Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0270
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

An AI Based Smart Cardiac Monitor

Akhila L S [1*],Sunitha G P [2]

[1*]Student, [2]Associate Professor, Department of Master of Computer Applications,
Jawaharlal Nehru national College of Engineering

akhilasgowda18@gmail.com, sunithagpise@jnnce.ac.in

Abstract

Machine learning (ML) is one of the subcategories of artificial intelligence, starts with data observations. The timeliness in arriving at the right decision in dealing with patients is another crucial factor commonly practiced in the medical discipline. In this area, AI, particularly the ML and deep learning, come in handy in assessing the condition based on the data that is usually bulky in the healthcare industry. Thus, high incidence of heart disease continues to claim several lives each year in India. The WHO has opined that stroke is one among the few diseases that can actually be foretold and hence prevented, had there been early measures put in place.The proposed system justify to predict cardiovascular disease with a lot more accuracy, using risk indicators incorporated with ML techniques such as decision trees. The specific records by which this study is assessable is the Kaggle UCI –Heart Failure, the dataset is composed of 12 attributes that could help in the prediction. Another critical phase within the performance analysis process is the preparation of the collected data. This has been achieved by the feature reduction, selection and prediction and is achieved by the implementation of decision tree algorithm and the constant accuracy of 96%.

Keywords: Artificial Intelligence, Feature Reduction, Machine Learning

1.Introduction

Cardiac diseases are a group of hazardous ailments, which, if not treated, may be fatal. Moreover, cardiovascular diseases are challenging to determine due to multiple downstream features including polygenic disorder, high symptoms, high cholesterol, irregular pulse, etc. The patient's past medical historycan be utilizedadvantageously to easily find risk factor information and sort a patient record by risk of heart disease. This one could easily predict the heart disease with the physician's skill without any occasional diagnostic tool, for example, ECG or ECHO tests. This will however depend with the ability of the doctor in thinking about these trails to estimate illness. Therefore, it is desirable to use a ML model to predict diseases, with the major decision being how to choose the system that assures the highest accuracy.

Heart attack is a severe medical condition characterized by sudden failure of the heart and an abrupt stoppage of blood circulation to the brain and to the different parts of the body. The outcomes are better when intervention is done early enough to increase the survival rate as well as to minimize the extent of neurological impairment. Nonetheless, the potential to predict potential cardiac arrest remains a difficult task owing to the unpredictable and multifactorial nature of the event. That is why accurate ability of predicting cardio arrest can improve the quality of treatment significantly and even save lives. The machine-learning (ML) techniques in predictive modellingcan be a solution to this challenge. When it comes to the selection of machine learning algorithms, the decision tree classifier is accessible to its simplicity, interpretability, and performance on clinical data inputs. This brings forward the decision tree classifier for the classification purpose.Compared with other classification methods, the decision tree algorithm has advantages in the application environment in clinical medical centres, because of its convenience and ability to classify new samples. The success rate of various decision tree (DT) implementations in this field increases with the volume of fresh data generated. While numerous research studies in diverse application domains primarily focus on using DT for model-free prediction and extracting rule sets with graphical representation, this may facilitate decision-making processes, particularly in the heart disease sub-domain.

1.1 Objectives

1. To design optimized machine-learning model to predict the cardiac problems.
2. To analyze the model developed with ECG dataset and other data.

2. Prior Art

Few of the prior classification techniques used to evaluate heart disease talked in the particular section. Mertozcan et.al [1] employed a decision tree algorithm to evaluate and predict heart disease.They construct and train the decision model using telehealth history data of 1190 patients. Alamgir A et.al'shighlights the use of AI technologies to predict cardiac arrest in any situation [2]. Jiaming Chen et.al.'s study [3] explores health-monitoring devices in smart communities, using advanced ML for ECG analysis to detect cardiac conditions. Ashraf Ewiset.al.'s paper [4] highlights the building of advanced ML algorithms for early detection and diagnosis of cardiovascular diseases. R Gomalavalliet.al.'s paper [5] presents a system developed with Lab-VIEW for alert reminders and continuous real-time monitoring geriatric patients. Sai Krishna Reddy et.al.'s research [6] is been induced in better predicting Cardiovascular Diseases byusing ML methods like Decision Tree and Naïve Bayes included with few risk factors. The dataset used for the overview of the study was the Heart Failure Dataset containing 13 items. Focusing on the probability of the cardiac arrest depending on the differences of regulated or unregulated variables in certain data sets with the help of ML algorithms is the purpose of the article [7] by R Karthikeyan et.al. The purpose of the study [8] by Hyeonhoon Lee et.al is to build and evaluate a machine learning-based real-time model for in-hospital cardiac arrest predictions using electro-cardiogram (ECG)-based heart rate variability (HRV) measures. The project [9] byDheepakG incorporates the topic of Internet of Things (IoT) and thus the information is available seamlessly regarding the matter of a remote supervisor. The scientific statement [10] by Antonis A. Armoundas outlines the current state of the art on the use of AI algorithms and data science in the diagnosis, classification, and treatment of cardiovascular disease.

As for the bulk records of existing work fo-cused on the UCI dataset, the majority of re-sults are informative, and the enhancement of classification methods is still a hotbed to gen-erate better prediction accuracy. However, dur-ing the survey it is made clear that the decision tree algorithm could predict cardiac arrest with higher accuracy. Thus, the motivation of this particular paper is to diagnose the heart disease accurately on using the machine learning algo-rithms like decision tree.

3. Proposed Methodology

This part will overview the operation ofan intelligent system that uses machine-learning algorithms to forecast a person's risk of cardiac arrest. It involves the following steps

3.1 Data collection:

The study uses the Kaggle UCI dataset [11], which is namely Heart disease dataset. It com-prises of 919 patient records, that is incorpo-rated into this comprehensive data. It

Age	Sex	ChestPain Type	RestingBP	Cholesterol	Fasting BS	Resting ECG	Max HR	Exercise Angina	Old Peak	ST Slope	Heart Disease
40	M	ATA	140	289	0	Normal	172	N	0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
37	M	ATA	130	283	0	Normal	98	N	0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0	Up	0

contains with 12 attributes and they are Patient's age, Patient's gender, type of Chest pain, RestingBP, cholesterol, Fasting blood sugar lev-el, Resting electrocardiogram, Achieved maxi-mal heart rate, Exercise-injected angina, old peak, ST_Slope, heart disease.

The following code snippet serves to load the datasets,

3.2 Data Pre-Processing:

Data pre-processing plays a keyrole in building an accurate classification model for cases of cardiac arrest. The Figure 2.2 deploys the workflow of the system. Following this conceptual idea behind building a decision tree.

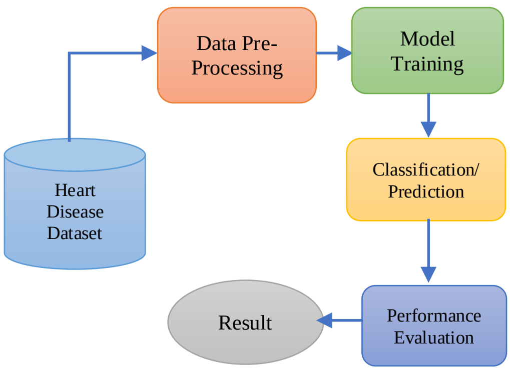

Figure 3.2: Work Flow Diagram

The ways to enhance the versatility and wholesomeness of this algorithm through some key data pre-processing techniques involves data cleansing, data modelling and data balancing.

3.2.1 Data Cleaning:

df = pd.read_csv('heart.csv') *//loading the csv data to a Pandas Data Frame*
df.head ()*//print first 5 rows of the dataset.*

The Table 3.1 spotlights the head of the suggesteddataset constituting of 5 rows specified with the values of various parameters.

Data cleaning is applied by removing missing values, outlier detection, removing duplicate entities.

i. Handling Missing Values:
- Imputation: Fill missing data including mean, median, or mode of the column.

$$x' = \begin{cases} x, & if\ x\ is\ not\ missing \\ mean(X), & if\ x\ is\ missing \end{cases} \quad \boldsymbol{eq-(1)}$$

Where *x* be the value of data entity and *x'* be the missing data
- Deletion: take off rows or columns with various missing data.

ii. Outlier Detection and Treatment:
- Z-Score: Remove or treat values that are a certain number of standard deviations away from the mean referred as std in the table 3.2.1.
- IQR Method: Remove or treat values outside 1.5 * IQR (Interquartile Range) from the first and third quartiles.

The values of all specified factors are noted in the Table 3.2.1

Table 3.2.1: Dataset Statistical Summary

Measure	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease	
count	918	918	918	918	918	918	918	0
mean	53.51089	132.3965	198.7996	0.233115	136.8094	0.887364	0.553377	0.1

std	9.432617	18.51415	109.3841	0.423046	25.46033	1.06657	0.497414	0.2
min	28	0	0	0	60	-2.6	0	0.3
25%	47	120	173.25	0	120	0	0	0.4
50%	54	130	223	0	138	0.6	1	0.5
75%	60	140	267	0	156	1.5	1	0.6
max	77	200	603	1	202	6.2	1	0.7

The Figure 3.2.2 provides a summary of descriptive statistics for a dataset containing various clinical parameters. Each row in the table represents a different statistical measure, and each column represents a different clinical parameter. Here is what each row and column represents:

- count: The number of observations for each parameter.
- mean: The average value of each parameter.
- std: The standard deviation, which measures the amount of variation or dispersion from the mean.
- min: The minimum value observed for each parameter.
- 25% (1st quartile): The value below which 25% of the data falls.
- 50% (median): The middle value, which separates the higher half from the lower half of the data.
- 75% (3rd quartile): The value below which 75% of the data falls.
- max: The highest value observed for each metrics.

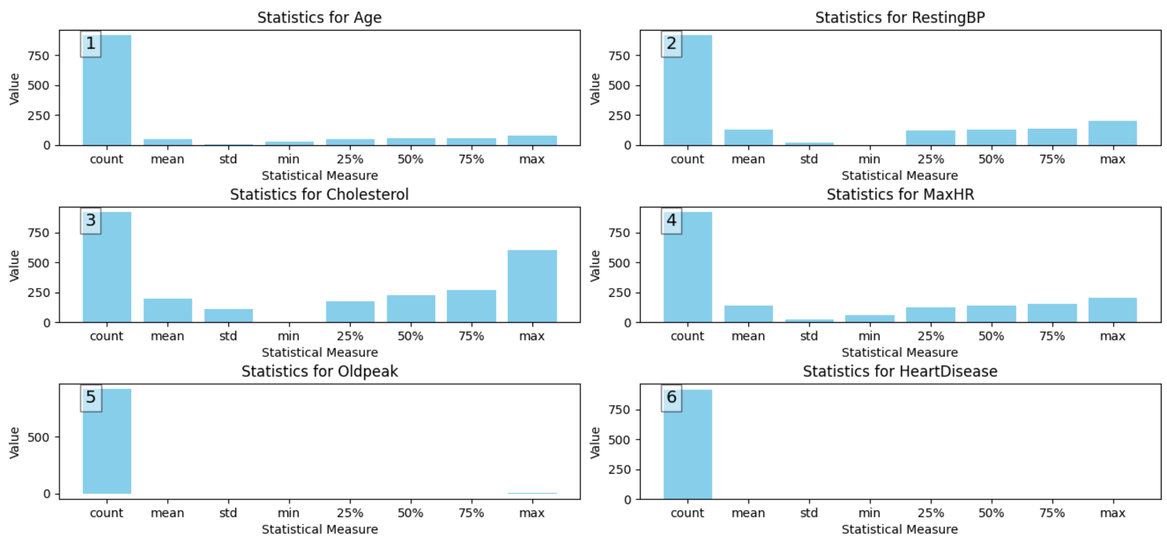

Figure 3.2.2: Plots of Value (y-axis) vs Statistical Measure (x-axis): (1) Statistics for Age, (2) Statistics for RestingBP, (3)Statistics for Cholesterol, (4) Statistics for MaxHR, (5) Statistics for Oldpeak, (6) Statistics for HeartDisease

3.3 Model Training with Decision Tree:

In order to provide decisions and forecasts that are as simple and accurate as possible, the decision tree algorithm builds a tree-shaped model of decisions by taking into account the characteristics of the provided data.

Step 1: Decision tree splitting criterion:

A Decision Tree Classifier defined as a supervised learning technique that divides the data set into further subgroups according to the features' outcomes. It is incorporated to the classification of data points.It uses criteria like Information Gain to split nodes.

a) Information-Gain for Feature Selection: Information gain is a powerful and commonly used criterion for feature selection.It measures the reduction in entropy or uncertainity from splitting a dataset based on a particular feature.

Information gain for a feature (A) is given by: $IG(D,A) = Entropy(D) - \sum_{v \in Values(A)} \frac{|Dv|}{|D|} Entropy(Dv) \quad \boldsymbol{eq-(2)}$

Where:

$$Entropy(D) = -\sum_{i=1}^{c} p \log_2(p) \boldsymbol{eq-(3)}$$

Entropy is measures the impurity or randomness in Dataset D, A is an attribute, *Dv* is the partition of *D* where attribute *A* has value *v*, and *c* is referred to the number of classes[6].

Features with higher information gain are considered more informative and are preferred for splitting the data.

After attaining the feature reduction, only six major parameters are assessed for evaluation and it includes age, cholesterol, maximum heart rate, fasting blood sugar, resting blood pressure and old peak as specified in the Table 3.2.1. Based on the six input parameters the prediction can be attained by the proposed approach.

Step 2: Training the model

Given a decision tree training dataset D, the training method divides the data iteratively, selecting the best split depending on the chosen criterion until the model fulfils a termination criterion. Entropy is taken into account by thedecision tree classifier as the criterion that trains the model and yields the model score.

Step 3: Hyper parameter Tuning

The hyper parameters of the model such as maximum depth, minimum samples split are optimized to enhance the performance of the model.

3.4 Prediction:

Based on the attribute values, the prediction starts at the decision tree's root and moves through the nodes to the leaf. The prediction is carried out by developing the decision tree algorithm.

The following code snippet describes how the Decision tree is implemented in the suggested approach:

A *DecisionTreeClassifier*is initialized with a variety of parameters specified to tailor its behavior in the code sample that is provided. Since the *ccp_alpha* is set to 0.0, the treeis not beingpruned. With the *class_weight* argument set to *None*, all classes are said to have the same weight. With *'entropy'* selected as the criterion}, the model gauges a split's quality based on information gain. The tree can only have four levels because the *max_depth* is only four. Since *max_features* is set to *None*, all features are taken into account when determining the optimal split. Furthermore, there is no restriction on the number of leaf node.

The following code snippet describes how the Decision tree is administred,

model = DecisionTreeClassifier(
ccp_alpha=0.0, *//no pruning is applied*
class_weight=None, *//all classes are supposed to have weight one*
criterion='entropy',*//measures the information gain*
max_depth=4, *//limiting the tree to 4 levels*
max_features= None, *//all the features are considered*
max_leaf_nodes=None, *//there is no limit*
min_impurity_decrease=0.0, *//split induces a decrease of impurity higher than or equal to this value*
min_samples_leaf=1, *// min number of samples required to be at a leaf node*
min_samples_split=2, *//min number of samples required to split an internal node*
min_weight_fraction_leaf=0.0, *//min weighted fraction of the sum total of weights required at a leaf node*
random_state=42, *// controls the randomness of the estimator*
splitter=‘best’)*//best split is chosen*

In the above given code, the decision tree guarantees that splits are made only if they reduce impurity; it measures the grade of splits using the entropy criterion.

With the entropy values, the following code will give the accuracy.

Prediction=model.predict(np.array(features).reshape(1, -1)) *//uses a trained model to make a prediction on a given set of features and np.array(features) converts the 'features' list to a NumPy array.*

Prediction,**score**=predict(X,y,features) *//to call a custom function named 'predict' to make a prediction and compute score.*

The code above converts the features to a suitable format and it will make prediction using the trained decision tree model, and the result is stored to the variable prediction.

3.5 Result analysis

The accuracy parameter, which most of the time determines the proportion of correctly categorized cases, sheds light on the classifier's degree of accuracy. Therefore, by computing parameters, the Confusion matrix aids in evaluating the effectiveness for the classification model. Certain attributes have been deduced from these characteristics, including F1-Score, recall, accuracy, and precision.
This system, the chosen output variable is a binary variable where 0 is the attribute of the person who will never develop heart failure while 1 is the attribute of the person with high propensity to develop heart failure. The confusion matrix of decision tree is given in table 3.5.1

Table 3.5.1: Confusion matrix

Value	Predicted0	Predicted1
Actual0	450	20
Actual1	15	434

Table 3.5.2: Major values of Performance Evaluation

Attributes	Value
Accuracy	96.20%
Precision	96.77%
Recall	95.74%
F1-Score	96.25%

3.6 Performance comparison

The suggested method has an astounding 96% accuracy rate in predicting the likelihood of cardiac arrest. The article provides an examination of predicting cardiac arrest using the model established with the ECG dataset and additional datasets, in addition to predicting an individual's likelihood of experiencing cardiac arrest using different characteristics.
The advantage of using real-time monitoring in the ECG dataset for cardiac arrest prediction and getting hold of the exact electrical signal of the heart, which may be opportune for intervention. Nevertheless, it is not easy to design and use, the data must be preprocessed beforehand and many computations are required. The proposed general indication method on the other hand utilizes twelve clinical parameters to assess the patient's state of health can be described as easy to practice and comprehend. It is more suitable for periodic appraisals and the first identification of the risks involved in a certain project. The application of these approaches naturally depends on the nature of the problem that is given, available resources, tools, and the field of medicine in which the prediction model can be utilized for prediction.

The following Figure 3.6.1 deploys the bar chart of the model performance of using ECG datasets referred from the study of Rajendra et.al[11] against other chosen 12clinical parameters for the prediction of cardiac arrest. The other 12 clinical parametersare referred to the data from the data collection [16], which performed better in comparison with numerous factors like accuracy, precision, recall and F1-Score values.

Table 3.6.1: Performance Comparison Table

Dataset	Accuracy	Precision	Recall	F1-Score
ECG	0.93	0.79	0.92	0.93
Heart Disease Data	0.96	0.96	0.95	0.96

The system's efficacy using ECG data and the data collected on cardiac illness at UCI [16]is compared in Figure 3.6.1. The comparison of the CA prediction outcomes is provided by the analysis. The analysis demonstrates that, in comparison to ECG data, the prediction of CA will be substantially more achievable with the UCI heart disease data collection [16].

One benefit of estimating CA using ECG datasets is that it allows for real-time monitoring and captures the heart's direct electrical activity, which can result in prompt interventions. On the other hand, the heart disease dataset is easier to use and analyse and offers a more comprehensive view of the patient's health state. Nevertheless, it is difficult to execute and demands a large amount of computational resources and data pre-processing. Although continuous ECG monitoring is more successful at capturing unexpected changes in the patient's health, this method is ideally suited for routine check-ups and risk classification. It is really challenging to make an easy forecast because the ECG data requires a lot more feature extraction techniques and pre-processing. The recommended data gathering [16] facilitates simple CA anticipation as soon as possible.

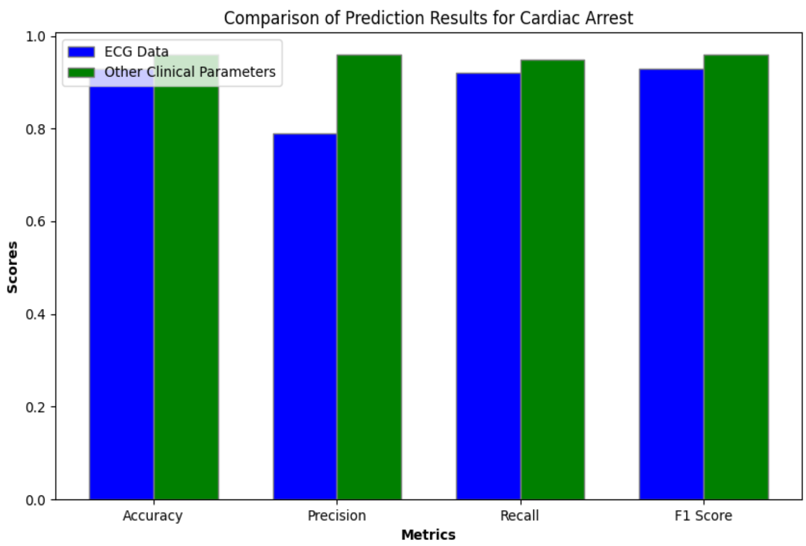

Figure 3.6.1: Bar graph plotting Scores (y- axis) v/s Metrics (x-axis)

4 Conclusion

Using a decision tree approach to predict cardiac arrest has the potential to improve patient outcomes. The goal of this project is to create an efficient predictive model by utilizing decision trees' advantages, such as their interpretability and user-friendliness. This initiative aims to significantly influence the medical care industry by tackling the pre-detection dilemma and offering practitioners practical insights.

5 Future Work

The majority of earlier research has focused on machine learning for early cardiac arrest prediction. A wearable device that records heart rhythm data will be the basis of the future system. It will transmit real-time sensor data via Bluetooth to a server or cloud, where a machine-learning algorithm will process it. The algorithm may include an Android application or web interface that will analyse the data and look for anomalies. Better yet, the device might instantly alert patients or medical staff, allowing for prompt intervention and care. Numerous advantages could occur if the future technology is compared to the cardiac monitoring methods used today. For instance, the system can be non-invasive, which eliminates the need for a face-to-face meeting, and it can provide continuous monitoring, making it possible to identify irregular or rare cardiac episodes.

References

1. Mertozcan, Serhatpeker et.al, A Classification and Regression tree algorithm for heart disease modellin,g and prediction, Healthcare Analytics, Science Direct, Vol. 3, Nov 2023, https://doi.org/10. 1016/j. health. 2022. 100130
2. AsmaAlamgir, Osama Mousa and Zubair Shah et.al, Artificial Intelligence in predicting Cardiac Arrest, JMIR Medicine informatics, IEEE vol. 9, iss.12,2021,https://medinform.jmir.org/2021/12/e3078
3. Jiaming Chen, Ali Valehi, and AbolfazlRazi et.al. al, Smart Heart Monitoring: The websites by using the early prediction ability and predictive analysis of ECG signals, IEEE, August-2019, 10. 1109/ACCESS. 2019. 2937875
4. Baghdadi et.al, Advanced machine learning techniques for cardiovascular disease early detection and diagnosis, Big journal of data, 2023, https://doi.org/10.1186/s40537-023-00817-1
5. R Gomalavalli, NithishGopi. G et.al, Smart Portable Cardiac Monitor using Lab View Application, Design Engineering, Research gate, Nov 2021, issue: 8, numbers of page: 16812 to 16817
6. V Sai Krishna Reddy et al. al, Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers, Journal of Physics: Conferences Scheduled for 2021
7. R Karthikeyan, D. VijendraBabu et.al, Cardiac Arrest Prediction using Machine Learning Algorithms, Journal of Physics: Annual Conference Series, London: Palgrave Macmillan, 2021, doi:10.1088/17426596/1964/6/062076
8. Hyeonhoon Lee, Hyun-Lim Yang et.al, Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in ICU, npj,DigitalMedicine,2023 https://doi.org/10.1038/s41746-023-00960-2
9. Dheepak g, Real-Time Proximity and Health Monitoring-Cum-Alert System with Ml-Enabled Predicting Model, JETIR Sep 2022, Volume 9, Issue 9.
10. Antonis A. Armoundas, Sanjiv M. Narayan, Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association, Volume 149, Issue 14, 2 April 2024; Pages e1028-e1050, https://doi.org/10.1161/CIR.0000000000001201
11. U.Rajendra Acharya, Hamido Fu-jita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals, Information Sciences, June 2017
DOI: 10. 1016/j. ins. 2017. 06. 027
12. Rishabh et al, Heart Diseases Prediction System Using CHC-TSS Evolutionary, KNN, and Decision Tree Classification Algorithm, Advances in Intelligent Systems and Computing, Springer, 2019,p.p.809-819
13. Yang F et.al, An Implementation of Naive Bayes Classifier, International Conference on Computational Science and Computational Intelligence, 2018, pp. 301-306.
14. Benllarchet.al, Improve Extremely Fast Decision Tree Performance through Training Dataset Size for Early Prediction of Heart Diseases, International Conference on Systems of Collaboration Big Data, Internet of Things & Security, 2019, pp. 1-5
15. Datasets: UCI Heart Disease Data, https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data/data