

# An Integration of Cardiovascular Event Data and Machine Learning Models for Cardiac Arrest Predictions

Krishna Prasad K.<sup>1</sup>, P. S. Aithal<sup>2</sup>, Navin N. Bappalige<sup>3</sup> and Soumya S.<sup>4</sup>

<sup>1</sup>Associate Professor & Post-Doctoral Research Fellow, College of Computer & Information Sciences, Srinivas University, Mangalore, Karnataka, India

OrcidID: 0000-0001-5282-9038; E-mail: [krishnaprasadkcci@srinivasuniversity.edu.in](mailto:krishnaprasadkcci@srinivasuniversity.edu.in)

<sup>2</sup>Vice Chancellor, Srinivas University, Mangalore – 575001, India.

OrcidID: 0000-0002-4691-8736; E-mail: [psaithal@gmail.com](mailto:psaithal@gmail.com)

<sup>3</sup>Associate Professor, Sahyadri College of Engineering & Management, Mangalore

OrcidID:0000-0003-1122-4897; E-mail: [nbappalige@gmail.com](mailto:nbappalige@gmail.com)

<sup>4</sup>Research Scholar, College of Computer & Information Sciences, Srinivas University, Mangalore, Karnataka, India and Faculty, Computer Science, Hira Women's College, Mangalore, Karnataka, India

OrcidID:0002-5431-1977; E-mail: [pkssoumya@gmail.com](mailto:pkssoumya@gmail.com)

**Area/Section:** Health Sciences.

**Type of the Paper:** Experimental Research.

**Type of Review:** Peer Reviewed as per [C|O|P|E](#) guidance.

**Indexed in:** OpenAIRE.

**DOI:** <http://doi.org/10.5281/zenodo.4679995>

**Google Scholar Citation:** [IJHSP](#)

## How to Cite this Paper:

Krishna Prasad K., Aithal P. S., Bappalige, Navin N., & Soumya S. (2021). An Integration of Cardiovascular Event Data and Machine Learning Models for Cardiac Arrest Predictions. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 5(1), 55-71. DOI: <http://doi.org/10.5281/zenodo.4679995>.

## International Journal of Health Sciences and Pharmacy (IJHSP)

A Refereed International Journal of Srinivas University, India.

© With Author.



This work is licensed under a [Creative Commons Attribution-Non-Commercial 4.0 International License](#) subject to proper citation to the publication source of the work.

**Disclaimer:** The scholarly papers as reviewed and published by the Srinivas Publications (S.P.), India are the views and opinions of their respective authors and are not the views or opinions of the SP. The SP disclaims of any harm or loss caused due to the published content to any party.

## An Integration of Cardiovascular Event Data and Machine Learning Models for Cardiac Arrest Predictions

Krishna Prasad K.<sup>1</sup>, P. S. Aithal<sup>2</sup>, Navin N. Bappalige<sup>3</sup> and Soumya S.<sup>4</sup>

<sup>1</sup>Associate Professor & Post-Doctoral Research Fellow, College of Computer & Information Sciences, Srinivas University, Mangalore, Karnataka, India

OrcidID: 0000-0001-5282-9038; E-mail: [krishnaprasadkcci@srinivasuniversity.edu.in](mailto:krishnaprasadkcci@srinivasuniversity.edu.in)

<sup>2</sup>Vice Chancellor, Srinivas University, Mangalore – 575001, India.

OrcidID: 0000-0002-4691-8736; E-mail: [psaithal@gmail.com](mailto:psaithal@gmail.com)

<sup>3</sup>Associate Professor, Sahyadri College of Engineering & Management, Mangalore

OrcidID:0000-0003-1122-4897; E-mail:[nbappalige@gmail.com](mailto:nbappalige@gmail.com)

<sup>4</sup>Research Scholar, College of Computer & Information Sciences, Srinivas University, Mangalore, Karnataka, India and Faculty, Computer Science, Hira Women's College, Mangalore, Karnataka, India

OrcidID:0002-5431-1977; E-mail: [pksumyya@gmail.com](mailto:pksumyya@gmail.com)

### ABSTRACT

**Purpose:** Predicting and then preventing cardiac arrest of a patient in ICU is the most challenging phase even for a most highly skilled professional. The data been collected in ICU for a patient are huge, and the selection of a portion of data for preventing cardiac arrest in a quantum of time is highly decisive, analysing and predicting that large data require an effective system. An effective integration of computer applications and cardiovascular data is necessary to predict the cardiovascular risks. A machine learning technique is the right choice in the advent of technology to manage patients with cardiac arrest.

**Methodology:** In this work we have collected and merged three data sets, Cleveland Dataset of US patients with total 303 records, Statlog Dataset of UK patients with 270 records, and Hungarian dataset of Hungary, Switzerland with 617 records. These data are the most comprehensive data set with a combination of all three data sets consisting of 11 common features with 1190 records.

**Findings/Results:** Feature extraction phase extracts 7 features, which contribute to the event. In addition, extracted features are used to train the selected machine learning classifier models, and results are obtained and obtained results are then evaluated using test data and final results are drawn. Extra Tree Classifier has the highest value of 0.957 for average area under the curve (AUC).

**Originality:** The originality of this combined Dataset analysis using machine learning classifier model results Extra Tree Classifier with highest value of 0.957 for average area under the curve (AUC).

**Paper Type:** Experimental Research

**Keywords:** Cardiac, Machine Learning, Random Forest, XBOOST, ROC AUC, ST Slope.

### 1. INTRODUCTION :

Cardiac arrest caused by coronary artery stiffness caused by the artery that carries oxygen rich blood to the heart is blocked by cholesterol plaques [1]. These cholesterol plaques are built up over the course of period, and eventually, when one of the cholesterol plaques ruptures there is a formation of blood clots that cause a blockage [2]. Because of this blockage, blood flow is cut off and oxygen-starved heart cells in the heart muscles start to die. When all the cells in the heart die, it causes cardiac arrest or a heart attack. A normal cardiac arrest indicates symptoms before they occur, symptoms such as tiredness, shoulder pain, nausea, crushing chest pain and sudden shortness of breath [3]. Alternatively, the silent cardiac arrest can be asymptomatic, and requires a prompt medical diagnosis. The asymptomatic cardiac arrest is mainly caused by diabetes and other similar disorders. Patients with diabetes and other similar disorders suffer from nerve damage, and such patients are always at risk of having silent cardiac arrest. The heart sends signals, which are known as cardiac biomarkers to other parts of the body by triggering symptoms by seeking help, while

the heart loses out oxygen flow. The nerve damage of patient prevents the cardiac biomarkers sent from the heart to other parts of the body and results in silent cardiac arrest. The machine learning classification techniques can be used to identify the possibilities of cardiac arrest in the patient. The process of categorizing a collection of data into groups is known as classification. Predicting the class of provided data points is the first step in the process. Classification techniques of machine learning are, Naïve Bays Classification, K-Nearest neighbor algorithms, Decision Trees, and Random Forest. Some features used as an input in classification techniques are Age, Sex, Chest Pain Type, resting BP, Cholesterol, Fasting Blood Sugar, resting ECG, MAX Heart Rate, Exercise Angina, Old Peak, and ST Slope. Using machine learning to predict cardiac arrest, initially the dataset must be loaded, then feature engineering must be performed, then feature selection, feature scaling, model selection, save the suitable model, and deploy the model [3].

## 2. RELATED WORK :

The authors investigated the high risk of cardiovascular disease and modified artificial plant optimization (MAPO) is applied on real life example i.e., cardiovascular disease, MAPO is a tool for measuring heart rate count using a fingertip video [3]. A determination of pulse wave velocity in an Outpatient can be used to assess the cardiovascular risks [4]. Machine learning and image fusion are used for predicting heart disease [5]. Wearable soft sensor transducer and patients interview results. Classifiers are compared with k-fold procedure. The cardiovascular status is divided into three categories: stable, non-critical unhealthy, and critical unhealthy [6]. Also, the noise robustness test is carried upon the data. A machine learning way of predicting patients with extreme dilated cardiomyopathy are more likely to experience cardiovascular events. (DCM) [7]. Prediction of risk in patients with diabetes using machine learning. A new approach for predicting Cardiovascular disorder has been developed in patients with type 2 diabetes. this model could identify the patients at high risks. The cardiovascular disease is predicted on the triage levels of the patient, and the prediction is done only at the time of triage. Atherosclerotic plaque tissues in RA (Rheumatoid Arthritis) are used for predicting cardiovascular disease in machine learning [8]. The possibility of Cardiac Arrest is predicted using machine learning classification methods. This model is specifically designed and trained with ICU data. The review summary of the related work is explained using Table 1.

**Table 1:** Review Summary of related work

Sl. No	Research Area	Research Focus	Reference
1	The cardiovascular disease is predicted in diabetic patients using pulse wave velocity and machine learning.	A determination of pulse wave velocity in an Outpatient can be used to assess the cardiovascular risks.	Rafael Garcia-Carretero et al. (2019) [3]
2	Artificial plant optimization technique to prevent heart disease. This technique evaluates based on MAPO to predict high risk.	This study has investigated the high risk of cardiovascular disease and modified artificial plant optimization (MAPO) is applied on real life example i.e. cardiovascular disease, MAPO is used to calculate the pulse count using fingertip video	Perna Sharma et al. (2019) [4]
3	Machine Learning is used to detect heart rate and the possibility of heart disease.	Classification model is used for prediction.	Manoj Diwakar et al.(2020) [5]
4	Conceptual design of a wearable soft sensor based on machine learning for noninvasive cardiovascular risk assessment. The models used here is predicts the collected data from the sensor.	Wearable soft sensor transducer and patient's interview results are collected then data set is analysed using Machine learning models, also Classifiers are compared with k-fold procedure. The data is divided into three categories: safe, non-critically unhealthy, and	Pasquale Arpaia et al. (2020) [6]

		critically unhealthy cardiovascular status.	
5	Machine Learning Solution on Spark for Using Patients' Social Media Posts to Detect Heart Disease. The social media post will help know the eating habits and life style of a patient.	A predictive framework for heart disease was developed using Apache Spark and Apache Kafka. Our real-time architecture is made up of three components: Stream Processing Pipeline, Online Prediction, and Stream Processing Pipeline.	Hager Ahmed et al. (2019) [7]
6	Clustering of heart disease and chemo informatics datasets using a machine learning algorithm.	For defined data points, the CBDCGAN + DBC approach will cluster mixed categorical and numerical attributes. The addition of synthetic samples to the training set greatly improves classification tasks, which are notoriously difficult in mixed datasets.	K. Balaji et al., (2020) [8]
7	Machine learning as a supplementary method for detecting cardiac arrest in emergency calls.	Predictions made on the quality of audio calls, and predictions performed before the end of the call.	Stig Nikolaj Blomberg et al. (2019) [9]
8	Machine learning was used to predict cardiovascular events in patients for up-coming year with severe dilated cardiomyopathy.	Using machine learning, a method for predicting cardiovascular events (DCM)	Rui Chen et al. (2019) [10]
9	A technique for predicting cardiac arrest in the patient using sensitivity analysis.	OHCA rescue system is used to predict the disease.	Samuel Harford et al. (2018) [11]
10	Out-of-hospital cardiac arrest predictions are performed. The data is collected for a certain period of time for certain locality.	A favourable machine learning model is prepared. The models help to differentiate the data set.	Yohei Hirano et al. (2020) [12]
11	Predictions of risks in type 2 diabetic patients. The diabetic data set can be sub merged with the training set.	The diabetic patient data set is evaluated for the predictions with help of evaluated data set using diabetic information.	Md Ekramul Hossain et al. (2021) [13]
12	Emergency in triage modeled for risk of heart disease. The triage helps to know the risk of cardiac arrest.	Low risk and high risk patients are predicted based on triage.	Huilin Jiang et al. (2021) [14]

### 3. OBJECTIVES :

The objects of this paper are;

- To detect cardiac arrest in a patient with the aid of data collected from various sources.
- To identify the possibility of cardiac arrest, caused by coronary artery disease with asymptotic or with symptoms
- To identify and analyse the possibility of heart disease using machine learning Classifier models through the collected sample dataset.
- In the implementation phase, to find the classifier model, to evaluate which performs better than other classifier models based on the accuracy.

#### 4. METHODOLOGY :

In this paper different dataset are considered for cardiovascular a problem which includes Cleveland Dataset of US patients with total 303 records, Statlog Dataset of UK patients with 270 records, and Hungarian dataset of Hungary, Switzerland with 617 records. These datasets are analyzed using various statistical measures.

#### 5. DATASET :

The description of data set is briefly described in this section; we are explaining the dataset used for cardiac arrest prediction model. Also, comprehensive overviews of the dataset's structure for real-time evaluation are made.

##### 5.1 Dataset Description

We have collected and merged three data sets, Clevel and Dataset of US patients with total 303 records, Statlog Dataset of UK patients with 270 records, and Hungarian dataset of Hungary, Switzerland with 617 records. This dataset is the most comprehensive data set with a combination of all three data sets consisting of 11 common features with 1190 records. This collection of data is used to train and evaluate machine learning algorithms. Descriptive data about our dataset are summarized in Table 2. The total 11 independent input features are described, age of the patients, age is described in years, gender of the patient either Male or Female, Male is denoted as one and female is denoted by 0, and gender is nominal variable and nominal is a categorical and does not follow any order, chest pain type is a type of chest pain what patient is experiencing, and it is categorizing in to 4 types, typical angina, atypical angina, non-anginal pain, and asymptomatic [9]. Typical anginal are chest pain, which is caused by reduced blood flow to the heart. It can be observed by an individual by a heaviness or tightness in the chest. Atypical angina, causes pain, but unrelated to heart, but it is caused by respiratory, musculoskeletal, and gastrointestinal or due to some heavy exercises [10]. Non – anginal pain is completely irrelevant to heart disease, and lastly, asymptomatic, without any symptoms. The resting blood pressure is measured in mm/HG [11]. Cholesterol measured in mg/dl. Fasting blood sugar is a Numeric variable in this dataset, here it is taken as 0 and 1. Fasting blood sugar is considered as one, if it is greater than 120 mg/dl else 0. Resting ECG is represented in 3 distinct values 0 for Normal, 1 for abnormality in ST-T wave, 2 for left ventricular hypertrophy. Abnormality in ST-T wave can be found in a heart patient and it is measured using ECG waves, left ventricular hypertrophy is, heart left pumping chamber has some stiffness and not pumping efficiently, and leads to cardiac arrest or heart attack [12].

**Table 2:** Patient Characteristics of Serious Cardiac Disease

Features	All Patients (n = 1190)	Patients with Events (n = 629)	Patients without Events (n = 561)
<b>Baseline</b>			
<b>Age, years (mean ± SD)</b>	54 ± 9	56 ± 9	51 ± 9
<b>Sex, n (%)</b>			
Male	909 (76%)	559 (89%)	350 (62%)
Female	281 (24%)	70 (11%)	211 (38%)
<b>Chest Pain Type</b>	3 ± 1	4 ± 1	3 ± 1
<b>Resting BP(mm/HG)</b>	132 ± 18	134 ± 20	129 ± 16
<b>Cholesterol</b>	210 ± 101	191 ± 120	232 ± 70
<b>Fasting Blood Sugar (mg/dl)</b>	0.2 ± 0.4	0.29 ± 0.45	0.11 ± 0.32
<b>Resting ECG</b>	0.69 ± 0.87	0.63 ± 0.86	0.75 ± 0.86
<b>Max Heart Rate (bpm)</b>	137.7 ± 25.5	129 ± 23.7	150 ± 22.7
<b>Exercise Angina</b>	0.38 ± 0.48	0.60 ± 0.48	0.13 ± 0.34
<b>Old Peak</b>	0.92 ± 1.08	1.33 ± 1.18	0.46 ± 0.73
<b>ST Slope</b>	1.62 ± 0.61	1.91 ± 0.51	1.29 ± 0.53

Maximum Heart Rate is calculated in Numerical Values [13–15]. Exercise angina is feeling of pain after an exercise and it is a nominal variable, if pain is observed then 1, else 0. The old peak is exercise induced ST depression compared with the state of rest. Then, the ST slope measured in terms of slope during peak exercise as shown in Fig.1, they are three types, up sloping, down sloping and horizontal. After peak exercise, for a normal patient blood pressure will be up sloping, if the blood pressure of the patient is down sloping or horizontal, there is a possibility of cardiac arrest in future or it can be observed that the patient has heart disease [16–20].

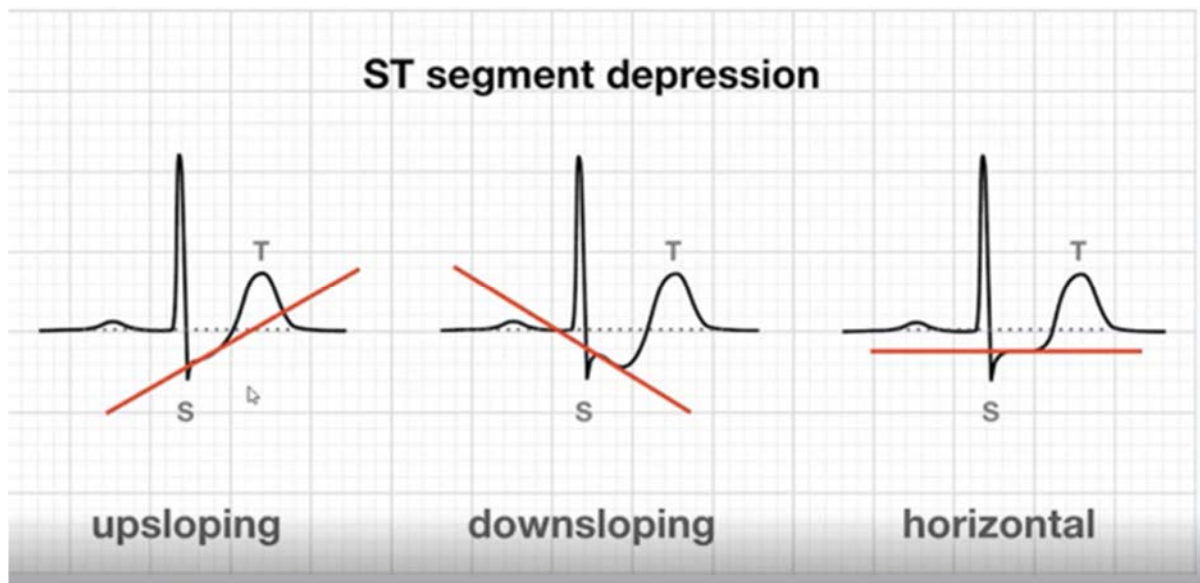


Fig. 1: ST - Slope depression

## 5.2 Building the Classifier Model using the dataset

The main goal of building model is to use several machine learning algorithms to achieve the highest accuracy. Model building consists of different stages : a) Data Pre-processing, b) Exploratory Data Analysis, c) Outlier Detection & Removal, d) Training and Test Split, e) Cross Validation, f) Model Building, g) Model evaluation & comparison, h) Feature Selection, and i) Model Evaluation.

### a) Data Pre-processing

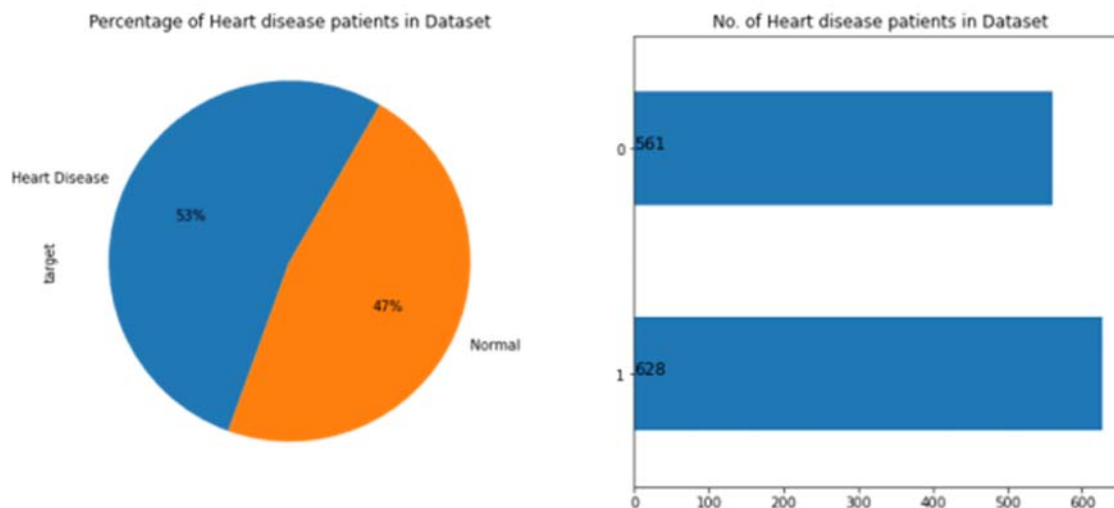
Data pre-processing are an essential step in machine learning to represent data suitable for the algorithms and classification. In this work, we have included necessary data pre-processing, such as missing values removal, feature encoding, and then transformed in to categorical variables and the generation of baseline characteristics of the study sample known as and prediction variables are given in the Table.2, and baseline characteristics are generated using a statistics tool R studio.

### b) Exploratory Data Analysis

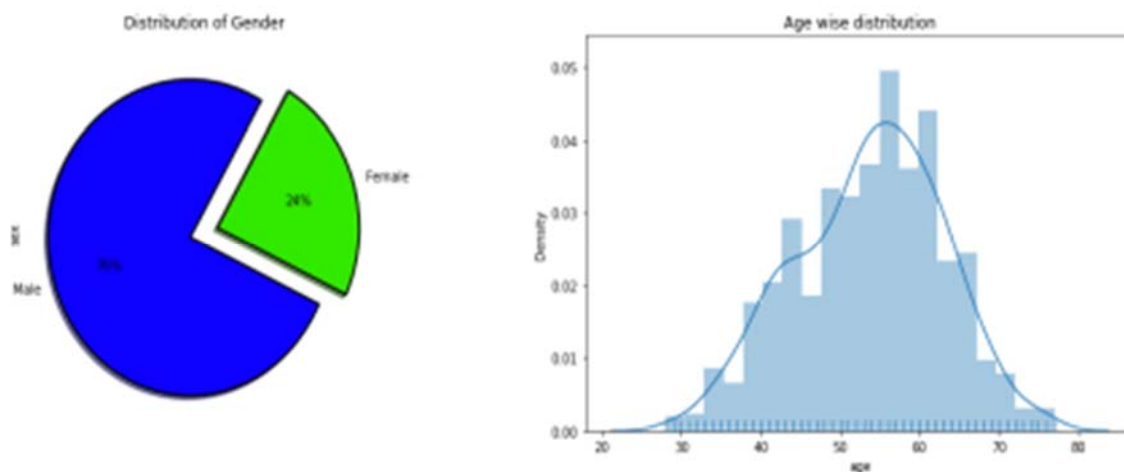
In this stage, we have analysed the shape of the dataset and generated the statistics of numerical and categorical features as shown in Table 2. And we also checked whether the selected dataset is balanced or not, as we can see in Fig. 2. The percentage of heart patients in data is balanced.

In this dataset (figure 2), 53% of patients are heart patients and 47% are normal patients. In addition, in numerical terms, the number of normal patients is 561 and number of heart patients are 628.



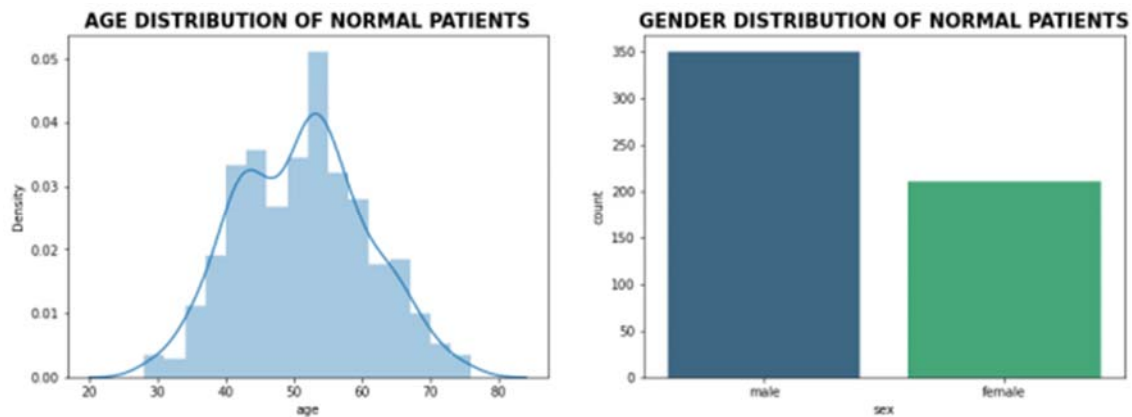


**Fig. 2:** Percentage and number of Heart disease patients in dataset



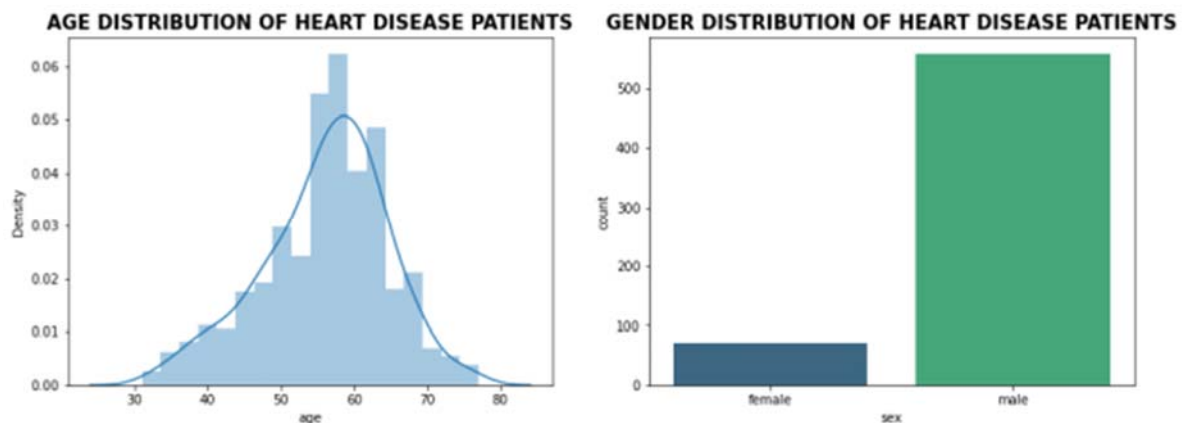
**Fig. 3:** Distribution of Gender and Age

The demographic part of the patients is also analysed based on the statistics generated in Table 2. The demographic analysis is performed on the age and gender of the patients, as shown in Fig. 3, showing that 76% of the distribution is male and 23% of the distribution is female. In addition, in the age wise distribution, mean age is 54 years. In Fig.3 the age distributions of healthy patients are shown and mean age is 51 for normal patients. In addition, the gender distribution of male and female is shown in Fig.3, as it is observed that numbers of normal patients are high in male patients.



**Fig. 4:** Distribution of Gender and Age for healthy patients

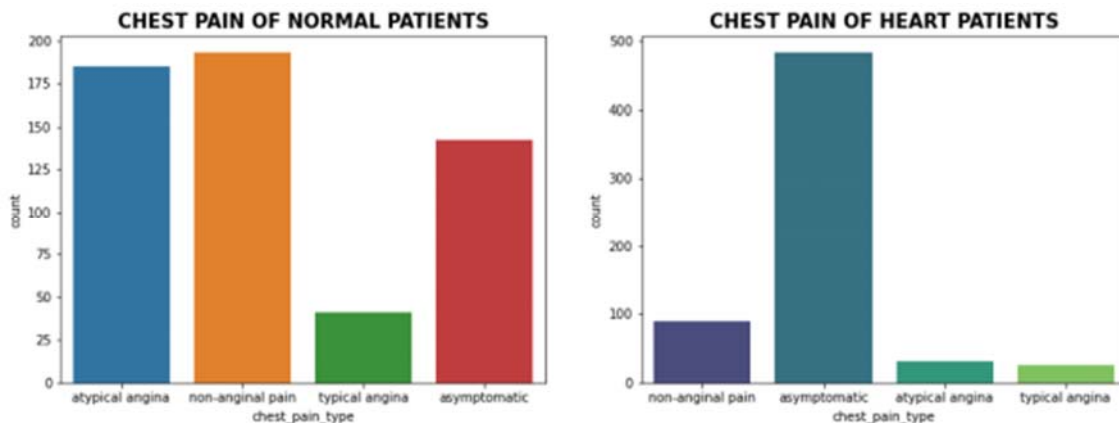
It is also observed in Fig.4, the age distribution of heart disease patients, which is high at the age of 56 and 60. And in the gender distribution of heart disease patients, it can be observed that male patients have a high risk of heart disease. In addition, female patients are low in number compared to males.



**Fig. 5:** Age and gender distribution of heart disease patients

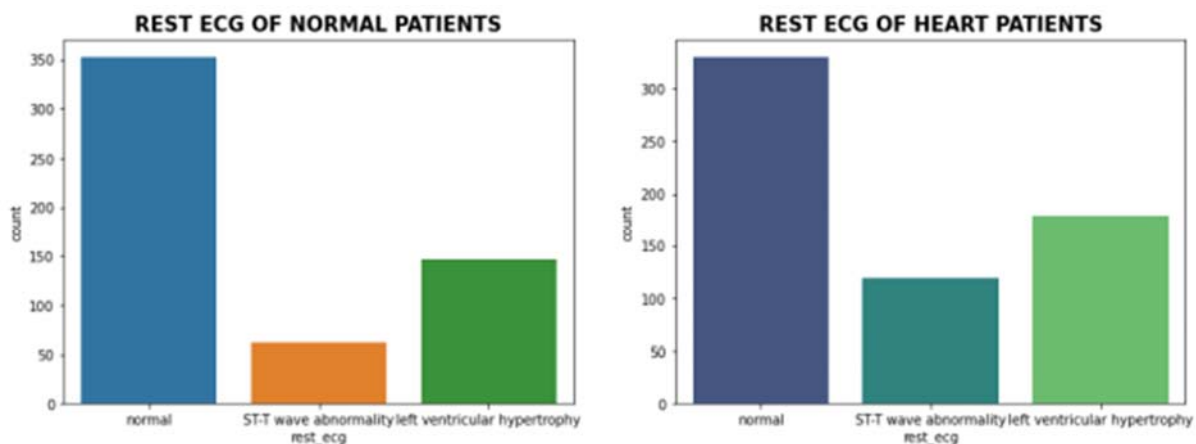
The chest pain of healthy patients is analysed in Fig.5, and all the four types of angina are analysed and it is observed, and shows that typical angina of normal patients are less. It is mainly observed that asymptomatic angina is high in heart disease patients. The ECG of normal and heart patients is compared, as shown in Fig.6, and it is observed that rest ECG of healthy and heart patients is normal, left ventricular hypertrophy and ST wave abnormality is much higher in heart patients. As shown in Fig.7, the ST slope of normal and heart patients is also analysed, after the exercise and if the patient has heart disease ST slope will be either down sloping or flat [21–25]. So, in our analysis, it is observed that ST slope of normal patients are up sloping and ST slope of a heart patient is flat. The distribution of numerical features in terms of pair plot is analysed in Fig.8, and in pair plot we have plotted based on the pairs cholesterol, resting blood pressure and age, as shown in the scatter





**Fig. 6:** Anginal pain analysis of normal and heart patients

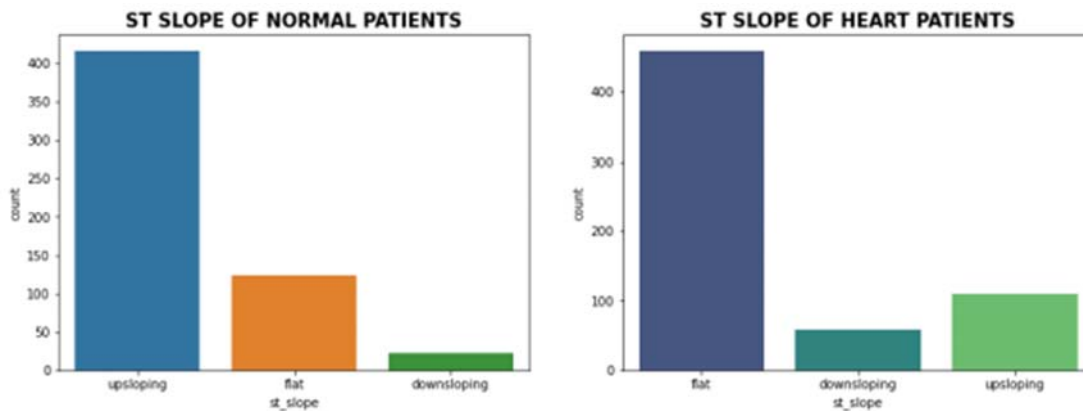
In pair plot analysis (Fig. 9) it is observed according to the dataset, as age increases the chances of cardiovascular disease and cardiac arrest risk increases. It is also observed that in terms of peak in age, the risk is higher when age of patient is more than 70. After the observation of scatter plot, the pear analysis done for resting blood pressure and Cholesterol, as shown in Fig.8, in this scatter plot, it is possible to see the outliers clearly, and it is observed that some of the patient's cholesterol and blood pressure is zero.



**Fig. 7:** Rest ECG analysis of Normal and Heart Patients

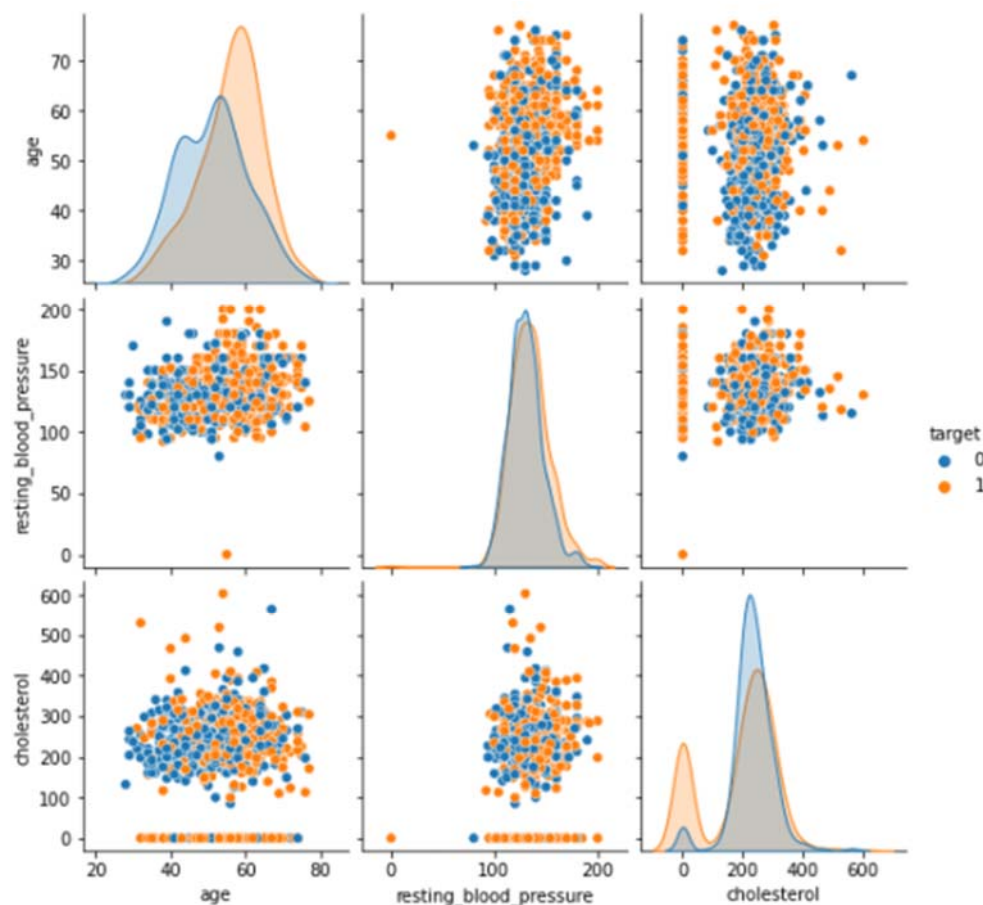
### c) Outlier Detection and Removal

The outliers can be removed and rectified from the data using the Z score, and outliers are sometimes unusual or usual data, it can be detected and filtered using Z score. After that data are segregated into features and target variables. After the segregation, the analysis is also done to find the correlations of the features with the target variable i.e., diabetes, as shown in Fig.9, in the correlations, it is observed that some values are positive and some are negative, and it can be observed that ST\_SLOPE Up sloping is negative and ST slope Flat is positive. ST slope flat has correlations with the diabetes and this positive



**Fig. 8:** ST slope analysis of Normal and Heart Patients.

The correlation shows the possibility of cardiac arrest in the patient. And ST\_slope value is closer to 1 so it is positive, and some other positive correlative variables are, ST\_Slope down slop, rest ECG left ventricular hypertrophy, sex\_male, ST depression, exercise-induced angina, fasting blood sugar, resting blood pressure, and age.

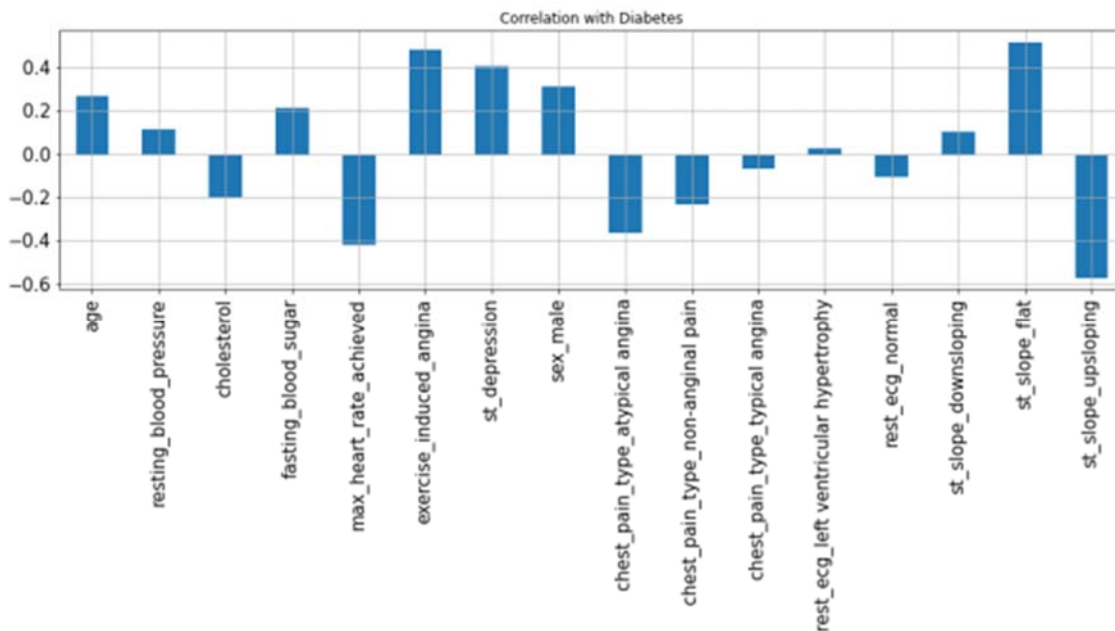


**Fig. 9:** Pair plot analysis of Patients.

#### d) Training and Test Split

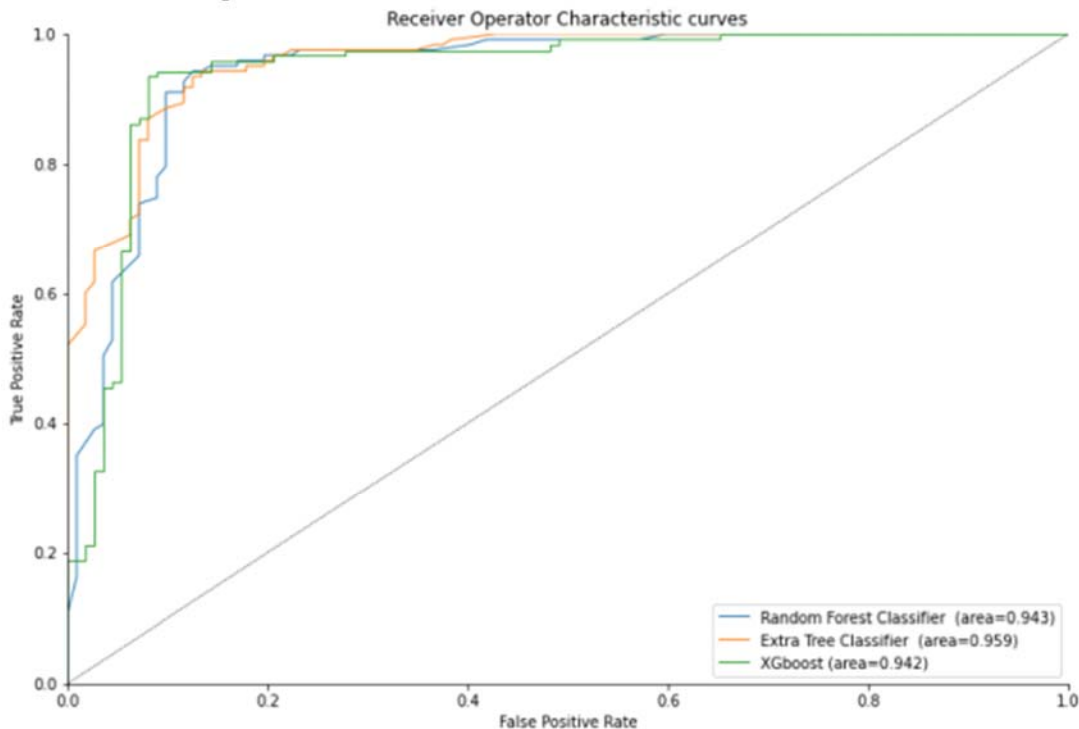
In the next phase, further analysis preferably can be done by performing train-test split of the dataset. Here, 80% is used as training data and 20% as testing data. The distribution of the target variable in the training

set is, 491 is heart patients and 446 is the number of healthy patients. In addition, in the test kit distribution of target variable,



**Fig. 10:** Correlations of all features with the target variable.

Heart patients are 123 and normal patients it is 112. In, training set total 937 and in Test kit 235 patients are included. After this phase, feature normalization is done



**Fig. 11:** ROC AUC curve.

The using MinMax Scalar method. We have used feature-normalization because, in the dataset the values are stored as either 0 or 1 for most of the variables. After the implementation of MinMax Scalar normalization

method for feature scaling [26], and we have selected only variables continuous from the training set, such as age, resting blood sugar, cholesterol, max\_heart\_rate\_received, and st\_depression, and these features are scaled down in the range of zero and same is applied to the test set.

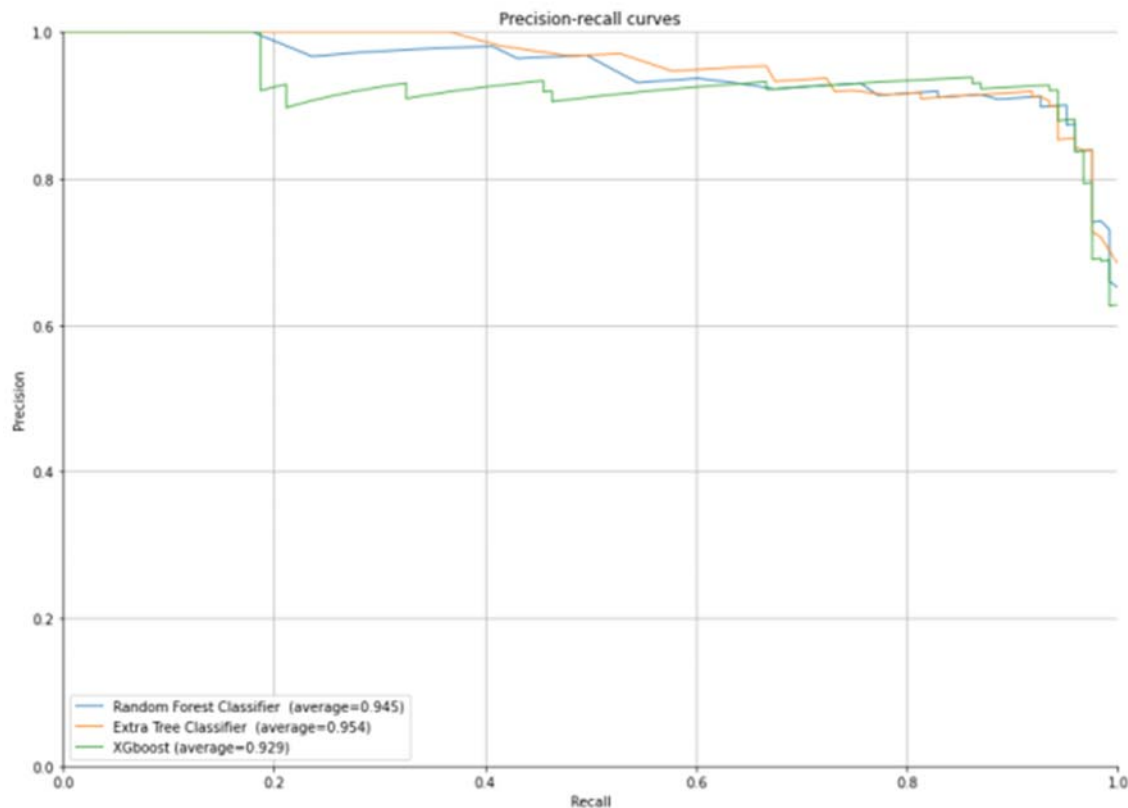


Fig. 12: Precision Recall curve.

#### e) Cross Validation

We construct different baseline models in this process and use 10-fold cross-validation to filter the best baseline models for use in level 0 of the stacked assembly system. Here, we have used all key machine learning algorithms. Various numbers of trees are used in classifiers to analyze the classifiers and performance is compared.

#### f) Model Building

In this phase, the Machine Learning classifiers are assigned by their criterion, Random Forest Classifier (criterion = 'entropy'), Multi-Layer Perceptron, nearest neighbor (n=9), Extra Tree Classifier (n\_estimators=500), XGBoost (n\_estimators=500), Support Vector Classifier (kernel='linear'), Stochastic Gradient Descent, Adaboost Classifier, decision Tree Classifier (CART), gradient boosting machine. These classifiers are built and analysed for selecting the best model.

#### g) Model Evaluation

We completed the most important evaluation metric for this problem domain during this phase, such as as sensitivity, specificity, Precision, F1-measure, geometric mean and Mathew correlation coefficient for a classifier model and we evaluated XGBoost Classifier is the best performer as it has highest test accuracy of 0.9191, sensitivity of 0.943 and specificity of 0.89 and highest f1-score of 0.9243 and lowest Log Loss of 2079. and finally, we have evaluated the models in ROC (receiver operating characteristic curve) under AUC(Area under the ROC Curve) curve [27–29].By plotting ROC, we can observe that, which machine learning algorithms have the highest area under ROC, and which ever machine learning algorithms highest area under ROC and whichever are more generalized.

Fig.11 clearly shows that Extra Tree Classifier has performed very well and more generalized, Extra Tree Classifier has achieved the highest average area under the curve (AUC) of 0.950. In the next step, we have evaluated the classifier model using precision recall curve as shown in Fig.11, precision recall curve is opposite to the AUC curve, and after the evaluation it can be clearly observed that Extra Tree Classifier has performed very well.

#### h) Feature Selection

In this phase, we have used several feature selection algorithms such as simple coefficient correlations, chi selector, RFE, Embedded LR selector, and Embedded Random Forestselector, Embedded LightGBselector, for selecting only those features, which are evaluated by the algorithms that, these features cause cardiac arrest in a patient [30–32].

**Table 3:** Feature selection

	Feature	Pearson	Chi-2	RFE	Logistics	Random Forest	LightGB M	Total
1	st_slope_flat	True	True	True	True	True	True	6
2	st_depression	True	True	True	True	True	True	6
3	max_heart_rate_achieved	True	True	True	False	True	True	5
4	exercise_induced_angina	True	True	True	False	True	True	5
5	cholesterol	True	False	True	True	True	True	5
6	age	True	True	True	False	True	True	5
7	st_slope_upsloping	True	True	True	False	True	False	4
8	sex_male	True	True	True	True	False	False	4
9	chest_pain_type_non-anginal pain	True	True	True	True	False	False	4
10	chest_pain_type_atypical angina	True	True	True	True	False	False	4
11	resting_blood_pressure	False	False	False	False	True	True	2

We have used a total of six algorithms for feature selection that causes cardiac arrest, after the evaluation, it is observed as given in the generation of Table.3 that all six algorithms have voted for ST\_Slope flat and ST\_depression. Furthermore, the Logistics algorithm has not voted for max\_heart\_rate achieved, and exercise\_induced\_angina, chi-2 has not voted for cholesterol. So, we have selected the features, which obtained five votes from the feature selection algorithms and we also selected st\_slope\_upsloping feature because it is a linked feature and cannot be bypassed. In this phase 7 features are selected.

#### i) Model Evaluation

After the feature selection, all steps starting from cross validation to model evaluation are repeated, and another new classifier model is implemented on the selected feature known as soft voting classifier, then the selected features again evaluated in ROC AUC curve as shown in Fig.13 and precision recall curve as shown in Fig. 14. and Extra Tree Classifier has obviously performed very well and, more generalized, has achieved the highest average area under the curve (AUC) of 0.957. In precision recall curve evaluation, as shown in Fig. 14, it can be clearly observed that Extra Tree Classifier has performed very well.

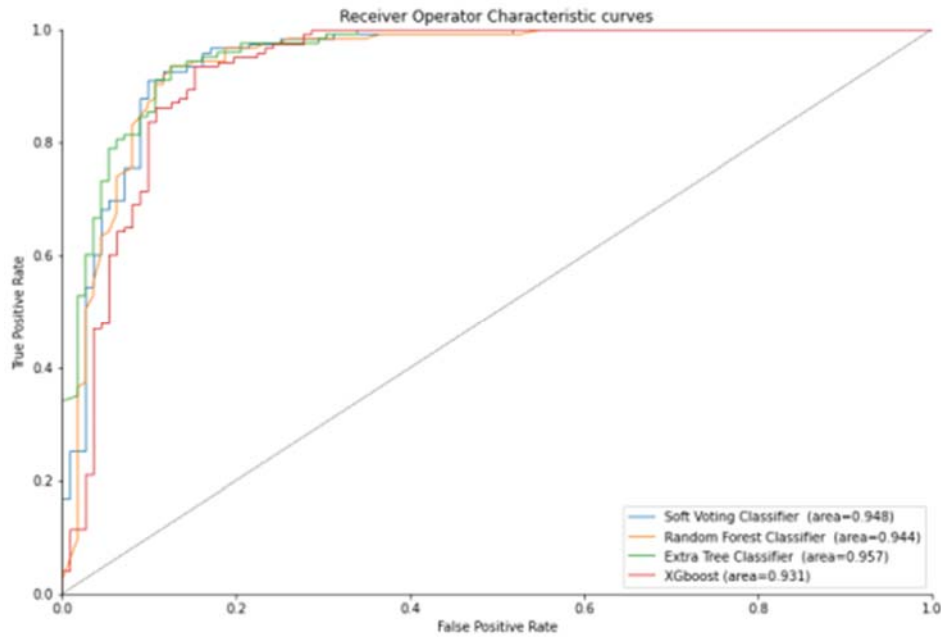


Fig. 13 : ROC AUC curve after feature selection.

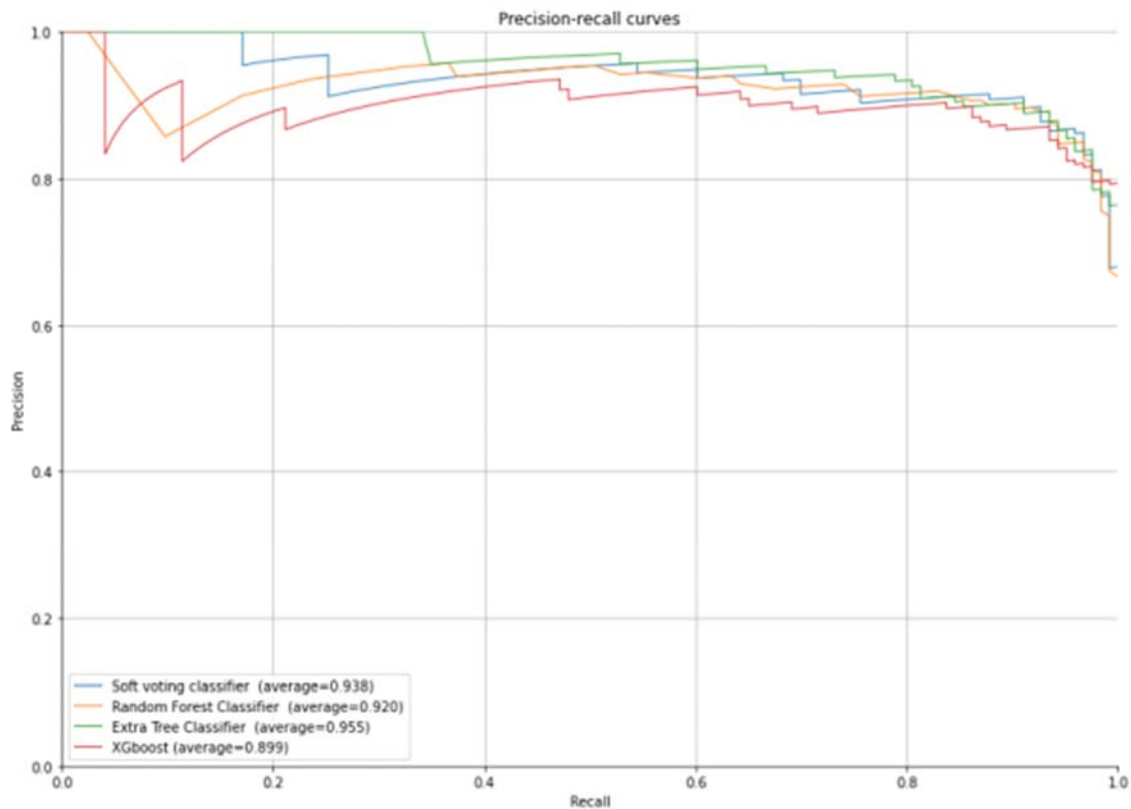


Fig. 14: Precision Recall curve after feature selection.

#### j) Feature importance

In the last phase, we have selected the features based on the attributes and contribution of these features to



the analysis. Fig. 15 shows the impact of each feature for cardiac arrest and their level of contributions.

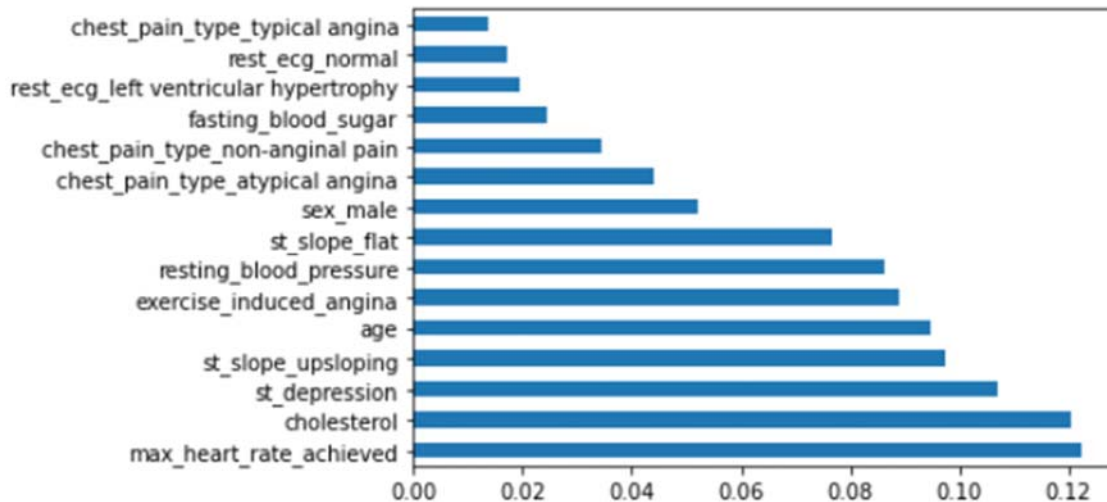


Fig. 15: Ranking of features

## 6. CONCLUSION :

In this study, the evaluation of cardiovascular disease is done using popular machine learning algorithms, data. The dataset used in this study are from famous Cleveland datasets of US, Statlog Dataset of UK, and Hungarian dataset of Hungary, Switzerland. The dataset collected is then verified for the changes by splitting into training and test set. Data are then given for feature extraction; this phase extracts 7 features, which contribute to the event. In addition, extracted features are used to train the selected machine learning classifier models, and results are obtained and obtained results are then evaluated using test data and final results are drawn. Extra Tree Classifier has the highest value of 0.957 for average area under the curve (AUC).

## REFERENCES :

- [1] Garcia-Carretero, R., Vigil-Medina, L., Barquero-Perez, O., & Ramos-Lopez, J. (2020). Pulse wave velocity and machine learning to predict cardiovascular outcomes in prediabetic and diabetic populations. *Journal of medical systems*, 44(12), 1-10.
- [2] Przewlocka-Kosmala, M., Marwick, T. H., Dabrowski, A., & Kosmala, W. (2019). Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: a classification based on machine learning. *Journal of the American Society of Echocardiography*, 32(12), 604-615.
- [3] Sharma, P., Choudhary, K., Gupta, K., Chawla, R., Gupta, D., & Sharma, A. (2020). Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning. *Artificial intelligence in medicine*, 102(11), 1-40.
- [4] Mishra, R. K., Tison, G. H., Fang, Q., Scherzer, R., Whooley, M. A., & Schiller, N. B. (2020). Association of Machine Learning-Derived Phenogroupings of Echocardiographic Variables with Heart Failure in Stable Coronary Artery Disease: The Heart and Soul Study. *Journal of the American Society of Echocardiography*, 33(9), 322-331.
- [5] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials Today: Proceedings*, 37(9), 3213-3218.

- [6] Arpaia, P., Cuocolo, R., Donnarumma, F., Esposito, A., Moccaldi, N., Natalizio, A., & Prevete, R. (2021). Conceptual design of a machine learning-based wearable soft sensor for non-invasive cardiovascular risk assessment. *Measurement*, 169(9), 1-29.
- [7] Ahmed, H., Younis, E. M., Hendawi, A., & Ali, A. A. (2020). Heart disease identification from patients' social posts, machine learning solution on Spark. *Future Generation Computer Systems*, 111(9), 714-722.
- [8] Balaji, K., Lavanya, K., & Mary, A. G. (2020). Machine learning algorithm for clustering of heart disease and chemoinformatics datasets. *Computers & Chemical Engineering*, 143(8), 1-13.
- [9] Blomberg, S. N., Folke, F., Ersbøll, A. K., Christensen, H. C., Torp-Pedersen, C., Sayre, M. R., ...& Lippert, F. K. (2019). Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*, 138(1), 322-329.
- [10] Chen, R., Lu, A., Wang, J., Ma, X., Zhao, L., Wu, W., ...& Liu, H. (2019). Using machine learning to predict one-year cardiovascular events in patients with severe dilated cardiomyopathy. *European journal of radiology*, 117(6), 178-183.
- [11] Harford, S., Darabi, H., Del Rios, M., Majumdar, S., Karim, F., Hoek, T. V., ...& Watson, D. P. (2019). A machine learning based model for Out of Hospital cardiac arrest outcome classification and sensitivity analysis. *Resuscitation*, 138(3), 134-140.
- [12] Reddy, K. H., & Saranya, G. (2021). Prediction of Cardiovascular Diseases in Diabetic Patients Using Machine Learning Techniques. In *Artificial Intelligence Techniques for Advanced Computing Applications* (pp. 299-305). Springer, Singapore.
- [13] Hirano, Y., Kondo, Y., Sueyoshi, K., Okamoto, K., & Tanaka, H. (2021). Early outcome prediction for out-of-hospital cardiac arrest with initial shockable rhythm using machine learning models. *Resuscitation*, 158(11), 49-56.
- [14] Hossain, M. E., Uddin, S., & Khan, A. (2021). Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. *Expert Systems with Applications*, 164(9), 1-13.
- [15] Jiang, H., Mao, H., Lu, H., Lin, P., Garry, W., Lu, H., ...& Chen, X. (2021). Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *International Journal of Medical Informatics*, 145(11), 1-7.
- [16] Khanna, N. N., Jamthikar, A. D., Gupta, D., Piga, M., Saba, L., Carcassi, C., ...& Suri, J. S. (2019). Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning-based tissue characterization. *Current atherosclerosis reports*, 21(2), 1-7.
- [17] Kochav, S. M., Raita, Y., Fifer, M. A., Takayama, H., Ginns, J., Maurer, M. S., ...& Shimada, Y. J. (2021). Predicting the development of adverse cardiac events in patients with hypertrophic cardiomyopathy using machine learning. *International Journal of Cardiology*, 327(11), 117-124.
- [18] Kim, E. Y., Lee, M. Y., Kim, S. H., Ha, K., Kim, K. P., & Ahn, Y. M. (2017). Diagnosis of major depressive disorder by combining multimodal information from heart rate dynamics and serum proteomics using machine-learning algorithm. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 76(2), 65-71.

- [19] Kumar, P. M., & Gandhi, U. D. (2018). A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Computers & Electrical Engineering*, 65(9), 222-235.
- [20] Kwon, J. M., Jeon, K. H., Kim, H. M., Kim, M. J., Lim, S., Kim, K. H., ...& Oh, B. H. (2019). Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes. *Resuscitation*, 139(2), 84-91.
- [21] Javan, S. L., Sepehri, M. M., Javan, M. L., & Khatibi, T. (2019). An intelligent warning model for early prediction of cardiac arrest in sepsis patients. *Computer methods and programs in biomedicine*, 178(6), 47-58.
- [22] Park, J. H., Do Shin, S., Song, K. J., Hong, K. J., Ro, Y. S., Choi, J. W., & Choi, S. W. (2019). Prediction of good neurological recovery after out-of-hospital cardiac arrest: a machine learning analysis. *Resuscitation*, 142(7), 127-135.
- [23] Przewlocka-Kosmala, M., Marwick, T. H., Dabrowski, A., & Kosmala, W. (2019). Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: a classification based on machine learning. *Journal of the American Society of Echocardiography*, 32(5), 604-615.
- [24] Seki, T., Tamura, T., Suzuki, M., & SOS-KANTO 2012 Study Group. (2019). Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. *Resuscitation*, 141(6), 128-135.
- [25] Signorini, M. G., Pini, N., Malovini, A., Bellazzi, R., & Magenes, G. (2020). Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring. *Computer methods and programs in biomedicine*, 185(1), 1-7.
- [26] Tantimongcolwat, T., Naenna, T., Isarankura-Na-Ayudhya, C., Embrechts, M. J., & Prachayasittikul, V. (2008). Identification of ischemic heart disease via machine learning analysis on magnetocardiograms. *Computers in biology and medicine*, 38(7), 817-825.
- [28] Kumari, C. U., Murthy, A. S. D., Prasanna, B. L., Reddy, M. P. P., & Panigrahy, A. K. (2020). An automated detection of heart arrhythmias using machine learning technique: SVM. *Materials Today: Proceedings*.
- [29] Yang, J., Tian, S., Zhao, J., & Zhang, W. (2020). Exploring the mechanism of TCM formulae for treating different types of coronary heart disease by network pharmacology and machining learning. *Pharmacological research*, 159(6), 1-15.
- [30] Geetha Poornima, K. & Krishna Prasad, K. (2020). Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 4(1), 25-39.
- [31] Gazala Khan (2020). The Impact of Healthcare Accreditation on Patients' Satisfaction: A Literature Review. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 4(1), 49-56.
- [32] Dias, Edwin & Mathew, Chris Sara. (2019). Respiratory Therapy Interventions Preventing the Need for Bronchoscopy in a 12 year old, with Right Lower Lobe Atelectasis -A Case Report. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 3(1), 21-24.

\*\*\*\*\*