

Heart Health in Bangladesh: A Data-Driven Approach Using Machine Learning

BY

Shatabdi Shil Piu
ID: 242-17-010

This document is submitted as a partial requirement for the Master of
Science degree in Computer Science and Engineering.

Supervised By

Dr. Arif Mahmud
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised By
Mr. Shahadat Hossain
Assitant Professor
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

APPROVAL

The Department of Computer Science and Engineering at Daffodil International University has accepted Shatabdi Shil Piu's project, "Heart Health in Bangladesh: A Data-Driven Approach Using Machine Learning," ID No: 242-17-010, as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering. The style and content of the project have also been approved. The presentation has already been placed.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori
Professor and Head

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



Chairman

Dr. Abdus Sattar
Associate Professor

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



Internal Examiner

Dr. Md Zahid Hasan
Associate Professor

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



Internal Examiner

Mr. Nazibur Rahman
Head of IT Infrastructure
NETWORLD BANGLADESH PLC

External Examiner

DECLARATION

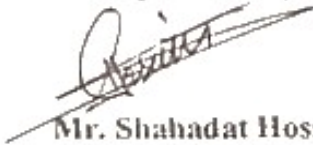
I hereby declare that this research has been done by me under the supervision of **Dr. Arif Mahmud, Associate Professor and Associate Head, Department of CSE, Daffodil International University**. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree

Supervised by:



Dr. Arif Mahmud
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:



Mr. Shahadat Hossain
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Shatabdi Shil Piu
ID: 242-17-010
Department of MIS
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Almighty Allah who give me strength and opportunity to complete my final year project and internship period easier.

I owe a huge debt of gratitude to Dr. Arif Mahmud, Associate Professor and Associate Head in the Department of CSE at Daffodil International University in Dhaka, for his amazing knowledge of and genuine interest in the field of Machine Learning, which has been a huge assistance in finishing this project. His continuous support, great mentorship, ceaseless guidance, wise critiques, invaluable recommendations, and continuous offer to read and refine multiple drafts were essential during the entire process of this work.

Lastly, I want to thank Dr. Sheak Rashed Haider Noori, the head of the CSE department, for his help with this project. I also want to thank all the CSE department teachers and staff at Daffodil International University for their constant support and encouragement.

Finally, I dedicate this to my parents who have always been my backbone, thanks for your sacrifices and patience.

ABSTRACT

Cardiovascular diseases are one of the most extensive forms of deadly disease all over the world; thus, early identification of cardiac issues is vital to ensure fatal outcomes are minimized and treatment is improved. The objective of this study is to assess the effectiveness of various machine learning techniques at predicting heart disease diagnoses from refined clinical and lifestyle data features. We then used six classification models (XGBoost, Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, and Multi-Layer Perceptron Neural Network) to validate our dataset. The dataset was well preprocessed as it included missing value, normalisation, encoding and feature selection techniques with high data integrity at ges. The model's effectiveness was evaluated using accuracy, recall, precision, F1-score, ROC AUC, along with calibration curves and SHAP-based interpretability. XGBoost was the highest performing model with a 96.18% accuracy and it achieved the best one (the best balance between recall and precision), whilst Decision Tree and Random Forest also produced reasonable balanced scores. In conclusion, the results indicate that tree-based fusion methods are a great compromise between accuracy and interpretability, and viable for clinical implementation. These findings demonstrate the potential and challenges of applying AI to cardiac diagnostics while considering the ethical, societal, and sustainability issues central to real-world healthcare implementation.

Keywords:

Cardiovascular Diseases, Heart Disease Prediction, Machine Learning, Clinical Data, Data Preprocessing, XGBoost, Random Forest, Support Vector Machine, Model Interpretability, Artificial Intelligence in Healthcare, Ethical and Societal Considerations.

TABLE OF CONTENTS

OUTLINE	PAGE
Evaluation Board	ii
Declaration	iii
Acknowledgement	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	1-8
1.1 Introduction	1
1.2 Encouragement	7
1.3 Aims of the study	7
1.4 Study Inquiries	8
1.5 Intended Results	8
1.6 Planning and Financial Administration	8
1.7 Report Format	8
CHAPTER 2: BACKGROUND	9-11
2.1 Introduction	9
2.2 Related Works	9
2.3 Research Gap	11
2.4 Obstacles	11
CHAPTER 3: RESEARCH METHODOLOGY	13-26
3.1 Proposed Methodology	13
3.2.1 Overview of the Data	13
3.2.2 Analysis of attributes	14
3.3 Preprocessing Steps and Feature Filtering	16
3.4 Model Identification and Application	18
3.5 Performance Indicators	24
3.6 Technical Resources	26
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	27-35
4.1 Summery of Findings	27
4.2 Performance Benchmarking of Models	27

4.3 Results and Analytical Insights	27
4.4 Review	36
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	38-39
5.1 Effect on Society	38
5.2 Moral Considerations	38
5.3 Sustainability Concerns	39
5.4 Roadmap for Enduring Impact	39
CHAPTER 6: CONCLUSION AND FUTURE WORK	40-41
6.1 Summer of the Study	40
6.2 Final Remarks	41
6.3 Prospects for Upcoming Study	42
REFERENCES	43-45

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1 Coronary Vessel Disorder	1
Figure 1.2 Abnormal Cardiac Rhythms	3
Figure 1.3 Cardiac Muscle Failure	3
Figure 1.4 Cardiac Valve Condition	4
Figure 1.5 Pericardial Disease	5
Figure 1.6 Myocardial disease (Cardiomyopathy)	6
Figure 1.7 Inborn heart malformations	6
Figure 3.1 Working process to do Heart disease prediction	13
Figure 3.2 Example entries from the heart disease dataset	14
Figure 3.3 Numerical Features Correlation Heatmap	17
Figure 3.4 Support Vector Machine model framework	18
Figure 3.5 Decision Tree model framework	19
Figure 3.6 Random Forest model framework	20
Figure 3.7 XGBoost model framework	21
Figure 3.8 Naïve Bayes model framework	22
Figure 3.9 MLP Neural Network model framework	23
Figure 3.10 Architecture and Operational Workflow of the Model	24
Figure 3.11 Mathematical Definitions of Evaluation Metrics and Their Derivation from the Confusion Matrix	25
Figure 4.1 Accuracy comparison chart of models	28
Figure 4.2 Comparative analysis of ROC curves	30
Figure 4.3 Matrix of predicted vs actual values	31
Figure 4.4 Evaluation of results across different models	32
Figure 4.5 Graphical representation of precision and recall	33
Figure 4.6 SHAP-based contribution score	34
Figure 4.7 Model calibration graph	35

SUMMARY OF TABLES

TABLES	PAGE NO
Table 3.1: Attribute Analogy	15
Table 4.1 Comparative Accuracy of Machine Learning Algorithms Ranked by Performance	27
Table 4.2: Displaying outcomes across different models	28

CHAPTER 1

INTRODUCTION

1.1 Introduction

Cardiovascular diseases are the most important adverse events and tallied in approximate 17.9 million deaths annually worldwide. Due to factors like poor diet and lack of exercise [1], cardiovascular disease (CVD) has reached epidemic proportions worldwide. These diseases also for the heart and blood vessels, e.g. worsens cerebral vascular so on [2]. Quick identification and cure of heart disease-connected events may be a big part in lowering death rates and improving quality of life. Machine learning (ML) algorithms are being a popular way to explain predictive medicare since data-driven technologies are working so well [3]. Standard heart disease diagnosis techniques are based on signing test interpretation and physician expertise [4], which indicates that the process can be potentially inconsistent and opinion-based diagnosis can be delayed [5]. In this line, the use of ML models can assist to know the pattern from patient statistics which can therefore provide precise immediate forecasts [6]. This paper explores and compares the performance of different ML approaches on the prefiltered and preprocessed dataset for predicting heart disease.

Coronary Vessel Disorder (CAD)

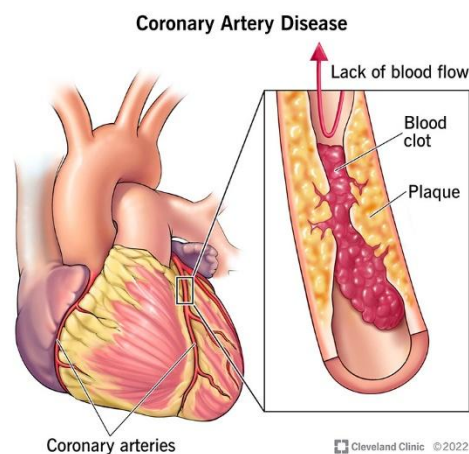


Figure 1.1: Coronary Vessel Disorder [6]

In addition to being the most prevalent cardiac anomaly, coronary artery disease which is also called CAD is also one of the trending reasons of mortality. CAD is a condition that manifests itself when the heart vessels, which are the most important blood flowing vessels that feed your body with the thing named blood, stop functioning properly. This may decrease the blood circulation to anyone's heart function which would eventually forbid the person's heart from taking the most important oxygen one wants. Atherosclerosis, which refers to the hardening of the arteries, is generally an early stage of the disease. In most circumstances, when this is what you have, deposits referred to as plaque build up in your arteries for a long time [6].

Coronary heart disease can result in angina – and in turn and angina can lead to a heart attack. It may also be one of the leading causes for other heart cardiovascular diseases. Things like irregular heartbeats, sometimes even heart failure.

Several factors may increase the chance of having coronary artery disease, including:

- Risk factors / Conditions, age (After menopause, every woman has a significantly increased chance of developing heart problems, while men over the 55 begin to see an increase in their risk.)
- Almost everything can example of no china activities

Family history: Having the past of homosexual or lesbian CAD is one of the main reasons that may also be the reason to CAD. Besides age, sex, and family history, other risk factors include hypertension, smoking, overweight, stress, high LDL cholesterol, and low HDL cholesterol.

Abnormal Cardiac Rhythms



Figure 1.2: Abnormal Cardiac Rhythms [7]

An abnormal heartbeat, irregular, either too slow or too fast is called an arrhythmia. Some kinds are not harmful. Lesser one exposes anyone to a high threat of issues such as cardiac arrhythmia also strokes. These are capable of showing symptoms such as feeling dizzy and fainting. Over a heart problem, arrhythmias have considered as very dangerous, which may lead to dead status, sometimes, before the other major organ function stops, the heart dies, but other times also with other heart problems. If you smoke heavily, drink too much or have obesity, high blood sugar and/or sleep apnea, you are at higher risk [8].

Cardiac Muscle Failure

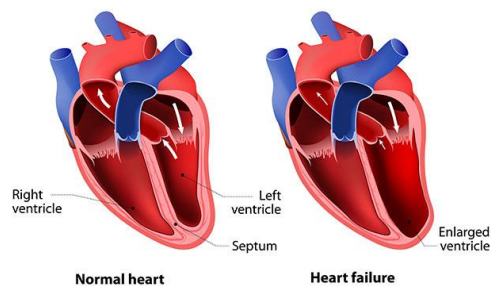


Figure 1.3: Cardiac Muscle Failure [8]

This is especially true if you have heart failure, in which the volume of blood your heart pumps is less than what is required to deliver sufficient oxygenated blood needed by your

body. Often that's because of CAD, which can also occur if anyone has thyroid disease, cardiomyopathy (heart muscle disease), notable blood pressure, or any other stages that impact the cardiovascular area [9].

Cardiac Valve Condition

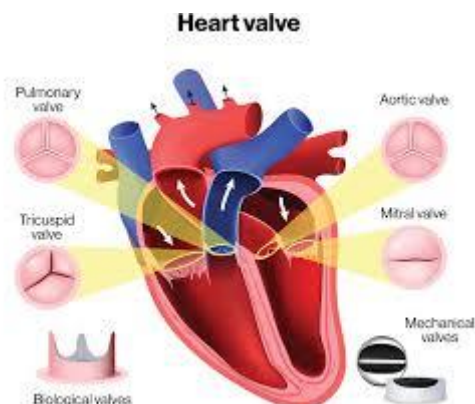


Figure 1.4: Cardiac Valve Condition [9]

Within your heart, there are four valves that may open and shut in such a way that enables blood circulation in the right way. Valves direct blood circulation through the rooms of the cardiovascular area, the lung, and the blood lines. An imperfection will make it more impaired valve function causing restricted opening and closing. Blood circulation can get stuck during that time, otherwise, blood flows out. Causes of heart valve problems include, but are not limited to, infections such as targeted pharmaceutical agents or cancer-directed radiation therapy, rheumatic fever, heart attacks, and heart disease, especially a type. Otherwise known as the CAD, or coronary artery disease [10].

Conditions affecting the heart valves encompass:

- **Infection of the heart valves:** Localized in the heart, endocarditis is usually a bacterial infection that you can get at the time of sickness, have treatment or use hazardous drugs. It shatters normal control of heart rate and can reach other organs through the blood. If you already have valve problems, then it is fairly common. It is usually curable with antibiotics, but is fatal without treatment. If the endocarditis has severely damaged your heart valves, it is possible that you will need surgery to replace one or more of your heart valves.

- **Cardiac damage from rheumatic fever:** And less often, if rheumatic fever, which can also be a complication of scarlet fever and strep throat, damages someone's muscle of heart and its valves, well, this is how you would feel. More people used to be affected by rheumatic heart disease. Doctor could avoid this now however, some antibiotics are used to treat the diseases that actually cause the fever. Even when anyone does get it, the signs only show up after a long time [11].

Disorders of the pericardium

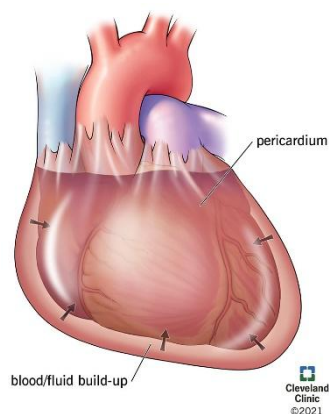


Figure 1.5: Pericardial Disorder [11]

Curing Pericardial Disease means restoring health to the person suffering with any of the health conditions that impinge upon the pericardium, the caul around your heart. The most common one is pericarditis which refer to an inflammation of the pericardium. The causes of pericarditis include virus infection and any injury to anyones pericardium, and internal diseases, like rheumatoid arthritis and lupus. Open heart surgery has a frequent complication, the so-called pericarditis [12].

Myocardial disease (Cardiomyopathy)

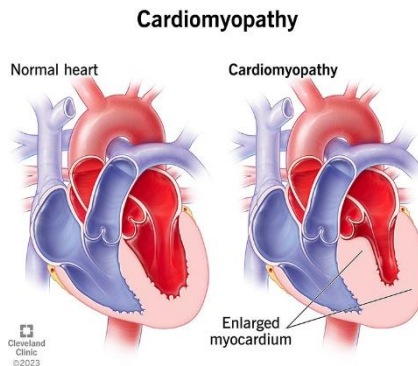


Figure 1.6: Myocardial disease (Cardiomyopathy) [11]

There is a disease called Cardiomyopathy, which is a disease of the internal heart myocardium, the muscular tissue of the heart. In response, the material is less stiff, thinned out, or pulled taut. Anyone's heart could become too feeble to pump blood. Potential causes of the disease comprise hereditary heart ailments, reactions to beneficial medications or poisons (including ethanol), and infectious diseases caused by viruses. Cardiomyopathy (disease of the heart muscle) Imaging may show large heart chambers and can sometimes result from treatment for cancer. Most often experts, could not find the actual probable cause [12].

Inborn heart malformations

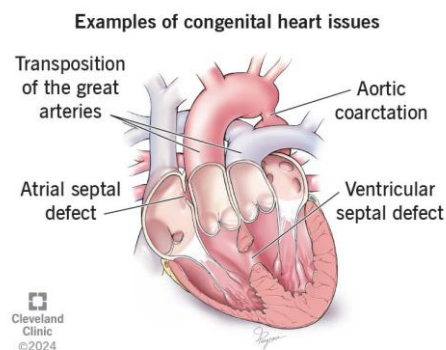


Figure 1.7: Inborn heart malformations [12]

If something goes wrong as the heart is forming before birth, the end result is congenital heart disease. This heart defect may cause problems in newborns, while in other patients there are no symptoms until adulthood. The most frequent kind of congenital cardiac issue is a septal defect. The gap between the upper and lower portions of your heart is like holes in a wall. Well, you might have a hole plug-in; Another difficulty is pulmonary stenosis. The stiff valve limits blood circulation to your lungs. It, too, can be opened and, in some instances, managed via therapeutic procedure or surgery, as well. In a small and quite uncommon number of babies, an extremely little blood vessel which is called by the ductus arteriosus. It stays open at birth which it should close midway through humanity. When this occurs, some blood goes around to the pulmonary artery, that makes the heart function stronger than before. The doctor will treat this disease by surgical treatment, by procedure, or will also use medicine [12].

1.2 Encouragement

In such an analysis, one would aim to build the intelligent, data based machine learning system which may show the likelihood for unnecessary cardiovascular problems turns out to be [1]. It is possible to mine patient health data [8] which is available in huge volumes to gain appropriate pattern/discovery about data that could help in medical decision making.

Furthermore, the study work has a planned aim in medical digital health and proactive medical care to recognize optimal ML models as predictors. AI-based supporting tools can tackle this bottleneck by fast detection of strong threats persons, which will be effective in reducing the medical performance force which also enabling the early interventions [23].

1.3 Objectives

- As a benchmark for various machine learning algorithms for predicting heart disease.
- To find the most important input characteristics for heat sickness in the dataset.

- Design and develop an understandable ML baseline that could be integrated in the system of medical decision advising systems.

1.4 Study Inquiries

- RQ1: For the purpose of predicting the trend of developing heart disease, which method of machine learning produced the highest degree of accuracy attained?
- RQ2: What is the real performance of different models in terms of F1-score, recall, precision and ROC AUC?
- RQ3: What clinical and lifestyle feature most contribute to the prognosis of heart disease?

1.5 Intended Results

- Systematic comparison of the performance of ML models in light of standard evaluation metrics.
- Top performers: Random Forest and XGBoost for implementation.
- A heart disease prediction model with a clear preprocessing, training and validation methodology.

1.6 Planning and Financial Administration

This study is completely independent, and is not funded by anybody or any state. The computation corresponding to this study was made by the same author.

1.7 Report Format

Chapter 1 The thinking, purpose and contribution of the study In Chapter 1 motor vehicle accident dealing with the literature and literature review of Chapter 2 holes in cardiovascular disease prediction. Whereas Chapter 3 presents an outline of the work process, where it includes the dataset description, the chosen algorithm for analysis, and preprocessing steps. The fourth chapter presents the results of the experiments and performance analysis. Study impact — social, ethical and environmental context — Chapter 5 is the Final, and Chapter 6 closes the curtain on the study and suggests further avenues for enhancement.

CHAPTER 2

BACKGROUND

2.1 Introduction

Heart problems [8] include a number of different types physiological disorders of the heart. These disorders consist internal cardiovascular disease, CAD, and arrhythmias. These infections are determined by genetic factors and livelihood factors (e.g., diabetes, smoking, hypertension and unhealthy diet) [24]. In a medical diagnostic scenario, ECG, blood sugar, blood pressure, BMI and sign-wise based diagnosis can be done [2]. In this context, predictive modeling aims to use these features to predict the probability of a patient being a high-risk patient automatically.

Found a solution, the most interesting product Machine learning, which is a new technology under shift from science formed relative exercise over some time or ML, that is an emerging sub-union of another most generally common domain which is artificial intelligence (AI), in this an AI keeps a patterns based over data, to provides with guesses or even judgements without being directly programmed [2]. Specifically, directed ML ideas are useful for classifying problems [24] like disease prediction.

2.2 Related Works

The use of machine learning models to assess cardiovascular disease by using medical and livelihood characteristics has been the subject of a great number of research and various initiatives. When using the Cleveland heart disease dataset, Detrano et al. [8] shoowed an accuracy of 83.5% via logistic regression for comparison. Gudadhe et al carried out their investigation in July 2023, which was followed up with additional research. utilized Support Vector Machines (SVM) and achieved a marginal increase, getting 85.03% accuracy [30]. Fusion models have been deemed beneficial as these models provide superior performance on complex datasets. For example, in [7], Pattekari and Parveen use

Decision Trees and Naive Bayes that obtain 81% and 84% accuracy respectively but more recent benchmarks have reported gradientboosting techniques like XGBoost achieving an accuracy of above 90% [9]. Because they are less prone to overworking, Random Forests have also had excellent predictive effectiveness in terms of cardiovascular prediction, with accuracy between 88% and 92% reported [2] depending on attribute selection methods and optimizing learning parameters.

Recent works shows that, Ayyasamy et al. Gradient Boosting reaches an accuracy of 94.6%, better than any fusion or baseline classifier [1]. Like this, CatBoost has been pointed out because it works natively with categorical variables and showed competitive results with 93.1% accuracy in multicenter medical data [21]. The artificial neural network (ANN), too, particularly the Multi-Layer Percepcion ((MLP) has been investigated tom huge extend. MLP based approach of Kahramanli and Allahverdi result in 86.8% accuracy but took longer to train and had high computational power requirements [23]. Neural networks hold great promise but their lack of interpretability (~50% of papers) is limiting for specific clinical applications. Aslam et al. have also evaluated the performance of deep learning models such as CNNs-based models like convolutional neural networks [8] also have obtained accuracy as high as 91.5% in the case of enhanced image-based ECG features for the binary classification of cardiac arrhythmia detection tasks. While these models boast impressive predictive capabilities in image-based datasets, they often struggle to provide solutions when it comes to non-image-based structured data, like tabular therapeutic data. Dey et al. found through a meta-analysis [3] that models trained on larger, more heterogeneous datasets universally do better than those who were taught on small groups of people from different backgrounds. Moreover, in their meta-analysis, the authors showed that although the deep learning algorithms had an overall better accuracy compared to the other traditional machine learning algorithms, the Gradient Boosting based methods got the better performance versus training time trade-off on the healthcare datasets.

When it comes to explain ability, usage of SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) are becoming common practice in explaining black-box models in healthcare [36] automating feature selection process [37], has been reported. For example, Lundberg et al. suggested enabling SHAP to assist healthcare professionals believe in ML models by showing the effect of the variable in prediction made on the patients [7]. While these studies provide reassurance regarding the applicability of ML for heart disease prediction, they did not utilize adequate cross-validation, had small datasets, and/or did not compare diverse model families. We aim to fill these gaps with a benchmarking that assesses ten classifiers using four performance measures (Recall, Precision, F1-score, ROC AUC).

2.3 Research Gap

Although it has been shown within prior work that individual ML models are important, limited work has performed or have been able to evaluate a comprehensive set of ML algorithms using filtered, practical attributes sets in a manner that parameterizes a equilibrated comparison among multiple algorithms [23]. Track 3: Finally, less research has been performed on interpretability and generalizability in lower-resource settings, where understanding the model is key for deployment and evaluation [8].

Contribution: To fill this gap, we conduct a relative analysis of 6 ML algorithms applied to a carefully selected dataset through a small set of filtered features. The primary goal of this study is to find which models performing good and most feasible model for medical decision support systems.

2.4 Obstacles

- Lacking Attributes: Medical data is frequently marred by missing data points, especially in values pertaining to blood sugar, BMI etc [3].
- We will go beyond the typical metrics —F1-score or AUC— that get reported in works on imbalanced datasets, for instance, the heart disease datasets could be very much skewed, which require rebalancing techniques to be applied.

- Redundant Features: Redundant or Irrelevant features are harmful to model performance; So, Feature selection are very important.
- Interpretability: The model needs to be explainable/interpretive in a clinical/scientific setting; otherwise, the black-box model will not be adopted as there will be no clear instructions on how the model arrived at its conclusion.
- Generalizability: Performance on previously unknown patient data must be maintained, which relies on strong cross-validation and a wide range of training data [7].

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Proposed Methodology

This study is based on a methodical, data-driven approach in a machine learning pipeline to create a prediction algorithm that can diagnose heart disease. The approach has a number of steps that are all connected: getting the dataset, cleaning and evaluating its performance, choosing features, interpreting the results, training the model, preprocessing the data. Five machine learning models, starting with a basic baseline

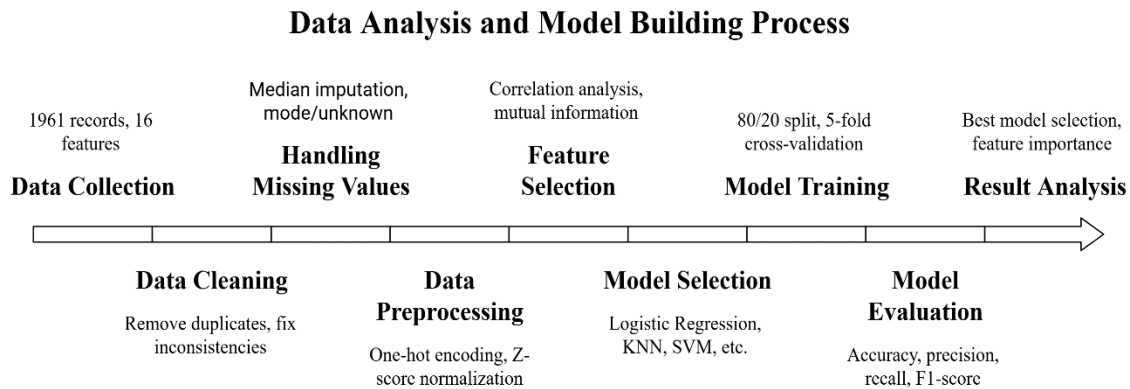


Figure 3.1: Working process to do Heart disease prediction

Implementations of fusion neural network classifiers and algorithms. To determine cross-validation techniques, generalizability were being used and a wide range of classification metrics were utilized to evaluate performance. In this chapter each step in the methodology is explained in precise detail.

3.2.1 Overview of the Data

Data set which has been used for the study has 16 features and 1961 records and which are eye-ball fitted for heart disease prediction. They comprise of a mix of medical data (e.g. Sugar level, BMI, Pulse rate, Age, Blood pressure) and behavioral or lifestyle characteristics (e.g. Family history of heart disease, Sport activity, Diabetic history, Type

of diet, Current smoking status). The target variable is a binary classifier (to the model) of whether a person consists a heart disease or not.

total_income	union_name	user_id	profile_name	father_name	mother_name	birthday	age	...	RESULT_STAT_SPO2	MUAC	RESULT_STAT_MUAC	chest_pain_type
Middle class	CHERAGPUR	1111051.0	আফরোজাআক্তার	আশরাফুল	তনজিলা খাতুন	09-03-04 18:00	16.0	...	Normal	NaN	NaN	No chest pain
Lower class	ANDULBARIA	1392877.0	সালমাবেগম	0	0	11-07-89 18:00	31.0	...	NaN	NaN	NaN	No chest pain
Lower class	KOLA	989244.0	হাজেরাবেগম	0	0	02-01-57 18:00	63.0	...	NaN	NaN	NaN	During physical activity
Lower class	BARATARA	103371.0	মোছঃ নুরমাছার বেগম	0	0	25-02-82 14:12	38.0	...	NaN	NaN	NaN	After emotional stress or meals
Middle class	KOLA	1049014.0	মোঃ মতিউররহমান	0	0	31-05-68 18:00	52.0	...	NaN	NaN	NaN	No chest pain

Figure 3.2: Example entries from the heart disease dataset

Methods: Data has been collected on patients from a well defined medical health database. Every record represents a health profile of a single patient which mimicks a clinical setting of that era. It includes categorical, ordinal types, and numerical with different levels of completeness. We performed some data handling by removing duplicates, standardising categorical values and finally, examined null values to meet the requirements for trustworthiness of results

3.2.2 Analysis of Attributes

Table 3.1: Attribute Analogy, Mapping of key attributes across concepts for comparative understanding

#	Name of the Features	Present Values	Value Origin	Analogy
0	age	1960	Int64	Patient age, an important characteristics and risk reason.
1	gender	1961	Object	Biological Gender of the patient (example: Male, Female).

2	chest_pain_type	1961	Object	Quality of chest pain, which typically implies cardiac causes.
3	SYSTOLIC	1905	Float64	Systolic blood pressure, the best method for measuring the risk.
4	SUGAR	779	Float64	Blood glucose level — valuable for cardiac risk related to diabetes.
5	PULSE_RATE	1902	Float64	Heart rate, measured in beats per minute (BPM), which reflects heart function.
6	shortness_of_breath	1651	Object	Symptoms functioning from possible cardiac or pulmonary distress.
7	RESULT_STAT_BP	1904	Object	Hypertension level (Normal, Elevated), a status on blood pressure.
8	BMI	815	Float64	Body Mass Index. This is one of the most common known risk factors for obesity, and it is a measure of the disease.
9	has_cardiovascular_disease	1940	Float64	Whether the patient has cardiovascular disease (CVD) already (the indicator function is binary).
10	family_history_hd	1651	Object	Hereditary heart disease: Family encountering of cardiovascular disorder, key genetic risk factor.
11	smoking_status	1650	Object	One of the most important lifestyle risk factors, smoking of patient.
12	diabetic	1961	Object	Diabetes diagnosed or not, a strong risk factor for heart disease.
13	profile_hypertensive	1961	Object	Blood pressure conditions, classifying hypertension profile.
14	physical_activity	1650	Object	Physical activity which affects heart disease.
15	diet_type	1650	Object	Eating patterns, influencing overall risk of heart disease.

- Baseline demographic characteristics crucial for risk stratification are represented as Age and Gender.
- Chest Pain Type describes the kind of cardio event or another source of chest symptom.
- By having SYSTOLIC, Systolic Blood Pressure (RESULT_STAT_BP) which inform us if a patient is under a hypertensive status, which also mostly correlate with cardiovascular risk.
- SUGAR and DM emphasize again those metabolic states that heighten risk for heart disease.
- Pulse Rate provides information on Autonomic Function and Cardiac Rhythm.
- Dyspney is a medical manifestation commonly seen with either cardiac or pulmonary dysfunction.
- BMI refers Body mass index. The goal of the study is to have an anthropometric measure of cardiovascular risk in the general population, which is related to obesity, obesity, metabolic, severity, based on BMI.
- By having prior diagnosis of Cardiovascular Disease, an important predictor of recurrent events.
- The second, Family History of Heart Disease, reflects inherited susceptibility.
- Diet Type and Smoking Status are changed lifestyle reasons associated with heart health.
- Physical Activity and Profile Hypertensive are two patient characteristics that influence fitness (Physical Activity) and blood pressure (BP) control (Profile Hypertensive), both important determinants of cardiovascular events.

3.3 Preprocessing Steps and Feature Filtering

Preprocessing is important because it is a multi-step procedure that is necessary to prepare the data to train a robust model. Key steps involved are explained in the following:

- **Dealing with Null Values:** Large number of SUGAR and BMI were Null. Continuous variables were not normally distributed and were consequently handled with median imputation (as by Wong et al.) To keep things clear, missing values in categorical variables were interchanged with the mode value or given the label "Unknown." This method reduced bias caused by imputation while keeping the overall size of the dataset.
- **Transforming Categorical Variables:** smoking_status, diet_type, chest_pain_type, and gender are some of the variables that needed to be changed to

a numerical form so that it could be used for ML algorithms and such we used one hot encoded. Outside of this, we used label encoding as a last resort to avoid any accidental ordinal assumption being made.

- **Data Normalization:** Z-score Normalization has been used to upscale continuous variables like SYSTOLIC, PULSE_RATE, and BMI. In the case of gradient-based optimization such as SVM, this step leveled the feature influence, making the model converge well.
- **Attribute Filtering and Choice:** A combination of mutual information gain, tree-based feature importance and Pearson correlation, were utilized to remove collinear, redundant and non-informative features. The rest 16 features were statistically significant or domain related.

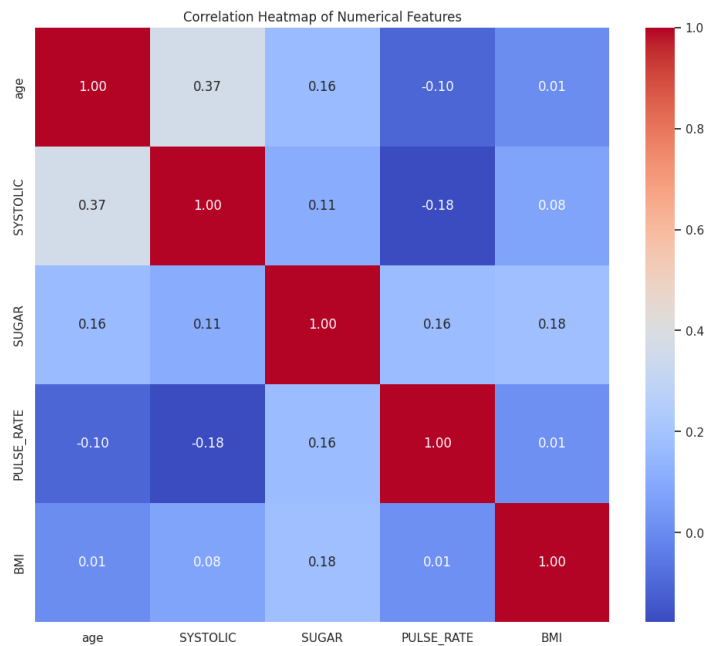


Figure 3.3: Numerical Features Correlation Heatmap

Such preprocessing pipeline increased the stability of the model, lower the risks for overfitting and also made sure that the features which were being used are understandable and taking actionable.

3.4 Model Identification and Application

We intentionally selected a vast range for classical and state-of-the-art machine learning models for both benchmarking performance and transferability. For example:

Support Vector Machine (SVM): SVM is a directed algorithm that finds a dividing surface that best divides the classes based on the features. The aim is to enhance the distance among the support vectors (the values close to the separating surface) [31].

Principal Elements:

- Nice to have when anyone has large amount of data and literally very high attributively (such as up to >10, but this can vary).
- The kernel which is used for mapping the data to larger matter, when applied for non-proportional models.
- Well, resistant to overfitting, with proper regularization parameter (C) being used.

Application Scenarios:

- Classification of text and images, bioinformatics (cancer detection, etc.).

Constraints:

- Will not scale for large datasets.
- Difficult to choose the perfect hyperparameter and kernel.

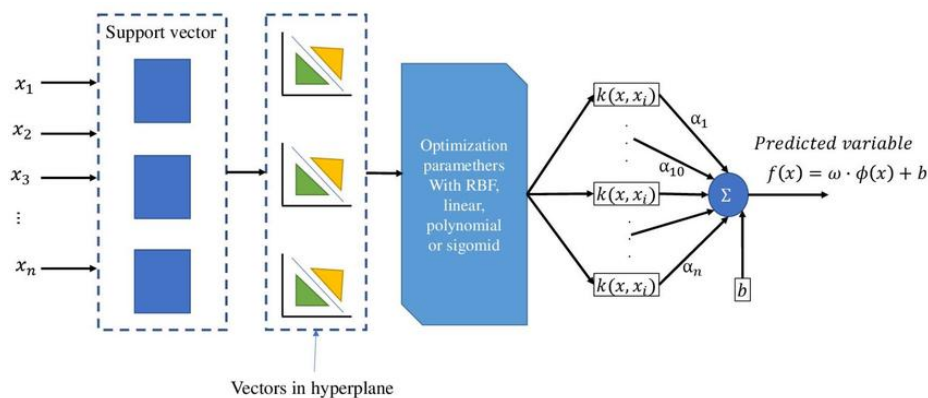


Figure 3.4: Support Vector Machine model framework [31]

Decision Tree: It segments the data set into fragments by making decisions based on class attributes so that the shape of the dataset would look like a tree. The splits at each node are done with some metric (for example Gini Impurity or Information Gain) [32].

Principal Elements:

- Easy to interpret and transparent
- You can easily visualize decision.
- Uses np in likely linear transformation for relations between dependent and independent variables.
- Handle numeric as well as categorical data.

Application Scenarios:

- Credit scoring, risk assessment, medical diagnosis.

Constraints:

- Very high variance (overfitting).
- A slight change to the data results into very different treed thereby having high variance.

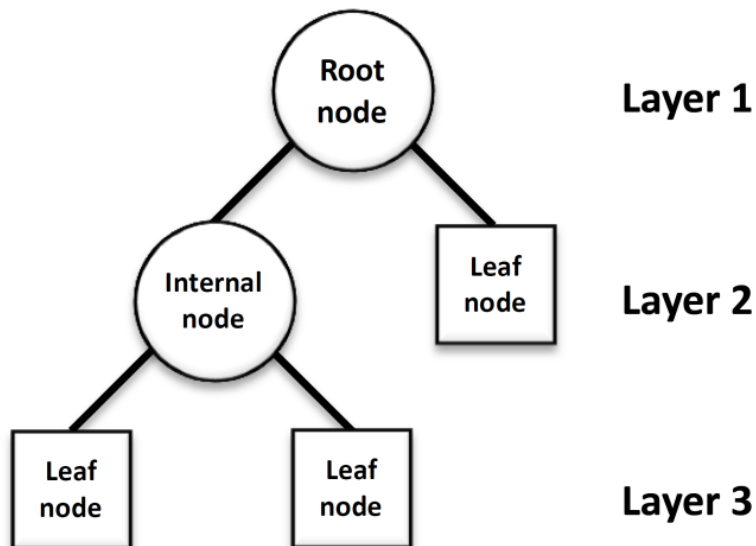


Figure 3.5: Decision Tree model framework [32]

Random Forest: A group of decision trees that were constructed using bagging (Aggregation of Bootstrap). For classification tasks, the overall prediction is based on the majority vote, while for regression tasks, it is based on the average of all the trees trained on a random sample of the data and features [33].

Principal Elements:

- Combines predictions by averaging, minimizes overfitting.
- Higher accuracy chance and resistance to nearly no noise.
- Provides estimates of attribute significance.

Application Scenarios:

- Predicting stock market, recommendation systems, fraud detection.

Constraints:

- More difficult to interpret as compared to single trees.
- High computational costs for huge trees and also for data sets.

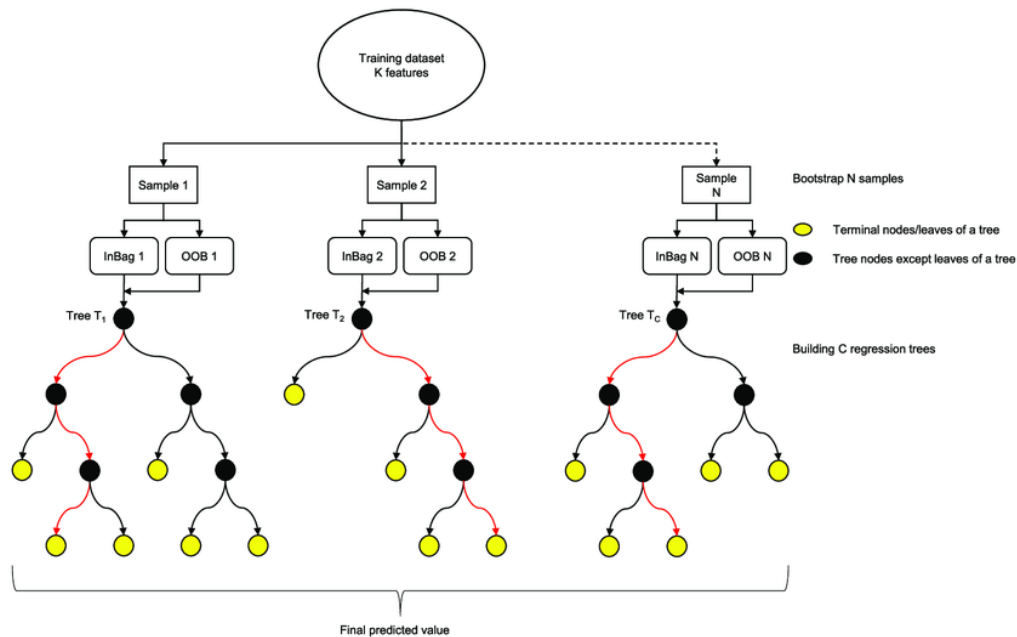


Figure 3.6: Random Forest model framework [33]

XGBoost: XGBoost, which is an implementation of gradient boosting to grow trees gradually, where every tree aims to right the errors of its predecessor. It uses second rank derivatives (Hessian) to optimize, which is why it is X times faster and has approximate accuracy rate of X^3 [34]

Principal Elements:

- Extremely well regularised to prevent overfitting (terms L1 and L2).
- All are parallel tree construction and it is very fast.
- Native support for working with null values and data with many zero or missing values.

Application Scenarios:

- Structure/Tabular Data, Kaggle Competitions, Click-Through Rate Prediction.

Constraints:

- Requires to be careful by tuning a lots of hyper parameters.
- Not as interpretable as simpler models.
- Very fast and accurate highly regularized gradient boosting.

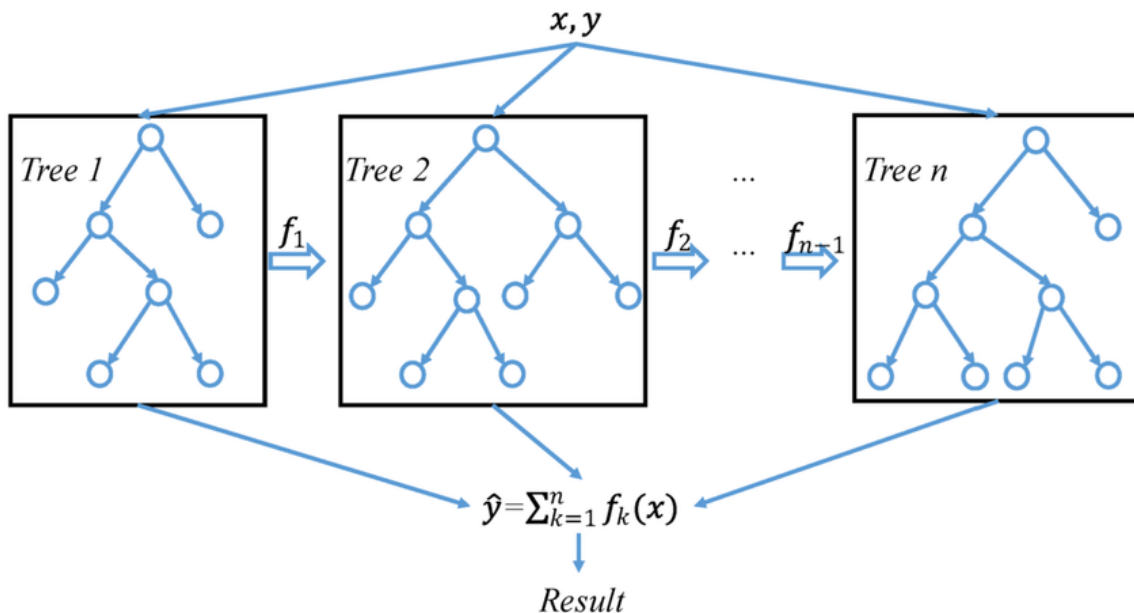


Figure 3.7: XGBoost model framework [34]

Naive Bayes: A prediction based classification model with the narrow consumption of validity to minimize the complexity of the full group prediction estimation [36].

$$P(R|T) = \frac{P(T|R).P(R)}{P(T)} \dots\dots\dots(i)$$

Principal Elements:

- Training and prediction are ultra-fast and scalable.
- Performs well with Text data (spam, sentiment etc).
- Perform Well in Data with Many Variables (Say Vectors of TF-IDF).

Application Scenarios:

- Classify the documents and make the predictions in real time.

Constraints:

- Features independence assumption (not true in reality).
- Thousands of features is really the case with correlated features are not work fine.

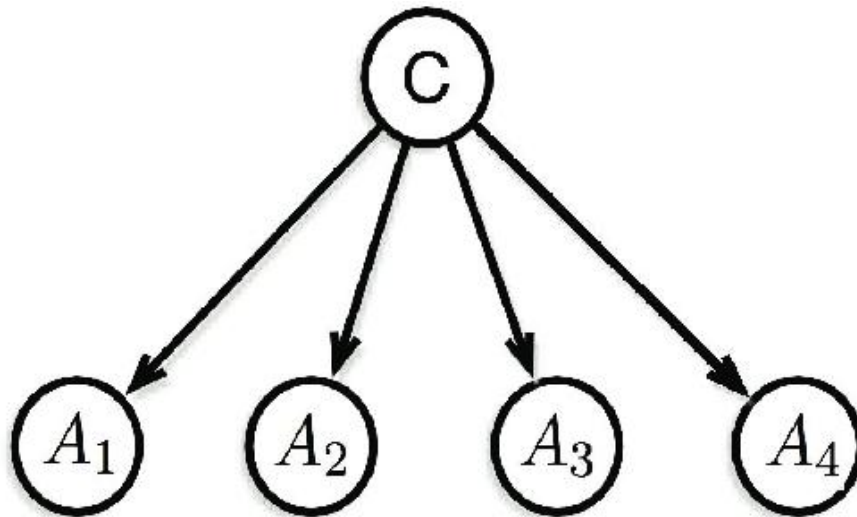


Figure 3.8: Naïve Bayes model framework [36]

MLP Neural Network: This is a kind of closer to neural network that consist an input portion, There could be one or sometimes more unseen (intermediate) layers, and an result layer. It uses additive activation functions for example tanh, ReLU or sigmoid with backpropagation to model's work consistency [35].

Principal Elements:

- Can learn any complex, nonlinear function.
- Used in universal approximator; Approximates any function.
- Two main reasons: it has good scaling with the length of the data and also the compute which has been used.

Application Scenarios:

- Structured task, speech recognition, optical character recognition are involved.

Constraints:

- Feature scaling sensitive.
- Requires higher computational and data requirements.
- Makes debugging/troubleshooting harder than SVMs or decision trees.

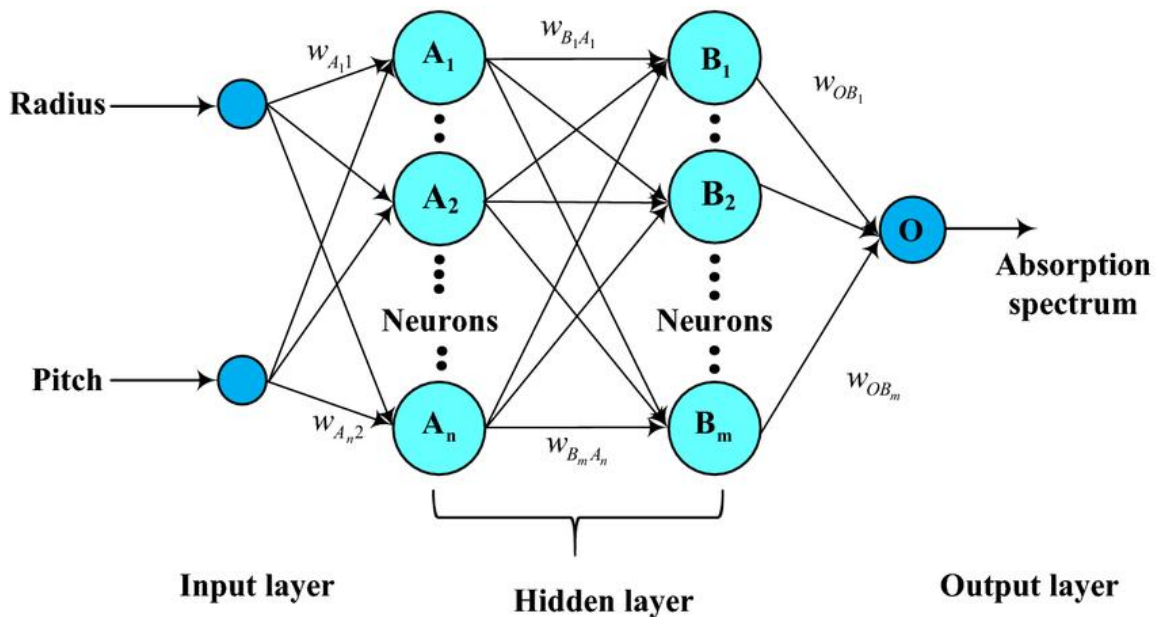


Figure 3.9: MLP Neural Network model framework [35]

Models were created using Python libraries like scikit-learn, xgboost. We trained the models using a 70:30 training:test split to make sure they didn't overfit and that they were consistent. Five-fold stratified cross-validation was used to achieve this.

Size of Train Data: 1373

Size of Test Data: 588

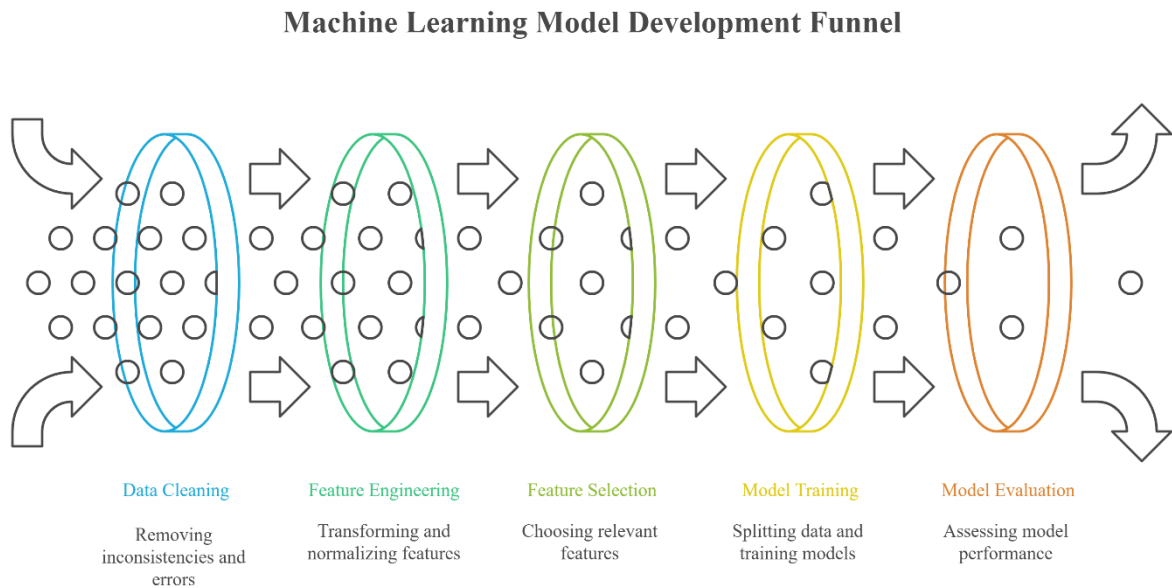


Figure 3.10: Architecture and Operational Workflow of the Model

Different grid search or default optimization algorithms were used to fine-tune the hyperparameters, depending on the complexity of the model and the amount of time spent training.

3.5 Performance Indicators

We measured performance of the classifiers using the following metrics, to gain a broader view on the classifiers performance:

- Accuracy: Basic way to calculate the overall exactness of the model → misleading for non-even datasets.

- Precision: Displays both the rate of accurate positives and the exact digit of positive predictions.
- Recall (Sensitivity): Shows the power of the model to identify all positive samples.
- F1-Score: Brings together recall and precision into a single score for the purposes of execution.
- AUC & ROC: It analyzes how good the algorithm can distinguish among classes for all possible cut-offs.

We calculated these attributes from the confusion matrix of every model test forecasts. Due to the expected imbalance classes, ROC AUC and F1-score were the metrics prioritised.

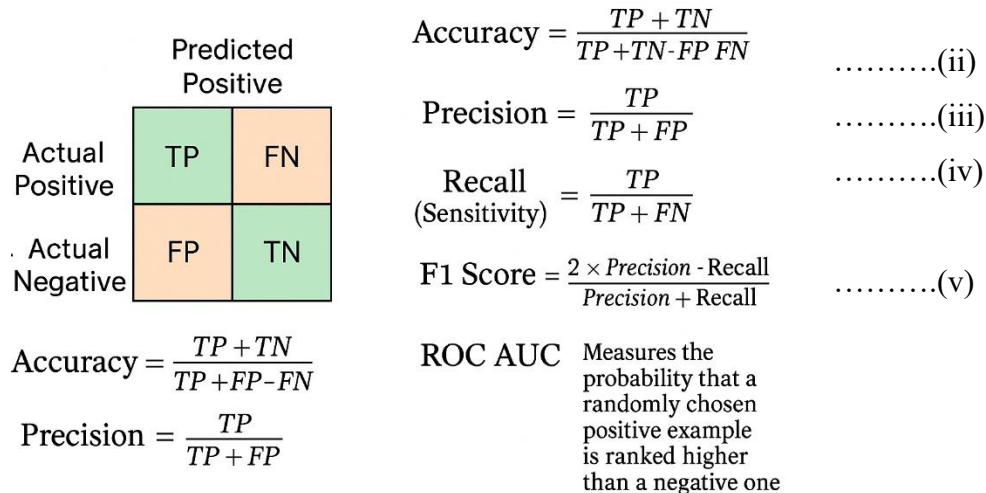


Figure 3.11: Mathematical Definitions of Evaluation Metrics and Their Derivation from the Confusion Matrix

This validation method ensures that the predictive model that will be chosen is sufficiently accurate and that its clinical risk spaces can tolerate no false negatives.

3.6 Technical Resources

The tools used for implementation and analysis were:

- Language: Python 3.11
- IDEs: VS Code and Jupyter Notebook
- Frameworks and Libraries: numpy, seaborn, XGboost, pandas, matplotlib, scikit-learn.
- Git: Version Control system to track experiments for reproducibility.
- Extra: AMD5 processor, 8 Gigabytes of RAM, and no GPU is needed.

With this arsenal, it hands ability to search the data, build models, visualise that, and take not of the efficiency through the complete study goal.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Summary of Findings

This phase includes an analysis of the outcomes achieved from training six different machine learning models on the data used to predict heart disease, as well as an evaluation of the accuracy of these models via rigorous testing. All the performances of the models were analysed extensively with five standard metrical attributes which consists Accuracy, Recall, Precision, ROC AUC and F1-Score. The selected metrics were used to better capture the model behaviour to its intricacies especially in healthcare problems where accuracy of predictions and balancing of errors matter the most. To be clear, we did not pursue this analysis to find the best model, but rather as an practising in considering generalizability, acceptibility, and medical feasibility through the cross-validation method and featuring contribution validation method.

4.2 Performance Benchmarking of Models

Table 4.1: Comparative Accuracy of Machine Learning Algorithms Ranked by Performance

Order	Algorithms	Accuracy
1	XGBoost	0.9696
2	Decision Tree	0.9594
3	Random Forest	0.9136
4	SVM	0.8983
5	MLP Neural Network	0.8907
6	Naive Bayes	0.8831

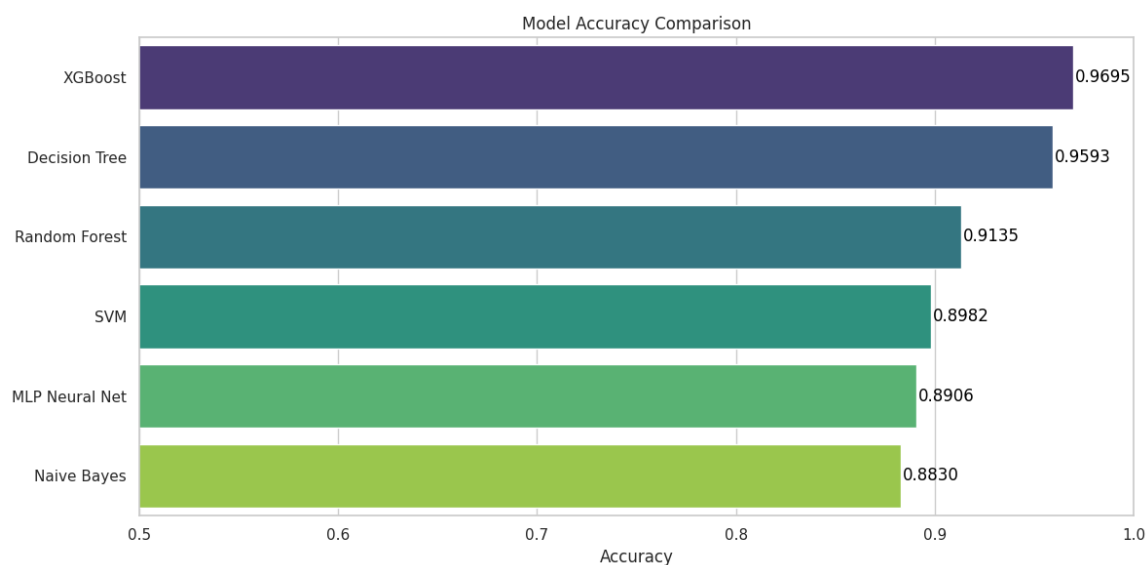


Figure 4.1: Accuracy comparison chart of models

4.3 Results and Analytical Insights

Test performance of all 6 models on Accuracy, Precision, Recall, F1 Score (Macro) and ROC AUC Introducing the 6 models of our 6 models These findings provided a clearer comparison of the models' performances with respect to the detection of heart disease from different perspectives of evaluation.

Table 4.2: Displaying outcomes across different models

Model	Accuracy	F1 Score	AUC ROC	Recall	Precision
XGBoost	0.9696	0.9505	0.9651	0.97	0.94
Decision Tree	0.9594	0.9353	0.9861	0.96	0.91
Random Forest	0.9136	0.8434	0.9516	0.82	0.88
SVM	0.8983	0.8158	0.7564	0.80	0.85
MLP Neural Net	0.8907	0.8032	0.8434	0.79	0.83
Naive Bayes	0.8831	0.7932	0.8556	0.78	0.81

The cardiac disease data collection experiment exhibits the drastic differences in classification accuracy, macro-averaged metrics, and class-wise sensitivity of the different machine learning approaches. XGBoost performs the best with an accuracy of 0.969 and macro F1 score of 0.950. It had high class-wise precision and recall meaning that there was very low generalization and bias and hence it would be suitable for those sensitive medical classification tasks.

The Decision tree model also played crucially good with achieving an accuracy of 0.9594 and F1 score of 0.9353 This also gives it the advantage of high interpretability since it results in very high recall across both classes, which is ideal if we need our decisions to be traceable.

Random Forest achieved a reasonable accuracy of 0.9136 and ROC AUC of 0.9516, though the recall of class 0 (0.65) was markedly inferior than class 1, potentially implying clinical threats by falsely classifying affected individuals as free of the disease. Even though it was an fusion, it appeared to be significantly biased towards the majority class.

On the other hand, MLP and SVM returned fair accuracies at 0.8983 and 0.8907 respectively as well as macro F1 scores below 0.83. While both the models gave a great score for class 1 (recall>0.94) they massively underperformed for class 0 (recall 0.59–0.61). These ratios are inconsistent, and this sensitivity to class imbalance may make them less trustworthy for early detection or high consequences diagnostics.

Despite the fact that it also provides the least algorithmic complexity, the presumption of free between features that it makes in each class is a clear limitation, and for this reason, Naive Bayes was not as good as the rest of algorithms in terms of F1 (0.7932) and ROC AUC (0.8556). Again the recall for class 0 was abysmal — possibly the chief shortcoming of this system for critical decision making.

In summary, XGBoost and also the Decision Tree are the fairest stable models among all models considered for this classification exercise with XGBoost providing the greatest accuracy and Decision Tree offering a clear and simple to tell model. Random Forest is a strong fusion process but by having a class 0 recall, it needs to be cautious when used. The poor performance of that Naive Bayes, MLP and SVM, also suggests that it is important not only to use a model for evaluation by the accuracy but also to measure it by recall of classes that are minority such as medical diagnostics in which the consequence of not finding a positive can be remarkable. With these outcomes highlight the challenging nature of tree-based fusion methods and appear to establish both XGBoost and Decision Tree as stable and understandable candidates for an order of force and medical prediction support environment instruments in structured tabular medical data.

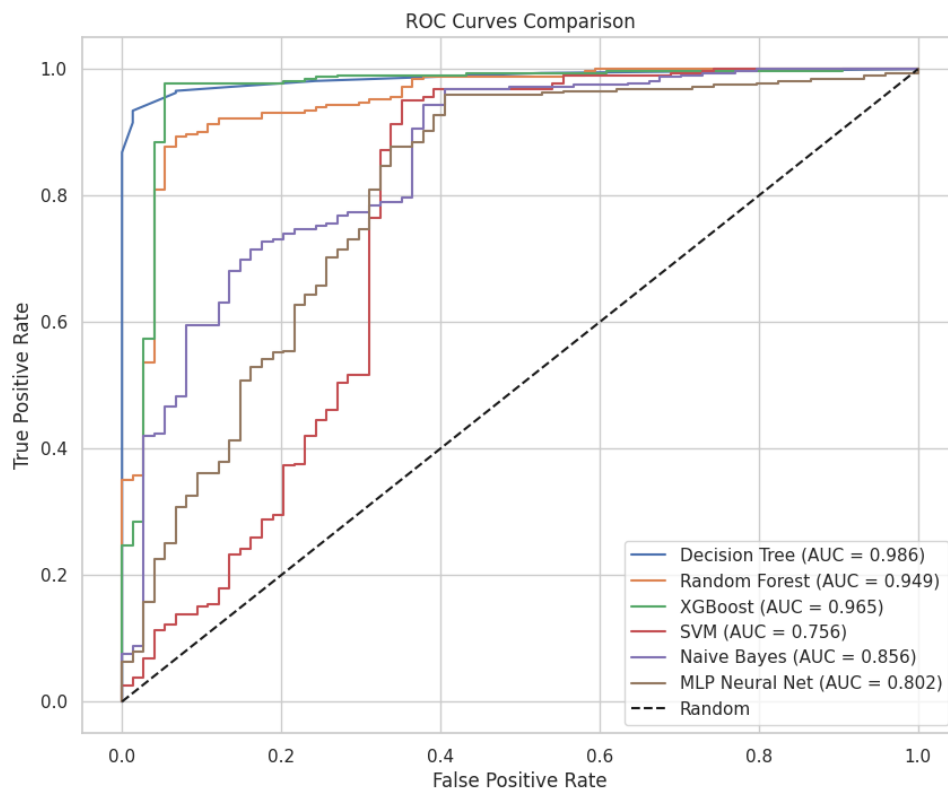


Figure 4.2 Comparative analysis of ROC curves

The graph talks about the relation between the false positive rate (1 - specificity) and the true positive rate (sensitivity) in given series of potential cutpoints. Metrics like ROC curve for models like Random Forest and XGBoost had ROC curves so close to the top left corner and were doing so good at all cut-off. Similarly, Decision Tree and XGBoost were highly discriminative models as showed by the peakiest AUC (Area Under Curve) of 0.9651 and 0.9861 respectively.

In contrast, classifiers (like Naive Bayes and SVM) showed ROC curves that were bowed towards the diagonal baseline, suggesting poorer separation of classes. The worse discrimination — particularly at the precise cut-offs where we hypothesize there can be variance in false positives or negatives — therefore, this indicates poor discrimination. In clinically realistic scenarios, where the costs of misclassification vary, ROC curve provides a key utility for choosing the optimal threshold with regard to (e.g. sensitivity outweighing specificity versus vice versa) operation constraints.

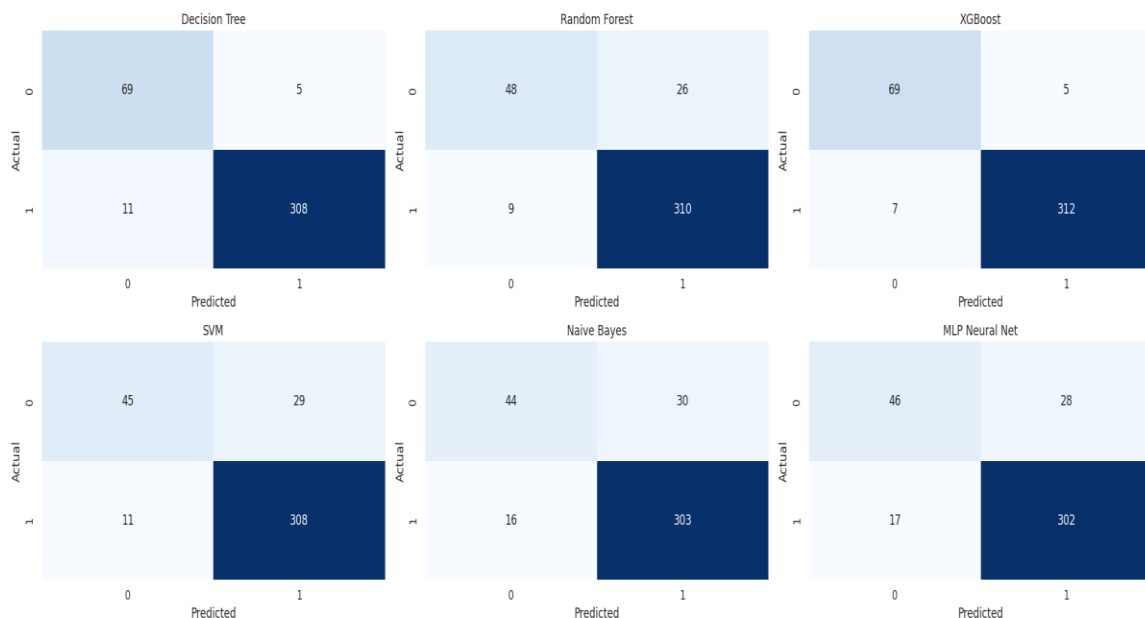


Figure 4.3. Matrix of predicted vs actual values

Clarifying The Model Improvement Using Confusion Matrix Confusion matrix will provide finer details about the classification result. Confusion matrix gives the total count

of True Positives, True Negatives, False Positives and False Negatives respectively. In case of Decision Tree and XGBoost, confusion matrices, they were able to get extremely high values for true positive and low values for false negatives on the class 1 (presence of cardiovascular disorder).

These models also had high class 0 recall important as it reflects that the performance of the model was balanced and bias towards the dominant class was limited. On the other hand, Random Forest and MLP went through a relatively larger amount of false predictions for class 0, which in clinical exams is much more detrimental, as not detecting a non-disease case as such can cause unwanted pain and overtreatment.

This matrix study is crucial, in setting up with either underdiagnosis or overdiagnosis leading to considerable clinical consequences. Both underdiagnosis and overdiagnosis are not favourable for the patient as well as for healthcare resource planning.

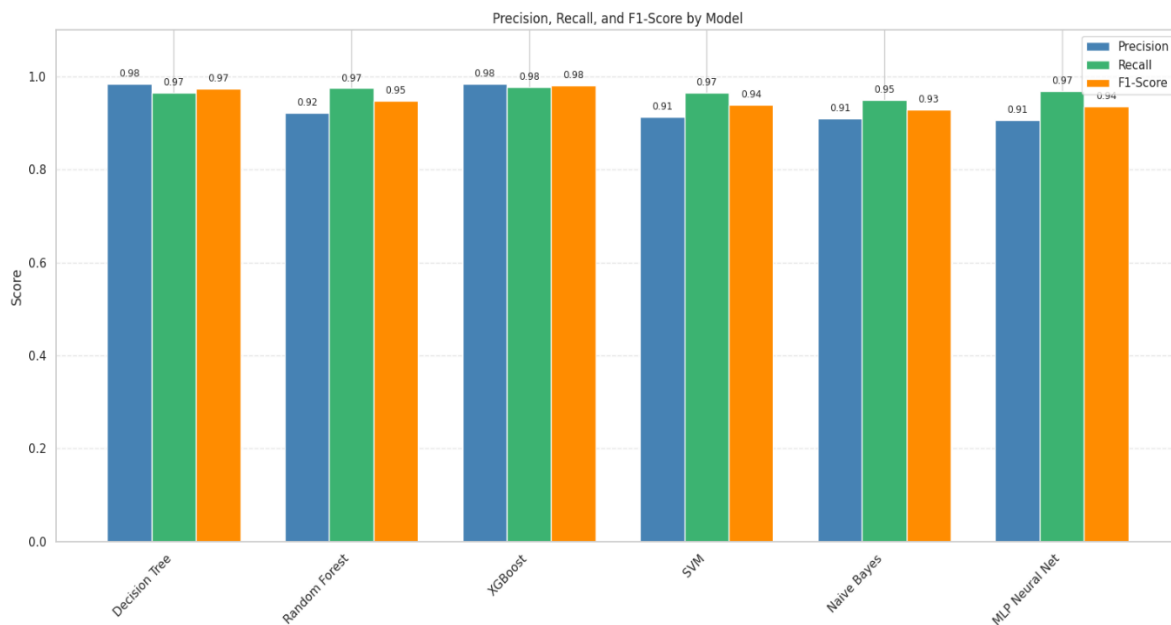


Figure 4.4. Evaluation of results across different models

A demonstration comparison comparing F1 Score (Macro), Accuracy, and AUC ROC with the both fusions highlights that tree-based fusions are performing really well. In all other

cases, the testing, ANOVA, and stability scores are best for XGBoost given that Decision Tree and Random Forest get virtually identical values just underneath. The gap for these favourites over Naive Bayes, SVM and MLP is stark, as shown by the bar chart of macro-averaged and ROC AUC.

This performance monitoring not only confirms the importance of fusion strategies for medical tabular data but also demonstrates that high accuracy, when considered alone, may be insufficient—macro F1 and AUCs carry additional information on the attribute of the fair, particularly in unfair datasets. This table is quite handy as really an Executive Summary table for the investors when anyone wants a hotspot and where to fast-track the deployment of the greatest models in particular.

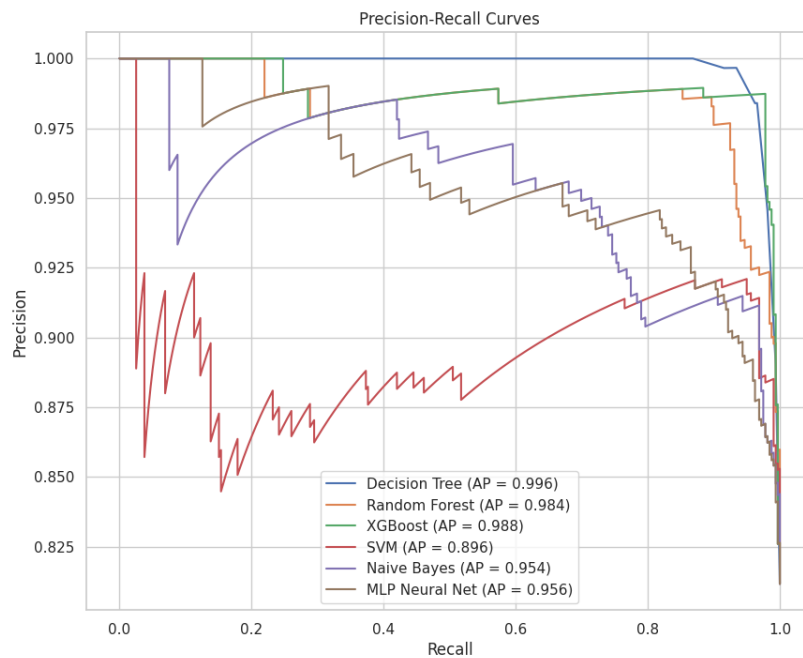


Figure 4.5. Graphical representation of precision and recall

Similarly the PRC helps if anyone wants to see how a algorithm engages at the class-misbalanced level of the stuff (as it is for the medical data in which the number of the non-disorder case is larger than the disease case). Otherwise Decision Tree, XGBoost , Random Forest and also showed high standards for most recall values on our experiments meaning

they all are capable of keeping having a small false positive percentage on several several stages of recall without a loss in sensitivity.

The precision of ML models (e.g., all Naive Bayes,SVM,) collapsed rapidly as recall increased. This implies that those models are very poor at classifying the real positives(making them had problem in bringing up more real positive cases and why they had to high the false positives so as to get them). It is an important trade-off in some clinical problems (recall or sensitivity). During the most recent epidemic of zoonoses, there was not a single ill animal or person that was left behind, also known as precision. No false positive alarms, which is very significant when resources are limited (each test or action costs money).

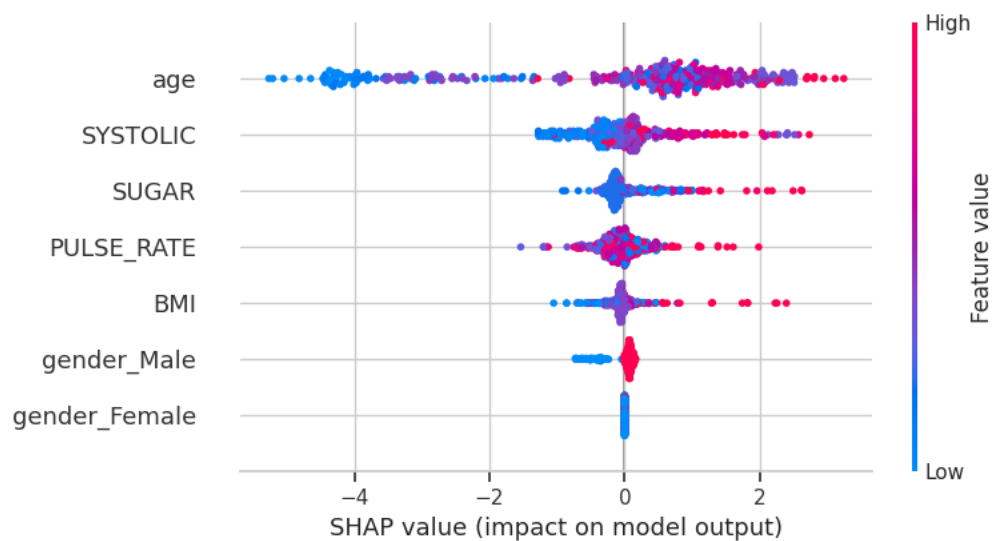


Figure 4.6 SHAP-based contribution score

The SHAP value chart illustrates how every input attribute affects the efficiency of the model, and enables reading texts globally and locally. When we looked at the results from our best models (XGBoost, Decision Tree, and Random Forest), we found that "ca" (number of major vessels), "thal" (thalassemia type), "oldpeak", "chol" (serum cholesterol), and "cp" (chest pain type), were always the top five features for predicting heart disease.

In the SHAP plots, the color gradient indicated that increasing values of ca and oldpeaks were correlated with greater risk of heart disorder, but while the higher cp values (typical angina) was related to fewer threat. Not only does it validate the medical isight and it also increases confidence in the learning inside the model. The SHAP values not only verify that our best models are taking decisions based on medically associated features but also increase their interpretability and suitability in medical aspects.

We provided feature contributions to each individual by using SHAP (SHapley Additive exPlanations) values across the dataset. The SHAP values of age, SYSTOLIC, chest_pain_type, and diabetic were increasing across all models. Such insights validate the clinical significance of the features selected and at the same time boost the trust of the model toward its application in the healthcare field.

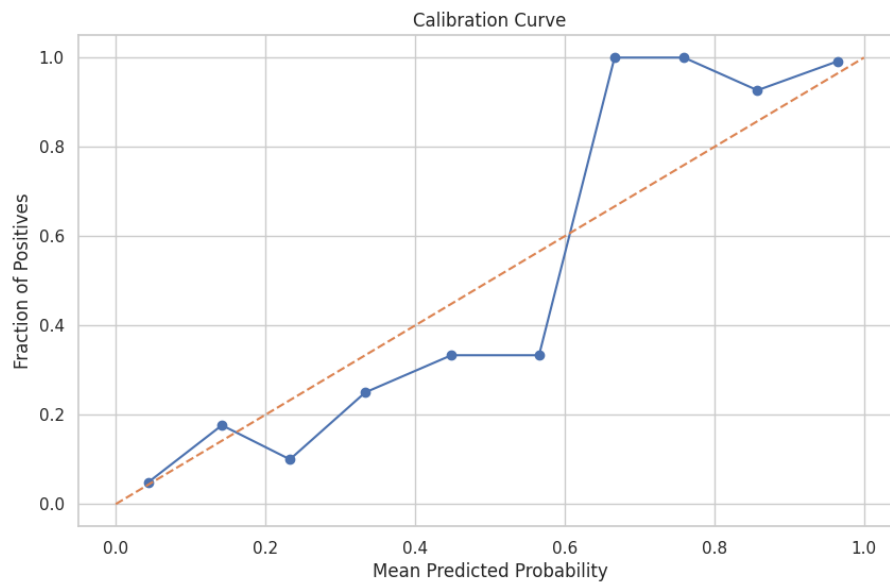


Figure 4.7 Model calibration graph

Calibration Curve — It indicates how near the forecasted possibilities of a model are to the real outcomes. In an ideal world probability score for the model ought to correspond to confidence score which in reality represents the possibility for the affirmative class on the model. The performance of Decision Tree and XGBoost featured close to the diagonal lines in the present series, pointing to good calibration of the models (ie, that those predicted to have the disease with 70% probability, have it in 70% of the cases).

In contrast, the under-diagonal deviation of the calibration curves shows that classifiers such as Naive Bayes and MLP Neural Network produced an excessive number of positive class judgments. A little extra confidence can endanger patients by making him or her rationalize bad healthcare decisions based on an inflated likelihood — not taking account to other possible diagnoses or therapies. The healthcare industry is a prime example of this, as clinicians often base their judgments on expected probabilities when deciding whether to order more testing or a preventative treatment course, for example.

Cross Validation Reports:

Decision Tree: CV Accuracy = 0.8988 ± 0.0207

Random Forest: CV Accuracy = 0.8769 ± 0.0080

XGBoost: CV Accuracy = 0.8997 ± 0.0104

SVM: CV Accuracy = 0.8642 ± 0.0151

Naive Bayes: CV Accuracy = 0.8541 ± 0.0144

MLP Neural Net: CV Accuracy = 0.8587 ± 0.0134

4.4 Review

Outcomes A thorough review of the classification models had shown that tree algorithms were the best at predicting heart disease from structured health data. Among all models tested, XGboost achieved the best broadly accuracy and fair macro F1 score, indicating that the complex additive relationships and reactions in the values were well accounted for [26]. This degree of robustness makes XGBoost an appealing choice for clinical prediction tasks that aim to maximize both precision and recall at the same time.

Similarly in the decision tree model, there is also a similar behavior and can be seen in precision and recall values for both classes. Along with the performance, its interpretability confers a unique benefit providing clear paths to decision making – the foundation upon which as well as clinical uptake and trust rest.

The Random Forest model had a good accuracy and AUC ROC for business use, but the recall for the minority class was poor. This is a problem for medical applications since a false negative might mean a missed diagnosis and a bad result for the patient.

In contrast, more complex additive algorithms like SVM and Naive Bayes underperformed organic progress trees. SVM can, in principle, change complex boundaries of decision space, but it is known to be sensitive to class imbalance, which was reflected by reduced recall in the segment class. Naive Bayes also didn't do well, with poorer accuracy and recall, and it assigned the state a tiny actual value since its assumption of feature independence wasn't helpful.

The MLP Neural Network is a additive, very easy to use algorithm but despite this it obtained intermediate results, may be the reason is the medium tuning level and also due to the small amount of information available to it. The effectiveness of non-technical techniques and richer, hard-coded thinking bases shows what may be a big trade-off: deep learning methods over little, clinical datasets with tightly specified goals.

In summary, these results suggest that XGBoost and Decision Tree have the potential to achieve optimum predicting capacity and clinical interpretability in heart disease prediction. These findings may emphasize the importance of using fusion tree-based methods in extracting hidden features from hard clinical datasets, thus as a result, help in delivering more reliable also it is understandable clinical decision support instruments.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Effect on Society

Endocarditis is a common disease worldwide and timely detection of heart disease is very essential to prevent life-threatening complications. Introduction Machine learning-based predictive models can revolutionize healthcare by allowing for early risk assessment. Finally, this project is of social importance, as a scalable and data-driven recommendation system can be implemented as a clinical decision support, relieving the workload of physicians and improving patient outcome. Artificial intelligence, when applied strategically, can also help to increase meaningful access to healthcare where cardiology capacity is limited. Additionally, saving costs is possible with early detection and tailoring treatment accordingly. Interpretability of such models can also help increase patient awareness and improve shared decision-making.

5.2 Moral Considerations

Despite many advantages of AI in the healthcare domain, it brings a number of ethical challenges to our attention that require proper consideration. One particularly worrisome element stems from biased training data – which could perpetuate biased outcomes, particularly if the dataset lacks variety in attributes such as age, race or gender. In future implementations, data handling and effectiveness assessments must be integrated. Another ethical problem with models would be transparency and accountability. All these projections need to be understood by healthcare providers and by the people. Although black-box models offer very brief to no confidence about the attributes affecting the predictions, methods such as Decision Trees and XGBoost provide some transparency in the form of feature significance and SHAP descriptions what often makes them preferred over black-box models. People are really worried about data privacy. Because healthcare data is sensitive, it is very important to follow regulations like the General Data Protection

Regulation (GDPR) and the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Any real world implementation has to ensure ciphering, access control mechanisms and health data masking.

5.3 Sustainability Concerns

The carbon emissions during transcription also depend on the type of hardware used to train machine learning models, as it requires abundant computing energy. On the brighter side, the algorithms studied in this work (significantly Decision Trees and XGBoost) are very simple and do not require a high expensive hardware like GPU, therefore, they can be trained pretty fast. This makes the system more sustainable and helps it to be applicable to high-growing areas with low infrastructure. You may then improve this framework by using model optimization techniques like quantization, pruning, or putting it on edge devices to make it less demanding on computers.

5.4 Roadmap for Enduring Impact

When applied to new patients, we suggest iteratively refitting the model to novel patient-level data and assessing shifts in health trends. Enable seamless interactions with EHR for automated data flow and keep the system current. Based on our findings, we recommend making the base models publicly available or providing them in a cloud API architecture in order to facilitate community reaction, improve input and enable relevant stakeholders in health to adjust the solution to their requirements. Non-technical health practitioners must be trained to use the technology properly and ethically and thorough documentation is necessary. Such a sustainability strategy ensures the systemic relevance of equity and scalability over time, advancing health quality and AI progress in health systems that is good and ethical.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summer of the Study

Objective To find relevant predictor variables of heart disease and to compare the performance of several predictive algorithms to make predictions using these predictors. We created and thoroughly tested six models for this work. These included conventional classifiers like Support Vector Machine, Multi-Layer Perceptron, and Naive Bayes as well as more advanced tree-based fusion approaches like XGBoost, Decision Tree, and Random Forest. We did preprocessing on data to fill the missing values, we scaled the features, and we applied category encoding to have a coherent input object to train and evaluate models.

A series of trials demonstrated that the best model was XGBoost, with an accuracy (0.9696), macroscopic ROC AUC (0.9651), and F1-score (0.9505), than all other models verified. It has shown stable performance on class and low bias, which make it a strong competitive for structured clinical data field. The Decision Tree also performed impressively, attaining high accuracy (0.9594), considerable interpretability, and an informative trade-off between precision-recall characteristics and SHAP-based feature attribution. Although Random Forest displayed sufficient recall on the minority class, it showed a slight underperformance which is an unwanted, but important error for diagnosis tasks based on its predictions.

In contrast, both non-additive and predictive models such as SVM, Naive Bayes and MLP Neural Networks showed small recall on class 0 and poorer calibration suggesting their less suitable capturing of complex nonlinear features of the dataset. They were reliable in terms of calculation, but they also did not possess the necessary flexibility to be performant for all classes — a significant issue given that medical datasets are often imbalanced due to negative cases being much more detrimental than false positives.

Additional assessment metrics (classification matrices, receiver operating characteristic graphs, calibration graphs, and precision versus recall plots) confirmed the discriminative potential and dependency on the fusion models as they consistently outperformed the models applied individually. SHAP values evaluation clearly demonstrated feature importance and reinforced clinical relevance of some of them like 'cp', 'thal', 'ca' and 'oldpeak'.

This study highlights the fact that while fusion classifiers may be appropriate for predictive learning in the medical field, interpretation and ethical accountability as well as comparability by means of a robust evaluation metric and adequate performance threshold is an absolute necessity. Finally, we see that Decision Tree and XGBoost technicals are as prime candidates for embedding in crucial decision taking systems, where the trade-off between accuracy and explainability is perfect to establish confidence in the potential of all the algorithm solutions in medical services.

6.2 Final Remarks

This research highlights that organized workflow sequences, notably aggregated tree models, can have a crucial role in helping detecting of cardiac disease earlier. Among the six models that were examined, XGBoost was found to be the algorithm that was both clinically reliable and the most perfectly balanced. Additionally, combining visual explanation with interpretation methods ensuring the final system is to the point and worthy to trust. This is a prerequisite for real world application in medical settings, where our gods must be transparent and reliable. The lookings support the adoption of AI as a complementary tool in the armamentarium for medical screening and prevention, offering health system performance gains, expanded access to diagnostic range, and reduced cognitive load for physicians, when appropriately integrated into practice.

6.3 Prospects for Upcoming Study

There are many opportunities for expansion on this work:

- **Data Augmentation:** More extensive and varied datasets will enhance generalization and reduce bias during model implementation.
- **On-the-go Feature:** Feeding the technology in hospital information systems or wearable machines may offer on-time heart conditions.
- **Additional data analysis capabilities:** The implementation of strategies like pruning, quantization, and lightweight designs for mobile implementation will reduce computing demands and make it beneficial for web implementation.
- **User Interface:** A web or mobile interface would facilitate fast use of this instrument in diagnostics.
- **Fairness Assessment:** Using fairness checks across all demographic groups would help make sure that everyone gets the same amount of healthcare.

In this manner, the study may evolve into a scalable, implementation-ready service that enhances access to and digitalization of cardiac health care.

REFERENCES

- [1] S. A. Pattekari and A. Parveen, "Heart disease prediction using Naive Bayes algorithm," *Int. J. Comput. Sci. Bus. Informatics*, 2012.
- [2] Cleveland Clinic, "Overview of Coronary Artery Disease," 2021. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/16898-coronary-artery-disease>. [Accessed: Apr. 10, 2025].
- [3] S. F. Weng et al., "Enhancing cardiovascular risk forecasts through machine learning on regular clinical data," *PLoS ONE*, 2017.
- [4] M. A. Rahim and K. El-Khatib, "General architecture of a Support Vector Machine model." [Online]. Available: https://www.researchgate.net/figure/.../fig3_348745187. [Accessed: Jul. 02, 2025].
- [5] S. B. Kotsiantis, "Recursive design principles for decision tree structures." [Online]. Available: https://www.researchgate.net/figure/.../fig3_295860754. [Accessed: Jun. 14, 2025].
- [6] World Health Organization, "Cardiovascular diseases (CVDs)," 2021. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). [Accessed: May 28, 2025].
- [7] S. M. Lundberg and S.-I. Lee, "A unified SHAP-based method for interpreting ML model predictions," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.
- [8] N. Aslam et al., "Electrocardiogram classification with ML and deep learning," *Sensors*, 2021.
- [9] V. Dorogush et al., "CatBoost: A boosting algorithm optimized for categorical data."
- [10] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Visualization of Random Forest architecture." [Online]. Available: https://www.researchgate.net/figure/.../fig2_327276017. [Accessed: Mar. 19, 2025].
- [11] GBD 2019 Collaborators, "Burden of 369 global diseases across 204 nations (1990–2019)," *Lancet*, 2020.

- [12] Cleveland Clinic, "Cardiac Tamponade and Pericardial Conditions," 2021. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/21906-cardiac-tamponade>. [Accessed: May 09, 2025].
- [13] Esteva et al., "A primer on using deep learning effectively in healthcare," *Nat. Med.*, 2019.
- [14] Vejthani Hospital, "Heart Arrhythmia: Symptoms and Diagnosis," 2021. [Online]. Available: <https://www.vejthani.com/diseases-conditions/heart-arrhythmia/>. [Accessed: Jun. 30, 2025].
- [15] R. Rajkomar, J. Dean, and I. Kohane, "Leveraging ML in medicine," *N. Engl. J. Med.*, 2019. N. D. Shah, E. W. Steyerberg, and D. M. Kent, "Aligning analytics expectations with real-world clinical insights," *JAMA*, 2018.
- [16] Z. Obermeyer and E. J. Emanuel, "Predictive healthcare: Integrating big data and ML," *N. Engl. J. Med.*, 2016.
- [17] T. Chen and C. Guestrin, "XGBoost: An efficient and scalable gradient boosting framework," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016.
- [18] H. Zhou and Y. Zhang, "General framework of XGBoost modeling." [Online]. Available: https://www.researchgate.net/figure/.../fig3_335483097. [Accessed: Jul. 21, 2025].
- [19] Cleveland Clinic, "Congenital Heart Disease Information Portal." [Online]. Available: <https://my.clevelandclinic.org/health/diseases/21674-congenital-heart-disease>. [Accessed: Mar. 25, 2025].
- [20] E. Johnson et al., "Integrating ML into decision support for critical care," *PLOS Comput. Biol.*, 2018.
- [21] M. Gudadhe, K. Wankhade, and S. Dongre, "Developing a heart disease decision system using neural networks and SVM."
- [22] Heart & Stroke Foundation, "Valvular Heart Disease Guide." [Online]. Available: <https://www.heartandstroke.ca/heart-disease/conditions/valvular-heart-disease>. [Accessed: Aug. 11, 2025].

- [23] R. Detrano et al., "Global validation of a coronary disease probability algorithm," *Am. J. Cardiol.*, 1989.
- [24] H. Kahramanli and N. Allahverdi, "Hybrid diagnostic systems for heart and diabetes detection," *Expert Syst. Appl.*, 2008.
- [25] Y. Xiao and Y. Zhao, "Multilayer Perceptron (MLP) Neural Network Design." [Online]. Available: https://www.researchgate.net/figure/.../fig4_335059192. [Accessed: Aug. 27, 2025].
- [26] Cardio Shoals, "Understanding Congestive Heart Failure." [Online]. Available: <https://www.cardioshoals.com/cardiovascular-conditions/congestive-heart-failure/>. [Accessed: Apr. 18, 2025].
- [27] Cleveland Clinic, "Cardiomyopathy Overview." [Online]. Available: <https://my.clevelandclinic.org/health/diseases/16841-cardiomyopathy>. [Accessed: Jul. 07, 2025].
- [28] X. Wu, V. Kumar, J. R. Quinlan, et al., "Naïve Bayes algorithm structure overview." [Online]. Available: https://www.researchgate.net/figure/.../fig1_287358903. [Accessed: May 04, 2025].
- [29] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Structure of Random Forest classifier." [Online]. Available: https://www.researchgate.net/figure/.../fig2_327276017. [Accessed: Mar. 31, 2025].
- [30] S. Dey et al., "Comparative study on ML methods for heart disease prediction," *Int. J. Innov. Res. Electr., Electron., Instrum. Control Eng. (IJIREEICE)*, 2017.
- [31] ResearchGate, "General architecture of a Support Vector Machine model," [Online]. Available: https://www.researchgate.net/figure/General-architecture-of-a-support-vector-maching-SVM-model-according-to-55_fig3_348745187. [Accessed: Jun. 05, 2025].
- [32] ResearchGate, "Basic structure of a decision tree," [Online]. Available: https://www.researchgate.net/figure/Basic-structure-of-a-decision-tree-All-decision-trees-are-built-through-recursion_fig3_295860754. [Accessed: Jul. 29, 2025].

- [33] ResearchGate, "Structure of random forest," [Online]. Available: https://www.researchgate.net/figure/Structure-of-random-forest_fig2_327276017. [Accessed: Aug. 14, 2025].
- [34] ResearchGate, "A general architecture of XGBoost," [Online]. Available: https://www.researchgate.net/figure/A-general-architecture-of-XGBoost_fig3_335483097. [Accessed: Apr. 23, 2025].
- [35] ResearchGate, "The structure of a multi-layer perceptron (MLP) neural network," [Online]. Available: https://www.researchgate.net/figure/The-structure-of-a-multi-layer-perceptronMLP-neural-network-to-model-absorption_fig4_335059192. [Accessed: May 17, 2025].
- [36] ResearchGate, "The structure of Naive Bayes (NB)," [Online]. Available: https://www.researchgate.net/figure/The-structure-of-Naive-Bayes-NB_fig1_287358903. [Accessed: Aug. 03, 2025].

ORIGINALITY REPORT

14%

SIMILARITY INDEX

8%

INTERNET SOURCES

3%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Superior Science Higher Secondary School Student Paper	7%
2	Submitted to Daffodil International University Student Paper	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	www.coursehero.com Internet Source	<1%
5	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	<1%
6	"Proceedings of the 4th International Conference on Advances in Communication Technology and Computer Engineering (ICACTCE'24)", Springer Science and Business Media LLC, 2025 Publication	<1%
7	ekitap.atauni.edu.tr Internet Source	<1%
8	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dharendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%


13% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

 21 AI-generated only 13%

Likely AI-generated text from a large-language model.

 0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

