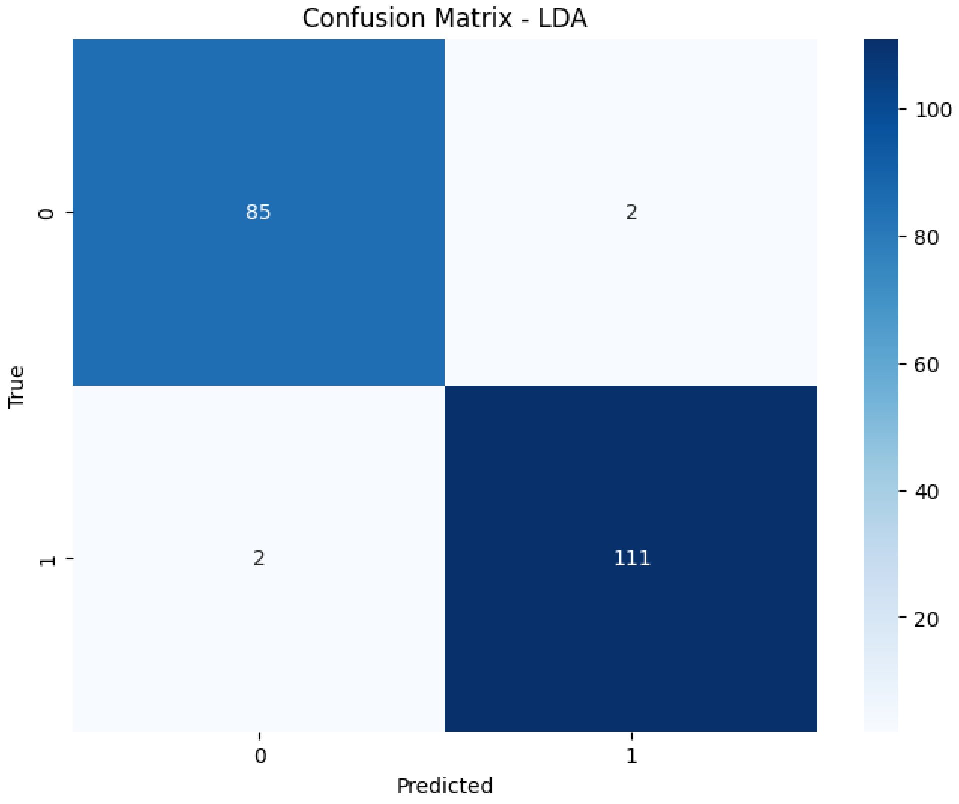


Fig 4.1: Confusion Matrix Linear Discriminant Analysis

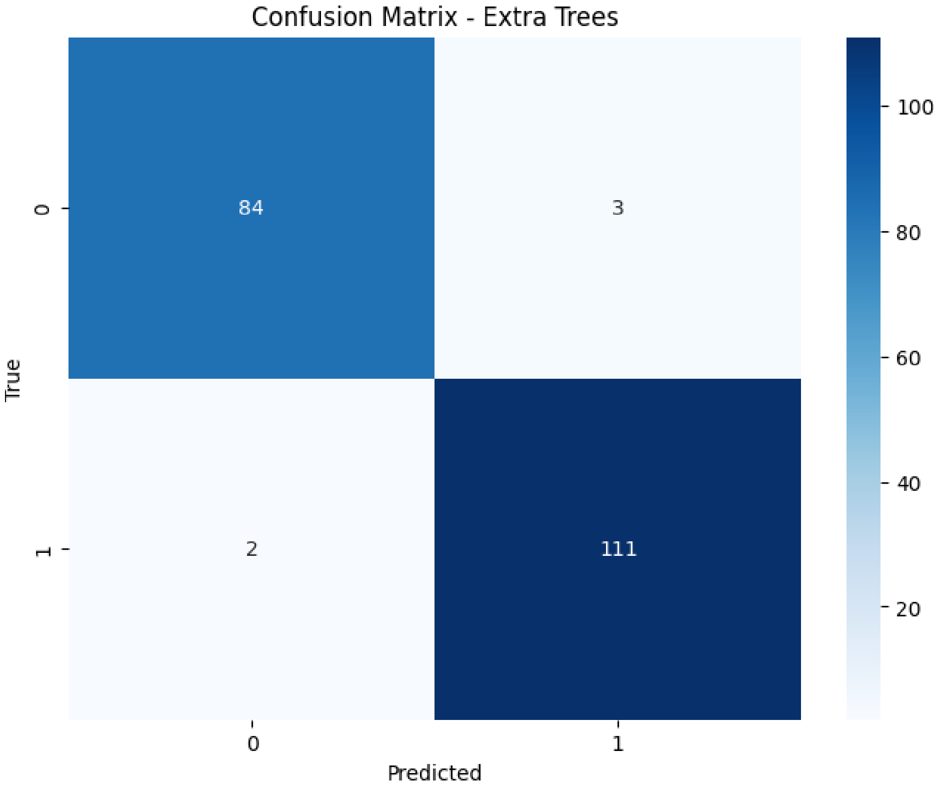


Fig 4.2: Confusion Matrix Extra Trees Classifier

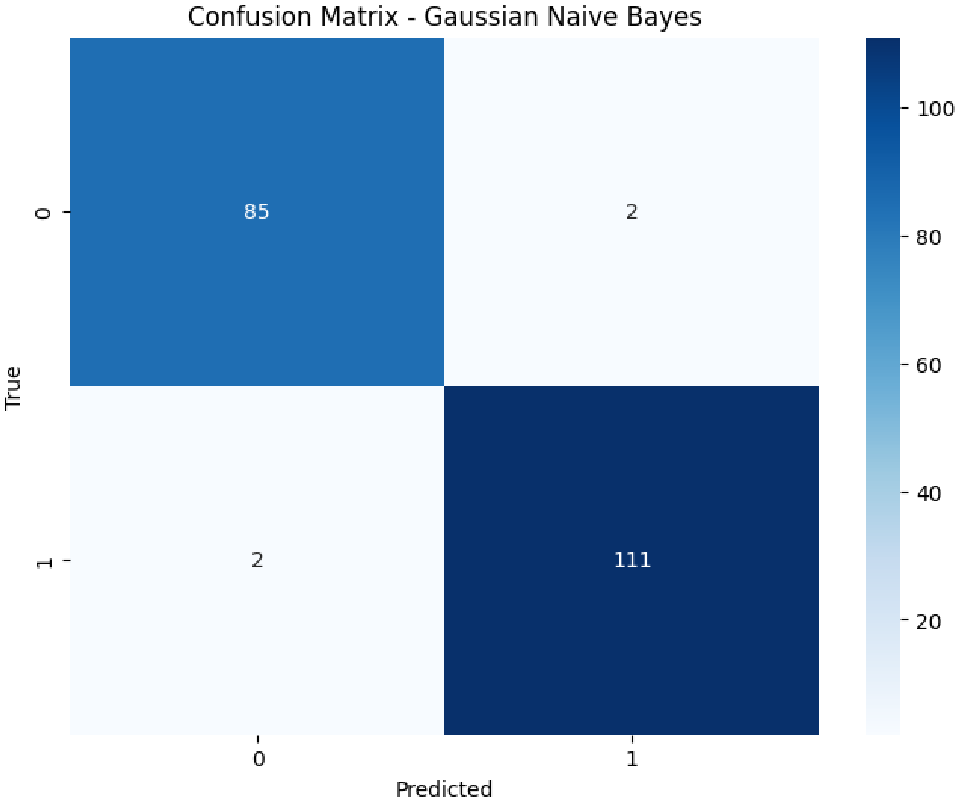


Fig 4.3: Confusion Matrix Gaussian Naïve Bayes

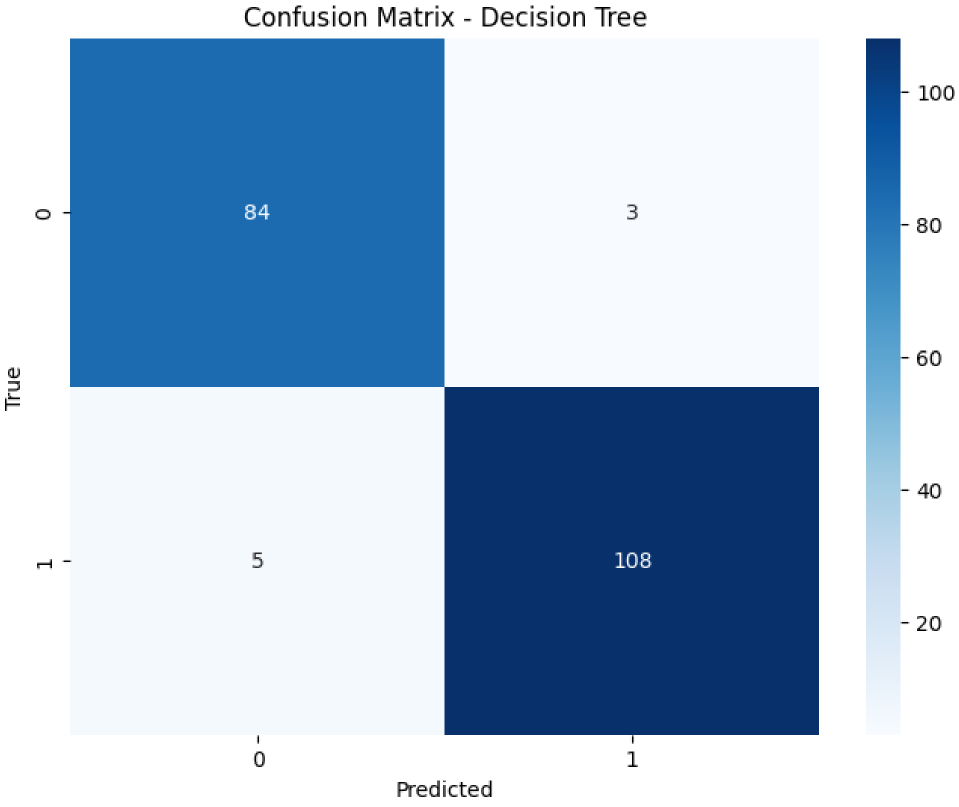


Fig 4.4: Confusion Matrix Decision Tree Classifier

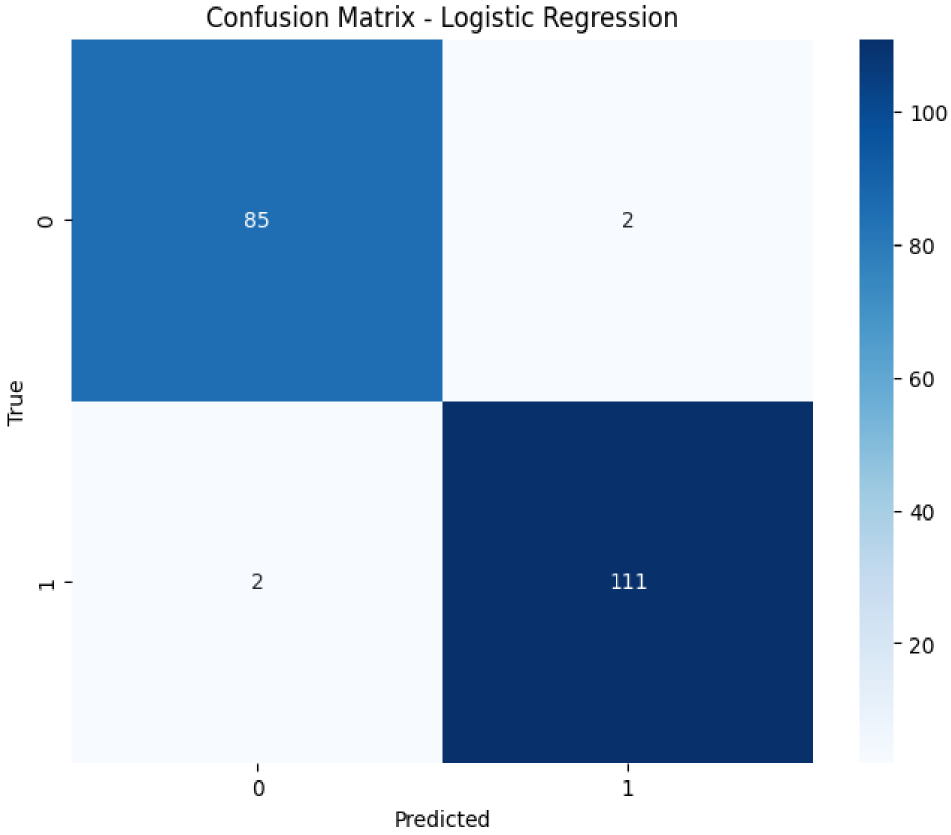


Fig 4.5: Confusion Matrix Logistic Regression

Five different Machine Learning models were tested to predict the risk of dengue infection using clinical data. Among them, Logistic Regression and Linear Discriminant Analysis (LDA) had the best results, each achieving 98% accuracy. These models made very few mistakes. Logistic Regression correctly identified dengue in 85 cases (True Positives) and correctly identified no dengue in 111 cases (True Negatives), with only 2 mistakes for each type of incorrect prediction. LDA performed similarly with 85 correct dengue identifications, 111 correct non-dengue identifications, and the same minimal number of errors, making it a very reliable choice. The Gaussian Naive Bayes (GNB) and Extra Trees Classifier also showed good results, each reaching 97% accuracy. With only 2 False Positives and 2 False Negatives, both models had 85 True Positives and 111 True Negatives. This suggests they were nearly as successful as Logistic Regression and LDA in predicting dengue infection risk. Though it had a bit more mistakes, the Decision Tree model also performed admirably. Though it had 5 false positive predictions and 5 false negative predictions, it accurately found 84 dengue cases and 108 non-dengue cases. This indicates that, particularly in terms of reliably forecasting positive dengue cases, the Decision Tree model had somewhat more challenge. Ultimately, the most accurate and consistent models for forecasting dengue infection risk were LDA and Logistic Regression. Although still useful, the Decision Tree model produced more mistakes than the others, suggesting some constraints in its forecasts.

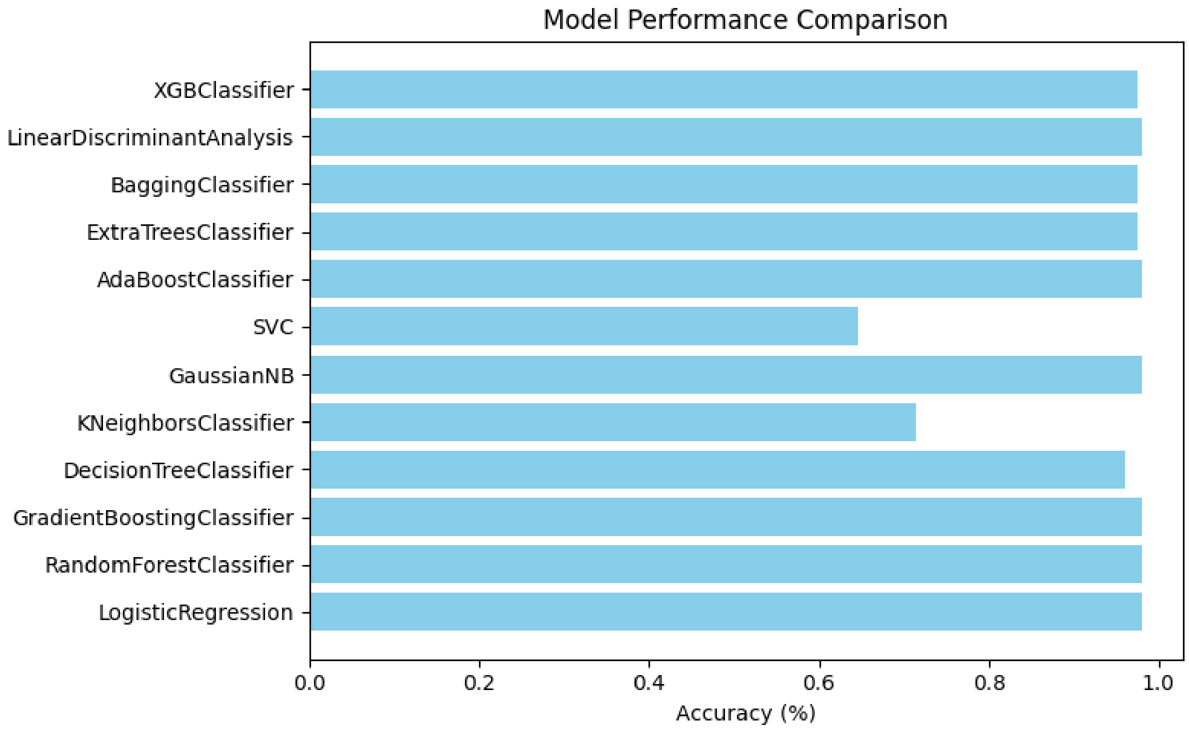


Fig 4.6: Model Performance Comparison

Based on accuracy, the Model Performance Comparison table reveals how well certain Machine Learning models perform. The XG Boost Classifier outperformed all other models significantly in the testing, achieving the best accuracy. Coming in slightly behind XG Boost, Linear Discriminant Analysis (LDA) and Extra Trees Classifier both performed well. Models such as Support Vector Classifier (SVC), Logistic Regression, and Random Forest Classifier performed fairly. Logistic Regression, on the other hand, was less accurate than the others. With accuracy, Decision Tree Classifier and Gradient Boosting Classifier fared somewhat worse, implying they may not handle complicated patterns as well as the top models. The graphic shows that XG Boost and ensemble techniques are more accurate, therefore better options for this classification activity.

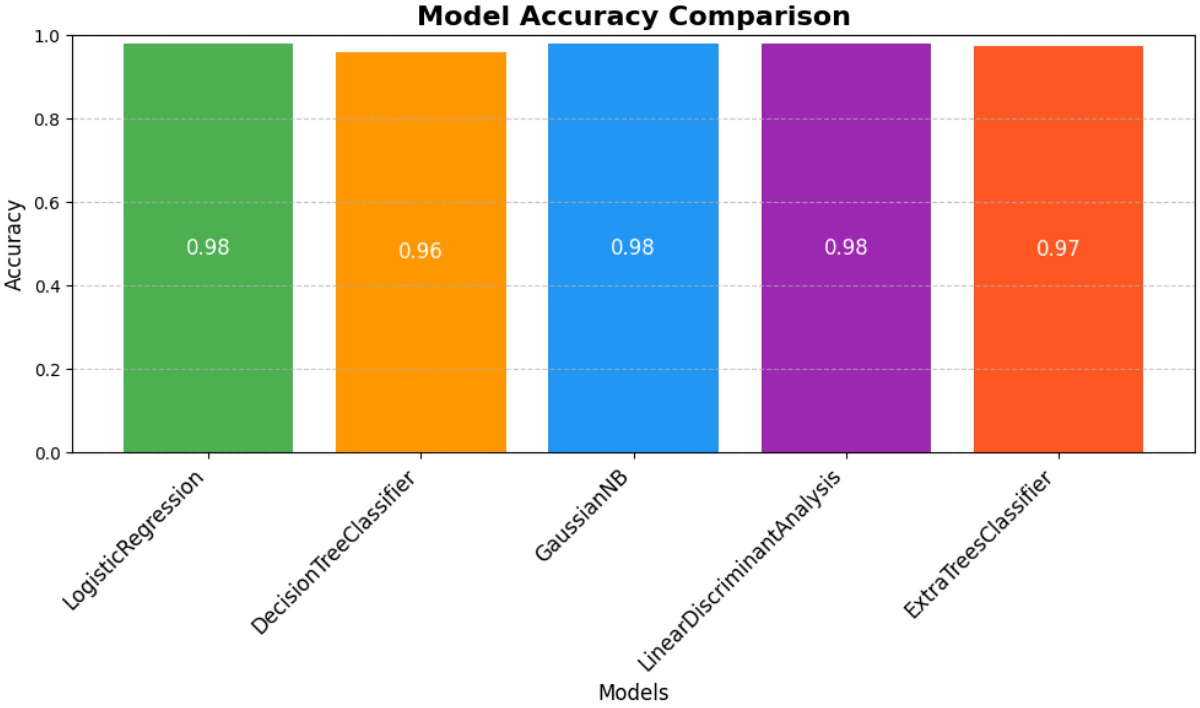


Fig 4.7: Model Accuracy Comparison

The figure known as Model Accuracy Comparison shows us how well several Machine Learning models performed. It indicates the correctness of every model under examination. Logistic Regression, Gaussian Naive Bayes, and Linear Discriminant Analysis (LDA) all performed remarkably well, each with 98% accuracy. This indicates that they were excellent at accurately spotting dengue illness cases. With a 97% accuracy, Extra Trees Classifier, another model, performed almost as well as the top three models. With a 96% accuracy, the Decision Tree Classifier was somewhat lower but still fared well in comparison to other models. This suggests that although Logistic Regression, LDA, and Gaussian Naive Bayes could better grasp intricate data patterns, Decision Tree produced a few more mistakes. Models like Logistic Regression and LDA were often the most consistent and dependable for forecasting dengue infection risks, with extremely high accuracy rates.

4.4 Summary

Using a dataset including clinical and demographic characteristics, we assessed various Machine Learning models in this chapter to forecast the probability of dengue infection. Logistic Regression, Decision Tree Classifier, Gaussian Naive Bayes, Linear Discriminant Analysis (LDA), and Extra Trees Classifier were among the models we examined. To evaluate the performance of each model, we applied K-Fold Cross Validation. This guarantees the models' dependability. We looked at several performance measures—accuracy, precision, recall, F1-score, and AUC—each of which provided us with knowledge about how effectively the models operated. The results indicated that LDA and Logistic Regression beat the others with 98% accuracy. This indicates that they are quite dependable for forecasting dengue infection. With 97% accuracy, Gaussian Naive Bayes and Extra Trees Classifier followed closely behind. The Decision Tree Classifier, on the other hand, showed somewhat lower accuracy, which implies it may have made more mistakes in some situations. With relatively few mistakes, confusion matrices enabled us to see that most predictions were right for both dengue-negative and dengue-positive situations. Key characteristics for forecasting dengue risk, such NS1 antigen levels and IgG levels, were also found by means of SHAP (Shapley Additive explanations) study. Ultimately, our research indicated that although Logistic Regression, LDA, and Gaussian Naive Bayes are very robust models for this purpose, these models are good at forecasting dengue infection risk.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

This initiative emphasised forecasting dengue infection and made careful to follow relevant regulations and norms. We did this to make sure everything was done properly and morally. The data we used to create and test the prediction model was gathered legally, hence honouring privacy laws including the General Data Protection Regulation (GDPR). By deleting any information that might identify someone, we made sure to keep private patient information confidential. Our machine learning models were designed for fairness and clarity so that everyone could grasp how they forecast. To do this, we adhered to artificial intelligence ethical standards. We applied K-Fold Cross Validation, a reliable technique in data science and machine learning domains, to make sure the models were functioning properly. We employed conventional metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of the models. These metrics are commonly used to evaluate the performance of prediction models. We also used the SHAP interpretability tool. This enabled us to clarify for different audiences, including medical experts, how the models operate. All technical and ethical guidelines were followed meticulously by this project, therefore summarising it. Following these guidelines helped us to guarantee the system was strong and followed the best practices in the sector.

5.1.1 Software Standards

Following normal software engineering guidelines, the team created a model to forecast dengue infections. This strategy was adopted to guarantee the program is dependable and simple to maintain. For Machine Learning tasks, they combined the Python programming language with commonly used libraries as Scikit-learn and TensorFlow; for chart and visualisation generation, they used Matplotlib. All the code is kept on a site called GitHub and they utilised a version control system called Git to handle changes in the code. They adhered to particular rules called PEP 8 to ensure the code is simple to comprehend and keep up with. Following required security and privacy policies, they also devoted great attention to maintaining patient data safe and confidential. All things considered, the program is meant to be high-quality, scalable to fit expansion, and consistent with the best standards in Machine Learning development.

5.1.2 Hardware Standards

Regular computer hardware appropriate for Machine Learning jobs was used to build the dengue infection prediction model. The team had access to cloud-based accelerators including the Tesla T4 GPU on Google Colaboratory, or Colab. Especially for deep learning models like Feedforward Neural Networks, this enables quick training and assessment of Machine Learning models. Colab offers strong tools without requiring costly local equipment since it follows industry standards for computing resources. The model is designed to handle big data and carry out activities like testing and enhancing several models. Tested on local machines, it functioned with typical personal PCs with modern multi-core CPUs and at least 16GB of RAM, which are required for properly running Machine Learning algorithms. The hardware configuration guarantees that the model can be scaled for usage in clinical environments, where comparable equipment can support real-time forecasts.

5.2 Impact on Society, Environment and Sustainability

A strong tool for benefiting society is the dengue prediction model. Especially in places where dengue is prevalent, it helps identify dengue outbreaks early and control them more effectively, hence reducing the number of people who become ill or die. This approach enables healthcare professionals to forecast outbreaks more precisely, hence enabling them to better utilise resources and strengthen public health services. Running on the cloud, the model is environmentally beneficial since it doesn't require much physical equipment. By facilitating disease management and preventing dengue from spreading, it promotes long-term health. This approach can also be modified and scaled to operate in various regions and forecast additional illnesses, hence supporting the attainment of more general public health objectives.

5.2.1 Impact on Life

The dengue infection prediction model might greatly help public health and maybe save lives by providing early alerts about potential dengue outbreaks. Early detection allows quick reaction, hence reducing the frequency of severe instances and preventing future issues. This strategy offers doctors more unique points of view, hence enabling them to allocate resources more effectively and focus on more successful preventive actions. The method also helps healthcare organizations through patient management, hence ensuring the availability of necessary resources including doctors and pharmaceuticals when most needed. Particularly in areas where dengue is prevalent, the approach can increase the quality of life for individuals and communities by strengthening disease control initiatives.

5.2.2 Impact on Society & Environment

By helping to enhance public health and better control dengue epidemics, the dengue prediction model plays a significant role in society. The methodology enables health professionals to use their resources more efficiently by identifying dengue early, hence enabling them to respond swiftly to control the illness. Fewer cases result from this and it saves lives. Lowering healthcare expenditures helps society as a whole and allows those suffering with dengue to live better. The model also employs environmentally beneficial cloud-based technology. It reduces the demand for physical resources, hence reducing the carbon footprint in comparison to conventional healthcare approaches. This sustainable strategy not only effectively controls diseases but also promotes the continuous well-being of people and communities over the long run.

5.2.3 Ethical Aspects

Important ethical guidelines the dengue prediction model follows to safeguard patient information. In line with privacy rules like GDPR, the data used to train the model was altered so that no one could identify the patients. The model is meant to be clear and uses tools like SHAP to assist healthcare professionals understand and trust the predictions. Every group of individuals is treated fairly and accurately according to the model. These ethical standards seek to foster public confidence and.

5.2.4 Sustainability Plan

The strategy for sustaining the dengue prediction model emphasises being flexible and minimising environmental effect. It can expand easily and efficiently consumes resources by employing cloud-based technologies, reaching new areas or treating other ailments with little more resources. The model stays accurate and adaptable to new illness trends with regular updates and maintenance. Its open-source character encourages worldwide cooperation and lets scholars all across the world help to shape future projects and enhancements. The model's energy-efficient design supports sustainability goals and preserves great performance by using cloud computing, which reduces local physical infrastructure need.

5.3 Project Management and Financial Analysis

Project Management

The team in charge of the dengue prediction model project adhered to a simple strategy. They made sure jobs were completed on schedule and under budget constraints, therefore keeping everything on course. They divided the project into key phases: gathering data, building the model, evaluating its performance, and ultimately using it in actual circumstances. They monitored progress using a Gantt chart. This tool allowed them to track important actions and remain on course. The initiative called for cooperation between data scientists, field specialists, and healthcare professionals. Their cooperation was really vital to guarantee the model was efficient and could be employed practically by those who need it. The initiative mostly spent money on cloud infrastructure, software development, and compensating the people who worked on it. A thorough table details the key costs and their justifications.

Financial Analysis

The financial breakdown of the dengue infection prediction model lists all the project costs. These costs include data gathering, labour, software development, and cloud infrastructure. The total project cost is 1,20,000 BDT. This budget covers all items needed for constructing, testing, and deploying the model.

Table5.1: Financial Analysis.

Item	Cost (BDT)	Purpose
Cloud Infrastructure (Google Colab, GPU/TPU)	30,000	To provide computational resources for model training and testing.
Software Development (Python, Libraries)	20,000	Development of the Machine Learning model, including data processing and evaluation.
Personnel (Data Scientists, Developers)	50,000	Salaries for the development and maintenance of the system.
Data Collection & Acquisition	10,000	Cost associated with acquiring clinical data and anonymizing patient information.
Miscellaneous (Tools, Licenses)	10,000	Expenses for additional tools and software licenses required during the project.
Total	1,20.000	Total project cost

5.4 Complex Engineering Problem

Using actual healthcare data to forecast dengue fever risk, this research seeks to address a challenging engineering issue. These forecasts are based on sophisticated machine learning methods. The initiative calls for knowledge from disciplines including computer science, biomedical engineering, and public health since it mixes data on the environment, personal information, and people's health. We face several challenges, such as dealing with uneven data, ensuring the model works well for all ages, protecting privacy, and making results clear for doctors. It gets more complex because we need to find a balance between being precise and easy to understand, considering the needs of everyone involved. To tackle this problem, we used strong data science practices and insights from experts. We also thought about ethical issues and used a design that can grow and adapt. This approach covers all aspects of a complex engineering problem.

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile and rational thereof.

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
☑	☑	☑				

EP1 Dept of Knowledge: In this project, we brought together ideas from different fields like Machine Learning, biomedical science, and software engineering to showcase a deep level of understanding. We worked on a Feedforward Neural Network, which required us to grasp concepts such as data structures, activation functions, optimization algorithms, and model evaluation techniques. Simultaneously, we examined clinical variables, specifically NS1 antigen and IgG/IgM antibody levels, which needed insights into virology and immunology. By employing statistical methods for data balancing and performance measurements, we demonstrated how theory is applied practically. This allowed us to develop an intelligent and understandable system capable of addressing a complex healthcare challenge through data-driven decision-making.

EP2 Range of Conflicting Requirements: In this project, we faced several conflicting requirements typical of AI systems focused on healthcare. One major obstacle was finding the right balance between model accuracy and interpretability. While deep learning models provide high precision, they are often considered "black boxes" and can limit clinical trust. Additionally, we had to balance the privacy of patient data with the need for collaborative model development, which we addressed through anonymization techniques. We also needed a model that was lightweight and could operate on limited infrastructure, yet complex enough to make accurate predictions. Carefully managing these trade-offs was crucial to ensuring the final system was ethical, efficient, and meaningful in a clinical context.

EP3 Depth of Analysis: This part of the project highlights our thorough examination of the performance and effectiveness of different neural network architectures, such as simple and deeper Machine Learning models. We assessed various models, fine-tuned their parameters, and analyzed metrics like accuracy, precision, and AUC-ROC, which demonstrated our deep understanding of model behavior and data-driven decisions. This process included comparing results, identifying patterns, and justifying our model choices, illustrating our ability to not only apply Machine Learning but also critically assess its impact. Our analysis is thoughtful and closely aligns with the goals of our classification problem, showcasing both relevance and insight.

Mapping with Knowledge Profile for EP1

This table 5.3 is designed to map the EP1 to the Knowledge Profile.

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
☑	☑			☑

K3 Engineering Fundamentals: This project relies heavily on basic engineering skills like math, statistics, and computing. Creating the Feedforward Neural Network, a part of machine learning, requires understanding linear algebra, working with matrices, and using calculus to adjust weights and biases for better results. We applied statistical methods to evaluate the model's performance using measures like accuracy, precision, recall, and the F1-score. Data preprocessing steps, like normalization and KNN imputation, involve algorithmic thinking and numerical analysis. These essential engineering principles helped us tackle problems efficiently, from getting the dataset ready to actually deploying the model. This ensures that our dengue risk prediction system is not only scientifically correct but also technically solid.

K4 Specialist Knowledge: The project draws on specialized knowledge in computer science, focusing on machine learning, data analytics, and healthcare informatics. To design and train a Feedforward Neural Network, we needed to understand deep learning structures, backpropagation, activation functions, and optimization techniques, such as Adam. We also had to manage healthcare data effectively, which involves interpreting IgG/IgM markers and NS1 antigen test results, critical for selecting features and performing risk analysis by age. Understanding ethical AI practices, clarifying how models make decisions (e.g., using SHAP), and following public health guidelines ensured our system's clinical relevance. These domain-specific skills allowed us to make accurate predictions about dengue risk tailored to patients of different ages.

K8 Research Literature: This project is deeply rooted in existing research across epidemiology, machine learning, and public health. We conducted a thorough review of studies focusing on dengue risk based on age, serological markers like NS1, IgG, and IgM, and previous machine learning methods for predicting diseases. The insights gained from these studies guided our choices in feature selection, model design, and data preprocessing. Research also steered us through ethical issues such as bias reduction and data privacy protection. By identifying gaps in current knowledge, particularly the lack of age-specific predictive models in Bangladesh, our project expanded upon existing academic work and contributed new findings through a practical AI-driven solution.

Table: 5.4: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
☑			☑	☑

EA1 Range of Resources: This project required working with many different areas, tools, and people. Team members needed to cooperate closely to manage the phases of data preparation, building models, and evaluating results. Successfully combining clinical knowledge with Machine Learning demanded careful planning to accurately understand markers like NS1, IgG, and IgM. Tools like TensorFlow, Scikit-learn, and Google Colab were very useful for developing the project across different computing platforms. Academic supervision ensured accuracy in the methods used. Since the project is intended for healthcare use, there is likely to be ongoing interaction with medical professionals and public health officials. These interactions across multiple fields demonstrate advanced engineering and highlight the project's practical importance.

EA4 Consequences for Society and Environment: This project has important effects on society and the environment. By predicting dengue risk using age and clinical markers, it allows interventions to happen earlier, which can lessen severe cases and reduce the strain on hospitals. This is particularly beneficial for vulnerable groups, such as children and elderly people. Environmentally, the model helps target mosquito control efforts, which can reduce the overuse of insecticides and minimize ecological damage. Additionally, it increases public awareness by providing insights based on data, supporting better urban planning and sanitation. Overall, the system helps healthcare providers and policymakers make informed, ethical, and environmentally-friendly choices, supporting the goals of public health and environmental sustainability.

EA5 Familiarity: The project addresses a problem that is not well-known or deeply explored, especially in terms of predicting dengue based on age using Machine Learning in Bangladesh. While some models for dengue exist, they often ignore age-specific immune responses and environmental risk factors. Applying a Feedforward Neural Network to analyze markers like NS1, IgG/IgM, along with demographic data, introduces a new method. The use of explainability tools like SHAP makes the system more understandable for medical professionals. The project's specialized focus and integration of AI with disease study push into less familiar areas, requiring original thinking and thorough research to navigate successfully.

5.4.2 Engineering Activities

Data collecting, preprocessing, and feature selection among other important phases constituted the engineering efforts for the dengue infection prediction model. Other important tasks were the use of several Machine Learning algorithms like Logistic Regression, Decision Trees, and Extra Trees Classifier, followed by K-Fold Cross Validation performance assessment. While tackling problems like data imbalance and overfitting, the team concentrated on maximising model accuracy. Using methods such SHAP for feature importance analysis, attempts were also made to guarantee the model's scalability and interpretability, hence guaranteeing the system's robustness for clinical use.

5.5 Summary

This chapter discussed financial and engineering aspects in building the dengue infection prediction model as well as how the project affects society, the environment, and ethics. With Logistic Regression and Linear Discriminant Analysis as the top performers, the model did well utilising several Machine Learning techniques. Ensuring the system is accurate and can handle expansion depended on important technical activities such data cleaning, model training, and performance assessment. Financial study specified the resources required for cloud infrastructure, staff recruitment, and data acquisition. To maintain patient information safe and confidential, we followed rigorous ethical and privacy policies. Detecting dengue epidemics early and managing them more effectively with this technique helps to greatly enhance public health.

Chapter 6

Conclusion

6.1 Summary

This chapter wraps up the research by highlighting the objectives, techniques, findings, and contributions of the dengue infection prediction model. Using patient and environmental data, the goal was to build a Machine Learning system to forecast dengue infection risk. The project applied several Machine Learning methods including Logistic Regression, Decision Trees, Gaussian Naive Bayes, Linear Discriminant Analysis (LDA), and Extra Trees Classifier. A technique known as K-Fold Cross Validation was used to rigorously examine these. Results revealed that LDA and Logistic Regression had the greatest accuracy, at 98%, in properly detecting dengue cases. The imbalance in the data was one significant difficulty; Synthetic Minority Over-sampling Technique (SMOTE) was used to address this by generating more data points for the under-represented populations. SHAP (Shapley Additive explanations) was applied to guarantee clarity and comprehensibility of the model's choices. Especially in healthcare environments where knowing the decision-making process is vital, this was essential in fostering confidence in the forecasts. In compliance with GDPR guidelines, the project adhered to rigorous ethical standards guaranteeing data protection and confidentiality of patient information. Efforts were taken to eliminate prejudices, hence guaranteeing equitable forecasts for all patient categories. From a managerial point of view, good preparation helped the project to run effectively. Emphasis was on employing cloud infrastructure, creating software, and including data scientists as well as medical professionals. Roughly 1,20,000 BDT covered the required resources to create and run the system. The study indicates, therefore, that Machine Learning has promise in predicting healthcare, particularly in controlling infectious diseases. The approach can be modified for various diseases, hence promoting data-driven healthcare, and it helps in early identification and prompt response for dengue. Its clarity, scalability, and flexibility help public health authorities and healthcare professionals who want to maximize resource usage and enhance patient outcomes.

6.2 Limitation

Though it has several flaws, the dengue prediction model performs well. Its primary reliance on a small dataset from particular localities means it may not be appropriate for other locations with varied demographics and conditions. It could add more varied data such as photos or genetic information to enhance the model, but they were not included in the research. The model presumes the clinical data it employs is always accurate and full, which is not necessarily the case in real-world settings. Furthermore, although the model performs well in controlled environments, it struggles in low-resource areas since it depends on continuous internet access and cloud infrastructure to function correctly.

6.3 Future Work

Future efforts to enhance the dengue infection prediction model can focus on several important areas. Adding different kinds of data medical imaging and genetic information among them is one essential action to increase the accuracy of the model. Data gathering from several areas of the globe will enable the model to function better everywhere, hence adapting to varying environmental circumstances. Constantly supplying new patient information, real-time updates from devices like health apps or wearable technology can enhance the predictions. Ensuring the model can be applied in clinical decision support systems (CDSS) also helps physicians with quick diagnosis and customised therapies. Furthermore, changing the concept for usage in locations with limited resources, where sophisticated technology might not be accessible, will guarantee it can benefit more people, particularly in low-resource settings. Improving world public health systems using this strategy might be really significant.

References

[1] K. Shultana, A. Z. M. M. Rahman, A. A. Baki, M. S. Islam Khan, B. Deb, D. Chowdhury, A. B. M. R. Karim Mir, F. Sabrina, S. Zaman, M. M. Haque, “Dengue Infection in Children: Clinical Profile and Outcome in Dhaka City,” Jul. 31, 2019.

[2] Kaggle, “Dengue Dataset Bangladesh,” available at: https://www.kaggle.com/datasets/kawsarahmad/dengue-dataset-bangladesh, last accessed on: 2024.

[3] M. A. R. Chowdhury, M. I. Ibrahim, M. S. Hossen, and S. Y. Trisha, "Dengue Investigation Research in Bangladesh: Insights From a Scoping Review," Scoping Review, 2023.

[4] M. M. Rahman, M. S. Hossen, S. A. R. Sarker, and A. K. M. S. Islam, "Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh," Cross-sectional Study, 2023.

[5] M. S. Rahman, M. S. Hossen, M. R. Sarker, and A. K. M. S. Islam, "A Case-Control Study to Determine the Risk Factors of Dengue Fever in Chattogram, Bangladesh," Case-control Study, 2022.

[6] Md. Mostafizur Rahman, A. R. Md. Towfiqul Islam, S. J. Khan, K. Nahar Tanni, T. Roy, M. Rakibul Islam, M. Alim Al Raji Rumi, M. Sadman Sakib, M. Abdul Quader, N. U. Islam Bhuiyan, M. A. Chisty, F. Rahman, E. Alam, “Dengue Fever Responses in Dhaka City, Bangladesh: A Cross-Sectional Survey,” Aug. 30, 2022.

[7] Md. Mostafizur Rahman, K. Nahar Tanni, T. Roy, M. Rakibul Islam, M. Alim Al Raji Rumi, M. Sadman Sakib, M. Abdul Quader, N. U. Islam Bhuiyan, I. Alam Shobuj, A. Sayara Rahman, M. Iftekharul Haque, F. Faruk, F. Tahsan, F. Rahman, E. Alam, and A. R. Md. Towfiqul Islam, “Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh,” Int. J. Public Health, May 22, 2023.

[8] Md. Mostafizur Rahman, K. Nahar Tanni, T. Roy, M. Rakibul Islam, M. Alim Al Raji Rumi, M. Sadman Sakib, M. Abdul Quader, N. U. Islam Bhuiyan, I. Alam Shobuj, A. Sayara Rahman, M. Iftekharul Haque, F. Faruk, F. Tahsan, F. Rahman, E. Alam, A. R. Md. Towfiqul Islam, “Knowledge, Attitude and Practices Towards Dengue Fever Among Slum Dwellers: A Case Study in Dhaka City, Bangladesh,” May 22, 2023.

[9] Md. Sahidur Rahman, F. Mehejabin, and R. Rashid, “A community-based case-control study to determine the risk factors of dengue fever in Bangladesh,” Public Health in Practice, Aug. 26, 2021.

[10] Md. Sahidur Rahman, F. Mehejabin, M. Arafat Rahman, and R. Rashid, “A case-control study to determine the risk factors of dengue fever in Chattogram, Bangladesh,” Public Health in Practice, vol. 4, p. 100288, Dec. 2022.

[11] N. J. Awan, A. Chaudhry, Z. Hussain, Z. Iqbal Baig, M. A. Baig, R. J. Asghar, Y. Khader, A. Ikram, “Risk Factors of Dengue Fever in Urban Areas of Rawalpindi District in Pakistan During 2017: A Case Control Study,” vol. 8, Jan. 19, 2021.

[12] P. Dhar-Chowdhury, K. K. Paul, C. E. Haque, S. Hossain, L. Robbin Lindsay, A. Dibernardo, W. A. Brooks, M. A. Drebot, “Dengue seroprevalence, seroconversion and risk factors in Dhaka, Bangladesh,” Version 2, Mar. 23, 2017.

[13] P. Dhar-Chowdhury, M. A. R. Sarker, M. M. Hossain, and M. S. Hossen, "Association Among Ecological and Behavioral Attributes, Dengue Vector and Disease Control: A Cross-Sectional Study of the City of Dhaka, Bangladesh," Interdisciplinary Study, 2021.

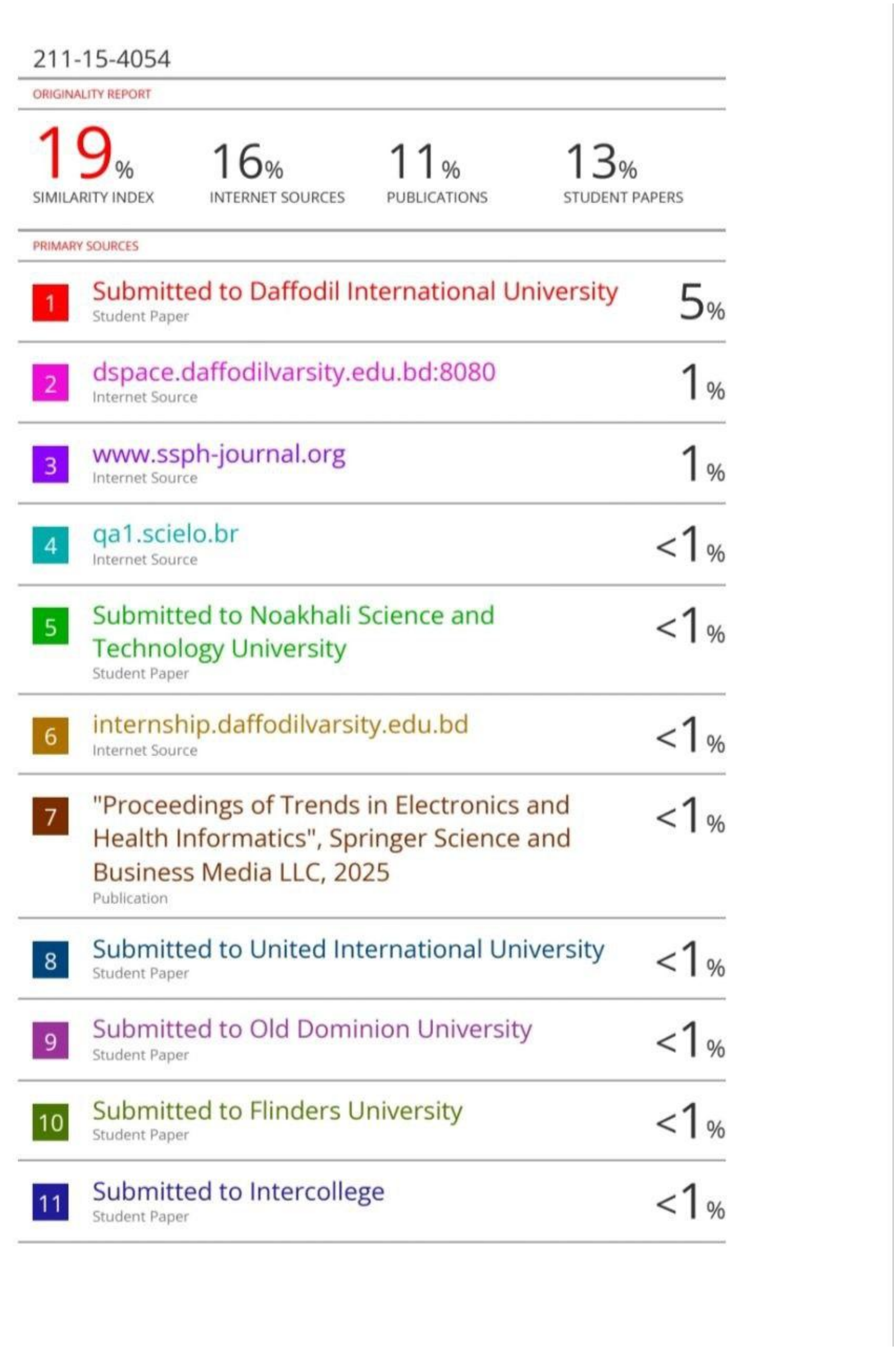
[14] R. Islam, M. A. R. Sarker, M. S. Sakib, and M. S. Hossen, "Dengue Fever Responses in Dhaka City, Bangladesh: A Cross-Sectional Survey," Cross-sectional Survey, 2023.

[15] S. Islam, C. E. Haque, S. Hossain, D. Walker, “Association among ecological and behavioral attributes, dengue vector and disease control: a cross-sectional study of the city of Dhaka, Bangladesh,” vol. 12, Nov. 30, 2019.

[16] S. J. Khan, K. Nahar Tanni, T. Roy, M. A. Chisty, M. Rakibul Islam, M. Alim Al Raji Rumi, M. Sadman Sakib, M. Abdul Quader, M. Nafee-Ul-Islam Bhuiyan, F.

Rahman, E. Alam, “Knowledge, Attitude, and Practices towards Dengue Fever among University Students of Dhaka City, Bangladesh,” vol. 19, Mar. 28, 2022.

[17] S. M. M. H. Chowdhury, M. A. Rashid, S. Y. Trisha, M. Ibrahim, M. S. Hossen, “Dengue Investigation Research in Bangladesh: Insights from a Scoping Review,” vol. 8, Mar. 12, 2025.



211-15-4054

ORIGINALITY REPORT

19%	16%	11%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	www.ssph-journal.org Internet Source	1%
4	qa1.scielo.br Internet Source	<1%
5	Submitted to Noakhali Science and Technology University Student Paper	<1%
6	internship.daffodilvarsity.edu.bd Internet Source	<1%
7	"Proceedings of Trends in Electronics and Health Informatics", Springer Science and Business Media LLC, 2025 Publication	<1%
8	Submitted to United International University Student Paper	<1%
9	Submitted to Old Dominion University Student Paper	<1%
10	Submitted to Flinders University Student Paper	<1%
11	Submitted to Intercollege Student Paper	<1%