**Group 12**

**Heart disease Prediction & ECG image classification of Cardiac Patients**

**Group Members**Abhay Krishnan, Aghil Kizhukkulathil, Akul Adichikkatt, Edwin Binu,Muhammed Hilal

**Overview**

Heart disease is a leading cause of death globally. Understanding predictive factors through data analysis can help in early diagnosis, improving preventive measures and treatment strategies, ultimately saving lives.

The coursework is divided into two parts

1. Predicting heart disease using machine learning techniques. (Tabular Dataset)
2. ECG image classification using Neural Network (Image Dataset)

**Objectives**

1. The Aim is to build predictive models using structured health data using various classifiers like Logistic Regression, Decision Trees, Random Forest… to analyze and predict heart disease using attributes such as cholesterol levels, chest pain types, and blood pressure
2. The is Aim is to applying deep learning techniques, such as convolutional neural networks (CNNs), these visual patterns can be automatically analyzed to classify different types of heart abnormalities, such as arrhythmias, ischemia, or myocardial infarction.

* ECG (Electrocardiogram) images contain distinct patterns that reflect various heart conditions.

**Data Collection**

*The source of Dataset 1 and Dataset 2 used for the analysis is from [Kaggle](*[*https://www.kaggle.com/*](https://www.kaggle.com/)*)*

Source Link and License Link

Dataset 1 - Predicting Heart Disease

Link of the Dataset *https://www.kaggle.com/datasets/mexwell/heart-disease-dataset*

Link of License [*https://creativecommons.org/licenses/by/4.0/*](https://creativecommons.org/licenses/by/4.0/)

Dataset 2 - ECG Image Classification

Link of the dataset *https://www.kaggle.com/datasets/evilspirit05/ecg-analysis*

Link of License [*https://www.mit.edu/~amini/LICENSE.md*](https://www.mit.edu/~amini/LICENSE.md)

**Dataset Description and Analysis**

Dataset 1

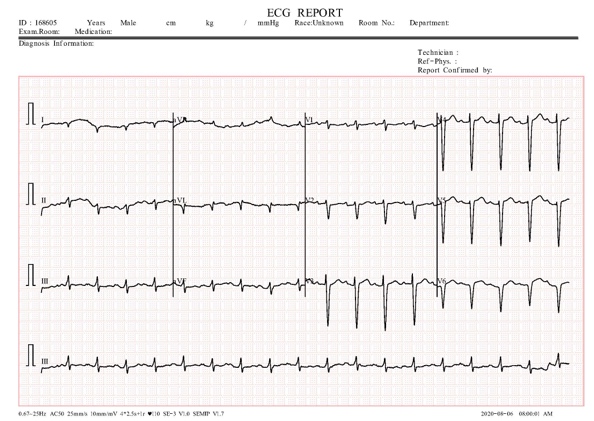
This image shows the overview of the tabular data, which showcases the attributes that we are using for the model: Age, Sex, Chest Pain Type, Resting Blood Pressure, Serum Cholesterol, Fasting Blood Sugar, Resting Electrocardiogram Results, Maximum Heart Rate Achieved, Exercise Induced Angina, Oldpeak (ST Depression), The Slope of Peak Exercise ST Segment, Class (Target)​A screenshot of a black screen

Description automatically generated

Dataset 2

This image represents the ECG of a MI patient and patient that have abnormal heartbeat

A graph of ecg

Description automatically generated

**Methodology**

**Dataset 1**

The data is loaded using pandas library, and it is proceeded with initial inspection by describing the dataset and visualization of correlation(using seaborn)

Then the data is preprocessed to check the missing values

The data is check for outliers using box plot

The data is feature scaled using Standard scalers as there are outliers found

The data is then split into training set and testing set

Decision tree parameter estimation is used to try out Different splits and Depths to find the best parameter using GridSearchCV Method Grid Search Cross-Validation is method that uses cross validation and Grid Search method to find the best hyper-parameter

* The best Parameter for Max Depth is \_ \_ \_ and Min sample split is \_ \_ \_

A diagram of a tree

Description automatically generated

The model is then fed a loop to predict heart disease of a patients with relevant attributes using different machine learning algorith Such as: DecisionTree, RandomForest, Logistic Regression, KNN, GradientBoosting, NaiveBayes, SVM and compare the accuracy between the model for better prediction.

To improve the accuracy added a K-fold algorithm with split 5 and ran the loop with it

KMeans clustering is ran through the data. To find the optimal K, has used Elbow Method. The Elbow method uses within-cluster-sum-of-square (WCSS) vs K value graph. The optimal K value is at the point where the graph forms an elbow

**Project 2**

The required libraries including pandas, os, tensorflow are imported

The dataset which is already split in the folder is imported and classified accordingly using os library

The data then is run through a traintest generation model, where the data is augumented and converted into a greyscale image

A CNN model is created using the tensorflow library to analyse the data

The model is trained for 25 epochs ensuring it learns effectively from the data

**Result**

The accuracy of the model is listed in the table below

A screenshot of a graph

Description automatically generated

The best model with the best accuracy is Random Forest with a Accuracy of 0.93

After performing K - Fold algorithm on the model there was a significant improvement in the accuracy, further validating the model

A black background with white numbers

Description automatically generated

After applying the K fold algorithm, the best model with the highest accuracy is Random Forest with a Accuracy of 0.95

K Means Clustering gave an output of

A black background with white text

Description automatically generated

Project 2

The model ran with a accuracy of \_ \_ \_