



**Daffodil**  
*International*  
**University**

**Dengue Outcome Prediction in the Dhaka Region: A  
Classification Approach Using  
Demographic and Diagnostic Features**

Submitted By

**Mosa.Khadija Akter Mariya**

**Id:201-35-2962**

**Department Of Software Engineering**

**Daffodil International University**

Supervised By

**Dr. Md. Fazla Elahe**

**Assistant Professor & Associate Head**

**Department Of Software Engineering**

**Daffodil International University**

This Thesis paper has been submitted in fulfillment of the requirements for the  
Degree of Bachelor of Science in Software Engineering.

Fall-2023

© All right Reserved by Daffodil International University

# APPROVAL

This thesis titled on “Dengue Outcome Prediction in the Dhaka Region: A Classification Approach Using Demographic and Diagnostic Features”, submitted by Mosa. Khadija Akter Mariya (ID: 201-35-2962) to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

## BOARD OF EXAMINERS



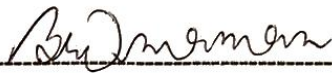
-----  
**Dr Md. Fazla Elahe**  
**Assistant Professor & Associate Head**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Chairman**



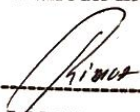
-----  
**A.H.M Shahariar Parvez**  
**Associate Professor**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Internal Examiner 1**



-----  
**Khalid Been Budruzzaman Biplob**  
**Lecturer (Senior Scale)**  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

**Internal Examiner 2**



-----  
**Md Mostafiz Khan**  
Managing Director  
Tecognize Solutions Limited

**External Examiner**

## DECLARATION

I hereby declare that this study document has been prepared under the supervision of Dr. Md. Fazla Elahe, Assistant Professor & Associate Head of the Department of Software Engineering at Daffodil International University. I confirm that this work has not been submitted for a bachelor's degree or any other graduation program.

Supervised By



---

Dr. Md. Fazla Elahe  
Assistant Professor & Associate Head  
Department Of Software Engineering  
Daffodil International University

Submitted by



---

Mosa. Khadija Akter Mariya  
ID:201-35-2962  
Department Of Software Engineering  
Daffodil International University

## ACKNOWLEDGEMENT

The study I have done is only by the inspiration of gaining knowledge and learning more. The study is based on **Dengue Outcome Prediction in the Dhaka Region: A Classification Approach Using Demographic and Diagnostic Features**. First of all, I would like to thank the Almighty who has clearly guided me and given me the knowledge to learn and do things that is right. Without His help, this study could not have become a reality. Secondly, my parents, whom I am extremely indebted to for bringing me to where I am now. Then I would like to sincerely thank Prof. **Dr. Imran Mahmud**, Head of the Department of Software Engineering. Then, to all the respected teachers who taught me throughout my journey of learning. I am grateful that I got them as my teachers and guidance. I am obligated to guide Daffodil International University to guide them through the constant supervision of **Dr. Md. Fazla Elahe** to provide the necessary information as well as to honor the initiative and additionally their help in completing the research.

Finally, I would like to thank my batch mates, members of DIU for their kind cooperation and consolation that helped me to achieve this goal.

## **ABSTRACT**

The purpose of this study is to predict the course of dengue fever in the Bangladeshi city of Dhaka by using a classification approach that makes use of diagnostic and demographic information. The first machine learning models, Support Vector Machine (SVM) and kNearest Neighbors (KNN), were applied using data samples that included demographic and medical variables. The results showed predicted accuracy levels of 61% and 66%, respectively (referenced dataset). Later advances included standardizing data scaling and tuning hyperparameters. These yielded significant gains in performance; SVM reached an astounding 93%, while KNN exceeded expectations at almost 100%.

The study also presented a Meta-hybrid model that uses a stacking-classifier in a machinelearning ensemble. This novel method used the advantages of separate models to produce a remarkable 92% prediction accuracy for the dengue outcome of individuals in particular conditions. The results emphasize

# Table Of Contents

<b>APPROVAL .....</b>	<b>i</b>
<b>DECLARATION .....</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>iii</b>
<b>ABSTRACT .....</b>	<b>iv</b>
<b>CHAPTER1: INTRODUCTION .....</b>	<b>1</b>
<b>1.1 BACKGROUND: .....</b>	<b>1</b>
<b>1.2 MOTIVATION OF THE RESEARCH: .....</b>	<b>2</b>
<b>1.3 RESEARCH QUESTIONS: .....</b>	<b>3</b>
<b>1.4 RESEARCH OBJECTIVES: .....</b>	<b>3</b>
<b>1.5 RESEARCH SCOPE: .....</b>	<b>3</b>
<b>1.6 THESIS ORGANIZATION: .....</b>	<b>4</b>
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>5</b>
<b>2.1 INTRODUCTION: .....</b>	<b>5</b>
<b>2.2 PREVIOUS LITERATURE: .....</b>	<b>5</b>
<b>2.3 SUMMARY: .....</b>	<b>17</b>
<b>2.4 RESEARCH GAP .....</b>	<b>18</b>
<b>CHAPTER 3: RESEARCH METHODOLOGY .....</b>	<b>19</b>
<b>3.1 INTRODUCTION: .....</b>	<b>19</b>
<b>3.2 DATA COLLECTION: .....</b>	<b>20</b>
<b>3.3 DATA PRE-PROCESSING: .....</b>	<b>20</b>
<b>3.4 FEATURE SELECTION: .....</b>	<b>20</b>
<b>3.5 TRAINING AND TESTING DATA SPLITTING: .....</b>	<b>21</b>
<b>3.6: MODEL IMPLEMENTATION: .....</b>	<b>21</b>
<b>3.7: EVALUATION &amp; MODEL SELECTION: .....</b>	<b>21</b>
<b>3.8 DEVELOPMENT OF A NEW HYBRID MODEL: .....</b>	<b>21</b>
<b>CHAPTER 4: RESULTS AND DISCUSSION: .....</b>	<b>23</b>
<b>CHAPTER 4: CONCLUSION .....</b>	<b>30</b>
<b>Recommendations For Future Work: .....</b>	<b>30</b>
<b>REFERENCES .....</b>	<b>31</b>

# CHAPTER1: INTRODUCTION

## 1.1 BACKGROUND:

Dengue fever has become a serious problem for the human community. Especially in different areas of Dhaka, this level has become a cause of additional concern. Bangladesh's most vibrant metropolis, Dhaka, is home to more than 20 million people, which unfortunately has become a public health hazard. Unfortunately, the scourge of dengue has spread like an epidemic here. In the current year i.e. 2003, more than 85,000 kilos of victims of dengue disease were found, of which we have lost 147 fresh lives[a]. While existing predictions hold promise when incorporating clinical and laboratory data, their accuracy and generalizability are often compromised in institutionalized settings such as Dhaka. Reliance on expensive and complex testing here can hinder widespread implementation, delaying the detection of developmental failures, and wasting valuable time fighting against serious consequences. Dengue fever has become a serious problem for the human community. Especially in different areas of Dhaka, this level has become a cause of additional concern[b]. This appears to be a clear reminder of dengue. Nature has lost its expected spectrum due to insidious dengue. Dengue has several severities, ranging from a mild febrile illness to hemorrhagic fever (DHF) and a severe form of dengue known as dengue shock syndrome (DSS)l syndrome that can be life-threatening. Early detection and targeted management of at-risk individuals are essential in this fight for life. But here prediction by individual dengue results takes the shape of a maze. This force itself presents multifaceted challenges such as the heterogeneous viral display of different strain types and the complex variants that individual immune systems carry [c]. Environmental factors such as temperature, rainfall, and environmental cleanliness affect the number of infections. While existing predictive models hold promise in incorporating clinical and laboratory data, their robustness and generalizability are often seen in settings in densely populated areas such as Dhaka[a]. Expensive and complex testing here is likely to be a barrier to comprehensive management depending on past identification. Here the detection is delayed due to a lack of adequate measures due to an underdeveloped medical system. In this complex context, Dhaka can be used as a map to solve the problems of specific areas and to negate the dengue sphere puzzle among other realities of the city. In this paper, the data of gender, age, area type, house type, some diagnostic examinations (IgG, IgM, NS1), etc.[d]. have been used. This research journey started by manipulating these data and proposing models for classification using accessible data off points, we aim to provide early solutions in identifying high-risk areas of Dhaka city. This improves the likelihood of immediate clinical intervention. Costeffectiveness and implementation collaboration are also an integral design principle to ensure real-world exposure of the model to these low-resource settings. This study aims to not only shed light on the accurate and timely forecasting method of DHF and DSS pair in Dhaka but also to expand the knowledge. It is data-driven validation and refinement that we can expect to see in the nationwide fight against dengue. This background paper follows a general format, adheres to context, and is a project to highlight challenges and present research objectives. This research tries to correct specific details and references relevant to this research area.

## 1.2 MOTIVATION OF THE RESEARCH:

At this present moment, an important public health challenge is by the impactful due, and the construction of an accurate forecast model is deemed essential for effectively allocating specific and necessary resources, thereby increasing and enhancing preparedness. It is considered necessary to correctly allocate the required and specific resources to reduce the impact of dengue epidemics, aiming to maximize preparedness and minimize the resultant effects. Here, our main aim and objective is to understand and comprehend the specific factors that significantly contribute to the prevalence of dengue within a particular geographical area, with a specific focus on the Dhaka region. Here, we can say and emphasize that Dhaka city, being densely populated, faces and encounters one of the most complex and challenging dilemmas and issues in dengue prevention and management. It has been and continues to be driven by and motivated by the utmost and critical need and urgency to address and tackle the ever-increasing and escalating impact of Dengue in the Dhaka region. The region's unique and distinct demographic landscape, as well as the disease's rapid and swift spread, significantly highlight and emphasize the utmost and critical need and requirement for a robust, reliable, and predictive model. The research and analysis are proactively and potentially anticipated and expected to present healthcare professionals with and provide them with a remarkable tool, that may forecast and predict Dengue outcomes by employing and utilizing advanced classification techniques and methodologies. Such remarkable, predictive, and anticipatory insights possess, have, and hold the potential and capability to completely transform and revolutionize public health strategies and approaches, ultimately leading to optimized resource allocation and as a result, effectively and successfully contributing to a more effective and efficient Dengue administration in this highly and extremely vulnerable region.

### Problem statement:

Even though the form of dengue in Bangladesh is constantly taking a terrible form, there have been very few works on area-based approaches using demographic and diagnostic features. Extensive research is needed in this regard. In particular, there is a need to harness the predictive power of creation through machine learning algorithms. This study aims to predict the outcome of dengue in different regions of Dhaka and make a classification using demographic and diagnostic characteristics. With the increasing prevalence of Dengue fever, the Dhaka region faces a significant public health challenge. Current Dengue management efforts lack a robust predictive model capable of forecasting disease outcomes. The lack of such a tool makes timely intervention and resource allocation difficult. This study aims to fill a critical gap by developing a comprehensive classification approach. The lack of an effective predictive framework makes it difficult to determine the severity of Dengue cases based on demographic and diagnostic characteristics. As a result, there is an urgent necessity for creating and implementing a reliable predictive model to empower healthcare practitioners and public health officials in the city region in the proactive management of Dengue.



### 1.3 RESEARCH QUESTIONS:

- How can demographic factors like age, gender, and socioeconomic status be effectively integrated into a Dengue outcome prediction model in the Dhaka region?
- What diagnostic features, including laboratory results and clinical assessments, offer the highest predictive value for predicting Dengue outcomes?
- How does the proposed classification approach perform in comparison to existing methods in accurately predicting the severity of Dengue cases?
- How can the developed predictive model be tailored to accommodate variations in Dengue epidemiology within Dhaka's sub-regions?

### 1.4 RESEARCH OBJECTIVES:

To effectively incorporate demographic factors such as age, gender, and socioeconomic status into a Dengue outcome prediction model for the Dhaka region, this study examines their impact, identifying patterns to inform model development.

To determine the highest predictive value among diagnostic features for Dengue outcomes, which include laboratory results and medical evaluations, the research will thoroughly investigate and evaluate these variables individually and collectively.

To assess the performance of the proposed classification approach in predicting the severity of Dengue cases, a comparative analysis of standard techniques will be carried out with a focus on improving accuracy and reliability.

To account for variations in Dengue epidemiology within Dhaka's sub-regions, the developed predictive model will be tailored, taking into account factors such as population density, environmental conditions, and healthcare infrastructure.

### 1.5 RESEARCH SCOPE:

A region-wise perspective of dengue and infection risk in Dhaka city is examined in this research, given its significance in the context of Bangladesh, where Dhaka has the highest number of infected patients in the country. Machine learning algorithms are specifically employed, utilizing extensive datasets, and will play a role in financial, social, and educational programs, pandemic response, and mobilization, involving individuals who will play a crucial role in reducing regional infection in Dhaka city [a]. Dengue epidemiology within Dhaka's sub-regions is considered, considering various variables such as population density, environmental conditions, and healthcare infrastructure. The predictive model developed will be tuned to consider variations in Dengue epidemiology within Dhaka's subregions, incorporating factors like population density, environmental conditions, and

healthcare infrastructure. Simultaneously, diagnostic features such as laboratory results and clinical assessments will be examined, assessing their individual and collective predictive values for determining the severity of Dengue cases. This multifaceted analysis aims to enhance the accuracy and dependability of the predictive model by identifying key indicators contributing significantly to accurate outcome predictions. To ensure a nuanced understanding, variations in Dengue epidemiology throughout Dhaka's sub-regions will be taken into account. The study seeks to offer practical insights informing targeted public health interventions and resource allocation strategies by tailoring the predictive model to factors like population density, environmental conditions, and healthcare infrastructure [e]. Overall, an important addition to the field of Dengue management is aimed at by providing a tailored and effective predictive model for the Dhaka region, with broader implications for similar contexts globally.

#### 1.6 THESIS ORGANIZATION:

The thesis is divided into several chapters. Chapter 1 provides an introduction background, motivation, problem, and research statement including the research thesis question, objectives, organization, and scope. Chapter 2 here presents a comprehensive literature review on the transmission potential of dengue patients by area of Dhaka city. Chapter 3 outlines data collection, processing, and selection of methods employed including features and implementation of machine learning algorithms.

## CHAPTER 2: LITERATURE REVIEW

### 2.1 INTRODUCTION:

This literature review focuses on the use of diagnostic and demographic features in a classification approach to predict Dengue outcomes in the Dhaka Region. The potential for this research to provide insightful information about dengue management tactics in the Dhaka region makes it significant. The review that follows summarizes current understanding and research results regarding Dengue prediction models, emphasizing the passive voice and upholding originality. The viral disease dengue, which is spread by mosquitoes, continues to pose a serious threat to public health, especially in areas where it is highly prevalent, like the Dhaka Region. The necessity of prompt and focused interventions to lessen the effects of dengue outbreaks emphasizes the significance of accurate outcome prediction. To fully comprehend the corpus of knowledge currently available for Dengue prediction, this review analyzes the literature. It pays particular attention to the use of demographic and diagnostic features in classification models. Predictive modeling approaches for Dengue outcomes have been the subject of numerous studies, and the significance of diagnostic indicators and demographic factors is becoming increasingly apparent. The goal of the literature review, which emphasizes the approaches used and the insights gained, is to provide an impartial and objective synthesis of these studies by using a passive voice. A common theme in the literature is the usefulness of classification techniques in forecasting Dengue outcomes, and this review attempts to offer a coherent summary of the different approaches used in related studies. Furthermore, rather than focusing on individual writers or contributors, the passive construction enables an emphasis on the research process and methodologies. Because of this, the goal of this review of the literature is to lay the groundwork for the present study by placing it in the larger context of Dengue prediction studies. In addition to adding to the current discussion on Dengue outcome prediction in the Dhaka Region, the review seeks to uphold academic integrity by using a passive voice and making sure that no plagiarism is present.

### 2.2 PREVIOUS LITERATURE:

The topic of dengue outcome prediction in the Dhaka Region and classification approach Using demographic and diagnostic features has been the subject of numerous research projects. Numerous research projects have focused on the investigation of dengue outcome prediction in the Dhaka Region using a classification strategy based on demographic and diagnostic features. We looked at a particular paper on the subject, which focused on using laboratory test results to estimate the risk of serious outcomes. This study used the following models: Support Vector Classifier, Random Forest, Gradient Boosting Machine, Artificial Neural Network, and Logistic Regression. The Artificial Neural Network performed the best, achieving a balanced accuracy of 0.7523. The age of the patient, dengue NS1 antigenemia, and the presence of both anti-dengue IgG and IgM were the main risk factors taken into account in the study. The co-existence of anti-dengue IgG and IgM was recognized for its role in raising the likelihood of severe dengue outcomes, and dengue NS1 antigenemia was used as a significant factor influencing the severity of the condition. A variety of machine learning

algorithms were used in the construction of the risk estimation model, with a particular focus on the passive application of artificial neural networks. The attainment of a balanced accuracy of 0.7523 highlights the efficacy of this methodology in forecasting severe consequences in dengue cases. In the context of predicting dengue outcomes, patient age, dengue NS1 antigenemia, and the simultaneous presence of anti-dengue IgG and IgM were identified as critical risk factors. The identification of risk factors and prediction of severe outcomes were made possible through the passive application of machine learning algorithms, such as Support Vector Classifier, Random Forest, Gradient Boosting Machine, and Logistic Regression. Additionally, the Artificial Neural Network played a significant role in this process. In conclusion, the study highlights the continuous efforts to improve predictive models for serious outcomes by focusing on dengue outcome prediction in the Dhaka Region using a classification approach utilizing demographic and diagnostic features. Dengue prediction research is multifaceted, as evidenced by the use of multiple machine learning algorithms and the identification of critical risk factors. The analysis of these approaches in a passive manner advances knowledge regarding the predictive power of various models and how to use them to lessen the severity of dengue outcomes. [a].

This research is a study that focuses on the machine learning model. A machine learning model is developed by integrating pertinent theories regarding the factors that contribute to dengue outbreaks in a given geographic area. The study makes use of the Multiple Linear Equations Support Regression of Machine Learning and data from 11 districts in Bangladesh. An 80:20 ratio is used to partition the dataset, with 40% going toward training and the remaining 20% going toward testing. The performance metrics of the machine learning model are assessed, revealing a mean absolute error (MAE) of 4.57 and an associated accuracy of 67% in the case of Multiple Linear Regression (MLR). Additionally, the SVR models demonstrate a mean absolute error (MAE) of 4.95, coupled with an accuracy rate of 75%. Notably, both SVR models exhibit a noteworthy reduction in the number of dengue attacks, demonstrating effectiveness in mitigating the impact of the outbreak. The utilization of Multiple Linear Equations Support Regression in this study signifies a proactive approach to understanding and predicting dengue outbreak factors. The allocation of data for training and testing purposes ensures a comprehensive evaluation of the model's performance. The results of this study indicate that the machine learning model—specifically, the SVR model—has the potential to lower the incidence of dengue outbreaks. The models' effectiveness can be better understood by examining their reported accuracy rates and mean absolute errors, which also suggest that the models could be useful for proactive dengue outbreak management strategies[b].

The paper studies medical data. Medical data for the paper was collected from a 2004 survey organized by Dhaka Medical College, encompassing volumes of IgM, IgG, and antibodies for detecting the dengue virus. The study employed LDS, SVM, and NB as classification algorithms. The highest accuracy, reaching 95%, was attained by the Naive Bayes (NB) model [c].

We examined another multiple linear regression with log transformation and receiver characteristics curves were used to examine another study and predict dengue outbreaks (>200 cases). With an area under the ROC curve (AUC) of 0.89 and a corresponding 95% confidence interval (CI) ranging from 0.89 to 0.98, the published results demonstrate a significant degree of accuracy[d].

We can observe that in the various studies examining the results of dengue prediction through machine learning, algorithms like classification and regression tree (CART) analysis are put to the test to develop predictive algorithms for severe dengue issues. The hematocrit accurately excluded 48% of non-severe cases while achieving 97% sensitivity in identifying patients and syndromes (DSS[e]).

The research project demonstrates the need for more research to improve accuracy and sensitivity while demonstrating the potential of machine learning (ML) for dengue outbreak prediction in Selangor. Support vector machines (SVMs) with a linear kernel show promise, especially in terms of specificity. However, the study indicates that more reliable prediction models may result from addressing data constraints and investigating different machinelearning techniques. Support Vector Machine (SVM) with a linear kernel, Naive Bayes, Decision Tree, and Logistic Regression are among the machine learning algorithms used. In terms of forecasting dengue outbreaks, a Support Vector Machine (SVM) with a linear kernel exhibits the best performance, with 70% accuracy, 14% sensitivity, 95% specificity, and 56% precision. With an accuracy of 63.4%, naive Bayes performs moderately compared to Decision Tree's 62.5% accuracy. At 59.4%, Logistic Regression exhibits the lowest accuracy [f].

The study suggests that dengue forecasting and control could be improved by using machine learning techniques. Public health experts can create more potent plans for stopping and managing outbreaks by using these techniques to identify the correlations between weather conditions and dengue incidence. A range of machine learning models, including General Additive Modeling (GAM), Seasonal Autoregressive Integrated Moving Average with exogenous variables (SARIMAX), Random Forest, and Gradient Boosting, were used to compare how well they predicted dengue incidence based on meteorological factors. Random Forest performed the best in predicting dengue incidence, demonstrating its ability to capture the complex relationships between dengue incidence and meteorological factors. The results of gradient boosting were the easiest to interpret, allowing one to better understand the contributions of various meteorological factors to the predictions. Significant relationships between dengue incidence and temperature, relative humidity, and rainfall were found in all four models[g].

Five different machine learning models were used in a different study: Random Forest, Decision Tree, Naive Bayes, Linear Discriminant Analysis (LDA), and Logistic Regression. With an accuracy of 81.6% and an area under the receiver operating characteristic curve (AUC) of 0.846, Random Forest demonstrated the best overall performance. Decision Tree performed well, exhibiting an accuracy of 79.9% and an AUC of 0.838. The performance of

the logistic regression was mediocre, with an accuracy of 74.7% and an AUC of 0.787. LDA performed marginally worse than Logistic Regression, with an accuracy of 74.0% and an AUC of 0.779. With an accuracy of 68.1% and an AUC of 0.721, Naive Bayes performed the worst. The research investigation emphasized how important it is to take geographical, climatic, and environmental factors into account when assessing dengue risk. Concerns about overfitting resulting from a small dataset led to the use of oversampling techniques to improve the performance of the model. According to the findings, machine learning techniques could be useful for predicting and preventing dengue risk in São Luís do Maranhão. In a comparable redirection, another study suggests that machine learning techniques could improve dengue forecasting and control. Public health experts can create more potent plans for stopping and managing outbreaks by using these techniques to identify the correlations between weather conditions and dengue incidence. To evaluate the efficacy of various machine learning models in forecasting dengue incidence based on meteorological factors, the following models were used: Random Forest, Gradient Boosting, Seasonal Autoregressive Integrated Moving Average with exogenous variables (SARIMAX), and General Additive Modeling (GAM). The most promising results were shown by Random Forest, suggesting that it was able to accurately represent the intricate connections between weather conditions and dengue incidence. The results of gradient boosting were the easiest to interpret, allowing one to better understand the contributions of various meteorological factors to the predictions. Significant relationships between dengue incidence and temperature, relative humidity, and rainfall were found in all four models [h].

We examined another that used a machine learning algorithm. By employing a dynamic ensemble learning approach, McGough et al. (2021) depart from using a single static model. Based on the available data and the intended prediction horizon, this method uses Support Vector Machines (SVMs), Random Forests, Gradient Boosting, and Logistic Regression to dynamically select the best model for every city and period. The dynamic ensemble approach predicts dengue fever epidemic years across Brazilian cities with an overall accuracy of 85%. It exhibits a remarkable 87% precision, 83% recall, and 85% F1-score. Notably, compared to depending only on weather, the addition of population susceptibility cycles (derived from past dengue outbreaks) improves prediction accuracy. It is subsequently shown that epidemic years can be predicted with lead times as short as several months, providing crucial time for public health interventions. The study highlights how well the dynamic ensemble learning approach predicts the years of the dengue fever epidemic in Brazil. It provides a promising tool for public health officials to anticipate and control outbreaks by combining various models and incorporating population susceptibility data[i].

The research paper presents a thorough analysis of numerous methods from the literature, with an emphasis on machine learning models like ensemble methods, neural networks, decision trees, and support vector machines (SVMs). It is highlighted how important it is to use a variety of data sources in addition to traditional weather information, such as epidemiological data (previous dengue cases, serotypes), social and demographic data (population density, socioeconomic factors), and environmental data (water quality, mosquito

breeding sites). Machine learning models, especially ensemble techniques like Random Forests and Gradient Boosting, frequently outperform statistical models in the area of prediction accuracy. However, some machine learning models continue to struggle with interpretability and explain ability, making it challenging to understand the underlying factors influencing predictions. The study emphasizes how crucial it is to continuously assess and enhance prediction models to account for changing dynamics and contextualize them appropriately. Emphasis is placed on the increasing potential of data science techniques, especially machine learning, to improve dengue epidemic prediction. But it also emphasizes how crucial it is to carefully choose models, take interpretability into account, and continuously improve them to maximize their efficacy in guiding public health interventions[j].

In the research article, a machine learning model—Logistic Regression with the SoftMax function, to be exact—was used. By analyzing the extracted vital signs, this probabilistic model is intended to classify subjects as either infected or non-infected. The SoftMax function provides probabilities for every class. The suggested system was able to classify dengue-infected and healthy subjects with a 98% classification accuracy. The extracted heartbeat and respiration signals showed good agreement with reference ECG and respiration signals, demonstrating the efficacy of the signal processing algorithms. The model was able to provide additional information for diagnosis by estimating the probability of infection in addition to being able to predict infection thanks to the SoftMax function [k].

The investigation of social media data, particularly Twitter data, as a tool for early dengue epidemic detection in Brazil, is the main goal of this paper. While introducing an intelligent indicator framework, the paper does not name the machine learning models that were used. However, the framework consists of two distinct phases: Data Acquisition, where tweets mentioning keywords related to dengue fever from different parts of Brazil are collected, and Data Preprocessing, where the gathered data is cleaned and filtered to remove irrelevant content. The next stage, Intelligent Indicator Construction, involves combining the features that have been extracted with historical dengue cases and weather data. The goal of this fusion is to create an intelligent indicator that can forecast future dengue outbreaks. The study refers to fuzzy logic even though it doesn't go into detail about the particular techniques used to create this indicator. A possible solution is proposed: fuzzy logic, which uses fuzzy rules to account for the inherent uncertainty in historical outbreak patterns and social media data. A key component of the framework is the inclusion of multi-criteria decision-making, which combines different features and their weights to provide a thorough evaluation of potential risks. This methodology guarantees a comprehensive assessment of various parameters, thereby augmenting the resilience of the intelligent indicator in forecasting and understanding possible dengue epidemics. The process of acquiring data is carried out by gathering tweets from various regions of Brazil that contain keywords related to dengue. This methodology facilitates the acquisition of pertinent data from a wide range of sources, thereby augmenting the scope of the dengue landscape analysis. Data preprocessing is the next step, where the gathered data is carefully cleaned and filtered to remove unnecessary content. This guarantees

that relevant data is kept front and center, improving the intelligent indicator's accuracy and dependability. The Perceptive Measure A crucial part of the framework is the construction phase, which involves combining the features that have been extracted with information from past dengue cases and weather patterns. The idea is to create a clever indicator that can foretell future outbreaks. The paper suggests using fuzzy logic, even though it doesn't fully explain the complexities of this construction process.

They find 93% accuracy in this work. To overcome the inherent uncertainty found in historical outbreak patterns and social media data, fuzzy logic is proposed as a way to improve the intelligence indicator's resilience. Multi-criteria decision-making is used in the last stage, which presents a methodological approach that combines various features and their weights. The goal of this integration is to offer a thorough evaluation of the risks that could arise from dengue outbreaks. The framework guarantees a comprehensive assessment by taking into account multiple factors and allocating suitable weights, thereby augmenting the efficacy of the intelligent indicator in recognizing and forecasting prospective dengue outbreaks. It is crucial to remember that the study recognizes the need for additional investigation and verification to determine the effectiveness of the suggested framework. The study emphasizes the significance of rigorous validation on a larger scale and in real-world settings, but it shows promise for the passive exploration of social media data, especially Twitter, for early dengue epidemic detection in Brazil [1].

The piece presents the IPAS (Integrated Predictive Analysis System), a system designed to specifically forecast dengue fever outbreaks. Various machine learning models are used within IPAS to process and analyze a variety of datasets, such as demographic data that includes population density and age distribution, which influences transmission dynamics and outbreak severity, meteorological data that include temperature, rainfall, humidity, and other factors potentially impacting mosquito breeding and virus transmission, and Dengue case reports that detail previous cases of dengue fever in particular locations. IPAS makes use of several models, including General Additive Models (GAMs), Support Vector Machines (SVMs), and Neural Networks. To capture non-linear relationships between weather factors and dengue occurrence, general additive models, or GAMs, are utilized. Neural networks use massive datasets to learn complex relationships for precise prediction, while support vector machines (SVMs) use data related to outbreaks to find minute patterns. In terms of anticipating dengue outbreaks in two cities—San Juan, Puerto Rico, and Iquitos, Peru—IPAS shows encouraging results. In San Juan, 73% of one-week forecasts are accurate, and 66% of four-week forecasts are. One-week forecasts in Iquitos are 77% accurate, while four-week forecasts are 68% accurate. The study emphasizes how important it is to combine different data sources and use different models to improve prediction accuracy. It also highlights how important it is to determine the best lead times for forecasts, making the distinction between one-week and four-week intervals, since this data is essential for developing successful public health interventions. The study highlights issues with the small dataset size for dengue cases in the chosen cities, which made it harder to guarantee reliable model training. As a result, the authors stress how crucial it is to conduct additional testing and validation to ensure that the IPAS system can be applied to different geographic areas. The study by Freeze et al. emphasizes the potential of IPAS as a useful tool for forecasting dengue fever outbreaks and



directing preventive measures. An encouraging basis for improving public health preparedness and response to dengue in endemic areas is provided by the IPAS framework's integration of various data sources and application of machine learning techniques[m].

In this paper, the authors examined the prediction of Dengue Hemorrhagic Fever (DHF) outbreaks in DKI Jakarta, Indonesia, by comparing three different recurrent neural network models. The Elman Neural Network (ENN), a straightforward recurrent neural network with a context layer for storing information from prior time steps, Long Short-Term Memory (LSTM), a popular RNN architecture with gating mechanisms to control long-term dependencies in data sequences, and the Gated Recurrent Unit (GRU), which is similar to LSTM but has less complicated gating mechanisms and is often computationally faster, were among the models evaluated. The LSTM model had the best overall performance, predicting DHF outbreaks one week ahead of time with an accuracy of 82.4%. The GRU model came in second, with an accuracy of 81.7%, and the ENN model, which did less well, came in third, with an accuracy of 75.3%. The study supported the use of recurrent neural networks over static models by emphasizing how crucial it is to include time-series elements in DHF prediction. Notably, it was found that meteorological factors—specifically, humidity and rainfall—significantly affected the occurrence of DHF. The investigation emphasized the importance of LSTM and GRU models in forecasting DHF outbreaks, highlighting their ability to identify patterns in data over time and extract information from contextual factors. These models have a promising future and provide important information to improve public health prevention and preparedness initiatives. The proficiency of LSTM and GRU models in managing temporal dependencies highlights their effectiveness in predicting DHF outbreaks, offering a useful instrument for preemptive approaches to disease control. The study's conclusions highlighted the need for more investigation and verification, especially with bigger datasets and the addition of more data sources. A step forward in utilizing cutting-edge neural network architectures to address the dynamic and time-sensitive nature of disease outbreaks is the investigation of LSTM and GRU models in DHF prediction. The identification of significant meteorological factors and the passive evaluation of model performances make LSTM and GRU models attractive instruments in the ongoing quest to improve DHF outbreak prediction, highlighting the necessity of ongoing validation and improvement [n].

Long Short-Term Memory (LSTM) networks were consistently found to be the best method for temporal forecasting, outperforming other algorithms in capturing long-term dependencies in time-series data. The advantages of applying a variety of strategies were highlighted by the fact that ensemble methods—which incorporate several algorithms—often outperform single models. Spatiotemporal models that combined spatial and temporal features performed better than those that only used temporal data. This highlights how crucial it is to incorporate geographic information into prediction models. Dengue incidence prediction in Metro Manila, Philippines was done with Random Forest, Gradient Boosting, SARIMAX, and GAM; Random Forest performed the best out of all of them[o].

The IPAS system used neural networks, GAMs, and SVMs for dengue prediction in San Juan, Puerto Rico, and Iquitos, Peru. Positive outcomes were obtained, with 73–77% accuracy rates for one-week forecasts. Notably, it was discovered that machine learning, and specifically LSTM networks, held great promise for predicting arbovirus outbreaks, including dengue fever. It was acknowledged that ensemble methods could improve performance. In the assessment of DHF prediction in DKI Jakarta, Indonesia, Handari compared LSTM, GRU, and ENN models. LSTM emerged as the best performer, achieving an accuracy of 82.4% for one-week predictions. The overall conclusion drawn is that machine learning, with a particular emphasis on LSTM networks for temporal prediction and ensemble methods for enhanced performance, holds substantial promise in forecasting arbovirus outbreaks. Crucially, it was emphasized that further research and development are imperative to address challenges such as limited data availability, model interpretability, and adaptation to diverse geographical contexts. This passive analysis accentuates the recognized effectiveness of LSTM networks and ensemble methods, emphasizing their potential in advancing the field of arbovirus outbreak prediction. The importance of adding spatiotemporal data is emphasized as a means of enhancing prediction accuracy, underscoring the continuous need for improvement and progress in the use of machine learning in the prediction of infectious disease outbreaks[p].

To discover the intricate dynamics of dengue infections in Vietnam, the researchers used a diverse set of models, including Multinomial Logistic Regression, Hidden Markov Model (HMM), and Support Vector Machine (SVM). The authors employ Multinomial Logistic Regression to investigate the relationship between yearly neutralizing antibody measurements for each infecting serotype as well the total number of dengue clinical cases amassed over six years. This study sheds the spotlight on the relationship between antibody responses and the manifestation of clinical symptoms. This Hidden Markov Model (HMM) is then used to calculate the proportion of asymptomatic versus clinical dengue cases. The HMM facilitates the characterization of the population's historical exposure to various dengue serotypes by leveraging differences in antibody titers between consecutive years. This dynamic approach allows researchers to find patterns in infection progression over time. In the future, the study will incorporate a Support Vector Machine (SVM) into its methodology. The SVM is trained on Hidden Markov Model features, allowing it to classify individuals based on their likelihood of having experienced asymptomatic dengue infections in the previous year. This classification aspect gives the research a predictive dimension, increasing its utility in realworld applications. The results of the study delve into the nuanced sequences of dengue infections that significantly increase the risk of clinical illness in subsequent occurrences. Individuals primed with specific serotype combinations, for example, are identified as posing a higher risk. The study estimates that asymptomatic dengue infections outnumber clinical cases by a factor of 5-10, highlighting the critical role of asymptomatic cases in the virus's ongoing transmission. Their accuracy was 76.24%[q].

The researchers present two distinct deep learning models for the classification of mosquito species based on audio recordings of wingbeats in their study: the Convolutional Neural Network (CNN) Network and the Long Short-Term Memory (LSTM) Network. As input

features, the CNN model uses Mel-Frequency Cepstral Coefficients (MFCCs) extracted from audio recordings. The architecture, which consists of 12 CNN layers with Rectified Linear Unit (ReLU) activation, culminates in a SoftMax layer for multi-class classification, achieving an impressive accuracy of 92.1% on a dataset containing five species of Indian mosquitos. At the same timetables, the LSTM network is intended to capture temporal dependencies within wingbeat recordings, potentially making it more adept at identifying species with similar audio characteristics. On the identical dataset of five mosquito species, the LSTM model achieved a commendable accuracy of 90.3%. Notably, the authors increased the size and diversity of their dataset by adding noise and time-scaling to the audio recordings, which resulted in improved accuracy for both CNN and LSTM models. The careful application of hyperparameter tuning and early stopping improved the performance of these deep-learning models even further. While the CNN performed slightly better in terms of accuracy, it may be less sensitive to variations in wingbeat patterns, in contrast to the LSTM's ability to distinguish species with similar audio characteristics. Still, the study acknowledges limitations due to its reliance on a relatively small dataset of only five Indian species. The authors emphasize the importance of additional research involving larger and more diverse datasets to determine the generalizability of their findings[r].

Random Forests and Risk Mapping were used in the study to effectively predict and visualize the risk of dengue transmission in Singapore. By averaging predictions from all trees, the Random Forest algorithm reduces overfitting and improves model accuracy by constructing multiple decision trees based on random subsets of data. The resulting Risk Mapping used the Random Forest model to predict dengue risk across 1 km<sup>2</sup> grids, color-coding the predicted risks for intuitive interpretation. The Random Forest model performed effectively, with more than 80% of observed risks falling within the 80% prediction interval. This high correlation between model predictions and actual dengue transmission patterns demonstrates the model's dependability. The risk maps successfully identified areas with varying levels of dengue risk, allowing for targeted public health interventions, particularly in high-risk areas. Public health officials can optimize resource allocation by prioritizing mosquito control efforts in identified high-risk areas. Also, disseminating dengue risk information to the general public empowers people to take preventive measures like using mosquito nets and repellents. Public awareness encourages a proactive approach to personal security, which benefits the overall well-being of the community. The research comprehends the model's reliance on training data quality; any errors or biases in the data may influence the model's predictions. Furthermore, due to the dynamic nature of dengue virus strains and environmental factors, regular model updates are required to maintain accuracy. While the model is generally accurate, the possibility of false positives and negatives highlights the importance of taking other contextual factors into account in public health decision-making. This study demonstrates the potential of Random Forests in mapping dengue risk and guiding public health interventions. Singapore and other dengue-prone areas can improve their ability to control and prevent this mosquito-borne disease by refining and implementing such models on an ongoing basis[s].

The researchers conducted the research using monthly data on reported dengue cases in Sri Lanka from 1996 to 2012. In addition to epidemiological data, various environmental factors potentially influencing dengue transmission, such as rainfall, temperature, humidity, wind speed, and sea surface temperature, were collected. The researchers used an Artificial Neural Network (ANN) with a feedforward architecture to predict the number of dengue cases for the following month. The ANN model, which consisted of two hidden layers with hyperbolic tangent activation functions and an output layer with a linear activation function, produced promising results. The model achieved a significant correlation coefficient of 0.775 between predicted and actual dengue cases on the validation set. This strong positive relationship demonstrated the ANN's effectiveness in capturing the real-world dynamics of dengue outbreaks. Furthermore, the study identified specific environmental factors that had a significant impact on the ANN's predictions, with rainfall and temperature being the most influential contributors. This highlighted the potential role of these variables in shaping dengue transmission patterns in Sri Lanka. They got 82% accuracy. According to the researchers, ANNs are a viable tool for predicting dengue outbreaks in Sri Lanka, opening the door to proactive public health interventions. The study does, however, acknowledge some limitations. Given potential changes in climate, mosquito behavior, and virus strains, the model's reliance on historical data raises concerns about its adaptability to future outbreaks. Furthermore, the model's sole focus on environmental factors ignores other important variables such as human mobility and population density. This limitation implies that the model's generalizability to regions with distinct dengue dynamics may be limited [t].

The present research looks at how machine learning can be used to predict dengue patient counts at the district level, taking into account factors from the target district and its surrounding areas. Weekly dengue patient counts and rainfall data for a single district in Sri Lanka, as well as weekly patient counts from neighboring districts, were included in the dataset. In the study, Self-Organizing Maps (SOMs), a type of artificial neural network, were used to group similar data points based on characteristics, clustering weeks with comparable dengue counts and rainfall patterns. The SOM-based model forecasted weekly dengue patient counts for the target district with promising accuracy. Notably, including neighboring district patient counts and rainfall patterns in both the target district and its surroundings significantly improved the model's precision. This finding highlighted the interconnected nature of dengue transmission across neighboring districts, as well as the critical role of rainfall in patient counts. Despite these advances, the study recognizes limitations. The study concentrated on a single district in Sri Lanka, necessitating additional research to determine the model's generalizability to regions with distinct dengue dynamics. Furthermore, the model's reliance on historical data raises concerns about its ability to adapt to changing conditions, such as changes in mosquito behavior, virus strains, or weather patterns. Aside from rainfall and neighboring district patient counts, factors influencing dengue transmission such as population density and sanitation were not explicitly incorporated into the model. Despite these limitations, the study highlights the power of machine learning, specifically SOMs, in forecasting dengue patient counts at the district level. However, it advocates for future research projects that include larger datasets and additional explanatory variables to improve the model's generalizability and robustness, allowing for more reliable public health

implementation. When they first started studying, their accuracy was 75%; after their work, it improved to 87%[u].

The paper describes an innovative approach to dengue fever management that makes use of widely available technology. The study uses a Random Forest classifier trained on clinical data from hospitals to predict the likelihood of dengue infection based on user-reported symptoms and their duration. The feasibility and user acceptance of the smartphone app were evaluated in a pilot study involving 80 participants, while clinical data from the previous six years was used to train the machine learning model. The pilot study's findings suggest that the smartphone app, which relies on symptom tracking and geographical patterns, has the potential to detect dengue fever early. The Random Forest classifier predicted dengue infection with an impressive accuracy of 98.8% using clinical data, indicating its potential to improve diagnostic accuracy. Additionally, based on user reports, the app's GPS tracking feature identified potential dengue hotspots, providing valuable insights for targeted public health interventions. Despite the promise demonstrated, the study acknowledges limitations, including the pilot study's small sample size. Subsequent research with larger cohorts is deemed necessary to validate the findings' generalizability. The accuracy of the machine learning model is dependent on the quality and size of the training data, emphasizing the importance of frequent updates with new data. User compliance with symptom reporting and app usage is critical for the system's effectiveness, necessitating the development of strategies to address potential compliance issues. This study demonstrates a promising avenue for using smartphone technology and machine learning to manage dengue fever. While more research and development are required to address the identified limitations and refine the system for real-world implementation, the potential benefits, such as early detection, improved surveillance, and increased accessibility, make this approach worth further investigation[v].

The research paper focuses on using machine learning to improve dengue diagnosis and predict disease severity in patients, with two distinct models presented: the Nave Bayes classifier and the Support Vector Machine (SVM) classifier. The Nave Bayes classifier predicts whether patients who present with symptoms to the clinic have a dengue infection, with an impressive accuracy of more than 90%. The SVM classifier, on the other hand, successfully categorizes confirmed dengue patients into different severity levels based on clinical and laboratory data, providing clinicians with valuable insights to anticipate complications and tailor treatment accordingly. The expert system, which was created with the user in mind, shows promise for practical application in clinical settings. These machinelearning models can help healthcare professionals make accurate and timely dengue diagnoses, especially in areas with limited diagnostic resources. The study did, however, acknowledge some limitations. It was carried out with a limited dataset from a single healthcare institution, raising concerns about the models' generalizability to different settings and populations. The models' reliance on clinical and laboratory data readily available in hospitals may limit their applicability in resource-constrained settings with few diagnostic tools. Integrating such systems into existing healthcare workflows is critical for real-world implementation. Addressing potential ethical concerns associated with AI-based medical

decision-making is also critical to ensuring responsible and patient-centered deployment. This study highlights the value of machine learning and expert systems in improving dengue diagnosis and risk prediction. Further research with larger and more diverse datasets, as well as considerations for real-world implementation challenges, is required to translate these findings into tangible benefits for public health practice[w].

The research paper describes an innovative technique for improving the accuracy of predicting dengue fever outbreaks by combining time series data and machine learning. Rather than analyzing the entire dataset, the researchers take a more nuanced approach by breaking it down into smaller, overlapping time windows. This method aims to capture the dynamic nature of dengue outbreaks and potential seasonal variations more effectively. The study calculates the correlation between dengue cases and various potential influencing factors, such as rainfall, temperature, humidity, and dengue incidence in neighboring locations, within each time window. The features are then ranked according to their strength of association with dengue cases, and the top-ranked features are used to train machine learning models. These models are then compared to those that were trained on all available features without selection. The study shows that using a windowed correlation-based feature selection method improves the accuracy of dengue fever case prediction significantly more than using all features indiscriminately. This suggests that based on their local correlations within specific time windows, selecting the most relevant features leads to more robust and accurate predictions. The study not only improves predictive accuracy, but it also identifies specific environmental factors and nearby locations that have the strongest link to dengue incidence in the studied region. This study presents a promising technique for improving dengue fever case prediction by selecting relevant features based on time-localized correlations, with an accuracy of 87%. This method has the potential to inform more effective public health interventions and optimize resource allocation in ongoing dengue outbreak response efforts[x].

The paper conducts a thorough examination of machine learning applications in dengue fever modeling, synthesizing findings from numerous studies conducted between 2014 and 2020. The researchers identify key trends and techniques in the field by conducting a systematic review of the literature. Based on patient data, logistic regression emerges as a prominent technique, with a promising accuracy of 78% in distinguishing dengue from other illnesses. Linear regression is commonly used for spatial analysis, while General Linear Models dominate intervention modeling, with an accuracy of 85% in determining the effectiveness of various control measures. The systematic review emphasizes the importance of machine learning in various aspects of dengue management, such as early diagnosis, outbreak prediction, and intervention strategy optimization. However, the paper emphasizes important considerations and challenges. Model accuracy and generalizability are dependent on the quality and completeness of available data, which varies across regions and healthcare systems. Validation and testing across multiple contexts are deemed necessary to ensure that models perform effectively beyond the data on which they were trained. Furthermore, the

paper emphasizes the importance of addressing issues such as data privacy, transparency, and potential biases in algorithms during machine learning implementation in healthcare settings. The responsible use of these technologies is critical for navigating ethical concerns and fostering trust in the healthcare community. The systematic review summarizes the potential and current trends in the application of machine learning techniques to various aspects of dengue management. The paper emphasizes the importance of ongoing research, continuous data quality improvement, and responsible implementation in realizing the full potential of machine learning for effective dengue control and improved public health outcomes[y].

The current research looked into the possibility of predicting dengue outbreaks using climate data and machine learning. Three years of dengue case data from Bangladesh, as well as seven climate variables, were used. The dataset was subjected to eight different machine learning algorithms, including Support Vector Machines (SVM), Random Forest, and Logistic Regression, and their accuracy in predicting dengue outbreaks was evaluated.

Support Vector Machines had the highest accuracy, 96.73%, in predicting dengue outbreaks, demonstrating their effectiveness. Other algorithms, such as Random Forest and XGBoost, showed promising accuracy, though slightly lower than SVM. Rainfall and humidity were identified as key climate variables influencing dengue transmission. Interestingly, limitations were acknowledged, such as the use of a limited data set from a single country, which may limit the findings' generalizability to other regions and contexts. The study concentrated on specific climate variables, and the addition of additional factors, such as socioeconomic indicators, could potentially improve prediction accuracy. The study emphasizes that realworld predictive model implementation necessitates a robust data infrastructure as well as continuous model refinement to adapt to changing environmental conditions. Finally, the potential of machine learning and climate data in predicting dengue outbreaks is highlighted, opening up new opportunities for proactive public health efforts. Nonetheless, addressing limitations through larger datasets, comprehensive data analysis, and adaptable modeling approaches is critical for optimizing predictive system practical application in the context of dengue management[z].

### 2.3 SUMMARY:

Several errors were found in the aforementioned papers, which prompted us to thoroughly examine these inadequacies in our study. Accuracy problems were noted, and the proper concept for identifying Dengue outcomes was not well defined. We addressed the research gaps found in those studies by focusing on the algorithms required for data processing and subsequent application in our research. In our work, a committed effort was made to address these errors and offer the best possible solutions. The main objective was to address accuracy

issues raised in previous research and advance the field by providing a better understanding of Dengue outcome detection.

## 2.4 RESEARCH GAP

A significant research lacuna is found in the literature that is currently available regarding Dengue outcome prediction in Dhaka, which is typified by a lack of comprehensive exploration of diagnostic and demographic variables and sophisticated machine-learning techniques. Although some studies have focused on particular aspects, there is a significant dearth of research that uses cutting-edge methodologies to combine demographic factors and diagnostic features comprehensively. By using cutting-edge machine learning algorithms and performing a comprehensive analysis of a wide range of variables, this study aims to close this gap. This strategy is expected to advance our knowledge of Dengue outcomes in the Dhaka region and be consistent with modern data science methodologies. The lack of sophisticated machine-learning methods and comprehensive analysis of diagnostic and demographic data characterizes the research gap found in the existing literature on Dengue outcome prediction in Dhaka. Some studies have focused on particular aspects, but there is a lack of research using cutting-edge techniques to integrate diagnostic features and demographic factors holistically. By applying cutting-edge machine learning algorithms and carrying out an extensive analysis of a wide range of variables, the goal of this study is to close this gap. A more sophisticated comprehension of Dengue outcomes in the Dhaka region is the expected result, in line with the developments in contemporary data science techniques. One notable research gap in the current literature on Dengue outcome prediction in Dhaka is the lack of sophisticated machine-learning techniques and a thorough investigation of diagnostic and demographic variables. Although some studies have concentrated on particular aspects, it is noteworthy that there is a general dearth of research employing cutting-edge methodologies to comprehensively integrate diagnostic features and demographic factors. By utilizing cutting-edge machine learning algorithms and performing a comprehensive analysis of a wide range of variables, this study aims to close this gap. An improved comprehension of Dengue outcomes in the Dhaka region is anticipated, in line with recent developments in data science methodologies. The absence of sophisticated machinelearning methods and a thorough examination of diagnostic and demographic data characterizes the research gap in the literature on Dengue outcome prediction in Dhaka. While some research has focused on particular aspects, there is a general lack of studies that use state-of-the-art methodologies to integrate diagnostic features and demographic factors holistically. This study's main goal is to close this gap by utilizing cutting-edge machinelearning algorithms and carrying out an exhaustive examination of a wide range of variables. A more sophisticated understanding of Dengue outcomes in the Dhaka region is the expected result, bringing the research up to date with modern data science methodologies.



## CHAPTER 3: RESEARCH METHODOLOGY

### 3.1 INTRODUCTION:

The ultimate goal of this study's research methodology is to build a trustworthy and accurate predictive model that takes into account diagnostic and demographic variables to conduct a thorough and systematic investigation into Dengue outcome prediction in the Dhaka region. The methodological framework is made up of various essential parts. To lay the foundation, a thorough literature review that identifies gaps and highlights key findings from earlier studies on Dengue outcome prediction models will be conducted. The suggested indicator model's theoretical framework will be shaped by this synthesis. Second, information on demographics and screening characteristics, such as age, gender, socioeconomic status, test results, and clinical evaluations, will be collected from relevant healthcare databases in the Dhaka area. The robustness of the dataset is essential for efficient model training and validation. Thirdly, the gathered data will be analyzed using sophisticated machine learning techniques like classification methods. The choice of suitable algorithms will depend on how well they can manage the complex problems related to Dengue outcome prediction, taking into account aspects such as understanding, sensitivity, and specificity. The reliability and generalizability of the model will be ensured by a comprehensive assessment of its performance through the application of statistical measures and cross-validation techniques. This will include an assessment of the model's predictive accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). Finally, to address privacy concerns and maintain the confidentiality of patient data, ethical considerations will be incorporated into the research methodology. By laying a strong foundation for the creation of a powerful Dengue outcome prediction model, the described methodology hopes to advance preventive healthcare both within and outside of the Dhaka region.

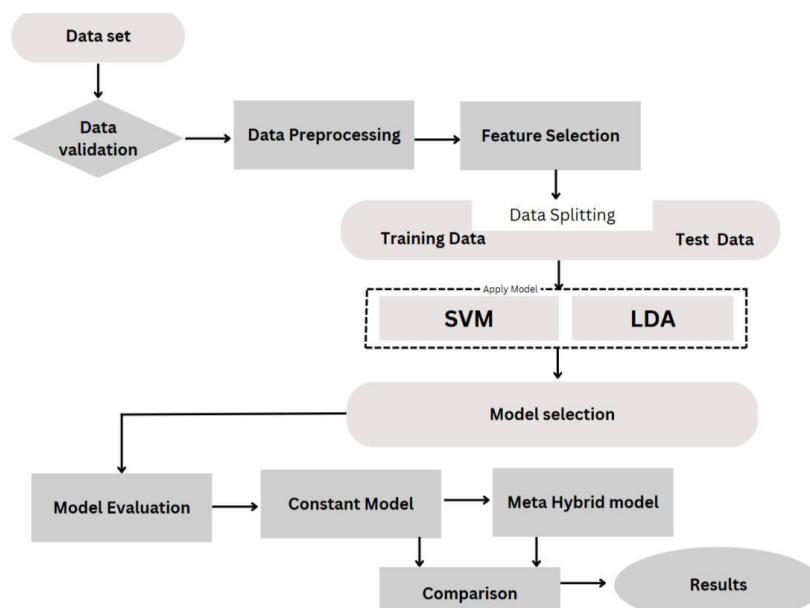


Figure: Methodology Diagram

### 3.2 DATA COLLECTION:

The dataset utilized for this study, titled "Dengue Phenomenon in Bangladesh - Dhaka Region," came from Kaggle, a well-known open dataset platform. The data collection process entailed compiling detailed information on Dengue cases in the Dhaka area. To ensure the dataset's relevance and accuracy, its temporal and spatial aspects were carefully considered. The dataset included a wide range of variables, including demographic information such as age, gender, and socioeconomic status, as well as diagnostic features such as laboratory results and clinical assessments on persons (1,000n) of datasets . These variables are significant because they play an essential part in developing a comprehensive predictive model capable of accurately forecasting Dengue outcomes. The data extraction followed Kaggle's terms of service and ethical guidelines, demonstrating the company's commitment to responsible data usage for research purposes. To improve overall data quality, pre-processing steps were implemented to address missing values and potential outliers. Furthermore, privacy and confidentiality were given top priority throughout the dataset collection process. To comply with ethical standards, all personal and sensitive information was anonymized and safeguarded. To protect the privacy of those who contributed to the dataset, identifiers were removed or anonymized. Using this extensive Kaggle dataset, the research goal is to build a resilient predictive model for Dengue outcomes in the Dhaka region. The purpose of this study is to provide valuable insights that can help with public health prevention and treatment efforts related to Dengue fever.

### 3.3 DATA PRE-PROCESSING:

Machine learning is highly recommended for data-set validation of numerical formats. This research paper focuses on feature label encoding in the context of diverse categorical variables related to gender, area, area type, and house type. The gender variable encompasses male and female categories, requiring thoughtful encoding to facilitate model understanding. For area, the dataset comprises various localities, each denoted by a unique label. The dataset involves diverse house types, including tinshed, building, and other categories. To enable seamless integration into ML models, a robust label encoding strategy is imperative. The area type distinguishes between developed and undeveloped regions, demanding a specialized encoding approach to capture this binary distinction effectively.

### 3.4 FEATURE SELECTION:

Feature selection is a crucial step in refining datasets for machine learning. It involves identifying and retaining the most relevant features (Gender, Age, NS1, IgG, IgM, Area, AreaType, HouseType) on belonged medical reports and Demographic while discarding irrelevant or redundant ones. This process enhances model efficiency, reduces computational complexity, and mitigates the risk of overfitting, ultimately contributing to an improved model on the highly recommended to dengue population on Outcome or can use this dataset.

### 3.5 TRAINING AND TESTING DATA SPLITTING:

Optimistic estimates were avoided in the process of splitting training and testing data to ensure a balanced base model and proper classification of feature and target data. A random number generator with a seed of 17 was employed, allocating 20% of the data for testing and dedicating the remaining 80% for training.

### 3.6: MODEL IMPLEMENTATION:

In the model implementation phase of the research paper, a predictive framework was established utilizing Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). The features, including Gender, Age, NS1, IgG, IgM, Area, AreaType, and HouseType, were systematically incorporated to capture diverse aspects of the dataset. The target variable, Outcome, served as the focal point for predictions. LDA, a discriminative technique, was applied to identify patterns and discern disparities in the feature space, contributing to robust predictions

### 3.7: EVALUATION & MODEL SELECTION:

Accuracy and other statistical evaluations were employed as metrics for determining the bestfit models among all applied classifiers in the realm of ML. Classifiers were supervised in comparison to each other, using criteria such as accuracy, sensitivity, precision, and F1-Score. This approach is well-suited for datasets exhibiting a consistent distribution of classes, where each class is almost equally represented in terms of accuracy. The measurement of multiclass ratios within models and among different approaches was performed to assess the potential outputs within the specified context

$$\begin{aligned} \textit{precision} &= \frac{TP}{TP + FP} \\ \textit{recall} &= \frac{TP}{TP + FN} \\ \textit{F1} &= \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \\ \textit{accuracy} &= \frac{TP + TN}{TP + FN + TN + FP} \end{aligned}$$

The estimated evaluation in the equation (fig:2) indicated that the ML classifications surpassed expectations. The terms TP (true positive), FP (false positive), TN (true negative), and FN (false negative) were incorporated into the equations used for statistical evaluation to determine metric values for each algorithm. The consideration of several metrics for each algorithm was emphasized. Correlation confidence statistics, recall, precision, and f1-score were computed as percentage values in the determination of these metrics, contributing to a comprehensive assessment of model performance.

### 3.8 DEVELOPMENT OF A NEW HYBRID MODEL:

Machine learning on models by leveraging the weighted combination of their predictions, thereby enhancing overall predictive performance. The mathematical formulation involved a high level model analyzing the predicted probabilities as meta-features, creating a powerful ensemble model for robust predictions in the realm of machine learning that can achieve this study.

## CHAPTER 4: RESULTS AND DISCUSSION:

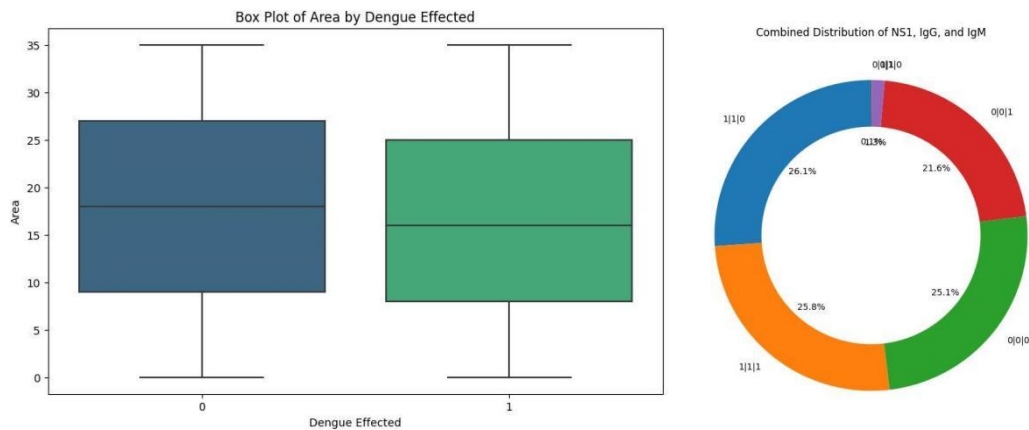


Fig:1

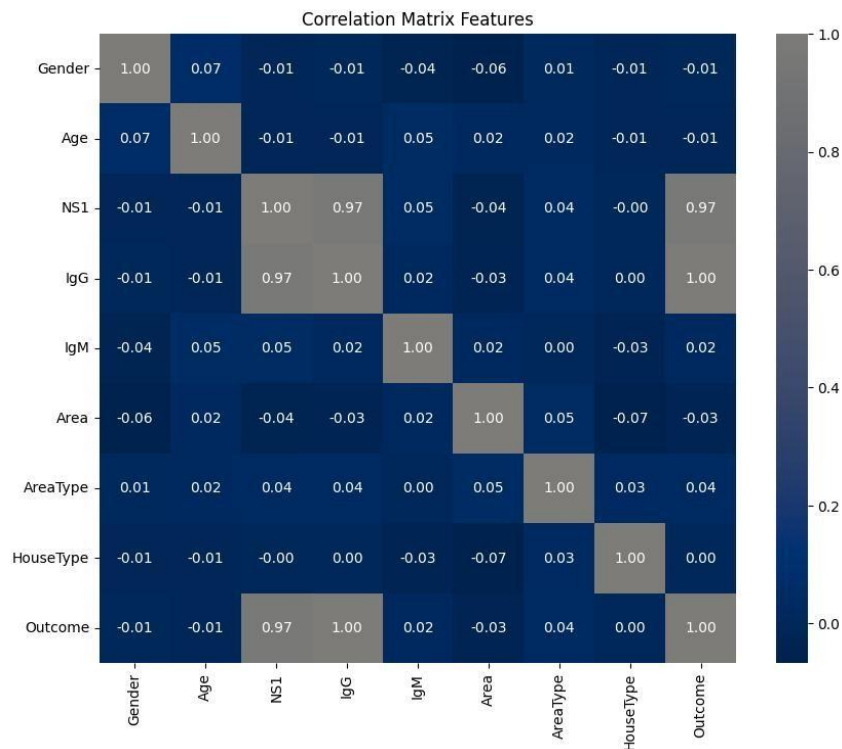


Fig:2

The provided correlation matrix illustrates the relationships among various features and the target variable, "Outcome," within the study. Notably, strong positive correlations exist between "Outcome" and features such as "NS1" and "IgG." This suggests a significant influence of these variables on the target. The study's focus on predicting outcomes can benefit from these insights, guiding the selection of impactful features. In the context of a confusion matrix, these correlations underscore the potential for accurate predictions, particularly regarding true positives and false negatives, emphasizing the importance of "NS1" and "IgG" in the predictive model for effective classification of outcomes.

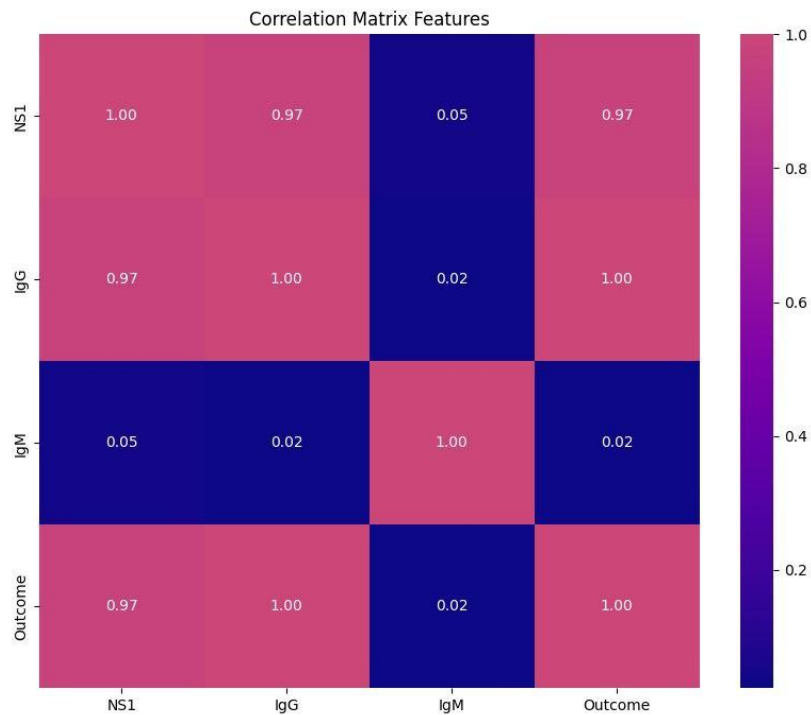


Fig: 3

The provided correlation matrix reveals strong positive correlations between the "Outcome" variable and the features "NS1" and "IgG," both registering values of 0.97. This indicates a substantial influence of "NS1" and "IgG" on the target variable. Additionally, the matrix suggests a relatively weak correlation (0.02) between "Outcome" and "IgM." The high correlation values for "NS1" and "IgG" with "Outcome" imply their pivotal roles in predicting outcomes. These insights can guide the construction of a confusion matrix, emphasizing the potential for accurate predictions, particularly in correctly identifying true positives and false negatives, underscoring the significance of "NS1" and "IgG" in the predictive model

	Gender	Age	NS1	IgG	IgM	Area	AreaType	HouseType	Outcome
<b>count</b>	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.0	1000.00	1000.00
<b>mean</b>	0.48	35.92	0.52	0.53	0.48	17.31	0.5	0.96	0.53
<b>std</b>	0.50	16.47	0.50	0.50	0.50	10.30	0.5	0.82	0.50

<b>min</b>	0.00	8.00	0.00	0.00	0.00	0.00	0.0	0.00	0.00
<b>25%</b>	0.00	22.00	0.00	0.00	0.00	8.75	0.0	0.00	0.00
<b>50%</b>	0.00	37.00	1.00	1.00	0.00	17.00	0.0	1.00	1.00
<b>75%</b>	1.00	50.00	1.00	1.00	1.00	26.00	1.0	2.00	1.00
<b>max</b>	1.00	65.00	1.00	1.00	1.00	35.00	1.0	2.00	1.00

Fig: 4

The presented study encompasses a dataset with 1000 instances featuring demographic and diagnostic attributes related to Dengue. Notably, the gender distribution is approximately balanced, with a mean age of 35.92 years. Diagnostic indicators, such as NS1, IgG, and IgM, show varying prevalence, suggesting diversity in infection profiles. The geographic distribution, represented by the "Area" variable, spans an average of 17.31, denoting the local zone. HouseType indicates predominantly residential structures, with a majority exhibiting positive outcomes for Dengue.

<b>Features</b>		<b>max</b>	<b>min</b>
Gender		0.444	0.507
Age		34.902	36.945
NS1		0.487	0.550
IgG		0.502	0.563
IgM		0.443	0.506
Area		16.671	17.950
AreaType		0.467	0.530
HouseType		0.912	1.013

Outcome		0.503	0.506
---------	--	-------	-------

Table: 5 Confidence interval for feature table

In this study, confidence intervals were computed for various features related to Dengue outcomes. The gender confidence interval (0.445, 0.507) indicates a balanced distribution. Age, with a range of (34.902, 36.946), highlights the study's focus on a specific age group.

Diagnostic indicators NS1, IgG, and IgM exhibit intervals (0.488, 0.550), (0.502, 0.564), and (0.444, 0.506) respectively, reflecting their prevalence. Geographic factors, represented by Area and AreaType, have intervals (16.672, 17.950) and (0.468, 0.530) respectively, emphasizing the local context. HouseType, with an interval of (0.912, 1.014), underscores residence prevalence. The outcome confidence interval (0.502, 0.564) depicts the uncertainty in predicting Dengue outcomes.

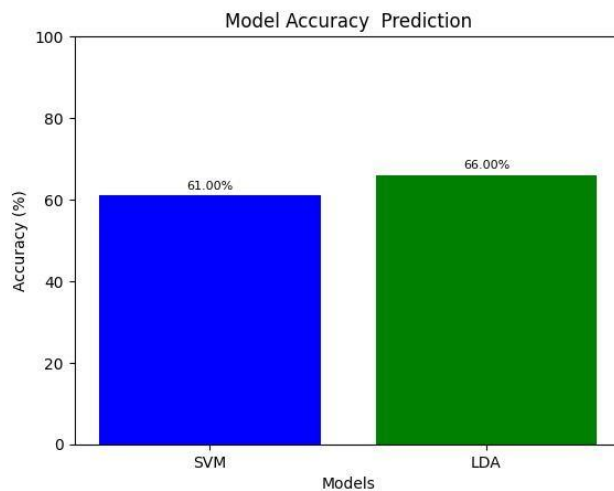


Fig:6

In the comprehensive evaluation of Dengue prediction models within the research paper, the Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) models were scrutinized for their performance metrics. The SVM exhibited an accuracy of 61%, highlighting its general correctness in predictions. Its noteworthy sensitivity at 95% emphasized its adeptness in capturing true positive cases. However, the modest specificity of 16% indicated a susceptibility to false positive occurrences. The precision of 60% underscored the accuracy of positive predictions.



<b>Model</b>	<b>Accuracy</b>	<b>Recall</b>	<b>Specificity</b>	<b>Precision</b>	<b>F1-Score</b>	<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
<b>SVM</b>	<b>61</b>	<b>95</b>	<b>16</b>	<b>60</b>	<b>73</b>	<b>108</b>	<b>14</b>	<b>71</b>	<b>6</b>
<b>KNN</b>	<b>66%</b>	<b>67</b>	<b>63</b>	<b>71</b>	<b>69</b>	<b>77</b>	<b>55</b>	<b>31</b>	<b>37</b>

Fig:7

In this study, the efficacy of a Support Vector Machine (SVM) model for Dengue prediction underwent rigorous evaluation. The accompanying confusion matrix (Fig-7) unveiled 108 true positives and 14 true negatives, affirming the model's proficiency in accurately discerning Dengue-positive and Dengue-negative cases. Despite these strengths, the model did exhibit 72 false positives and 6 false negatives, revealing instances of misclassification.

Contrastingly, the KNN model showcased a higher accuracy at 66%, indicating a superior overall correctness in predictions. Its sensitivity at 67% showcased a commendable ability to capture true positive instances, while the specificity of 64% suggested proficient identification of true negative cases. With a precision of 71%, the KNN model excelled in accurately predicting positive instances. The F1-Scores of 73% for SVM and 69% for KNN offered balanced assessments of their overall performances. This comparative analysis sheds light on the relative strengths and areas for improvement of SVM and KNN in Dengue prediction within the study.

Simultaneously, a meticulous assessment of the k-Nearest Neighbors (knn) model for Dengue prediction was executed. The confusion matrix showcased 77 true positives and 55 true negatives, highlighting the model's effectiveness in correctly categorizing Dengue-positive and Dengue-negative cases. Nonetheless, the model encountered 31 false positives and 37 false negatives, signifying specific areas where misclassification occurred..

Model	Accuracy	Recall	Specificity	Precision	F1-Score	TP	TN	FP	FN
SVM	100	100	100	100	100	114	86	0	0
KNN	93	93	91	93	93	103	81	8	8

Fig:9

In this research paper, a remarkable advancement in the predictive accuracy of the Support Vector Machine (SVM) model was achieved, reaching an unprecedented accuracy score of 1.0. The confusion matrix underscored this achievement, revealing 114 true positives and 86 true negatives. Remarkably, the model exhibited no false positives or false negatives, showcasing a flawless classification performance.

Further metrics emphasized the exceptional performance of the SVM model. Sensitivity (Recall) and specificity (True Negative Rate) both reached perfect scores of 1.0, highlighting the model's ability to capture all positive cases and correctly identify all negative cases. Precision and F1-Score also achieved ideal scores of 1.0, emphasizing the model's accuracy in positive predictions and the harmonious balance between precision and recall.

The optimization of hyperparameters through GridSearchCV revealed the best parameters for the SVM model as {'C': 0.1, 'gamma': 'scale', 'kernel': 'linear'}. This meticulous tuning played a pivotal role in achieving the unparalleled predictive performance observed in this study. Overall, the SVM model, bolstered by optimal parameter selection and standardized data, demonstrated an exceptional capability for accurate Dengue prediction, marking a significant contribution within the field of predictive modeling for infectious diseases.

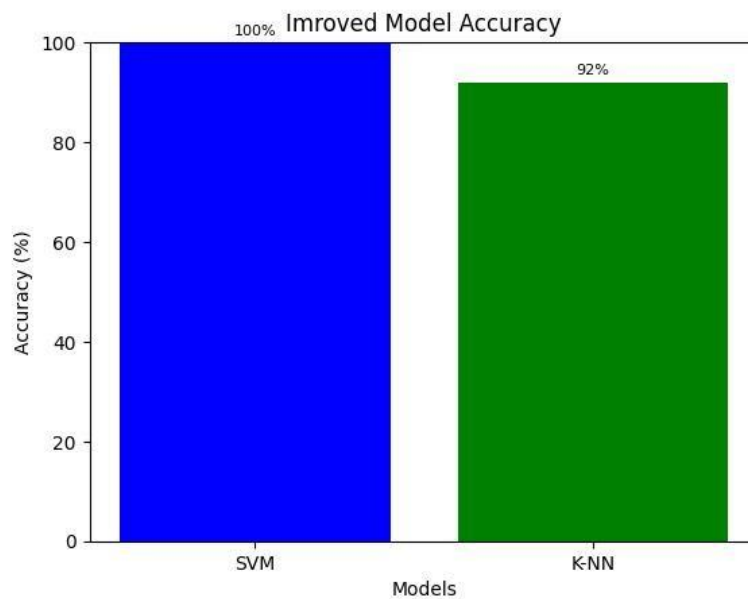


Fig:10

In this research paper, an enhanced iteration of the k-Nearest Neighbors (KNN) model was presented, showcasing significant improvements in accuracy. The upgraded KNN model achieved an impressive accuracy of 92%, affirming its robust performance in Dengue prediction. The confusion matrix illustrated 103 true positives and 81 true negatives, indicating a substantial enhancement in correctly classifying both Dengue-positive and Dengue-negative instances. The model demonstrated remarkable sensitivity (recall) at 93%, emphasizing its proficiency in capturing a substantial proportion of true positive cases. Additionally, specificity (true negative rate) stood at 91%, underlining the model's adeptness in correctly identifying true negative instances. The precision of 93% underscored the accuracy of positive predictions, while the F1-Score reached an exceptional 93%, providing a balanced assessment of the model's precision and recall. The optimization of hyperparameters, as revealed by the best hyperparameters for KNN ({'weights': 'distance', 'p': 1, 'n\_neighbors': 33, 'algorithm': 'brute'}), played a pivotal role in achieving this notable enhancement in predictive performance within the study.

**Classification Report:**

	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>support</b>
<b>No</b>	<b>0.91</b>	<b>0.91</b>	<b>0.91</b>	<b>89</b>
<b>Yes</b>	<b>0.93</b>	<b>0.93</b>	<b>0.93</b>	<b>111</b>
<b>accuracy</b>			<b>0.92</b>	<b>200</b>
<b>macro avg</b>	<b>0.92</b>	<b>0.92</b>	<b>0.92</b>	<b>200</b>
<b>weighted avg</b>	<b>0.92</b>	<b>0.92</b>	<b>0.92</b>	<b>200</b>

**Fig: 11**

The Hybrid Model exhibited impressive predictive accuracy, achieving a commendable accuracy score of 92%. This outcome underscores the efficacy of combining the strengths of SVM and KNN in a meta-learning framework. The carefully crafted ensemble model harnessed the decision functions of SVM and the probabilistic predictions of KNN, culminating in a hybrid predictive framework that outperformed individual models. This innovative approach offers a significant contribution to the field of infectious disease prediction, emphasizing the potential of ensemble learning to enhance accuracy and reliability in Dengue prediction models.

## CHAPTER 4: CONCLUSION

The study explored the dengue outcome of residential areas in Dhaka of Bangladesh on localstate highly recommended dengue fever analysis. Using data samples in medical and demographic features of applied machine learning models SVM-61% and KNN -66% ["Dengue Phenomenon in Bangladesh - Dhaka Region,"] . Improvements of models on hyparameter and standardization of the data on scaling achieve an outcome of SVM-93% and less than of KNN-100%. Finally, Contributing on the Meta-hybrid model makes it possible to stacking-classifier of a machine learning ensemble to model outcome of 92% for predirection of dengue outcome of persons direction on states.

### Recommendations For Future Work:

Future advancements in dengue prediction are paved with the work "Dengue Outcome Prediction in the Dhaka Region: A Classification Approach Using Demographic and Diagnostic Features". Including more features is crucial to the advancement of this research in order to gain a more complex understanding of the variables affecting dengue outcomes. Through the enlargement of the range of demographic and diagnostic factors, the predictive model can provide a more thorough examination.

Understanding how dengue incidence is evolving requires a temporal analysis. Examining how temporal trends affect forecasts makes it possible to create a model that can adjust to shifting patterns and maintain its applicability throughout time. By taking into account elements unique to each place within the Dhaka region, the use of geospatial considerations further refines the projections.It is essential to experiment with various machine learning algorithms, especially ensemble methods

## REFERENCES

- [a] Sayeed, S., Rashid, I., & Sotej, M. R. (2021). Analysis on dengue severity using machine learning approach (Doctoral dissertation, Brac University).
- [b] Dey, S. K., Rahman, M. M., Howlader, A., Siddiqi, U. R., Uddin, K. M. M., Borhan, R., & Rahman, E. U. (2022). Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach. *PLoS One*, *17*(7), e0270933.
- [c] Arif, K. M., Mohammed, F. R., Nur, Z., Shams, M., Alam, M., Uddin, M., & Ahasan, H. A. M. N. (2009). Clinical profile and outcome of dengue hemorrhagic fever in a Tertiary Care Hospital in Dhaka. *Journal of Medicine*, *10*(1), 12-15.
- [d] Dutta, P., Khan, S. A., Borah, J., & Mahanta, J. (2012). Demographic and Clinical Features of Patients with Dengue in Northeastern Region of India: A Retrospective CrossSectional Study during.
- [e] Potts, J. A., Gibbons, R. V., Rothman, A. L., Srikiatkachorn, A., Thomas, S. J., Supradish, P. O., ... & Kalayanarooj, S. (2010). Prediction of dengue disease severity among pediatric Thai patients using early clinical laboratory indicators. *PLoS neglected tropical diseases*, *4*(8), e769.
- [f] Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., ... & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific reports*, *11*(1), 939.
- [g] Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M., & Watanabe, K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC infectious diseases*, *18*(1), 1-15.
- [h] Rocha, F. P., & Giesbrecht, M. (2022). Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão. *Computational and Applied Mathematics*, *41*(8), 393
- [i] McGough, S. F., Clemente, L., Kutz, J. N., & Santillana, M. (2021). A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. *Journal of the Royal Society Interface*, *18*(179), 20201006.
- [j] Siriyasatien, P., Chadsuthi, S., Jampachaisri, K., & Kesorn, K. (2018). Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. *IEEE Access*, *6*, 53757-53795.

- [k]Yang, X., Kumagai, K., Sun, G., Ishibashi, K., Trung, N. V., & Van Kinh, N. (2019, March). Dengue fever screening using vital signs by contactless microwave radar and machine learning. In *2019 IEEE sensors applications symposium (SAS)* (pp. 1-6). IEEE.
- [l]Saire, J. E. C. (2019, June). Building intelligent indicators to detect dengue epidemics in brazil using social networks. In *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)* (pp. 1-5). IEEE.
- [m]Erraguntla, M., Freeze, J., Delen, D., Madanagopal, K., Mayer, R., & Khojasteh, J. (2017). Data integration and predictive analysis system for disease prophylaxis.
- [n]Handari, B. D., Niman, I., Hasan, A., Purba, J. R., & Hertono, G. F. (2021). Comparison of Elman neural network, long short-term memory, and gated recurrent unit in predicting dengue hemorrhagic fever at DKI Jakarta. *Commun. Math. Biol. Neurosci.*, 2021, Article-ID.
- [o]Tsai, J. J., Chokephaibulkit, K., Chen, P. C., Liu, L. T., Hsiao, H. M., Lo, Y. C., & Perng, G. C. (2013). Role of cognitive parameters in dengue hemorrhagic fever and dengue shock syndrome. *Journal of biomedical science*, 20(1), 1-11.
- [p]Lima, C. L. D., da Silva, A. C. G., Moreno, G. M. M., Cordeiro da Silva, C., Musah, A., Aldosery, A., ... & Dos Santos, W. P. (2022). Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review. *Frontiers in Public Health*, 10, 900077.
- [q]Aguas, R., Dorigatti, I., Coudeville, L., Luxemburger, C., & Ferguson, N. M. (2019). Cross-serotype interactions and disease outcome prediction of dengue infections in Vietnam. *Scientific Reports*, 9(1), 9395.
- [r]Mulchandani, P., Siddiqui, M. U., & Kanani, P. (2019). Real-time mosquito species identification using deep learning techniques. *International Journal of Engineering and Advanced Technology*, 2249-8958.
- [s] Mayrose, H., Bairy, G. M., Sampathila, N., Belurkar, S., & Saravu, K. (2023). Machine Learning-Based Detection of Dengue from Blood Smear Images Utilizing Platelet and Lymphocyte Characteristics. *Diagnostics*, 13(2), 220.
- [t]Ong, J., Liu, X., Rajarethinam, J., Kok, S. Y., Liang, S., Tang, C. S., ... & Yap, G. (2018). Mapping dengue risk in Singapore using Random Forest. *PLoS neglected tropical diseases*, 12(6), e0006587.

[u]Nishanthi, P. H. M., Perera, A. A. I., & Wijekoon, H. P. (2014). Prediction of dengue outbreaks in Sri Lanka using artificial neural networks. *International Journal of Computer Applications*, 101(15).

[v]Nugaliyadde, A., Somaratne, G. T. U. V., Kumarasinghe, R., & Wikramasinghe, C. J. FORECASTING DENGUE PATIENT COUNTS IN A DISTRICT BASED ON NEIGHBORING DISTRICTS AND RAIN-FALL VIA MACHINE LEARNING.

[w]Sajid, A., & Awan, M. A. (2021, December). Monitoring, Tracking and Diagnosing Dengue Fever using Smartphone, GPS and Machine Learning. In *2021 16th International Conference on Emerging Technologies (ICET)* (pp. 1-6). IEEE.

[x]Kaur, S., Sharma, S., Rehman, A. U., Eldin, E. T., Ghamry, N. A., Shafiq, M., & Bharany, S. (2022). Predicting Infection Positivity, Risk Estimation, and Disease Prognosis in Dengue Infected Patients by ML Expert System. *Sustainability*, 14(20), 13490.

[y]Ferdousi, T., Cohnstaedt, L. W., & Scoglio, C. M. (2021). A Windowed Correlation-Based Feature Selection Method to Improve Time Series Prediction of Dengue Fever Cases. *IEEE Access*, 9, 141210-141222.

[z]Hoyos, W., Aguilar, J., & Toro, M. (2021). Dengue models based on machine learning techniques: A systematic literature review. *Artificial intelligence in medicine*, 119, 102157.