International Journal of Research in Engineering, Science and Management
Volume 5, Issue 8, August 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Prediction of Coronary Artery Disease using an Ensemble Model

P. T. Parikshith Belliappa[1*], Raksha Anand[2], Rinki Kiran[3], S. Sandeep[3], J. Vimala Devi[5]

[1,2,3,4]*Student, Department of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*

[5]*Associate Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*

***Abstract*: The field of medical analysis follows an exponential curve in the area of scientific knowledge, yet things are not very optimistic and assured from the perspective of patients as many diseases are discovered quite late by doctors. The lack of comprehensive data on relevant risk factors for coronary artery disease (CAD) has limited the ability to predict the risk of developing CAD in large populations. The challenge lies in the data complexity and correlations in terms of prediction using conventional techniques. Coronary artery disease is predicted by uploading the electrocardiogram (ECG) reports of patients into the machine learning model. Multiple classifiers like k-nearest neighbors, Logistic Regression, Support Vector Machine, and Voting Based Ensemble Classifier are implemented, and based on the acceptable criteria on the accuracy, the model will be finalized. This model will help medical professionals in diagnosis of cardiac diseases, to detect whether a patient has/had Myocardial Infarction, Abnormal Heartbeat, or the patient is hale and healthy by inferring the ECG reports.**

***Keywords*: Image segmentation, Gaussian filtering, Otsu thresholding, KNN, Logistic regression, SVM, Ensemble classifier.**

1. Introduction

Heart is the primary organ of the human body which pumps blood through the network of arteries and veins which comprises the cardiovascular system. The coronary arteries are present along the surface of the heart providing oxygen-rich blood to the heart muscle. Sometimes deposits of cholesterol and other substances on the artery wall termed as plaque, can narrow these arteries which eventually reduces the blood flow. This causes coronary artery disease or otherwise known as ischemic heart disease. It is one of the most common types of heart disease.

Heart disease is found to be one of the biggest causes of morbidity and mortality among the population of the world. It has remained the leading cause of death at the global level for the last 20 years. Every year,12 million deaths occur worldwide due to heart disease. World Health Organization has estimated that mortality rate caused by heart diseases will mount to 23 million cases by 2030 (estimates one out of four deaths is due to CAD). While Coronary Artery Disease is a significant cause of death and disability, it is preventable.

Machine learning has the potential to disrupt the medical industry by opening up new ways to handle healthcare data, transforming patient care, and streamlining administrative processes. Huge medical records, which previously required a human intervention, can now be used as input data for machine learning in medical diagnosis.

At present, there are plenty of algorithms available that could detect and predict heart anomalies from clinical reports based on several criteria of the patients such as age, gender, blood pressure, medical history, etc. This paper aims to analyze several machine learning techniques implemented in recent years for diagnosing heart disease. The aim is to implement the various supervised learning algorithms like K-Nearest Neighbor (KNN), Logistic Regression, Support Vector Machine, Gaussian Naive Bayes, Random Forest and XGBoost in the prediction model using the Electrocardiogram (ECG) images of the patient and determine the presence of coronary artery disease. The Electrocardiogram (ECG) images undergo data pre-processing, feature extraction which generates 1-D signals. These signals are given as inputs to the trained machine learning model. A comparative analysis of the different machine learning models is performed by the voting ensemble classifier and the best classifier is chosen for the final prediction.

The datasets have been obtained from Mendeley Data Repository. It is a research data repository which allows researchers to make their research data publicly available. The dataset contains the images of ECG signals of healthy individuals, patients diagnosed with myocardial infarction and coronary artery disease (abnormal heartbeat). This dataset is employed in the training and testing of the model involved in prediction of coronary artery disease.

The related work is précised under the Section 2. Section 3 encapsulates the methodology and architecture of the proposed model. The results thus obtained are mentioned in the Section 4. The conclusions driven from the project is specified in Section 5.

2. Related Work

Sushmita et al. [1] demonstrated a machine learning algorithm which analyses ECG data and heart disease prediction. This method designed a model that made use of supervised machine learning to find anomalies in an ECG report

*Corresponding author: parikshith.18cs098@cambridge.edu.in

and by further analysis, it predicted the chances of a patient suffering from a certain heart disease. There were six supervised machine learning algorithms such as Logistic regression, Decision tree, Nearest neighbor, Naïve Bayes, Support Vector Machine, Artificial Neural Network, were used to distinguish between normal and abnormal ECG. The dataset was divided into two parts. In order to train the model, 75% of the data was put into one group and the rest 25% data in another group for testing. Cross Validation and Random Train Test Split was used in order to avoid any kind of anomalies or repetitions which gave accurate results. The results were then compared with each other for a better understanding. Analysis showed that Logistic regression was the best algorithm to be used for Right Bundle Branch Block while Decision Tree gave the best result for Myocardial Infarction and for Sinus Bradycardia. All the algorithms except Nearest Neighbour could be used in case of Sinus Tachycardia. Lastly, Naive Bayes algorithm gave the best score for coronary artery disease.

Divya et al. [2] proposed a method that predicted the risk of coronary heart disease (CHD) using machine learning algorithms like Random Forest, Decision Trees, and K-Nearest Neighbors. The work included a pre-processing extensive approach which was practiced on a subset of Framingham Heart Study (FHS) dataset. Resampling, normalization, classification and prediction were the steps that were included in this approach. The analysis of the various machine learning algorithms estimates its use in prediction by developing various classification models.10-Fold cross-validation was performed to generate randomness. The performance measures thus evaluated showed that Random Forest obtained an accuracy of 96.8% with an execution time of 1.3969 seconds. Decision Tree had the least execution time of 0.81 seconds but gave lesser performance than random forest, whereas the execution time of KNN was as high as 1.9 seconds with similar performance. Thus, Random Forest obtained a robust prediction model which aimed at providing a valuable assistant in the health management sector.

M. Kavitha et al. [3] demonstrated a hybrid machine learning model which predicts heart disease. This method designed a hybrid model consisting of Decision Tree and Random Forest to predict the heart disease. The combined model was developed and implemented based on the probabilities of the Random Forest. Cleveland database was considered with 70% datasets as training datasets and 30% as testing dataset. The result analysis showed that heart disease detection was effective when the Random Forest algorithm was employed along with a hybrid model. Decision Tree achieved an accuracy around 79% while Random Forest gave 81% accuracy. The Hybrid model displayed an accuracy of 88%.

Lee et al. [4] performed a study which aims at building a prediction model to detect the occurrence of coronary artery calcification. It is one of the risk factors related to coronary artery disease. Ensemble-based methods were employed to build the model in order to handle imbalanced dataset. The dataset used was a regular medical check-up dataset from Samsung Medical Center. The study classified people into the high-risk group and normal group depending on the follow-up checks of the coronary artery calcification. The features underwent preprocessing together with feature selection by a wrapper-based approach. Ensemble Classification algorithms included Decision Tree, Logit Boost, MultiBoostAB and Bagging. The ensemble strategies followed were to selection of the best classifier with highest accuracy, top-k-based method to extract top-k-features frequently used and finally a voting-based method to select the most accurate classifier by voting. The analysis showed that LogitBoost was the best classification algorithm as it displayed the best performance with an accuracy of 74.07%. The voting-based model showed good accuracy whereas the top-k-based model had greater accuracy then classifier-based model. Thus, the usage of ensemble-based classification method provided appropriate and better accuracy which was about 5% higher than other imbalanced classification method. Even without laborious pre-processing and parameter tuning, a cost-sensitive approach to make precise classifiers was developed.

Zhang et al. [5] proposed this paper to develop and build a machine learning system to diagnose Coronary Artery Disease at an early stage. This research encompasses the usage of two ensemble learning algorithms named XGBoost and Random Forest to construct the classification model. The proposed method involved various feature processing technologies. These included the following: feature smoothing, encoding, construction and feature selection. The performance of the models like the accuracy and precision improved along with the usage of these processing technologies. Further processing on the Z-Alizadeh Sani dataset was done by employing the algorithms like Synthetic minority oversampling technology (SMOTE) and adaptive synthetic (ADASYN) Algorithm. These overcame the problem of unbalancing. The stability of the model was tested and performed using the 10-fold cross validation. The experimental results showed that the best performance for prediction of coronary artery disease was obtained by the XGBoost model with an accuracy of 94.7%. This inferred that the model resulted in a stable and effective ability to identify CAD patients.

Amanda H Gonsalves et al. [6] proposed this paper to find the most effective Machine Learning (ML) model in predicating Coronary Heart Disease (CHD) by using historical medical data. The research paper made use of three supervised learning techniques namely Naive Bayes, Support Vector Machine and Decision Tree. These helped in discovering correlations in data thus contributing to improving the prediction rate. The dataset used was the South African Heart Disease dataset consisting of 462 instances. The dataset was initially pre-processed to eliminate noise. Later, it was used to train the predictive models using the Machine Learning techniques like SVM, Naive Bayes and Decision tree. The performance of the classification models was measured using the confusion matrix. The result analysis depicted that the Naive Bayes outperformed Decision Tree and Support Vector Machine with highest accuracy. The prediction model obtained was used to develop a mobile application which helped people track their health leading to early detection of CHD. The limitations of the paper included not having enough instances and the CHD data being unbalanced.

V. V. Ramalingam et al. [7] presented a survey paper which predicted heart disease using various machine learning techniques and also analyzed their performance. An ensemble model was created with the following supervised learning algorithms such as Random Forest, Naïve Bayes, K-Nearest Neighbor, Decision Tree, Support Vector Machines. On the research conducted it could be known that there is a very large scope for machine learning models in predicting any heart related diseases or coronary artery disease. Every algorithm had given extremely well accuracy in some cases and poor accuracy in some other cases. Decision tree model worked fine except cases which includes overfitting whereas random forest had solved the problem of overfitting by employing multiple algorithms. Naïve Bayes model's classifier were exceptionally fast and have performed well. SVM had performed exceptionally well. Systems based on machine learning algorithms have performed exceptionally very accurate in predicting the heart diseases. Still a lot of research could be done on how to handle high dimensional data and overfitting.

3. Methodology

Figure 1 illustrates the working of the proposed model. Once the ECG images are uploaded into the model, they are transformed into the required format which are employed by the machine learning algorithms to analyze and predict the outcome.

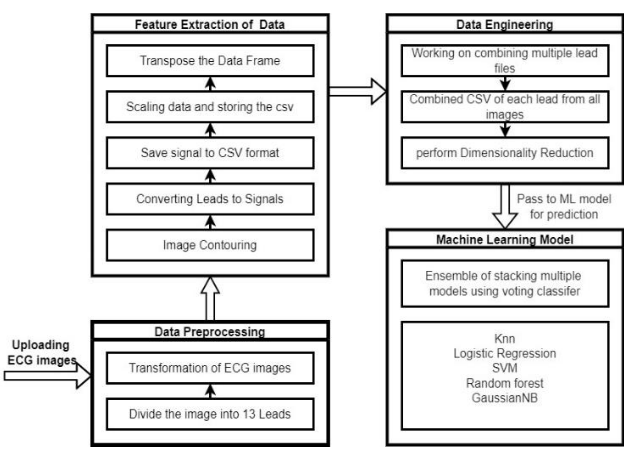

Fig. 1. System architecture

The flow of the process is carried out in following segments:

A. Data Preprocessing

The initial step involves transforming the ECG images into leads and perform pre-processing on these leads by removing grid lines and other noise present in the images. This has been achieved by making use of certain methods such as rgb2gray, gaussian filtering and Otsu thresholding. The rgb2gray method helps in converting the image to gray scale as the conversion to a single channel grayscale image is important for thresholding. It has been followed by the process of Otsu thresholding which performs automatic image thresholding in the simplest form and the algorithm returns a single intensity threshold that separates pixels into two classes, foreground and background. Gaussian filter is a low pass filter which has been used for reducing noise and blurring regions of an image. The algorithm can now detect the borders of the objects from these pixels.

The data processing is as follows:

1) An ECG image has been uploaded into the model

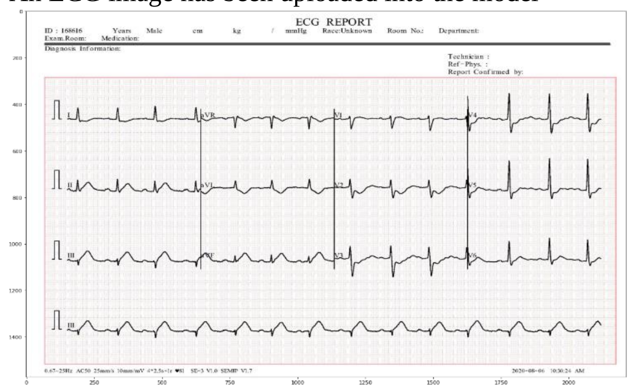

Fig. 2. ECG image of a patient

2) The ECG image has been split into 13 Leads

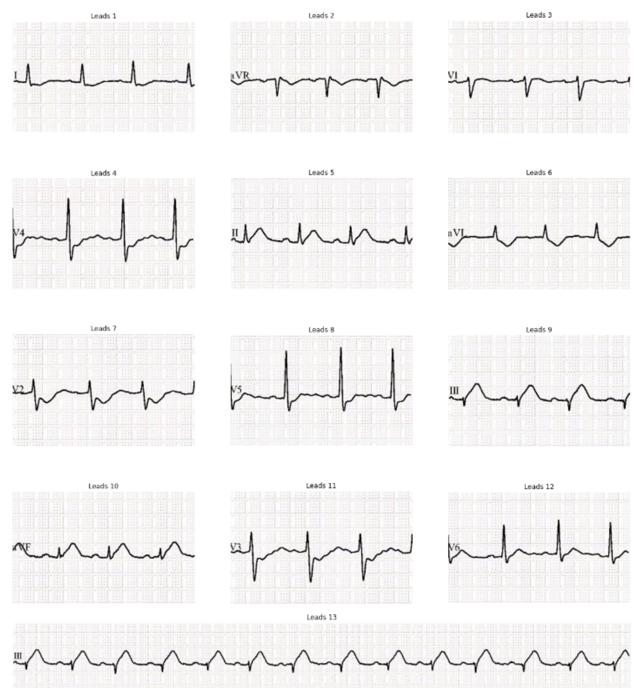

Fig. 3. 13 Leads extracted from the ECG image

3) The required twelve leads have been transformed into grayscale images. Otsu thresholding has been performed to separate the foreground and the background. Gaussian filtering has been used for smoothening the images.

B. Feature Extraction

The resultant image obtained after pre-processing has been further subjected to contouring in order to identify the structural outline of objects present in that image which in turn helps in identifying the shape and flow of the objects.

The necessary signals are then extracted from the images in the form of values. These values are stored in a .csv file which have been further scaled using MinMaxScaler without changing the shape of the original distribution. This obtains a normalized signal within a specific range from which the values are considered for representing the peaks and troughs of the ECG waves.

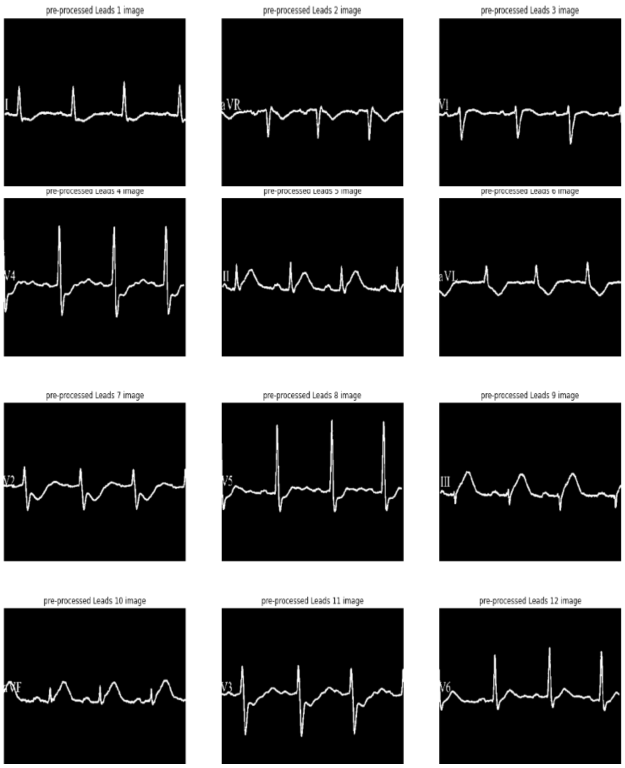

Fig. 4. Transformed leads

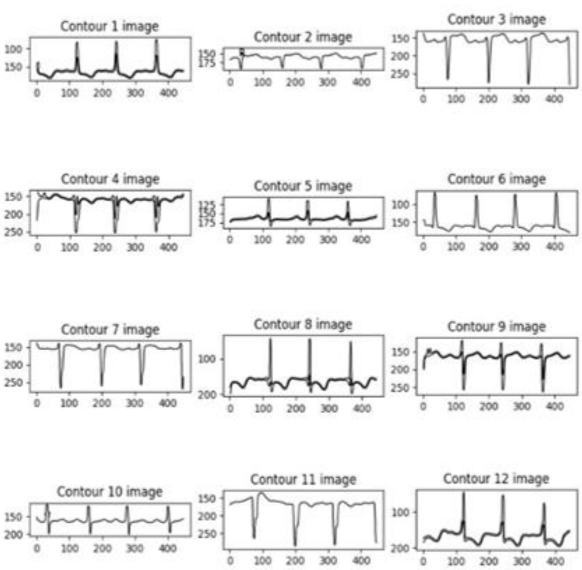

Fig. 5. Contour leads

	0	1	2	3	4	5	6	7	8	9	...	245	246	247	248	249	250	251	252	253
X	0.745093	0.744958	0.752378	0.764513	0.757202	0.754713	0.765575	0.770066	0.75244	0.732998	...	0.769579	0.781019	0.775996	0.759667	0.744735	0.73452	0.740976	0.749191	0.753299

1 rows × 255 columns

Fig. 6. 1D signals of the ECG images

Transposing these signals into one dimensional signal can be used for further analysis. These one-dimensional signals are used by the machine learning model to predict the condition of the patient.

C. Data Engineering

In this step we have both csv files and the cropped lead (1-13) to work on. In order to find out the variations in the data, the total variance and variance among the data components have been identified. These variances are sorted using the natsort function to obtain their natural order which makes it easier to understand the variations in the data. Few supervised classification algorithms such as K-Nearest Neighbor, Logical Regression, Support Vector Machine, etc., have been employed in the model to process the data stored in the CSV file and after processing and analyzing the data. The data in the csv files undergo Dimensionality reduction is performed on the data which further increases the accuracy rate of the model.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	-1.678782	-0.412932	0.684989	-1.306078	0.270179	0.585918	-0.422890	0.246461	0.044889	0.353991	0.927250	-0.620871	-0.090258	-0.775913	0.231839	-0.618439	0.038086
1	-1.582305	0.519510	-1.246505	1.651600	-0.130831	0.390073	-0.444086	0.292156	-0.459205	-0.087866	1.383836	-0.473842	0.420504	-0.256876	0.365666	0.320517	0.387268
2	-1.312178	0.180662	1.240179	-1.867477	-0.230707	0.241249	0.066655	-0.377587	-0.309773	0.543821	0.935326	0.619717	1.338468	-0.499635	0.559353	-0.378007	0.324931
3	-1.113164	-1.741394	1.115683	0.202218	-0.244324	0.613945	0.247207	0.067990	0.225652	-0.272218	-1.262435	-0.807793	0.519322	-0.414492	0.060801	0.238216	0.255922
4	-2.464051	-2.282357	1.125929	0.978918	-0.233772	-0.056228	0.424466	0.205385	0.030233	0.291353	-1.129254	-1.016363	0.380888	0.253396	-0.496416	0.578266	0.268729
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
932	-1.267053	-1.429970	-2.282429	-0.591794	0.935615	0.093106	0.466886	0.540627	0.061973	-0.007981	-0.183685	-0.161472	0.331055	-0.443831	-0.298484	-0.997683	-0.170435
933	-0.914378	-0.302846	-0.269675	-0.218922	0.502654	-1.050568	-0.519431	1.524155	0.342077	0.389508	-0.708046	0.316125	0.259628	-0.004201	-0.021326	0.549809	-0.173944
934	-0.203484	-0.385792	0.330116	0.911653	-0.321411	0.038475	-0.213118	-0.501001	-0.175490	0.800086	0.024483	0.092339	-0.251265	0.494278	-0.184421	0.237742	0.233529
935	1.416122	-0.646671	0.244334	0.161198	0.113901	0.917797	0.391670	1.114875	-0.144792	-0.168619	-0.005679	-0.150314	0.898686	-0.979840	0.441496	0.002612	-0.453228
936	-1.961520	2.263005	0.408932	-0.790544	0.571374	-0.491889	0.371044	0.200214	0.470579	-0.190302	-0.922355	0.083533	-0.237807	0.912569	-0.361271	0.316327	-0.495830

937 rows × 101 columns

Fig. 7. Dimensionality reduction

This data is processed and analyzed by the model by applying different machine learning techniques without any target to check predictions and using this on Realtime for the test data.

D. Machine Learning Algorithms

The one-dimensional values obtained by transforming the ECG images of patients are given as input into the algorithms used in the model. The algorithms used are:

K Nearest Neighbor Algorithm:

- Step 1: Load the important libraries
- Step 2: Import dataset and extract the dataset into train and test
- Step 3: Select the number K of the neighbors
- Step 4: Calculate Euclidean distance of K number of neighbors
- Step 5: Take the K nearest neighbors as per the calculated Euclidean distance
- Step 6: Among these k neighbors, count the number of the data points in each category
- Step 7: Assign the new data points to that category for which the number of the neighbor is maximum
- Step 8: Coming up with predictions
- Step 9: Evaluating model's performance

K Nearest Neighbor model has been used for making predictions based on the ECG images imported. The required libraries are imported and loaded into the model. The one-dimensional values are then fed to the model and are extracted as training data which contains attributes and labels, and test data which only contains attributes. A constant K is selected which is represented as the number of nearest neighbors. The Euclidean distance is calculated using the formula mentioned below and the K nearest neighbors to the test datapoints are chosen. Based on the class with majority neighbors, the model classifies the data and predicts the outcome by identifying the class the data best fits to. The K Nearest Neighbor model is evaluated based on its performance.

$$ED(x,y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

Support Vector Machine Algorithm:

- Step 1: Load the important libraries
- Step 2: Import dataset and extract the X variables and Y separately.
- Step 3: Divide the dataset into train and test
- Step 4: Initializing the SVM classifier model
- Step 5: Fitting the SVM classifier model
- Step 6: Coming up with predictions
- Step 7: Evaluating model's performance

Support vector machine is well suited for image classification. The required libraries have been imported and loaded into the model. The one-dimensional values are then fed to the model which contains attributes and labels. This is the training data which has been used for training the SVM model. The test data is mapped within a specific range of values to produce accurate outcomes. The test data is classified and the prediction is made. The SVM model is evaluated based on its performance.

Logistic Regression Algorithm:

- Step 1: Load the important libraries
- Step 2: Import dataset and extract.
- Step 3: Divide the dataset into target and feature variables.
- Step 4: Split the dataset into train set and test set.
- Step 5: The model is trained by passing the train features and target class values and it implements classification.
- Step 6: Evaluating the performance of the model.

Logistic regression classifier falls under supervised learning that models the probability of a dependent variable. It is mainly employed for classification problems and is easy to implement. The libraries have been imported and the data is loaded and preprocessed. It is then split into target variable and feature variables. The best parameters are identified by using the Grid Search CV function from the given parameter grid. The dataset is then split into test and train data.

The fit method takes the data values(X) and the label (y) as arguments to fit the model accurately. The prediction is done on the test data using the predict method. The model is thus evaluated returning the accuracy and the classification report is displayed.

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

Random Forest Algorithm:

- Step 1: Random records or samples are selected from the dataset.
- Step 2: The algorithm builds a decision tree based on these records.
- Step 3: Choose the number of trees required in the algorithm and repeat steps 1 and 2.
- Step 4: For a new record, each tree in the forest predicts a value or category to which the record belongs.
- Step 5: Finally, the new record is assigned to the category that wins the majority vote.

Random forest is a type of supervised machine learning algorithm based on ensemble learning. It is used for both classification and regression problems. A forest is made up of trees and more number of trees results in a more robust forest. The algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. Since each tree does not consider all the features, the feature space is reduced. Random forest works on the Bagging principle.

$$RFfi_i = \frac{\sum_{j \in all\, trees} normfi_{ij}}{T}$$

Gaussian Naïve Bayes Algorithm:

- Step 1: Separate the training data by class.
- Step 2: Summarize the dataset by calculating the mean and standard deviation.
- Step 3: The dataset is separated into rows by class and statistics are calculated on each subset.
- Step 4: Calculate the Gaussian Probability Density Function.
- Step 5: Calculate Class Probabilities using the Gaussian probability density function and the statistics followed by multiplying the obtained probabilities.

A Gaussian Naive Bayes algorithm is a special type of NB algorithm. It is specifically used when the features have continuous values. It is used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset.

$$P(X|Y=c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{\frac{-(x-\mu_c)^2}{2\sigma_c^2}}$$

In order to improve the model predictions, model stacking has been done which combines the outputs of multiple models and running them through meta-learner which is another machine learning model. A voting-based classifier trains the ensemble of various models and predicts the outcome (class) depending on the class with the highest probability. The front-end makes use of an open-source application framework in Python language known as Streamlit which is used to build instant web applications. With the help of this framework, we designed a front-end application which allows the user to upload the ECG images, pass this image to the trained model and obtain results. The results are then displayed to the user which can be beneficial for further diagnosis.

4. Experimental Result

The Use case diagram is shown below, it shows the interaction between the web application and the users. It describes the functions and scope of a system. When the user uploads an ECG image, the image will undergo the transformation to finally obtain the one-dimensional values generated from the leads and the output of each process is displayed to the user before displaying the final prediction.

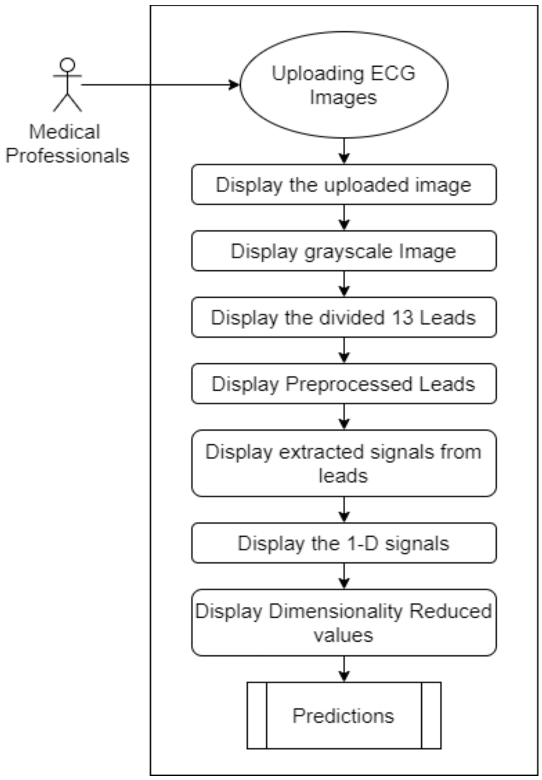

Fig. 5. Use case diagram

This paper has demonstrated the sequential steps of processing and transforming the ECG images into one dimensional signal through methods such as image segmentation where the ECG images have been divided into 13 leads followed by the smoothing process which uses a Gaussian filter to reduce the noise from the images. Otsu Thresholding method has been carried out to differentiate between the foreground and background (pixels) which is essential in the contouring and scaling process to generate two-dimensional signals. These are then normalized to one-dimensional signals. The algorithms in our machine learning model have been employed to process the one-dimensional signals thus generated and have returned the predictions along with the accuracy, f1 shape, recall and support. The ensemble classifier also known as voting classifier chooses an algorithm using the acceptable criteria and thus the final predictions will be made by the developed model. The final predictions unfolds whether the ECG image uploaded characterizes normal heartbeat, abnormal heartbeat or myocardial infarction.

The following results have been obtained by the machine learning algorithms for four classes:

The results obtained for the Abnormal Heartbeat class has been shown in the table below for each machine learning model implemented.

Table 1
Results for dataset with Abnormal Heartbeat

Metrics	KNN	LR	SVM	Ensemble Classifier
Accuracy	0.83	0.78	0.87	0.92
Precision	0.92	0.85	0.77	0.92
Recall	0.65	0.56	0.92	0.91
F1-score	0.76	0.69	0.84	0.92
Support	105	105	119	80

The results obtained for the Myocardial Infarction class has been shown in the table below for each machine learning model implemented.

Table 2
Results for dataset with Myocardial Infarction

Metrics	KNN	LR	SVM	Ensemble Classifier
Accuracy	0.83	0.78	0.87	0.92
Precision	0.95	0.88	1.00	1.00
Recall	1.00	1.00	1.00	1.00
F1-score	0.97	0.94	1.00	1.00
Support	94	94	125	72

The results obtained for the Normal class has been shown in the table below for each machine learning model implemented.

Table 3
Results for dataset with Normal Heartbeat

Metrics	KNN	LR	SVM	Ensemble Classifier
Accuracy	0.83	0.78	0.87	0.92
Precision	0.74	0.75	0.84	0.86
Recall	0.93	0.86	0.84	0.92
F1-score	0.82	0.80	0.84	0.89
Support	112	112	140	79

The results obtained for the History of Myocardial Infarction class has been shown in the table below for each machine learning model implemented.

Table 4
Results for dataset with History of Myocardial Infarction

Metrics	KNN	LR	SVM	Ensemble Classifier
Accuracy	0.83	0.78	0.87	0.92
Precision	0.74	0.63	0.93	0.93
Recall	0.70	0.67	0.68	0.83
F1-score	0.72	0.65	0.78	0.88
Support	61	61	80	48

This renders in the early prediction and diagnosis of the coronary artery disease thus signifying to be of huge benefit. It has been found to be accurate and a lot more efficient compared to the traditional methods. The frontend implementation is efficient and user-friendly which poses as an added advantage in deciphering the required details and interpretations of the result.

5. Conclusion

Coronary artery disease prediction can be done by applying various algorithms using the patient reports. However, in this project, the focus was more on identifying and extracting patterns from Electrocardiogram image reports. By digitizing ECG records, the need for time-consuming manual intervention to comprehend the report can be eliminated. With digitization, the automation of diagnosis and analysis can be achieved more quickly. All the algorithms used in the model give out results with varied accuracy and their performance measures specify the most suitable classifier that can be utilized for prediction. The concept of automation gives an upper-hand in speeding up the prediction process at the users' convenience leading to timely assistance and diagnosis of the disease.

In the project, it has been observed that the accuracy of the prediction model is increased to 0.92 using the ensemble classifier by stacking up varies machine learning algorithms such as K Nearest Neighbors, Logistic Regression, Support Vector Machine, Gaussian Naïve Bayes and Random Forest algorithms when compared to the machine learning algorithms such as K nearest Neighbors (accuracy – 0.83), Logistic Regression (accuracy – 0.78) and Support Vector Machine (accuracy – 0.87). The classification metrics for the four class groups; Abnormal Heartbeat, Myocardial Infarction, Normal Heartbeat and History of Myocardial Infarction are obtained by the model for the Machine Learning algorithms using which the given data is grouped to the class which fits best and hence the prediction is made.

References

[1] Sushmita Roy Tithi, Afifa Aktar, Fahimul Aleem, Amitabha Chakrabarty, "ECG Data Analysis and Heart Disease Prediction using Machine Learning Algorithms".

[2] Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik, "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms".

[3] M. Kavitha, G. Gnaneswar1, R. Dinesh, Y. Rohith Sai, R. Sai Suraj, "Heart Disease Prediction using Hybrid machine Learning Model".

[4] Heart Disease Facts - Center for Disease Control and Prevention, https://www.cdc.gov/heartdisease/facts.html

[5] Shasha Zhang, Yuyu Yuan, Zhonghua Yao, Xinyan Wang, Zhen Lei, "Improvement of the Performance of Models for Predicting Coronary Artery Disease Based on XGBoost Algorithm and Feature Processing Technology".

[6] Amanda H. Gonsalves, Fadi Thabtah, Rami Mustafa A. Mohammad, Gurpreet Singh, "Prediction of Coronary Artery Disease using Machine Learning: An Experimental Analysis".

[7] V.V Ramalingam, Ayantan Dandapath, M Karthik Raja, "Heart Disease Prediction using Machine Learning techniques: a survey".

[8] Coronary Artery Disease – Center for Disease Control and Prevention, https://www.cdc.gov/heartdisease/coronary_ad.html

[9] Coronary artery disease – mayo clinic, https://www.mayoclinic.org/diseasesconditions/coronary-artery-disease/symptoms-causes/syc-20350613

[10] Jihyun Lee, Hye Jin Kam, Ha-Young Kim,Sanghyun Yoo, Kyoung-Gu Woo, Yoon-Ho Choi, Jeong Euy Park and Soo Jin Cho, "Prediction of 4-year risk for Coronary Artery Calcification using Ensemble-based Classification".

[11] Ayatollahi, H., Gholamhosseini, L. & Salehi, M. Predicting coronary artery disease: a comparison between two data mining algorithms. BMC Public Health 19, 448 (2019).