Early Cardio Vascular Disease Detection Using Machine Learning and Explainable AI

Author: Bhargava Dandu

A thesis submitted in partial fulfilment of the requirements

for the degree of

M.Sc Data Analytics

Supervisor:

Dr. Zohaib Ijaz (Dundalk Institute of Technology)

August 2025

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Sc Data Analytics is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signature: Bhargava Dandu

August 2025

ABSTRACT

Cardiovascular diseases are the leading causes of the death worl wide by contributing 32% of deaths world wide. In Majority of the fatal cardiovascular disease cases, the medical industry experts identify the symptoms at the final stage. So, early detection of the cardiovascular disease are crutial. In this research we use Machine Learning (ML) and Explainable AI(XAI) to identify which parameters to be considered to identify the disease in early stages. Using a dataset that we have obtained from UCI machine learning repository, we have implemented Random forest classifier to predict the cardio vascular disease of a patient and achieved a model performance of 62%. We further tuned the model with Ml techniques and created 2 additional models with a better model accuracy with 80% of F1 score for the Gradient boosting classifier.The novelty of this research is how we handled the NULL values and creating features that will help ML models to discover deeper trends in the dataset. The Explainable AI(XAI) techniques are used to understand the decision making of the model and factors influencing the predictions. The future work of this research is refining the classifications based on the newly created features and further classifying the severity of the heart disease of the patient using them.

Keywords- Cardio Vascular Diseases(CVDs), SHapley Additive Predictions(SHAP), Machine Learning (ML), Eplainable AI (XAI), Local Interpretable Model-agnostic Explanations (LIME), Gradient Boosting, Precision, Recall, F-1 Score.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to our supervisors, Dr. Zohaib Ijaz and Dr. Abhishek Kaushik, for their invaluable guidance and support throughout this research. Their expertise and insights were instrumental in shaping the research questions and developing a well-structured methodology. The clarity and depth of their feedback challenged me to refine my ideas, think critically, and approach problems from multiple perspectives, significantly enhancing the overall quality of this work.

I am particularly grateful for their patience and willingness to provide detailed advice at every stage of the project. Their encouragement and constructive criticism not only helped me overcome challenges but also inspired me to maintain high academic and research standards. The knowledge and skills I gained under their mentorship have contributed greatly to my professional growth and have strengthened my confidence in conducting independent research.

I also appreciate the time and effort they dedicated to reviewing drafts, guiding experimental design, and offering practical suggestions that improved the rigor and clarity of the study. This research would not have reached its current level without their continuous support, guidance, and mentorship. I am deeply thankful for their role in making this academic journey both enriching and successful.

Git hub repository: GIT HUB Project link

Table of Contents

List of Tables

List of Figures

ABBREVIATIONS

- **ECG**: Electrocardiogram.
- **CVD**: Cardio Vascular Disease.
- **HDD**: Heart Disease Dataset.
- **DL**: Deep Learning.
- **XAI**: Explainable AI.
- **ML**: Machine Learning.
- **SHAP**: SHapeley Additive exPlanations.
- **LIME**: Local Interpretable Model-Agnostic Explanations.
- **SMOTE**: Synthetic Minority Oversampling Techniques.
- **FCBF**: Fast Correlation Based Filter.
- **MRMR**: Minimum Redundancy Maximum Relevance.
- **EHR**: Electronic Health Records.
- **ASCVD**: Atherosclerotic Cardio Vascular Diseases.
- **CAD**: Coronary Artery Disease.
- **MI**: Myocardial Infarction (Heart Attack).
- **CHF**: Congestive Heart Failure.

Chapter 1

Introduction

As per statistics of World Health Organization (WHO), Cardiovascular diseases (CVDs) cause 32% of global deaths worldwide and 28.1% deaths in India. The severity of the patient depends on the number of years he is taking to recover after getting diagnosed. Studies show 85% of the patients recover within one year of treatment, but only 55% of the patients recover after two years of treatment, and only 33% of patients revover within 5 years of treatment [1].

Cardiovascular diseases (CVDs) are one of the major causes of the deaths worldwide. Using Machine learning models many researchers developed complex models to identity early signs of heart diseases in patients. But, the black box models are very complex to interpret and we cannot the reason behind the prediction. In such situations, even though we develop a model with high accuracy, medical industry experts would lack trust on it, because the false positives and false negatives are so valuable in medical industry. So, to overcome this problem statement, we are trying to incorporate Explainable AI techniques (XAI) to make the models transparent. If the models are transparent and we are able to interpret their predictions then medical industry experts would develop trust, confidence in their decision making by using our models.

Artificial Intelligence (AI) and Machine Learning (ML) technologies are reshaping the medical industry every day. So, in this research, we aim to use Machine Learning techniques and Explainable AI (XAI) techniques for early detection of cardiovascular diseases. We implement the ML techniques on the dataset obtained from UCI machine learning repository and further apply XAI techniques for further explanations of model's predictions. These techniques clearly outline which feature is influencing the model's prediction which makes the decision making process more transparent and reliable.

In our research, we are planning to incorporate widely used XAI frameworks like SHAP (SHapeley Additive exPlanations) and LIME (Local Interpretable Model-agnositic Explanations). These tools helps us to interpret the model's classification by highlighting the individual feature contributions towards the model outcome. This will provide transparency in our model's classifications and help us to gain trust from medical industry experts [2]. By incorporating XAI techniques, we can see transparency in results. This will help the medical industry experts to understand which parameters contribute more for final classification.

1.1 Background

In medical field electrocardiogram provides accurate metrics of the heart rate. It is the electrical activity of the heart[3]. The ECG values cannot be tracked for all the individuals. There are other factors that we can be considered to tell if the person can be suffering from heart disease or not. On the other hand, cardiovascular diseases has been constantly increasing due to various reasons. Factors like poor diet, sedentary lifesyle, smoking, stress and genetics. In countries like India, the count has been increased a lot in past 10 years because lack of awareness on the above factors and also because of poor medical infrastructure [4]. Along with adressing these factors one of the main consideration is early detection of these heart related diseases.

In recent years, machine learning in the healthcare has shown tremendous path breaking solutions. ML techniques have developed solutions to the major day to day problems that every individual is dealing with. Researchers have been training ML models that can understand hidden relationships in the data where it is hard for human intelligence to understand. There ML models have been predicting risk, classifications, and other important target variables. Various ML techniques have been improving the performance of these models and giving better results over the period of time.

In healthcare industry, a single misclassification can lead to many unbearable consequenses to the patient. A false positive can result in unnecessary stress, medical tests, lengthy procedures to the patient where they are not required. In the same way, a false negative can result in missed diagnosis to the patient where it can lead to the death of patient. Achieving 100% accuracy for any model is not possible. So, explaning the models that are being used in medical industry are very important.

There are Explainable AI techniques (XAI) that explains the model's decision making behind the prediction/classification. With this the medical industry experts will build trust on the model that we have build even though the model is not having 100% accuracy.

1.2 Problem Definition

There are many Machine Learning models, and deep learning models that are showing strong predictive capabilities in detecting cardiovascular diseases. There are many robust Deep Learning (DL) models that are showing more than 99% accuracy as well. The major limitation with these deep learning models is that we cannot estimate the logic behind it's prediction. So, this lack of transparency in the model's decision can lead to lack of trust ont hem. This lack of interpretability is problematic in medical industry becuase, the decisions that are made in this industry must be justified with clear reason and accountability. This will ensure patient's safety.

Also, majority of the ML and DL models that have been built for predicting the heart disease predictions are being trained on the same dataset. Since we are dealing with patient's data we are suppose to consider various parameters while training the models. But all the models that were trained are using only the standard features present in the dataset which is available in the UCI machine learning repository.

The problem can be framed as follows:

- **To create new features using feature engineering and transformation techniques to better enhance the predictive performance of machine learning models for early detection of cardiovascular diseases.**
- **To implement Explainable AI (XAI) techniques such as SHapeley Additive exPlanations (SHAP), to improve the model transparency and provide better insights which can improve the trust of medical practitioners.**

1.3 Data Analytics Solution

The proposed solution involves in using feature engineering techniques and XAI techniques for creating new features and explaining their importance in the model's decision making. We have extracted heart disease dataset from the UCI machine learning repository and performed feature engineering on the existing features. After performing feature engineering techniques, we have implemented the ML models on the newly created features which is capable of distinguishing between patients at risk of heart disease and those who are not in the risk of heart disease. Once we have created the new features, we aimed to use ML techniques like Cross validation, Ensemble learning, and Hyperparameter tuning to enhance the model performance. Once we got the better model performance, we managed to implement explainable AI techniques in order to explain the reasoning behind the model's classification. SHAP has been used to provide explanation for the model's behaviour around individuals predictions. These two solutions will help medical industry experts to better understand what parameters are to be considered inorder to classify a patient's heart health. So, the model that has been build in this research can influence the further research direction in detecting heart diseases of the patients.

1.4 Research Methodologies

The research methodologies used are as follows:

- **Literature Review**: We have reviewed almost 25 research papers of existing works on the ML and XAI applications on the cardiovascular diseases and out of them 15 research papers contains of the research performed on the same UCI machine learning repository.

- **Dataset acquisition**: We aimed to perform the machine learning and XAI techniques on the publicly available datsets so, we have taken the heart disease dataset from the UCI machine learning repository for our research.

- **Dataset Preprocessing**: Since we are dealing with the patient data, each and every record is very important and contains unique information. So, we aimed not to delete the records that did not contain complete information. We instead used regression techniques to fill the missing/NaN vallues in this dataset. This is one of the novel approach that we have taken in this research to prepare the dataset for modelling.

- **Feature Engineering**: Along with the existing 13 features + 1 target variable present in the dataset, we have created 6 additional features that will better help model to understand the hidden patterns present in the dataset that will result in better classification of risk of patients with heart disease.

- **Model Implementation**: Implementing multiple Machine Learning (ML) models and classifying the risk of patients suffering from heart diseases.

- **Model Development**: Further implementing ML techniques like hyperparameter tuning, Cross validation, Ensemble learning techniques on the model to give better results.

- **Integration of XAI techniques**: Applying SHAP and LIME to explain the predictions of the final trained models.

- **Evaluation metrics**: Evaluating the model performance to see which model performed well overall.
- **Discussion and future work**: Since we have taken a novel approach in data preprocessing and model implementation, we have outlined few techniques that can be performed in future as extension of this research. We have outlined the discussions highlighting the trade-offs between bias and variance outlining the future areas of developlment.

1.5 Conclusion

Cardiovascular disease represents an overall global health challenge resulting millions of deaths annually. Some of those critical challenges are early detection and long-term recovery, in which we have focussed on early detection in this research. While machine learning provides robust solutions by developing models that can perform prediction and classification. It's lack of interpretability is the current barrier that we are planning to overcome in this research. The outcome of this study is expected to contribute robust machine learning model for early detection of cardiovascular diseases and through integration of Explainable AI techniques like SHAP and LIME we develop models that are not only accurate but also transparent. So, on broader goal this research aligns with responsibility of inntegrating AI in healthcare - ensuring that technological advancements are used for meaningful improvements of patients health.

1.6 Report Structure

This research report consists of the following sections Section 2: **Literature review**, Section 3: **Exploration of Dataset and Methods**, Section 4: **Machine Learning Techniques**, Section 5: **Explainable AI techniques**, Section 6:**Results and Evaluation**, Section 7: **Conclusion and Future work**, Section 8: **Acknowledgment** to our supervisor and followed by **References**.

Chapter 2

Literature Survey

Authors in [5] has used good data preprocessing techniques before applying ML models. The researcher has incorporated ensemble techniques like Bagging, Boosting to get the final predictions and narrowed down the final predictions using voting technique. In their research, they have used standard heart disease dataset to make predictions whereas in real-time the heart disease data may contain some uncertainties .

In the same way authors in [6], have implemented ML and DL models on standard dataset with only 303 instances. For industry advancement we must try to integrate real-time dynamic data and train our model to make better classifications .

Authors in [7] have focused on building a pipeline to integrate live patients records as an input and classify the patients. They have used Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) for data pre-processing and extracted important features. In this research they have used novel techniques to handle the missing values and they have used covariance based multivariate outlier detection procedure to detect the outliers and remove them.

In our research, As a novel approach, we aim to train our model using the same dataset, but we are trying to incorporate XAI techniques like SHAP and LIME to explain our model predictions. This results in better understanding for both the doctors and researchers, which variables play an important role in classifying the presence of heart disease.

Authors in [8], implemented GridSearch CV technique used for hyperparameter tuning. They have used five fold cross validation technique to train their model and employed soft voting ensemble method to increase model predictions. Though they have used hyperparameter tuning, they did not focus much in feature engineering techniques.

Similarly, authors in [9], has focused on using feature selection techniques to extract appropriate features for model training. But, the feature selection techniques used in this model are static and they may lead to increase biasness in the predictions.

So, in our research we aim to use feature engineering techniques to create new features and try to use XAI techniques to know which feature contributed more for model predictions.

Authors in [10], tried to classify the Hate Speech(HS) by using machine learning and deep learning models and evaluated the models performance. They have used Google Zigsaw and HateXplain to train different machine learning and deep learning models respectively. Out of the existing machine learning models, LSTM model has shown the best performance when trained on Google Zigsaw dataset with an impressive accuracy and recall score. To achive this result, they have used Count Vectorization and TF-IDF techniques for feature extraction. After pre-processing they have implemented decison tree, KNN, random forest, multinomial naive Bayes, LR and LSTM models on the google jigsaw dataset. They have implemented BERT + MLP and BERT + ANN models on the HateXSpeech dataset. Further to evaluate the model's predictions they have implemented XAI techniques like LIME on BERT variants and not other models because Google jigsaw dataset is not annotated like HateXSpeech dataset. By using XAI techniques, they further fine tuned the model for 50 epochs.Though this paper is not related to cardio vascular disease, this paper briefly explained about XAI techniques and how they play major role in interpretting model's predictions. Based on these techniques we can have better understanding of the model's algorithm and individual feature's contribution on the final prediction.

Authors in [11], used the Random Forest machine learning model to predict the mortality of the patient. They major contribution with this research is the feature extraction techniques used and the implementation of XAI techniques like SHAP to explain the reasoning behind the model predictions. For the feature selection, they have used LASSO regression(Least Absolute Shrinkage and Selection Operator) to shrink the less relevant features to 0. Then further they implemented the RFE technique to recurssively eliminate the unwanted features and reduced the dimentionality of the training data. By using these techniques they have achieved good model performance and further implemented XAI techniques to explain the probability of the features which contribute to the model predictions.

Authors in [12], performed the standard feature engineering technique which is Principle Component Heart Failure (PCHF) feature engineering technique and converted the standard features to reduced features. This technique extract the important components from the dataset and they are further implemented into the ML models. By using this technique they have achieved good results and the Decision trees, Random Forest algorithms have outperformed the other ML models (LR, SVM, NB, MLP). From their research we can consider two major gaps, first they did not explain the importance of newly created features which we include in our research. Secondly, the real time heart disease data is very dynamic and can be varying from patient to patient. So, we can propose few feature engineering techniques on multiple sub groups of patients as our future study.

In our research, we aim to create new features, and check the LASSO Regression values of them and proceed to implement our ML model on it. We can also further use the same XAI techniques used in this paper and evaluate the importance of newly created features.

Authors in [13], aimed to integrate AI framework with ML and IoT technologies. This approach is highly used in medical industry these days. They have compared the classification models within each and also with/without using the XAI techniques. Out of them, they have realised that models with XAI techniques are more trust worthy, when compared with the one that do not use XAI. They have used ealuation metrics like AUC and ROC curve, sensitivity and specificity to evaluate the model performances. The SVM classifier score high accuracy among all the implemented models. Techniques like feature selection, explainable feature weight initialisation, normalisation and optimisation have been used in this paper along with XAI. We can critically appreciate their approach where they tried to integrate the ML technologies with IoT devices. They have also used XAI technology which acts like a bridge for medical industry and researchers for both of them to better understand how the ML models are classifying patients. They have considered secondary evaluation done by doctors to check if the model has taken right parameters to make the predictions. In this particular research, they have mentioned their limitation that they trained their model with a dataset that contains limited variables. They have suggested to implement same techniques on other complex datasets as their future study. We can bridge this research gap in our thesis by implementing similar techniques in our dataset.

Authors in [14], have applied algorithmic models to analyze ECG signal in an eiffective way. They used 2 deep learning algorithms Mobile net V2 and VGG16 to classify ECG images. To achieve this they have implemented the Deep Learning models on the ECG images dataset. We can critically appreciate their approach of reducing the dimensionality of the dataset. Because Deep Learning models have complex algorithms and we do not know what happens inside those algorithms. So reducing the dimentionality using brightness, contrast, gamma, hue, saturation and central-crop is very brilliant. But, They did not explain the model's predictions. They are focussing on deep learning techniques and hyperparameter tuning techniques to optimise the model performance, but they are not considering which feature contributed more for model prediction. If they can find out that, they can gather a dataset accordingly which reduces the effort in dataset pre-processing. Authors in [15], used 4 different ML models to classify the heart disease of patients. They have taken 80/20 split of the train/test dataset and trained, Naïve Bayes, Random Forest, Decision tree and Logistic regression ML models. They have used F1- Measure, Precision, Accuracy and recall as their evaluation metrics to evaluate the performance of the model. Based on their research, they have classified that Random forest has the highest accuracy among the models. In our research we aim to consider this research as base and work their future work by implementing Explainable AI techniques on ML model to explore reasoning of model's classification.

Table 2.1. State-of-the-art review of related research papers for cardiovascular disease prediction using ML and XAI (Part 1)

Paper	**Techniques used in this paper**	**Techniques we can implement in our project**	**Critical evaluation**	**Novel Approach**	**Year**
The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction	Researchers used preprocessing (outlier detection/removal), Random Forest with bagging and voting, AdaBoost. Evaluated using AUC, ROC, precision, recall, and accuracy.	We can adopt preprocessing (outlier removal), Random Forest with bagging, and boosting (AdaBoost). Evaluate with multiple metrics.	This work used only standard datasets. Real-time data uncertainties were not considered.	Even though we also use standard datasets, we plan to integrate XAI (SHAP, LIME) for interpretability.	June 2022
Enhancing Heart Disease Prediction Accuracy Through ML Techniques and Optimization	Used GridSearchCV for hyperparameter tuning, 5-fold cross-validation, and soft voting ensembles.	We can apply GridSearchCV, tuning, and ensemble voting.	Feature engineering was not emphasized—only standard dataset features were used.	Our focus will include feature engineering for early detection by creating new features from existing attributes.	April 2023
Advanced Machine Learning Techniques for Predicting Heart Disease: A Comparative Analysis Using the Cleveland Heart Disease Dataset	Logistic Regression baseline, Random Forest, Gradient Boosting, XGBoost, LSTM.	Train both traditional ML and advanced DL models (LSTM), compare performance.	Dataset size (300) too small for robust generalization; real-world integration missing.	Apply hyperparameter tuning (Bayesian, Random Search), expand dataset with engineered features.	September 2024
Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers	Entropy-based feature engineering, ICA, PCA, LDA, ensemble classifiers.	Apply missing value imputation, outlier removal, entropy-based feature selection.	Good nested evaluation pipeline; no integration with real-time datasets.	Combine feature engineering + XAI to understand influence of engineered features.	June 2022
Feature Selection Strategies for Optimized Heart Disease diagnosis using ML and DL models	Feature selection: Mutual Information, Chi-Square, ANOVA F-test.	Implement appropriate feature selection depending on model complexity.	Static feature selection $\rightarrow$ potential bias.	Create new features (e.g., Age $\times$ LDL cholesterol) and evaluate their impact with XAI.	March 2025

Table 2.2. State-of-the-art review of related research papers for cardiovascular disease prediction using ML and XAI (Part 2)

Paper	Techniques used in this paper	Techniques we can implement in our project	Critical evaluation	Novel Approach	Year
Social Media Hate Speech Detection Using Explainable Artificial Intelligence (XAI)	LSTM, Decision Tree, KNN, Random Forest, Naive Bayes, Logistic Regression, BERT + MLP/ANN. Applied LIME for XAI.	Implement SHAP/LIME to evaluate feature contributions.	Not related to CVD, but shows strong role of XAI.	Apply SHAP/LIME to check feature influence in CVD dataset.	August 2022
Artificial Intelligence for Cardiac Diseases Diagnosis and Predicting Using ECG Images on Embedded Systems	DL models (MobileNetV2, VGG16) for ECG classification.	Explore DL models for ECG dataset classification.	Achieved 95% accuracy, but no interpretability (black-box).	Add hyperparameter tuning + XAI for ECG models.	August 2022
Feature-Enhanced Machine Learning for All-Cause Mortality Prediction in Healthcare Data	Random Forest, LASSO regression, RFE, SHAP for explanations.	Use SHAP for feature importance, dimensionality reduction.	Good feature elimination pipeline; but creating new features may capture hidden variance.	Engineer new features + validate them with SHAP.	March 2025
Heart Disease Prediction using Machine Learning	Compared Naïve Bayes, Random Forest, Decision Tree, Logistic Regression with 80/20 split.	Adopt 80/20 split, test Random Forest in detail.	Only standard ML comparison, no XAI.	Apply XAI on trained models.	April 2020
XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques	Integrated ML + IoT + XAI (SVM, feature selection, normalization).	Apply SHAP for model transparency.	Strong IoT + XAI integration; dataset lacked robustness.	Extend with Statlog dataset + engineered features.	December 2022
Explainable Artificial Intelligence (XAI) Approach to Heart Disease Prediction	SHAP, LIME + user-friendly interface for doctors.	Use LIME to evaluate feature influence.	Good XAI integration; dataset too small.	Extend study with deep learning models (CNN, LSTM).	July 2024
XAI-Augmented Voting Ensemble Models for Heart Disease Prediction: A SHAP and LIME-Based Approach	Ensemble: XGBoost + LightGBM + Bayesian optimization, SHAP + LIME.	Compare hyperparameter tuning (Bayesian vs. others).	Strong Bayesian tuning; limited dataset features, interpretability challenges.	Apply to larger dataset + engineered features.	October 2024
XAI–reduct: Accuracy preservation despite dimensionality reduction for heart disease classification	Dimensionality reduction with SHAPASH + DALEX XAI.	Use advanced XAI to study feature importance.	Selected only 4 features → too limited.	Use more features with XAI explanations.	April 2023
Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning	Principal Component Heart Failure (PCHF) technique + Decision Trees, Random Forest.	Apply PCHF + XAI to evaluate new components.	Reduced features only; ignored patient-level variability.	Engineer subgroup features + validate with XAI.	May 2023

Authors in [16] have used feature selection techniques to select 10 important features that are highly effective in predicting heart diseases. They have performed ML techniques with Synthetic Minority Oversampling Technique (SMOTE) to these selected features. In this research, they have used ANOVA (Analysis Of Variance), and chi-square tests to get choose the selected features. They have implemented 10 different Machine Learning models and out of them XG Boost algorithm has scored the highest accuracy in predicting the heart disease. They have implemented all these models on the publicly available dataset of Egypt and Saudi Arabia dataset. The incomplete records in the medical data are considered as "dark data". In this reasearch they have managed to delete all the dark data and implemented the models. As a novel approach from this study, we are aiming to pre process the dark data by not deleting them and implement the Machine Learning (ML) techniques and Explainable AI (XAI) techniques.

Authors in [17] have used unique feature selection techniques and developed Machine Learning (ML) models. This approach is highly appreciable because they have implemented Fast Correlation Based Filter (FCBF) and Minimum Redundancy Maximum Relevance (mRMR) for feature selection. The further steps of this reasearch is similar to the other studies, but as the novel approach of this research they are aiming to integrate this decision support system in the health care clinics to detect the heart failure of patients at an early stage.

Authors in [18] they have performed research is predicting Myocardial Infarction of patients by considering two different datasets. One of the selected is the dataset that we are using in this research, the other one is Framingham dataset. The outcome of this research is they have trained supervised Machine Learning models that can be able to detect if the patient is prone to heart disease or not. They have performed only binary classification of the dataset, where we are performing multi class classification by further dividing the patients with heart disease into 4 different classes, mild heart disease, moderate heart disease, severe heart disease and very severe heart disease. This research only shows the outcome if the patient is suffering from heart disease or not, this is only helpful to identify patients that are going to be suffering from heart disease. But the approach we are taking in our research will help us to classify the patients and this can also help us to analyze the treatment tha can be given to the patient in before hand. This can be a very time efficient measure that can be implemented.

Authors in [19] have performed research on the UCI heart disease dataset by implementing supervised learning algorithms to predict the heart disease in the patient. In this research, they focussed only on implementing the ML models and they have performed comparitive analysis to check which algorithm is giving better results. Further they have used ensemble techniques to improve the model performance. We can appreciate their approach of using the ensemble techniques to improve the model performance. From this research, we can incorporate the usage of ensemble techniques in our research, but we can implement boosting technqiues as a novel approach.

Authors in [20] have performed research on a dataset that is not publicly available. They have extracted the Electoric Health Records (EHRs) of patients through proper authorizations and permissions. Then they have performed analysis and ML implementation on that data. We can appreciate them for performing analysis on the raw EHR data instead of already avaialble dataset. The reason is because, people's body vitals have been constantly changing and concluding the facts based on 10 years old datasets is not ideal. The body vitals of current generation is not the same to people of decade ago. The outcome of this research is they have perfromed ML techniques on the extracted EHR data and they have predicted the patients suffering from Atherosclerotic Cardio Vascular Diseases (ASCVD). In this reaserch they have used mean value imputation to fill the missing values and they have used multiple imputations to fill the missing values. From this research we can consider to impute the missing values of few variables by the mean value. As a novel approach we can use ML model to predict the missing values based on the variables that are not haivng any missing values.

Chapter 3

Exploration of Dataset and Methods

3.1 Dataset Collection

The data in this dataset is collected by clinical data collection from the patients who were taking angiographic test. Angiography is a test conducted to check arteries of human heart by injecting dye into the blood. After angiography, the researchers have extracted the test results from the Electronic Health records (EHRs). Then they have extracted 14 important features from the overall features of the dataset. Then they have anonymised the extracted data for research purpose. Further they have aggregated the data as per region.

3.2 Dataset Description

We have obtained this licensed dataset from UC Irvine Machine Learning Repository. This dataset is formed by gathering patient information who are undergoing angiogram treatment in Cleaveland, Hungary, Switzerland and Long Beach VA. The raw format of these individual data files contains 76 attributes and 303, 294, 123, and 270 instances for Cleaveland, Hungarian, Switzerland, and Long Beach VA respectively. Out of the 76 attributes, researchers have extracted 14 attributes that are ideal for predicting heart diseases. The target variable is one of the 14 attributes. In the below table 3.1 Description of Dataset Features describes the full form of all the features and contains the definitions of all the features.

Table 3.1. Description of Dataset Features

Feature	Description
age	Age of the patient
gender	Sex of the patient
chest_pain	Type of chest pain (0–4, increasing severity)
resting_BP	Resting blood pressure (mmHg)
serum_cholestrol	Serum cholesterol level (mg/dl)
fasting_blood_sugar	Fasting blood sugar (0 = normal, 1 = high)
resting_ecg	ECG result (0 = normal, 1 = ST-T abnormality, 2 = LV hypertrophy)
max_heart_rate	Maximum heart rate achieved
exercise_angia	Exercise-induced angina (yes/no)
old_peak	ST depression induced by exercise
slope	Slope of the ST segment during exercise
no_of_major_vessels	Number of major blood vessels (0–3)
thal	Thalassemia (3 = normal, 6 = fixed defect, 7 = reversible defect)
target	Diagnosis (0 = no disease, 1 = mild, 2 = severe)

3.3 Exploratory Data Analysis

After performing the Uni-variate analysis on the features present in our dataset, we got an overview of how data is distributed in the dataset. Here are the key points that outlines the features as follows:

- The dataset is created by extracting ECG results from male patients, the male patients are twice greater than female patients.
- The age of the patients present in the dataset is normally distributed between 30 years and 75 years.
- The patients with chestpain severity 4 are 50% in this dataset.
- serum cholesterol has normal distribution but there are few outliers. Some values in the dataset are containing 0, which is technically not possible. So, we shall fill the serum_cholesterol values using MLP regression.
- The fasting blood sugar of majority of the patients is under normal conditions (<= 120 mg/dl), only 45(approx) patients blood sugar level is higher (>120mg/dl).
- maximum heart rate of patients is showing normal distribution with a mean around 150- 160 beats per minute.
- Since majority of the patients are involved in physical activity, the patients without exercise angia are 2x times than patients suffering from exercise angina.
- The slope of peak exercise ST segment contains 3 different classes and out of them downsloping class patients are very less in number.

3.3.1 Univariate Analysis

Age Distribution:

- The below figure 3.1 Age distribution represents the distribution of age of the patients that were participated in this research. We can see that the histogram is not having even distribution. The majority of the people lie within the middle age, 40-70 years of age.

- we can see that the majority of the frequency is between 50-60 years of age. We can consider that this age group is the average age group where any person can affect with caridac related issues. So, this dataset aligns strongly with this statement.

- we can observe that the patients under 20 and patients above 80 years are not many. Based on this, we can interpret that people under 20 years are hard to prone to heart related issues and maybe people who got affected by heart diseases cannot make it till 80. This can also be a dataset sampling error.

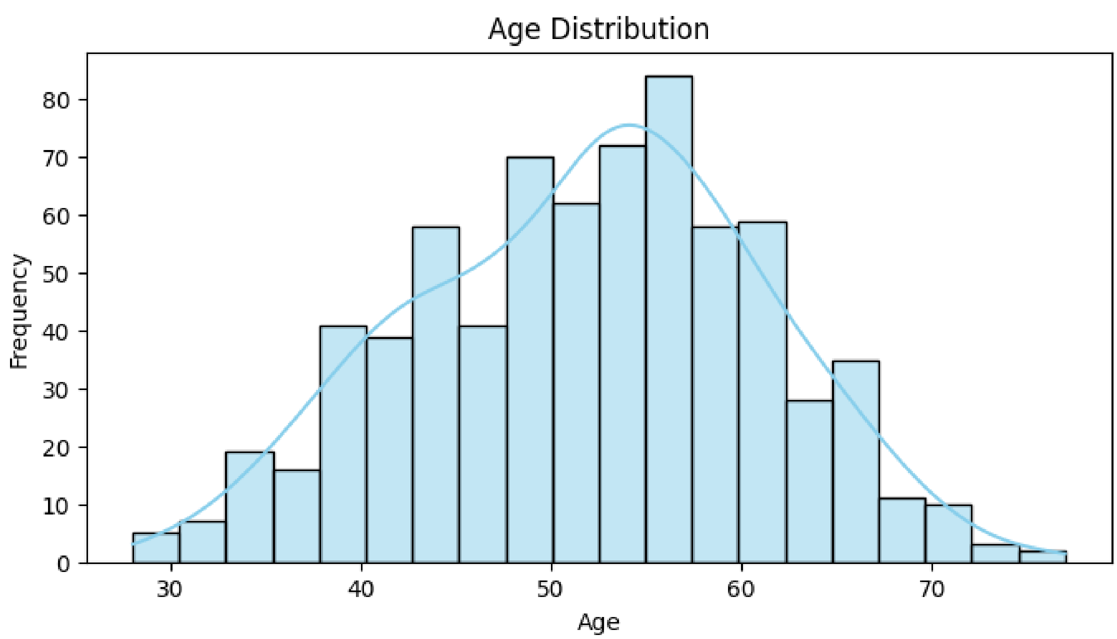

Figure 3.1. Age Distribution

Gender Distribution:

- In the below figure 3.2 Gender distribution graph, there is a gender imbalance in this dataset and nearly 3/4th of the patients that are participated is male and remaining 1/4 participants are female. This leaves us with a ratio of 3:1 for male:female in this dataset.
- studies show that males are more prone to heart attacks than females. This might be the reason this dataset contains more male patients data.
- there might be a possibility that the model can understandn only male patients patterns and not female patient's data. So, if that is the case, we can do future work on collecting heart disease dataset with female patients.

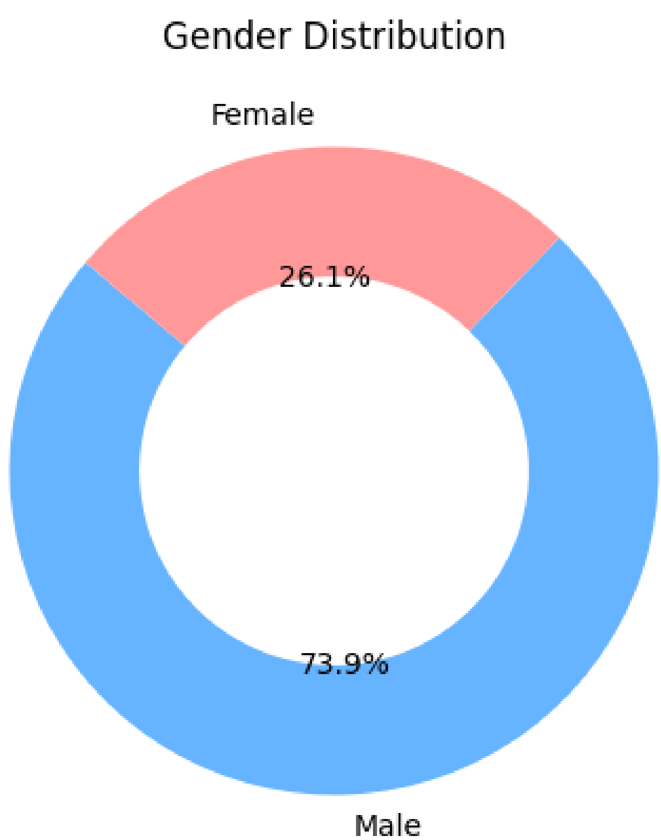

Figure 3.2. Gender Distribution

Chest pain:

- In the below figure 3.3 Chest pain distribution graph, majority of the patients in this dataset are having chestpain type 4, which means that there is no symptoms. This is the main reason that we have selected this dataset for this study because we can see the trend of data for early detection of cardiovascular diseases.

- The type 1 patients are also less in number which also means that we can proceed with using this dataset for early detection of cardiovascular diseases. Because type 1 is patients that are suffering from severe chest pain, and this dataset does not consist of more people who are already suffering from chest pain and who are present in the emergency room.

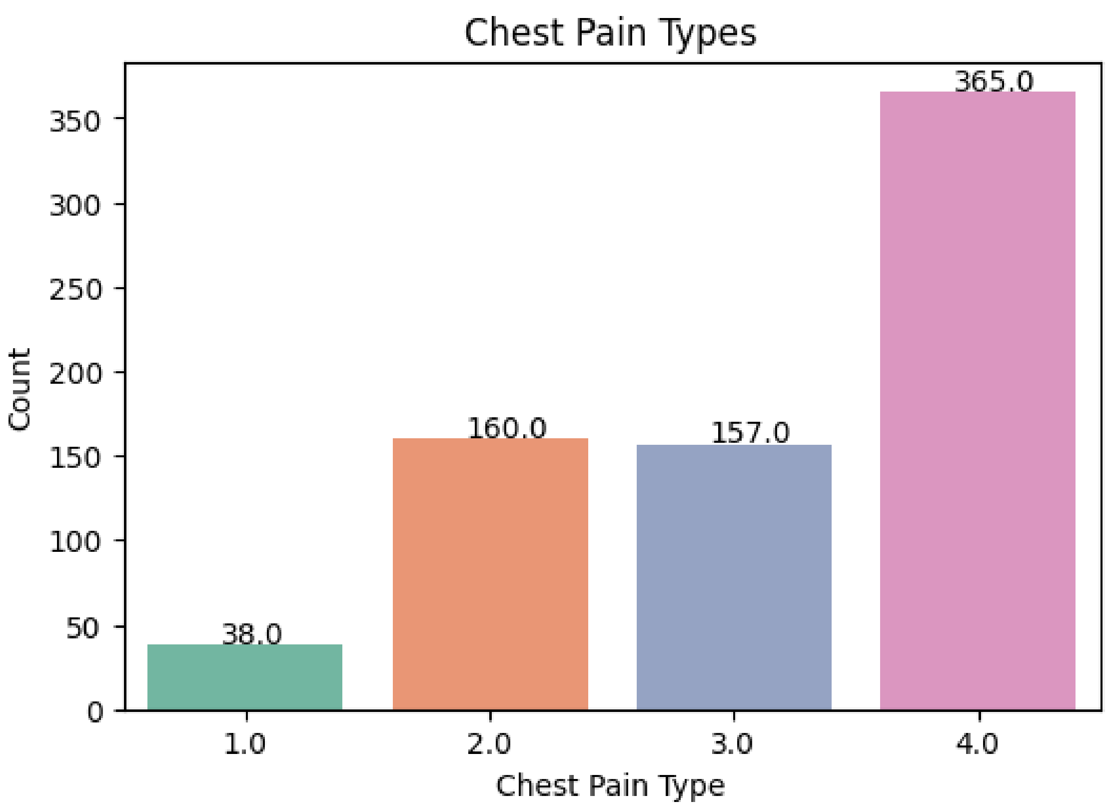

Figure 3.3. Chest Pain Distribution

Exercise Induced Angina

- The below figure 3.4 Exercise induced angina distribution shows that 66.3% of the patients did not report experiencing the exercise-induced angina, This finding is inline with the previous chest pain feature where majority of the patients fall under type 4 chest pain where the patients did not report that they have been suffering from chest pain.
- The remaining 33.7% statistics shows that the patients might be suffering from severe heart pain and they have reported it accordingly. This group of people's data can be used to check the model understanding, because majority of the predictions from this pool will show that the patient might be have a risk of gettig cardiovascular disease.

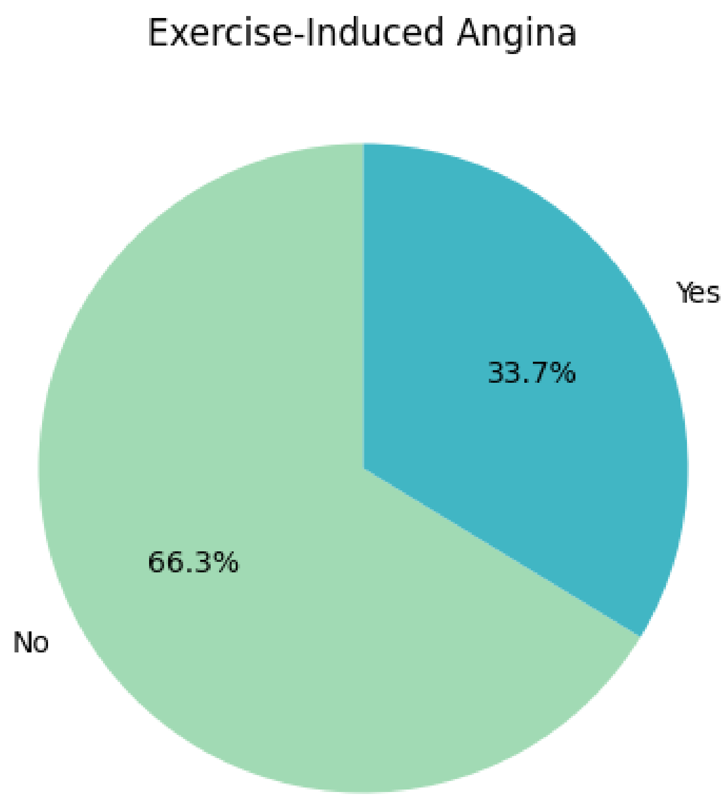

Figure 3.4. Exercise Induced Angina Distribution

Thallasemia:

- In medical terms, thallasemia not only refers to the blood disorder of the patient, in this research it refers to the blood flow to the heart.
- In the below figure 3.5 Old Peakokay Distribution, The 3.0 represents that there is no irregular blood flow detected. 6.0 means that there is irreversible defect detected, this usually shown in the people who suffered from heart attack. 7.0 means there is reversible defect detected, this means that the patient is having normal bloodflow in normal times and defective blood flow during the stress.
- in this figure 3.5 we can see that majority of the patients are not having any defects or there are patients that are having reversible defects. This is the right proportion of patients to deal with early prediction of heart data.
- Fixed defect patients (6.0) shows that they already suffered from heart disease and these are the patients that model has to focus to understand that they have high probability in suffering from a heart disease in future.

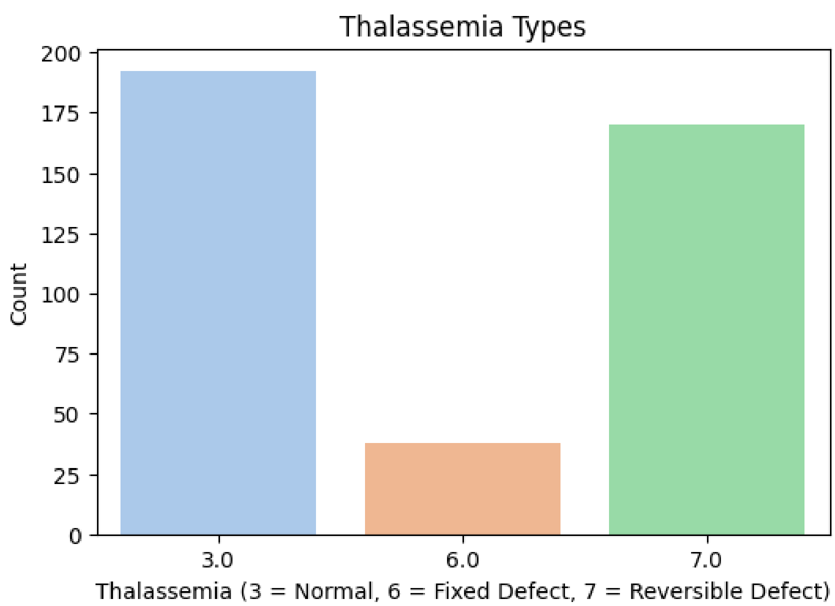

Figure 3.5. Thallasemia Distribution

Old Peak:

- Old peak measures the depth of the ST segment depression on ECG after exercise when compared to rest. It is the primary sign for heart attack where heart muscle do not get enough oxygen.

- In figure 3.6 Old peak distribution, we can see that the patient data from this dataset is right skewed with a notable tail. The data is normally distibuted with the majority of the values at 0. But there is a significant right skewness observed for the values 4,5 and 6.

- This indicates that majority of the patients show little to no ST depression.

- For the early detection of heart disease the model has to clearly understand the patients with low/no depression values.

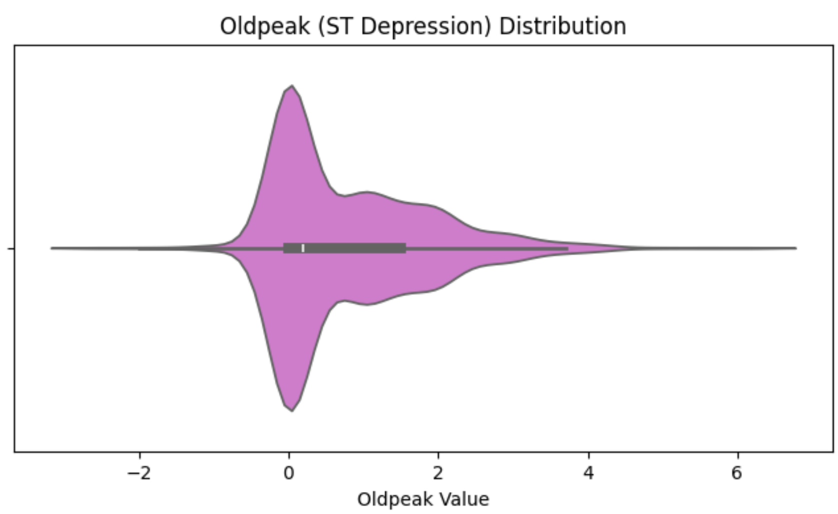

Figure 3.6. Old Peak

The above univariate analysis of the variables and the dataset description given in the table 3.1 helps us to understand more about the description of the features and their distribution. Since we have got a better understanding of what class patients are present in our dataset, we can now proceed into data pre processing and followed by model implementation.

3.4 Data Pre-processing

In this research, one of the novel approach we have taken is the way of how we performed data pre=processing. Since we deal with medical data, every medical record is important. So, we did not delete a single record from the dataset that we have collected. For the columns that have less than 5% of missing records in them, we have replaced them with the median, median or mode values depending on the distribution of the column. For the columns that are containing missing values of more than 30% we have used KNN imputation approach to fill the NULL values.

After handling the NULL values, we found out that, there are few records in serum_cholesterol column that are zero. Technically, serum_cholesterol cannot be zero. So, we used Multi Layer Perceptron Regression model to predict the values that are zero, by figuring out hidden relationships of the dataset. We have used age, resting_bp and the max_heart_rate columns to predict the serum_cholesterol values. Further we replaced the original serum_cholesterol values with the predicted values. In the below figure 3.7 Methodology, we show the flow of methodology that we used in this research.

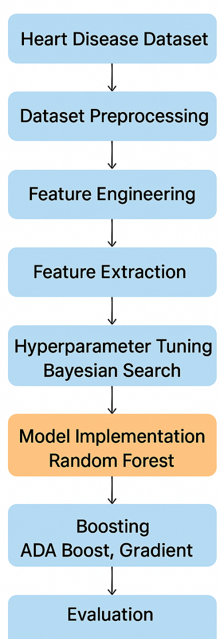

Figure 3.7. Methodology.

Chapter 4

Machine Learning Techniques

4.1 Machine Learning Techniques

After pre-processing the dataset, we have now build a model that predicts the presence of heart disease in a patient. We have used random forest model to build a base-line model, because since our dataset has mixed data types, and there were many outliers in this dataset as well. Further now we shall proceed to fine tune this model by using machine learning techniques to improve the model efficiency.

4.2 Hyperparameter Tuning

In this research, we have used bayesian search to check for the right hyperparameters instead of grid search or random search. The reason we have used bayesian search is this technique is computationally efficient when compared to other searching options. Also, provides best hyperparameter values by efficiently searching for them. After performing the hyperparameter tuning, the ideal parameters to tune the model are as follows: criterion = "log_loss", n_estimators = 50, random_state = 42.

4.3 Feature Engineering and Feature Extraction

In this research, we have used feature engineering concept to create new features with the help of existing features. So that, we can train the model with new set of features, rather than other models that are all using the existing features. So, we have created four new features of risk ratios, by calculating the risk percentage from the existing columns. Here are the newly created columns:

- age, we have changed the age to categorical by splitting the age values into 4 different bins, young age, middle age, and old age.
- cholesterol risk flag, where cholesterol >than 240
- High Blood Pressure risk flag, where resting blood pressure is >than 140.
- fitness index = Thal/Age.
- old_peak_slope = old peak * slope.
- age_exangia = age* exercise angia.

We then proceeded to perform feature extraction to extract only required features. Also to not overfit the model by training with more feature information. We have implemented Recursive Elimination Feature Extraction to extract the required features. After extracting the require features we then re-implemented the model to check it's performance. The model performance has slightly improved from 60.2% accuracy to 62% after feature extraction and 66% after hyperparameter tuning.

4.4 Model Implementation

In this dataset, we have mixture of categorical and numerical rows, we have also created new categorical features from the existing numerical features. So, we have aimed to use a non-parametric model and we have selected random forest classifier to classify our dataset. The dataset is slightly complex and have many hidden relationship between the features, and there are few outliers in the dataset, considering all these parameters we implemented random forest model. The base-line model got 60.2% accuracy. Further we have tuned the model by performing hyperparameter tuning, feature engineering and feature extraction and imiproved the model accuracy to 66%

4.5 Ensemble Learning

In this research we are using the random forest model as the base-line model and in order to improve the model performance, we implemented boosting technique by bringing all the weak learners together to train the model and reduce the bias of the model. The boosting techniques also helps us to reduce the variance of the model by focusing only on the important areas of the dataset, instead of focusing on the entire dataset and increasing the variance.So, as an extension of our baseline model we implemented Ada boosting, and Gradient boosting technique. The Gradient boosting technique has outperformed Ada Boosting by improving the model performance by 82% accuracy.

4.6 Model Deployment

4.6.1 Overview

After we created the final models we have saved the models using the pickle library. The critical step is to deploy these models so that the healthcare practitioneres, and researchers will be able to see the performance of these models. So, we aim to deploy these models to not only show the performance of these metrics but also to transparently show the features that were contributing to the model's decision making. For this purpose we have taken a light-weight, interactive and userfriendly dashboard developed using streamlit.

4.6.2 Deployment Framework

: We have selected streamlit for deployment because, it is very interactive and user friendly. It simply allows the user to transform machine learning pipelines into fully functional web applications with minimal effort.

The deployed application meets the following interactive objectives:

- Select the model between random forest, AdaBoost and Graditent Boosting.
- Display evaluation metrics including the confusion matrix, and classification report.
- Display interactive visualizations, precision-recall curves, feature importance and permutation importance.

4.6.3 Dashboard Design

: The dashboard that we designed will contain the below outlined key section:

- **Dataset Preview**: This section will show the dataset overview and displays the first few records of the dataset to give the overview of the datatypes.
- **Model Selection:** This section will give an option to the user to select one of the model (Random Forest, AdaBoost, Gradient Boosing) that we have trained through a dropdown menu.
- **Performance Metrics**: We show the Accuracy score of the selected model, classification report that shows the precision, recall and f1 score. We also show the confusion matrix showing predictions vs actual values of each and every class.
- Evaluation Curves: We plot the ROC curves to show the perfromance of the model for each class.
- **Feature Importance and Permutation Importance**: We have plotted a bar chart to show the feature importance of the dataset. We have then plotted

a bar plot to show the permutation values of the features that contributed to the final decision.

Further detailed explanation is provided in the results and evaluation chapter i.e., chapter 6, where each model results have been evaluated.

4.6.4 Technial Implementation

- We have deployed the model by using **streamlit** and we have developed the code using **python** language.
- The **dashboard.py** script loads the preprocessed dataset. we have used **@st.cache_data** to reduce the reload times.
- Further we have used **MinMaxScaler** to normalize numerical features. We have created age_group, high_chol flags, and fitness_index additional features using existing features.
- we have further trained 3 models: **Random Forest, AdaBoost classifier, Gradient Boosting**
- we have further evaluated the model performance by calculating the confusion matrix, and classification reports. We have plotted the ROC curves to show model performance on individual classes.
- We have further deployed the model using the following command **Streamlit run dashboard.py**

4.6.5 Benefits of deployment

The deployed dashboard can be **accessible** to all the medical industry professionals and researchers. The graphs and outputs shows the **Transparency and Interpretability** of the models that we have developed. It is a single platform where we can **flexibly** check the performance of multiple models.

Chapter 5

Explainable AI Techniques

5.1 Importance of Explainable AI in Medical Industry

When we use complex machine learning or deep learning algorithms fir classification or predictions we must also think of the performance of the model when we deal with medical industry problems. The reason is because, a false positive in medical industry can cause a lot of stress in the patient where the patient has to undergo a lot of unwanted procedures, tests and lot of stress. In the same way, a false negative in medical industry can lead to missed treatment for the patient. This can be fatal. So, when we aim to predict or classify for a problem in medical industry, we must make sure performance is very important. Achieving 100% accuracy all the time is not possible, if we deal with black box models for finding solutions we must also make sure what made the model to make this decison. **Explainable AI** (XAI) focuses on providing explanations to the model's decision making. This increases the transparency for the medical experts which improves the trust in the model that we build for thier problem.

5.2 SHAP: SHapeley Additive exPlanations

The most widely used technique in XAI is **SHapeley Additive exPlanations** (SHAP), where we assign score for the feature's contribution towards model's prediction. This will help us to understand which feature has pushed the model towards positive prediction and which feature has pushed the model towards negative predictions[21]. The key benefits of the SHAP are outlined as follows:

- **Local Interpretability**: This explains the individual predictions by evaluating the feature contributions towards the final prediction.
- **Global Interpretability**: This aggregates all the individual prediction's score to see which feature has more influence towards the prediction on a whole.
- **model-agonistic**: This technique can be applied to any machine learning model or deep learning model, where we can understand any model, even black box models.

5.3 Implementation of SHAP in this Research

In this research, we have implemented the XAI techniques like SHAP explaining the feature contribution to the final classification. There are many black box models applied on the heart disease dataset, but since we are dealing with the medical data, it is very important to explain the reasoning behind the classification. This will increase the trust and reliability in the medical industry experts.

The below figure 5.1 Explanation of Gradient Boosting model, shows the contribution of the newly created features towards the model classification. We can see that the chest_pain feature has the highest contribution towards the model classification.

The below figure 5.2 Explanation of Ada Boost model shows that for the ada boost model, the num_major_vessels feature has more contribution towards the final classification. Comparing with the gradient boosting model, it has equal importance along

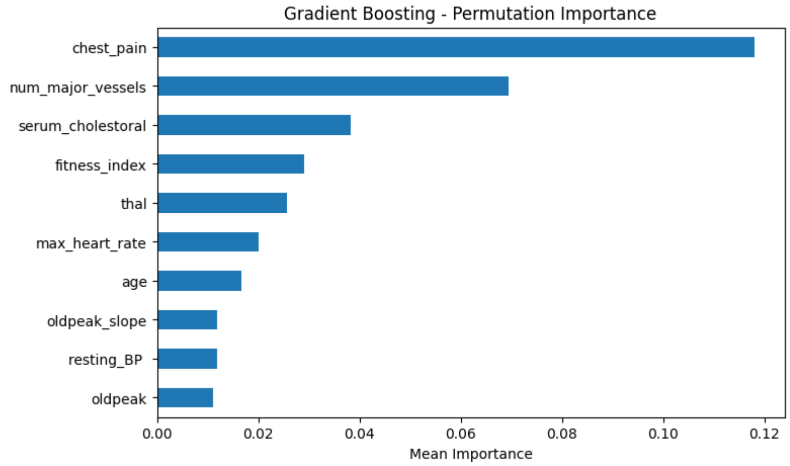

Figure 5.1. Explanation of Gradient boosting Model.

with the chest_pain feature. Along with the positive contributions we also have features that has negative contributions towards classification. Resting_blood_pressure has negative score for classifying a patient suffering from heart disease or not.

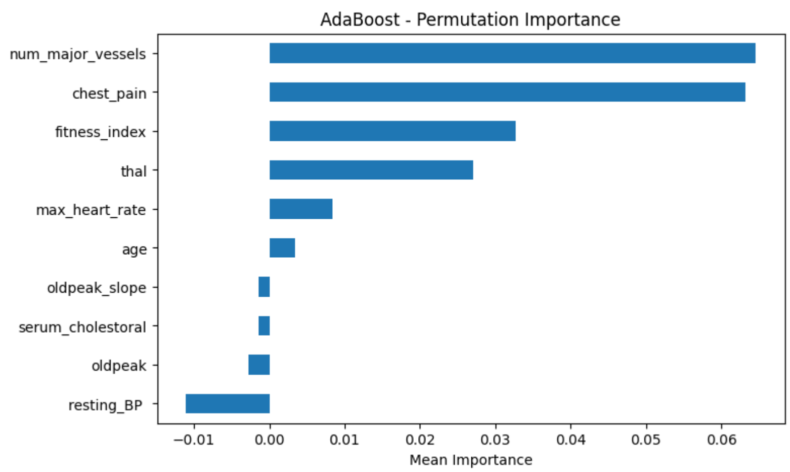

Figure 5.2. Explanation of Ada Boost Model.

In the below figure 5.3 Explanation of Random Forest Classifier, there are very less features that has positive scores. This can mean that the random forest is not able to sub classify the patients into mild, severe, and very severe heart disease classes. For all the three models chest_pain is the only feature that has greater contribution for classifying the patients with heart disease or not.

So, these Explainable AI techniques help the medical industry experts and researchers to focus on which features are very complex in terms of predicting the

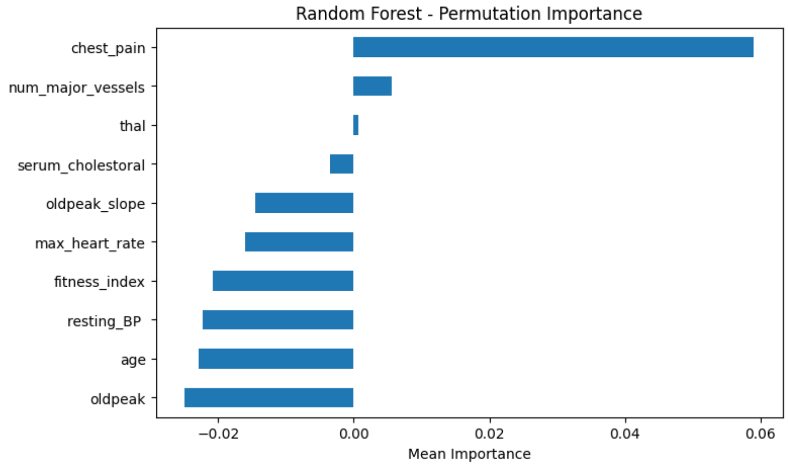

Figure 5.3. Explanation of Random Forest Classifier.

heart disease of the patient.

5.4 Benefits of XAI for Healthcare

The integration of XAI techniques shows the below mentioned benefits for the healthcare industry.

- **Research insight**: For the researchers, who are interested in performing research in this topic, they can easily understand which feature to deeply investigate.
- **Decision support**: Upon investigation, once they get to know the feature contributions, they will make a decision for which feature to include and which feature to exclude.
- **Clinical trust**: Based on the feature explanations, the medical industry experts will get trust on model's predictions.

Chapter 6

Results and Evaluation

6.1 Results

The analysis of the cardiovascular disease dataset revealed several key insights in predicting heart diseases. The baseline random forest model has acheived a overall F1 score of 76% performing well in particular with the class 0 true positives. We got a recall value of 90%. However, it didn't perform well in the class 2 and class 3 patients. The reason for this might be because, we do not have enough data to train our model to further classify them into severity of the heart disease. The below table 6.1 shows Evaluation metics of machine learning models on the cardiovascular disease dataset.

Table 6.1. Evaluation metrics of machine learning models on the cardiovascular disease dataset

Model	**Class 0**	**Class 1**	**Class 2**	**PR AUC**
Random Forest	76% / 90%	68% / 55%	0% / 0%	0.80
AdaBoost	75% / 75%	69% / 69%	0% / 0%	0.565
Gradient Boosting	80% / 88%	72% / 64%	15% / 12%	0.82

The Ada Boost shows slightly lower performance when compared to the random forest model. It has a slightly better F1 score of 75% and 69% for the class 0 and 1 but the PR AUC score is not well. It has 56.5% score with which we can assume that the Ada Boost model is poorly predicting the target variable.

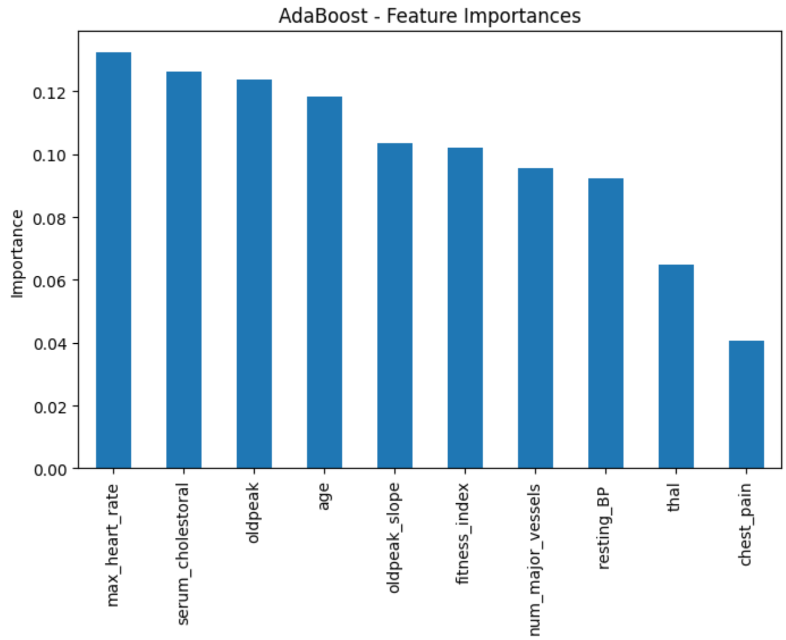

Figure 6.1. Ada Boost Features Importance

In contrast, the gradient boosting model has the highest peformance among all the three models with a F1 score of 80% and recall scores of 88% and 64% for class 0 and 1 respectively. The Random forest model has better recall score class 0 but the gradient boosting did well with the class 1. This mean that gradient boosting is better understanding the hidden relationships in dataset among the three models.

Feature Engineering,Newly created features such as cholesterol risk flag, high blood pressure indicator, fitness index, old peak-slope interaction, and exercise angia interaction helped the models to capture clinically relevant risk factors and complex relationships that are not present originally in the dataset.

Explainable AI, Including Features importance and permutations importance highlighted the chest_pain severity as the most influential predictor across all the models. Other features like number of major vessels, resting blood pressure showed importance in ada boost model but not in random forest model and vice versa. This shows the non-linear feature interaction shown by boosting technique. These explanations enhance the trust and reliability among the medical industry experts by clearly showing the factors that drives the final predictions. Despite these improvemets the higher classes predictions are still challenging. All the three models

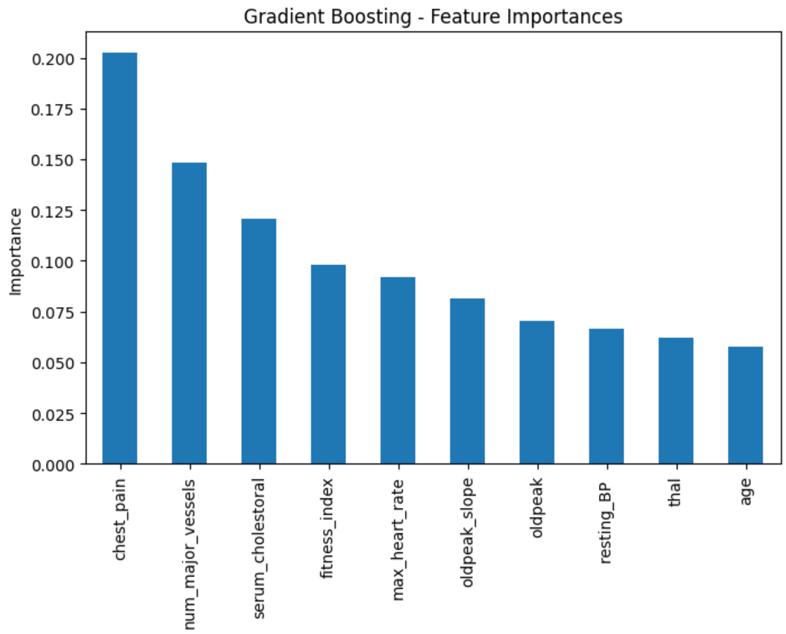

Figure 6.2. Gradient Boosting Features Importance

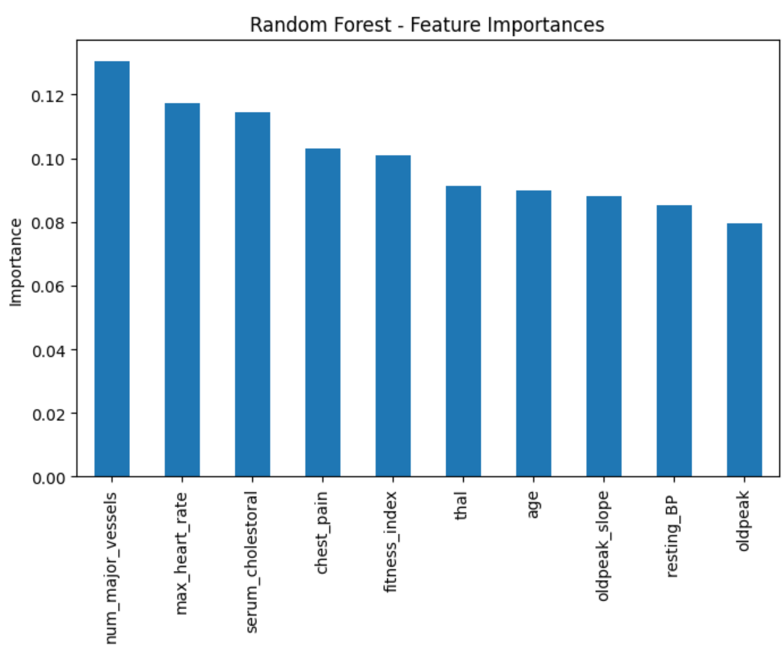

Figure 6.3. Random Forest Features Importance

struggled to predict the class 2,3 and 4. This means the models are not able to further classify the severity of the heart disease.

6.2 Evaluation

In this research, we have taken Precision, Recall, AUC and ROC curve metrics as our standard metrics to evaluate our model performance. This is a multi class dataset with 5 classes. Some of the classes were easy to classify but model was not able to find the hidden relationships in few classes.

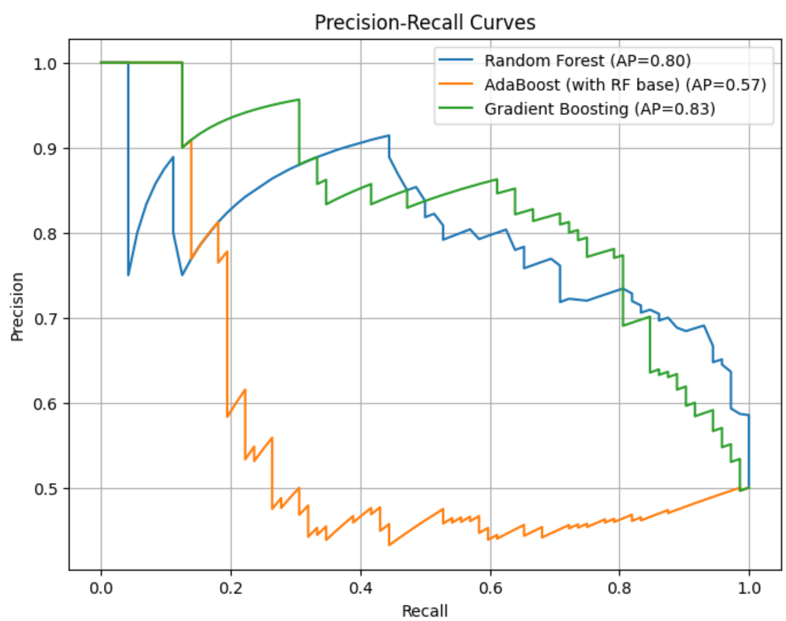

Figure 6.4. Precision Recall Curves

- Random Forest Classifier: The Baseline random forest classifier achieved 90% of recall score for the 0 class. That means it is able to predict very well with the patients that do not contain heart disease. Overall the F1 score is 76% which shows a balanced performance of the model but the downside of the model is that, it was not able to classify the 2 and 3 classes. The PR AUC score is 80% which is a very good performance.

- Ada Boost classifier: The Ada boost model has slightly lesser performance than random forest classifier but it has got reasonable results for the classes 0 and 1 with 75% and 69% accuracy. The Ada boost classifier has lower PR AUC score when compared to all the three models with the score of 0.565. This model is not able to classify the positives of the target variable.

- Gradient Boost classifier: The gradient boosting model has the highest performance in all the evaluation metrics. It has highest recall value for both 0 and 1 classes with the score of 88% and 64% respectively. It is able to predict the class 1 records better than random forest model. The weighted f1 score is 80% which is the highest in all the three models.

6.3 Model Predictions

- **Output 1**: In the below table 6.2 shows the class 0 predictions where we have given a sample input of the record present in the dataset which falls under class 0. The patient falls under mixed clinical picture. The model is not confident

Table 6.2. Clinical interpretation of class 0 patient features

Feature	Value	Clinical Interpretation
age	35	Young adult.
chest_pain	0	Typical angina (or asymptomatic). This is a significant risk factor.
resting_BP	120	Normal blood pressure.
serum_cholesterol	250	High cholesterol (Hyperlipidemia). A major risk factor for heart disease.
max_heart_rate	170	High maximum heart rate achieved for age 35. This could indicate good fitness or, in a clinical context, might be a response to stress.
oldpeak	0.5	Minimal ST depression induced by exercise relative to rest. This is a mildly abnormal result.
num_major_vessels	0	No major blood vessels colored by fluoroscopy. This is a very good sign, indicating no significant blockages.
thal	2	Likely indicates 'fixed defect' (old heart attack) or 'reversible defect' (ischemia). This is a significant risk factor.
fitness_index	17	This appears to be a high value (possibly METs or other capacity score), suggesting good functional capacity.
slope_of_peak	1	Upsloping ST segment during peak exercise. This is a normal and desirable response.

about this predictions so, we got only 38.5% probability for class 0. Since this is the highest probability shown, the model has classified this patient falls under class 0. But the overall probability of patient having heart disease is 61.5% which is 2 times more. Since this is a multiclass problem, we did not get accurate result for this record. Below figure 6.5 shows the probabilities of the class 0 predictions.

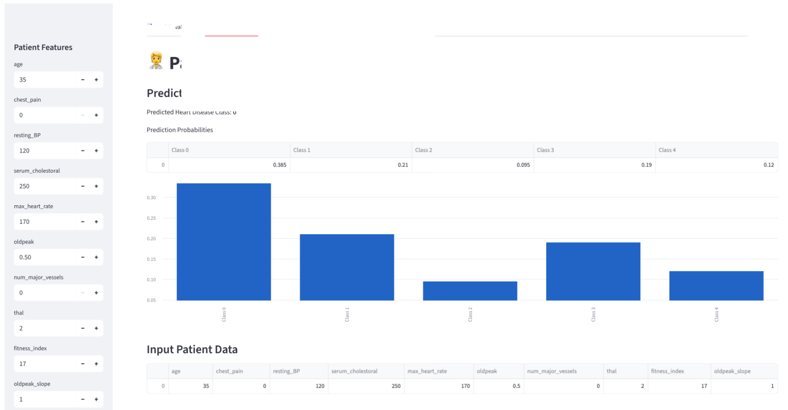

Figure 6.5. Probability of individual class Predictions - Output 1

Further this patient can be flagged to check for more diagnosis, the main advantage of this report is it shows the probability of the output. This helps the end user to make decision based on this output.

- **Output 2**: In the below table 6.3 that shows the clinical interpretation of patient features for the predicted class 2. The table shows the description of the input that we have given to check which class does the patient falls under. According to the input that we have given, the patient is 50 years of age with some border line health vitals. There are 0 blocked vessels for the patient and the patient has good max_heart_rate for their age. But there are some risk flags like serem_cholesterol levels with 250 mg/dl. The Thallesemia is also at 2 which means reversible defect.

Table 6.3. Clinical Interpretation of Patient Features for Predicted Class 2

Feature	Value	Clinical Interpretation
Age	50	Middle-aged adult. Risk of heart disease increases steadily after age 45.
Chest Pain Type	1	Typical angina. Chest pain related to reduced blood flow to the heart. A significant symptom.
Resting Blood Pressure (mm Hg)	120	Normal blood pressure. This is not a contributing risk factor for this patient.
Serum Cholesterol (mg/dl)	250	High cholesterol (Hyperlipidemia). A major, modifiable risk factor for coronary artery disease.
Maximum Heart Rate	150	Good maximum heart rate achieved for age 50, often associated with better cardiovascular fitness.
Oldpeak	1.00	Mild ST depression induced by exercise. This is an abnormal ECG response that can indicate ischemia.
Number of Major Vessels	0	No major vessels with significant blockages seen on fluoroscopy. This is a very strong positive indicator and often the most significant factor in a low-risk prediction.
Thalassemia (Thal)	2	Likely indicates a 'reversible defect' meaning some areas of the heart muscle show inadequate blood flow under stress (ischemia). A risk factor.
Fitness Index	50	A moderate value, suggesting average functional capacity.
Slope of Peak Exercise ST Segment	2	Flat slope. This is an abnormal and concerning ECG finding, often associated with coronary artery disease.

This means that there is stress that the heart is undergoing during physical activity and it is recovering in rest. The ST_slope is 2 which means that there is some abnormality during physical exercise.

So, the model has predicted that the patient falls under class 2 with over 70% of probability which means that the patient might falls under moderate heart disease.

In the below figure 6.6 where it shows the probability of individual class predictions of class 2, where the model confidently predicted that the patient falls under class 2 with over 70% of probabilty. Due to the above better health vitals that were mentioned above, the model has predicted the probability of over 12% and 18% that the patient might falls under class 0 and class 1. In this scenario the medical industry experts must consider the border line flags into consideration to make the right predictions and we can clearly see that our model is taking right parameters into considerations for predicting.

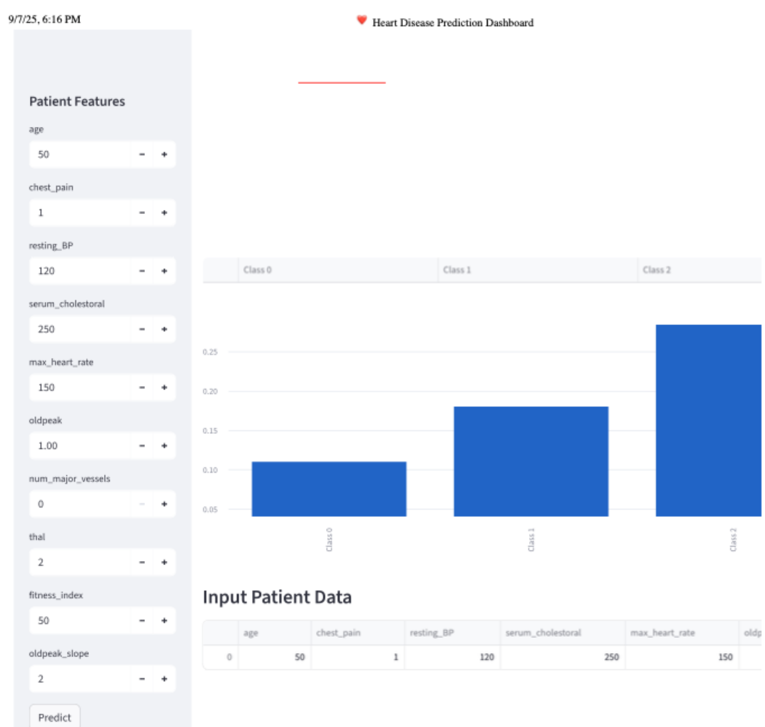

Figure 6.6. Probability of individual class Predictions - Output 2

- **Output 3**: In the below table 6.4 which shows the clinical interpretation of the patient features of class 3. In the below table we can see that the patient has 4 different sevearity levels that gives highest possibility for cardio vascular diesease of class 3. The patient had chest_pain type 1, which is very strongly assosiated with heart disease, borderline cholesterol of 209 and old_peak 1 which is compeletely abnormal during physical activity. The thallesemia is 2 which is a very significant risk factor as well. The fitness_index is 0 which is a very poor vital that shows there is no physical fitness for the patient.

Table 6.4. Clinical interpretation of patient features (Case 3)

Feature	**Value**	**Clinical Interpretation**
Age	46	Young adult. Early onset of heart disease is concerning.
Chest Pain Type	1	Typical angina. Chest pain strongly associated with heart disease.
Resting BP	135	High-Normal or Stage 1 Hypertension. A risk factor.
Serum Cholesterol	209	Borderline High. A modifiable risk factor.
Max Heart Rate	160	Good maximum heart rate for age 46.
Oldpeak	1.0	1.0 mm of ST depression. This is a significantly abnormal ECG response to exercise, indicating a high likelihood of ischemia (inadequate blood flow to the heart). This is one of the strongest predictors in the dataset.
Num Major Vessels	0	No major vessels blocked. This is a positive sign but is being overruled by other strong negative indicators.
Thal (Thalassemia)	2	Likely indicates a 'fixed defect'. This suggests an old, prior heart attack (myocardial infarction) where heart muscle tissue has been permanently damaged. This is a very significant risk factor.
Fitness Index	0	A very low value, suggesting very poor functional capacity and inability to perform well on the stress test.
Slope of Peak	1	Upsloping ST segment. This is a normal finding, but it is contradicted by the highly abnormal oldpeak value.

The below figure 6.7 that shows the probability of individual class prediction for class 3, the model has highest weightage of 40% that the patient falls under class 3. The model had similar predictions for class 1, 2, 3 as well.

Our model is able to strongly predict that this patient falls under the risk of getting heart disease but it is not likely to differentiate between class 1, 2, and 3 because of borderline health vitals. There are few conflicting vital signs that is confusing the model like number_major_vessels in this example. For the medical industry experts to trust this model, we have to further train our model with the similar patients records that will help it to understand the hidden features and give results with more probability towards the predicted class.

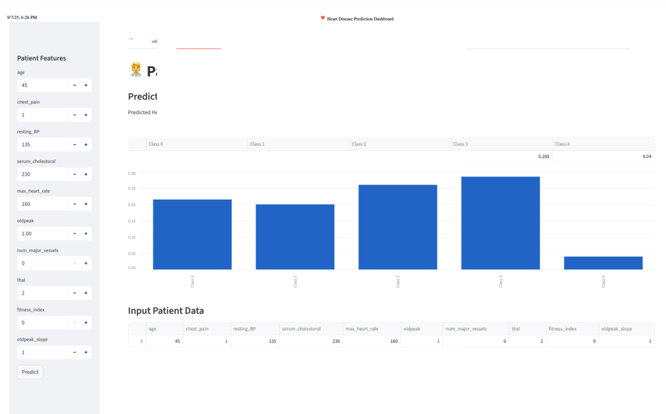

Figure 6.7. Probability of individual class Predictions - Output 3

- **Output 4**: In the below table 6.5 where it shows the clinical interpretation of patient features of class 3. The below patient's input that we have given have several health risk factors where the patient is suppose to be placed in class 4 risk but the model predicted class 3. This is the limitation of our model where we are suppose to further tune the model with further extreme patients so that it can classify the patients into extreme risk class. This limitation is occuring because we do not have enough patients data after 80 years of age where the model is not able to classify such patients data. In this example, the patient has chest_pain type as 3 which is considered as non-anginal pain, severe serum_cholesterol levels with 300mg/dl, 160 blood pressure with hypertension, ST_Slope value as 3 which is very abnormal, number_major_vessels as 3

  is considered to be triple vessel disease. The patient also has very poor

Table 6.5. Clinical interpretation of patient features (Case 3)

Feature	Value	Clinical Interpretation
Age	75	Advanced age. This is a major non-modifiable risk factor. The risk of CAD increases significantly with age.
Chest Pain Type	3	Non-anginal pain. While not the classic "typical angina," any chest pain in this context is taken seriously and can be a symptom of heart disease.
Resting BP	160	Stage 2 Hypertension. This is severely high blood pressure, putting immense strain on the heart and arteries.
Serum Cholesterol	300	Severely High Cholesterol. This is a critical level, significantly accelerating the buildup of plaque in the arteries (atherosclerosis).
Max Heart Rate	120	Low maximum heart rate. For a 75-year-old, this is low and suggests the patient could not tolerate much physical stress on the heart, often due to existing disease.
Oldpeak	3.00	Profound ST depression. This is a markedly abnormal ECG finding. It indicates severe ischemia (the heart muscle is severely starved of oxygen during stress). This is one of the strongest predictors in the dataset.
Num Major Vessels	3	Three vessels diseased. This means three out of the four major coronary arteries have significant blockages. This is known as triple-vessel disease and is a very serious condition requiring aggressive intervention.
Thal (Thalassemia)	3	'Reversible defect'. This means a large area of the heart muscle shows inadequate blood flow under stress but recovers at rest. This is direct evidence of widespread, significant ischemia.
Fitness Index	0	Very poor functional capacity. The patient has extremely low exercise tolerance, consistent with severe heart disease.
Slope of Peak	0	Downsloping ST segment. This is the most abnormal and ominous finding on an exercise ECG, strongly associated with significant coronary artery disease.

fitness_index of 0, Thallesemia as 3 which is irreversible heart ischemia.

In the below figure 6.8 which shows probability of individual class predictions - class 4, the model has wrongly predicted that the patient falls under class 3. The reason the model was not able to predict accurately is because this model has limitation of not categorizing higher classes i.e., class 3 and class4. But since the model is giving probabilities the medical industry experts can understand why the model is classifying that way.

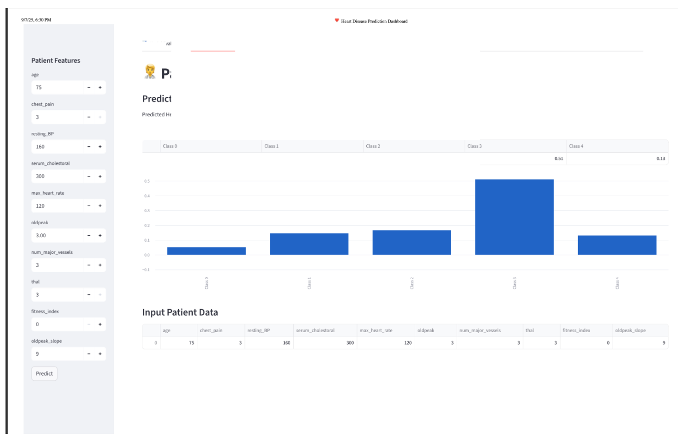

Figure 6.8. Probability of individual class Predictions - Output 4

6.4 Limitations of the Model

Currently we trained the model on the real time data where we have limited records of age below 20 and above 80 years of age, where the model is not completely trained on the data. To overcome this limitation we can colaborate with medical industry experts to synthetically generate the data to further train the data. Since the real heart disease data is very limited, we can try to synthetically generate meaningful data of patients and train the model with good amount of permutations. This will help the model to better understand the permutations on every class of patients. This will result in creating a robust model which performs better with real-time data.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this research, we have taken a novel approach of creating new features that has not been explored in any other research and achieved a good F1 score of 80%. We have build computationally efficient models to classify the target variables. Currently we are able to distinguish the patients between no heart disease (0) and heart disease (1,2,3, and 4) but the further classification of mild, moderate, severe, and very severe heart diseases are not being classified. We have created new features such as age group, high cholesterol, high blood pressure, fitness index, old peak slope interaction, and age-exercise angina where these features allows the model to capture the hidden trends present in the heart disease dataset. This feature engineering strategy significantly improves the machine learning models to better predict the heart disesase dataset. Among the tested models, the ensemble techniques for random forest, AdaBoost and Gradient boosting has given better results than the base-line models. This demostrates that the feature engineering technique is not only useful to improve prediction of the dataset, but also improves the classification of the target variable. Further the integration of XAI techniques like SHAP has provided better explanations for the model's decisions and provided a transparent view of individual's features towards the contribution of final decision. This is very critical in medical industry where trustworthiness and interpretability are two fundamentals.

In this research, SHAP revealed that chest_pain, number_major_vessels are the key contributors for the positive predictions. This inlines with the existing medical knowledge of ours.

Overall, this study demonstrates good feature engineering techniques in combine with intepretable machine learning models which offers a decision support system for early detection of cardio vascular diseases.

7.2 Discussion and Future Work

The future work of this study can be creating features that helps the model to understand the realationships between higher classes (differences in class 2,3,4) and improving the performance of classification. Researchers can also fine tune the models to explore the hidden trends in the newly created features based on the SHAP outputs. We can also perform future work on the novel approach that we have taken in this research that is the feature engineering techniques. Researchers can create new features that will be able to further explore the trends in the dataset and better classify the severity of the heart disease. The future research can be focussed on building a model that can be able to classify the class 2 and 3 records as well. Another line of future work can be done to enhance the predictability by using the deep learning models. These advanced models will show better performance after fine-tunig them. We need not have to worry about the algorithm as they are black box models, we integrate XAI techniques for model interpretability. In terms of explainability perspective, we can involve Local Interpretable Model Agnostic Explanations (LIME) techniques to validate model predictions. This technique will further help the researchers to understand the reasoning behind the model predictions. Finally, we can extend this research in terms of integrating with live medical data. This will further bridge the gap between machine learning and real-time usage on medical data. This will increase the trust in medical practioners so they incorporate AI in healthcare industry and finally contribute to more effective management of cardiovascular diseases.

Bibliography

1. Sethi, A., Dharmavaram, S. & Somasundaram, S. *Explainable artificial intelligence (xai) approach to heart disease prediction* in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (2024), 1–6.

2. Rezk, N. G., Alshathri, S., Sayed, A., El-Din Hemdan, E. & El-Behery, H. XAI-Augmented Voting Ensemble Models for Heart Disease Prediction: A SHAP and LIME-Based Approach. *Bioengineering* **11,** 1016 (2024).

3. Aggarwal, R., Podder, P. & Khamparia, A. in *Biomedical data analysis and processing using explainable (XAI) and responsive artificial intelligence (RAI)* 91–103 (Springer, 2022).

4. World Health Organization. *Cardiovascular Diseases* https://www.afro.who.int/health-topics/cardiovascular-diseases (2025).

5. Absar, N. *et al. The efficacy of machine-learning-supported smart system for heart disease prediction* in *Healthcare* **10** (2022), 1137.

6. Shrestha, D. Advanced Machine Learning Techniques for Predicting Heart Disease: A Comparative Analysis Using the Cleveland Heart Disease Dataset. *Applied Medical Informatics* **46** (2024).

7. Rajendran, R. & Karthi, A. Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers. *Expert Systems with Applications* **207,** 117882 (2022).

8. Chandrasekhar, N. & Peddakrishna, S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes* **11,** 1210 (2023).

9. Ahmad, B., Chen, J. & Chen, H. Feature selection strategies for optimized heart disease diagnosis using ML and DL models. *arXiv preprint arXiv:2503.16577* (2025).

10. Mehta, H. & Passi, K. Social media hate speech detection using explainable artificial intelligence (XAI). *Algorithms* **15,** 291 (2022).

11. Lee, H. & Tsoi, P. Feature-Enhanced Machine Learning for All-Cause Mortality Prediction in Healthcare Data. *arXiv preprint arXiv:2503.21241* (2025).

12. Qadri, A. M., Raza, A., Munir, K. & Almutairi, M. S. Effective feature engineering technique for heart disease prediction with machine learning. *IEEE Access* **11,** 56214–56224 (2023).

13. Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N. & Alarfaj, F. K. XAI framework for cardiovascular disease prediction using classification techniques. *Electronics* **11,** 4086 (2022).

14. Mhamdi, L., Dammak, O., Cottin, F. & Dhaou, I. B. Artificial intelligence for cardiac diseases diagnosis and prediction using ECG images on embedded systems. *Biomedicines* **10,** 2013 (2022).

15. Sharma, V., Yadav, S. & Gupta, M. *Heart disease prediction using machine learning techniques* in *2020 2nd international conference on advances in computing, communication control and networking (ICACCCN)* (2020), 177–181.

16. El-Sofany, H., Bouallegue, B. & El-Latif, Y. M. A. A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific reports* **14,** 23277 (2024).

17. Alom, Z., Azim, M. A., Aung, Z., Khushi, M., Car, J. & Moni, M. A. *Early stage detection of heart failure using machine learning techniques* in *Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021* (2021), 75–88.

18. Gupta, S. K., Shrivastava, A., Upadhyay, S. P. & Chaurasia, P. K. A machine learning approach for heart attack prediction. *International Journal of Engineering and Advanced Technology (IJEAT)* **10,** 1–11 (2021).

19. Gupta, C., Saha, A., Reddy, N. S. & Acharya, U. D. *Cardiac Disease Prediction using Supervised Machine Learning Techniques.* in *Journal of physics: conference series* **2161** (2022), 012013.

20. Ward, A. *et al.* Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. *NPJ digital medicine* **3,** 125 (2020).

21. SHAP Documentation. *An introduction to explainable AI with Shapley values* https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html (2025).