

Application of Random Forest and Decision Tree Classifier Approach for The Survival of Heart Patients using Characteristic Extraction *

Prabhudutta Ray, Dr. Raj Rawal , Dr. Ahsan Z. Rizvi

(Affiliation of authors): Prabhudutta Ray, Computer Science & Engineering, IAR University, Gandhinagar, India

Email address of corresponding author (phdrayprabhu@gmail.com)

Dr. Raj Rawal, Gujarat Pulmonary and Critical Care Medicine, Gandhinagar, India

Email address: dr.rajrawal@gmail.com

Dr. Ahsan Z. Rizvi , Computer Science & Engineering, IAR University, Gandhinagar, India

Email address : ahsan.rizvi@iar.ac.in

How to cite this paper:

Prabhudutta Ray, Dr. Raj Rawal and Dr. Ahsan Z. Rizvi, " Application of Random Forest and Decision Tree Classifier Approach for The Survival of Heart Patients using Characteristic Extraction ," *International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies*, Vol. no. 1, Iss. No 1, S No. 001, pp. 1–11, July 2024.

Received:25/06/2024

Revised: 20/07/2024

Accepted: 31/07/2024

Published: 31/07/2024

Copyright © 2024TheAuthor(s).
This work is licensed under the
Creative Commons Attribution
InternationalLicense (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Abstract

In modern time cardio pulmonary diseases are increasing at an alarming rate. It is a range of serious disorder that affects both the heart (cardio) and the lungs (pulmonary) at times resulting in death. In spite of continuous development in the field of medical science – both in the surgical and oral drugs, death due to cardiac arrest has remained a major cause of headache for the global population. Research suggests that mortality due to cardiac failure is mainly occurring in the early morning due to high blood pressure fluctuations. An advanced technology like machine learning has shown a new horizon of hope in the medical arena. ML carefully helps to select the patterns best suitable for data sets with various types of algorithms along with different correlation techniques. There are many features that are responsible for the heart disease like age, BP, sodium, cretonne, ejection fraction etc. After identifying important features by applying feature selection and ranking techniques the selected features are used as an input for the different ML algorithms. After implementing Random Forest algorithm it was observed that two sets of data tree emerged as comparison of results. Firstly a Decision Tree Classifier was selected to control the entire data sets, as well as to check the results with the generated tree for the selected samples. After generating selected samples its results got compared with the classification graph of the decision tree to achieve better result. It was observed that ML algorithms like Random Forest and Decision Tree Classifier Techniques can helps to predict the major reasons for cardio vascular diseases.

Keywords

Cardiovascular Diseases (CVD), LR, SVM Machine Learning (ML),

1. Introduction

The heart is the main source of supply of blood throughout the different parts of the body. The illness in the heart occurs due to aging, different parts of the coronary arteries are damaged and blocked for fats and cholesterol. When different parts of organ not received sufficient blood due to inability of the heart to pump then several disorders are noticed in the body like heart attack, high BP, cerebral attacks, brain death, etc. CVD diseases are the main reason for global death, affecting an estimated near about 18 million people globally. During cardiologist investigation, it was found that smoking is the major reason for COPD diseases and it is responsible for around 9 in every 10 cases. The toxic and harmful chemicals during smoking can damage the lining of the lungs and airways of the heart alveoli. COPD is a major cause of disability, and it is the fourth leading cause of death in the United States according to the CDC. Chronic obstructive pulmonary disease (COPD) is the name for a collection of lung diseases that includes chronic bronchitis, chronic obstructive airway disease, and emphysema. People who are suffering from COPD face major difficulties in breathing, primarily due to the narrowing of their airways; this is called airflow obstruction [1]. Spirometry can detect COPD before symptoms are recognized. Spirometry results can easily find the severity of the diseases. Spirometry measures how much air a patient can breathe out. Individual patients are at danger due to raised BP, dextrose, lipids as well as heavyweight and fattiness. All mentioned factors are the primary cause of heart attack. Cardio cardiovascular system consists of a network of blood vessels which includes veins, arteries, capillaries, etc. These blood vessels responsible for deliver blood all part of the body and any irregularities in usual blood flow from lungs part find different coronary abnormality. Especially in the low-income and middle – income territories more than 75% of deaths occur from cardiovascular diseases. Further due to stroke, sudden heart attack, and chest pain 80% of the deaths occur. Advances and development of health care system patient data are available for designing predictive models for cardiovascular diseases. The contributing factors to heart diseases are such as blood pressure, diabetes, smoking habits, high cholesterol, obesity, genetic predisposition, and age, even if the precise and actual cause of heart diseases is still unknown. So it is not easy to identify the heart diseases correctly. Various techniques related to data mining, ML, and neural networks have been used for the cause CVD. The nature of the CVD illness is perplexing, what is more, subsequently; the malady must be taken care of for the patients cautiously. The use of medical science and information technology are heavily utilized for finding out the different natures of heart diseases. Though the present healthcare system is “data rich” knowledge outcome is very poor. Finding hidden connections and patterns in healthcare data is still missing for successful data analysis methods. Exploratory data analysis is important for analyzing data that helps in finding strong correlations among attributes, and forecasting the future and hidden patterns. Furthermore, deep learning and AI techniques are used to correctly identify heart diseases at the early stage for heart disease prediction. Complex measurement can easily handled using ML and the prediction system can be made more intelligent and effective [2].

2. Related Work

Different researchers apply several types of techniques to solve coronary problems related predictions and develop classification models. **Table - 1** below describes related work, feature selection method, classifier used, and inference taken from the paper in tabular form.

Related work	Method	Classifier used	Inference
Hybrid HRFLM based techniques are used.	Discussed and designed new hybrid HRFLM techniques.[3]	RF and linear model	Accuracy achieved 88.7%
wearable system is based heart disease prediction.	Proposed the wearable system based CVD prediction[4]	NN, DNN,	Preprocessing feature selection and classification are performed
Hybrid ML technique used for heart diseas prediction	Hybrid ML technique that combines classification and regression.[5]	RF, DT	Accuracy achieved 88.7% .
Prediction of CVD disorder	Describe a web-based solution for CVD disorder[6]	Electrocardiogram; LSTM and CNN	MDCNN achieves an accuracy of 98.2.

DNN and linear SVC-based prediction techniques.	DNN and linear SVC-based integrated feature selection with gradient varnishing or explosion techniques.[7]	DNN,SVC	Achieved 98.56 % accuracy.
Early detection of CVD.	Suggested a steps for detection of CVD problems associated with the lifestyle.[8]	Feature selection, prediction model, classification algorithms.	Prediction accuracy is 99.1%.
Heart disease disorder	Model based on a recursion-enhanced random forest model for heart disease disorder.[9]	linear model with a random forest	Achived high accuracy and low classification error
NN based approach for CVD prediction	Proposed an ANN based techniques for CVD disease prediction[10]	DNN,NN	Accuracy 98.4% achieved.
Feature Ranking Approach for heart diseases	Describe feature ranking techniques for CVD prediction using the Kaggel data set of 303 patients using AI approaches [11]	KNN,CNN,RF,DNN	Good accuracy scores for CNN for prediction.
CVD-based disease using specific features	Predict heart disease using BMI is one of the important characteristics.[12]	SMOTE,KNN	Results shows 99.1%, 98.0% and 95.5 % accuracy gain.
ML based techniques for CVD prediction	compare the results of several data mining techniques.[13]	DT,NB,MLP,KNN,etc	shows SVM using the boosting technique outperforms.
Evolutionary Algorithm-based CNN for Predicting CVD	Proposed CNN-JSO techniques for predicting cardiac diseases[14]	CNN-JSO	Predicting cardiac diseases with 94.12 % accuracy.
Range of peripheral factors for heart disease prediction	Estimation of Prediction for Heart Failure Chances Using Various Machine Learning Algorithms.[15]	NB,NN,RF,KNN	Extracts key information to increase the accuracy
CVD prediction for coronary blockages using mathematical rough set theory.	Developed CVD by RST and an RNN-based approach. [16]	RST-RNN	RST-RNN has shown highly accurate results for predicting heart diseases.
Heart diseas prediction using modeling frame work	Developed a predictive steps for big data clinical settings. [17]	LR, RF, SVM, NB	Good results in predictive modelling of CVD..
Class imbalance in heart diseas prediction	Different learning rates were evaluated for the performance of MLP by using the sampling methods. [18]	RF,SVM, LR	Performance of MLP are evaluated
Prediction of CKD Disease using Data Mining Techniques	Applying different ML algorithm to a dataset for diagnosis.[19]	DT, LR, SVM, NB and NN.	Performance measurements of different methods.
wavelet transform to find features.	An analytic wavelet transform to detect epileptic features as input for training the model.[20]	ECG, SVM	Accurate in diagnosing the CVD.

3. Mathematical equations, subsections, tables, and figures

Correlation analysis helps for evaluating the relationship between two features of the dataset and to identify the strength of correlation among the features. The degree of relationships and the estimation of the correlation coefficient range between +1 to -1. An estimation of +1 shows a perfect and strong correlation between two features. The correlation coefficient goes towards 0 indicate the association between the features will be increasingly delicate. The degree of correlation relationships pointed out as negative and positive with aid of its operator notation. In real time analysis correlation techniques like Spearman’s, Pearson and Kendall’s are highly used [21].

A) Pearson’s correlation:

This correlation is highly recommended and used for the estimation of the relationship among the various features of the data elements and brings the high degree with respect to correlation.

Where

$$r_{x,y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}} \text{----- (1)}$$

n = number of observation
x_i = value of x (for ith observation)
y_i = value of y (for the ith obseration)

B) Spearman’s Rank correlation:

This is a non –parametric test that is used for checking the degree of correlation between the two features. This is the best type of correlation test that evaluated on a scale to provide the best possible outcome. This correlation used the following formula

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \text{----- (2)}$$

Where

ρ = Spearman’s rank corelation
d_i = difference between the rank of corresponding variables
 n = number of observation

C) Kendall’s Rank Correlation: -

This is a non –parametric type of correlation which helps in evaluating the dependencies among the different features. This correlation used the following formula

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \text{-----(3)}$$

Where

n_c = number of corelation
n_d = number of discordant

Using the above correlation techniques helps in finding the relation among the features. Features getting arranged or sorted out is the course toward changing unsavory data into features subject to the condition best correlation scores are considered. Next section focuses on the detailed methodology used for feature selection. It contains several stages like feature description, data preprocessing, feature selection, feature weight computation and model evaluation [22].

3.1 Feature Description

This section focuses on the description of the features describe in the **Table - 1**.

SL NO	Features	Description	Data types
1	age	Age of the patient in a year	Normal
2	sex	Gender of the patient(1 = Male 0 = Female)	Normal
3	pheight	Height of the patient	Normal
4	pweight	Weight of the patient	Normal
5	smoker	Smoking habit	Normal

6	alcoholic	Consume alcohol	Normal
7	mnocig	Number of cigerates	Normal
8	cp	Chest pain type	Normal
9	trestbps	Resting blood pressure(in mm Hg on admission to the hospital)	Normal
10	chol	Serum cholestrol in mg/dl	Normal
11	fbs	(Fasting blood sugar >120 mg/dl)(1 = true , 0 = false)	Normal
12	restecg	Resting electrocardiographic results	Normal
13	thalachh	Maximum heart rate achieved	Normal
14	exng	Exercise-induced angina	Normal
15	oldpeak	ST segments that induced by exercise relative to rest	Normal
16	slope	Pesk exercise ST segments	Normal
17	ca	Number of major vessels that colored by fluoroscopy	Normal
18	thal	Defect values	Normal
19	oxyzen	Oxegen level	Normal
20	output	Persons suffering from heart diseases = 1, Persons does not suffering from heart diseases	Normal

3.2 Feature Selection:

Features identification was based on experiments performed by collected data set from the B.M.J hospital Ahmadabad. We performed a set of experiments that dealt with combination of several features with different classification models like KNN, NB, DT, LR, and NN to identify significant features. In below Table-2 shows the features obtained from based on the performance classification model. [23-25]

3.3 Heart Diseases Prediction

This study aims to evaluate which heart diseases classification algorithm performs the best. This section summarizes all the study's findings and introduces the best performer based on various performance indicators. The best classifier is identified using the available data set with all 19 variables once the performance is obtained [26, 27]. **Second**, the best classifier for heart disease is found merely by considering 14 parameters that strongly correlated with the output value. Finally, this study has contrasted with various performance indicators of multiple ML algorithms for the available data set [28, 29].

3.4. Feature Description and Score

Below table shows the identified features, descriptions of each features and its corresponding score also.

Table 3: Feature Specification and Score

SL NO	Specs	Score
0	age	22.032070
1	sex	7.687854
2	pheight	40.415796
3	pweight	0.474771
4	smoker	5.065624
5	alcoholic	0.346179
6	mnocig	485.716942
7	cp	61.006780
8	trestbps	14.615828
9	chol	31.836538
10	fbs	0.132231
11	restecg	3.150748
12	thalachh	187.801154
13	exng	39.365492
14	oldpeak	72.285098
15	slope	9.492574
16	ca	65.585300

17	thal	5.657280
18	oxyzen	0.038461
19	output	135.000000

Table -4: Below represents 10 best features in descending order with respective scores by applying select best algorithm

SL NO	Specs	Score
6	mnocig	485.716942
12	thalachh	187.801154
19	output	135.000000
14	oldpeak	72.285098
16	ca	65.585300
7	cp	61.006780
2	pheight	40.415796
13	exng	39.365492
9	chol	31.836538
0	age	22.032070

3.5 . Feature importance

By applying XGBoost classifier ranking the features accordingly

- Feature: 0, Score: -0.00000
- Feature: 1, Score: 12.44483
- Feature: 2, Score: 0.00000
- Feature: 3, Score: -0.00000
- Feature: 4, Score: 93.32225
- Feature: 5, Score: 86.50811
- Feature: 6, Score: 26.74607
- Feature: 7, Score: 3.28535
- Feature: 8, Score: 0.00000
- Feature: 9, Score: 0.00000

Next we show the corresponding graph in respect to the features

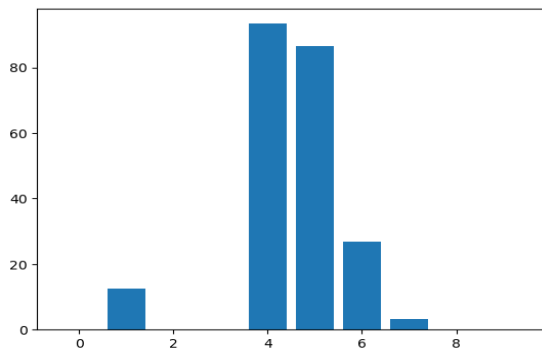


Figure: 2: Graph for feature scores

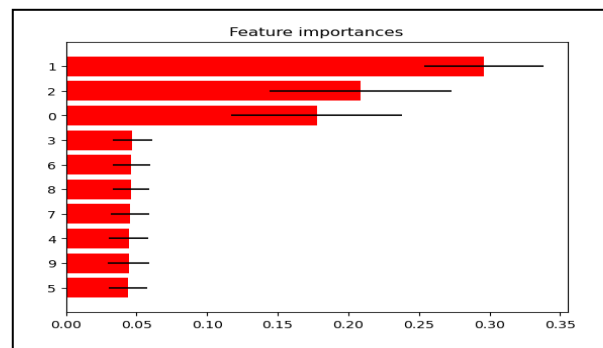


Figure: 3: Feature Importance Graph

In below figure 4 and 5 represent correlation matrix and heat map representation for the features

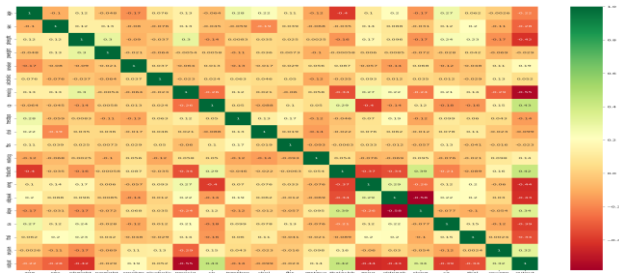


Figure 4: Correlation Matrix

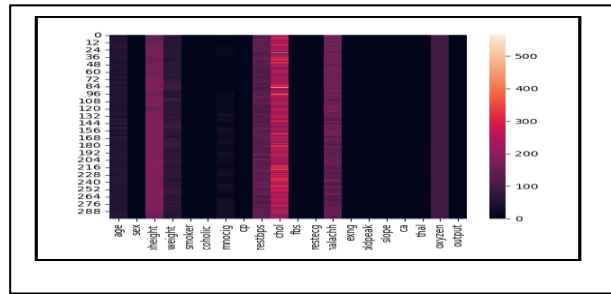


Figure 5: Heat Map

After performing above correlation and heat map analysis it is important to find the feature importance analysis for that we are using Random forest and Decision tree analysis and the result of the corresponding analysis are show [30]below.

```
rf.feature_importances_
```

```
array([0.02603599, 0.0999771 , 0.09619568, 0.02726774, 0.02692796, 0.08886828, 0.0431
5289, 0.09686836, 0.10378119, 0.01794767,0.03113806, 0.10472665, 0.02195891, 0.075459
44, 0.03014939, 0.04104958, 0.02863086, 0.0203159 , 0.01954834])
```

```
rf.feature_importances_.shape
```

```
(19,)
```

After that we are applying estimator to find the best feature importance which are as follows.

```
rf = RandomForestClassifier(n_estimators=2)
```

```
rf.fit(X,y)
```

```
RandomForestClassifier(n_estimators=2)
```

```
rf.feature_importances_
```

```
array([0.01626696, 0.00445712, 0.02596203, 0.0079985 , 0.,0., 0., 0.0215382 , 0, 0.01
645401,0.02109637, 0.01120308, 0.75215825, 0.01945793, 0.0.01744962, 0.03365852, 0.02
41303 , 0.02816909])
```

In the Random Forest classifier the classification performed internally by decision tree classifier. The results of the classification are shown below

```
rf.estimators_
```

```
[DecisionTreeClassifier(max_features='auto', random_state=1001812315),
```

```
DecisionTreeClassifier(max_features='auto', random_state=1586819166)]
```

From the above analysis it is important to check the feature importance of both the decision tree which are as follows

```
print(rf.estimators_[0].feature_importances_)
```

```
[0.          0.          0.03865324 0.01214767 0. 0.0. 0.04307641 0.          0.01262008
 0.          0.02240617 0.62536553 0.03891587 0.          0.03489924 0.06731704 0.04826
06 0.05633818]
```

```
print(rf.estimators_[1].feature_importances_)
```

```
[0.03253392 0.00891424 0.01327083 0.00384933 0. 0. 0. 0. 0.02028795 0.04219274 0.
0.87895098 0. 0. 0. 0. 0. 0.]
```

From the above it can be proved that our computed feature importance is correct with the value 0.01626696. From the estimator 2 returned value of feature importance is 0.03253392. Applying the following mathematical calculation we get the same value as listed in the feature importance.

```
print ((0+0.03253392)/2)
```

```
0.01626696
```

Next we applying the decision tree classifier for displaying the decision tree

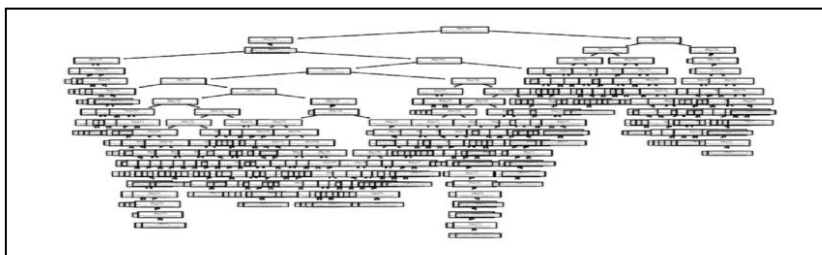


Figure 6: Decision tree classifier tree for the data set

Next we represent the feature importance list for the decision tree which are as follows

```
clf.feature_importances_
array([0.03271084, 0.08033034, 0.09180493, 0.03275711, 0.03437808,0.10547884, 0.04567
666, 0.0795971 , 0.10351878, 0.01632109, 0.0243667 , 0.11025557, 0.02095394, 0.059862
9 , 0.01848446, 0.06236492, 0.0362739 , 0.02413682, 0.02072702])
```

Next we perform the classification of the 15 samples and two classes and applying 2 features only that generate the tree which are as follows X,y =

```
make_classification(n_samples=15,n_classes=2,n_features=2,n_informative=2,n_redundant=0,random_state = 0 )
```

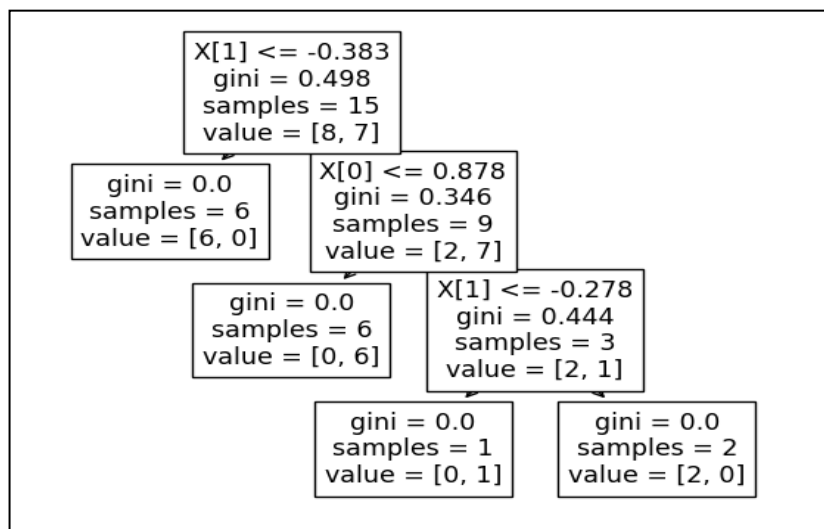
```
clf = DecisionTreeClassifier()
```

```
clf.fit(X,y)
```

```
DecisionTreeClassifier()
```

```
from sklearn.tree import plot_tree
```

```
plot_tree(clf)
```

**Figure 7:** Generate the tree for 15 samples

Next we show the feature importance of the corresponding tree which is as follows

```
clf.feature_importances_
array([0.23809524, 0.76190476])
```

After the above analysis we apply the full data set to check the classification results which are as follows

```
X,y = make_classification(n_samples=300, n_classes=2,n_features=19,n_informative=2,n_redundant=0,random_state = 0 )
```

```
clf = DecisionTreeClassifier()
```

```
clf.fit(X,y)
```

```
DecisionTreeClassifier()
```

```
from sklearn.tree import plot_tree
```

```
plot_tree(clf)
```

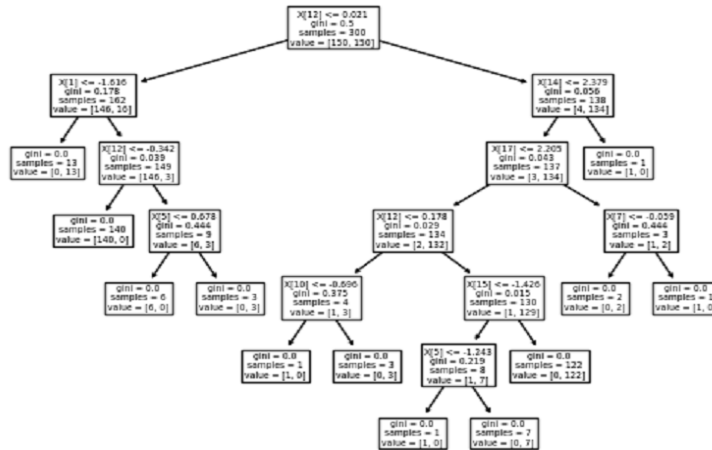


Figure 8: Decision Tree Classification

clf.feature_importances_

```
array([0. , 0.15306874, 0. , 0. , 0. , 0.03833333, 0. , 0.00888889, 0. , 0. , 0.01 ,
0. , 0.77151504, 0. , 0.01266335, 0.0015641 , 0. , 0.00396654, 0. ])
```

Performance Graph:

In this work, several ML models are developed and used cross-validation with 10-fold to choose the best models. Models with high accuracy scores are analyzed accordingly[31,32]. All the accuracy scores of the models are shown in **Table 2**.

Table -2: Comparison among ML Algorithms

SL No.	Algorithm	Accuracy score
1	Logistic regression	95%
2	Decision Tree	85%
3	SVC	96%
4	Random forest	90%
5	k-nearest neighbor	91%

From the above table-2, it is found that the support vector machine classifier gives a good accuracy core as compared to other algorithms. Further based on the same input vector $a = [[35,1,153,68,1,1,11,2,130,250,0,1,187,0,3.5,0,0,2,98.6]]$ all the algorithms correctly classify the output [30, 31].

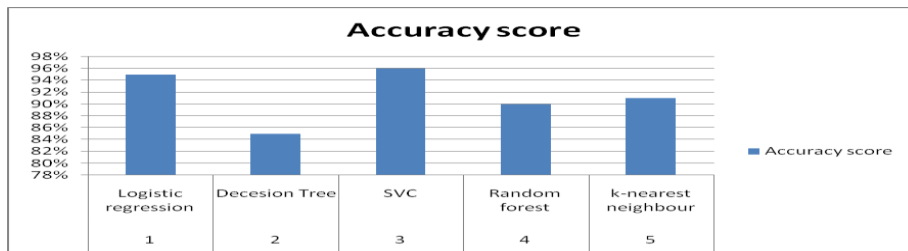


Figure 12: Comparison graph for ML algorithms with accuracy

Discussion: It was observed that using Random Forest algorithm itself creates two Decision tree for comparison of results. First we use the decision tree classifier for the entire data set then check the results with the generated tree for 15 samples. After that obtained results will be compared with the decision tree classification graph. Using the estimator several feature importance results are generated to find the good feature importance value 0.01626696. Using random forest estimator the listing of feature importance value are shown 0.03253392 0.00891424 0.01327083 0.00384933 0. 0. 0. 0. 0. 0.02028795 0.04219274 0. 0. 0.87895098 0. 0. 0. 0. 0. 0. 0. for better outcome of the feature importance results [32, 33].

4. Conclusion:

The aim is to utilize the maximum computation power of the machine and help the cardiologist to take quicker decision by identifying the good features that contributes more during prediction of heart diseases. For this research work we have used medical data set collected from the hospitals that originally contains 20 attributes and 303 rows. The following points are giving clear idea for objectives of this research work

- To enable the cardiologist in making a quick or fast and accurate decision regarding the best suitable treatment of the patients suffering with cardiac diseases or cardiac problem.
- To decrease the mortality ratio and to help the medical practioners with medical data containing 20 attributes with 303 rows collected from the hospitals.
- After analyzing the exploratory data analysis and data preprocessing steps several feature analysis with the help of statistical measures were applied to identify the good features which will be useful to detect and reduce chances of heart diseases.
- Processing and analyzing of raw data, ML techniques were used with the help of various algorithms to make a clear demarcation of heart diseases, as said earlier it will help to reduce the ratio of death among patients suffering with cardiac related problems.
- The algorithm used that included LR, NB, SVC, and XGBoost for better classification.

Future work: For better and accurate diagnosis of heart diseases deep learning network with various other improved optimization technique along with increasing the size of data set is suggested to be helpful for heart related diseases. to bring revolutionary or challenging improvements for treating of heart diseases ML and different optimization techniques , can also be taken into consideration.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1]. Anders Løkke Icon,Ole Hilberg Icon, Peter Lange Icon,Rikke Ibsen,Gunilla Telg Icon,Georgios Stratelis , “Exacerbations Predict Severe Cardiovascular Events in Patients with COPD and Stable Cardiovascular Disease–A Nationwide, Population-Based Cohort Study “ Pages 419-429 | Received 16 Nov 2022, Accepted 27 Mar 2023, Published online: 31 Mar 2023.
- [2]. Maaïke Giezeman, Josefin Sundh , Åsa Athlin, Karin Lisspers, Björn Stållberg, Christer Janson , “ Comorbid Heart Disease in Patients with COPD is Associated with Increased Hospitalization and Mortality – A 15-Year Follow-Up”, Pages 11-21 | Received 23 Jun 2022, Accepted 05 Oct 2022, Published online: 08 Jan 2023.
- [3]. Institute of Health Metrics and Evaluation. GBD Compare 2010. <http://vizhub.healthdata.org/gbd-compare/>. Accessed April 30, 2014
- [4]. Junejo, A.R., Shen, Y., Laghari, A.A., Zhang, X. and Luo, H., 2019. Molecular diagnostic and using deep learning techniques for predict functional recovery of patients treated of cardiovascular disease. *IEEE Access*, 7, pp.120315-120325.
- [5]. Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y.R. and Suraj, R.S., 2021, January. Heart Disease Prediction using Hybrid machine Learning Model. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1329-1333). IEEE.
- [6]. Khan, M.A., 2020. An IoT framework for heart disease prediction based on MDCNN classifier. *IEEE Access*, 8, pp.34717-34727.
- [7]. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in *IEEE Access*, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
- [8]. Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7, pp.81542-81554.
- [9]. Nadakinamani, R.G., Reyana, A., Kautish, S., Vibith, A.S., Gupta, Y., Abdelwahab, S.F. and Mohamed, A.W., 2022. Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Computational Intelligence and Neuroscience*, 2022.
- [10]. Nikam, A., Bhandari, S., Mhaske, A. and Mantri, S., 2020, December. Cardiovascular Disease Prediction Using Machine Learning Models. In 2020 IEEE Pune Section International Conference (PuneCon) (pp. 22-27). IEEE.
- [11]. Praneetha, M., Jesudoss, A. and Mayan, A., 2021, May. Cardiovascular Disorder Prediction using Machine Learning. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1665-1670). IEEE.
- [12]. Rahim, A., Rasheed, Y., Azam, F., Anwar, M.W., Rahim, M.A. and Muzaffar, A.W., 2021. An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases. *IEEE Access*, 9, pp.106575-106588.
- [13]. S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 204-207, doi: 10.1109/ISCC.2017.8024530
- [14]. Samir, A.A., Rashwan, A.R., Sallam, K.M., Chakraborty, R.K., Ryan, M.J. and Abohany, A.A., 2021. Evolutionary algorithm-based convolutional neural network for predicting heart diseases. *Computers & Industrial Engineering*, 161, p.107651.
- [15]. Ashwin Chavan, Bhairavi Chitnavis, Poorva Wadkar, Md. Ubaid Khan, " Estimation of Prediction for Heart Failure Chances Using Various Machine Learning Algorithms", *International Research Journal of Engineering and Technology (IRJET)* Volume: 10 Issue: 04 , Apr 2023.

- [16]. Talasila, V., Madhubabu, K., Madhubabu, K., Mahadasyam, M., Atchala, N. and Kande, L., 2020. The prediction of diseases using rough set theory with recurrent neural network in big data analytics. *International Journal of Intelligent Engineering and Systems*, 13(5), pp.10-18.
- [17]. Nitten S. Rajliwall; Girija Chetty; Rachel Davey, “Chronic disease risk monitoring based on an innovative predictive modeling framework”, *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-8,2017.
- [18]. Pinar Yildirim, “Chronic Kidney Disease Prediction on Imbalanced Databy Multilayer Perceptron”, *IEEE 41st Annual Computer Software and Applications Conference*, pp. 193-198, 2017.
- [19]. Maryam Soltanpour Gharibdousti, Kamran Azimi “Prediction of Chronic Kidney Disease Using Data Mining Techniques”, *Proceedings of the Industrial and Systems Engineering Conference*, 2017.
- [20]. M. Sharma, Tan, RS. Acharya, U.R. A new method to identify coronary artery disease with ECG signals and time-Frequency concentrated antisymmetric biorthogonal wavelet filter bank,, *Pattern Recognition Letters*, Vol. 125, pp. 235-240(2019).
- [21]. Chauhan A, Jain A, Sharma P, Deep V. Heart disease prediction using evolutionary rule learning, in 2018 4th International conference on computational intelligence & communication technology (CICT) (pp. 1–4). *IEEE*; 2018.
- [22]. Dey L, Mukhopadhyay A. Biclustering-based association rule mining approach for predicting cancer-associated protein interactions. *IET Syst Biol*. 2019;13(5):234–42.
- [23]. Dua, D., Graff, C. *UCI machine learning repository* [[http:// archi ve. ics. uci. edu/ml](http://archi.ve.ics.uci.edu/ml)]. Irvine, CA: University of California, School of Information and Computer Science; 2019.
- [24]. Domadiya N, Rao UP. Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case study on the heart diseases. *Sādhanā*. 2018;43(8):1–9.
- [25]. Domadiya N, Rao UP. Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data. *Procedia Comput Sci*. 2019;148:303–12.
- [26]. Fitriyani NL, Syafrudin M, Alfian G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. *IEEE Access*. 2020;8:133034–50.
- [27]. Han J, Pei J, Kamber M. *Data mining: concepts and techniques*. Elsevier; 2011.
- [28]. Ibrahim SP, Sivabalakrishnan M. An enhanced weighted associative classification algorithm without preassigned weight based on ranking hubs. *IntJ Adv Comput Sci Appl*. 10(10); 2019.
- [29]. Ibrahim SS, Sivabalakrishnan M. An evolutionary memetic weighted associative classification algorithm for heart disease prediction. In *Recent Advances on Memetic Algorithms and its Applications in Image Processing* (pp.183–199). Springer, Singapore; 2020.
- [30]. James SL, et al. Global, regional, and national incidence, prevalence, and yearslived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*392 (10159), 1789–1858; 2018.
- [31]. Jabbar MA, Deekshatulu BL, Chandra P. Graph based approach for heart disease prediction. In *Proceedings of the third international conference on trends in information, telecommunication and computing*. New York, NY: Springer. 2013. p. 465–474.
- [32]. Kannan AG, Castro TARVC, BalaSubramanian R. A comprehensive study on various association rule mining techniques; 2018.
- [33]. Khan SA, Yadav SK. Class-based associative classification using super subsets to predict the by-diseases in thyroid disorders. in *International conference on advances in computational intelligence and informatics* (pp. 301–308). Springer, Singapore; 2019.