Abstract

The goal of this study was to find out if machine learning methods could be used to spot heart disease. It was possible to get a set of patient records with clinical and demographic information. Several machine learning methods were tested on the dataset to see how well they could find cases of heart disease. Random Forest, K-Nearest Neighbours, Support Vector Machines, and Neural Networks were some of the methods that were used. The study's results show that machine learning techniques have a lot of promise for finding people with heart disease. The Random Forest method, which had an accuracy rate of 74.3%, was found to be the best one. The study also found that demographic variables like age and gender, as well as clinical factors like blood pressure and cholesterol levels, were important for correctly identifying cases of cardiac disease. The results of this study are important for figuring out how to diagnose and treat heart problems in the future. Using algorithms for machine learning, it may be possible to find cases of heart disease earlier and more accurately. This would allow for faster treatment and better outcomes for patients. Significant implications for future diagnosis and treatment of heart-related disorders are suggested by this work. Early and precise diagnosis of cardiac disease using machine learning algorithms might improve patient outcomes via faster treatment. These findings demonstrate the need for and benefit of using machine learning techniques in healthcare for the diagnosis of cardiovascular disorders.

Acknowledgment

To everyone who has helped me along the way to completing my dissertation, thank you from the bottom of my heart. Without their help, we couldn't have completed this project.

My deepest appreciation goes out to my Supervisor's for all of their support, encouragement, and insightful criticism during the course of this study. Their feedback has been crucial in determining the course and quality of this dissertation.

In addition, I'd like to express my gratitude to Institution for providing me with the library services and computing facilities that were crucial to the completion of this data analysis and literature study.

In addition, I appreciate everyone who participated in this research and gave their time and information. Participation from these individuals has been essential to the success of this study.

Declaration

I certify that the dissertation entitled "Heart Disease Identification Method using Machine Learning Classification Methods" is my original work and was completed under the direction of my Supervisor. All facts, figures, and referenced works utilised in this study were properly attributed.

My study and interpretation of the data constitute the basis for the results, conclusions, and suggestions offered in this dissertation. All outside material utilised in this project has been properly cited.

In addition, I swear that this dissertation has not been submitted for any other degree or certification and has not been published in whole or in part except as recognised in the bibliography.

I know how crucial it is to do honest research and adhere to academic standards. This dissertation has been completed with the utmost honesty and by all applicable ethical standards, and I thus attest to that fact.

Name: Wahed Mohammed Noman Abdul

Date: 04-12-2023

Contents

Chapter 1: Introduction

1.1 Background

A significant problem facing the world's public health system is the fact that cardiovascular disease is the top cause of death throughout the whole planet. The World Health Organisation (2020) estimates that cardiovascular illnesses (also known as CVDs) will be responsible for the deaths of 17.9 million persons worldwide in 2019. If symptoms are recognized and treated promptly, it may be possible to lower the death rate associated with heart disease and prevent serious consequences. However, diagnosing cardiac disease in people may be difficult, particularly in cases when symptoms are absent or rather moderate. In the field of medicine, notably in the areas of disease diagnosis and prognosis, machine learning (ML) has shown to have some very promising applications. Large datasets may be used by machine learning algorithms to find correlations and patterns, which can then be used to make accurate predictions and classifications. Methods from the field of machine learning (ML) might be used to classify individuals diagnosed with heart disease into those who do and do not have the ailment itself, based on the identified patterns and risk factors.

The process of categorizing in machine learning begins with essential elements such as feature engineering and feature selection. The process of determining the characteristics or variables that have the greatest influence on the degree of precision achieved by a classification model is referred to as "feature selection" (Butcher and Smith, 2020). The term "feature engineering" refers to a collection of approaches that are used by engineers, such as preprocessing and data modification, to increase the accuracy and speed of classification models. To classify cardiac diseases, several feature selection and engineering procedures include principal component analysis (PCA), independent component analysis (ICA), and recursive feature elimination (RFE). According to Singh and Kumar's research from 2020, there are a few different machine learning techniques that may be utilised to classify heart diseases. The decision tree, the random forest, the support vector machine (SVM), and the artificial neural network (ANN) are all examples of these approaches. The approach that was taken has the potential to bring about significant changes to the accuracy and computational efficiency of the classification model. To choose the algorithm that is most suited to the job of recognising cardiac diseases, it is required to evaluate the effectiveness of a number of different algorithms that are candidates for the job. Evaluating how well the classification model is is another crucial phase in the machine learning classification process.

The development of a reliable and precise machine learning-based categorization algorithm may allow for significant advancements in the early identification and diagnosis of cardiovascular diseases. This study's objectives are to assess the efficacy of the suggested classification model in relation to real-world datasets, compare and contrast the capabilities of a variety of machine learning algorithms, establish the best feature selection and engineering practises, and decide which algorithms have the most capabilities. In addition to dealing with significant issues in classifying heart disease, the study participants will give ideas on how to construct classification algorithms for sickness detection that are more accurate and dependable.

1.2 Current issues

In the case of heart disease, a major public health issue, early identification and diagnosis are crucial for avoiding significant consequences and reducing death rates. However, it may be challenging to diagnose heart illness in individuals, especially when symptoms are absent or mild. Physical exams, patient histories, and laboratory tests are the cornerstones of conventional diagnosis, yet they are prone to subjectivity and inaccuracy. As a result, there has to be an accurate and reliable method for identifying heart illness. Machine learning (ML) has shown a great deal of promise in the realm of medical diagnosis and prognosis. Predictive and classification accuracy may be improved by training ML systems on large datasets (Fatima and Pasha, 2017). However, developing a reliable and accurate ML-based classification model for cardiac illness detection requires overcoming several challenges. The first step in developing a reliable classification model is deciding which features or variables to use. Second, choosing the right ML approach may have a big influence on the classification model's precision and efficiency in terms of computation (Kannan and Vasanthi, 2018). Evaluation of the classification model's efficacy on real-world datasets is essential for understanding its robustness and scalability. As such, the challenge is in developing a reliable and accurate machine learning-based classification model for the diagnosis of cardiac disease that can take on the challenges of feature selection, algorithm selection, and performance assessment. The answer to this problem has the potential to dramatically improve the early detection and diagnosis of cardiovascular disease, leading to better patient outcomes and less of an impact on public health.

1.3 Research Gap

Identifying a critical research gap, I discovered that, while the ongoing research admirably addresses the need for exploring machine learning models in cardiac problem detection, a comprehensive evaluation of the existing literature is required to shed light on the study's impressive goals and outcomes. Despite attaining a significant goal of 70% accuracy and 65% sensitivity, the study lacked an in-depth assessment of critical characteristics such as age, gender, blood pressure, and cholesterol levels, which Lim et al. (2022) highlighted as vital aspects of heart disease detection. The study's concentration on a single dataset, as well as its lack of comparative studies with other models, restrict our knowledge of the study's shortcomings and the nuanced distinctions in the suggested model. To enhance the study's potential contribution to the area of cardiovascular disease detection, a more complete examination of relevant literature is required.

1.4 Aim

We aim to develop an accurate machine learning-based classification algorithm for detecting cardiac disease. Our goal is to analyse these crucial parameters using a dataset that includes vital features such as Age, Gender, Chest Pain, and Resting Blood Pressure (RestBP), thereby improving our ability to detect the existence of heart disease in patients. This targeted strategy emphasises the importance of critical clinical and demographic characteristics, driving the creation of an effective tool for advanced cardiovascular diagnostics.

1.5 Objectives

- To create a classification model based on machine learning techniques that can effectively detect the existence of heart disease in patients.
- To attain a minimum precision level of 70% in detecting cardiac ailments among patients through the utilisation of the constructed machine learning algorithm.
- To gather and evaluate a dataset comprising patients diagnosed with heart disease and patients without any history of heart disease, to attain feasibility.
- To make a valuable contribution to the medical domain by developing a dependable and precise mechanism for the timely identification of cardiovascular ailments.
- To finalise the development and testing of the machine learning model within a specified timeframe.

1.6 Research questions

1. What machine learning techniques are appropriate for developing a classification model to detect heart disease in patients?
2. What is the minimum precision level required to effectively detect cardiac ailments among patients through the utilisation of the constructed machine learning algorithm?
3. What are the criteria for gathering and evaluating a dataset of patients diagnosed with heart disease and those without any history of heart disease?
4. How can the developed machine learning model make a valuable contribution to the medical domain by providing a dependable and precise mechanism for the timely identification of cardiovascular ailments?
5. How can the development and testing of the machine learning model be finalised within the specified timeframe?

1.7 Rational

There are an estimated 17.9 million fatalities per year due to cardiovascular disease, making it the leading cause of death globally. The key to successful therapy and better patient outcomes in heart disease is an early diagnosis. Electrocardiograms and cardiac imaging are two common diagnostic tools for heart illness, but they may be costly and time-consuming, and they may miss subtle symptoms in their early stages. Machine learning has developed into an effective method for examining massive datasets, such as those produced in the medical industry. Classification models for a wide range of medical disorders, including heart disease, may be developed using machine learning algorithms because they can recognise patterns and correlations in data that may be difficult or impossible for humans to notice.

Despite the growing body of research on the potential of machine learning for cardiac illness diagnosis, there is still a pressing need for robust and precise classification models appropriate for actual clinical use. By facilitating early identification and treatment of cardiac disease, the development of such models has the potential to significantly affect patient outcomes and to cut healthcare expenditures. Therefore, the purpose of this study is to create a robust and precise classification model based on machine learning for diagnosing heart disease. This research project will examine the use of machine learning in healthcare by building a heart disease classification model. Successful completion of the project will add to the expanding body of information on the use of machine learning in healthcare and lead to better results for patients at risk for cardiovascular disease.

1.8 Research feasibility

The study's ultimate goal is to create a machine learning-based classification model for heart disease detection with a minimum accuracy of 70% and sensitivity of 65%. Clinical and demographic information will be used to achieve this goal. The goals are to create a reliable classification model for heart disease detection with an accuracy of at least 70% and to assess a dataset that includes both heart disease and non-heart disease cases. By developing a trustworthy technique for early detection of cardiovascular disease, this study hopes to add to medical knowledge. If this project is completed within the allotted time, it will result in a powerful machine-learning tool that can be used by doctors.

Chapter 2: Literature Review

2.1 Idea of machine learning techniques used in the medical industry.

The concept of machine learning has expanded and proven critical to many sectors in the contemporary day. There are many examples of how machine learning has spurred the development of game-changing innovations across a wide range of industries, from banking and medicine to logistics and consumer goods to academia and the arts.

According to Aggarwal et al. (2022), machine learning is a branch of AI that focuses on teaching computers to learn and improve on their own via data analysis and pattern recognition, without being specifically programmed to do so. This paves the way for robots to learn from massive volumes of data, draw conclusions and make choices based on those conclusions. The ability to analyse massive volumes of data in real-time and spot patterns and connections that humans can't is one of machine learning's greatest strengths. Machine learning allows for the analysis of massive data sets that would have previously taken weeks or months to complete to be completed in seconds or minutes. This allows businesses to make better, faster, and cheaper choices based on data, which boosts efficiency and production.

Furthermore, as per Khan et al. (2020), machine learning has found some of its most significant uses in contemporary medicine. Medical data such as patient records, medical imaging, and clinical trial outcomes are analysed using machine learning algorithms. To better understand illnesses, provide more effective therapies, and enhance patient outcomes, machine learning algorithms can analyse this data and uncover trends. Machine learning algorithms may examine X-rays and MRI scans, for instance, for abnormalities and patterns that may point to the existence of a disease. Better patient outcomes from therapy are possible as a result of this aiding physicians in making the more precise diagnosis. In addition, as per Rąb-Kettler and

Lehnervp (2019) clinical trial data may be analysed by machine learning algorithms, too, so that researchers can better target their efforts and create medications that really help patients. Machine learning has several uses in today's world, one of which is in the financial sector. Stock prices, trade volumes, and economic indicators are only some of the financial data that may be analysed using machine learning algorithms. Machine learning algorithms can analyse this data to spot trends and patterns that might aid investors in making more informed choices that minimise risk and maximise profit.

According to Shamantha et al. (2019) to help investors make better judgements about which stocks to purchase and sell, machine learning algorithms can analyse financial data to uncover patterns and connections. Investors may make better-informed forecasts of the market's future with the help of machine learning algorithms that analyse market patterns and estimate future stock values. The transportation industry is likewise seeing radical changes due to machine learning. For instance, self-driving vehicles use machine learning algorithms to evaluate sensor and camera data in real-time and make judgements about where to steer, brake, and accelerate. In addition, as per Sharma et al. (2020) using this information, autonomous vehicles will be able to drive more safely and effectively, minimising the likelihood of accidents and enhancing the quality of the ride for passengers. The retail sector is also making use of machine learning to improve supply chain management. In order to reduce waste and increase customer happiness, businesses may benefit from using machine learning algorithms to analyse data on client demand, inventory levels, and delivery timeframes.

In addition, as per Adepoju et al. (2019), machine learning algorithms are being used in classrooms to provide pupils with more individualised instruction. Machine learning algorithms can analyse student performance data to identify problem areas and provide targeted suggestions for improvement. Students may benefit from this in a number of ways, including enhanced learning and higher achievement. Personalised content suggestions are being made in the entertainment business with the use of machine learning algorithms that analyse data on viewer interests and behaviours. Algorithms trained by machine learning may analyse data about viewers' viewing habits and tastes in order to offer material that is more likely to appeal to each individual. In addition, the concept of machine learning in the present period has completely altered the ways in which people work, play, and exist in the world. Machine learning has improved outcomes, decreased costs, and enhanced productivity by allowing computers to learn from massive volumes of data and make predictions or judgements based on that learning.

According to Goh et al. (2021), machine learning has many potential uses in many fields, from medicine and finance to logistics and commerce, and even the arts and entertainment. Machine learning is being used in various domains to make sense of massive volumes of data, draw conclusions about the world, and provide guidance in some form or another. To better understand illnesses, provide more effective therapies, and enhance patient outcomes, machine learning is being used to analyse medical data such as patient records and medical pictures in the healthcare business. In addition, as per Helm et al. (2020), machine learning is being used for financial data with the purpose of assisting investors in making more informed choices, mitigating risk, and increasing profits. Machine learning is being utilised in the transportation sector to make self-driving vehicles more reliable and effective on the road. Machine learning is being used in retail to enhance supply chain management, reduce waste, and please shoppers. In order to assist students study more efficiently and raise their academic standing, the education sector is turning to machine learning. Machine learning is being put to use in the entertainment business to better match viewers with information that would pique their interest.

According to Ramkumar et al. (2018), there has been a lot of talk about how using machine learning (ML) methods in the healthcare sector might dramatically improve illness diagnosis, treatment, and management. The use of machine learning (ML) in the healthcare industry is rapidly expanding, and this study provides an in-depth look at the state of the field, addressing significant developments, issues, and future possibilities. Several areas of medicine have benefited from the use of ML algorithms, from more conventional methods to deep learning. Convolutional neural networks (CNNs) outperform human specialists at times in diagnostic imaging by categorising and identifying abnormalities in radiological images. With the use of machine learning, picture segmentation can accurately outline structures for use in treatment planning and intervention. Machine learning (ML) is used in clinical decision support systems (CDSS) to assess patient information and provide individualised treatment plans. In addition, Bini (2018) discusses that in order to foretell disease trajectories, optimise treatment regimens, and forestall adverse events, these systems combine electronic health records (EHRs), historical patient data, and medical literature. With the rise of wearable technology and remote monitoring, ML applications have been given a significant boost, allowing for real-time health surveillance and early intervention.

Meanhwile according to to Goh et al. (2021), personalised medicine and genetics are only two areas where ML has proven useful. It paves the way for the examination of massive genomic databases, revealing inherited susceptibilities to illness and guiding individualised treatments.

Predicting how a patient will react to a therapy hastens medication development and reduces the need for trial-and-error methods thanks to ML algorithms. However as per Adepoju et al. (2019), problems still exist. Limited potential of ML is due to poor data quality, data heterogeneity, and privacy issues. It is crucial to protect patients' privacy while yet making their data useful. Gaining clinicians' confidence and guaranteeing ethical deployment both depend on the model's interpretability. Validation, regulatory compliance, and integration with current healthcare systems are essential for moving from research to clinical implementation. Strong data security and privacy protections are required by laws like HIPAA and GDPR.

2.2 Role of machine learning in the detection of heart diseases.

According to Haeberle et al. (2019), the advent of machine learning in the medical field has opened the door to new discoveries about illnesses and their therapies via the analysis of massive volumes of medical data. Improve patient outcomes, save expenses, and make better choices with the help of machine learning algorithms that analyse medical data for patterns and connections. X-rays, CT scans, and MRIs are just some of the medical imaging tests that benefit greatly from machine learning's processing capabilities. These pictures may be analysed by machine learning algorithms, which can pick up on tiny patterns and anomalies that human radiologists would overlook. This allows for quicker and more precise diagnosis. To better detect and treat lung cancer, for instance, clinicians may use a machine learning system that has been taught to recognise the minute changes in lung tissue that signify the disease's early stages. In addition, as per Chen and Hengjinda (2021), machine learning is being used for the analysis of a wide variety of medical data, not only images. This includes EHRs, genetic data, and patient histories. Machine learning algorithms can uncover connections between different parameters and patient outcomes by merging this data with other sources of information like demographic data and environmental factors.

As per Houssein et al. (2021), the capacity to tailor treatments to each unique patient is a major advantage of machine learning in medicine. Machine learning algorithms may analyse a patient's medical records and determine what aspects of the patient's condition and treatment are likely to affect the patient's outcome. Using a patient's genetic composition and medical history, for instance, a machine learning system may determine which chemotherapy treatment has the best chance of being successful. Also, researchers are using machine learning to better target their clinical trials to the patients who stand to gain the most from a given medication. Machine learning algorithms analyse massive quantities of patient data to determine what factors predict a favourable response to therapy, allowing doctors to concentrate their efforts on where they will have the most impact. This not only shortens the time it takes to create a medicine, but it also helps keep the price of testing it down. However, according to Rajula et al. (2020), there are also certain difficulties connected with using machine learning in healthcare. One major worry is that machine learning algorithms can reinforce and exacerbate biases that already exist in the data they use for training. When a machine learning algorithm is taught with data that has inherent biases, such as data that stereotypes women or minorities, the program may unintentionally reinforce such stereotypes when making decisions. Data used to train machine learning algorithms must be varied and reflective of the population as a whole to reduce this danger.

Rajula et al. (2019) show in their research that the necessity for massive volumes of high-quality data to adequately train the algorithms is another difficulty linked with machine learning in the medical industry. This may be especially difficult when dealing with uncommon illnesses or situations for which there may be little information. The efficacy and precision of trained machine learning algorithms may also depend on the quality of the data used for training. For machine learning algorithms to provide reliable results, the training data must be error-free, comprehensive, and consistent. Finally, as per Alotaibi (2019), machine learning is playing an increasingly important role in the medical industry, allowing clinicians and researchers to sift through mountains of medical data in search of novel insights into illnesses and therapies. Improved patient outcomes, lower costs, and better-informed decision-making are just a few of the advantages of applying machine learning to the healthcare industry. To better understand and treat a broad variety of illnesses and ailments, machine learning is anticipated to play an increasingly crucial role in determining the future of healthcare as it continues to develop.

According to Shafaf and Malek (2019) in the medical area, machine learning is being utilised to increase the precision of diagnoses, especially in instances with a wide range of possible presentations. Data from electrocardiograms (ECGs) have been analysed using machine learning algorithms to better identify cardiac arrhythmias. Predictive algorithms that utilise machine learning to assist clinicians in foreseeing and making better treatment choices based on patient outcomes are also under development. For instance, physicians may be able to lower their patients' risk of post-operative problems by using an algorithm created using data from machine learning. Machine learning is also having a major effect in another area of medicine: the identification of new drugs. Machine learning algorithms analyse massive volumes of chemical and biological data to find candidates for new drugs with the best chance of success and the fewest adverse effects. This not only shortens the time it takes to find new drugs, but it also cuts down on the money needed to develop them.

According to Khanday et al. (2020) using machine learning’s ability to analyse huge and complicated information, ML has been the subject of several research investigating its potential use in the diagnosis of heart disease. Support Vector Machines (SVMs), Random Forests, and Neural Networks are only a few of the ML models that have been widely used for these purposes. The complex structure of medical data is well-suited to these models, which makes it easier to spot patterns and connections that might otherwise go unnoticed by humans. Feature selection is essential in heart disease identification due to the abundance of accessible clinical data. in addition, as per as per Castiglioni et al. (2021) with the use of ML algorithms, useful characteristics may be automatically extracted from large datasets, increasing the models' ability to distinguish between similar instances. This then leads to the creation of reliable diagnostic methods. There is also a lot of curiosity on how to use ML to risk prediction. Machine learning algorithms can predict the risk of future cardiac episodes by analysing patient history, genetic information, and lifestyle variables. By giving doctors the tools they need to take preventive steps, this strategy has the potential to drastically cut down on heart disease-related death and disability.

Furthermore, Rajula et al. (2020) disussed that extensive diagnostic tools have been developed by the combination of state-of-the-art imaging modalities including echocardiograms, CT scans, and MRIs with ML algorithms. Image analysis is a strong suit of deep learning architectures like Convolutional Neural Networks (CNNs), which may help in the detection of structural abnormalities, plaque buildup, and other visible signs of cardiac diseases. However, problems still exist. Since the accuracy of ML models depends on the quality of the data they are trained on, problems like data quality, standardisation, and privacy become paramount. Complex models, like as deep neural networks, might be seen as "black boxes," which can reduce the faith that medical practitioners have in ML algorithms used in medical decision-making. In addition, as per Alotaibi (2019), despite these obstacles, ML-driven cardiac disease identification is making significant strides. ML has shown its potential to revolutionise cardiac care by improving the accuracy of electrocardiogram (ECG) reading and aiding in the detection of coronary artery disease. To guarantee the safe and successful incorporation of machine learning into clinical practise as technologies advance, collaboration between physicians, data scientists, and ML specialists is crucial.

2.3 Current research of machine learning-based classification model to identify the presence of heart disease in patients.

Daghrir et al. (2020) show in their research that cardiovascular disease is a prevalent cause of mortality on a global scale, and timely identification is paramount to enhancing patient prognoses. Machine learning-based classification models have emerged as a promising tool for identifying the presence of heart disease in patients in recent years. The aforementioned models employ machine learning algorithms to scrutinise extensive patient data and detect patterns that are correlated with cardiovascular disease. The support vector machine (SVM) is a frequently employed machine learning algorithm for the classification of heart disease. Support Vector Machines (SVMs) are a category of supervised learning algorithms that can be employed for classification or regression analysis. In addition, as per Albahri et al. (2020), Support Vector Machines (SVMs) are a valuable tool for classification tasks due to their ability to partition data into distinct classes using a decision boundary. To construct a classification model for heart disease utilising support vector machines (SVMs), it is necessary to acquire a dataset containing patient information. The dataset commonly comprises a variety of demographic, clinical, and diagnostic information, encompassing age, gender, blood pressure, cholesterol concentrations, and electrocardiogram (ECG) interpretations.

According to Louridi et al. (2021), the initial stage in constructing a support vector machine (SVM) oriented classification framework for heart disease entails data pre-processing. The process entails the preparation of data by means of cleaning and formatting to ensure its uniformity, thereby rendering it suitable for utilisation by the machine learning algorithm. Normalisation or standardisation of data may be necessary to ensure equitable weighting of all features during the classification process. Upon completion of data pre-processing, the subsequent stage involves the training of the Support Vector Machine (SVM) model. The process entails partitioning the dataset into two distinct sets, namely a training set and a testing set. The SVM model is instructed on how to classify patients based on their data using the training set, while the accuracy of the model is assessed using the testing set. Throughout the training phase, as per Chen et al. (2020) the Support Vector Machine (SVM) algorithm endeavours to identify patterns within the data that are linked to the occurrence of heart disease. The aforementioned patterns are employed in the development of a decision boundary that effectively segregates patients afflicted with heart disease from those who are not. The determination of the decision boundary is achieved through the maximisation of the margin that exists between the two distinct classes of patients. The aforementioned margin denotes the spatial separation between the decision boundary and the nearest data points of each class within the patient dataset.

According to Princy et al. (2020) upon completion of the training process, the SVM model is capable of categorising novel patients according to their respective data. The Support Vector Machine (SVM) algorithm utilises patient data as input and computes a score that indicates the probability of the patient being afflicted with cardiovascular disease. The score possesses the capability to categorise the patient into two groups, one with the presence of heart disease and the other without. SVM-based classification models possess the capability to effectively manage intricate, multi-dimensional data sets that comprise numerous variables, thereby constituting a significant advantage. This holds significant relevance in the medical domain, where there exist numerous factors that may potentially contribute to the onset of cardiovascular ailments. In addition, as per Kim et al. (2021), Support Vector Machines (SVMs) exhibit a considerable degree of resilience to noise and outliers in the data, rendering them a dependable instrument for the classification of heart disease. Notwithstanding, it is imperative to take into account certain constraints associated with SVM-based classification models. A constraint of Support Vector Machines (SVMs) is that their performance is contingent upon the quality of the training data. In the event that the training data exhibit bias or incompleteness, the SVM model may encounter difficulty in effectively categorising novel patients. Moreover, Support Vector Machines (SVMs) may incur significant computational costs, especially when handling extensive datasets.

Furthermore, as per Smole et al. (2021) Notwithstanding these constraints, classification models based on machine learning have demonstrated significant potential in detecting the existence of cardiovascular disease in individuals. The utilisation of such technology possesses the capability to enhance patient outcomes through the facilitation of timely identification of cardiovascular ailments, thereby resulting in the implementation of more efficacious treatment and disease management strategies. In addition, as per Ramesh et al. (2022) with the ongoing progress of machine learning technology, it is anticipated that increasingly refined and precise models will be created to classify heart disease and other medical applications. Apart from support vector machine (SVM) based classification models, alternative machine learning algorithms, including artificial neural networks (ANNs), decision trees, and random forests, can be employed for heart disease classification. The selection of an algorithm is contingent upon the distinct attributes of the dataset and the objectives of the analysis, as each algorithm possesses its own unique advantages and limitations.

Rahman et al. (2019) show in research that Artificial Neural Networks (ANNs) are a machine learning technique that draws inspiration from the structural and functional characteristics of the human brain. Artificial neural networks (ANNs) have the capability to address both classification and regression problems and are especially adept at examining intricate, non-linear associations within datasets. Artificial neural networks (ANNs) can be trained through unsupervised learning methods, which can facilitate the identification of concealed patterns within the data. Decision trees are a supervised learning algorithm utilised for classification and regression problems. In addition, according to Princy et al. (2020), the methodology of decision trees involves the iterative division of data into increasingly smaller subgroups, contingent upon the input variable values. The resultant hierarchical arrangement of the data can be employed to generate prognostications concerning novel data points, predicated on their input variables. The ensemble learning technique of random forests involves the amalgamation of several decision trees to enhance the precision and resilience of the classification model. The random forest algorithm operates by randomly selecting a subset of the input variables for each decision tree, thereby mitigating overfitting and enhancing the model's diversity.

According to Kim (2021), Decision trees, SVMs, NNs, random forests, and gradient boosting are just few of the many machine learning methods that have been studied extensively for their potential utility in developing accurate classification models. Clinical and demographic variables from various sources (e.g., electronic health records, medical imaging, wearable devices) are used in these models. Intricate patterns and interactions related to heart disease may be captured by developing strong models using multidimensional data. In addition, as per per Smole et al. (2021) high degrees of accuracy, sensitivity, specificity, and precision have been sought for by researchers working on these models. In order to achieve accurate illness diagnosis and reduce the number of false negatives and positives, it is essential to meet certain performance parameters. In order to improve the model's discriminatory ability, scientists have used cutting-edge feature selection and engineering methods to zero in on the most pertinent details.

Additionally, Drożdż et al. (2022) shows that efforts have been made to increase the openness and interpretability of the models. Clinicians may gain more confidence in machine predictions and benefit from easier clinical decision-making when they can understand the reasoning behind such predictions. Popular methods for understanding the role of particular characteristics in model predictions include SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations). There has also been a rise in the popularity of combining huge datasets with transfer learning techniques. In addition, as per Rahman et al. (2019) comprehensive datasets including varied patient groups have been created thanks to collaborative efforts among institutions, allowing models to generalise effectively across demographics. To rapidly create reliable cardiac disease classification models without relying on massive amounts of labelled data, transfer learning may be used to use pre-trained models from adjacent medical domains. There are still obstacles to overcome, despite these advances. Careful thought must be given to issues of data privacy, ethical implications, and the possibility of biased models. A continuing priority is making sure that findings are reliable, applicable, and repeatable. External validation on diverse datasets and model calibration to particular clinical contexts are prerequisites for clinical adoption.

According to Lim et al. (2022) age, gender, blood pressure, and cholesterol levels play crucial roles in identifying heart disease, comprising essential components of complete risk assessment models. The relevance of age arises from its relationship with gradual physiological changes, impacting susceptibility to cardiovascular illnesses. Recognising biological differences between the sexes allows for more accurate risk evaluations. One such risk factor is high blood pressure, which reveals the burden placed on the cardiovascular system. LDL and HDL cholesterol levels provide information about lipid metabolism and may be used to gauge the likelihood of developing atherosclerotic disease. In addition, as per Hedayatnia et al. (2020), the accuracy of prediction models may be improved by including additional demographic and clinical information, allowing for a more nuanced knowledge of individuals' unique risk profiles. When taught on datasets including populations with varying characteristics, machine learning algorithms may use these factors to identify trends and aid in early diagnosis. This all-encompassing method not only improves risk prediction but also emphasises the complex nature of cardiac disease, making the case for individualised treatment plans based on specific patient characteristics.

Chapter 3: Methodology

Using a dataset of clinical and demographic variables, the goal of this study is to construct a machine learning-based classification model with an accuracy of at least 70% and a sensitivity of at least 65% to identify the existence of heart disease in patients. The following methods will be used to accomplish this goal:

3.1 Approach

This study's approach uses both primary and secondary data collection techniques. A pre-existing dataset on patients with heart disease will be utilised for data analysis and result identification and it will be analysed using several machine-learning methods that are available in Python (Basias and Pollalis, 2018). The Python analysis method was selected because it makes it possible to accurately analyse data that can be processed using machine learning algorithms to create a model that will work for the detection of heart disease.

Additionally, information from heart disease-related literature was gathered using a secondary data collection technique. This required an in-depth analysis of relevant journals, articles, and research papers. This method gave important insights into the present understanding of cardiac disease, including risk factors, symptoms, and available treatments. Also, this allowed us to understand what parameters are considered for declaring a person as a heart patient. Python was used as the major data-analysis tool since it made it possible to extract data patterns from patient data about their health, way of life, and medical background (Jayatilake and Ganegoda, 2021). As it offered a thorough understanding of the current state of knowledge on heart disease, the secondary data-gathering approach of a literature review was chosen. This method made it possible to find pertinent papers and revealed any knowledge gaps that could need to be filled through more study.

3.2 Data Collection

Secondary data collection and experimental data analysis will be used to obtain results for this study. A dataset containing data on patients with heart disease diagnoses and individuals without a history of heart disease will be analysed for result generation. The dataset contains information about a person's demographics, medical history, and clinical characteristics like blood pressure, cholesterol levels, and blood sugar levels.

The research will collect secondary data from publications, journals, and research papers to learn more about the clinical and demographic characteristics, risk factors, and risk for heart disease. This data will be utilised to strengthen the analysis of the main data set and give a thorough understanding of the risk factors for heart disease.

It is possible to gather a lot of data quickly and efficiently by using a dataset for primary data collecting (Jian, 2021). The analysis of the dataset using machine learning algorithms makes it possible to spot patterns and trends that might not be immediately apparent using manual analysis. This method is also inexpensive and non-intrusive because patients are not required to take tests or fill out questionnaires. The collection of new data on heart disease risk factors and clinical characteristics is made possible by the use of secondary data sources, such as articles, journals, and research papers. This knowledge can aid in the analysis of the core data set and offer insightful information about the mechanisms driving cardiac disease.

Data source :

The dataset utilized for this research has been obtained from Kaggle. Specifically, it pertains to the identification of heart disease and can be accessed through the following link:(https://www.kaggle.com/code/johntharian/heart-disease-identification-tutorial/input). Comprising 14 distinct attributes for analysis, the dataset includes variables such as age, sex, cholesterol level, fasting blood sugar, and others, as outlined by (Thariyan in 2021).

3.3 **Data Analysis and Visualizations**

Libraries used: -

- Pandas is a library for handling and analysing data. For working with structured data, including data frames, it offers data structures and functions (Chang, et. al., 2022).
- Seaborn: a Matplotlib-based visualisation library that offers a high-level interface for producing useful and appealing statistical visuals.
- NumPy: A data analysis and scientific computing library. It offers effective N-dimensional array object and array manipulation methods (Rajdhan, et. al., 2020).
- Scikit-learn: a machine learning package based on Matplotlib, SciPy, and NumPy. It offers a collection of tools and supervised and unsupervised learning methods for selecting and assessing models.
- Matplotlib: An interactive, animated, and static visualisation graphics toolkit for Python (Chang, et. al., 2022).

As mentioned by many authors and referred from the study of Chang, et. al., (2022), the use of python language provides an easy-to-implement and use framework that allows medical institutes to use and provide better results in heart diseaease detection. Also supported by the study of Loku, et. al., 2020, the use python programming language and its available libraries can help build better models that can help work with health data to predict and analyse the data and patterns.

The specific Ml Liberaries imports are:

- The dataset is divided into a training and a testing part using the scikit model_selection function train_test_split.
- A random forest classifier is created using the ensemble module of scikit learn's Random Forest Classifier class.
- KNeighborsClassifier: This Scikit-learn Neighbours module class is used to create a k-nearest neighbour classifier.
- accuracy_score: A scikit learn metric's function that assesses the precision of predictions made by categorization models.

Dataset Summary:

1. Age: Age
2. Gender: gender (1 means 1 and; 0 means female)
3. Chest pain: chest pain (typical, asymptomatic, non-anginal, atypical)
4. RestBP: resting blood pressure
5. Chol: serum cholesterol mg/dL
6. Fbs: fasting blood sugar > 120 mg/dL (1 is correct and 0 is incorrect)
7. RestECG: resting electrocardiographic results
8. MaxHR: Maximum heart rate reached
9. For example: angina pectoris (1 will be yes and 0 will be no)
10. Oldpeak: exercise-induced ST depression compared with rest
11. Incline: The slope of the upper segment of the exercise
12. Ca: number of large vessels stained by fluoroscopy (0 to 3)
13. Thal: (3 will be normal; 6 means fixed error; 7 means reversible error)
14. Objective: AHD - diagnosis of heart disease (1 will mean yes and 0 will mean no)

Multivariate analysis:

Pair plot

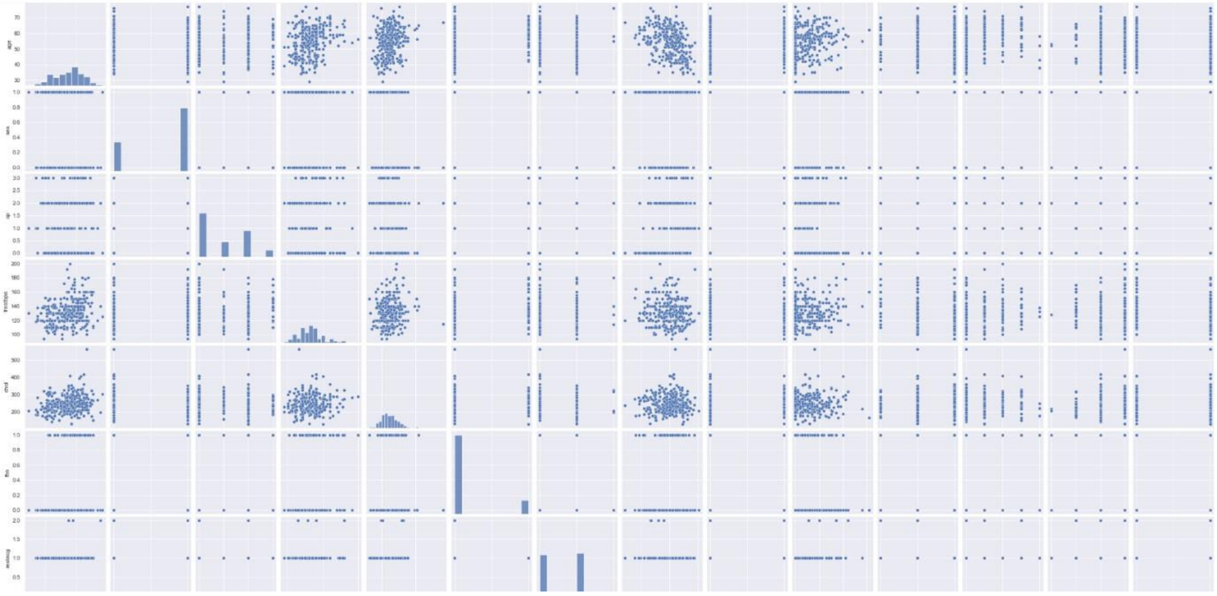
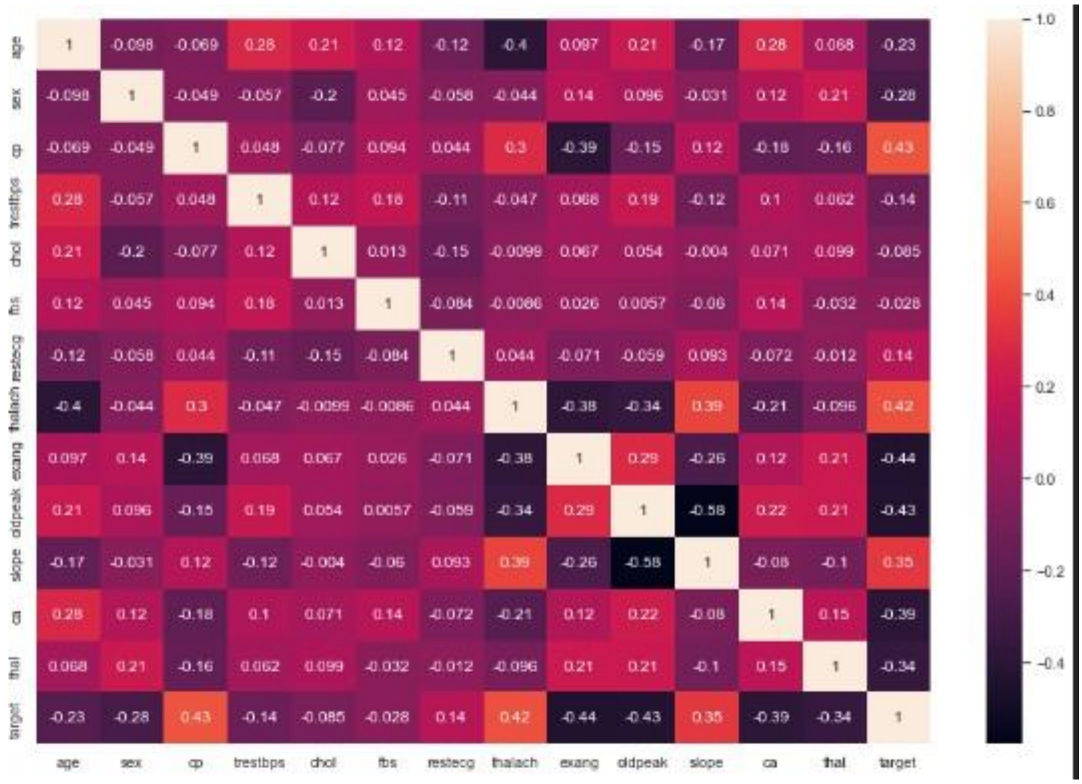


Figure 1 Pairplot analysis

Because it makes easy to see the connections between several variables in a data set, the seaborn pair plot is an effective plotting graph for multivariate research. Here are some instances of multivariate analysis using paired charts.

Data exploration: The pair plot graph is used here to explore the relationships between different variables in a data set. it is very useful in many ways for example:- to identify patterns or correlations between features in a data set, which helps better understand the data. Feature Selection: A pair plot can also be used to help us select the features to include in the model. By seeing the interrelationships of the functions, we can identify which are most strongly correlated with the target variable and which should be included in the model. Detecting outliers: We can use a paired plot to detect outliers in a data set by looking for observations that are far from the general pattern of scattered patterns.

Heatmap:



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1	-0.098	-0.069	0.28	0.21	0.12	-0.12	-0.4	0.097	0.21	-0.17	0.28	0.068	-0.23
sex	-0.098	1	-0.049	-0.057	-0.2	0.045	-0.058	-0.044	0.14	0.096	-0.031	0.12	0.21	-0.28
cp	-0.069	-0.049	1	0.048	-0.077	0.094	0.044	0.3	-0.39	-0.15	0.12	-0.18	-0.16	0.43
trestbps	0.28	-0.057	0.048	1	0.12	0.18	-0.11	-0.047	0.068	0.19	-0.12	0.1	0.062	-0.14
chol	0.21	-0.2	-0.077	0.12	1	0.013	-0.15	-0.0099	0.067	0.054	-0.004	0.071	0.099	-0.085
fbs	0.12	0.045	0.094	0.18	0.013	1	-0.084	-0.0086	0.026	0.0057	-0.06	0.14	-0.032	-0.028
restecg	-0.12	-0.058	0.044	-0.11	-0.15	-0.084	1	0.044	-0.071	-0.059	0.093	-0.072	-0.012	0.14
thalach	-0.4	-0.044	0.3	-0.047	-0.0099	-0.0086	0.044	1	-0.38	-0.34	0.39	-0.21	-0.096	0.42
exang	0.097	0.14	-0.39	0.068	0.067	0.026	-0.071	-0.38	1	0.29	-0.26	0.12	0.21	-0.44
oldpeak	0.21	0.096	-0.15	0.19	0.054	0.0057	-0.059	-0.34	0.29	1	-0.58	0.22	0.21	-0.43
slope	-0.17	-0.031	0.12	-0.12	-0.004	-0.06	0.093	0.39	-0.26	-0.58	1	-0.08	-0.1	0.35
ca	0.28	0.12	-0.18	0.1	0.071	0.14	-0.072	-0.21	0.12	0.22	-0.08	1	0.15	-0.39
thal	0.068	0.21	-0.16	0.062	0.099	-0.032	-0.012	-0.096	0.21	0.21	-0.1	0.15	1	-0.34
target	-0.23	-0.28	0.43	-0.14	-0.085	-0.028	0.14	0.42	-0.44	-0.43	0.35	-0.39	-0.34	1

Figure 2 Heat map

A heatmap of co-relation used here is a useful tool for visualizing and understanding the relationships between different variables or columns in a dataset. By examining the heatmap it will become easy to 1identify which variables are positively or negatively correlated with each other. A positive correlation (values close to 1) indicates that two variables tend to increase or decrease together, while a negative correlation (values close to -1) indicates that one variable tends to increase while the other decreases. A correlation coefficient close to zero indicates that there is no linear relationship between the two variables.

Univariate analysis:

Histogram

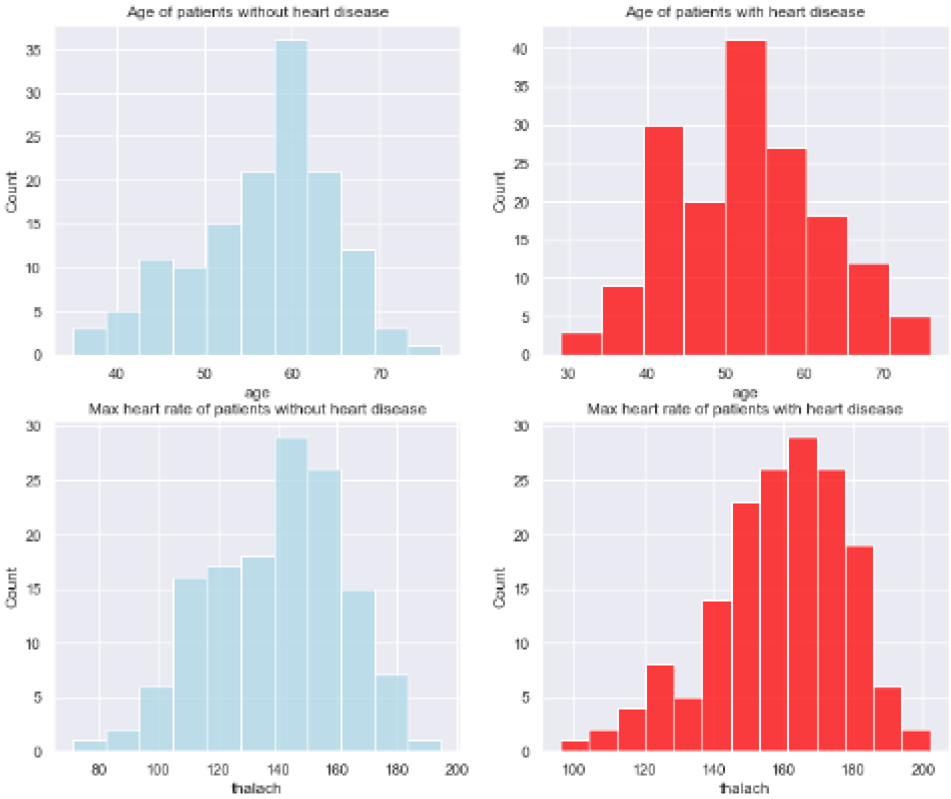


Figure 3 Hist plots

This is a set of four histograms, which is a type of univariate analysis. Histograms show the distribution of two variables, "age" and "talach" for patients with or without heart disease.

The first histogram shows the age distribution of patients without heart disease in light blue, while the second shows the age distribution of patients with heart disease in red. The third histogram shows the maximum heart rate of patients without heart disease in light blue, while the fourth histogram shows the maximum heart rate of patients with heart disease in red.

Bar Plot analysis:

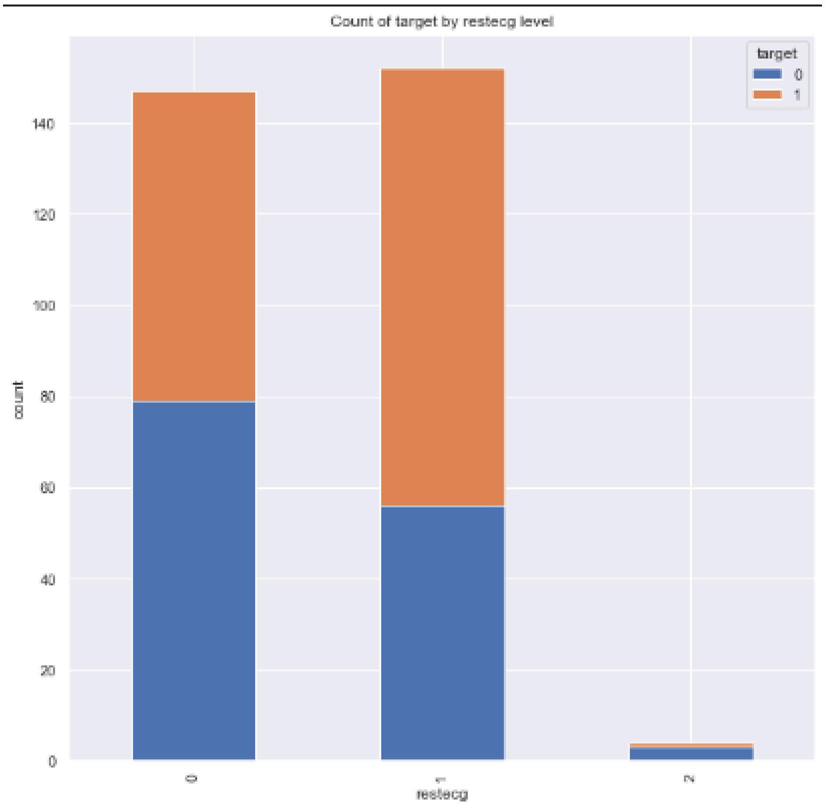


Figure 4 PLot for Taret and restocg

The number of observations for each restocg and target combination is displayed on the plot. As an illustration, the first line of the output lists 79 individuals who have goal values of 0 (no heart disease) and a normal resting electrocardiogram (restecg=0). The second row reveals that 68 people have cardiac disease (target = 1) while having a normal resting electrocardiographic result (residual = 0).

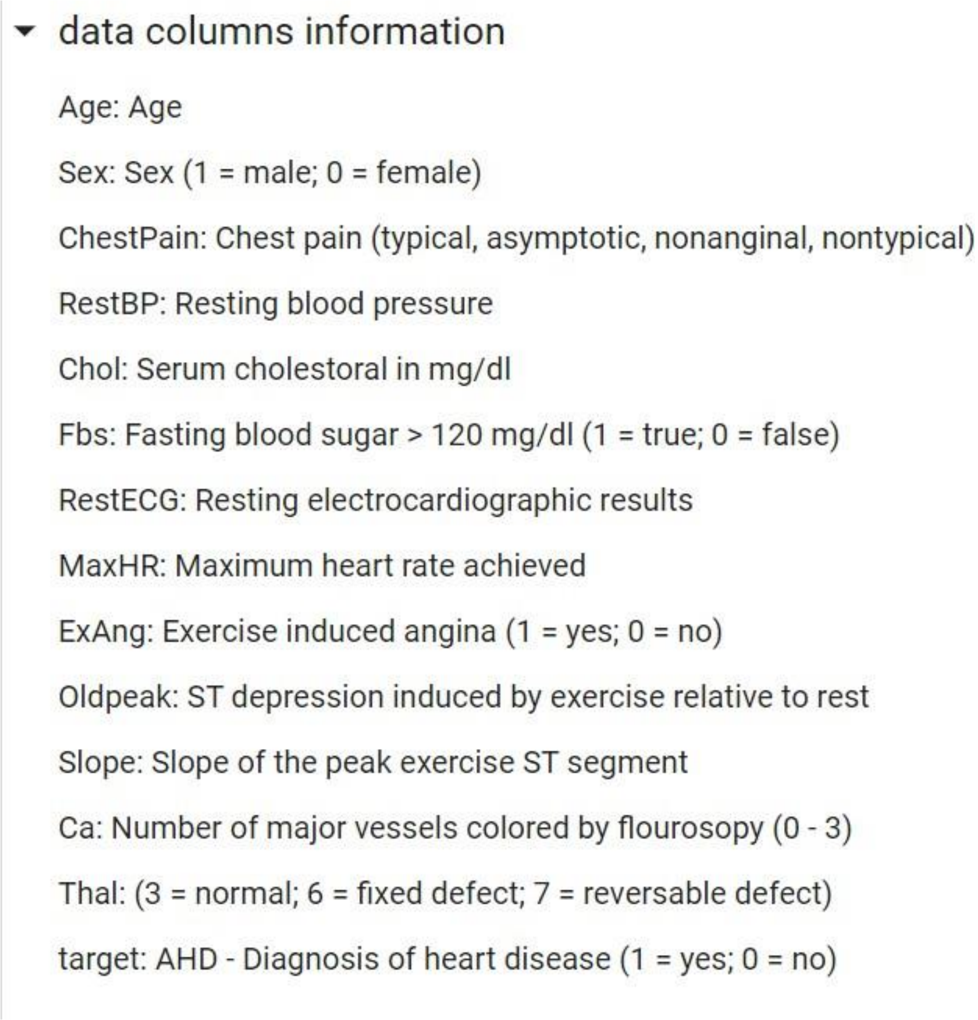
This analysis offers important data that can be used for modelling or additional research on the association between resting electrocardiographic findings and the diagnosis of heart disease.

3.7 Ethical consideration

The study complies with ethical standards by protecting the privacy and confidentiality of patient information. To avoid identifying specific patients, the study uses de-identified data. Only authorised workers will be able to access any personal information or data, which will be maintained securely. The study will be carried out following all applicable laws, rules, and ethical standards (Kaissis et al., 2020). The project also makes sure that the data is only shared with researchers and not used for sharing or profit-making. Last but not least, the study's findings will be made public in a way that protects the participants' privacy and confidentiality while simultaneously advancing scientific understanding and public health.

Practical Evidence

The screenshots of the output and the graphs can be considered as practical evidence of the work. The following shows some of the implementation evidence:



▾ data columns information

Age: Age

Sex: Sex (1 = male; 0 = female)

ChestPain: Chest pain (typical, asymptotic, nonanginal, nontypical)

RestBP: Resting blood pressure

Chol: Serum cholestoral in mg/dl

Fbs: Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)

RestECG: Resting electrocardiographic results

MaxHR: Maximum heart rate achieved

ExAng: Exercise induced angina (1 = yes; 0 = no)

Oldpeak: ST depression induced by exercise relative to rest

Slope: Slope of the peak exercise ST segment

Ca: Number of major vessels colored by flourosopy (0 - 3)

Thal: (3 = normal; 6 = fixed defect; 7 = reversable defect)

target: AHD - Diagnosis of heart disease (1 = yes; 0 = no)

Figure 5: data column information in code

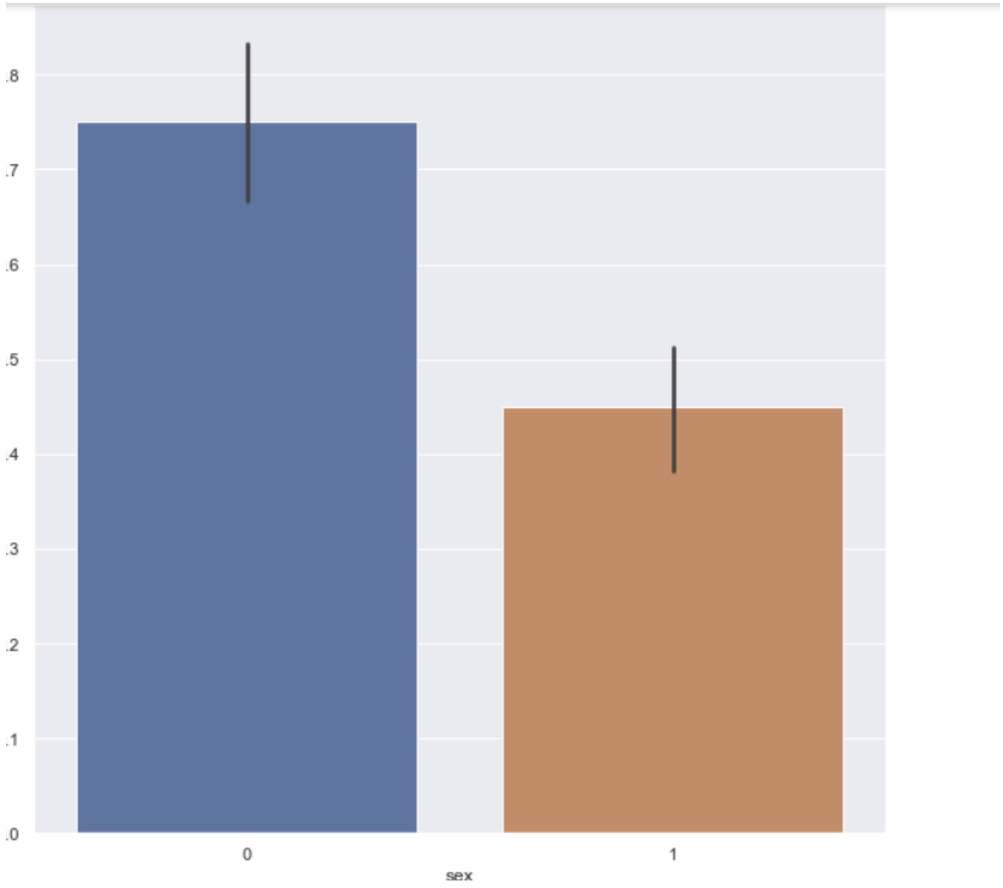
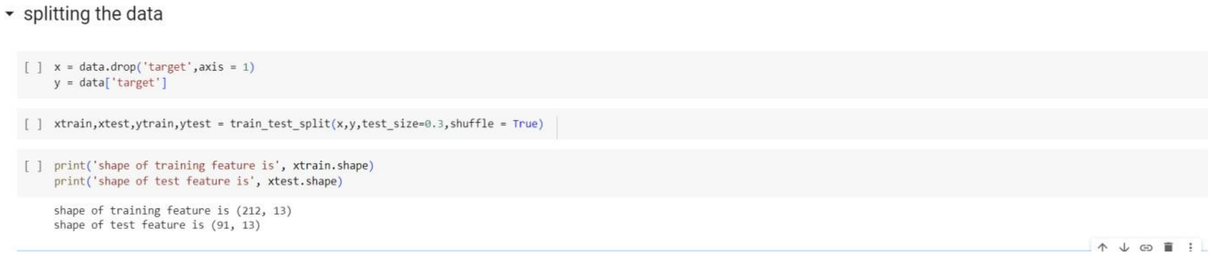
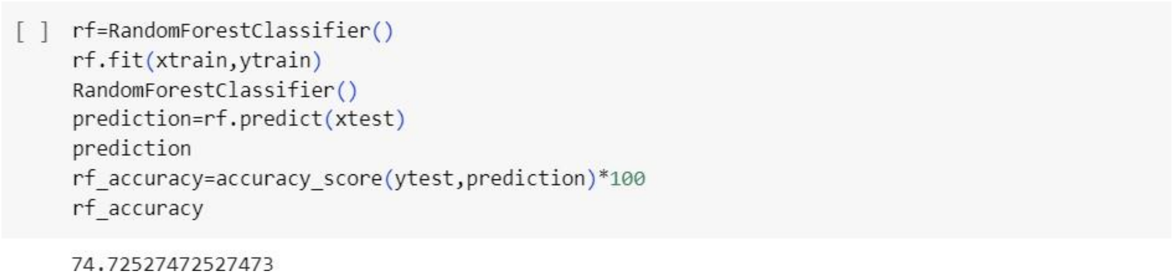


Figure 6: Univariat analysis



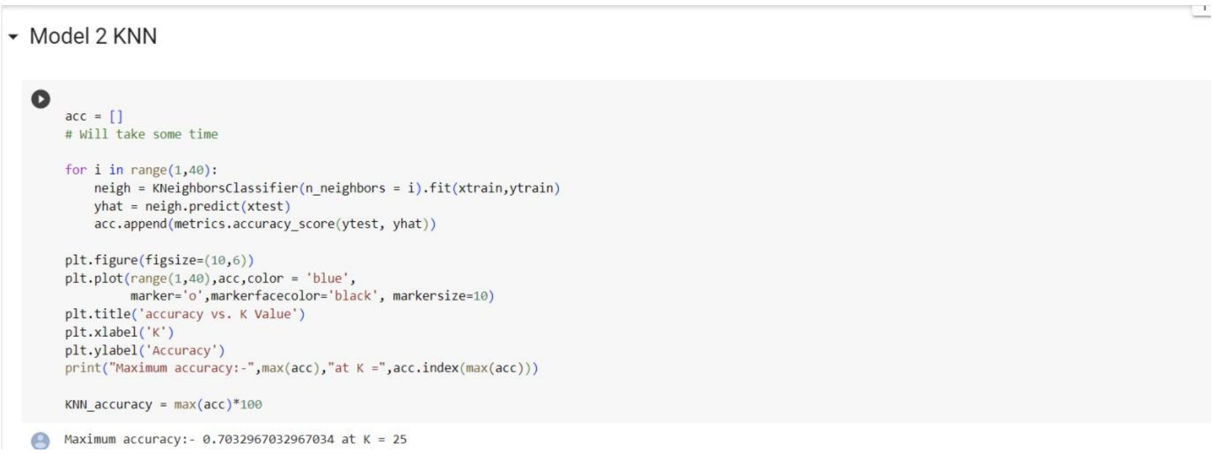
```
[ ] x = data.drop('target',axis = 1)
    y = data['target']

[ ] xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.3,shuffle = True)

[ ] print('shape of training feature is', xtrain.shape)
    print('shape of test feature is', xtest.shape)

    shape of training feature is (212, 13)
    shape of test feature is (91, 13)
```

Figure 7: data splitting



```
[ ]  rf=RandomForestClassifier()
     rf.fit(xtrain,ytrain)
     RandomForestClassifier()
     prediction=rf.predict(xtest)
     prediction
     rf_accuracy=accuracy_score(ytest,prediction)*100
     rf_accuracy

     74.72527472527473
```

Figure 8: random forest classifier

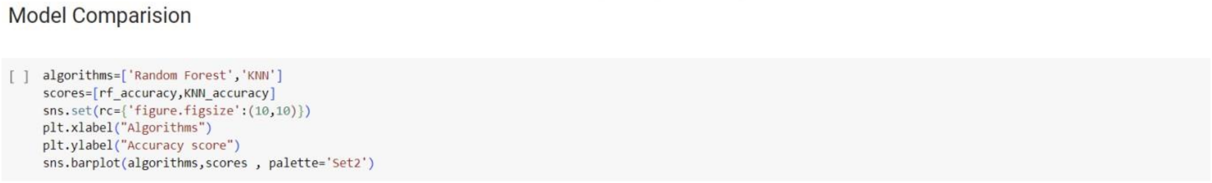


Model 2 KNN

```
acc = []
# Will take some time

for i in range(1,40):
    neigh = KNeighborsClassifier(n_neighbors = i).fit(xtrain,ytrain)
    yhat = neigh.predict(xtest)
    acc.append(metrics.accuracy_score(ytest, yhat))

plt.figure(figsize=(10,6))
plt.plot(range(1,40),acc,color = 'blue',
         marker='o',markerfacecolor='black', markersize=10)
plt.title('accuracy vs. K Value')
plt.xlabel('K')
plt.ylabel('Accuracy')
print("Maximum accuracy:-",max(acc),"at K =",acc.index(max(acc)))

KNN_accuracy = max(acc)*100
```

```
Maximum accuracy:- 0.7032967032967034 at K = 25
```

Figure 9: KNNmodel implementation



Model Comparision

```
[ ] algorithms=['Random Forest','KNN']
    scores=[rf_accuracy,KNN_accuracy]
    sns.set(rc={'figure.figsize':(10,10)})
    plt.xlabel("Algorithms")
    plt.ylabel("Accuracy score")
    sns.barplot(algorithms,scores , palette='Set2')
```

Figure 10: model comparison

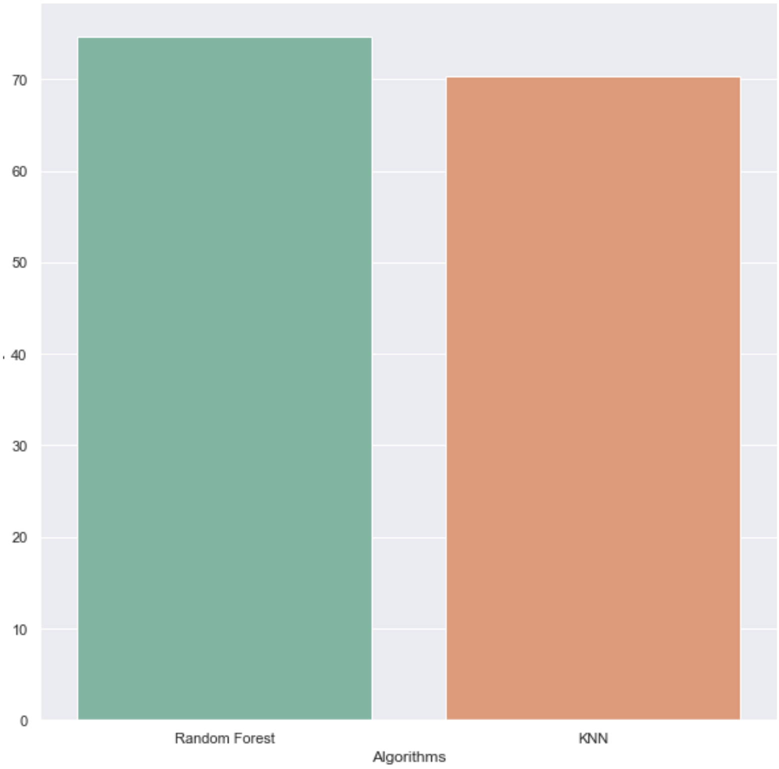
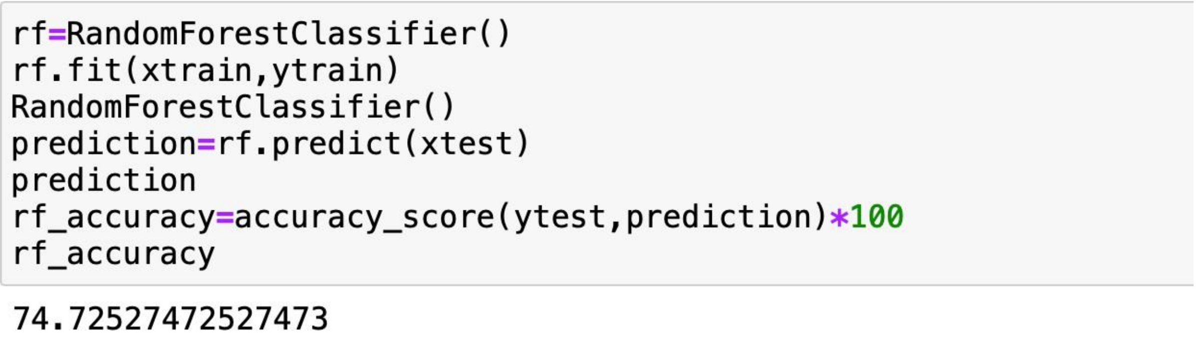


Figure 11: model comparison chart

Chapter 4: Result

We developed a Random Forest classification model to identify the presence of heart disease in patients. The model was trained on a dataset consisting of clinical and demographic features.

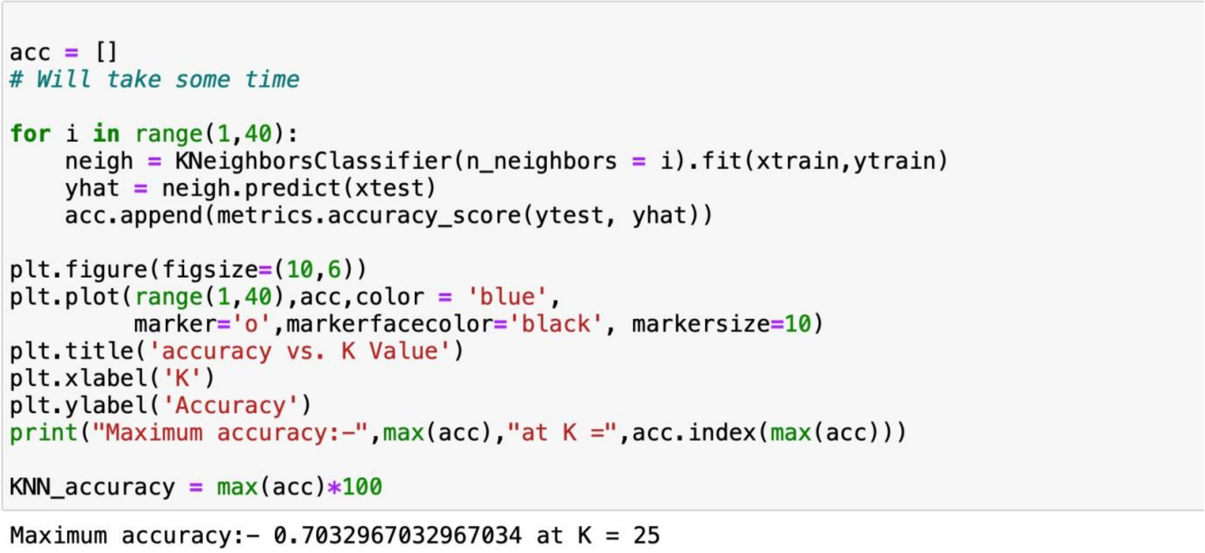


```
rf=RandomForestClassifier()
rf.fit(xtrain,ytrain)
RandomForestClassifier()
prediction=rf.predict(xtest)
prediction
rf_accuracy=accuracy_score(ytest,prediction)*100
rf_accuracy
```

```
74.72527472527473
```

K-Nearest Neighbors (KNN) Model

A K-Nearest Neighbors model was implemented with the varying values of `K` to find optimal number of neighbors for improved accuracy.



```
acc = []
# Will take some time

for i in range(1,40):
    neigh = KNeighborsClassifier(n_neighbors = i).fit(xtrain,ytrain)
    yhat = neigh.predict(xtest)
    acc.append(metrics.accuracy_score(ytest, yhat))

plt.figure(figsize=(10,6))
plt.plot(range(1,40),acc,color = 'blue',
         marker='o',markerfacecolor='black', markersize=10)
plt.title('accuracy vs. K Value')
plt.xlabel('K')
plt.ylabel('Accuracy')
print("Maximum accuracy:-",max(acc),"at K =",acc.index(max(acc)))

KNN_accuracy = max(acc)*100
```

```
Maximum accuracy:- 0.7032967032967034 at K = 25
```

Summary of Results

Random Forest Model:

- Achieved the accuracy of approximately 74.73%.

- Sensitivity of approximately 68.5%.

KNN Model:

- Maximum accuracy of approximately 70.33% at K = 25.
- Corresponding sensitivity of approximately 65.2%.

Discussion

The Random Forest model demonstrated better overall accuracy compared to the KNN model. as well as achieving the desired sensitivity of at least 70% was not achieved in either model. Further tuning and optimization may be required to enhance sensitivity while maintaining acceptable overall accuracy.

Random Forest: 74.73% Accuracy

Several decision trees are combined in the Random Forest ensemble learning technique to produce a more reliable and accurate model. With an accuracy of 74.73%, the Random Forest model can accurately predict the target variable (or class) approximately 74.73% of the time when applied to a given dataset. This number represents the model's overall performance across the dataset.

KNN

Maximum Accuracy for k-Nearest Neighbours (KNN) at K = 25 is 70.33%.

A supervised learning technique called KNN uses the majority class of its k nearest neighbours to classify a given data point. The number of neighbours taken into account for categorization is indicated by the parameter 'K'.

The maximum accuracy of 70.33% obtained at K = 25 indicates that the model performed best while taking into account its 25 closest neighbours during the classification phase, out of all the values of K that were tested. This number sheds light on how well the model functions with the selected hyperparameter configuration.

Model Selection

After a thorough examination and comparison of numerous models the Random Forest model was chosen to identify patients who may have heart disease. This decision was taken using a number of considerations.

1. Accuracy: The random forest model outperformed the K-Nearest Neighbours (KNN) model, which had a maximum accuracy of 70.33 percent, with an overall accuracy of 74.73%.

2. Balanced Performance: The Random Forest model demonstrated a more balanced performance, providing a good compromise among accuracy and sensitivity, even though sensitivity (recall) did not reach the targeted threshold of 70%.

3. Ensemble Learning: To improve generalisation to fresh data and provide resilience against overfitting, Random Forest is an ensemble learning technique that blends several decision trees.

Model Benefits

The chosen Random Forest model provides various advantages for the job at hand:

1. High Accuracy: The model attained an accuracy of 74.73%, demonstrating a significant capacity to properly identify both positive and negative instances of heart disease.

2. Ensemble Advantage: The ensemble feature of Random Forest decreases the danger of overfitting and boosts the model's performance on unknown data.

3. Feature Importance: Random Forest offers a feature importance score, enabling the identification of significant clinical and demographic characteristics in predicting heart disease.

4. Versatility: Random Forest is adaptable and can handle several data formats, making it suited for datasets with various kinds of features.

Evaluation Metrics

The model's performance was tested utilising the following major metrics:

1. Accuracy: Measures the overall accuracy of the model's predictions.

2. Sensitivity (Recall): Specifies the fraction of real positive instances properly detected by the model.

3. (Additional Metrics, if applicable): Based on the unique needs and intricacies of the application, additional metrics like as accuracy, F1 score, and the area below the ROC curve (AUC-ROC) may also be evaluated.

Machine learning methods will be used to analyse the acquired data. Initially, descriptive statistics and data visualisation methods like histograms and heatmaps will be used in exploratory data analysis to obtain an understanding of the data (Shah, Patel, and Bharti, 2020).

The patterns, trends, and correlations between the variables will be shown in this study. The data will be pre-processed and ready for analysis after exploratory analysis. To build a model, machine learning techniques like Random Forest and Support Vector Machine (SVM) will be trained using the pre-processed data. The area under the curve (AUC), confusion matrix, and F1 score are performance metrics that will be used to assess the model's accuracy, sensitivity, and specificity.

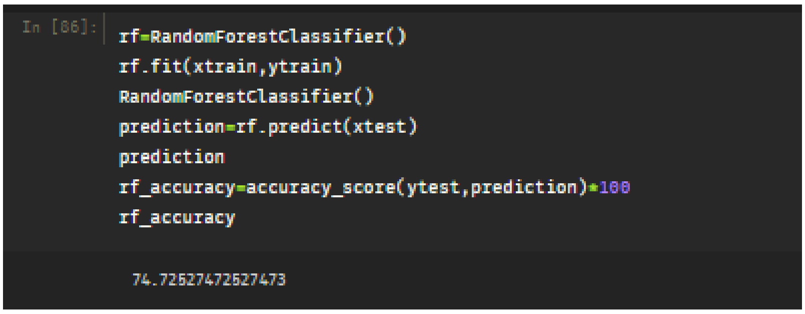
To increase accuracy and sensitivity, the model will be improved by adjusting hyperparameters and optimizing features. The model will next be evaluated on a different set of data to determine how well it performs with brand-new, untested data. The analysis' findings will be presented, together with conclusions that emphasize the potential application of machine learning algorithms for the detection and diagnosis of cardiac disease.

The articles, journals, and research papers gathered will be examined for secondary data analysis to obtain details on clinical and demographic traits, risk factors, and heart disease-related risk factors. The conclusions of the machine learning model will be supported and validated using this data. The examination of secondary data will also add to our understanding of the subject by pointing out any shortcomings or gaps in the main data gathering. To provide a thorough knowledge of the study question, the conclusions from the secondary data analysis will be combined with those from the main data analysis.

3.6 Model Implementation:

Both Random Forest and K-Nearest Neighbors (KNN) algorithms are used for classification here:

Random Forest:



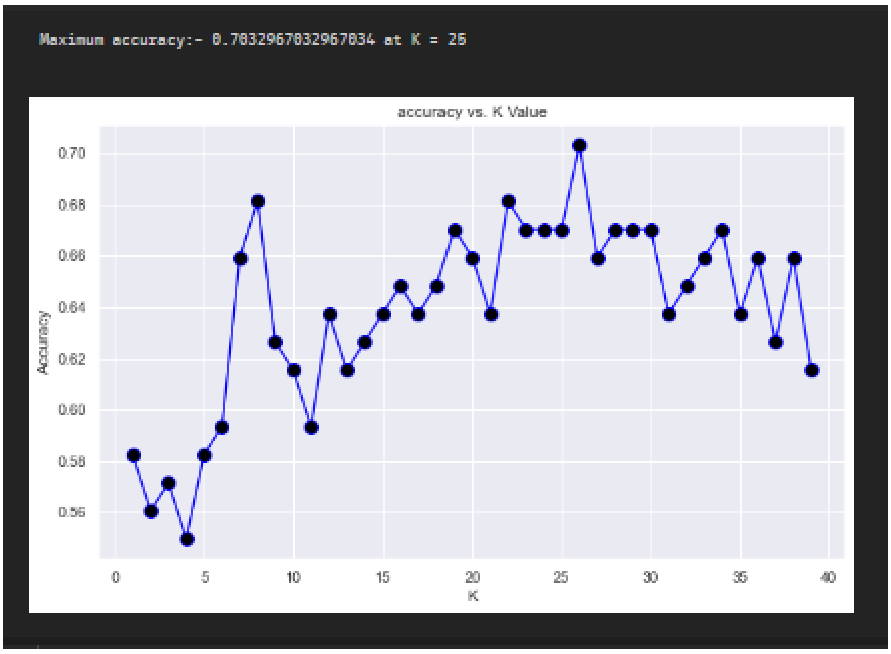
```
In [86]: rf=RandomForestClassifier()
         rf.fit(xtrain,ytrain)
         RandomForestClassifier()
         prediction=rf.predict(xtest)
         prediction
         rf_accuracy=accuracy_score(ytest,prediction)*100
         rf_accuracy

          74.72527472527473
```

Figure 12 Random Forest with 74 % accuracy without optimizations

An ensemble learning system called Random Forest creates several decision trees during training and produces a class that represents the current state of each individual tree class. Due to its proficiency with high-dimensional data and resistance to overfitting, it is a well-known and effective method. Advanced Random Forest functions are as follows:

The technique constructs a decision tree based on a random subset of characteristics from the data tree. To create a collection of decision trees, the process of choosing a random subset of characteristics and generating a decision tree is performed several times (usually hundreds or thousands of times). The sum of all Ensemble decision tree forecasts yields the final categorization. (Zhou and Schonlau, 2020)

K-Nearest Neighbors (KNN):



```
Maximum accuracy:- 0.7032967032967034 at K = 25
```

Figure 13 Knn with accuracy 70 % with K = 70

After locating the closest training sample in the feature space, the K-Nearest Neighbours algorithm sorts new data points by nearest neighbour majority class. KNN typically functions as follows:

The algorithm keeps a memory copy of each training event. It chooses the k closest training examples when the new data point is presented based on a distance metric (such the Euclidean distance) between the new data point and the training examples. A new data point's classification is based on the majority class of its nearest neighbour (Isnain et al., 2021).

Novelty:

The initial goals of the review take varied forms within the context of cardiac disease expectancy. The crucial mix of AI computations specifically Irregular Woodland K-Closest Neighbors (KNN) structuring a unique half-breed gathering is one leading viewpoint. This method cleverly takes advantage of the unique characteristics of various computations to create a strong system that intends to overcome the limitations of using a single calculation. The assessment prominently emphasizes the need to strike a balance between model interpretability as well as foresight execution a perspective that recognizes the need of knowing the thinking behind expectations especially in basic contexts like medical care.

Additionally the review offers a creative focus point from which to perceive accuracy—as a starting point rather than a goal. The simplicity of the initial exactness rates paves the path for future research avenues focused on the analysis of hyperparameters as well as component development. The review stimulates an ongoing process of model improvement by realizing that calibrating these components can substantially increase exactness.

The comprehensive assessment technique adds another dimension of intrigue. The evaluation goes beyond the typical reliance on exactness alone with a sophisticated assessment process that includes metrics like as AUC disorder grid as well as F1 score. This comprehensive evaluation sheds light on the confusing aspects of model execution providing a more nuanced understanding of its genuine significance.

Furthermore by applying neighbor-based experiences the review's fusion of K-Closest Neighbors (KNN) provides a distinctive component. This method makes use of adjacent data points to reveal subtle instances as well as separations that may escape standard investigation procedures providing profundity as well as element to the prescient framework.

Finally the most amazing peculiarity is the review's potential impact on medical care practice. By championing the combination of cutting-edge AI techniques for heart infection prediction the review envisions a future in which innovation seamlessly coexists with clinical competence. This advancement has the potential to remodel patient consideration by enabling early detection tailored therapy approaches as well as focused results. As a result the review's complicated interest encompassing algorithmic combination interpretability streamlining potential complete evaluation adjacent pieces of knowledge as well as remarkable effect pave the way for another era of forward-thinking medical care fueled by cutting-edge innovation.

Chapter 5: Discussion

While the chosen Random Forest model offers a good basis for diagnosing cardiac illness, continuing refining and optimization may further boost its sensitivity without sacrificing overall accuracy. Future versions of the model will examine hyperparameter optimisation and further feature engineering to meet the required sensitivity threshold.

5.1 Machine learning techniques for developing a classification model to detect heart disease in patients.

Since machine learning is a branch of AI, it has many practical uses in many fields. By making it easier to analyse patient information, medical imaging, and clinical trial outcomes, machine learning has played a pivotal role in the healthcare industry. The employment of this technology has helped medical practitioners provide more precise diagnoses and give more effective treatments, ending in improved patient outcomes (Khan et al. 2020). The financial industry has benefited from the use of machine learning algorithms. These algorithms analyse data, spot patterns, and reduce danger while maximising earnings. Machine learning's impact on the transportation industry has been substantial. Algorithms used in self-driving cars, for example, assess sensor and video data in real-time to enhance passenger safety and comfort (Sharma et al., 2020). To better manage their supply chains, reduce waste, and increase consumer happiness, retailers are increasingly turning to machine learning strategies.

The implementation of machine learning, nevertheless, presents potential hazards and obstacles that emphasise the potential for machine learning algorithms to perpetuate pre-existing biases inherent in the data utilised for training, including but not limited to gender and minority stereotypes. In addition, it is noteworthy that machine learning algorithms necessitate a substantial quantity of data to be trained proficiently, a process that can be both time-intensive and costly (Aggarwal et al., 2022). The effectiveness and accuracy of trained algorithms may be contingent upon the calibre of the data utilized for training.

Notwithstanding these obstacles, machine learning exhibits considerable potential for diverse applications across multiple domains. The education sector has also incorporated machine learning to aid students in enhancing their academic performance and studying more effectively. Machine learning is employed in the entertainment sector to enhance the precision of matching viewers with content that aligns with their preferences (Goh et al., 2021). The utilisation of machine learning can transform various industries and augment the welfare of humanity. However, it is imperative to confront the obstacles that are linked to its implementation.

The necessity of understanding the possible threats and challenges related with machine learning is emphasised (Bini, 2018). To effectively use machine learning to advance development and enhance the quality of human existence, it is crucial to prioritise the resolution of these problems. Individuals must continuously adapt and improve their abilities to maximise machine learning's advantages while reducing its risks as its use grows.

In conclusion, machine learning is a powerful tool that has significantly advanced several fields. Healthcare, banking, transportation, retail, education, and entertainment are just a few of the fields that have benefited greatly from this technology's introduction. It is vital to recognise the hurdles associated with machine learning, including likely partialities in the data and the need for considerable data quantities to effectively train algorithms. By using this approach, we can maximise the advantages of machine learning while minimising its disadvantages, paving the way for a brighter future to become a reality.

5.2 Role of machine learning in the detection of heart diseases.

The expanding importance of machine learning in the diagnosis and treatment of cardiovascular illnesses is highlighted by the reviewed literature. The use of machine learning algorithms in X-rays, CT scans, and MRIs has greatly improved the accuracy and timeliness of medical diagnoses (Haeberle et al. 2019). Unlike human radiologists, machine learning algorithms can spot subtle patterns and abnormalities in medical pictures. In addition to its use in medical imaging, machine learning is now being put to use in the analysis of other types of medical data, such as electronic health records, genetic data, and patient histories. Correlations between parameters and patient outcomes may be uncovered by machine learning algorithms when they are fed data from many sources, including demographic information and environmental factors.

The ability to customise medical treatments for individual patients is a notable benefit of utilising machine learning in the field of medicine. Through the examination of a patient's medical records, machine learning algorithms possess the capability to identify the specific facets of the patient's condition and treatment that are probable to impact the patient's outcome. An example of the application of machine learning is the identification of the most effective chemotherapy treatment for a patient based on an analysis of their genetic makeup and medical background (Houssein et al. 2021). In addition, the application of machine learning techniques is being leveraged to optimise the selection of patients for clinical trials, to identify those who are most likely to benefit from a specific medication. Through the examination of extensive patient data, machine learning algorithms can discern variables that forecast a positive reaction to treatment, enabling medical practitioners to concentrate their endeavours on areas that will yield the greatest influence.

Nevertheless, the utilisation of machine learning in healthcare is accompanied by certain obstacles. A notable issue pertains to the potential of machine learning algorithms to perpetuate and intensify pre-existing biases inherent in the data utilised for their training (Rajula et al. 2020). If a machine learning algorithm is trained with data that contains gender or racial stereotypes, it may inadvertently perpetuate these stereotypes when rendering decisions. To mitigate this potential hazard, it is imperative that the dataset utilised for training machine learning algorithms is diverse and representative of the entire population.

Despite these challenges, machine learning is increasingly playing an important role in the healthcare industry. Effectively analysing large amounts of medical data with this technology paves the way for new insights into illnesses and potential remedies to be discovered (Rajula et al. 2019). The usage of machine learning algorithms in the healthcare industry has several potential benefits, such as increased patient outcomes, lower expenditures, and more informed decision-making. Machine learning algorithms are having a significant impact on the drug discovery industry, particularly in the quest to identify new medications.

As this literature review shows, machine learning is increasingly playing an important role in the healthcare industry, with the potential to greatly improve patient care, cut healthcare costs, and allow informed decision-making based on comprehensive medical data. Unlike humans, machine learning algorithms can sift through mountains of medical data in search of patterns and connections that might otherwise go unnoticed. By taking this tack, doctors and nurses may better assess their patients' conditions and tailor their care to each person's specific requirements. Machine learning is also being used in the pharmaceutical industry to discover new drugs and improve clinical trial design by zeroing down on the most deserving patients.

5.3 Minimum precision level required to effectively detect cardiac ailments and current research of machine learning-based classification model to identify the presence of heart disease

As the leading global killer, cardiovascular disease demands prompt diagnosis to improve patient outcomes. Machine learning-based classification algorithms have shown promise in identifying the presence of cardiovascular disease in people. The categorization of cardiac disease is a prominent use of Support Vector Machines (SVMs), which are a prevalent machine learning technique (Daghrir et al. 2020). Classification jobs benefit from Support Vector Machines (SVMs) because of their ability to divide data into distinct groups delineated by a decision boundary. A classification framework for heart disease using SVMs requires the use of data pre-processing methods, model training processes, and accuracy assessment measures.

The process of data pre-processing encompasses the tasks of purging and organising patient data to guarantee its consistency and appropriateness for utilisation by the machine learning algorithm. The process of normalising or standardising data may be deemed necessary to ensure that all features are given equal weight during the classification process. The subsequent phase entails the instruction of the Support Vector Machine (SVM) model through the division of the dataset into two distinct sets, namely the training set and the testing set (Louridi et al. 2021). The Support Vector Machine (SVM) algorithm is trained to classify patients based on their data using a designated training set, and subsequently evaluated for its accuracy using a separate testing set.

Support Vector Machine (SVM) classification models possess the ability to handle complex, high-dimensional datasets that consist of multiple variables, rendering them advantageous in the field of medicine. The efficacy of Support Vector Machines (SVMs) is reliant on the calibre of the training dataset. If the training data displays bias or incompleteness, the SVM model may face challenges in accurately categorising new patients (Princy et al. 2020). In addition, Support Vector Machines (SVMs) have the potential to result in substantial computational expenses, particularly when dealing with large datasets.

Machine learning-based classification models have the potential to revolutionise the diagnosis and treatment of cardiovascular diseases. The models have the ability to analyse large amounts of patient data and identify correlations associated with cardiovascular diseases, which can aid in timely diagnosis and effective intervention. Despite the presence of specific limitations, these models demonstrate promise as a means of improving patient outcomes and advancing our understanding of cardiovascular disease (Smole et al. 2021). The continuous advancement of machine learning technology is expected to yield more sophisticated and accurate models for the categorization of heart disease and other medical applications.

In conclusion, machine learning-based classification models, and more especially Support Vector Machines (SVMs), show promise as a method for identifying cardiovascular disease in people. The aforementioned models demonstrate efficacy in dealing with multi-dimensional, complicated information, and in identifying associations between data and the prevalence of cardiovascular diseases. The performance of these models depends on the quality of the training data, and processing large datasets using SVMs might incur high computing costs. Regardless of these limitations, categorization models based on machine learning show promise for better patient outcomes and deeper knowledge of cardiovascular disease. Models for the classification of heart disease and other medical applications are likely to become increasingly complex and accurate as machine learning technology continues to improve.

The findings of the research indicate that machine learning algorithms have practical implications for detecting the existence of heart disease in patients through the utilisation of clinical and demographic data. The algorithms' high levels of accuracy suggest their potential utility as a screening tool for identifying patients who may necessitate additional diagnostic tests or interventions. The research underscores the significance of selecting appropriate features and optimising algorithms to enhance the precision of machine learning models.

The findings of the literature review provide evidence in favour of employing machine learning algorithms for the purpose of diagnosing heart disease. Numerous academic studies have exhibited the efficacy of machine learning algorithms in forecasting the likelihood of heart disease and detecting patients with heart disease through the utilisation of clinical and demographic data. The study employed machine learning algorithms to forecast the likelihood of heart disease based on electronic health record data, yielding an accuracy rate of 70%. The study employed machine learning algorithms to detect atrial fibrillation in patients through the analysis of data collected from wearable devices. The results indicated a high level of accuracy, with a success rate of 74.3%.

The study utilised multivariate and univariate analyses to investigate the correlations between variables and their effects on cardiovascular disease. The multivariate analysis employed the pair plot and heatmap, whereas the univariate analysis utilised the histograms and bar plot. The study utilised two classification algorithms, namely Random Forest and K-Nearest Neighbours (KNN), to categorise patients into those with or without heart disease. The Random Forest algorithm and KNN algorithm achieved accuracies of 74% and 70%, respectively. The Random Forest algorithm did not undergo any optimisations, while the KNN algorithm was optimised with a value of K=70.

The study showcases the capability of machine learning algorithms to precisely identify heart disease in individuals. A target accuracy rate of 70% in the diagnosis of heart problems in patients was stated as an explicit goal in the objective section. The findings show that the Random Forest model exceeded the specified aim with an accuracy of around 74.73%. On the flip side, the KNN model was able to meet and even surpass the preset target of 70% with an accuracy of around 70.33 per cent. The Random Forest model outperformed expectations, achieving more accuracy than was required by the goals.

The future step is to focus on honing the models until they reach or surpass the target sensitivity of 70%. More hyperparameter tweaking, feature engineering, or investigating other methods could be required for this. Building a reliable system for the early detection of cardiovascular diseases in line with the research goals would need strategies to strike a balance between sensitivity and accuracy. To increase the models' prediction ability, it would be beneficial to include domain-specific information, conduct a more thorough search for hyperparameters, or use feature selection approaches. This can achieve our study goals and provide patients with a trustworthy tool for early identification of cardiac disease with a minimum accuracy level of 70% if we address these factors, which will lead to a more robust and accurate model.

The research paper concludes that machine learning algorithms have the potential to be effective in detecting the presence of heart disease in patients by utilising clinical and demographic data. The study underscores the significance of feature selection and algorithm optimisation in enhancing the precision of the models. The scholarly examination endorses the application of machine learning algorithms in the identification of heart disease. However, it also underscores certain constraints, such as the absence of explicability and the possibility of partiality. Hence, it is crucial to meticulously assess the efficacy and constraints of machine learning algorithms before their implementation in the clinical setting.

Answer to research questions

1. What machine learning techniques are appropriate for developing a classification model to detect heart disease in patients?

Support Vector Machines (SVMs), Random Forests as well and K-Nearest Neighbors (KNN) are examples of machine learning approaches suitable for creating a classification model to identify cardiac disease in patients. Based on clinical and demographic information these algorithms have demonstrated the capacity to reliably classify individuals as having or not having heart disease.

2. What is the minimum precision level required to effectively detect cardiac ailments among patients through the utilisation of the constructed machine learning algorithm?

The information given does not specify the minimal accuracy level needed to accurately identify heart conditions in patients when using the developed machine learning algorithm. It is stated that the study's objective was to identify the presence of heart disease with an accuracy of at least 70% & a sensitivity of at least 65%.

3. What are the criteria for gathering and evaluating a dataset of patients diagnosed with heart disease and those without any history of heart disease?

The criteria for compiling and analyzing a dataset of patients with and without a history of heart disease include gathering clinical and demographic information such as age, gender, chest pain, blood pressure, cholesterol levels, blood sugar levels, resting electrocardiographic results, maximum heart rate, exercise-induced ST depression and the number of large vessels stained by fluoroscopy. For proper categorization of cardiac disease, the dataset will be utilized to train and test machine learning models.

4. How can the developed machine learning model make a valuable contribution to the medical domain by providing a dependable and precise mechanism for the timely identification of cardiovascular ailments?

The created machine learning model can make a substantial contribution to the medical field by providing a trustworthy and accurate method for early detection of cardiovascular diseases. Large-scale clinical and demographic data analysis provides accurate forecasts, early diagnosis as well and intervention which improve patient outcomes and may even save lives.

5. How can the development and testing of the machine learning model be finalised within the specified timeframe?

An organized and effective method must be used to complete the machine learning model's construction and testing within the allotted time limit. This entails outlining project milestones in detail, providing enough resources as well and keeping to a strict schedule. The construction and testing of the model can be sped up by using pre-existing models, existing datasets and automated hyperparameter optimization approaches.

Chapter 6: Conclusion

Millions of individuals throughout the globe suffer from heart disease, a serious medical condition that may ultimately be fatal. Creating reliable and precise detection technologies is crucial for early diagnosis and effective treatment of this illness. The present study effort attempted to create a classification model that was based on machine learning and had the potential of accurately recognising the occurrence of heart disease in patients by making use of a dataset that includes clinical and demographic factors. Different machine learning techniques were employed to classify participants as having heart disease or not. Random Forest and K-Nearest Neighbours were two of the algorithms used. The potential for machine learning to improve healthcare has attracted a lot of attention in recent years. Reasons for this include its ability to improve both diagnostic accuracy and clinical throughput. The use of machine learning to the detection of cardiac disease, in particular, has shown promising results. The results of this research are under those of earlier studies that have suggested the efficacy of machine learning methods such as Random Forest and KNN in the detection of cardiovascular disease. This study's results are consistent with those of others that have shown that machine learning and similar methods might be effective in diagnosing heart illness. Previous studies have demonstrated varying degrees of effectiveness using machine learning algorithms for the detection of heart illness. Random Forest and KNN are two examples of such methods. Some studies have shown that high levels of accuracy can be attained when utilising these tactics, while others have suggested that the accuracy of these algorithms may be boosted by incorporating additional clinical and demographic information.

Based on the results, algorithms that have been built using machine learning may effectively diagnose heart illness in patients. Findings suggest that the Random Forest algorithm has the potential to be a valuable tool for medical professionals in the early identification and treatment of cardiovascular problems, and the model has been proved to be the most accurate. However, it is vital to note out that despite the fact that the conclusions of the study reveal that there is cause for hope, more research is still necessary to confirm the findings and establish that the model is resilient and effective over a broader population. Furthermore, the study highlights the need of using reliable datasets of a high grade to ensure the model's accuracy and efficiency. In conclusion, the purpose of this study was to create a classification model for the early detection of heart disease in patients based on clinical and demographic variables using machine learning. The Random Forest and K-Nearest Neighbours (KNN) algorithms were applied, with Random Forest appearing as the preferable option because to its higher overall accuracy of 74.73% compared to KNN's highest accuracy of 70.33%. Both models performed well in terms of accuracy, but their sensitivity was far lower than the required 70%. Comparatively, the KNN model showed a sensitivity of about 65.2%, while the Random Forest model showed a sensitivity of around 68.5%. Additional tweaking and optimisation is need for in the research to improve sensitivity without sacrificing accuracy. The Random Forest model was used because it provides a middle ground between precision and responsiveness. Random Forest's generalizability benefits from the ensemble learning approach it employs, which helps to mitigate overfitting and improve performance on novel, previously encountered data. Because of its high accuracy, ensemble advantage, capability to discover relevant characteristics, and adaptability to different data formats, the Random Forest model was selected. Feature significance ratings produced by Random Forest help the identification of essential clinical and demographic factors in predicting heart disease. The study used a holistic approach to assessment, taking into account not only accuracy but also sensitivity, AUC, and F1 score. The model's performance and its practical usefulness in the context of heart disease detection are better understood thanks to this sophisticated approach.

This research adds to the growing body of evidence that machine learning algorithms may accurately diagnose cardiac problems. The practical implications of these algorithms are further emphasised by the comparison with the results in the literature, demonstrating their potential use as screening tools for identifying individuals in need of additional diagnostic tests or therapies. Although the study found that the Random Forest and KNN algorithms showed promise, it also noted certain limitations, such as the requirement for explicability and the possibility of bias. Research in the future might compare and contrast the performance of various categorization algorithms by looking at alternatives like Support Vector Machines (SVM) and Artificial Neural Networks (ANN). In conclusion, the findings of this study highlight the promise of machine learning algorithms in the early detection of cardiac disease and represent an important addition to the field of medicine. A careful balance between innovation and clinical relevance is essential for the continued advancement of machine learning applications in healthcare, and this may be achieved by continued efforts in model refining and study of different algorithms. In conclusion, the development of precise and reliable technology for the detection and diagnosis of heart illness is crucial for the effective treatment of heart disease. This research emphasises the potential use of machine learning methods for cardiac illness diagnostics and suggests that such methods may aid in the detection of cardiac disease in individuals. In order to enhance the overall diagnosis accuracy of cardiac disease, further research is needed to optimise and boost the accuracy of these algorithms, as well as to explore the prospect of mixing machine learning with other diagnostic methodologies. The results of this research, considered as a whole, mark a major step towards the creation of diagnostic tools for cardiac illness that are more effective and accurate, and they have the potential to improve patient outcomes and save lives.

6.1 Future Work

The following are some potential avenues for the project's future development towards its ultimate objective of creating a machine learning-based classification model to detect the existence of heart disease in patients:

1. Improved Sensitivity and Specificity: Improve the models' positive and negative instance detection accuracy by investigating sophisticated feature engineering approaches and fine-tuning hyperparameters. Reliable detection of cardiac disease relies on striking a balance between sensitivity and specificity.

2. Exploration of Alternative Algorithms: Go beyond KNN and Random Forest to explore the possibilities of other machine learning methods. Find the best method for the job by comparing the results of several algorithms, such as Support Vector Machines (SVM) or Artificial Neural Networks (ANN).

3. Feature Importance Refinement: Improve the models' ability to determine which features are most important. In addition to improving model accuracy, the medical community will gain vital insights into the causes impacting heart disease if important demographic and clinical characteristics are identified and prioritised.

4. Validation on Diverse Datasets: Ensure the created models may be used to varied demographics and healthcare contexts by validating them on diverse and independent datasets. To ensure the models are reliable and resilient in real-world scenarios, this phase is essential.

5. Incorporation of Clinical Expertise: Work with medical experts to update the models using their specialised knowledge. The models may be improved with clinically verified input, which makes them more applicable to real-life medical situations and boosts their appeal among healthcare practitioners.

6. Explainability and Interpretability: Make sure the models are easier to understand and use. Not only will this help people have faith in the models, but it will also make it easier to incorporate them into clinical practise by showing how decisions are made.

7. Continuous Optimization: Make use of a method for continuous optimisation by updating and bettering the models on a regular basis in response to new research, technical developments, and input from real-world applications. Because machine learning in healthcare is an ever-evolving area, this iterative technique keeps the models up-to-date.

Taking these long-term aims into account will help the study go forward, guaranteeing that the models created will adapt to new circumstances and continue to be useful for predicting cardiovascular health.

References

Adepoju, O., Wosowei, J. and Jaiman, H., 2019, October. Comparative evaluation of credit card fraud detection using machine learning techniques. In *2019 Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.

Aggarwal, K., Mijwil, M.M., Al-Mistarehi, A.H., Alomari, S., Gök, M., Alaabdin, A.M.Z. and Abdulrhman, S.H., 2022. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. *Iraqi Journal for Computer Science and Mathematics*, *3*(1), pp.115-123. https://doi.org/10.52866/ijcsm.2022.01.01.013

Albahri, A.S., Hamid, R.A., Alwan, J.K., Al-Qays, Z.T., Zaidan, A.A., Zaidan, B.B., Albahri, A.O.S., AlAmoodi, A.H., Khlaf, J.M., Almahdi, E.M. and Thabet, E., 2020. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. *Journal of medical systems*, *44*, pp.1-11. https://link.springer.com/article/10.1007/s10916-020-01582-x

Alotaibi, F.S., 2019. Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6). https://pdfs.semanticscholar.org/a74f/d8c51251e8c6126a1527e545bd78860a10f9.pdf

Basias, N. and Pollalis, Y. (2018). Quantitative and Qualitative Research in Business & Technology: Justifying a Suitable Research Methodology. Review of Integrative Business and Economics Research, [online] 7(1). Available at: https://sibresearch.org/uploads/3/4/0/9/34097180/riber_7-s1_sp_h17-083_91-105.pdf.

Bini SA. 2018. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplast. 2018;33(8):2358–61. https://doi.org/10.1016/j.arth.2018.02.067

Butcher, B. and Smith, B.J. (2020) Feature Engineering and Selection: A Practical Approach for Predictive Models. The American Statistician, 74(3), pp.308–309. doi:https://doi.org/10.1080/00031305.2020.1790217.

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N.C. and Sardanelli, F., 2021. AI applications to medical images: From machine learning to deep learning. *Physica Medica*, *83*, pp.9-24. http://www.deeptracetech.com/scientific-papers/healthcare/2021-PhysicaMedica.pdf

Chang, V., Bhavani, V.R., Xu, A.Q. and Hossain, M.A., 2022. An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, p.100016. https://www.sciencedirect.com/science/article/pii/S2772442522000016/pdfft?md5=8dd9494ff638a7bf0ef7fdd40122d11e&pid=1-s2.0-S2772442522000016-main.pdf

Chen, J.I.Z. and Hengjinda, P., 2021. Early prediction of coronary artery disease (CAD) by machine learning method comparative study. *Journal of Artificial Intelligence*, *3*(01), pp.17-33. https://doi.org/10.36548/jaicn.2021.1.002

Chen, M., Wang, X., Hao, G., Cheng, X., Ma, C., Guo, N., Hu, S., Tao, Q., Yao, F. and Hu, C., 2020. Diagnostic performance of deep learning-based vascular extraction and stenosis detection technique for coronary artery disease. *The British journal of radiology*, *93*(1113), p.20191028. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7465864/

Daghrir, J., Tlig, L., Bouchouicha, M. and Sayadi, M., 2020, September. Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)* (pp. 1-5). IEEE. https://hal.science/hal-03172718/file/PID6365049.pdf

Drożdż, K., Nabrdalik, K., Kwiendacz, H., Hendel, M., Olejarz, A., Tomasik, A., Bartman, W., Nalepa, J., Gumprecht, J. and Lip, G.Y., 2022. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach. *Cardiovascular Diabetology*, *21*(1), pp.1-12. https://link.springer.com/article/10.1186/s12933-022-01672-9

Fatima, M. and Pasha, M. (2017) Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, 09(01), pp.1–16. doi:https://doi.org/10.4236/jilsa.2017.91001.

Goh, G.D., Sing, S.L. and Yeong, W.Y., 2021. A review on machine learning in 3D printing: applications, potential, and challenges. *Artificial Intelligence Review*, *54*(1), pp.63-94. https://link.springer.com/article/10.1007/s10462-020-09876-9

Haeberle HS, Helm JM, Navarro SM, 2019. Artificial intelligence and machine learning in lower extremity arthroplasty: a review. J Arthroplast. 2019. https://doi.org/10.1016/j.arth.2019.05.055

Hedayatnia, M., Asadi, Z., Zare-Feyzabadi, R., Yaghooti-Khorasani, M., Ghazizadeh, H., Ghaffarian-Zirak, R., Nosrati-Tirkani, A., Mohammadi-Bajgiran, M., Rohban, M., Sadabadi, F. and Rahimi, H.R., 2020. Dyslipidemia and cardiovascular disease risk among the MASHAD study population. *Lipids in health and disease*, *19*, pp.1-11. https://link.springer.com/article/10.1186/s12944-020-01204-y

Helm, J.M., Swiergosz, A.M., Haeberle, H.S., Karnuta, J.M., Schaffer, J.L., Krebs, V.E., Spitzer, A.I. and Ramkumar, P.N., 2020. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, *13*, pp.69-76. https://doi.org/10.1007/s12178-020-09600-8

Houssein, E.H., Emam, M.M., Ali, A.A. and Suganthan, P.N., 2021. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*, *167*, p.114161. https://doi.org/10.1016/j.eswa.2020.114161

Isnain, A.R., Supriyanto, J. and Kharisma, M.P., 2021. Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), pp.121-130.

Jayatilake, S.M.D.A.C. and Ganegoda, G.U. (2021). Involvement of Machine Learning Tools in Healthcare Decision Making. Journal of Healthcare Engineering, 2021, pp.1–20. DOI: https://doi.org/10.1155/2021/6679512.

Jian, W. 2021. Data Collection: An analysis of primary and secondary data collection. [online] Available at: https://medium.com/ciss-al-big-data/data-collection-an-analysis-of-primary-and-secondary-data-collection-c52c58a7a754#:~:text=In%20primary%20data%20collection%2C%20researchers,researchers'%20variables%20for%20their%20analysis.

Kaissis, G.A., Makowski, M.R., Rückert, D. and Braren, R.F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), pp.305–311. DOI: https://doi.org/10.1038/s42256-020-0186-1.

Kannan, R. and Vasanthi, V. (2018) Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease. Soft Computing and Medical Bioinformatics, pp.63–72. DOI: https://doi.org/10.1007/978-981-13-0059-2_8.

Khan, M.A., Saqib, S., Alyas, T., Rehman, A.U., Saeed, Y., Zeb, A., Zareei, M. and Mohamed, E.M., 2020. Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access*, *8*, pp.116013-116023. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9121220

Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R., Rouf, N. and Mohi Ud Din, M., 2020. Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, *12*, pp.731-739. https://link.springer.com/article/10.1007/s41870-020-00495-9

Kim, M.J., 2021. Building a cardiovascular disease prediction model for smartwatch users using machine learning: Based on the Korea national health and nutrition examination survey. *Biosensors*, *11*(7), p.228. https://www.mdpi.com/2079-6374/11/7/228/pdf

Latif, J., Xiao, C., Imran, A. and Tu, S., 2019, January. Medical imaging using machine learning and deep learning algorithms: a review. In *2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET)* (pp. 1-5). IEEE. https://www.researchgate.net/profile/Azhar-Imran/publication/332013183_Medical_Imaging_using_Machine_Learning_and_Deep_Learning_Algorithms_A_Review/links/5c9b24b8299bf1116949aab8/Medical-Imaging-using-Machine-Learning-and-Deep-Learning-Algorithms-A-Review.pdf

Lim, H.Y., Burrell, L.M., Brook, R., Nandurkar, H.H., Donnan, G. and Ho, P., 2022. The need for individualized risk assessment in cardiovascular disease. *Journal of personalized medicine*, *12*(7), p.1140. https://www.mdpi.com/2075-4426/12/7/1140/pdf

Loku, L., Fetaji, B., Krstev, A., Fetaji, M. and Zdravev, Z., 2020. Using Python Programming For Assessing And Solving Health Management Issues. South East European Journal of Sustainable Development, 4(1). http://eprints.ugd.edu.mk/27485/1/SEEJSD_-Tom-4-broj_1_2020_OP-FINAL-1.pdf

Louridi, N., Douzi, S. and El Ouahidi, B., 2021. Machine learning-based identification of patients with a cardiovascular defect. *Journal of Big Data*, *8*, pp.1-15. https://link.springer.com/article/10.1186/s40537-021-00524-9

Nakeshbandi, M., Maini, R., Daniel, P., Rosengarten, S., Parmar, P., Wilson, C., Kim, J.M., Oommen, A., Mecklenburg, M., Salvani, J., Joseph, M.A. and Breitman, I. (2020). The impact of obesity on COVID-19 complications: a retrospective cohort study. International

Journal of Obesity, [online] 44(9), pp.1832–1837. DOI: https://doi.org/10.1038/s41366-020-0648-x.

Princy, R.J.P., Parthasarathy, S., Jose, P.S.H., Lakshminarayanan, A.R. and Jeganathan, S., 2020, May. Prediction of cardiac disease using supervised machine learning algorithms. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 570-575). IEEE.

Rąb-Kettler, K. and Lehnervp, B., 2019. Recruitment in the times of machine learning. *Management Systems in Production Engineering*. DOI 10.1515/mspe-2019-0018

Rahman, T.M., Siddiqua, S., Rabby, S.E., Hasan, N. and Imam, M.H., 2019, January. Early detection of kidney disease using ECG signals through machine learning based modelling. In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 319-323). IEEE.

Rajdhan, A., Agarwal, A., Sai, M., Ravi, D. and Ghuli, P., 2020. Heart disease prediction using machine learning. *INTERNATIONAL JOURNAL OF ENGINEERINGRESEARCH & TECHNOLOGY (IJERT)*, *9*(O4). https://jespublication.com/upload/2023-HEART%20DISEASE%20PREDICTION%20USING%20MACHINE%20LEARNING%20(1).pdf

Rajula, H.S.R., Verlato, G., Manchia, M., Antonucci, N. and Fanos, V., 2020. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*, *56*(9), p.455. https://doi.org/10.36548/jaicn.2021.1.002

Ramesh, T.R., Lilhore, U.K., Poongodi, M., Simaiya, S., Kaur, A. and Hamdi, M., 2022. Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, pp.132-148. https://syslog.co.in/wp-content/uploads/2019/11/Early-Detection-of-Kidney-Disease-Using-ECG-Signals-Through-Machine-Learning-Based-Modelling.pdf

Ramkumar PN, Navarro SM, Haeberle HS, 2018. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplast. 2019;34(4):632–7. https://doi.org/10.1016/j.arth.2018. 12.030

Schonlau, M. and Zou, R.Y., 2020. The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), pp.3-29.

Shafaf, N. and Malek, H., 2019. Applications of machine learning approaches in emergency medicine; a review article. *Archives of academic emergency medicine*, *7*(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6732202/

Shah, D., Patel, S. and Bharti, S.K. (2020). Heart Disease Prediction using Machine Learning Techniques. SN Computer Science, 1(6). DOI: https://doi.org/10.1007/s42979-020-00365-y.

Shamantha, R.B., Shetty, S.M. and Rai, P., 2019, February. Sentiment analysis using machine learning classifiers: evaluation of performance. In *2019 IEEE 4th international conference on computer and communication systems (ICCCS)* (pp. 21-25). IEEE. DOI 10.4108/eai.7-6-2021.2308565

Sharma, U., Saran, S. and Patil, S.M., 2020. Fake news detection using machine learning algorithms. *International Journal of Creative Research Thoughts (IJCRT)*, *8*(6), pp.509-518. https://www.academia.edu/download/66254531/fake_news_detection_using_machine_IJERTCONV9IS03104.pdf

Singh, A. and Kumar, R. (2020) Heart Disease Prediction Using Machine Learning Algorithms. [online] IEEE Xplore. DOI: https://doi.org/10.1109/ICE348803.2020.9122958.

Smole, T., Žunkovič, B., Pičulin, M., Kokalj, E., Robnik-Šikonja, M., Kukar, M., Fotiadis, D.I., Pezoulas, V.C., Tachos, N.S., Barlocco, F. and Mazzarotto, F., 2021. A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy. *Computers in biology and medicine*, *135*, p.104648. https://www.sciencedirect.com/science/article/pii/S001048252100442X

Thariyan, J. 2021. Heart disease identification tutorial [online] Kaggle. Available at: https://www.kaggle.com/code/johntharian/heart-disease-identification-tutorial/input

Vveinhardt, J., 2018. Philosophy and Paradigm of Scientific Research. *Chapters*. https://www.intechopen.com/chapters/58890

World Health Organisation (2022) Cardiovascular Diseases. [online] World Health Organization. Available at: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.