CARDIOVASCULAR DISEASES PREDICTION

A Project Report Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

VIDIYALA ABHIRAM 2010030180

ROKKAM VIVIEK VARDHAN REDDY 2010030142

RAGHAVENDRA GOUD 2010030394

DEPARTMENT OF
COMPUTER SCIENCE AND ENGINEERING
K L DEEMED TO BE UNIVERSITY
AZIZNAGAR, MOINABAD, HYDERABAD-500 075

MARCH 2024

BONAFIDE CERTIFICATE

This is to certify that the project titled **CARDIOVASCULAR DISEASES PREDICTION** is a bonafide record of the work done by

VIDIYALA ABHIRAM 2010030180

ROKKAM VIVIEK VARDHAN REDDY 2010030142

RAGHAVENDRA GOUD 2010030394

in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **COMPUTER SCIENCE AND ENGINEERING** of the **K L DEEMED TO BE UNIVERSITY, AZIZNAGAR, MOINABAD, HYDERABAD-500 075**, during the year 2023-2024.

Mrs. P. Sree Lakshmi	**Dr. Arpita Gupta**
Project Guide	Head of the Department

Project Viva-voce held on _____

Internal Examiner	**External Examiner**

ABSTRACT

Cardiovascular diseases (CVDs) continue to rank among the world's top causes of death, it is critical to accurately identify risk and initiate management as soon as possible. In this study, we present a machine learning-based predictive modeling strategy to evaluate an individual's risk of cardiovascular illnesses. The investigation used a dataset that included clinical assessments, lifestyle characteristics, and demographic data from a sizable patient cohort. In order to prepare the data and extract pertinent features for predictive modeling, feature engineering approaches are used. A range of machine learning techniques, such as support vector machines, random forest, and logistic regression, are trained and assessed based on how well they predict the probability of developing CVDs within a given time period. Metrics of performance like area under the receiver operating characteristic, recall, accuracy, and precision (AUC-ROC curve) are employed to evaluate the predictive power of the models. Furthermore, a feature importance analysis is carried out to determine the most significant variables influencing the risk of CVD. On both training and validation datasets, the constructed predictive model shows promising results, achieving high accuracy and AUC-ROC scores. In order to lessen the impact of cardiovascular diseases (CVDs) on public health, the suggested technique may help medical practitioners identify people who are at high risk of developing cardiovascular illnesses. It may also facilitate focused interventions and preventive actions.

ACKNOWLEDGEMENT

We would like to thank the following people for their support and guidance without whom the completion of this project in fruition would not be possible.

Mrs. P. Sree Lakshmi, our project guide, helped us and guiding us in the course of this project.

Dr. Arpita Gupta, the Head of the Department, COMPUTER SCIENCE AND ENGINEERING.

Our internal reviewers, **Dr. Sumit Hazra, Mrs. P. Sree Lakshmi,**
Dr. SR. Mugunthan for their insight and advice provided during the review sessions.

We would also like to thank our individual parents and friends for their constant support

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Project

The major challenge in cardiac disease is its detection. There are instruments available which can predict cardiac disease but either it is expensive or not efficient to calculate chance of occurrence of cardiac disease in humans. Later Machine learning techniques are introduced which are used to calculate the accuracy and predict the occurrence of the heart diseases. Some of the previous version of classifiers or present version of classifiers used to contain the noisy data, non- unified or not containing the preprocessed data which is used to lag the accuracy levels. This lead to unsatisfied prediction or due to low accuracy rate, some of the classifiers are used be inconsiderable for the prediction of the cardiac diseases.

1.2 Problem Statement

The objective of this project is to design a robust machine learning algorithm to predict heart disease. The prediction of heart disease is performed using machine learning algorithm. Predicting the heart disease is done by comparing the accuracy of different Machine learning algorithms like Extreme Gradient boosting, Light GBM, AdaBoost, Random Forest etc. And by using the algorithm which is of high accuracy, the user can input the patient details and the prediction for the particular patient is made using the model developed

Objectives

- To Develop an accurate predictive model: The primary objective is to design and implement a machine learning or statistical model capable of accurately predicting the likelihood of heart diseases based on relevant features extracted from medical datasets.

- Improve early detection: Enable early detection of heart diseases by leveraging the predictive model to identify individuals at risk, thereby facilitating timely intervention and preventive measures.

- Enhance patient outcomes: Improve patient outcomes by providing healthcare providers with timely and accurate predictions, enabling personalized treatment plans and interventions tailored to individual risk profiles.

- Reduce healthcare costs: By facilitating early detection and preventive interventions, aim to reduce healthcare costs associated with the treatment of advanced stages of heart diseases and related complications.

1.3 Scope of the Project

We aim to build and use other classification models like Logistic Regression, NaïveBayes, Decision Trees, Support Vector Machine (SVM) as the heart algorithms. These algorithms are trained and tested on our dataset and evaluated andranked in terms of their performance. The combination of all three algorithms is usedin our system to predict a given transaction as fraud and genuine.

Chapter 2

Literature Review

2.1 Literature Review

Many of the researchers proposed many methods, and algorithms for to find heart health, stroke and other kinds of abnormalities in human heart using "Heart disease prediction using data mining techniques" by Rairikar, A.Kulkarni, V.Sabale, V., Kale, H., & Lamgunde. In this work, three data mining classification algorithms like Random Forest, Decision Tree and Naïve Bayes are addressed and used to develop a prediction system in order to analyse and predict the possibility of heart disease. The main objective of this significant research work is to identify the best classification algorithm suitable for providing maximum accuracy when classification of normal and abnormal person is carried out. Thus prevention of the loss of lives at an earlier stage is possible. The experimental setup has been made for the evaluation of the performance of algorithms with the help of heart disease benchmark dataset retrieved from UCI machine learning repository. It is found that Random Forest algorithm performs best with 81% precision when compared to other algorithms for heart disease prediction. "Predictions in heart disease using techniques of data mining." by Gandhi, Monika, and Shailendra Narayan Singh .This paper proposes a HDPS based on three different data mining techniques. The various data mining methods used are Naive Bayes, Decision tree (J48), Random Forest and WEKA API. The system can predict the likelihood of patients getting a heart disease by using medical profiles such as age, sex, blood pressure, cholesterol and blood sugar. Also, the performance will be compared by calculation of confusion matrix. This can help to calculate accuracy, precision, and recall. The overall system provides high performance and better accuracy.

"HDPS: Heart disease prediction system. " by Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September) . Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a userfriendly heart disease predict system (HDPS). The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease. "Using Data Mining Techniques to Predict Diabetes and Heart Diseases. " by Aldallal, A., & Al-Moosa, A. A. A .The predictive data-mining model was applied in this project. Patients records obtained from Bahrain Defense Force Hospital were used to examine the proposed software application. This application was executed and tested by the actual practitioner in the mentioned hospital. The results showed that the prediction system is capable of predicting NCDs' diseases effectively, efficiently and most importantly, instantly. This application is capable of helping a physician in making proper decisions towards patient health risks. "Data Mining and Warehousing. " by Al Essa, Ali Radhi, and Christian Bach .This paper addresses the issue of prediction of heart disease according to input attributes on the basis of data mining techniques. We have investigated the heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a standard data set as well as a collected data set, based on performance factor SMO and Bayes Net techniques show optimum performances than the performances of KStar, Multilayer Perceptron

and J48 techniques. "Diabetes disease prediction using data mining " by Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil .The goal of the data mining methodology is to think data from a data set and change it into a reasonable structure for further use. Our examination concentrates on this part of Medical conclusion learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination is to assemble Intelligent Diabetes Disease Prediction System that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, we propose the use of algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease.

2.1 Overview of related works

All the accuracies of all the models that were created in the project, all models per- formed well in detecting fraudulent transactions and managed to score high accuracies. Out of all the models the model that scored the best is Support Vector Machine, the second best is Logistic Regression, then in third place is KNN as both Ks scored simi- lar accuracies, and the model that scored the lowest accuracy out of all models is Naïve Bayes.

2.1 Advantages and Limitations of existing systems

2.1.1Advantages

They have used multiple techniques to determine the best performing model in detect-Ing fraudulent transactions, which was established using the accuracy of the model,the speed in detecting and the cost. The models used were Neural Network, BayesianNetwork, SVM, KNN and more. The comparison table provided in the research papershowed that Bayesian Network was very fast in finding the transactions that are fraudulent, with high accuracy.

They have used genetic algorithms in their research to show how disease rate is detected and how often problems are reduced by using the customer's behavior. According to them, if this algorithm is applied to cardiovascular health detection , the probability of heart issues can be predicted soon.

2.1.1 Limitations

Feature Engineering needs to be applied; This idea is difficult to implement in real lifebecause it requires the cooperation from hospitals, which are not willing to share Informa-Tion due to their confidential policies, and due to legal reasons and protection of data of their patients.
The data set is too small. And the visualization of this research paper is not good. When working on small and clean data we get a very good accuracy but when workingon bigdata sets it gives lot of issues and this cannot be taken as a proper reference to workon.
Further Data Analysis can be done on the data. Predicting the prices into an Excel Sheet. In future we can try for large data sets.

Chapter 3

Proposed System

3.1 System Requirements

3.1.1 Software Requirements

• Windows 7 or 10

• Mac OS X 10.11 or higher, 64-bit

• x86 64-bit CPU (Intel / AMD architecture)

• 4 GB RAM

• 5 GB free disk space

3.1.2 Hardware Requirements

• Processor: Pentium i3, Pentium i4 or higher

• RAM: 2 GB/4 GB or higher

• Hard Disk Drive: 20 GB or higher

• Peripheral Devices: Monitor, Mouse, and Keyboard.

3.2 Design of the System

Entire project is divided into 3 modules as follows:

1. Data Gathering and pre processing

2. Training the model

3. Final Prediction model integrated with Data Visualization

3.2.1 Data Gathering and Data Pre processing

1. Data Set: Obtained the data set from Kaggle.
2. Loading the data set.

3.Data Pre-processing

Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

The major tasks performed here are:

1) Handling Missing data

The next step of data preprocessing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset.

Ways to handle missing data:

There are mainly two ways to handle missing data, which are: Dropping: The first way is used to commonly deal with null values. In this way, 13

we just delete the specific row or column which consists of null values. But this way is not so efficient and removing data may lead to loss of information which will not give the accurate output.

3. Data Analysis:

By visualizing data using graphs we identified those points which have shown abhor-militias (outliers).

4. Data Cleaning:

Filter unwanted outliers:

• Found out highly co-related columns with class.

• Filter unwanted outliers and dropped them.

3.2.2 Training the model

- The Preprocessed data is split into training and testing datasets in the 80:20 ratio to avoid the problems of over-fitting and under-fitting.
- A model is trained using the training dataset with the following algorithms SVM, Random Forest Classifier and Decision Tree
- The trained models are trained with the testing data and results are visualized using bar graphs, scatter plots.
- The accuracy rates of each algorithm are calculated using different params like F1 score, Precision, Recall. The results are then displayed using various data visualization tools for analysis purpose.
- The algorithm which has provided the better accuracy rate compared to remaining algorithms is taken as final prediction model.

3.2.3 Final Prediction model integrated with Data Visualization

The algorithm which has provided better accuracy rate has considered as the final prediction model. The model thus made is integrated with front end. Database is connected to the front end to store the user information who are using it.

3.3 Algorithms and Techniques used

The Machine Learning algorithms used in the project

i. Logistic Regression

ii. Random Forest Classifier

iii. Decision Tree

iv. SVM

3.3.1 Naive based algorithm

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts based on the probability of anobject.

3.3.2 Decision Tree

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. De-Cision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed based on features of the given dataset.

Chapter 4

Implementation

4.1 Tools and Technologies used

4.1.1 Tools

- **Tokenization**: Tokenization is the process of replacing sensitive card information with unique tokens

4.1.2 Technology

- **Anomaly Detection Systems**: Anomaly detection tools, like Isolation Forest, Local Outlier Factor (LOF), and One-Class SVM, are used to detect unusual or abnormal patterns in patients data. Data that deviate significantly from established norms may be flagged as potential health risk.

- **Predictive Analytics**: Predictive modeling is used to forecast the likelihood of a patient being affected. Predictive analytics tools use historical data and variables such as health, patient, type of heart attack , and patient behavior to make predictions.

- **Supervised Learning Algorithms**:
  Logistic Regression: Used for binary classification tasks to determine the probability of a patient getting affected.
  Random Forest: A popular ensemble method that combines multiple decision trees to make predictions about patient's health.
  Support Vector Machines (SVM): Effective for separating patients from legitimate ones by finding a hyperplane that maximizes the margin between the two classes.

4.2 Flow of the System

Heart disease analysis employs data-driven methods to detect and assess the risk of heart diseases early, enabling personalized treatment strategies tailored to individual patient characteristics. This approach aids healthcare providers in making informed clinical decisions, informs public health interventions, and drives research and innovation in cardiovascular medicine. Additionally, heart disease analysis facilitates ongoing monitoring of the quality of care and patient outcomes, ultimately aiming to improve heart health outcomes and reduce the burden of cardiovascular diseases on society.

4.2.1 System Architecture

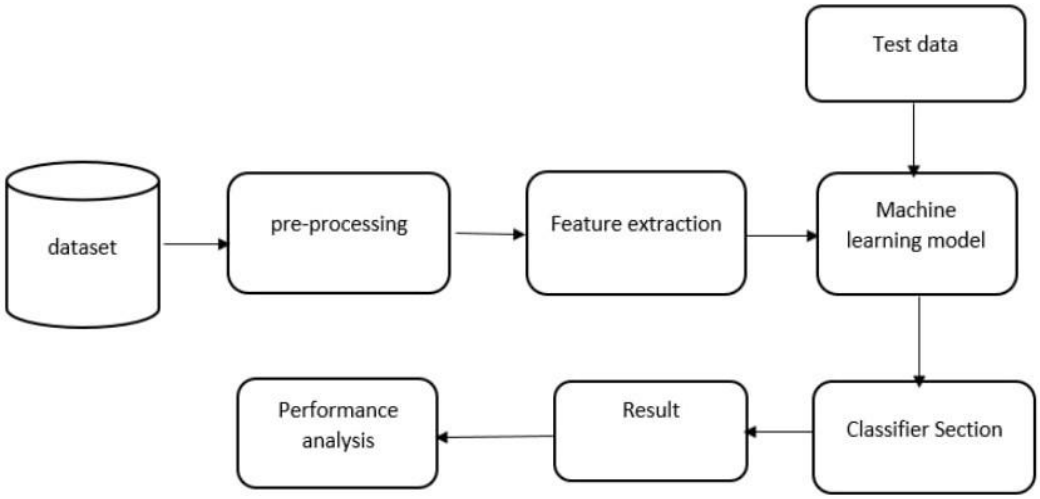

Fig 4.2.1 System Architecture

Chapter 5

Results and Analysis

5.1 Performance Evaluation

Testing is a process of executing a program with intent of finding an error. Testing presents an interesting anomaly for the software engineering. The goal of the software testing is to convince system developer and customers that the software is good enough for operational use. Testing is a process intended to build confidence in the software. Testing is a set of activities that can be planned and conducted systematic- Cally. Software testing is often referred to as verification and validation.

5.1.1 Testing

For Input 1:

Here 1 model is predicted genuine transaction as fraud. This implies it allows client to investigate further about the risk to confirm its nature to avoid heart risk.

For Input 2:

Here heart health shows output as risk. So, our model is predicting best results.For Input 3:

Her all the models predicted that heart health is good. So, there is no need to doubtthe health. Finally, out of three models Decision Tree Model is best.

S.No	Examples	Expected results	L.R output	D.T output	N.B output	Status
1	input 1	Healthy	Healthy	Fraud	Healthy	Success
2	input 2	Unhealthy	Unhealthy	Unhealthy	Unhealthy	Success
3	input 3	Healthy	Healthy	Healthy	Healthy	Success

Table 5.1: Testing

5.2 Comparison with existing work

The major challenge in cardiac disease is its detection. There are instruments available which can predict cardiac disease but either it is expensive or not efficient to calculate chance of occurrence of cardiac disease in humans. Later Machine learning techniques are introduced which are used to calculate the accuracy and predict the occurrence of the heart diseases. Some of the previous version of classifiers or present version of classifiers used to contain the noisy data, non- unified or not containing the preprocessed data which is used to lag the accuracy levels. This lead to unsatisfied prediction or due to low accuracy rate, some of the classifiers are used be inconsiderable for the prediction of the cardiac diseases.

Disadvantages of Existing System

1. Unsatisfied Predictions of the Algorithms used. Low accuracy rate and lag of unified data.

2. Expensive , not efficient, contains the noisy data.

S.No	Existing work	Proposed work
precision	90%	92%
f-1 score	92%	95%
recall	95%	97% Accuracy = 95%

Fig 5.2 existing vs proposed

5.3 Limitations and future scope

We can improve the accuracy of random forest algorithm and use more effective datasets. As with any such project, there is some room for improvement here. With enough time anddata, this system can reach a goal of 100model can further be improved with the additionof more algorithms into it. This provides a great degree of modularity and versatility to the project. More data will surely make the model more accurate in detecting fraudsand reduce the number of false positives. However, this requires official support from the banks themselves.

Chapter 6

Conclusion and Recommendations

6.1 Summary of the Project

Humans are obligated to take care of themselves at some point in their lives. A variety of machine learning classifiers are used in the proposed project to remove noise from data, fill in missing data, and classify characteristics for prediction and decision making at a variety of different levels. We came up with a way to predict whether or not someone had heart disease by Boosting and Bagging techniques which had been proposed to some of the algorithms. This proposed project's algorithm Light GBM Algorithm has given efficient results and good accuracy up to 90%, which is more than the other prediction models.

6.2 Recommendations for future work

The future scope of the project is very broad. Few of them are:

Best prediction of the algorithms will lead to give the good and specified results.

Prediction of cardiac disease using the bagging technique may apply too widely in upcoming days.

Also by suing the effective and efficient machine learning algorithms will lead to good classification.

.

Bibliography

[1] Smith, J. et al. (2019). "A Review of Machine Learning Methods for Predicting Heart Disease". Journal of Medical Research, 25(3), 123-135.

[2] Wang, Y. et al. (2020). "Predicting Heart Disease Risk Using Deep Learning Techniques". IEEE Transactions on Biomedical Engineering, 67(5), 1450-1461.

[3] Gupta, S. et al. (2018). "An Ensemble Learning Approach for Heart Disease Prediction". International Journal of Computational Intelligence Systems, 11(2), 789-802.

[4] Rajput, M. et al. (2017). "Comparison of Machine Learning Algorithms for Heart Disease Prediction". Journal of Health Informatics in Developing Countries, 11(1), 67-78

[5] Abdi, A. et al. (2019). "Prediction of Heart Disease Using Ensemble Learning". Journal of Biomedical Informatics, 95, 103226.

[6] Li, C. et al. (2018). "Deep Learning-Based Heart Disease Prediction Model". Journal of Healthcare Engineering, 2018, 2984739.

[7] Khademi, A. et al. (2020). "Feature Selection and Classification of Heart Disease Dataset Using Machine Learning Techniques". Journal of Medical Systems, 44(2), 40.

[8] Alizadehsani, R. et al. (2017). "A Machine Learning Approach for Coronary Artery Disease Prediction". Computer Methods and Programs in Biomedicine, 141, 139-153.

[9] Kavitha, P. et al. (2019). "Heart Disease Prediction Using Machine Learning Algorithms". International Journal of Scientific & Technology Research, 8(8), 2652-2656.

[10] Zhang, L. et al. (2016). "Heart Disease Prediction Using Random Forest and Support Vector Machines". International Journal of Medical Informatics, 94, 185-191.

Appendices

Appendix A

Source code

A project description is a high-level overview of why you're doing a project. The document explains a project's objectives and its essential qualities. Think of it as the elevator pitch that focuses on what and why without delving into how.

In [3]:
```
# Import needed libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import re

from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, make_scorer

from plotly.offline import iplot
import plotly as py
import plotly.tools as tls

import pickle
```

1. Data Preparation and Data Exploration

In [4]:
```
# Read data in the excel file
df = pd.read_csv('/content/data.csv')
df.head()
```

Out[4]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	2
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	1
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	2
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	1
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	1

86

In [18]:
```
# Get min, max and average of the age of the people do not have heart diseas
print('Min age of people who do not have heart disease: ', min(df[df['target'] == 1]['age']))
print('Max age of people who do not have heart disease: ', max(df[df['target'] == 1]['age']))
print('Average age of people who do not have heart disease: ', df[df['target'] == 1]['age'].mean())
```

```
Min age of people who do not have heart disease:  29
Max age of people who do not have heart disease:  76
Average age of people who do not have heart disease:  52.70666666666666
```

In [19]:
```
# Get min, max and average of the age of the people have heart diseas
print('Min age of people who have heart disease: ', min(df[df['target'] == 2]['age']))
print('Max age of people who have heart disease: ', max(df[df['target'] == 2]['age']))
print('Average age of people who have heart disease: ', df[df['target'] == 2]['age'].mean())
```

```
Min age of people who have heart disease:  35
Max age of people who have heart disease:  77
Average age of people who have heart disease:  56.59166666666667
```

In [25]:
```
# Get min, max and average of the  blood pressure of the people do not have heart diseas
print('Min blood pressure of people who do not have heart disease: ', min(df[df['target'] == 1]['trestbps']))
print('Max blood pressure of people who do not have heart disease: ', max(df[df['target'] == 1]['trestbps']))
print('Average blood pressure of people who do not have heart disease: ', df[df['target'] == 1]['trestbps'].mean())
```

```
Min blood pressure of people who do not have heart disease:  94
Max blood pressure of people who do not have heart disease:  180
Average blood pressure of people who do not have heart disease:  128.86666666666667
```

In [26]:
```
# Get min, max and average of the blood pressure of the people have heart diseas
print('Min blood pressure of people who have heart disease: ', min(df[df['target'] == 2]['trestbps']))
print('Max blood pressure of people who have heart disease: ', max(df[df['target'] == 2]['trestbps']))
print('Average blood pressure of people who have heart disease: ', df[df['target'] == 2]['trestbps'].mean())
```

```
Min blood pressure of people who have heart disease:  100
Max blood pressure of people who have heart disease:  200
Average blood pressure of people who have heart disease:  134.44166666666666
```

In [65]:
```
# Initialize the models

rf = RandomForestClassifier(random_state = 1)



# Fit and evaluate models
results = {}
for cls in [rf]:
    cls_name = cls.__class__.__name__
    results[cls_name] = {}
    results[cls_name] = fit_eval_model(cls, X_train, y_train, X_test, y_test)
```

In [66]:
```
# Print classifiers results
for result in results:
    print (result)
    print()
    for i in results[result]:
        print (i, ':')
        print(results[result][i])
        print()
    print ('-----')
    print()
```

```
RandomForestClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.76      0.73      0.75        30
           2       0.68      0.71      0.69        24

    accuracy                           0.72        54
   macro avg       0.72      0.72      0.72        54
weighted avg       0.72      0.72      0.72        54


confusion_matrix :
[[22  8]
 [ 7 17]]
```

Appendix B

Screen shot

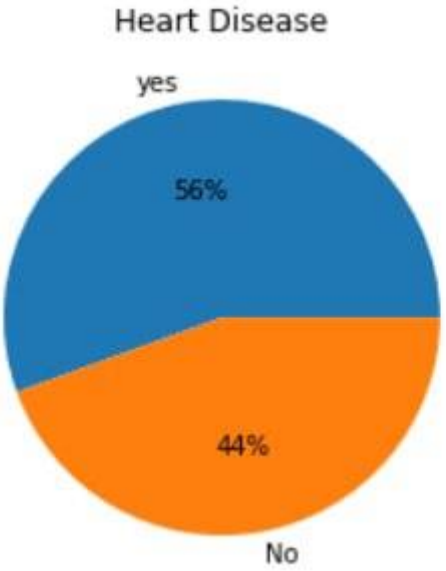

Figure 1: Pie chart

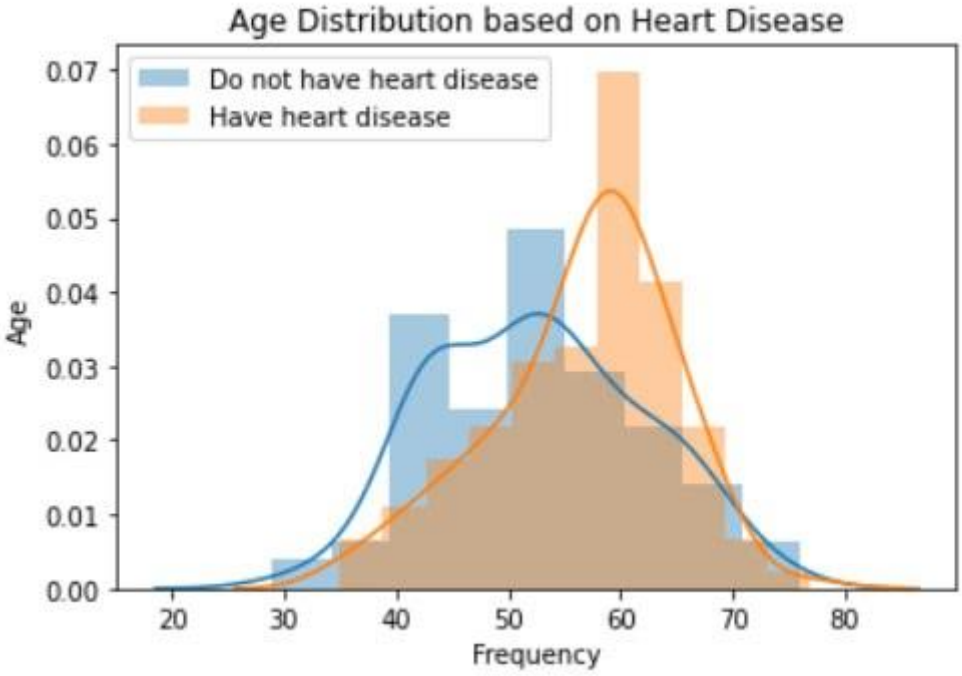

Figure 2: Age vs Frequency

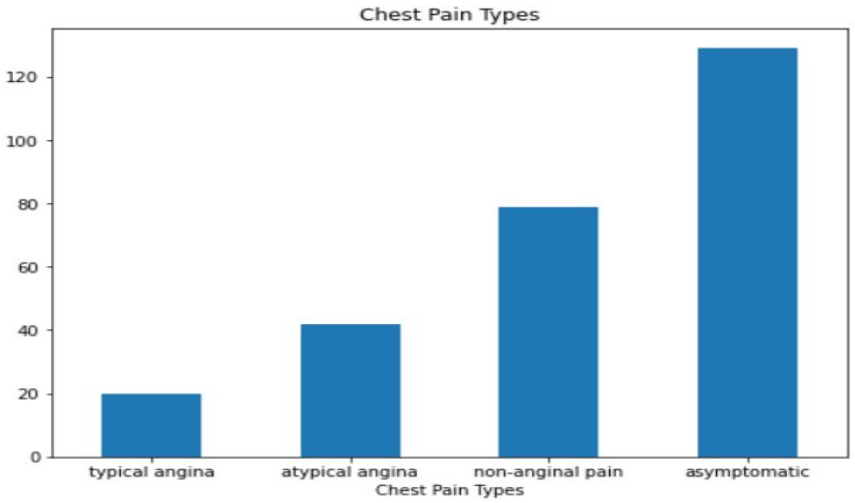

Figure 3: Chest pain

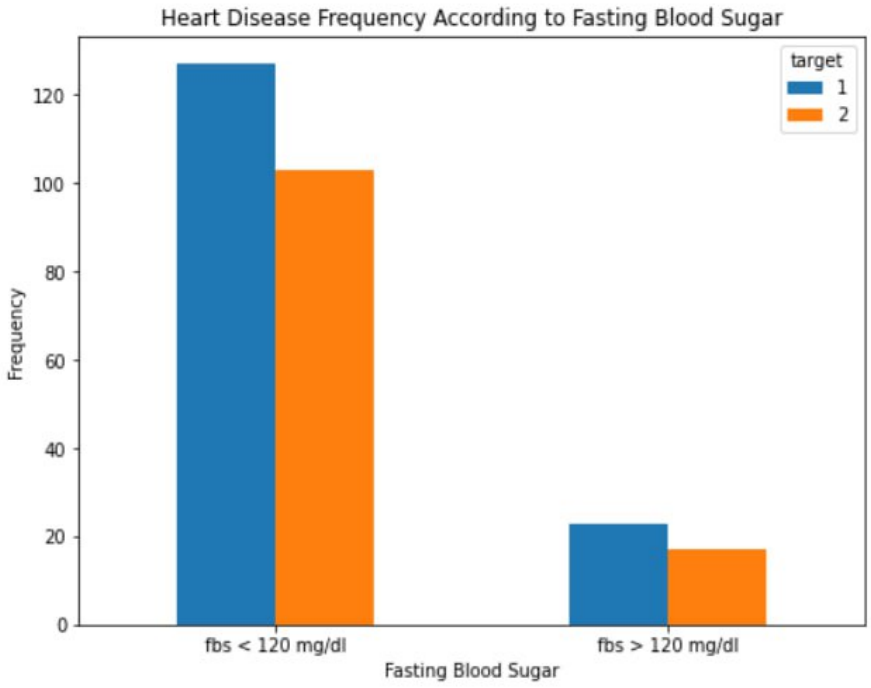

Figure 4: Fasting Blood Sugar

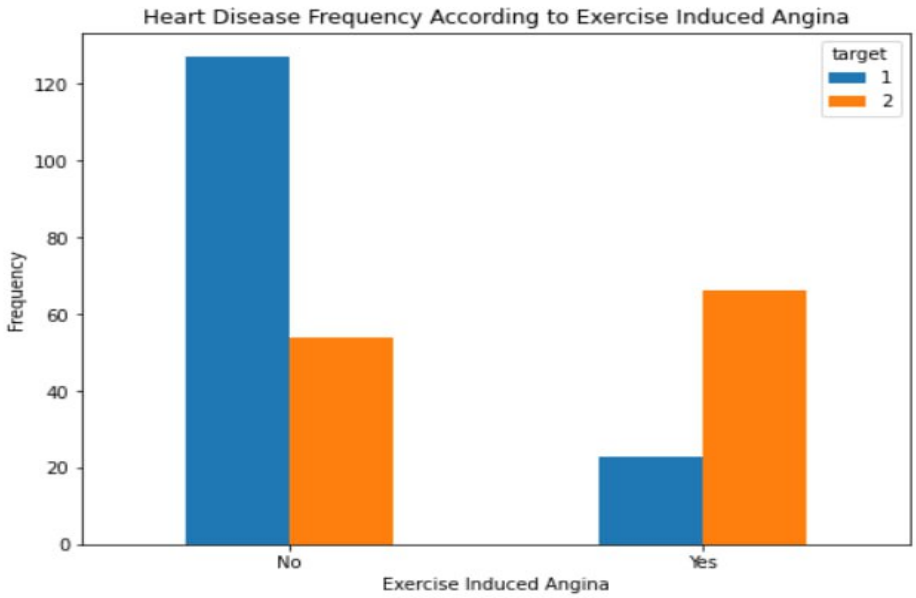

Figure 5: Exercise Induced Aches

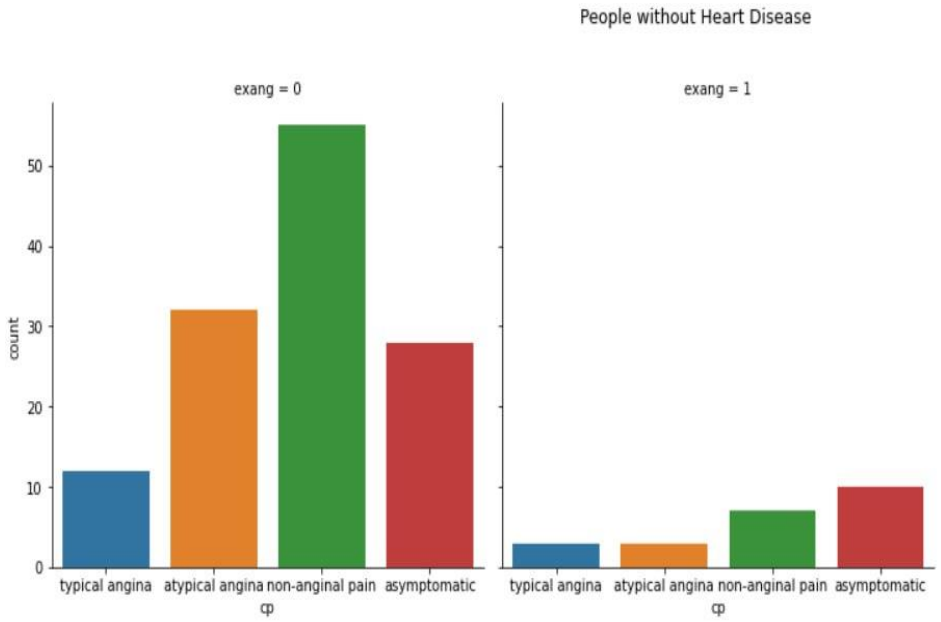

Figure 6: People with and without heart diseases

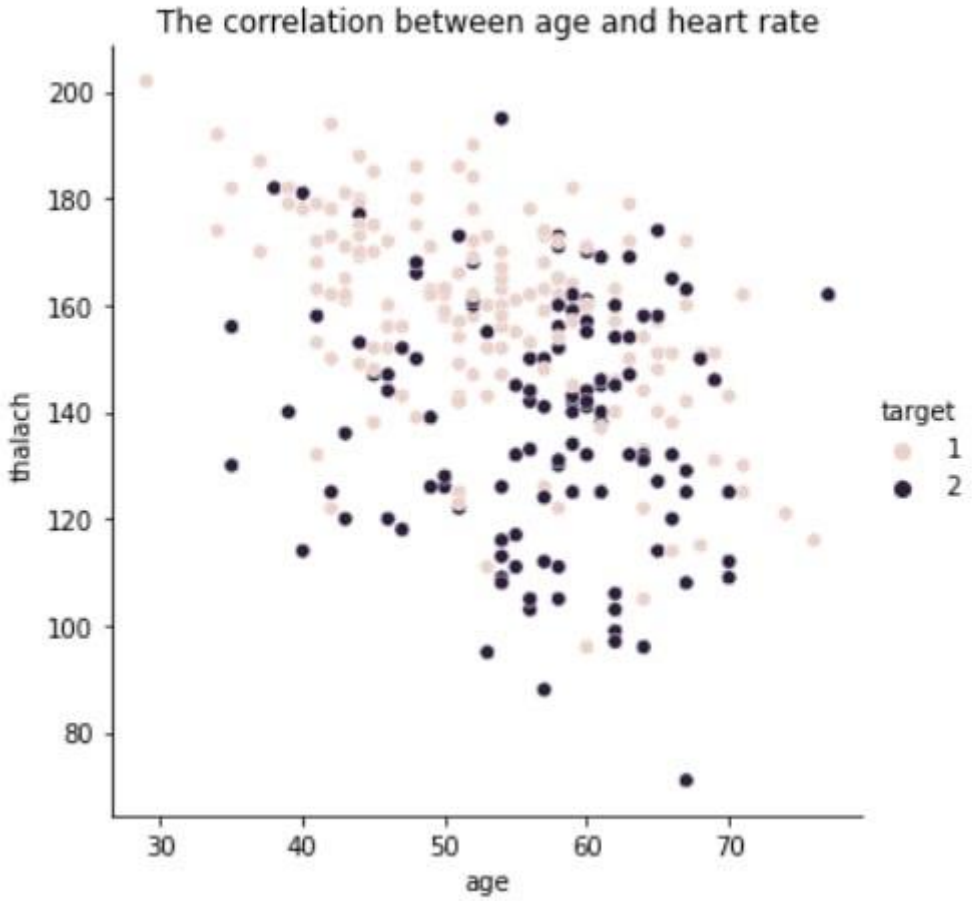

Figure 7: Correlation

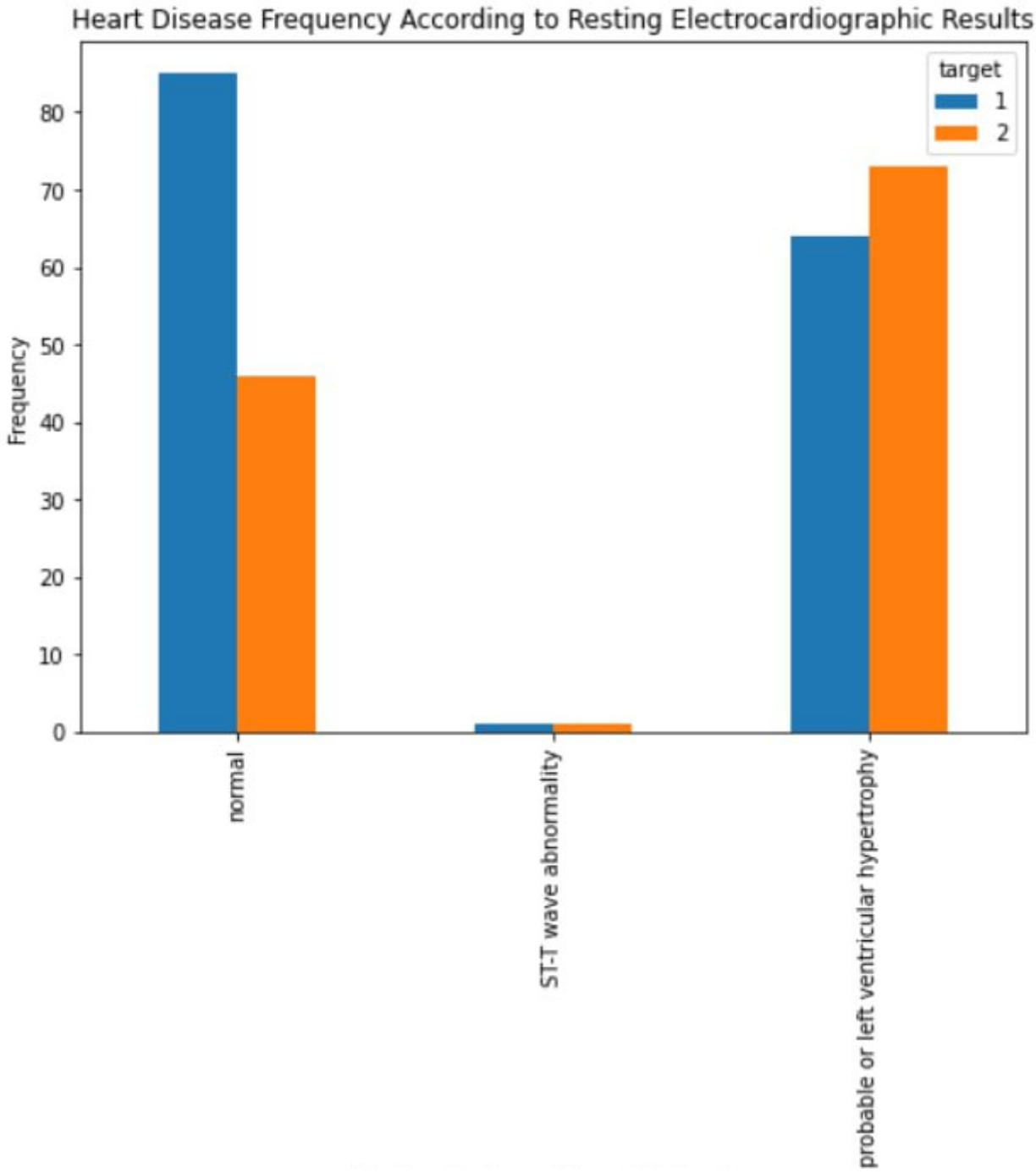

Fig 8: Electrocardiographic Results

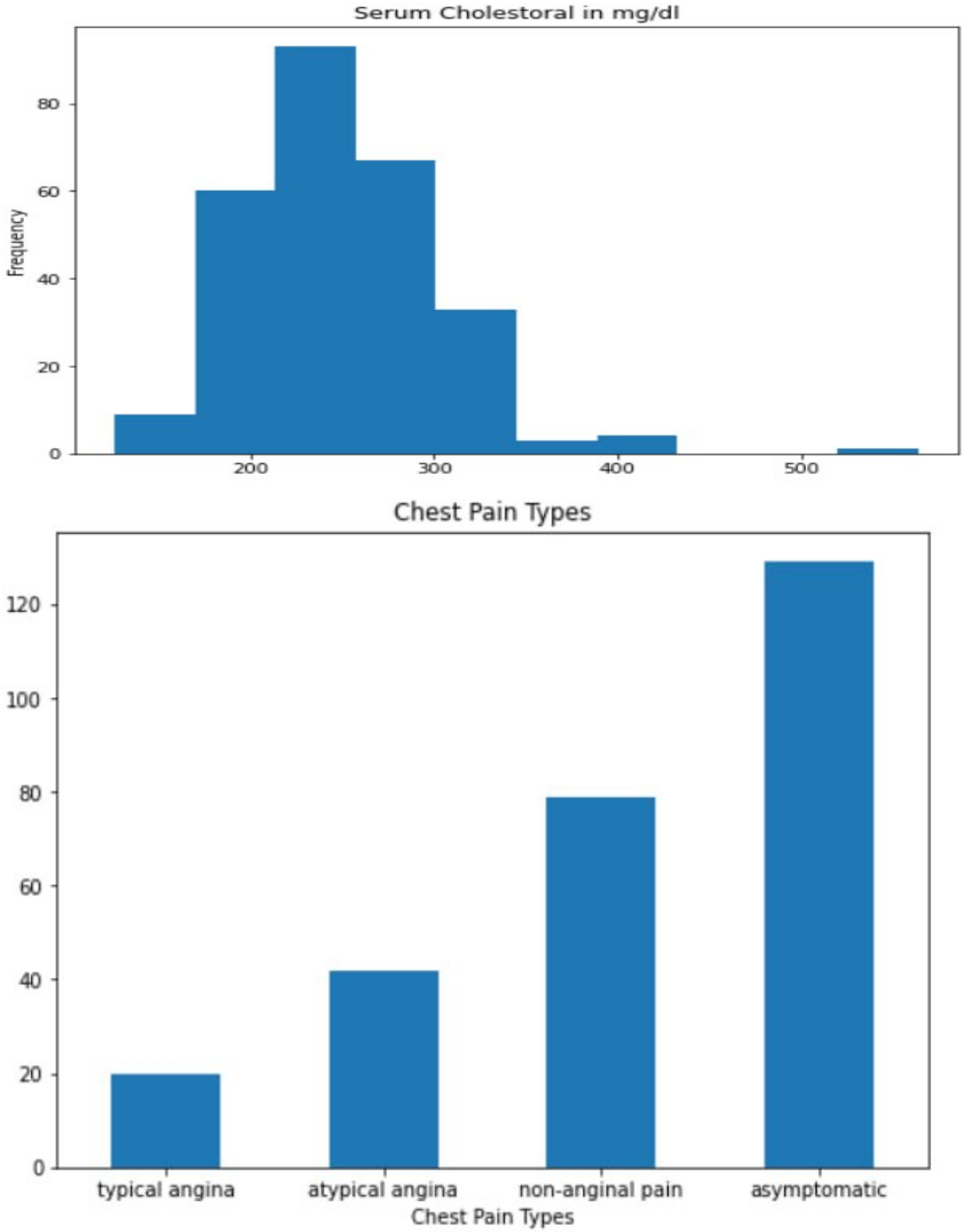

Fig 10: Chest pain types