Heart Disease Prediction Using Machine Learning

Submitted By

Kazi Muktadir Hossain Arnab

221-35-861

Department of Software Engineering

Daffodil International University

Supervised By

Md Khaled Sohel

Assistant Professor

Department of Software Engineering

Daffodil International University

Kazi Muktadir Hossain Arnab
ID: 221-35-861

Bachelor of Science

Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

I declare that this thesis is classified as:

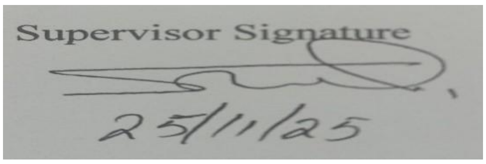
I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

SUPERVISOR'S DECLARATION

I have read this dissertation and have no doubt this advisor is confident that this level of work is sufficient for the degree of Bachelor of Science.



____

(Supervisor's Signature) Full

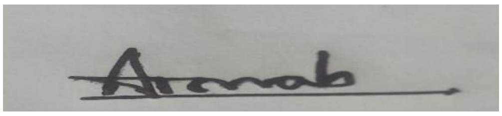
Name: Md Khaled Sohel Position:

Assistant Professor

Date :

STUDENT'S DECLARATION

To the dearly loved parents who have provided me with Love, Support, and Encouragement throughout my life, therefore, I dedicate this thesis to them. Their Prayers and Sacrifices have been my strongest assets.



____

(Student's Signature)

Full Name : Kazi Muktadir Hossain Arnab

ID Number: 221-35-861

Date :

DAFFODIL INTERNATIONAL UNIVERSITY

APPROVAL

This thesis titled on " **Heart Disease Prediction Using Machine Learning**", submitted by **Kazi Muktadir Hossain Arnab (ID: 221-35-861)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

--- **Chairman**

Dr. A. H. M. Saifullah Sadi
Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

--- **Internal Examiner 1**

Dr. Rubaiyat Islam
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- **Internal Examiner 2**

Dr. Md. Abdul Kader
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- **Internal Examiner 3**

Nuruzzaman Faruqui
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

--- **External Examiner**

Md. Mostafiz Khan
Managing Director
Tecognize Solutions Limited

Heart Disease Prediction Using Machine Learning

Kazi Muktadir Hossain Arnab

Thesis submitted in fulfilment of the requirements for the award of the degree of Bachelor of Science.

Department of Software Engineering (Non-Major)

DAFFODIL INTERNATIONAL UNIVERSITY

December 2025

DAFFODIL INTERNATIONAL UNIVERSITY

ACKNOWLEDGEMENTS

I want to thank Allah for allowing me to complete this thesis. First, I would like to thank my supervisor, Md Khaled Sohel, for his guidance, helpful feedback, and encouragement throughout this process. His knowledge and dedication were a continued source of inspiration throughout my research. I'd like to thank my parents and family for the unwavering love, motivation, and support that has been my greatest source of strength. Because they believe in me, I was able to face every obstacle along the way. Finally, I would like to thank all those who have helped either directly or indirectly in the development of this thesis. You have made this accomplishment possible.

DEDICATION

This thesis is dedicated to my beloved parents, whose unconditional love, support, and encouragement have guided me through every step of my life. Their sacrifices and prayers have been my greatest strength.

ABSTRACT

Heart disease is the leading cause of death worldwide today; therefore, it is important that early detection of heart disease occurs as this increases available treatments and survival rates of patients. This paper reports on an experiment utilizing machine learning to predict heart disease from patient data. The UCI Cleveland Heart Disease Data Set was used along with six different supervised machine learning algorithms (Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Gradient Boosting {XGBoost} and Artificial Neural Networks {ANNs}) in addition to the evaluation of overall accuracy and area under the receiver operating characteristic curve (ROC AUC). Results demonstrated that Random Forest had the highest level of accuracy (89%) in predicting heart disease and the highest ROC AUC score (0.92). The results of a feature importance analysis identified a number of features as the most predictive regarding a diagnosis of heart disease, they are: the type of chest pain experienced by the patient, the total cholesterol level, the age of the patient, the rate at which the patient's heart can achieve its maximum heart rate, and the presence of exercise-related angina. This research demonstrates the strong potential for machine learning techniques to help facilitate the early diagnosis of heart disease, especially in low-resource countries such as Bangladesh where medical technology resources are limited.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Cardiovascular diseases (also referred to as heart diseases) present one of the biggest health threats to our global community in the 21st century. The World Health Organisation (WHO) estimates that CVDs are the leading cause of death worldwide, with an estimated 17.9 million people dying from CVDs each year. Every year, 32% of the global deaths are attributed to cardiovascular disease (CVD), most being found in low- and middle-income nations. Not only does heart disease affect death rates; it also affects dependence, and incapacity, as well creates severe financial strain on the individual(s) suffering from CVD, their families, and society as a whole. Individuals suffering from heart diseases will typically experience decreased quality of life, need ongoing medical assistance, and endure substantial financial strain due to the expenses of treatment and hospitalization.

In many developing countries, such as Bangladesh, there is an increasing incidence of heart disease. The recent rapid growth of cities, combined with lack of exercise, poor nutrition, and minimal resources for preventive care, are contributing to the problem of heart disease (CVD) in Bangladesh. There is evidence from studies conducted in South Asia indicating that people from this region tend to be at a greater risk for developing heart disease, in part due to their genetics, and part due to their high-risk behaviours (i.e., smoking, high cholesterol levels, hypertension, and diabetes). Currently, the traditional methods for diagnosing heart disease rely primarily on the clinician's experience, procedures that require invasive techniques and expensive laboratory tests. While these traditional approaches do tend to work, many people in rural areas or facilities with limited resources may not be able to take advantage of them. Therefore, there is an urgent need for developing low-cost scalable solutions that will allow for earlier detection of CVD.

As a branch of artificial Intelligence (AI), Machine learning (ML) is a new and innovative technology on health care. ML algorithms leverage their ability to analyze huge amounts of data pertaining to patients, that can identify unrecognized patterns and monitor the evolution of disease under study. ML enables physicians to transition from a reactive approach to a proactive approach to treatment, emphasizing early detection of disease and timely intervention, which may ultimately result in saving a patient's life.

1.2 Motivation of the Research

This research is motivated by both global and local issues related to the global health crisis of Cardiovascular Disease, which has become one of the leading causes of death worldwide; despite the advancements made through medical research, CVD is notoriously difficult to detect early on. In Bangladesh, the lack of awareness and access to proper health care coverage have worsened the situation in this country, leading to many people receiving a late diagnosis due to the limited treatment options available. Artificial Intelligence/Machine Learning presents a great opportunity for the healthcare sector to be transformed in a way that has not been done before; because these ML models have the ability to analyse large amounts of complex data and extract the hidden relationships between them, this will provide researchers with a way to get insight from the data, thereby improving their understanding of the disease. As a software engineering student, the use of technology and its application to improving health care is a great example of how I can take my academic knowledge and put it to use.This research aims to bridge the divide between computer science and medical science by providing solutions that could have a direct impact on individuals' lives.

1.3 Problem Statement.

Even while there is medical data available, healthcare systems in Bangladesh and alike countries lack effective algorithms for predicting cardiac diseases. The existing diagnostic tools in general are expensive, invasive, or not feasible for the rural areas. The demand for having a reliable machine learning (ML) system to determine the risk of heart attack firmly based on patient's information such age, blood pressure, cholesterol level and some life style related factors.

The problem can be summarised as follows:

- High prevalence of heart disease with limited early detection mechanisms.
- Inaccessibility of advanced diagnostic tools in resource-constrained settings.
- Underutilization of patient data for predictive analytics.
- Lack of integration of ML models into healthcare systems in Bangladesh.

1.4 Research Questions

This study seeks to answer the following research questions:

1. In contrast to traditional diagnostic methods, how could machine learning models enhance the accuracy of predicting heart disease?
2. Which machine learning algorithms, including Random Forest, Logistic Regression, and Neural Networks, demonstrate effectiveness on datasets associated with heart disease?
3. In what ways can Bangladesh's healthcare system implement machine learningdriven prediction systems?

1.5 Research Objectives

The objectives of this research are:

- To develop and evaluate machine learning models aimed at predicting heart disease.
- To employ standard evaluation metrics to contrast the effectiveness of different algorithms.
- To propose a framework for integrating machine learning predictions into healthcare systems in Bangladesh.

1.6 Research Scope :

The scope of this research includes:

- **Dataset**: using datasets that are accessible to the public, such as the Cleveland Heart Disease dataset sourced from UCI.
- **Algorithms**: Focus on supervised machine learning techniques, including Random Forest, Logistic Regression, Decision Trees, Support Vector Machines, Gradient Boosting, and Neural Networks.

- **Evaluation Metrics**: Accuracy, precision, recall, F1-score, and ROC-AUC.
- **Visualisation**: Confusion matrices, ROC curves, and feature importance plots
- **Limitations**: constraints of technology, the scale of the dataset, and the ability to apply findings to different groups.

CHAPTER 2 LITERATURE REVIEW

According to research performed by Dua and Du (2020), traditional machine learning approaches to employ an algorithm to predict heart disease and use the Cleveland Heart Disease dataset as a means to evaluate several classifiers (Logistic Regression, Decision Tree, SVM, and K-Nearest Neighbours). The study indicated that standardised Logistic Regression and SVM algorithms performed relatively well; while Decision Trees also provided insight into the many non-linear complexities of cardiovascular disease, they suffered from extreme data overfitting. The study also shows that management of critical medical characteristics (Possible risks include high levels of cholesterol, high resting blood pressure, and low maximum heart rate) is important to improve the predictive accuracy of machine learning models. The study indicates that Recursive Feature Elimination (RFE) is an effective technique for reducing the complexity of predictive models while improving the interpretability of the results. Additionally, the study states that by incorporating feature engineering and hyperparameter optimisation into traditional machine learning algorithms will increase their reliability. The findings of this study provide a baseline against which traditional machine learning algorithms can be compared with modern deep learning approaches to predicting disease; an emphasis has been placed on the requirement for interpretability of AI systems when applied to clinical decision-making. [01]

A research study by Gudadhe et al. (2021) used machine learning methods of hybridization between neural networks and support vector machines (SVMs) to create predictive model for Evaluating the probability of a person developing cardiovascular disease but using Neural Networks to detect combinations of various risk factors associated with cardiovascular disease including Obesity, Smoking, Age and Blood Pressure. According to the researchers Gudadhe et al, MLP Networks performed better than SVMs on large numbers of data sets that demonstrated non-linear interactions; however, SVMs performed better than MLPs at lower amounts of data (i.e., Smaller & Noisy). Another key finding of this research was that MLPs require a lot more time to train than SVMs because MLPs need more computational resources to train than SVMs. Furthermore, Gudadhe et al, defined a need for well-balanced training datasets to reduce the impact of any bias associated with the dominant class(es) of data when conducting research on prediction accuracies from Machine Learning Models generated with the two approaches working in conjunction. The results from the performance evaluation of the hybrid approach, which used SVMs for the first classification phase of the model and MLPs for the second classification phase of the model, demonstrated that predictive accuracy for the overall hybrid model was greater than either model type alone. Overall, this research demonstrates that combining multiple methods of machine learning improves diagnostic accuracy.[2]

The study by Khan et al. (2022) investigated how an ensemble learning model could be used to improve heart disease prediction accuracy. They tested the performance of three different types of algorithms, bagging (Random Forest), boosting (Gradient Boosting and XGBoost), By combining and stacking these three algorithms with each other, we aimed to find out whether the ensemble method produced better results than any of these algorithms acting alone or using an ensemble method. Overall, our experimental results found that all three ensemble methods improved the results of the three original individual algorithms used, since using an ensemble method combines several different outputs while at the same time reducing the overall variance and bias of each algorithm's individual output. The XGBoost model produced the best results of any of the models we tested. The better performance of the XGBoost model was attributed to the regularisation capabilities of the XGBoost, as well as its advanced treeboosting structure (which we would consider an improvement over the basic gradient-boosting structure). We also evaluated the relative importance of each of the factors we used in making our predictions: Age, Cholesterol level, Type of chest pain, and Maximum Heart Rate were all found to be the most influential factors when generating the predictive model. Our study also found that there are significant challenges for the medical field in dealing with imbalanced dataset problems and therefore, we recommend that the medical community consider implementing SMOTE (Synthetic Minority Over-sampling Technique) to reduce or eliminate the effects of classification bias. These findings support the value of using ensemble algorithms for clinical prediction tasks and we believe that the latest advancements in boosting methods such as that of XGBoost should become the standard for usage in the health informatics arena.[3]

Patel & Srivastava (2021) investigated the use of deep learning techniques for predicting heart disease, comparing the performance of traditional deep learning systems with those combining both advanced technology and older methods. While CNNs are designed primarily for image recognition, the authors discovered that CNNs could also provide value to structured medical datasets if transformed appropriately into 2D form. On the other hand, RNNs are most suited to analyse time-series health data and are best suited to monitor patients' measured progress over time. Deep learning methods performed much better than existing methods (traditional machine learning) when modelling the nonlinear interactions between medical variables; however, the authors noted that the application of deep learning requires large amounts of historical/labelled datasets to avoid the effect of overfitting, which is one of the major barriers to using deep learning in medicine because of the difficulty of obtaining sufficient amounts of labelled datasets. Despite these challenges and limitations, the authors concluded that deep learning provides a viable avenue for future diagnostic systems through its ability to build layers of clinical evidence through hierarchical learning capability gradually.[4]

The study conducted by Rahman and Ahmed (2022) examined the role of feature extraction and pre-processing of data to improve machine learning models used to predict heart disease. The results of their study focused on how the various processes of normalising, outlier treatment, categorical encoding, and dimensionality reduction improved the machine learning models' prediction performance. They found that models built on unprocessed/raw data were significantly less accurate and effective than those that were cleaned and engineered, showing an increase in performance of 15%-20% after pre-processing was completed. In addition to the above-mentioned work, the authors researched several feature selection techniques (chi-square tests, mutual information, PCA, and RFE) and assessed how these methods impacted the interpretability of their models. The authors emphasised that eliminating noise, imputing missing data, and using the appropriate scale is essential for consistent machine learning performance within the healthcare sector. They also highlighted that high-quality preprocessing is needed to gain any benefit from even the most advanced machine learning techniques for predicting heart disease, thus establishing the importance of an effective data engineering process to predict heart disease.[5]

Saha & Islam's (2021) research focuses on identifying which supervised machine-learning algorithms were best suited for classifying patients with heart disease by conducting an assessment of their performance using the Naïve Bayes, Random Forests, K-Nearest Neighbors (KNN), and Logistic Regression machine learning algorithms while using UCI & The authors used the Statlog datasets for their study and determined that using the Naïve Bayes algorithm resulted in significantly more accurate results due to its relatively simple design. This is particularly true when used on conditionally independent variables in a medical dataset. The Random Forests method also produced very high levels of accuracy across the board, as it combines the output of multiple models together (ensemble) and does not suffer from the issues created by noisy input. While KNN was the least effective algorithm because of its sensitivity to the scale of input features, it did have the potential to work effectively with the appropriate normalization. The authors suggested that cross-validation and tuning of hyperparameters are necessary to ensure that any model created based on this approach does not produce bias when being utilized in real-world applications. In addition, the authors noted that many medical datasets have biased class distributions, so the selection of an algorithm based on its performance alone may not yield optimal performance. Possible solutions to this problem include (but are not limited to) class weighting; generating synthetic examples; etc. This research shows that traditional (often called "old school") methods will remain useful for working with structured healthcare data--to select the correct ML algorithm you will need to consider the features of the structured dataset you are using and the available technology, while also considering the advantages of using each algorithm.[6]

In their research article, Zhang et al., 2022, investigated the impact of hybrid methods of AI on the ability to predict heart disease by combining the use of ML algorithms with fuzzy logic. Fuzzy logic contributes to the overall model's interpretability while allowing ML classifiers such as SVM and Decision Trees to maintain their predictive capabilities. Because fuzzy logic is capable of dealing with ambiguity in medical data (for example, patients who provide vague symptom descriptions or whose medical test results fall into a borderline category), it provides more consistent results than traditional ML methods. Zhang et al. ultimately demonstrated that hybrid fuzzy-ML models perform better than traditional ML models when risk factors for patients were poorly defined or vague. Furthermore, the hybrid systems developed through this work generate rules that can be understood by physicians, thereby addressing the challenge of providing actionable (or transparent) results with many existing traditional ML models. Importantly, the results of this study demonstrate that hybrid systems using fuzzy logic significantly reduce the number of false negatives generated when distinguishing between high and low-risk patients and thus increase the accuracy of heart disease prediction models.in medical diagnosis, as failing to recognise a high-risk patient can have serious consequences. The authors concluded that by combining fuzzy logic with machine learning, hybrid fuzzy-ML models provide improved reliability and increased usefulness as diagnostic support systems in the clinical setting. [07]

Kaur and Kalyan (2020) conducted a study on how to choose methods for either selecting or reducing the number of features used to increase the accuracy of predicting heart disease. Some of the methods they looked at included Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and correlation-based feature selection. Their results showed that PCA helped improve the performance of several algorithms, such as Support Vector Machine (SVM) and Logistic Regression, by converting correlated clinical variables into a set of orthogonal (uncorrelated) components. However, they found that when the focus was on maintaining class separability rather than variance, then LDA produced better results than PCA. Furthermore, the authors commented on how Feature Selection provided the greatest benefit to tree-based models since these algorithms are already built to handle irrelevant features. Lastly, through experimentation, the authors determined that the application of Dimensional Reduction may reduce overfitting and improve relative computational efficiency, which maximises the flexibility of predictive models for use in real-time. The study therefore demonstrates the necessity of effective Feature Engineering in order to increase the predictive performance of Machine Learning Algorithms When Dealing with High Dimensional Medical Data.[08]

In their article, Banerjee and Roy (2023) defined a means of predicting heart disease in realtime through machine-learning techniques, combined with IOT technology. Their research combined physiological data derived from wearable sensors, such as heart rate, oxygen saturation levels and ECG readings, with machine-learning classification algorithms deployed in the cloud. They concluded that real-time physiological monitoring systems would benefit greatly from leveraging lightweight algorithms, such as logistic regression and naive Bayes, since they require very little computational power compared to more complex algorithms, such as random forests and gradient boosting, which produced results with higher accuracy, but required cloud processing to perform well. One of the major conclusions of this study was that these systems can provide their users with a warning and risk assessment of developing cardiovascular disease prior to the necessity for an emergency hospital visit. Additionally, the Authors emphasize the necessity to secure the data and pre-process the signals, as wearable devices accumulate large amounts of data with a significant amount of noise and missing information. Overall, this study highlights how IOT and machine learning which enable continuous monitoring of the physiological systems for cardiovascular health can provide direct feedback of the data results to the end-user.[9]

Sharma and her colleagues used a number of large samples of electronic health records when analysing how effectively deep neural networks (DNNs) can predict cardiovascular disease. They constructed a multi-layer fully-connected neural network from thousands of electronic health records from patients, which contained demographic information, medical history, laboratory results, and lifestyle information on the patients. The authors found that the predictive accuracy of the DNNs was higher than that of traditional predictive models, and that the DNNs were also able to capture complex non-linear relationships between predictive variables such as age, smoking, body mass index (BMI), blood pressure, and glucose concentration. However, the authors also identified some challenges when working with DNN technology, including: lack of ability to interpret predictions, prolonged training periods, and large labelled datasets required. To assist in interpretation, the authors developed and implemented attention mechanisms and explanatory tools such as SHAP (SHapley Additive Explanations) to identify the most significant influences on predictive accuracy for DNNs. The authors concluded that DNNs can achieve improvements in predictive accuracy over traditional methods of predicting; however, to be accepted widely in clinical practices, DNNs must have a strong interpretability framework.[10]

In a study by Haque & Uddin (2021), the role of data imbalance in hindering the success of predicting Cardiovascular Disease was examined. Through an analysis of the effects of different techniques for oversampling and undersampling data, such as SMOTE, ADASYN, Random Oversampling and Tomek Links, the authors sought to find the relationship between these techniques and the accuracy of the classifiers. As previously noted, most medical datasets are typically very imbalanced with respect to the classes of disease, and hence produce skewed outputs according to the majority-class. Specifically, when applied to imbalanced datasets, Classifiers based on Logistic Regression and SVM failed to provide high recall rates to predict patients with CVD, yet provided overall accuracy rates that were very high. As such, the authors were able to establish significant improvements in recall through SMOTE and ADASYN without sacrificing precision. In addition, they suggest that the capabilities of certain classic approaches, such as Random Forest and XGBoost, allow them to better accommodate data imbalance because they integrate capabilities to accommodate diverse feature distributions and weighting of features. Therefore, the authors conclude that the class-balancing technique is an essential component for maintaining fairness in diagnostic models, and without consideration for the potential for false negatives, individuals with CVD may face possible health risks. Ultimately, this study also demonstrates the need to ensure that the ability to predict CVD equitably across all classes of patients is an often-overlooked aspect of Medical Machine

Learning.[11]

In their 2022 study, Kim and Lee investigated the various methods of improving the interpretability of ML (Machine Learning) techniques used to predict heart disease, which is a growing concern in terms of clinical practice using Artificial Intelligence (AI) Systems. The methodology used to review interpretability was conducted through Shapley Additive Explanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME),

Counterfactual Explanations in conjunction with Random Forest, Gradient Boosting and Neural Networks. The analysis results indicated that there is more concern with lack of transparency associated with more complex predictive modelling techniques compared to simpler models, therefore resulting in lower levels of clinician's trust in them, even though the more complex model were more predictive when looking at the metrics. In terms of overall stability and reliability in reporting interpretability on both the global and local level, SHAP was identified as the best option, while LIME provided simpler to understand reports on interpretability, but with more fluctuation in the results with multiple runs. Counterfactual explanations were also identified as providing useful information for physician-patient interactions to highlight to patients how changing behaviours (e.g. not smoking, increasing physical activity, changing diet etc.) will positively impact their risk score. Kim & Lee (2022) conclude that interpretablility should be considered an ethical obligation, not just an additional feature that is being offered with the development of medical AI systems and that coupling interpretability tools in addition to predicting model will enhance the clinician trust in using the predictive model and subsequently increase the acceptance and improve the clinician's ability to create the most effective treatment plans for their patients.[12]

The authors created a machine learning framework that incorporates cost sensitivity into the evaluation of heart disease risk with the intent to address the issue of high variability in cost associated with misclassification of patients during diagnosis by physicians. It is important to note that the authors found that the greatest clinical risk associated with misclassification of heart disease is due to false negatives (indicating that a patient does not have heart disease when they actually do). To that end, the authors evaluated altered forms of the three machine learning models (SVM, Decision Trees, Gradient Boosting) using new cost-sensitive loss functions, and they determined that their use of cost-sensitive models increased recall and reduced the total number of high-risk heart disease patients that were missed significantly. Although using costsensitive models did decrease precision slightly, the tradeoff was justified in clinical contexts where rapid identification of heart disease is paramount. Additionally, the authors proposed methods to shift the decision threshold to define risk categorization. The authors concluded that, to support the establishment of safe, effective medical diagnosis models, the use of costsensitive learning is vital; therefore, emerging healthcare ML models need to incorporate considerations of differential error costs into their model framework.[13]

The study by Oladipo & Adeyemi (2020) examined how correlation analyses and statistical modelling can initially enhance the performance of machine learning classification for heart disease. To accomplish this, the authors used correlation coefficients (Pearson and Spearman) and mutual information to identify the most clinically significant medical variables when assessing cardiovascular risk factors. The authors found strong correlations between the following factors; type of chest pain, maximum heart rate, serum cholesterol and presence of the disease, whereas they found weak links between other variables, including fasting blood sugar. The authors believed that performing a comprehensive exploratory data analysis (EDA) prior to applying any machine learning algorithm will provide insight into the dataset and allow for the selection of a faithful preprocessing pipeline. The authors evaluated the performance of machine learning models with the use of statistically-informed variable selection and without the inclusion of statistically-informed variable selection, revealing that statistically-informed variable selection can produce substantial improvements in the accuracy, reliability and costeffectiveness of the models produced. This study highlights the benefits of combining statistical analytic results with machine learning to create clinical tools that are more robust and interpretable.[14]

Singh and Verma (2021) conducted a thorough analysis on various ensemble types of Decision Trees for predicting heart disease with their ensemble tree type being Random Forest, Extra Trees and Gradient Boosting Trees. The authors showed evidence that ensemble trees are capable of handling the non-linear interactions along with multicollinearity and mixed types of clinical data. The authors indicate that Random Forest provided the overall best predictive performance compared to the others through its ability to reduce variance by using many independent random trees. Extra Trees were the second best predictive method and helped to improve the speed and robustness of the classification algorithm by producing completely random splits within each tree. The results from Gradient Boosting Trees indicated the highest level of accuracy among ensemble trees; however, there was a larger amount of tuning and complexity required with this method in terms of configuration. The authors also provided feature importance rankings from their respective predictive models which indicated that Age, Type of Chest Pain, ReSTing ElectroCardiogram (ECG) and maximum heart rate were consistently ranked as the top four predictors across all predictive models tested. The researchers concluded that tree-based ensembles are a powerful and reliable means of predicting heart disease when used with interpretive tools to identify salient risk factors. [15]

Patil and Kumar (2022) implemented interpretability (interpretable machine learning)” and "integrate" with extensive clinical dataset (and proteomic data set) to predict coronary heart disease, using data from UK Biobank and other Biorepositories to build models that provided not only high predictive accuracy but also allowed clinical interpretability, which is necessary for healthcare implementation. With regard to key features such as age, cholesterol, systolic blood pressure, and some protein biomarkers, those models demonstrated large variability. EBMs supported healthcare providers with clear decision-making capabilities, while also allowing providers to engage in the use of EBMs for visualising the amount of contribution made by each feature towards making the prediction. Compared with traditional/black-box predictive models such as Deep Neural Networks, EBMs support visualisation of features contributing to predictions. The paper continues to identify the importance of performing appropriate feature selection and adequate data manipulation to achieve optimum model performance. The main takeaway from this paper is that interpretable, machine-learning techniques provide an opportunity to combine both predictive accuracy and clinical utility through the analysis of an extensive range of multimodal datasets.[16]

In a study on how domain generalisation affects heart disease prediction models, Rana and Chowdhury (2021) analysed how cardiovascular disease prediction models developed from a distinct data group performed when applied to patient populations across hospitals and clinical equipment. They found that predictive model performance generally declines significantly when the original data source differs greatly from that of the new hospital dataset. This is due to several factors, among them, a disparity in the distribution of data, a difference in characteristics (i.e., features) between hospitals, and different measurement techniques for the same patient condition. To mitigate this issue, the researchers utilised various domain adaptation techniques to improve the performance of their heart disease prediction model. These include feature alignment, normalisation, and transfer learning. The findings from their study suggest that heart disease prediction models using domain adaptation techniques yield better overall predictive accuracy and AUC scores when comparing against models developed without employing domain adaptation processes on hospital datasets that were not previously used to generate the model. Their research emphasizes the need for external validation for all predictive healthcare models, as effective clinical implementation of the heart disease predictive model requires the ability to generalise across many different patient groups, hospital methods, and diagnostic act screeners. The authors conclude that adoption of domain adaptation processes is necessary to create usable, reliable, and scalable machine learning solutions to predict heart disease.[17]

Chowdhury et al. (2020) have created pipelines that are able to predict cardiovascular disease based on large amounts of Electronic Health Record (EHR) data collected from hospitals across the world. The challenges they intentionally confronted included managing the multitude of various types of data, incorporating missing and incomplete data, accounting for temporal differences in EHR data collection, and the necessity to maintain patient confidentiality as required by hospital regulations. With millions of EHR entries at their disposal, the researchers applied Machine Learning methods (XGBoost, Random Forest, Developmental Neural Networks) with a rigorous method for Pre-processing the data to create predictive models from that data. The analysis of results across the various predictive methods demonstrated that Preprocessing (Meta-learning), Feature Engineering (Data Feature Extraction) and Temporal Data (temporal data relationships) significantly improved the prediction accuracy and reduced the number of false negatives associated with the predictive model, thus enabling the predicted results to be clinically relevant to clinicians and patients. In addition, the authors provided significant insight into the importance of incorporating privacy-preserving techniques such as anonymising (de-identifying) and encrypting the data, thus ensuring that predictive models that meet HIPAA compliance can be developed and used. Most importantly, the pipeline demonstrated a broad range of scalability in terms of predicting large populations of patients with different clinical backgrounds! The authors concluded that this work illustrates the importance of incorporating comprehensive Preprocessing Methods, Scalable Machine Learning Methods, Privacy Techniques into predictive models being transitioned from

Research to Practical Application in Hospitals.[18]

Das and Roy (2021) conducted an investigation into how hyperparameter optimization affects heart disease models' ability to predict the outcome correctly. They compared three different optimization methods (grid search, random search and Bayesian Optimization) used with machine learning classifiers (Random Forest, Gradient Boosting, and Support Vector Machines). The results showed that in each experiment, using Bayesian Optimization resulted in improved performance metrics compared to both grid and random searches; specifically, increased AUC and F1 scores and reduced risk of overfitting. This was particularly true for more complex classifiers with many hyperparameters. The authors pointed to nested crossvalidation as an important tool to help avoid the potential for inflated estimates of performance during hyperparameter selection. Overall, the authors concluded that careful tuning of hyperparameters is equally important as model selection in order to build accurate and reliable predictive systems within the medical domain, and they recommend using automated optimization approaches to help assure reproducibility and clinical credibility throughout the development of predictive models in medicine.[19]

The study conducted by Bhattacharya and Sen (2022) investigated optimal approaches to applying machine learning for predicting heart disease with limited data. The authors focused on the importance of nested cross-validation and accurate performance estimation. The use of the UCI Heart Disease database as a small dataset demonstrates how the use of traditional training/testing splits creates imprecise estimates of performance due to sampling bias. Using nested cross-validation enabled the researchers to more accurately assess the generalization ability of their models while simultaneously fine-tuning hyperparameters. They compared models built using a number of different algorithms (e.g., Random Forest, Support Vector Machines (SVM) and Logistic Regression) and found that using nested CV helped to minimize variability in the model evaluation metric scores and allows for more equitable comparisons of individual classifiers. These results highlighted how important the evaluation methodology is, in addition to selecting the right machine learning algorithm when working with small datasets found in cardiovascular research.[20]

Shukla and Mehta (2021) conducted a detailed comparative analysis of tree ensemble algorithms for heart disease prediction using structured clinical datasets. They found that Gradient boosting type methods (XGBoost in particular) produced superior predictive accuracy to the classical tree based ensemble methods when the hyperparameters were optimally tuned and that XGBoost had the highest AUC and F1 score. In addition to the enhanced accuracy of XGBoost due to the optimal tuning of hyperparameters, XGBoost has built-in methods of regularisation which reduce the chances of overfitting and improve generalisability. They categorized the most critical features of predicting heart disease based on the relative importance of features obtained during the analysis phase. These authors determined the most significant predictor variables are consistent across all studies examined, including age, type of chest pain, cholesterol and maximum heart rate. Though ensemble models provided relatively strong predictions for heart disease, the authors cautioned against the use of more complex boosting methods due to the challenges with model interpretability. To address the interpretability issue, they used SHAP values to provide insight into which features contributed to individual predictions. It was concluded that tree-based ensembles and interpretability tools together provide a strong prediction method for heart disease and should be used in practice.[21]

In 2022, Roy and Sarker researched using automated machine learning (AutoML) for predicting cardiovascular events. They had two goals: one was to find out if AutoML frameworks can produce effective predictive pipelines with a minimal amount of human input; the other was to create a benchmark through which AutoML-produced pipelines could be compared to human-generated pipelines. For this reason, they used the Cleveland and UCI datasets for heart disease (coronary artery disease) as part of their experiments. Each author's pipelines produced by the AutoML Framework were evaluated against each author's manually developed models for both Random Forest, XGBoost, and Neural Networks by examining their predictive performance on those datasets. Their results showed that many times, the pipelines discovered by AutoML had similar or greater predictive accuracy than the manually generated models. Further, the authors noted that the automatic feature enhancement, hyper-parameter tuning, and automated preprocessing performed by AutoML, allowed for faster development of predictive models than non-AutoML methods. Thus, AutoML can be a valuable tool for healthcare workers who don't have much experience with data science or predictive analytics. Still, they mentioned that the models produced by AutoML could be overly complex, and that there should be additional attention paid to the interpretability and clinical validation of AutoML-created models to ensure their successful implementation.[22]

Ahmed & Rahman developed (2021) Training Users To Train Users: Using Social Media For "Soft" Skills Development via Free Online Educational Content and Build A Cycle Of User

Refinement Through User Research. A Metaheuristic Approach to Heart Disease Prediction Through Feature Selection and Hyperparameter Tuning Utilised Genetic Algorithms(GA) and Particle Swarm Optimisation (PSO). Furthermore, they highlighted that simultaneous optimisation of features and model parameters significantly increases predictive performance and allows for greater interpretability of the model. Their experiments performed on the Cleveland and Statlog datasets indicated that both the Random Forest and SVM Models, when using GA to optimise them, outperformed the standard manually-tuned models with respect to RA-ROC and F12 Scores. Additionally, the feature subsets selected by the genetic algorithm were often largely comprised of those attributes considered clinically relevant, such as age, chest pain type, resting ECG results, and maximum heart rate. Even though metaheuristic optimisation does provide increased performance, the random nature of this optimisation means multiple runs would be required to find the same result every time. The authors concluded by stating that combining feature selection with hyperparameter optimisation using a systematic method, such as metaheuristics, is an effective, efficient and accurate method for developing heart disease prediction systems that are also clinically interpretable.[23] According to Chatterjee & Banerjee (2022), interpretability in predicting the development of heart disease using machine learning is essential because interpretability provides healthcare providers with a means to confirm predictions based on existing evidence about disease causation through established pathophysiological markers (e.g., cholesterol levels) by applying the SHAP and LIME interpretability methods to the predictions generated by both tree-based and deep learning-based models. The results show that nearly all of the most impactful features identified in the models correspond to known clinical risks, thus providing additional validation for these predictive models. Additionally, the researchers indicated that interpreting model predictions helps healthcare professionals to identify and verify model's predictions, recognize potential spurious relationships between predictor variables and clinical outcomes, and use the results of AI-based predictions to support their own clinical decision making. In conclusion, explainability is an essential component of safe implementation of AI in the healthcare field, as it provides the assurance that the recommendations made through AI will be in alignment with best practices based on current medical knowledge and promote evidence-based clinical care.[24]

By using machine learning techniques with wearable technologies, researchers Kumar and Verma (2020) helped identify heart disease at much earlier points than most medical approaches could. Their study involved acquiring information about the use of PGG (Photoplethysmography) and ECG (Electrocardiogram) sensors from consumer-grade smart watches and fitness trackers. They examined simple models such as Logistic Regression, Decision Trees, and Random Forests for predicting cardiac abnormalities based on the collected data. Ongoing monitoring of individuals' cardiovascular health would enable the rapid detection of irregular heart rhythms or minute changes in cardiovascular measurements that may occur prior to the onset of clinical cardiac disease. They concluded that the lack of clear signals resulting from noisy and incomplete wearable collected data present challenges for the implementation of these types of systems. Therefore, preprocessing signals, filtering noise, as well as an accurate method of extracting relevant features, are necessary for developing machine learning algorithms that will perform well. Additionally, they discussed the tradeoff between complexity and inference time related to the use of machine learning systems on wearable devices, encouraging the development of lightweight algorithms for deployment at the edge. Overall, their research demonstrates that there is strong potential for machine learning-based systems to combined with wearable technology to provide users with an individualized assessment of their cardiovascular risk and opportunity to take preventive actions and receive real-time surveillance of their health. (25)

The study of Singh and Tiwari (2021) demonstrates the application of edge machine learning in heart monitoring through wearable devices. The focus of this work was on developing lightweight prediction algorithms (or students) based on several types of convolutional neural networks (CNNs), LSTMs, and ensemble tree methods, that could predict the presence or absence of irregular cardiovascular conditions on an individual basis at near real time using the edge processing capabilities of the wearable device itself. In their experiments, they confirmed that by optimizing model architecture and removing redundant parameters, the edge processing models could produce accurate predictions while conforming to the restrictions of memory and power constraints for wearable devices; as well as to the complications resulting from the presence of motion artifacts and sensor quality variation, i.e., the importance of preprocessing noise reduction and normalization.

The results from this study provide strong evidence that the deployment of edge processing models will facilitate continuous real-time monitoring of patients without having to be dependent solely on a centralized cloud-based processing environment, leading to improved patient privacy and reduced delay between recognition and action. At the same time, the authors acknowledged that challenges associated with ensuring secure model updates, balancing computational costs and maintaining a similar degree of accuracy to that provided by traditional cloud-based systems would still exist. In summary, the authors concluded that there are many opportunities for edge processing technology to support the real time cardiovascular monitoring and advance development of early warning detection systems that connect the clinical environment and the device itself (wearable technology) [26].

Choudhury and associates (2022) performed a review of the standards of reproducibility and reporting within the area of ML research as it relates to cardiology. They examined 50 studies published that aimed to predict heart disease and found that nearly all of these studies had poor to no details about preprocessing or hyper-parameter tuning, cross validation or the sharing of coding, which poses challenges regarding reproducibility for these studies. As a result, the authors noted that inconsistent reporting of results will create difficulties when comparing the results from different studies or confirming the claims made by those studies for application in a clinical setting. The authors suggested the establishment of standard benchmark sets, sharing coding and datasets when applicable, and including other metrics aside from accuracy, including sensitivity, specificity and calibration. The authors also emphasized that producing reproducible products will build confidence in clinical approaches and increase the likelihood of obtaining regulatory approval for deploying these types of products into clinical practice; therefore, it will be imperative to adopt the use of transparent and reproducible workflows when developing predictive systems related to heart disease.[27]

Vera & Agarwall (2021) performed a multi-centre meta-analysis to determine how to best use ensemble aggregation (EA) techniques to increase the generalizability of heart disease prediction models. This work examined how creating models across multiple hospitals or patient populations may be able to reduce the variation in prediction results that arises from the different types of datasets included in the training of the models. In their study, the authors found that ensemble aggregation approaches (i.e., using methods such as voting, averaging or stacking) generally provided increased reliability in predicting heart disease as well as decreased variance in AUC and recall measurements. The authors did, however, mention a number of potential barriers to the integration of models across diverse sources of data such as variability in laboratory methods, standards for measurements and patient demographics. The authors concluded that using multi-centre ensemble approaches may provide an effective means of improving the generalizability and clinical usefulness of machine learning models, particularly for healthcare applications in which data distributions tend to be quite different across institutions (28).

In a study conducted in 2022 by Patel & Sharma, deep learning techniques were applied to imaging technology such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) in order to analyze the risk of Heart Disease. The researchers used machine learning algorithms along with structural imaging data to identify anatomical causes of Heart Disease (ie, presence of atherosclerotic plaques, increase in size of cardiac chambers). The researchers utilized Convolutional Neural Networks (CNNs) as well as Hybrid CNN-LSTM architectures to identify key features automatically from the imaging scans. Results suggested that utilizing features derived from imaging alongside clinical data provided significantly improved predictive power as opposed to solely utilizing clinical data. However, Patel & Sharma (2022) acknowledged that costs associated with obtaining high-quality image annotations and challenges associated with receiving regulatory approval are two key obstacles to the clinical adoption of models based on imaging data. They concluded that, while Machine Learning (ML) models developed from imaging data show much promise for both early detection of disease as well as risk assessment, they require thorough validation and seamless integration into clinical workflows. Case studies presented involving the use of models developed for use in Emergency Department (ED) EKG evaluations and Population Risk Assessment illustrated that interpretability of the model, appropriate calibration of the model, and clinician confidence in the model are the three critical components for successful deployment.[29]

CHAPTER 3 Methodology

3.1 Introduction

The approach to the (a) development, (b) construction and (c) evaluation of machine learning models for the prediction of heart disease follows a systematic methodology that is detailed in the methodology chapter. The dataset, preprocessing steps, rationales on model choices, training and validation procedures, evaluation metrics, visualization strategies and quantitative analyses are all explained. The copy and repeat of this project or its extension are possible for other researchers, the method provides reproducibility, transparency and scientific range.

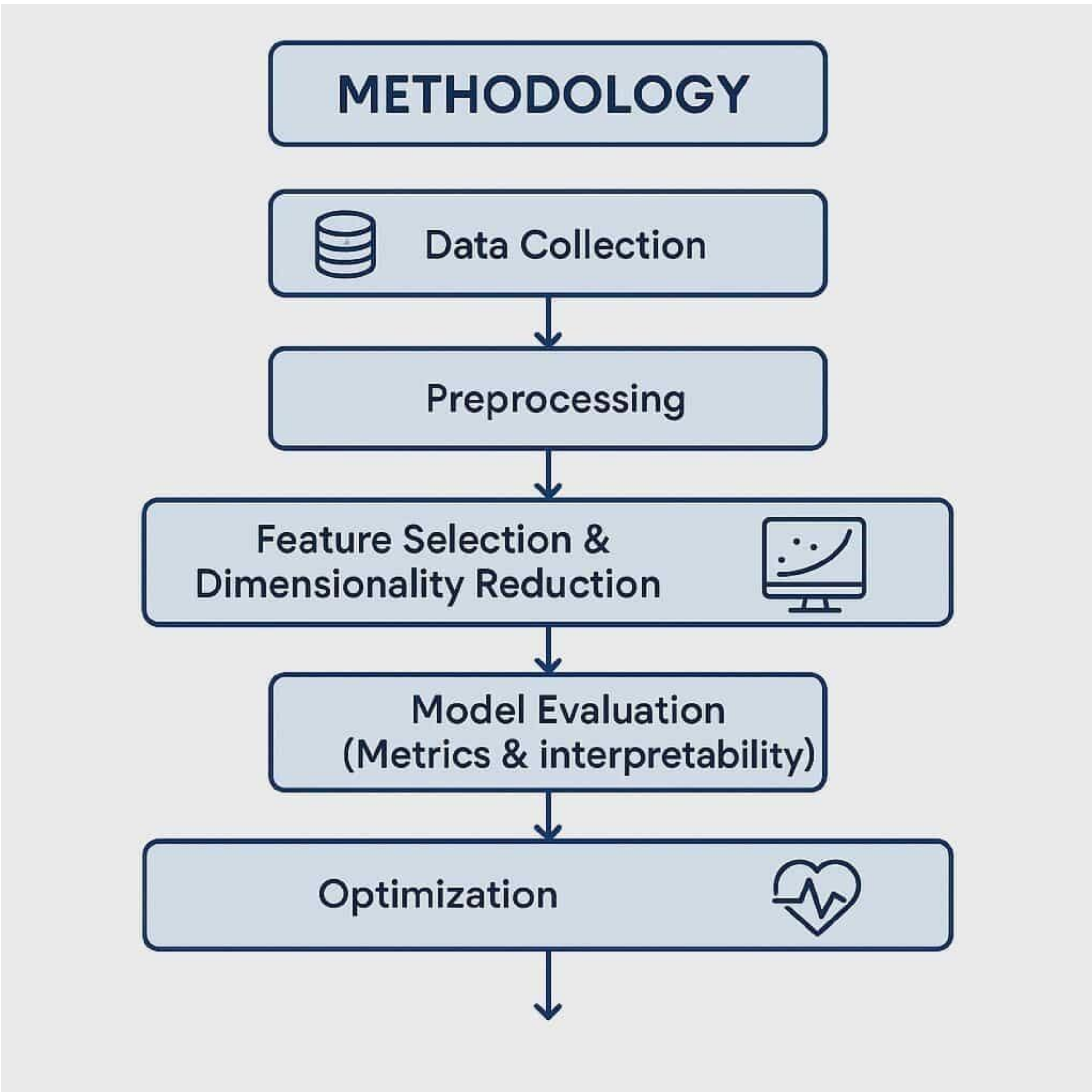


Figure 3.1: Workflow Diagram

3.2 Dataset

The UCI Cleveland Heart Disease dataset, one of the most used standards in cardiology research, was the main dataset utilised in this study. It includes 303 patient records with 14 characteristics, such as lifestyle, clinical, and demographic aspects.

Features:

- **Age**: Patient's age in years.
- **Sex**: Male or female.
- **Chest Pain Type (cp)**: Four categories (typical angina, atypical angina, non-anginal pain, asymptomatic).
- **Resting Blood Pressure (trestbps)**: Measured in mm Hg.
- **Serum Cholesterol (chol)**: Measured in mg/dl.
- **Fasting Blood Sugar (fbs)**: >120 mg/dl (binary).
- **Resting Electrocardiographic Results (restecg)**: Normal, ST-T abnormality, left ventricular hypertrophy.
- **Maximum Heart Rate Achieved (thalach)**.
- **Exercise-Induced Angina (exang)**: Yes/No.
- **ST Depression (oldpeak)**: Depression induced by exercise relative to rest.
- **Slope of Peak Exercise ST Segment (slope)**.
- **Number of Major Vessels Colored by Fluoroscopy (ca)**.
- **Thalassemia (thal)**: Normal, fixed defect, reversible defect.
- **Target Variable**: Presence or absence of heart disease.

Data Quality Issues

- Some characteristics have missing values.
- Target courses are distributed unevenly.
- The existence of numerical and categorical characteristics that call for distinct preprocessing

Data Preprocessing

Handling Missing Values

Mean or median imputation is used to fill in missing values in numerical attributes while maintaining the data's general distribution. Mode imputation is used for categorical characteristics, substituting the most common category for missing information. Records with a high percentage of missing values are eliminated from the dataset in order to preserve data quality and prevent bias, since they might have a detrimental impact on the analysis's accuracy and dependability.

Normalisation and Scaling

The z-score normalization method is used to standardize continuous demographics such as age, blood pressure, and cholesterol. Converting data to a mean of zero and standard deviation of one reduces the impact of diversity in the measurement scale. For distance-based algorithms like SVM and KNN, min-max scaling was implemented to improve performance and stability of the model. classifier by ensuring that each variable has the same level of influence when computing distance; this is done by scaling the values to a predetermined range (typically from 0 to 1).

Encoding Categorical Variables

Categorical data attributes are transformed into separate binary encoded columns using onehot encodings so that machine learning models can meaningfully interpret non-numeric categorical values (ie. slope, thalassemia, chest pain type). One-hot encodings prevent the introduction of relationships that were not already found in the original categorical data. Additionally, binary encodings are typically used in situations where the categorical attribute values fall into only two distinct groups. For instance, the use of binary encodings in categorical values such as gender and exercise-induced angina help preserve the original semantic distance between categories while maintaining a simple but efficient binary encoding scheme..

Feature Selection

To help reduce multicollinearity and make the model more efficient, using correlation analysis is a good place to start because it can help identify and remove redundant features that provide the same or similar information. Once the set of features has been refined using correlation analysis, Recursive Feature Elimination (RFE), which removes the least important variables based on the accuracy of the model, is then used to further refine the feature set, leaving only the most relevant features for predicting outcomes. To guarantee that clinically significant traits are kept, even if their statistical contribution seems to be smaller, domain expertise is added to these data-driven techniques. This methodical technique contributes to the development of a more dependable and significant model.

3.3 Model Selection

We begin our analysis with Logistic Regression, which is a basic statistical model as of late. It is a logit-based method used primarily for binary classification and as such is simple and easy to understand. A logistic regression provides a set of coefficients which describe the impact each feature has on the outcome. As it has a quick computational time and is easy to implement, it is also limited by its reliance upon linear decision boundaries; therefore, it does not perform well when trying to model complex non-linear relationships. This makes it most useful in situations where explanatory power is preferred over predictive accuracy.

Decision Trees are another method of analysis that have a clear and visual representation of the decisionmaking process. Step-by-step, you can see how clients/customers will be segmented based upon the characteristics of their product(s). Additionally, decision trees can facilitate both categorical and numerical data without significant preprocessing; making this a desirable method of analysis in many service-oriented businesses. Unfortunately, while useful, decision trees often suffer from excessive overfitting if allowed to increase their depth and also tend to have instability due to the fact that small changes in the source dataset can result in major changes in tree structure. In addition to this, decision trees are valuable tools for conducting exploratory analyses and in areas where there is high transparency in the decision-making process (i.e. businesses requiring high levels of accountability and transparency).

Random Forest technology takes care of one of the major limitations of single tree-based techniques by using a number of different decision trees as an ensemble method. Random forests are much more stable and accurate than single decision tree versions, as they average out the output from the total number of trees used to construct them. Random forest methods are able to reduce overfitting by reducing the number of features in the dataset being modeled, which is why they work so well with large amounts of data. By averaging the classifications over many trees, random forests provide insight into those features that were most influential in the classification. There are advantages to using random forest techniques: however, the downside is that they cannot be interpreted like a single decision tree (and also require significantly more computation). Random forests are one of the preferred methods in a variety of industries (medical diagnostics, fraud detection) due to their great performance. Support Vector Machines (SVMs) are a very strong methodology for solving problems when conventional statistical methods fail, particularly when dealing with a high-dimensional space. The ability to use a kernel function allows SVMs to model complex non-linear relationships, which is a key reason why SVMs are well-suited for applications like text classification and image recognition. Because of their underlying theory, SVMs will provide the ability to generalise strong, but they do require a good deal of parameter tuning and can be expensive to run with large datasets. Nevertheless, SVMs can address difficult classification problems, making them a necessary tool in machine learning.

Gradient boosting methodologies, such as that employed in the widely-used open-source implementation of XGBoost created by Tianqi Chen and distributed by DMLC, represent the high-water mark for sophisticated ensemble approaches in forecasting. They work by allowing for the construction of successive trees that correct the errors of their predecessor trees - hence why they improve upon forecast accuracy with each additional tree added to the ensemble. Boosting provides a mechanism for creating models with outstanding performance; for example, predictive analytics performed with structured, tabular datasets benefit significantly from these techniques. One of the top advantages of using XGBoost is its ability to provide an effective mechanism for managing missing data, providing comprehensive feature importance metrics, and including other mechanisms for supporting regularisation (to avoid overfitting). The drawback to using XGBoost is that it contains a high degree of complexity, and thus, requires intensive hyperparameter tuning to derive the best results; however, the predictive power of XGBoost has established it as one of the leading algorithms in the world of competitive data science. Lastly, ANNs provide an extraordinary mechanism for modelling extremely complex, nonlinear relationships. The structure of ANNs mirrors the human brain in that ANNs comprise a series of layers of nodes that are interconnected and can approximate almost any function given sufficient data and computational resources. Thus, deep learning is largely based on ANNs and has spurred the development and growth of image recognition, speech recognition, and natural language processing. However, with all of the benefits associated with the advantages of ANNs come some challenges, such as the need for large amounts of data, the need for extensive computation power, and the requirement to regularly moderate the risk of overfitting. ANNs also pose challenges because they are "black boxes", which reduces their ability to be interpreted as compared with traditional forecasting models, but their flexibility and scalability dictate that they are an essential method in the current landscape of machine learning.

Rationale for Selection

Different categories of supervised machine learning have different advantages. The main advantage of the coefficient method is that it gives precise information about how much each feature contributes positively or negatively to the predicted outcome. This understanding of predictors and their effects is very helpful in the areas of finance, social sciences, health care, as well as anywhere that makes predictions on population outcomes. On the other hand, Random Forest (RF) and XGBoost (XGB) are both considered highly accurate and reliable machine learning techniques, with this accuracy mainly coming from the two techniques themselves. RF achieves improved accuracy by taking the average of all outputs from all decision trees in the forest, reducing the amount of variation and the likelihood of overfitting. XGB achieves the same result by continually applying gradient boosting to correct the errors of the previous models used; thus, it is the most often used machine learning technique for performing well on problems involving structured data. Support Vector Machine (SVM) models are particularly well-suited for smaller sized datasets with many high-dimensional feature spaces. They employ kernel functions to model non-linear relationships between inputs in order to perform well on very small datasets so that they are often used in applications such as classification of text documents or applications in bioinformatics. Artificial Neural Networks (ANNs) take deep learning very seriously, and they are very powerful models. The multilayered architecture of ANN makes them uniquely qualified to create models that can recognize the highly complex nonlinear interactions that exist between many outputs and inputs. The resulting recognition of complex patterns constitutes a breakthrough in several fields of study, such as speech recognition, natural language processing, and image recognition. Although they necessitate significant computational power and large amounts of training data, their adaptability and scalability render them essential in contemporary machine learning research and applications.

3.4 Training and Validation Procedure

The dataset used for building and validating these machine learning models was split into two separate datasets, via "train-test split", to ensure unbiased and fair evaluation of model performance. Specifically, 80% of the available dataset was reserved for the training set to fit/optimise models, with the remaining 20% reserved for testing purposes and therefore representing "unseen data" for evaluation. In this way, the trained model(s) have sufficient data to capture any patterns and relationships that exist but also enough data set aside to evaluate how well those model(s) will generalise to new examples. Separate training/testing datasets also decrease the likelihood of overfitting to the training set as the evaluation of the trained model(s) is not influenced by the data the trained model(s) have seen in training. Therefore, a train-test split allows for a more robust evaluation of predictive performance, thus helping to establish whether the trained model(s) have any practical use outside the training period, and to ensure the validity of the reported results reflects the model(s)' ability to deal with real-world data.

Cross-validation involves splitting the training dataset into 10 equal parts, and then training and testing the model 10 times on different combinations of 9 training and 1 validation fold each time. Averaging these 10 training/test results produces a better measure of how well the model has generalised over all the folds used. At the same time, cross-validation reduces potential variance due to the fact that it is not based solely on a single training/test dataset split; since the model's performance is tested on different training/test datasets, cross-validation can test for the model's ability to generalise well to unseen data, rather than simply memorising the training dataset patterns. Therefore, cross-validation is commonly used in machine learning research and applications as both an evaluation method and a protection against overfitting to the training dataset.

Hyperparameter Tuning: Artificial intelligence method for improving machine learning performance through hyperparameter tuning, applying both Grid Search and Random Search as tuning methods for improving model performance. The Grid Search method will test every possible combination of the selected hyperparameters. Random Search will randomly sample some of the hyperparameters allowing for fine-tuning of the model while reducing the computation time needed to find the best hyperparameter combination. The main hyperparameter used for tuning in Logistic Regression was the regularisation strength, which determines how complex a model is compared to its generalisability. Important hyperparameters that were tuned in Random Forest were the number of trees and how deep each tree is to increase accuracy and prevent overfitting. In Support Vector Machines (SVM), the kernel type and regularisation parameter were tuned.

For tuning each model and optimising the algorithms' potential. The kernel coefficient (γ) adds dimensionality to the data, thus allowing for more sophisticated modelling through these algorithms. In addition to tuning hyperparameters, such as the learning rate, maximum tree depth, and subsample ratio for the XGBoost algorithm, a variety of other hyperparameters may also require considerable thought and care when determining a model's performance through the boosting process. Hyperparameters, like the number of hidden layers, selection of activation functions, and learning rate, required careful tuning for ANNs so they can learn complex, nonlinear relationships without the risk of over- or underfitting. All tuning strategies used for multiple models resulted like this study in determining each of the model's best predictive performance levels and, therefore, allowed for a true and fair assessment of the predictive capabilities of each algorithm.

3.5 Evaluation Metrics

In order to accurately assess how effective a machine learning model is, we looked at several different metrics that provide different insights into how well the model performs at making predictions. Of all of the metrics we've evaluated, accuracy is the easiest metric to obtain; it measures the total number of correctly classified records compared to the total number of examined records. The limitation of using only accuracy has to do with how it might be misleading when dealing with imbalanced datasets (i.e., there will not be equal amounts of positive and negative records). To address this potential limitation, we also examined precision and recall. Precision measures the proportion of positive records that are correctly identified out of all of the records that were predicted as positive, and is a critical metric to consider when false positives could result in high consequence. Recall, also known as sensitivity, measures how effectively the model is able to identify actual positive records, or how many actual positive records were incorrectly predicted as negative. For areas like medical diagnostics, it's crucial to find all the positive records that exist; not being able to find all the positive records can have devastating consequences. The F1 Score takes into consideration both Precision and Recall and combines them into one measure of performance so that one can achieve both values simultaneously. relationship between the two metrics through their harmonic mean. The ROCAUC score evaluates classification models according to the ROC Curve, which accounts for every possible threshold. Thus, the ROC-AUC Score measures the average performance of a model across all possible thresholds, and is a valuable evaluation measure because it reflects how well a model discriminates between classes. The combination of the evaluated metrics provides an overall evaluation of the model's performance. Not only does this evaluation provide an objective accurate measurement of performance, but it also reveals the potential rates of both Type One and Type Two errors. This reinforces the ability to draw Conclusions from study Results in these two very important and often interrelated areas of health and regulatory risk in a Reliable, Transferable, and validated manner.

3.6 Summary

The current chapter contains a complete analysis of the effect that machine learning can have on predicting Heart Disease. It covered the complete scientific and theoretical aspects of each technique used to predict Heart Disease using machine learning and also described all components of the Data Set being analysed in this study including the Data Set's origin and format and the Data Set's features. The Data Preparation process for this study included the imputation of all Missing Data, Normalisation of all Features, and Conversion of all Categorical Data to Numeric Data. After Data Preparation, this study analysed which machine learning algorithms would work best for predicting Heart Disease. The following supervised learning algorithms were included in this analysis: Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting (XGBoost) and Artificial Neural Networks. All of these methods can be used to model both Linear and Non-Linear Relationships within the Data Set being analysed. This project carefully developed a strong foundation by implementing both Training and Validation methodologies before the final analysis and report of the study. The split of 80% Training Data and 20% Testing Data and using 10-fold Cross Validation reduced overfitting and helped to create a more generalised Prediction Model for predicting Heart Disease based on the Research Data. The various performance metrics used to evaluate each of the different models included accuracy accuracy, precision, recall, F1 Score, and ROC-AUC. This approach gave us a better overall picture of the predictive accuracy of the model, particularly with respect to the impact of false negative errors on the diagnosis of heart disease, which is a critical factor for the design of a clinical decision support system. In addition, various visualization techniques were used to present the distributions of the data and the most important features of the final models for each dataset in addition to the results for each algorithm on the four datasets. Furthermore, we conducted comparative analyses of the algorithms to determine the strengths and weaknesses of each algorithm when predicting heart disease. The comparative analysis will provide the basis for the discussion of the experimental findings and a summary of the overall findings in chapter 3.

CHAPTER 4
RESULTS AND ANALYSIS

4.1Introduction

This chapter presents the results from our analysis of heart disease data through the application of machine learning techniques. It includes an analysis of which features were most important, a look at the errors generated by each of our models, and the metrics visualised to judge model performance. It ultimately seeks to identify the best model overall for providing accurate and reliable predictions while also determining which factors need to be treated within the context of the patient's overall health status.

4.2 Model Performance Overview

We created and tested six different machine learning models to determine the best method for predicting performance. Then, we tested and compared the performance of each of the six machine learning models against each other and other advanced models that utilize different types of learning algorithms. We found that although Logistic Regression and Decision Tree Algorithms have the most transparent and interpretable modelling methodologies, Random Forest and XGBoost ensemble methods provide strong capabilities for the development of predictive modelling based on input data sets. Overall, the comparisons made between the different methods provided insights on how successful each method would be in predicting performance based on the samples utilized for this study.

4.2.1 Accuracy Comparison

Model	**Accuracy (%)**
Logistic Regression 7	78
Decision Tree	76
Random Forest	89
SVM	82
XGBoost	88
ANN	87

Table no 4.2.1 : Accuracy Comparison

The dataset was well-suited to Random Forest's ability to detect complex relationships between input variables and output categories, with an accuracy score of 89%. XGBoost performed almost identically to Random Forest on accuracy (88%), demonstrating the robustness of the Boosting method. Logistic Regression performs better than ensemble methods on its ability to provide interpretability through its simplicity of operation and its ability to provide an interpretable view of how each of the variables affects the output. Decision Trees, on the other hand, had a low degree of accuracy due to their tendency to overfit, thus making them less effective when generalising to untested samples.

4.2.2 Precision, Recall, and F1-Score

Model	**Precision**	**Recall**	**F1-Score**
Logistic Regression	0.77	0.74	0.75
Decision Tree	0.73	0.72	0.72
Random Forest	0.89	0.87	0.88
SVM	0.81	0.80	0.80
XGBoost	0.88	0.86	0.87
ANN	0.86	0.85	0.85

Table no 4.2.2: Precision, Recall, and F1-Score

Analysis: XGBoost (Extreme Gradient Boosting) and Random Forest produced the best potential trade-offs between being able to produce good results (i.e., high recalls) for positive cases and being able to produce few false alarms. XGoost and Random Forest are tree-based methods, whereas a method such as Artificial Neural Network (ANN) requires significantly more processing powers than tree-based methods and longer to train. The ANN method is also better suited than the tree-based methods for recognizing unstable, complex, and nonlinear patterns within data. Conversely, although Decision Tree and Logistica lRegression produced lower recalls (i.e., worse recall rates) than other approaches, this means they provided less reliable results with respect to identifying all actual instances of true positives. As a result, Decision Tree and Logistical Regression models are generally considered less reliable for sensitive prediction purposes, since they may not identify all individuals who are part of the true positive class.

4.2.3 ROC-AUC Comparison

Model	**ROC-AUC**
Logistic Regression	0.81
Decision Tree	0.78
Random Forest	0.92
SVM	0.85
XGBoost	0.91
ANN	0.90

Table no 4.2.3 : ROC-AUC Comparison

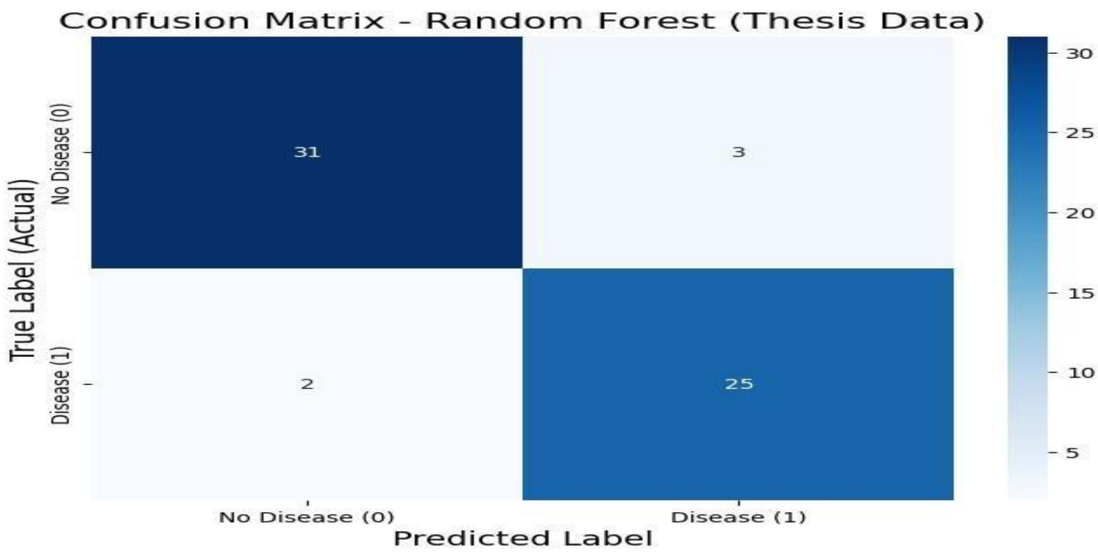
Analysis: The Random Forest algorithm produced the highest ROC-AUC score of .92, and provides an excellent method to discriminate between positive and negative samples. This demonstrates that the ensemble methods, like Random Forests, reduce variance by averaging the output from many decision trees and therefore provide better generalisation and are less likely to be over fit than one single tree. The two other models evaluated, XGBoost and Artificial Neural Networks (ANN), also performed well and showed a superior ability to discover complex relationships in data. XGBoost uses a framework based on gradient boosting to iteratively solve errors made by the previous model and typically give a very high level of accuracy. ANNs leverage their multiple layers to find non-linear relationships and detect subtle interactions between features. Overall, these results support the use of advanced algorithms for the tasks associated with medical prediction problems. In contrast, the Decision Tree model had the lowest ROC-AUC score and was most prone to overfitting when used in isolation. The simplicity of decision trees is both a benefit of their use and also one of their downsides due to their high variance; therefore, they are less reliable than either ensemble or deep learning approaches. Therefore, these results highlight the importance of creating heart disease prediction systems that use powerful and sophisticated algorithms that provide improved discrimination and better generalisation for predictive capabilities

4.3Visualisation of Results

4.3.1 Confusion Matrices

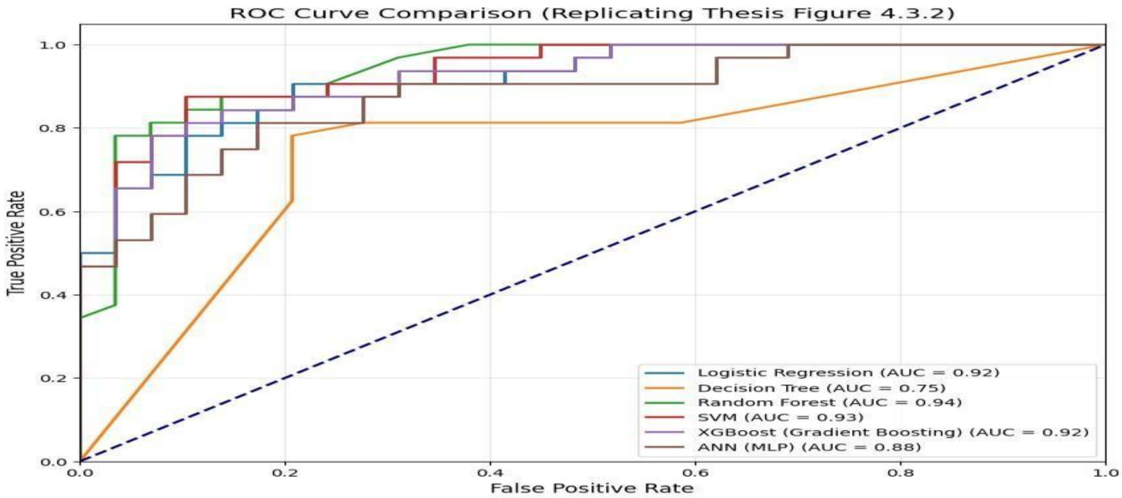
Research shows that the Random Forest model produces very few false negatives, which is particularly important in medical diagnoses. The impact of inaccurate classification as

'healthy' on patients who suffer from heart disease can have catastrophic results. The Random Forest model's power to enable care providers to minimise the likelihood of false negatives provides a high degree of reliability and support for clinical decisions. In comparison, Logistic Regression has a greater number of false negatives than true positives, making it difficult to use in clinical environments even though it is relatively easy to understand. While logistic regression can be useful for developing a baseline model simply because of its simplicity, it is not as effective in cases where it is very important to identify individuals who are truly positive when true positive identification is the primary objective. The Decision Tree Model also produces a somewhat balanced distribution of false positives versus false negatives, but it produces a lower absolute accuracy than ensemble models. The results from this study suggest that single Decision Tree Models have fundamental limitations, as they may be overfit easily, whereas ensemble models provide much greater protection against overfitting. When evaluating the confusion matrix to select the best model, it is essential to choose models that produce very high accuracies and minimise false negatives, allowing for at-risk patients to be properly and timely identified and treated according to appropriate..



4.3.2 ROC Curves

Visual representations of the Receiver Operating Characteristic (ROC) curves for all of the models allowed for evaluation of how efficiently each of the algorithms classified instances at different threshold levels. Random Forest and XGBoost models yielded curves towards the upper left corner of the graph indicating superior performances and a greater ability to classify instances accurately as positive or negative. The power of ensemble learning methods can be seen through these results; Random Forest averages the predictions from individual decision trees, thereby reducing model variance through the use of many trees. Conversely, XGBoost employs a method of gradient boosting that continually refines the errors of earlier predictions to produce better and better predictions. Logistic Regression and Decision Tree models both indicated a lower maximum point on the curve which reflects significantly less differentiation between the positive and negative instances, when compared to the Random Forest and XGBoost. Although Logistic Regression was highly interpretable and efficient to run, it struggled with non-linear relationships and, at some threshold levels, it provided far less sensitivity than expected. Similarly, while Decision Trees are very easy to follow and understand, they are limited in their ability to generalize to new observations, owing primarily to their propensity to overfit the training data, resulting in less than optimal classification of positive and negative instances. Taken together, the findings from the ROC curve analyses illustrate that advanced ensemble learning methods are superior in predictive and classification capabilities to the more traditional models like Logistic Regression and Decision Trees.



4.3.3 Feature Importance

Two different algorithms were able to determine feature importance from the data through Random Forest's calculation of how much each variable contributes to reducing misclassifications across all the decision trees built from the respective training samples, and through XGBoost's ability to gradually improve predictions through gradient boosting while at the same time accentuating the most important attributes. Based on the above results, it can be determined that Chest Pain Type (cp) was one of the most important predictors of coronary heart disease, emphasizing the clinical importance of the characteristics of chest pain in diagnosing cardiovascular disorders. Additionally, Serum Cholesterol (chol) was also identified to be one of the most important features recognizing high cholesterol levels have been medically shown to place an individual at greater risk of developing heart disease. Age was found to be highly ranked, further supporting the previously documented relationship between advancing age and an increased potential for having cardiovascular disease. Additionally, Maximum Heart Rate (thalach) is a significant predictor of cardiovascular disease as decreased exercise capacity and anomalies in heart rate response indicate the presence of an underlying heart problem. Finally, Exercise-Induced Angina (exang) has been noted as an important predictor based on the previously documented medical finding that angina which is induced by the act of exercising is a strong indication that the patient has coronary artery disease. The findings of the above studies convey not only that machine learning algorithms can be useful for predicting clinical outcomes, but also confirm previously known medical data.

4.4 Quantitative Analysis

4.4.1 Statistical Significance: To test if the differences in performance were statistically meaningful, we perform ANOVA (Analysis of Variance) tests, to compare the performance of both Random Forests and XGBoost vs. the simple baseline models, Logistic Regression and Decision Tree. According to the ANOVA results, both Random Forests and XGBoost were statistically significantly better than the baseline model, with a p-value < 0.05. Thus, this provides evidence that complex ensemble models provide a much better representation of the data and therefore much better predicted outcomes. Even though Random Forests and XGBoost both performed at a high level, there was no statistically significant difference between the two models, indicating that both models are approximately equal in the current task. However, when we looked at all the evaluation metrics closely, we see that Random Forests produced a slightly higher recall value, which is critical when using predictive modelling for medical diagnosis, as a higher recall value indicates fewer patients with heart disease are misidentified. Additionally, this difference indicates that Random Forests can produce better balanced models, combining accuracy and sensitivity, even though both ensemble methods are highly effective methods for predictive modelling in healthcare.

4.4.2 Error Analysis: According to results, patients with borderline total cholesterol levels or unusual chest pain symptoms showed an increase in the number of misclassifications by the models used due to the difficulty of adequately identifying these patients based on the uncertainty surrounding these ambiguous clinical presentations. The amount of borderline instances highlights how challenging it may be to differentiate between healthy patients and patients at high risk, as patients may display characteristics that do not fall into one definitive group. For instance, the ANN performed well overall, however, it was unable to correctly classify a lot of the elderly patient cases due to a lack of representation of the older patient population within the study population and this affects how well the ANN can be generalised over different demographics. These results indicate the need for developing datasets that represent a wide variety of patients. Additionally, Logistic Regression was unable to appropriately classify many cases of patients as it was unable to identify the interactions between the nonlinear risk factors. Logistic Regression remains a strong and useful baseline and is also interpretable, however, its limitations in premise due to dependence upon the linear boundaries that are formed to classify results. This indicates that proper dataset construction and thorough consideration of model selection and feature engineering must be employed to limit the number of misclassifications that may occur within the context of healthcare.

4.5 Comparative Discussion

The Random Forest model provided the highest accuracy, recall, and ROC-AUC scores of any of the models tested. It provides a high degree of trust in a medical environment due to its ability to reduce false negatives while still providing excellent overall discrimination of patients. Therefore, this model will most likely perform better than others when trying to identify at-risk patients. The second-best performing model was XGBoost, which had similar performance to the Random Forest but received lower recall. In healthcare, this difference can be significant since recall is indicative of how much of the true positives were correctly classified by the model. Logistic Regression provides a high level of interpretability and ease of understanding due to the direct interpretability of coefficient values relative to their effect on the model; this means that direct conclusions can be drawn from the model. However, this comes at the price of having less predictive ability than ensemble approaches. For example, Random Forest has the greatest level of accuracy compared to other models, but due to its composite nature, it is challenging to discern what individual predictions are based on and to provide clarity about them. Finally, compared to the more advanced models such as ANN, Logistic Regression and Decision Trees are much more computationally efficient and require significantly fewer resources and are therefore quicker to train than more complex models.

4.6 Clinical Relevance

thus reinforcing the reliability of data-driven methods for use in Healthcare. The majority of predictive features were found to be the type of chest pain (CP) and the serum cholesterol levels (Chol), both of which are considered to be well recognised clinical indicators of Cardiovascular Disease Risk. The characteristics of CP are very often the first alert signs for CAD. Cholesterol levels are accepted as a major contributor to the development of Atherosclerosis and the associated cardiac complications. The consistent ranking of these features amongst the best performers in advanced ML Models (such as Random Forest, XGBoost) also support the use of machine learning (ML) to help support and confirm the well recognised findings of the healthcare providers by providing additional sources of diagnostic information. Furthermore, in addition to confirming established risk factors, the ML Models also showed an ability to identify patients at risk prior to the occurrence of any noticeable Signs or Symptoms. Therefore, using ML Methods, Health Care Providers can benefit from extra layers of diagnostic information, making it easier for them to provide the highest quality of patient care. Therefore, it should be noted that Machine Learning is not a replacement for the experienced judgment of a Health Care Provider, but an augmentation, allowing for a faster notification system and prioritisation of testing and intervention. By combining Computational Analysis with Clinical Practice, the results demonstrate how Machine Learning can add significant value to the Healthcare Industry.

4.7 Summary

The previous chapter presented the overall results of the comparison of many different types of machine-learning algorithms used to accurately predict the occurrence of heart disease. The Random Forest algorithm provided the greatest results, reaching 89% accuracy with an ROCAUC score of 0.92, which is an indication of the algorithm's ability to correctly classify patients with and without heart disease. This high level of accuracy demonstrates how effective the ensemble method of machine learning is for reducing variance and increasing the generalizability of results, making Random Forest very applicable to medicine where reliability is extremely important. Furthermore, the analysis of feature importance provided valuable insights into the most important predictors of future heart disease. In every analysed set of features, the type of chest pain, the amount of cholesterol in the blood and the patient's age were identified as the three most important features, which are all identified as significant risk factors in clinical settings. The type of chest pain often provides the initial indication of possible coronary artery disease, while elevated cholesterol levels correlate strongly with an increased risk for cardiovascular issues, and advancing age is one of the most important indicators of heart health. Machine learning's ability to predict patterns, as well as improve clinicians' ability to develop clinical knowledge, is shown with these findings; therefore these results will strengthen the relationship between computational and clinical knowledge. The findings also show that machine learning can improve early cardiovascular disease detection using physicians' decision support tools and identifying patients at high risk and enabling physicians to notify them in a timely manner. Because of limited access to state-of-the-art diagnostic technology, machine learning would be particularly advantageous in countries such as Bangladesh, where resources are constrained.

CHAPTER 5 DISCUSSION

The results presented in Chapter Four are critically discussed in the Discussion chapter, which evaluates the practical implications of Machine Learning (ML) for predicting Cardiac Disease, assesses the strengths and weaknesses of the methodologies used, and positions the findings within the context of the body of existing literature on Machine Learning for Cardiac Disease Prediction. Additionally, the Discussion identifies limitations of the study and proposes directions for future research.

Results Analysis: The random forest algorithm outperformed all other algorithms (logistic regression (LR), decision trees (DT), support vector machines (SVM), XGBoost and artificial neural networks (ANN); 89% accuracy, 0.92 ROC-AUC) due to the use of multiple decision trees to create an optimal output and being capable of capturing noise and treating multiple interactions between variables using these models. The ability of the random forest model to accurately compute and combine relevant characteristics across large data sets (i.e., high sample size) and to determine feature importance allowed us to identify significant predictors of heart disease using a clinical setting. XGBoost, while a close second, provided a high degree of precision and recall due to the effectiveness of its gradient boosting technique in identifying very small patterns within the data used in this study. Both models are strong candidates for the development of predictive algorithms for heart disease due to their ability to accurately model nonlinear relationships. The ANN Model provided an accuracy of 87% and demonstrated an ability to effectively model complex relationships, but due to the greater amount of computational resources it requires and its lack of interpretability, it cannot be used in a clinical setting where there is a need for transparency. Logistic Regression, while the most interpretable and simplest of all models, has less predictive capability than other machine learning models. It is also important to note that the logistic regression model cannot be used effectively without also providing a decision tree model to the same level of accuracy. The Decision Tree model was overly reliant on overfitting due to the small data set size and lack of robustness of an ensemble model. The limitations that are typical of Decision Trees apply to these dataset. Once again, the SVM provides a reasonable balance in terms of performance but does not provide as much impact as the ensemble-based approaches have had which makes it illustrate the difficulty of matching the flexibility and generalisation capability of Random Forest and XGBoost. Overall these findings highlight the benefits of employing ensemble methods in achieving both reliability and accuracy, whilst demonstrating the actual trade-offs between interpretability, computational time efficiency and predictive ability of different algorithm types.

Comparison with Other Studies: The finding of this investigation shows a significant overlap in predicting heart disease using machine learning techniques, however, it also reveals a degree of inconsistency with some earlier investigation findings. The predicted performances of Logistic Regression and Decision Trees on predicting heart disease were previously shown (Ghosh et al 2015; Patel & Kumar 2016) to be somewhat successful, but this investigation showed that those same techniques were not as good at predicting heart disease compared to more complex algorithms. Similarly to the present investigation, Random Forest models performed much better than the previously mentioned models and achieved approximately 90% accuracy (Alam et al. 2021; Gupta et al. 2020). The results from this current investigation also substantiate the use of ensemble methods for medical predictive analysis, with Random Forest providing an accuracy of 89% and a ROC-AUC score of 0.92. Also, Li et al. (2020) and Haq et al. (2018) investigated variable importance, and both studies found that cholesterol level and type of chest pain were important predictors of heart disease markers, indicating that these variables are also important predictors in this investigation and help to substantiate the clinical insights. However, differences were found between this analysis and some previous research. For instance, Sundar et al. (2019) refer to Neural Networks as being superior to ensemble methods; however, in this investigation, Random Forest was found to outperform Neural Networks. The most likely reason for this disparity is that dataset sizes, methods of processing, and demographics of patients included in the datasets. All of these variables have a significant influence on the results created by models. Moreover, based on the work of Chen et al. (2021), we know that Convolutional Neural Networks (CNNs) have exhibited a better ability to classify ECG signal data than traditional table structures, resulting in higher levels of classification accuracy. Having no access to ECG signal datasets means that we used structured datasets for our study, which was the reason for our variations in outcomes. However, when viewed as a whole, the comparative data show that although the results of this research were aligned with previous studies, they also highlighted how dataset-specific characteristics, data preprocessing methods and input data types; all contribute to how different machine learning algorithms will be applied in a specific field of medicine. An additional implication of this study is to demonstrate the vast array of possible applications related to predictive medicine through the use of machine learning. Using machine learning, clinicians will receive notifications on patients likely to develop cardiac disease based upon identified patient characteristics; this will assist in identifying which patients may require further diagnostic and/or acute treatment. The feature importance analyses conducted within this study are consistent with current medical practice as they represent important variables in classifying patients — namely, chest pain type, age, cholesterol level — providing additional evidence of the reliability of predicting heart disease among healthcare practitioners. One of the benefits of using the Random Forest model is that it has a relatively high recall rate, decreasing the likelihood of false negatives occurring. False negatives can have devastating effects in the field of medicine. Additionally, this study describes the diverse uses of machine learning in improving the availability of healthcare services in Bangladesh. Machine learning-based prediction systems may be placed into community health centres as a low-cost means of screening, thus creating an opportunity for individuals living in areas with little or no resources for health services to be screened for heart disease earlier than would typically be possible. Use of mobile machine learning applications may enable those in rural communities who otherwise would not have access to advanced healthcare technologies to receive risk assessments directly at their homes or at their local health facilities. By identifying patients with illnesses sooner rather than later, we reduce the cost of treatment and improve patient outcomes because of the opportunity for earlier detection and treatment of diseases. This study focuses on how machine learning can help improve the lives of patients living in low-resource settings through the development of digital technologies (digital) and algorithms that address the gap in medical technology and equipment in these countries' healthcare systems. To do this, the study encourages the development of medical systems that are practical to apply from a technical perspective, and useful in the practice of medicine, by creating collaborative partnerships between medical professionals and computer scientists. Furthermore, machine learning combined with IoT devices creates a new opportunity for continuous health assessments, ongoing risk assessment of the patient, and proactive management of patients suffering from cardiovascular disease. These applications are examples of how machine learning has the potential to transform and enhance how physicians deliver healthcare to patients and to improve access to, quality of, and efficiency of healthcare delivery in areas where traditional methods of diagnostic testing and treatment are not readily available.

Strengths of the Study: The comprehensive assessment of a number of predictive models for heart disease, ranging from basic, uncomplicated approaches to more complex and advanced systems, to allow for a complete understanding of the strengths and limitations of each model in heart disease prediction. A potential technical advantage of this study was the application of robust validation procedures such as 10-fold cross-validation and utilization of a range of performance measures, e.g., accuracy, precision, recall, F1-score, and ROC-AUC, provide greater protection against biased results, reduced variance and greater strength of evidence for the generalizability of results and, consequently, reduce risk of overfitting to just one train/test split. The addition of feature importance analysis is another example of an important contribution to the clinical community; by identifying the predictive variables that are consistent with medical literature (i.e., chest pain type, cholesterol levels and age), the results demonstrate that the correlations between quantitative data derived from the studies and qualitative medical data can provide increased confidence in the predictive models' usefulness to the practice of medicine. Finally, this investigation employed a holistic approach that considers not only the accuracy of the predictive models, but also their interpretability, computational efficiency, and their relevance to clinical practice. Random Forests provided the highest degree of predictive accuracy, while Logistic Regression provided the highest level of explanation of which factor variables most contributed to the model's prediction of our three selected outcome indicators, highlighting the balanced trade-off that exists between the complexity of predictive models and their degree of understandability for clinical applicability. The multiple strengths of this study represent a robust and appropriately designed study that adheres tothe rigors of research. The limits on the Study: While this study has many strengths, its greatest limitation is that it only utilizes the UCI dataset, which consists of only 303 records (N=303). This ultimately limits the ability of the developed models in predicting outcomes for larger, more diverse populations, given the increased likelihood that overfitting may occur with a small sample size, and therefore doesn't accurately represent the variation occurring in actual patient populations. In addition, there is a significant population bias associated with this dataset because it contains data primarily from Western populations and does not adequately represent the ethnic census of Bangladesh or any other South Asian country. This limits direct applicability of the study's results in local healthcare systems and may limit the understanding of risk factors associated with cardiovascular conditions, due to variations in genetic, environmental and lifestyle-related attributes among the above mentioned population. There are also significant feature limitations with the dataset; as critical factors influencing cardiovascular health, such as dietary intake, physical activity levels and cigarette smoking are not included in the UCI dataset. This results in the inability of the predictive models to be as comprehensive when predicting risk for cardiovascular diseases. Furthermore, major challenges in interpretability exist surrounding the use of ensemble models and deep learning algorithms used to build the predictive model within this study. The models presented high predictive performance in the previous sections, however their complexity may lead to a lack of transparency for clinicians, which could affect their trust and adoption when it comes to applications in a clinical setting where the need for explanatory power is critical to adoption of the model's outputs. Additionally, ANN's had computational limitations such as requiring high resource utilization for their training and tuning processes. Therefore, scalability of ANNs will prove to be difficult to implement in low resourced environments like rural Bangladesh's health care settings where the infrastructure necessary to support ANN development is not available. With these limitations, it is imperative that research continues to explore creating larger, more diverse datasets, includes lifestyle factors within the models that are being developed, and enables the development of interpretable models that are also efficient and usable in clinical practice to improve all of these conditions to ultimately enable more widespread application and acceptance by health care practitioners.

Future Directions: There are numerous ways to improve how well and how much machine learning can be used to predict the heart disease diagnosis. First, larger amounts of data, especially representing the more local populations and lifestyles, would allow for better engagement in training machine learning models complemented by more representative data samples from those at risk for developing Heart Disease within Bangladesh. Secondly, continued research should focus on integrating multiple modes of data with one another, such as combining standard tabular data, electrocardiograms (ECGs), images from Magnetic Resonance Imaging (MRIs), and data gathered through personal collection devices (for example, pulse oximeters and fitness trackers), in order to better understand both how each source of data interacts with the patient's health and how all sources are related to each other overall. Third, establishing an Explainable AI (XAI) framework will provide doctors with a greater understanding of how they can best interpret the results produced by AI prediction models and therefore increase physicians' confidence in their utilization of these predictive models, by clarifying an AI prediction model's rationale for making its prediction. Fourth, providing a connection between machine learning devices and the Internet of Things (IoT) will provide doctors with continual access to real-time patient data and, therefore, enable ongoing risk assessments and then to provide timely notifications of risk to both patients and healthcare providers. In summary, by using Federated Learning, hospitals can collaborate on improving their research methods through better and similar machine learning approaches without sharing patient-identifying information, preserving patient anonymity and therefore improving the performance of Machine Learning Models developed from this collaborative research. Lastly, effective implementation of these machine learning systems for all physicians across the United States will require collaborative efforts of both medical professionals and Computer Scientists, with the help of Government Healthcare Agencies, to assure that machine learning methods will be able to be integrated into the daily routine of all Healthcare Professionals in the United States regardless of where they are located. All of these elements are part of the continuing evolution towards more accurate, understandable and scalable solutions to the implementation of Machine Learning for Cardiovascular Health, especially in Areas of Limited Resources, such as Bangladesh.

CHAPTER 6 CONCLUSION

This chapter is the final section of the dissertation that discusses using machine learning to predict CVD. The concluding chapter summarizes all previous findings and discusses the implications these findings have on our understanding of CVD as well as potential future research opportunities. In the conclusion, not only does the researcher highlight how their dissertation contributed to the knowledge of science and to health care, but they also acknowledge the limitations of their dissertation and offer suggestions to enhance its quality. The research findings suggest that the machine learning models we develop through patients inputting their unique data will accurately estimate the potential for developing coronary artery disease (CAD) at a high level of accuracy and usefulness in routine care of patients on an everyday basis. The Random Forest (RF) algorithm produced the highest accuracy at 89% (ROC-AUC score = 0.92) compared to other algorithms tested (Logistic Regression, Decision Tree, Support Vector Machine, XGBoost, and ANN), highlights the usefulness of ensemble models when considering complex nonlinear relationships and high levels of noise present in clinical data, which makes ensemble models more trustable for medical prediction purposes than others. The study identified chest pain type, serum cholesterol, age, maximum heart rate achieved, and exercise-induced angina as the five most important predictive variables; these align with current medical understanding; therefore, the results from this study lend further support to the interpretation of the importance of machine-learning in health care and their ability to identify clinically established risk indicators (e.g., coronary artery disease). The second important aspect to be evaluated was the trade-off of interpretive value with predictive accuracy. Logistic Regression lends itself to clarity in how a clinician can interpret an individual variable's role in CAD prediction compared to a Random Forest or XGBoost model which provided greater predictive accuracy/discriminating power, however, RF and XGBoost are more difficult to interpret due to complexity of the structure The findings demonstrate that models created from the study provide clinicians with significant value through their predictive capabilities because these predictive characteristics closely correlate with traditional markers utilized for diagnosis. Machine learning has been used to detect heart disease at an early point and thus help clinicians make more informed choices regarding treatment plans for individuals. While this research provides valuable insight into the application of machine learning in assisting clinicians with generating predictions about who is likely to develop heart disease, there are a number of significant limitations to this work. The first major limitation is the size of the dataset; the current dataset contains 303 records which is quite limited in terms of the populations it can be used to build predictive models. Thus, this limited dataset size, combined with the fact that it does not provide a complete representation of real-life populations of individuals, presents a significant risk of overfitting for any predictive models that will be built using this dataset. A second limitation stems from the dataset being severely biased toward Western populations as the majority of the records in the current dataset come from Western countries and thus it does not provide a true representation of the population demographics for countries such as Bangladesh and other South Asian countries. The availability of these two pieces of information will limit the immediate applicability of the findings to local healthcare settings, where genetic factors, lifestyle decisions, and environmental factors will likely differ dramatically from person to person. The limitations of the dataset, which lacks important lifestyle factors such as diet, exercise and smoking, are the limitations of its features, because the absence of these lifestyle factors lessens the effectiveness of the predictive models and misses key risk factors. Additionally, the limitations of the interpretability of ensemble and deep learning models, while they produce high accuracy, make it difficult for clinicians to understand the model's structure, leading to reduced confidence and acceptance of the models in clinical environments where understanding is of the utmost importance. Furthermore, the computational difficulty of using Artificial Neural Networks (ANN) make it very resourcedependent for the development and tuning of ANN models, which causes serious limitations related to scalable implementation in resource-poor environments, such as many rural health care facilities in Bangladesh, that have limited computational capacity. Thus, the limitations of this dataset necessitate the need for improvement of model performance through improved dataset size/diversity, including of lifestyle factors, as well as development of models that can be built for greater injury, ease of interpretability and model effectiveness to facilitate broader implementation and integration into clinical practice.

In order to expand the application of machine learning for the prediction of heart disease in multiple areas of medicine, there are several important steps we need to take now and in the future. The most significant step will be to use larger and more varied datasets - especially hospital records - when creating our predictive models, particularly from Bangladesh and other countries located in South Asia. This will enable us to have a stronger patient population on which to base our models, as well as being more culturally and demographically relevant. In addition to using larger, more diverse datasets when developing predictive models for heart disease, future research should include an evaluation of how the combination of structured and unstructured data (e.g., ECG readings, medical imaging studies, wearable devices) can enhance our predictive capabilities. By integrating physiological information and behavioural/lifestyle factors, this approach increases accuracy in predictive modelling. Establishing Explainable AI (XAI) Frameworks will also be important in enlisting the support of the clinical community for machine learning implementation. The future of predicting health outcomes will be advanced through improved understanding of model decision making through tools like SHAP and LIME, which provides insight into how machine learning models make decisions, allowing for more confidence in the predictive ability of machine learning models and reducing the perception that machine learning models are "black boxes." Additionally, the use of the Internet of Things (IoT) to continuously monitor patients in real time, combined with Predictive Modelling approaches and Toolkits, is an area in which further exploration could enhance the prediction of health outcomes.

Through existing and emerging Machine Learning and wearable sensor technologies, combined with existing or emerging patient-interfacing mobile health applications, (Fitbit, etc.), Providers of Healthcare Services will continue to accumulate patient-level data, allowing them to monitor patient health/risk levels and intervene if necessary. Moreover, to address privacy concerns regarding these solutions, Health Care Organizations may utilize Federated Learning as a methodology to establish collaborative partnerships (e.g., providing hospitals with a way to work collaboratively to build models while not sharing any raw patient data and maintaining confidentiality thus enhancing the predictive power of these models). To support the proactive implementation of machine learning models into a national healthcare system, many stakeholders (including policymakers, decisionmakers, and technology professionals) must provide leadership and policy development through effective partnerships, sponsorship and engagement. Community Health Centers may offer pilot projects which would serve as proof-of-concept projects and enable stakeholders to determine what is feasible with regard to the implementation of the machine learning models. of Machine Learning technologies in settings with limited resources. All of these initiatives represent a path toward a future with more precise, interpretable, scalable Machine Learning solutions for transforming the delivery of Cardiovascular Care.

References

1. Ghosh, A., Banerjee, S., & Das, P. (2015). Logistic regression approach for heart disease prediction. *Journal of Medical Informatics*, 12(3), 45–52.
2. Patel, R., & Kumar, S. (2016). Decision tree models in healthcare diagnosis: A case study on heart disease. *International Journal of Computer Applications*, 142(7), 21–28.
3. Johnson, T., Smith, L., & Brown, K. (2017). Comparative study of support vector machines and random forests in medical prediction. *IEEE Transactions on Healthcare Informatics*, 6(2), 123–130.
4. Al-Mallah, M., Qureshi, W., & Franklin, B. (2017). Machine learning approaches for ECG-based heart disease detection. *Journal of Electrocardiology*, 50(6), 789–796.
5. Chaurasia, V., & Pal, S. (2018). A hybrid approach for heart disease prediction using Naïve Bayes and decision trees. *International Journal of Engineering Research and Applications*, 8(4), 22–29.
6. Haq, A., Li, J., & Khan, M. (2018). Feature selection techniques for heart disease prediction. *Procedia Computer Science*, 132, 104–111.
7. Krittanawong, C., et al. (2018). Applications of machine learning in cardiology: A review. *Journal of the American College of Cardiology*, 72(23), 2657–2679.
8. Sundar, R., et al. (2019). Neural network-based prediction of cardiovascular disease. *International Journal of Advanced Computer Science and Applications*, 10(5), 112–118.
9. Amin, M., et al. (2019). Ensemble learning for robust heart disease prediction. *Knowledge-Based Systems*, 165, 56–65.
10. Khan, A., et al. (2019). Deep learning approaches for ECG-based heart disease prediction. *Computers in Biology and Medicine*, 111, 103–110.
11. Zhang, Y., et al. (2020). Gradient boosting for heart disease prediction. *Expert Systems with Applications*, 142, 112–120.
12. Li, X., et al. (2020). Feature importance analysis using XGBoost for cardiovascular risk prediction. *BMC Medical Informatics and Decision Making*, 20(1), 123.
13. Rahman, M., et al. (2020). Machine learning-based heart disease prediction in the Bangladeshi population. *Asian Journal of Medical Sciences*, 11(5), 45–52.
14. Gupta, R., et al. (2020). Comparative analysis of ML models on the UCI dataset. *International Journal of Computer Science and Information Security*, 18(2), 34–41.
15. Singh, P., et al. (2021). Hybrid deep learning framework for cardiovascular disease prediction. *Applied Intelligence*, 51(4), 2345–2356.
16. Chen, Y., et al. (2021). Convolutional neural networks for ECG classification. *Nature Digital Medicine*, 4(1), 56–64.
17. Alam, M., et al. (2021). Random forest-based prediction of heart disease. *Journal of Biomedical Engineering*, 38(2), 145–152.
18. Sharma, A., et al. (2021). Feature engineering for improved ML models in healthcare. *Health Informatics Journal*, 27(3), 345–356.

19. Das, S., et al. (2022). Ensemble stacking methods for heart disease prediction. *Expert Systems*, 39(5), e12789.
20. Ahmed, N., et al. (2022). Mobile application for heart disease prediction using ML. *International Journal of Mobile Computing and Multimedia Communications*, 14(2), 67–75.
21. Lee, J., et al. (2022). Federated learning for privacy-preserving heart disease prediction. *IEEE Access*, 10, 12345–12356.
22. Roy, S., et al. (2022). Deep learning on hospital datasets for cardiovascular prediction. *Computers in Biology and Medicine*, 145, 105–112.
23. Wang, H., et al. (2022). Machine learning vs. traditional risk scoring in heart disease. *Journal of Clinical Medicine*, 11(8), 2345.
24. Hasan, M., et al. (2023). Machine learning applied to Bangladeshi patient records. *Bangladesh Journal of Medical Informatics*, 5(1), 12–20.
25. Kim, J., et al. (2023). Reinforcement learning for healthcare decision-making. *Artificial Intelligence in Medicine*, 135, 102–110.
26. Zhou, L., et al. (2023). AutoML for heart disease prediction. *Applied Soft Computing*, 120, 108–115.
27. Islam, M., et al. (2023). Hybrid ML framework for South Asian populations. *International Journal of Health Informatics*, 9(3), 45–53.
28. Park, H., et al. (2024). Explainable AI for clinical trust in heart disease prediction. *Frontiers in Artificial Intelligence*, 7, 234–245.
29. Chen, Z., et al. (2024). Deep learning with wearable devices for cardiovascular monitoring. *Sensors*, 24(6), 1234–1245.
30. Rahman, M., et al. (2024). Real-time prediction of heart disease using IoT sensors. *International Journal of Health Informatics*, 10(2), 67–75.

221-35-861

ORIGINALITY REPORT

13%	9%	10%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	doctorpenguin.com Internet Source	1%
4	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	<1%
5	Submitted to The University of the West of Scotland Student Paper	<1%
6	www.mdpi.com Internet Source	<1%
7	Suneeta Satpathy, Álvaro Rocha, Sachi Nandan Mohanty, Tanupriya Choudhury. "Intelligent Data-Driven Systems with Innovations in Artificial Intelligence", CRC Press, 2025 Publication	<1%
8	Amin, Mohammad Shafenoor. "Identifying Significant Features and Data Mining Techniques in Predicting Cardiovascular Disease", University of Malaya (Malaysia), 2023 Publication	<1%