

**PREDICTION OF CARDIOVASCULAR DISEASES USING MACHINE
LEARNING**

BY
MD. OSMAN HOSSAIN
ID: 232-25-018

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

Supervised By

Mr. Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc
Department of CSE
Daffodil International University

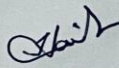


DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
January 2025

APPROVAL

This Project titled “**PREDICTION OF CARDIOVASCULAR DISEASES USING MACHINE LEARNING**”, submitted by MD OSMAN HOSSAIN, ID No : 232-25-018 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **11-01-2025**.

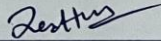
BOARD OF EXAMINERS



Dr. Sheak Rashed Haider Noori, PhD
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

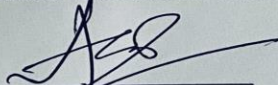
Chairman



Dr. Md. Zahid Hasan, PhD
Associate Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

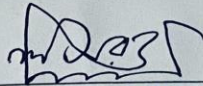
Internal Examiner



Dr. Arif Mahmud, PhD
Associate Professor & Director MIS

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Mohammed Nasir Uddin, PhD
Professor

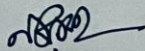
Department of Computer Science and Engineering
Jagannath University

External Examiner

DECLARATION

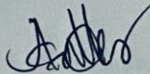
I here by declare that, this thesis has been done by us under the supervision of **Mr. Narayan Ranjan Chakraborty, Associate Professor and Associate Head, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

Supervised by:



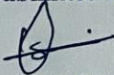
Mr. Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:



Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc
Department of CSE
Daffodil International University

Submitted by:



MD. OSMAN HOSSAIN
ID: 232-25-018
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I really grateful and wish my profound our indebtedness to **Mr. Narayan Ranjan Chakraborty, Associate Professor and Associate Head**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine Learning” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

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

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

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

Abstract

A condition of the heart or blood vessels is called cardiovascular disease (CVD). Globally, cardiovascular diseases (CVDs) constitute the primary cause of mortality. Coronary heart disease is among the most frequent cardiovascular conditions. Additional examples of coronary heart disease include aortic disease, peripheral artery disease, strokes, and decreased or obstructed blood supply to the heart muscle. Another important fact regarding cardiovascular disorders is that they are typically linked to an accumulation of fatty deposits within the arteries and a heightened risk of blood clots. In this work, I developed a cardiovascular disease prediction model using machine learning techniques, with 92.62% and 89.05% accuracy rates, respectively, for the Stacking Classifier and k-Nearest Neighbors algorithms. This model is performed exceptionally well and demonstrates the potential utility of machine learning as a tool for cardiovascular disease. In addition to Random Forest and Decision Tree, I evaluated several other algorithms, including Support Vector Machine and Logistic Regression and an ensemble model using Voting Classifier. Here, a Support Vector Machine could provide better accuracy. Our findings show that Stacking Classifiers, KNN, and machine learning have the potential to predict cardiovascular disease and imply that these technologies may prove beneficial in enhancing the disease's detection and management.

TABLE OF CONTENTS

CONTENTS	PAGE
Approval	i
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
Table of contents	v
List of Figures	ix
List of Tables	x
CHAPTERS	
CHAPTER 1: INTRODUCTION	01-05
1.1 Introduction	01
1.2 Motivation	02
1.3 Rationale Study	03
1.4 Research Questions	03
1.5 Research Objective	04
1.6 Expected Outcome	05
1.7 Layout of Research	05
CHAPTER 2: BACKGROUND STYDY	06-10
2.1 Preliminaries	06

2.2 Related Works	06
2.3 Challenges	10
2.4 Research Summary	10
CHAPTER 3: RESEARCH METHODOLOGY	11-24
3.1 Introduction	11
3.2 Data Description	11
3.3 Data Source	12
3.4 Data Pre-processing	12
3.4.1 Feature Engineering	13
3.4.2 Data Splitting	13
3.5 Data Analysis and Visualization	13
3.6 Research Subject	16
3.7 Proposed Methodology	17
3.8 Used Machine Learning Models	18
3.8.1 Logistic Regression	18
3.8.2 Stacking Classifier	18
3.8.3 Decision Tree	20
3.8.4 Random Forest	21
3.8.5 KN-Neighbors	22
3.8.6 Support Vector Machine	23
3.8.7 Voting Classifier	24
CHAPTER 4: RESULT AND DISCUSSION	25-32

4.1 Introduction	25
4.2 Performance Evaluation	25
4.2.1 Precision	26
4.2.2 Accuracy	26
4.2.3 Recall	26
4.2.4 F1 Score	27
4.2.5 ROC Curve	27
4.2.6 AUC Curve	27
4.3 Features Importance Analysis	28
4.4 Result Analysis	29
4.5 Model Performance Measurement	32
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY	33-35
5.1 Introduction	33
5.2 Impact on Society	33
5.3 Environmental Aspects	33
5.4 Ethical Aspects	34
5.5 Sustainability	34
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FUTURE RESEARCH	36-39
6.1 Introduction	36
6.2 Limitations	36

6.3 Future Work	36
6.4 Conclusions	36
REFERENCES	38-39
PLAGIARISM REPORT	40

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Missing Value count	14
Figure 3.2: Age-wise affected people	14
Figure 3.3: Data Distribution of each column.	15
Figure 3.4: Data Correlation Matrix	16
Figure 3.5: Research Flow	17
Figure 3.6: Logistic Regression ROC Curve.	18
Figure 3.7: Stacking Classifier ROC Curve.	19
Figure 3.8: Confusion Matrix of Stacking Classifier	19
Figure 3.9: Work Flow of Proposed Model	20
Figure 3.10: Decision Tree ROC Curve	20
Figure 3.11: Random Forest Classifier ROC Curve	21
Figure 3.12: K-Nearest Neighbors Classifier ROC Curve	22
Figure 3.13: Confusion Matrix of K-Nearest Neighbors Classifier.	22
Figure 3.:14 Support Vector Machine	23
Figure 3.15: Voting Classifier ROC Curve	24
Figure 4.1: Stacking Classifier Features Importance	28
Figure 4.2: Decision Tree Features Importance	28
Figure 4.3: Random Forest Features Importance	29
Figure 4.4: Accuracy of all Classifiers.	30
Figure 4.5: ROC AUC Score of all Classifiers	31
Figure 4.6: Prediction Time	31

LIST OF TABLES

TABLES	PAGE NO
Table 3.1: Features description of the dataset	11
Table 4.1: Accuracy of all models	29
Table 4.2: Score of ROC, AUC	30
Table 4.3: Algorithms Performance Measurement	32

CHAPTER 1

INTRODUCTION

1.1 Introduction

Cardiovascular disease (CVD) is still a significant threat in the quickly changing healthcare landscape, taking millions of lives annually and heavily taxing health systems across the globe. Conventional methods of diagnosis and therapy frequently cannot accurately predict the onset, course, and specific outcomes of a given patient. Allow me to present machine learning (ML), a branch of artificial intelligence with enormous potential to change how we think about cardiovascular health. This study explores the potential applications of machine learning to various aspects of cardiovascular disease. It explores how machine learning (ML) algorithms help with risk assessment, diagnosis, and illness progression prediction, improving patient outcomes and making the most use of healthcare resources. I will examine case studies and current research that demonstrate the application of machine learning in cardiovascular care, and I will talk about the successes and problems encountered in this cutting-edge sector. I aim to describe how cardiovascular care is infused with machine learning techniques. Machine learning (ML) is enhancing current procedures and opening the door for innovative diagnostic and treatment approaches. Examples include predictive analytics for risk assessment and the sophisticated interpretation of cardiac imaging. To demonstrate how machine learning (ML) improves clinical decision-making and customizes treatments for each patient, this paper will examine a number of ML applications in cardiovascular medicine. Some research articles use ensemble approaches and machine-learning techniques to create automated prediction systems for heart disease severity. It also emphasizes the significance of artificial intelligence in diagnosing cardiovascular disease. Each model's accuracy, precision, recall, and f-score were used to assess the performance.[2] This study aims to show how machine learning revolutionizes cardiovascular medicine. The report highlights another piece of research that uses patient sensations and health markers to predict disease.

The study examines the effectiveness of various algorithms, such as Random Forest, Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes, for classifying cardiovascular sickness. The article concludes that the Random Forest algorithm produces the maximum accuracy.[14] I will examine the potential of machine learning to challenge existing illness management paradigms, talk about the difficulties in combining technology and healthcare, and look forward to a time when cardiology and artificial intelligence will work together to provide more accurate, predictive medical care

1.2 Motivation

The promise of machine learning to address a significant global health concern drives research into predicting cardiovascular diseases (CVD). Given that cardiovascular illnesses are the primary cause of death globally, improved methods of prevention, early identification, and treatment are required. Machine learning is a potent tool for evaluating intricate health data, which makes it possible to create more accurate prediction models for a larger population. This method offers a shift towards individualized healthcare by considering unique aspects like genetics and lifestyle. It can also be integrated with developing technologies to enable comprehensive health monitoring. Furthermore, early intervention and customized treatment strategies made possible by machine learning can drastically save healthcare expenditures. Additionally, it may reveal particular risk factors and care-related obstacles in various communities, leading to more equitable healthcare solutions. It is estimated that cardiovascular disease (CVD) causes 17.3 million deaths annually, or almost one-third of all fatalities worldwide. CVD is expected to grow more widespread as the population ages, and lifestyle factors, including smoking, eating poorly, and not exercising, become more common. The economic expenses of CVD are estimated to be between 3 and 4 percent of global GDP. Therefore, research on CVD is crucial to create measures that effectively lower the disease's burden. This is why we are working on it, and the outcomes will help save lives, money, and time. The application of machine learning to CVD prediction is a ground-breaking advancement in healthcare that seeks to improve results, reduce expenses, and offer more customized and inclusive treatment.

1.3 Rational Study

The need to overcome the shortcomings of present healthcare practices and harness the potential of cutting-edge technology is the foundation for the research on predicting cardiovascular diseases (CVD) using machine learning. Cardiovascular illnesses continue to pose a severe threat to global health, and current prediction models frequently fall short of accurately capturing the complex interactions between hereditary, lifestyle, and environmental factors. A sophisticated approach to data analysis is provided by machine learning, which makes it possible to identify risk factors and subtle patterns that conventional methods would overlook. This improves prediction accuracy and opens the door to individualized preventative and treatment plans that customize healthcare for each patient. Furthermore, machine learning can significantly increase the effectiveness and accessibility of healthcare by expanding the availability of sophisticated predictive tools, even in environments with low resources. Integrating AI with cutting-edge health technology like wearables and telemedicine makes this research especially pertinent. This work intends to contribute significantly to machine learning and healthcare by filling a research gap. The findings have the potential to change patient outcomes and healthcare practices worldwide.

1.4 Research Objectives

In this paper, the primary objective is to design a system that is affordable, accurate, works with simple inputs, and is efficient. The dataset has first undergone preprocessing. Subsequently, I put several Machine learning algorithms into practice and assembled them. The system is optimized using feature selection and some more techniques.

1. Assemble and curate large datasets with a range of data related to cardiovascular health, including genetic markers, lifestyle factors, medical history, and clinical measurements (such as cholesterol and blood pressure). Take the required preparatory steps to handle missing values, outliers, and data standardization.

2. From the information, choose the data points most helpful in forecasting cardiovascular diseases. Look into dimensionality reduction techniques and feature selection algorithms to streamline computation and enhance model interpretability.

3. Build and train machine learning models, such as logistic regression, support vector machines, random forests, and using preprocessed data. The models can be improved by modifying the hyper-parameters and evaluating other performance measures, including accuracy and area under the receiver operating characteristic curve.

4. Utilize what you've learned to estimate a person's risk of cardiovascular disease. Analyze the models' accuracy in identifying high-risk individuals and classifying risk categories. Consider using ensemble methods or integrating multiple models to improve forecast accuracy.

Through the accomplishment of these goals, the study hopes to contribute to the development of precise and dependable machine learning models that can aid medical professionals in identifying patients who have a cardiovascular disease. This will allow for the development of tailored preventive strategies and targeted interventions that will reduce the severity of these illnesses and enhance patient outcomes.

1.5 Expected Outcomes

This work aims to demonstrate the effectiveness and accuracy of machine learning combined with a stacking classifier, KNN, to predict cardiovascular disease.

This study will contribute to our understanding of cardiovascular disease by identifying previously unknown complicated patterns and relationships. These models make Personalized healthcare tactics possible by enabling customized preventative and treatment regimens based on individual risk profiles. By facilitating the early identification of high-risk patients and facilitating prompt and efficient interventions, this customization also contributes to the improvement of preventive healthcare. This research aims to integrate machine learning into healthcare, which will lead to revolutionary advances in patient care, medical research, and public health policy.

1.6 Layout of Research

A research report's format usually consists of multiple important sections that work together to give readers a concise, well-organized study summary. The structure of the paper is as follows:

Chapter 01: Background Study. Introduction, Motivation, Rational Study, Research Question, Research Objectives, Expected Outcomes, Layout of Research.

Chapter 02: Research Methodology, Background Study. Preliminaries, Related Works, Challenges Research Summary.

Chapter 03: Result & Discussion, Introduction, Data Description, Data Source, Data Pre-processing, Feature Engineering, Data Splitting, Data Analysis and Visualization, Research Subject, Proposed Methodology, Logistic Regression, Stacking Classifier Decision, Tree Random Forest, Classifier K-Neighbors, Classifier Support Vector Machine, Voting Classifier

Chapter 04: Conclusion, Introduction, Performance Evaluation, True Negative, False Positive, Precision, Accuracy, Recall, Score, ROC Curve, AUC Curve, Features Importance Analysis, Result Analysis, Model Performance Measurement.

Chapter 05: Future Work, Introduction, Impact on Society, Environmental Aspects, Ethical Aspects, Sustainability.

Chapter 06: References, Introduction, Limitations, Future Work, Conclusion

CHAPTER 2

Background Study

2.1 Preliminaries

The first actions and arrangements required before commencing a research investigation are referred to as the preliminary stages. To prepare for the research on "Prediction of Cardiovascular Diseases using Machine Learning," a thorough foundation covering many facets of machine learning and cardiovascular diseases (CVD) is necessary. First and foremost, one must be thoroughly aware of cardiovascular disease (CVD), including its forms, epidemiology, risk factors, pathophysiology, and modern techniques for diagnosis and treatment. It is essential to comprehend this to recognize the necessity of better prediction methods. A basic understanding of machine learning is also required, including its fundamental ideas, methods, and particular uses in the healthcare industry. We used machine-learning approaches to facilitate the construction of models. I also evaluated the models' speed across several platforms. I conducted some comparative analysis with other papers in the subsequent section. Some eminent scientists making fantastic progress in this area were our sources of inspiration. For this research to be understood clearly, it is also crucial to define a literature review correctly.

2.2 Related Works

Uddin et al. say that to forecast CVD, the study uses a hybrid model for feature selection in conjunction with a neural learning technique. The writers contrast several categorization techniques, including Random Forest, XGBoost, Multi-Layer Perceptrons, and Artificial Neural Networks (ANN) (MLP). The results suggest that after employing a feature selection approach, ANN demonstrated 75% accuracy, which is comparable to Random Forest. Furthermore, it was discovered that the Decision Tree Classifier performed superior to other machine learning methods when combined with ANN and hard voting.[1] On the other hand, Alalawi et al. Talk about their research article, which uses ensemble approaches and machine-learning techniques to create automated prediction systems for heart disease severity in diagnosing cardiovascular disease.

The performance was evaluated using each model's f- score, recall, accuracy, and precision. Thus, the Random Forest model yielded the most outstanding results in the dataset on cardiac illnesses (94% accuracy), while the Gradient Boosting model yielded the best results in the dataset on cardiovascular disease (73% accuracy, 73% recall, 73% F1-score, and 74% precision).[2]

Mahmoud et al. This study compares seven machine learning algorithms for the prediction of cardiovascular disease: Random Forest, Decision Tree, Support Vector Machine (SVM), Adaptive Boosting (Adaboost), Naive Bayes, K-Nearest Neighbors (KNN), and Logistic Regression. In this instance, the machine learning models' input data generate three different assessment measures.[3]

Kumar et al. This study employs machine learning methods, including ensemble models, Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Logistic Regression and Random Forest to Predict Cardiovascular Illnesses. The paper discusses the performance of these algorithms and the importance of various features in predicting CVDs. Here, KNN has the best accuracy, 72.28%, then the stacking algorithm, 72.16%.[4]

Sufian et al. provide a thorough framework for predicting cardiac illness using cutting-edge machine learning methods, emphasizing ensemble learning, data balancing, and feature selection. Additionally, this study shows considerable gains in accuracy and efficacy compared to earlier methodologies. 87% in logistic regression, 95% in XGBoost, 83% in decision trees, 90% in random forests, randomized search CV random forests, and grid search XGBoost, and 91% in the ensemble model were the accuracy rates attained by this study model.[5]

Nadakinamani et al. This study's machine learning-based cardiovascular illness prediction system uses popular machine learning algorithms such as REP Tree, M5P Tree, Random Tree, Linear Regression, Naive Bayes, J48, and JRIP to classify cardiovascular datasets. When the performance of the suggested system was evaluated using a range of criteria,

The Random Tree model had the best accuracy of 100%, the lowest MAE of 0.0011, lowest RMSE of 0.0231, and fastest prediction time of 0.01 seconds. [6]

Kadam et al. In this study, the author compares machine learning techniques—such as image processing, acoustic signal processing, and classification algorithms—to forecast cardiovascular illnesses. The random forest classifier, which has an accuracy of 90.16% on the provided dataset, is the most appropriate. On the other hand, the accuracy of logistic regression, naïve Bayes, and decision trees is lower. The KNN's lowest accuracy in this case is 67.21%. [7]

Garg et al. This study paper primarily focuses on using machine learning algorithms to detect cardiovascular diseases. The study uses supervised learning algorithms, such as Random Forest and K-Nearest Neighbor (K-NN) to classify individuals with heart disease. K-NN predicts with 86.885% accuracy, while Random Forest achieves 81.967% accuracy. [8]

Chauhan et al. The primary focus of this research study is the use of machine learning algorithms for the early identification of cardiovascular diseases. The accuracy of many techniques is compared, including Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression. In this instance, logistic regression is determined to be the most successful method, with an accuracy rate of 89%. [9]

Tiwari et al. offer this study to assess the effectiveness of several machine-learning methods and explore the viability of utilizing machine learning to predict cardiac events. According to the findings, Random Forest is the most reliable technique for predicting cardiovascular disease, followed by Logistic Regression and KNN. Also, it discusses the use of the sigmoid function to convert predicted response values into probability values. The sigmoid function transforms predictions into probabilities, showing an S-shaped curve. [10]

Kumar et al. This study focuses on the analysis and prediction of cardiovascular disease. This time, it examines how well machine learning tree classifiers predict Cardio Vascular

Disease (CVD) and finds that, with an execution time of 1.09 seconds, ROC AUC Score of 0.8675, and accuracy of 85%, the Random Forest classifier fared better than the others.[11]

Rani et al. This study suggests a hybrid decision support system based on the patient's clinical features for early heart disease diagnosis. To handle missing values, apply a hybridized feature selection approach that combines the multivariate imputation via the chained equations method and the Genetic approach (GA) with recursive feature elimination. Using the Cleveland heart illness dataset from the UCI machine learning library, the system outperforms existing heart disease prediction algorithms with an accuracy of 86.6%. [12]

In the following discussion, Kumar G. et al. Their research proposes a model for predicting cardiac disorders, which uses machine learning techniques to facilitate efficient decision-making in the healthcare sector. It investigates the application of methods like logistic regression, Random Forest, Support Vector Machine, Gradient Boosting, and Naive Bayes classifier. Metrics like specificity, sensitivity, and classification accuracy are used to assess how well these models work. The significance of data preprocessing and machine learning in the diagnosis of cardiovascular disorders is emphasized in the study. [13]

Rubini PE et al. The creation of an application that predicts disease based on patient symptoms and health markers is highlighted in the study. To classify cardiovascular illness, the study analyzes the efficacy of several algorithms, including Random Forest, Logistic Regression, Support Vector Machine (SVM), and Naïve Bayes. The paper concludes that the Random Forest algorithm yields the highest accuracy.[14]

Lastly, Louridi et al. This study examines the accuracy, precision, sensitivity, recall, and other key performance indicators of several machine-learning algorithms, including SVM, Naïve Bayes, KNN, and others. The study highlights how crucial preprocessing is to machine learning, especially when substituting missing data with mean values. According to the paper's conclusion, preprocessing is essential to getting better outcomes from machine learning applications used in medical diagnosis, especially in predicting cardiovascular disease.[15]

2.3 Challenges

This research has several difficulties. The study's effectiveness and dependability depend on addressing several issues faced by machine learning research on cardiovascular disease prediction. One of the main problems is the availability and quality of the data, which is frequently problematic because of dataset consistency and privacy issues. Furthermore, the research has to address how difficult it is to handle various data types—like genetic data and medical records—and successfully combine them for machine learning analysis. Another major problem is choosing and controlling the complexity of several machine learning algorithms, particularly when determining which one best suits the unique requirements of cardiovascular disease prediction. In general, to ensure the precision, ethics, and therapeutic value of machine learning applications in cardiovascular disease prediction, these problems must be resolved.

2.4 Research Summary

A significant development in the use of artificial intelligence in healthcare is the paper "Prediction of Cardiovascular Diseases using Machine Learning." To develop machine learning models to forecast cardiovascular diseases (CVD), the leading cause of mortality globally, this study thoroughly analyzes various data sets. To increase the accuracy and efficacy of CVD prediction, researchers are looking for patterns and risk factors that traditional statistical analysis cannot detect. One of the main areas of attention is the practical use of these machine learning models in clinical settings, with a particular emphasis on helping medical professionals with early diagnosis and individualized treatment planning to enhance patient outcomes. This work adds significantly to the crucial topic of cardiovascular disease prognosis. It comprehensively analyzes how machine learning can transform CVD diagnosis and treatment.

CHAPTER 3

Research Methodology

3.1 Introduction

This methodology section describes the methodological strategy used to forecast cardiovascular diseases (CVD) using machine learning techniques. The study entails gathering pertinent data, preparing it for analysis, choosing suitable machine learning models, training and validating these models, and assessing their effectiveness.

3.2 Data Description

Essential variables are medical history (BPMeds, prevalentStroke, prevalentHyp, diabetes), lifestyle factors (is_smoking, cigsPerDay), and demographic data (age, sex). The presence or absence of cardiovascular disease is the outcome variable.

Table 3.1: Features description of dataset

Features Name	Description	Type
age	Patient age	Nominal
sex	Male or Female	Continuous
is_smoking	Smoke or not	Nominal
cigsPerDay	Day wise cigarettes counts	Continuous
BPMeds	Blood Pressure	Nominal
prevalentStroke	experienced a stroke in the past or not	Nominal
prevalentHyp	Is patient hypertensive or not?	Nominal
diabetes	Is this patient contains diabetes or not?	Nominal
totChol	Level of Cholesterol in total	Continuous
sysBP	This Patient Systolic Blood Pressure	Continuous
diaBP	This Patient Diastolic Blood Pressure	Continuous

BMI	Patients body mass index	Continuous
heartRate	Patients heart rate condition	Continuous
glucose	Level of glucose	Continuous
TenYearCHD(Predictvariable)	Predict term	Binary

3.3 Data Source

The information is utilized in this study to forecast cardiovascular disease and to compare the findings with those of earlier studies. It includes a significant amount of patient data, including medical records. Kaggle gathered the dataset. The dataset is used to create validation, testing, and training sets. The clean version of the cardiovascular diseases dataset has the following shape: instance 3390 and features 14.

3.4 Data Pre-processing

To build a more accurate machine learning model, data preprocessing is required. Data preprocessing is the term for the data cleaning procedure. I will remove all NAN values from our dataset. This process is also known as data wrangling. This involves figuring out which data could be more consistent, noisy, and missing. The transformation and modeling preparation of raw data is a crucial stage in machine learning. This procedure entails standardizing or normalizing the data for uniformity, addressing mistakes and missing values, and creating features to improve model performance. It also entails identifying the most pertinent features and minimizing their quantity to avoid overfitting and streamline computations. Another crucial step is to convert categorical data into a numerical representation. Additionally, balancing unequal datasets receives particular attention. I have used techniques like SMOTE, Label Encoder, and Removing outliers, which are very useful for this dataset

3.4.1 Feature Engineering

Computers cannot comprehend the expression of a dataset's nominal qualities. Therefore, I need to transform the data into a format that machines use. The yes and no values have been converted to 1 and 0, respectively, using Sklearn label encoding. Machine learning models' performance is significantly impacted by feature engineering, a critical process in which domain expertise is used to identify and choose the most pertinent characteristics from raw data. This procedure entails extracting new features from the available data, converting features into a format better suited for modeling, and using feature selection to determine which characteristics are most pertinent to the model. Simplifying the model without losing essential information also entails addressing missing values in the dataset, translating categorical variables into a numerical format, and minimizing the number of features.

3.4.2 Data Splitting

To train, validate, and test a machine learning model, the available dataset must be split into various sets, known as data splitting. This division is essential for an impartial assessment of a model's performance. Depending on the size and makeup of the dataset, the split's precise proportion may change.

Here, two portions of the dataset are separated: one for testing and one for modeling. Seventy percent of the data is used for training the model, and just thirty percent is used for testing. Seventy-five percent of the data is used for training the model, and just twenty-five percent is used for testing. Stratify the assertion that training and testing sets represent every class equally.

3.5 Data Analysis and Visualization

This graph, which I created using `show_missing()`, shows the percentage of missing data in each column of the `df_copy` DataFrame (or a subset of its columns). The plot's bars are color-coded to make them easier to see.

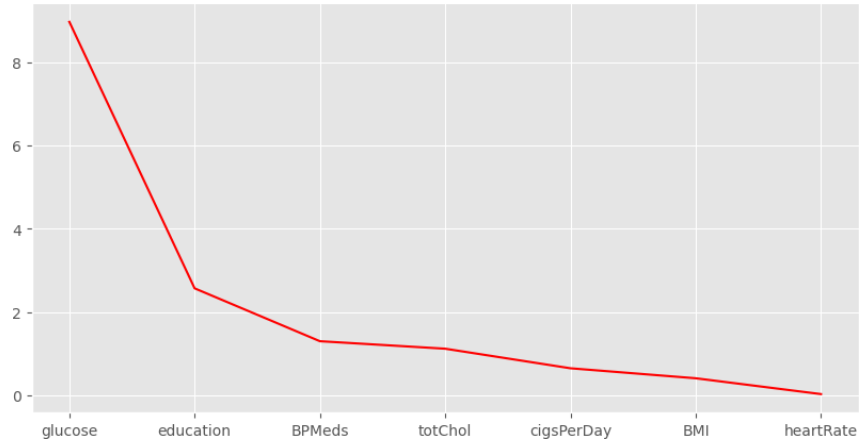


Figure 3.1: Missing Value Count

This figure shows the age distribution in the dataset based on the probability of getting coronary heart disease in the next ten years. In this figure, you may compare the age distribution of individuals with and without the risk of CHD because each age group will have its density curve.

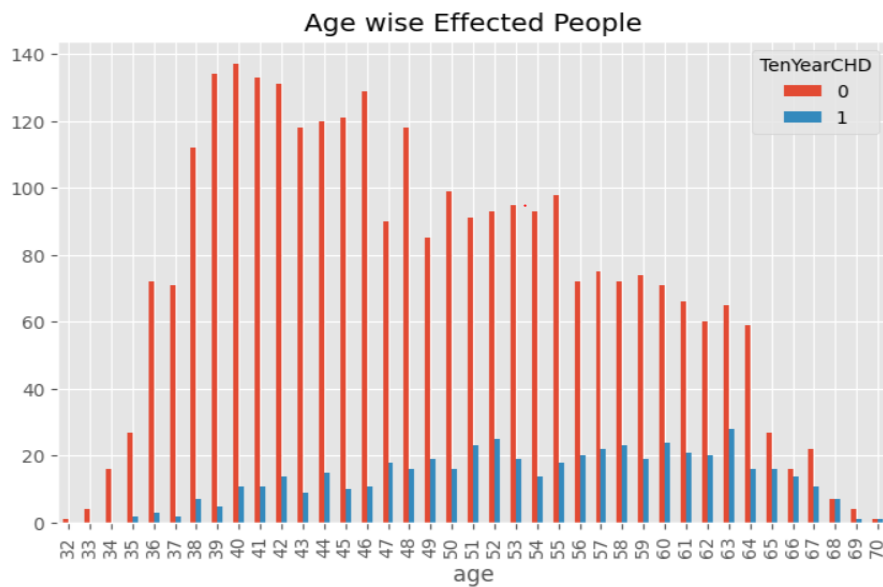


Figure 3.2: Age-wise affected people

In addition to understanding the distribution of data and data exploration, we can view a visual representation of the data distribution for each column in this instance.

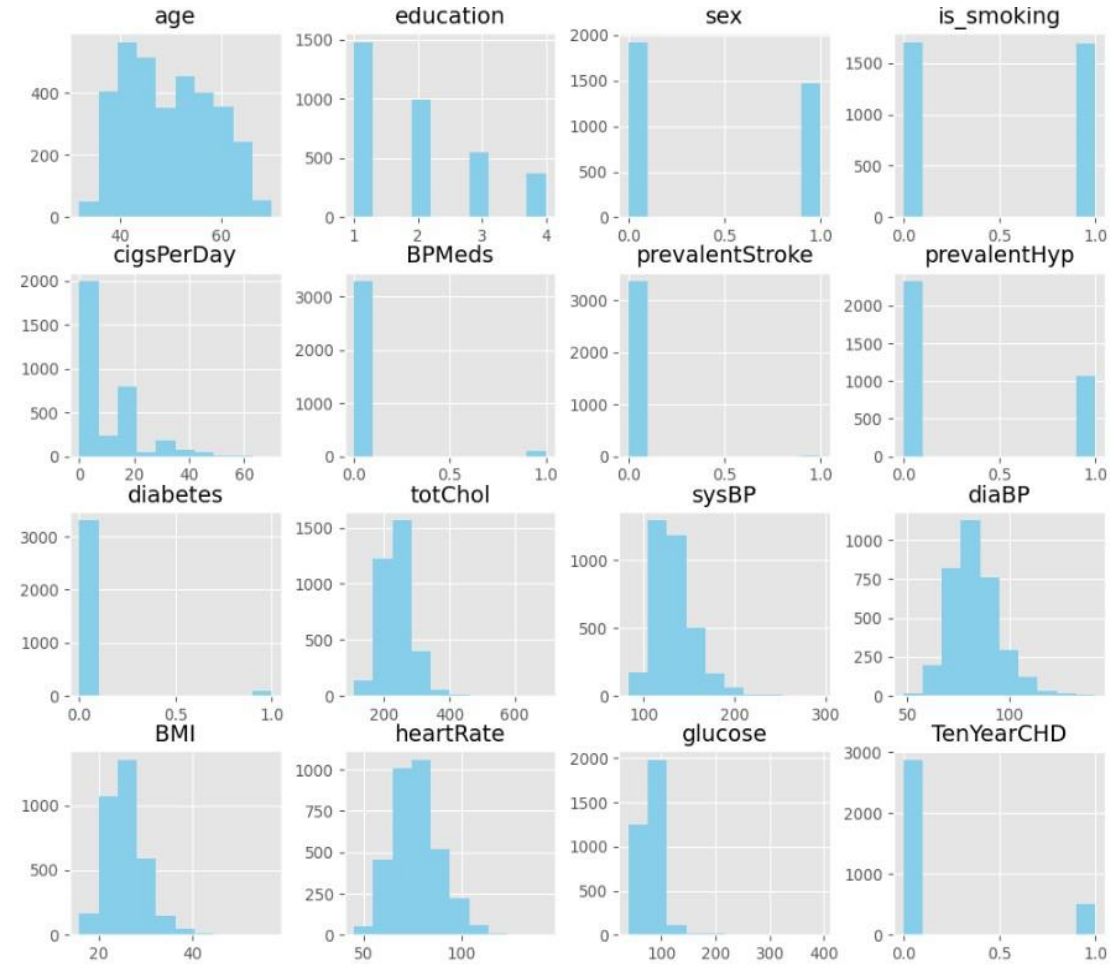


Figure 3.3: Data Distribution of each column.

In this case, I have produced a heatmap depicting a DataFrame's correlation matrix. This is quite helpful for machine learning feature selection or comprehending the relationships in a dataset for statistical analysis.

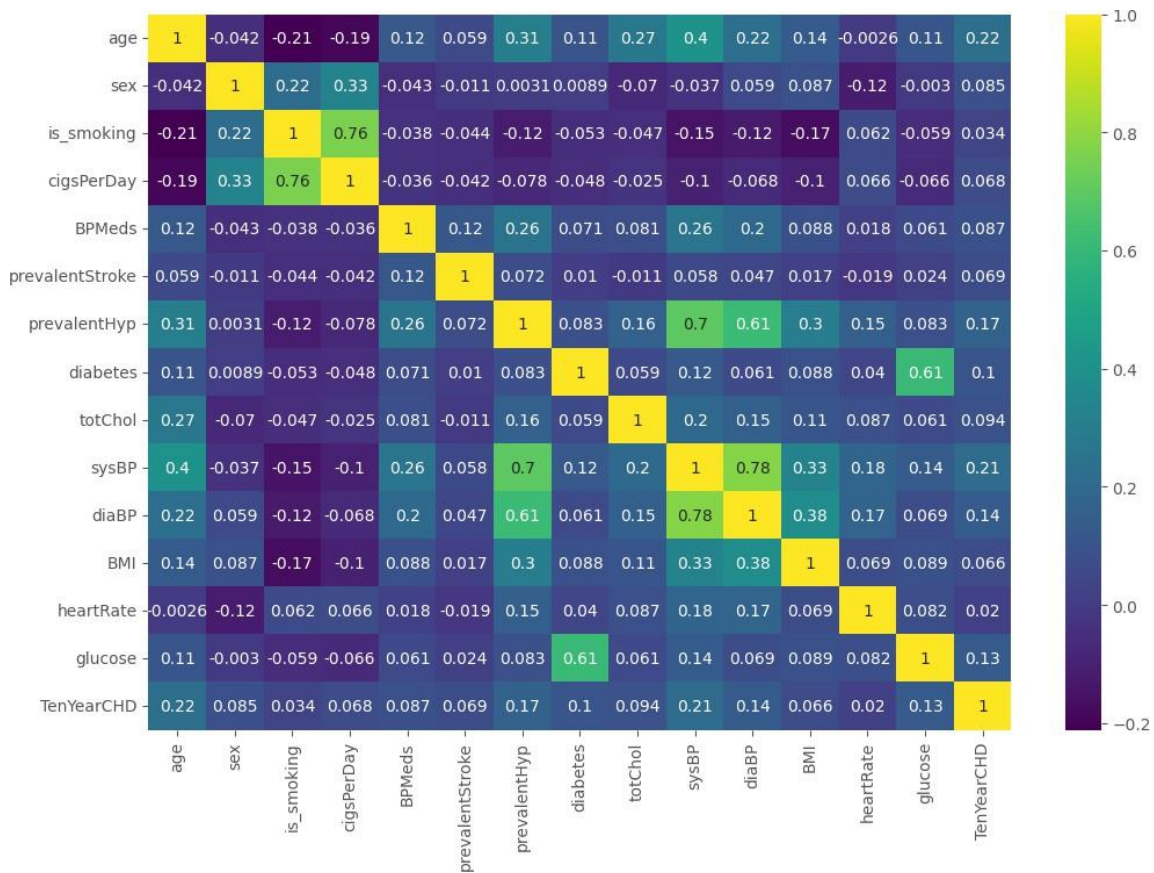


Figure 3.4: Data Correlation

3.6 Research Subject

Modeling cardiovascular disease prediction with machine learning is the focus of my study. The primary area of interest for this research is machine learning techniques. Models for the experiment are created using Google Colaboratory. Seaborn library, NumPy, pandas, sci-kit-learn, and pandas were used to develop an ML-based prediction model. The general programming language utilized was Python. I used Google Colaboratory to build the models and run the simulations.

3.7 Proposed Methodology

This suggested model diagram would be a visual aid to illustrate the data's path through the system. It would describe how different sources of data are compiled. The process of converting unprocessed data into an analytically ready format is clearly shown in this picture, as is the progression toward the final product. This predictive analysis indicates the probability and prediction of cardiovascular illnesses.

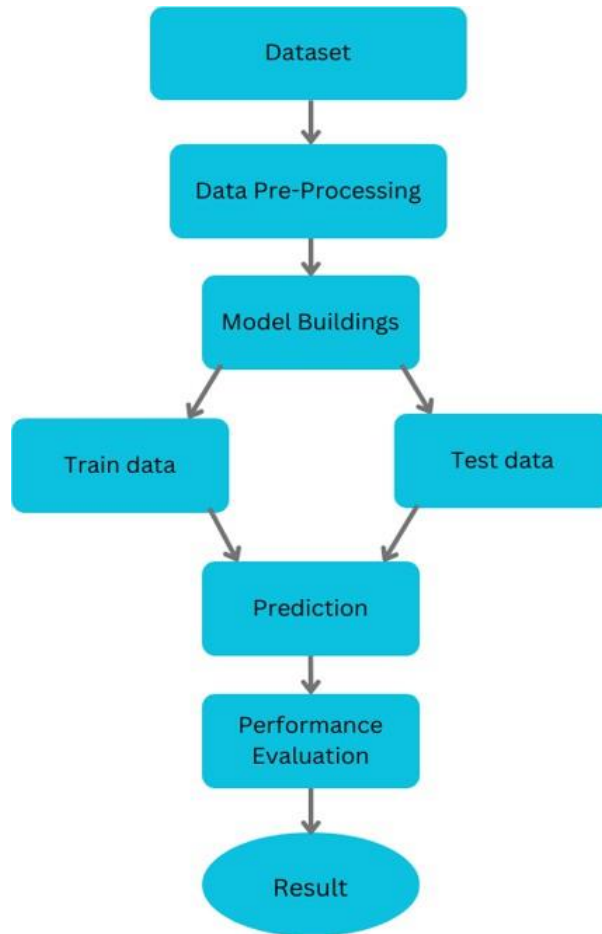


Figure 3.5: Research Flow

3.8 Used Machine Learning Model

3.8.1 Logistic Regression:

Logistic regression is a fundamental and well-liked machine learning technique that works particularly well for classification applications. Although logistic regression has model extensions for multiclass classification, it is fundamentally a statistical model that models a binary dependent variable using a logistic function. Although it can be expanded to multiclass classification, its main application is binary classification issues (1/0, Yes/No, True/False). When using binary classification, an event is expected to occur (class 1) if the output probability is more than a threshold (usually 0.5); if it is less than the threshold, the event is forecast not to occur (class 0).

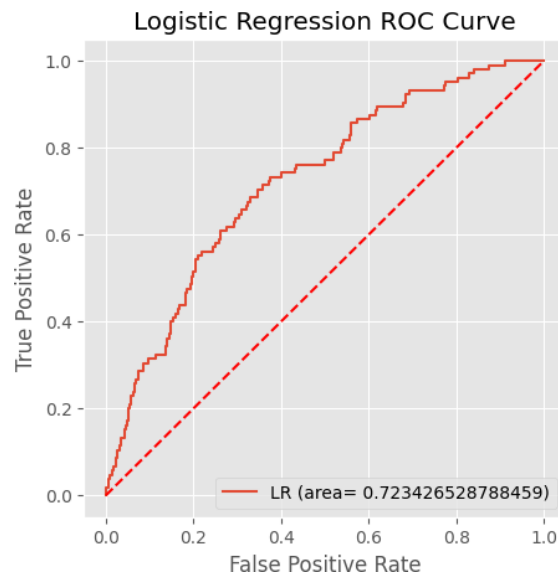


Figure 3.6: Logistic Regression ROC Curve.

A logistic regression model assists us in resolving the scenario where the output can only take the values 1 and 0 by using the sigmoid function. Here, the accuracy is 87.77%.

3.8.2 Stacking Classifier

This is my proposed model. An ensemble machine learning method called stacking, or stacked generalization, combines several base models, or learners, to enhance prediction performance.

A range of methods can be employed as basis models in stacking, such as decision trees, random forests, support vector machines, k-nearest neighbors, and more. Using the same dataset, this model is independently trained.

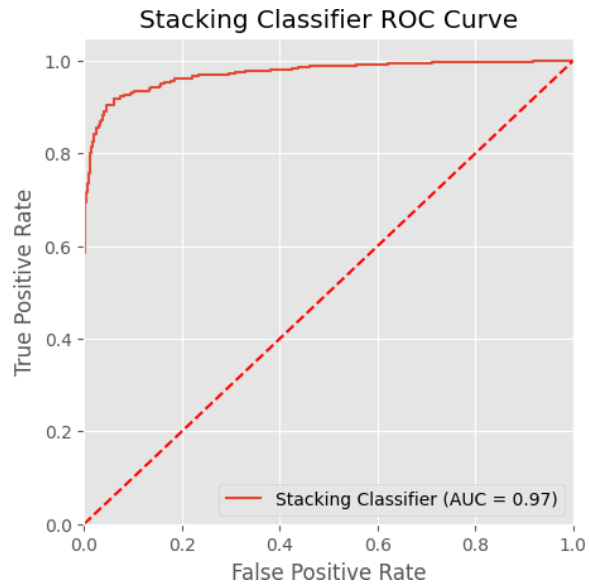


Figure 3.7: Stacking Classifier ROC Curve.

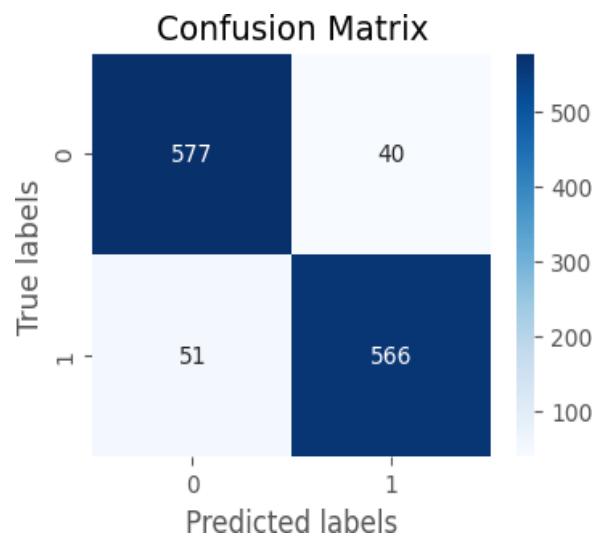


Figure 3.8: Confusion Matrix of Stacking Classifier

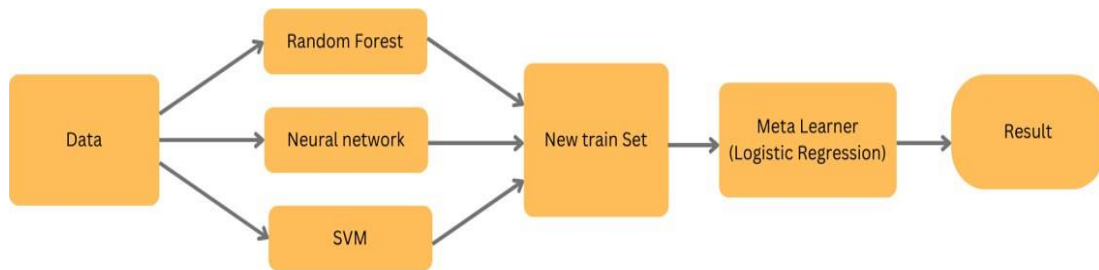


Figure 3.9: Work flow of proposed model

As a final estimator, we employed random forest, neural network, and SVM in our model. The accuracy of 92.62% achieved here is the highest for this dataset and is rather good.

3.8.3 Decision Tree

Regression and classification are two uses for the flexible and simple-to-understand machine learning technique: the decision tree.

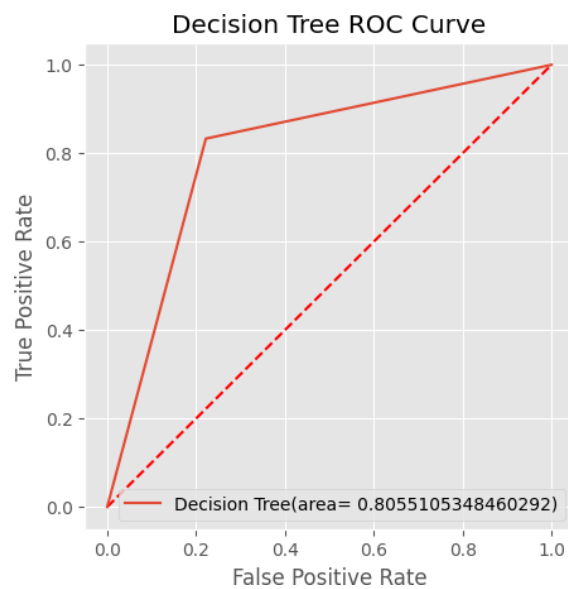


Figure 3.10: Decision Tree ROC Curve

It's a supervised learning system that separates the data into subsets based on key attributes and then learns to make predictions recursively. Decision trees are popular because they are helpful for decision-making and the creation of insights, and they are simple to comprehend and depict. According to our research paper, the Decision Tree's accuracy is 80.55%, which is not good compared to other algorithms.

3.8.4 Random Forest Classifier

For machine learning applications, including regression and classification, we know Random Forest is a well-liked ensemble learning technique. It is renowned for its excellent accuracy and resilience and for handling sophisticated datasets. To produce more dependable and accurate predictions, Random Forest constructs several decision trees and aggregates their forecasts. To arrive at a final prediction, predictions from many models are integrated. The Random Forest algorithm integrates the estimates from several decision trees.

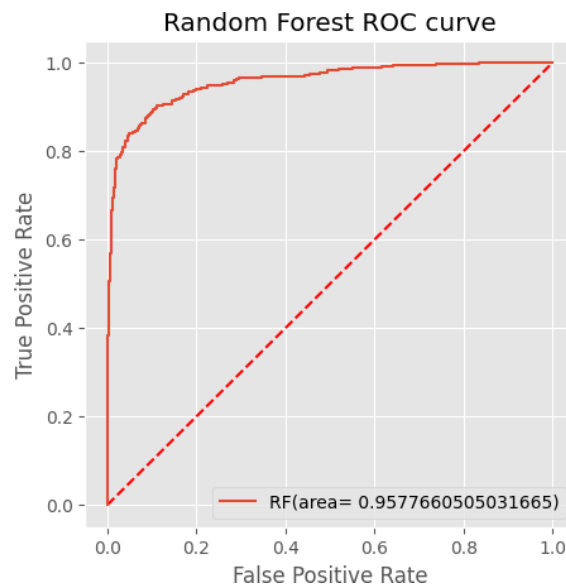


Fig 3.11: Random Forest Classifier ROC Curve

In my study report, I discovered that the Random Forest algorithm has an accuracy of 88.16%, which is good accuracy in comparison to other algorithms.

3.8.5 K-Neighbors Classifier

The K-Nearest Neighbors (KNN) Classifier is a simple yet effective machine-learning method for regression and classification. It belongs to the family of instance-based or lazy learning algorithms and is well-known for being easy to use. In the feature space KNN uses, data points are represented as vectors with many features. Each data point is a single point in this feature space

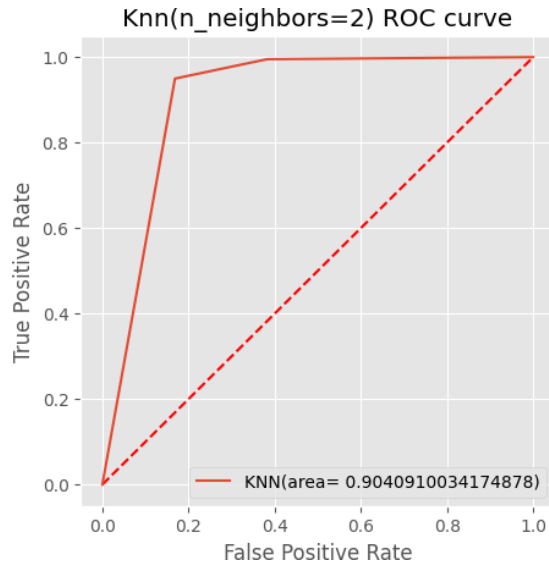


Figure 3.12: K-Neighbors Classifier ROC Curve

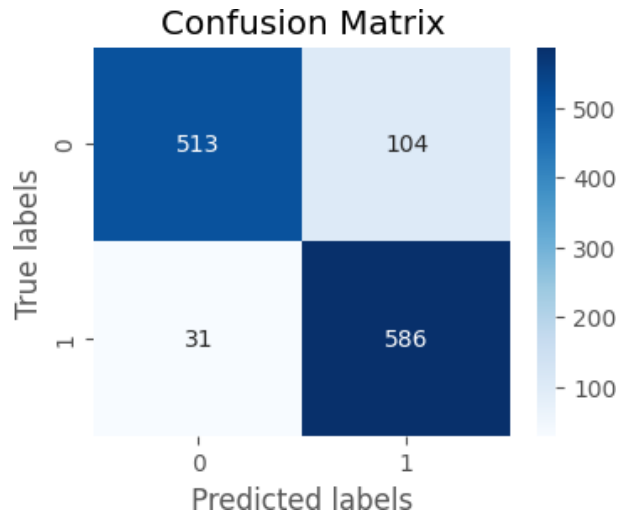


Figure 3.13: Confusion Matrix of K-Nearest Neighbors Classifier.

A positive integer indicates how many closest neighbors should be considered for forecasting. The method will consider the three nearest neighbors, for instance, if $K=3$. The accuracy of the K-Nearest Neighbors Classifier in this research article is 89.05%. This yields a second-height accuracy among the algorithms and gives good accuracy compared to others.

3.8.6 Support Vector Machine

Support Vector Machine is a powerful and versatile machine-learning technique for regression and classification applications (SVM). SVM is renowned for its capacity to manage intricate data patterns and establish precise decision limits. To anticipate continuous values or effectively divide data points into multiple groups, it locates the hyperplane that best fits the data. SVM is beneficial for linearly separable data, which may be divided into two or more classes using a plane (in 3D), a hyperplane (in higher dimensions), or a straight line (in 2D). However, SVM can also handle non-linear data

(explained below) by using kernel functions.

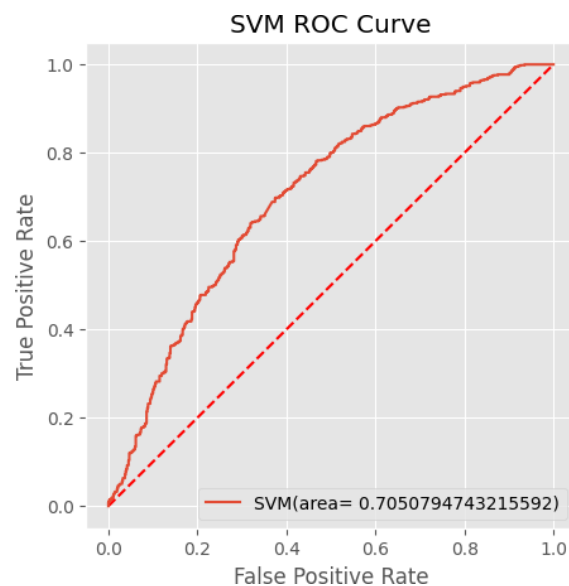


Figure 3.14: Support Vector Machine ROC Curve

According to this study article, the support vector machine's accuracy is 65.72%. This algorithm has the lowest accuracy and does not perform well compared to all the others.

3.8.7 Voting Classifier

The Voting Classifier is an ensemble learning method that combines the predictions from different machine learning models to produce a final prediction. This kind of ensemble approach increases overall forecast accuracy by utilizing a variety of models. The Voting Classifier amalgamates the forecasts from several underlying models or classifiers. These foundational models may have several types or even be the same type with altered hyperparameters.

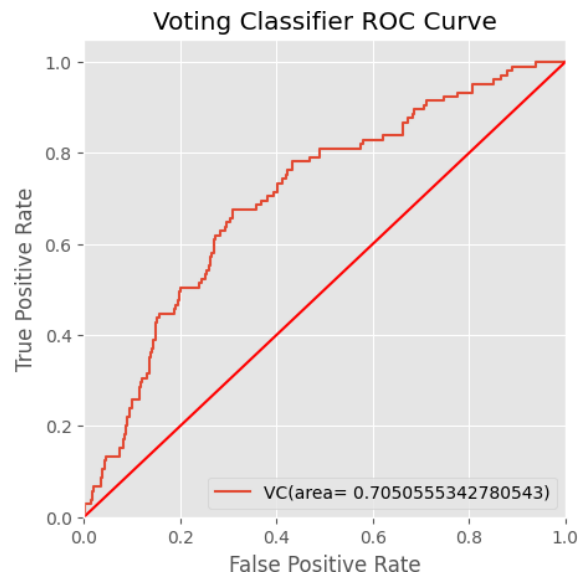


Figure 3.15: Voting Classifier ROC Curve

I have found that the voting classifier's accuracy in this study article is 87.77%. This one offers good accuracy compared to other algorithms. Five algorithms were combined here: decision trees, Support Vector Machines, Random Forests, K-neighbors classifiers, and Logistic Regressions.

CHAPTER 4

Result and Discussion

4.1 Introduction

In the preceding section, I covered the dataset, dataset processing methods, and machine learning models. This section will describe the evaluation process and the outcomes of the models that use the processed data. To determine which machine learning algorithm produces the best accuracy, several were tested, and the results are being examined.

4.2 Performance Evaluation

To determine the optimal classification algorithm, this study's training on the testing dataset involves trying a variety of classifiers.

A range of performance evaluation measures are used to assess classifiers. In machine learning, evaluation determines a model's correctness, efficacy, and efficiency. Typically, it involves applying predefined metrics to compare the model's predictions with the actual outcomes. Accuracy, precision, recall, and F1 score are crucial metrics for classification tasks; mean squared error (MSE) or mean absolute error (MAE) are essential for regression tasks. Recognizing the benefits and drawbacks of the model can influence future work toward enhanced performance on untested data. These measurements are derived from the true positives, false negatives, and false positives parameters of each classifier's confusion matrix.

A True positive (TP) occurs when the observed and anticipated values are both positive. When the expected values are negative, but the observed values are positive, this is known as a false negative (FN).

True Negative: When both the expected and observed values are negative, the situation is said to be true negative (TN).

False Positive: A negative observation is the foundation for an optimistic forecast when the observed values are negative even though the expected values are positive. The following evaluation metrics are computed to determine each classifier's performance based on the values of TP, FN, TN, and FP of the confusion metrics:

4.2.1 Precision: Precision is a metric for positive forecast accuracy. The precision of a machine learning model, or its ability to provide good predictions, is one measure of the algorithm's effectiveness. The percentage of genuine positives to all optimistic predictions is known as precision.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

4.2.2 Accuracy: The standard metric in machine learning is accuracy. It's a statistic used to assess how well a model performs in categorization tasks. Accuracy is the proportion of all input samples to all correctly predicted samples. It can be calculated by dividing the total number of forecasts by the number of correct guesses.

$$Accuracy = \frac{TP+TN}{TP+FP+FP+FN} * 100\% \quad (2)$$

4.2.3 Recall: Recall measures the completeness of optimistic predictions. A machine learning model correctly classifies the proportion of data samples from all samples for a class as being in the positive class of interest.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4.2.4 F1 Score: Precision and recall are harmonic means. When memory and precision must be balanced, this may be helpful.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

4.2.5 ROC Curve: A receiver operating characteristic (ROC) curve shows the overall classification levels of a categorization model. A visual aid for evaluating the effectiveness of classification models in machine learning is the Receiver Operating Characteristic Curve or ROC Curve. It provides insight into the trade-off between accurately recognizing positives and incorrectly identifying negatives at different threshold levels by plotting the True Positive Rate (Sensitivity or Recall) against the False Positive Rate. The curve makes comprehending how well a model differentiates across different groups easier. ROC curves become particularly helpful when there is a significant variation in the outcomes of several categorization error kinds or when datasets are imbalanced.

4.2.6 AUC Curve: AUC, or area under the curve, has been used to gauge performance based on numerous factors. The area under the ROC (Receiver Operating Characteristic) curve, or AUC for short, is a machine learning metric used to evaluate classification models' performance. The AUC is beneficial since it is unaffected by the dataset's class distribution and classification cutoff. Plotting the True Positive Rate against the False Positive Rate for various threshold values is what this curve does. The AUC's single probability value measures the model's capacity to distinguish between positive and negative classes. An AUC of 1 denotes a perfect model, but an AUC of 0.5 means the model performs no better than random guesswork. Because the AUC is not affected by the class distribution or the classification threshold, it is a reliable metric for comparing models, especially when dealing with unbalanced datasets. The more accurately the model can classify the various classes, the better its AUC.

4.3 Features Importance Analysis

In machine learning, feature importance aids in identifying how each input information affects a model's prediction. It is essential for comprehending the model, simplifying it by eliminating aspects that aren't as vital, enhancing performance, and learning more about the data. The evaluation of feature importance can be achieved by employing tree-based techniques such as Random Forest and looking at coefficients in linear models. Below, we can see the figure for the essential features in our model.

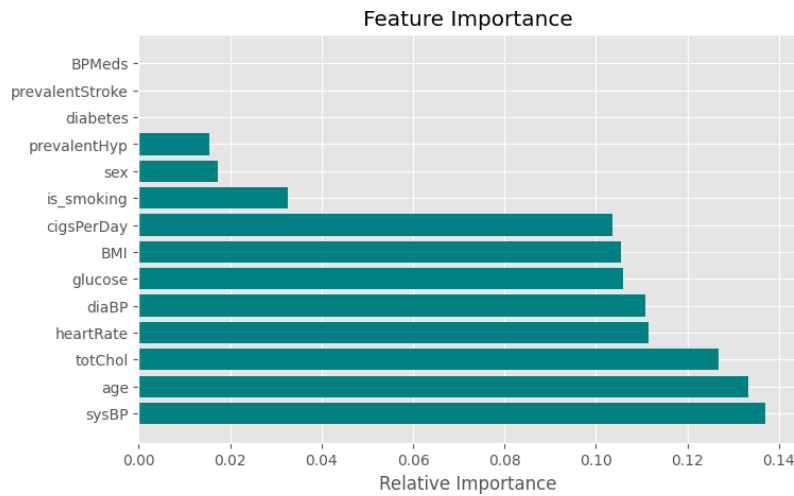


Figure 4.1: Stacking Classifier Features Importance

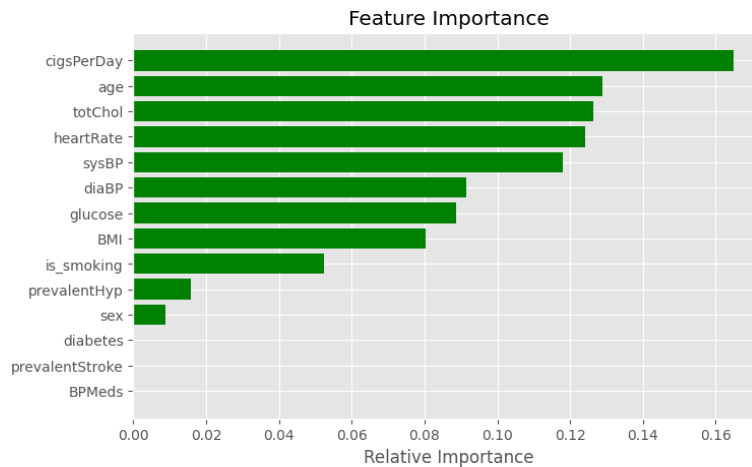


Figure 4.2: Decision Tree Features Importance

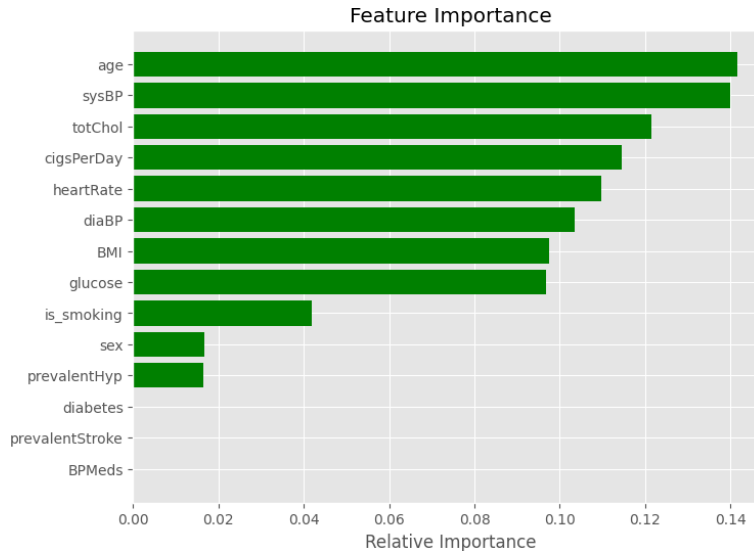


Figure 4.3: Random Forest Features Importance

4.4 Result Analysis

Based on trials employing a range of machine learning classification methods, the cardiovascular disease dataset will be used to detect cardiovascular disease automatically. The classifier's output following feature selection and data processing is displayed in Table II below. The model with the most robust performance in predicting the disease was the Stacking Classifier, which came in second at 92.62% accuracy. SVM had the lowest accuracy of any model, at 65.72%.

Table 4.1: Accuracy of all models

Algorithms	Accuracy
Logistic Regression	87.77%
Stacking Classifier	92.62%
Decision Tree	80.55%
Random forest	88.16%
K-Nearest Neighbors Classifier	89.05%
Support vector Machine	65.72%
Voting Classifier	87.77%

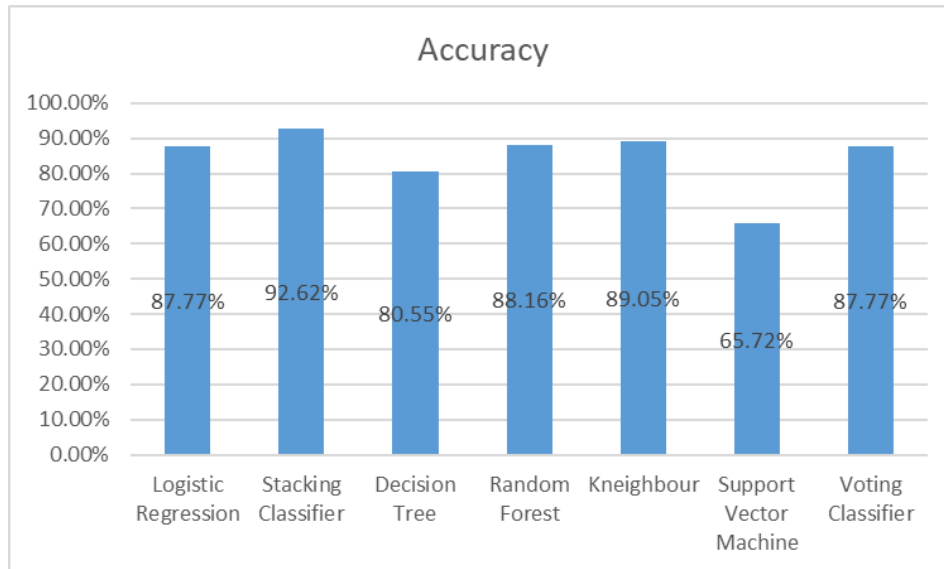


Figure 4.4: Accuracy of all Classifiers.

I will now discuss the ROC AUC Score, which was derived from experiments using several machine-learning methods for classification.

Table 4.2: Score of ROC, AUC Score

Algorithms	ROC AUC Score
Logistic Regression	72%
Stacking Classifier	97%
Decision Tree	80%
Random forest	95%
K-Nearest Neighbors Classifier	90%
Support vector Machine	70%
Voting Classifier	70%

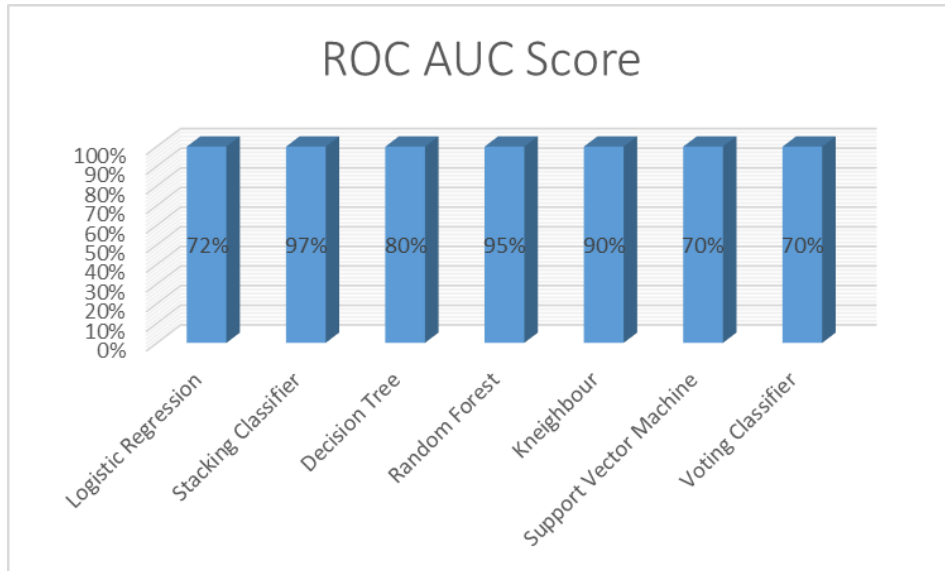


Figure 4.5: ROC AUC Score of all Classifiers.

Figure 14 displays the classifier's output following feature selection and data processing. In terms of the disease's ROC AUC Score, the Stacking Classifier model performed best, with a 98% accuracy rate. SVM had the lowest ROC AUC Score of any model, at 70%, followed by the Voting Classifier, at 67%.

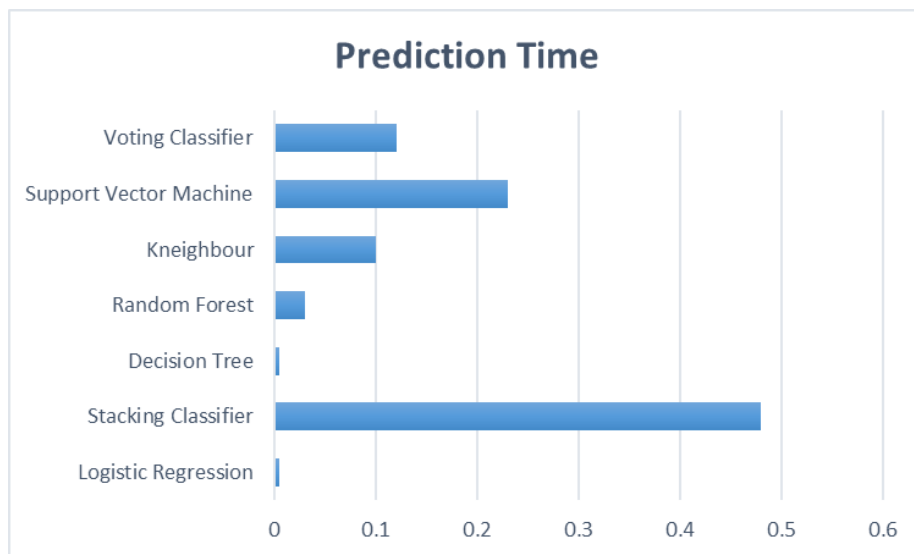


Figure 4.6: Prediction Time

In this research, we have calculated all model time predictions. In the above chart, the algorithm that needs the most time is the Stacking Classifier, and the algorithm that needs the least time is the Decision Tree and Logistic Regression.

4.5 Model Performance Measurement

Evaluating a model's overall performance, particularly in prediction accuracy, efficiency, and efficacy, is known as performance measurement in machine learning. It is crucial to comprehend the benefits and drawbacks of a model to guide future advancements. While analyzing imbalanced datasets, a confusion matrix helps visualize prediction accuracy and get other vital metrics such as precision, recall, and the F1 score. The ROC-AUC score is an additional crucial measure for classification models. The problem specifics, the data at hand, and the model under consideration affect which evaluation methods are applied.

Table 4.3: Algorithms Performance Measurement

Algorithms	Accuracy	Precision	Recall	F1-Score
Logistic regression	87.77%	0.88	1.00	0.93
Stacking Classifier	92.62%	0.92	0.94	0.93
Decision Tree	80.55%	0.82	0.78	0.80
Random forest	88.16%	0.91	0.85	0.88
K-Nearest Neighbors Classifier	89.05%	0.95	0.83	0.88
Support Vector Machine	65.72%	0.68	0.60	0.64
Voting Classifier	87.77%	0.88	1.00	0.93

Performance measurement is a crucial step in the machine learning process. It directs feature engineering, model selection, and hyperparameter tweaking. The type of data determines which metrics and techniques are used.

CHAPTER 5

Impact on Society, Environment, Ethical Aspects, and Sustainability

5.1 Introduction

I looked at the potential societal effects of this study. This chapter has covered the patient's impact on cardiovascular disease, cultural ramifications, morality, and patient benefits. We have now discussed the project's viability and room for expansion in order to assist more people down the road.

5.2 Impact on Society

The field of machine learning-based research on the prediction of cardiovascular disease is expected to have significant societal effects, especially in the medical domain. Its most significant impact is improving cardiovascular disease prevention and early diagnosis, patient outcomes, and mortality rates. Offering customized risk assessments and treatment plans enhances personalized healthcare and increases medical therapies' efficacy. The potential for lower healthcare expenditures is a significant outcome of this research since early therapies can avoid the need for later, more costly treatments. Furthermore, the knowledge gained from this study is beneficial for planning and guiding public health policies, enabling more focused health campaigns and preventative actions. In addition, it discusses healthcare fairness and accessibility, which aids in customizing health treatments for underprivileged groups and various demographics. Furthermore, the study promotes healthy lifestyle choices by increasing public knowledge of cardiovascular disorders.

5.3 Environmental Aspects

Environmental factors, including lifestyle choices, socioeconomic position, and other ecological factors, significantly influence the risk and prevalence of cardiovascular illnesses. These essential elements add to the complexity of

understanding and treating cardiovascular disease. Developing comprehensive strategies for preventing and managing cardiovascular illnesses requires addressing these environmental factors. When a patient is afflicted with a sickness, they must endure tremendous suffering to identify the condition and receive treatment. They had to go to the doctor's chambers and wait a lengthy time to see the doctor. The patient's precious time is wasted in this way. Sometimes, it results in patients dying because they don't receive the proper care promptly. However, patients can recover quickly if they employ AI-based technology to diagnose their illness. Once more, it facilitates prompt diagnosis of diseases by medical professionals, allowing patients to receive appropriate recommendations and treatment on time. Developing comprehensive strategies for preventing and managing cardiovascular illnesses requires addressing these environmental factors.

5.4 Ethical Aspects

Many significant ethical considerations must be made when researching cardiovascular illnesses to preserve the integrity of the study and its societal usefulness. Informed consent is essential, requiring that participants are fully informed about the research's goals, methods, and rights. Another crucial component is preserving the privacy and confidentiality of patient data and susceptible health information, which necessitates following data protection regulations and using secure data processing techniques. Several researchers created fictitious datasets or altered data to improve their study outcomes. It is unethical in the extreme. This kind of research can potentially ruin patients' lives in medicine. Inadequate care can result in patient death. Therefore, we must strictly uphold these moral principles.

5.5 Sustainability

Sustainability in researching the potential applications of machine learning to predict cardiovascular diseases (CVD) refers to the significant consequences on humans. I considered the sustainability plan for our research because of this. I plan to incorporate reinforcement learning into our model in the future. This will

continually release new models by gathering fresh data, training, and applying our model. We will be able to forecast this illness more precisely through this technique.

CHAPTER 6

Conclusions and Future Work

6.1 Introduction

In this chapter, I summarized my findings and discussed my plans for this research. You can use a sophisticated method or improve the model's efficiency. Thus, this chapter provides a concise summary of the project's conclusions. The final section of the chapter is a list of the references used for the research.

6.2 Limitations

- The dataset needs to be more extensive and more accurate.
- The perfect prediction may occasionally have a problem.
- In that case, you need to determine how to get more precise findings and which prediction model works best.
- When applying it to real-world situations, difficulties could occur.

6.3 Future Work

Numerous studies have addressed this subject, and I attempted to predict which algorithms would produce the most efficient results. In the future, I want to use these models in a real-world scenario and work on creating a more precise model that can be used to treat cardiovascular disease in patients in clinics, hospitals, and other healthcare settings before it becomes worse.

6.4 Conclusion

To predict cardiovascular disease (CVD), machine learning classifiers, including Random Forest, Decision Tree, Logistic Regression, Stacking Classifier, Voting Classifier, Support Vector Machine (SVM), and K-nearest neighbors (KNN) were employed in this study. The suggested approach, which classified patients with cardiovascular disease using a machine learning classifier called Stacking Classifier, beat all other classifiers examined in terms of accuracy, achieving a higher 92.62% with a ROC AUC score or

ROC AUC score of 97%. So, this model will play a vital in the medical sector. Lastly, this study will help a lot of cardiovascular disease patients.

REFERENCES

- [1] Uddin, M. I., Ismom, A. A. A., Anik, Z. H., & Arefin, M. N. (2022, December). Cardio Vascular Disease prediction using machine learning with improved feature selection. In 2022 12th International Conference on Electrical and Computer Engineering (ICECE) (pp. 336-339). IEEE.
- [2] Alalawi, H. H., & Alsuwat, M. S. (2021). Detection of cardiovascular disease using machine learning classification models. *International Journal of Engineering Research & Technology*, 10(7), 151-7.
- [3] Mahmoud, H. K. M. (2023). Comparison between Machine Learning Algorithms for Cardiovascular Disease Prediction, *Scientific Journal for Financial and Commercial Studies and Research*, Faculty of Commerce, Damietta University, 4(1)1, 909-926.
- [4] Kumar, D. (2020). Cardiovascular disease prediction using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 46-54.
- [5] Sufian, M. A. (2023). AI Models for Early Detection and Mortality Prediction in Cardiovascular Diseases. *Authorea Preprints*.
- [6] Nadakinamani, R. G., Reyana, A., Kautish, S., Vibith, A. S., Gupta, Y., Abdelwahab, S. F., & Mohamed, A. W. (2022). Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. *Computational intelligence and neuroscience*, 2022.
- [7] Kadam, M. A., Patil, S., Pethkar, P., Shikare, R., & Sarnayak, S. (2023). A cardiovascular disease prediction system using machine learning. *Journal of Pharmaceutical Negative Results*, 7216-7225.
- [8] Garg, A., Sharma, B., & Khan, R. (2021). Heart disease prediction using machine learning techniques. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012046). IOP Publishing.
- [9] Chauhan, Y. J. (2020). Cardiovascular disease prediction using classification algorithms of machine learning. *International Journal of Science and Research*, 9(5), 194-200.
- [10] Tiwari, P., Singh, S., & Singh, J. Prediction of Cardiovascular Disease Using Machine Learning and Data Science.
- [11] Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020, March). Analysis and prediction of cardio vascular disease using machine learning classifiers. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 15-21). IEEE.
- [12] Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, 7(3), 263-275.
- [13] Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari, V. (2018, March). Prediction of cardiovascular disease using machine learning algorithms. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* (pp. 1-7). IEEE.
- [14] Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., & Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. *Annals of the Romanian Society for Cell Biology*, 904-912.

[15] Louridi, N., Amar, M., & El Ouahidi, B. (2019, October). Identification of cardiovascular diseases using machine learning. In *2019 7th mediterranean congress of telecommunications (CMT)* (pp. 1-6). IEEE.

ORIGINALITY REPORT

8%

SIMILARITY INDEX

6%

INTERNET SOURCES

3%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University
Student Paper

3%

2

[dspace.daffodilvarsity.edu.bd:8080](https://dspace.daffodilvarsity.edu.bd/8080)
Internet Source

2%