

Comparative analysis and feature importance of machine learning and deep learning for heart disease prediction

Priyanka Gupta, Dambarudhar Seth

Department of Computer Science and Engineering, SRM Institute of Science and Technology (SRM IST), Ghaziabad, India

Article Info

Article history:

Received Jul 26, 2022

Revised Sep 18, 2022

Accepted Sep 29, 2022

Keywords:

Deep learning

Feature importance

Heart disease

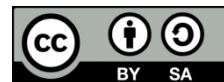
Hyperparameter tuning

Machine learning

ABSTRACT

Cardiovascular disease (CVD) or heart disease is one of the main reasons for early death, even at young age and that too often sudden. If it is detected more accurately, much before it seriously affects the individual, life can be saved through proper medication and changes in lifestyles. In this work different machine learning classifiers and a deep learning algorithm multi-layer perceptron (MLP) were applied on two different datasets, Framingham heart study dataset and UCI heart disease dataset for prediction of heart disease. These algorithms were optimized using hyperparameter tuning and compared for their performance measures and prediction accuracies. For different features, feature importance scores were calculated using machine learning algorithms. The features were ranked according to their scores. Out of various classification algorithms, random forest algorithm has shown the best results with prediction accuracy of 97.13% for Framingham dataset. MLP has shown good performance for both the datasets.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Priyanka Gupta

Department of Computer Science and Engineering, Faculty of Engineering and Technology

SRM Institute of Science and Technology

Ghaziabad, Uttar Pradesh, India

Email: priyanka6gupta@gmail.com

1. INTRODUCTION

Machine learning (ML) techniques have been used in healthcare system for predicting diseases [1], [2]. Cardiovascular disease (CVD) or heart disease is one of the main reasons for early death even at young age and that too often sudden. As stated by American Health Association (AHA), one out of four deaths are due to CVD. A large section of population is suffering from heart disease and as per World Health Organization (WHO), 17.9 million deaths were due to CVDs in 2019. CVD is a common problem. Various factors are considered major risk factors associated with the occurrence of heart disease like smoking, diabetes, high blood pressure, alcohol, obesity, and level of cholesterol. Predicting heart disease is challenging but if the disease detection is done at an earlier stage and preventive measures are taken at an earlier point of time then it can go a long way in reducing mortality due to heart disease. Earlier detection of heart diseases helps to save the lives of people to a great extent. In the healthcare domain, machine learning techniques plays a major role in prediction of heart disease. Recently, neural network techniques are also being used in this respect. Use of neural networks helps in analyzing large volume of medical data in an efficient manner. Multilayer perceptron is a kind of neural network consisting of layers having different number of neurons in each layer.

Singh *et al.* [3] have proposed the development of prediction system for heart disease. Lutimath *et al.* [4] have utilized Naïve Bayes (NB) classifier and support vector machine (SVM) with radial kernel for prediction of heart disease. SVM with radial kernel was found better for detecting the heart disease. Latha and Jeeva [5] have presented the use of ensemble learning techniques such as majority voting, boosting and bagging

for improving the prediction accuracies of heart disease. In study by Sarkar [6] and Mohan [7] authors have proposed hybrid model for predicting heart disease. The hybrid model has shown good prediction accuracy. Some studies have presented the comparative analysis of different classifiers for predicting the heart disease [8]-[15]. Miao *et al.* [16] have applied an Ensemble learning approach on 4 different datasets. Ayon *et al.* [17] have performed the comparison of various algorithms on two different datasets, stalog and Cleveland heart disease dataset. Deep neural network (DNN) has shown the best results on stalog dataset while SVM has shown the best performance on Cleveland dataset. Fitriyani *et al.* [18] have presented a model for predicting heart disease and compared two datasets, stalog and Cleveland heart datasets. The Cleveland dataset has shown the best prediction accuracy. Escamila *et al.* [19] have proposed the use of chi square test for feature selection and principal component analysis (PCA) for reducing the dimensionality in predicting the heart disease. Rani *et al.* [20] have proposed the hybrid technique which combines genetic algorithm (GA) and recursive feature elimination (RFE) for selecting the features. Various machine learning algorithms were utilized for predicting the heart disease. Among all, Random Forest has shown the best results with an accuracy of 86.6%. Bharti *et al.* [21] have presented the use of machine learning (ML) and deep learning (DL) algorithms for predicting the heart disease. Deep learning has shown the best results with an accuracy of 94.2%. Bhoyar *et al.* [22] have utilized MLP for predicting heart disease and compared UCI and Cardiovascular disease dataset. The best accuracy of 87.30% was achieved with cardiovascular disease dataset. Nahiduzzaman *et al.* [23] have presented MLP and SVM for classifying heart disease into two class and five class. For two-class SVM has shown best accuracy of 92.45% and for five-class MLP has shown the best results with accuracy of 68.86%.

By reviewing the works performed by various authors it has been find out that the performance of the existing systems is comparatively less. So, in this work we presented a model which will try to optimize the performance of the model by hyperparameter tuning. The optimized algorithms were applied on two different datasets and were compared for their performance measures. Moreover, feature importance score was estimated for each feature to delineate features having higher significance. Earlier works on feature importance have used statistical methods such as correlation coefficient, and chi square test for finding significant features. In this study machine learning algorithms were used for the same.

The paper is organized in the following sequence. Datasets used are discussed in section 2. In section 3 proposed methodology with its architectural diagram are discussed. The results of the experiments are discussed in section 4. Lastly in section 5 conclusion is discussed.

2. DATASET USED

2.1. UCI heart disease dataset description

We have used is in this study the Cleveland dataset (UCI, 1990) [24] which was obtained from the (UCI) depository of machine learning (ML) containing the dataset of heart disease which comprises of 4 autonomous databases provided by 4 autonomous medical institution. The Cleveland dataset have 303 instances and 13 attributes. Out of 13 attributes, 5 of these consist of numerical values and 8 of these consists of categorical variables Table 1. The heart disease diagnosis attribute is categorized into either presence or absence of heart disease. Presence is denoted by '1' and absence of disease is denoted by '0'.

Table 1. UCI dataset attributes

S No.	Attributes
1	Age- Patient age is considered in years
2	Sex/Gender: Male is taken as 1 and Female is taken as 0
3	Blood Pressure at Rest (Trestbps)
4	Maximum attained value of Heart Rate (Thalach)
5	Chest pain (Cp)-Chest pain is classified into 4 categories 1. Typical angina 2. Atypical angina 3. Non-anginal pain 4. Asymptomatic
6	Fasting blood sugar If the serum level of fasting blood sugar is more then 120mg/dl it is taken as 1 else, it is taken as 0
7	Serum Cholesterol levels (in mg/dl) (Chol)
8	Resting ECG (Restecg)- . Normal is taken as 0, . ST-T wave abnormality (T-inversion, ST-elevation or depression of more than 0.05 mV) is taken as 1 . Left Ventricular Hypertrophy by Ester's criteria is taken as 2
9	Oldpeak-ST depression induced by exercise in comparison with the state of heart
10	Exercise induced angina (Exang)-If present it is taken as 1 else 0
11	Slope of ST segment at peak exercise (Slope) . Upsloping is taken as 1. Flat as 2. Downsloping as 3
12	Number of Major vessels colored by Fluoroscopy (Ca)- Ranges from 0 to 3
13	Obtained Defect (Thal) It depicts status of heart by three different values . Normal is taken as 3. Fixed defect is taken as 6. Reversible defect is taken as 7

2.2. Framingham heart study dataset description

The dataset used in this study is Framingham heart disease dataset [25] from Kaggle. The dataset has 4,240 instances and 15 attributes which are described in the Table 2. The ten-year coronary heart disease (CHD) signifies the target attribute. It is categorized into two classes ‘0’ denotes the absence of risk of CHD and class ‘1’ denotes the presence of CHD.

Table 2. Framingham dataset attributes

S No.	Attributes	Description
1	Sex	0: male, 1: female
2	Age (Years)	Age of Patient in years
3	Current Smoker	1: If patient is Current Smoker 2: If patient is Non-Smoker
4	CigsPerDay	Number of Cigarettes the person smokes in a day
5	BPMeds	Patient on BP Medication 0: If patient not on BP Medication 1: If patient is on BP Medication
6	Prevalent Stroke	Patient had a previous stroke 0: If patient is not having previous stroke 1: If patient is having a previous stroke
7	Prevalent Hyp	Patient is Hypertensive or not 0: If patient is not Hypertensive 1: If patient is Hypertensive
8	TotChol	Total Cholesterol level
9	SysBP	Systolic Blood Pressure
10	DiaBP	Diastolic Blood Pressure
11	Diabetes	Patient is Diabetic or Non-Diabetic
12	BMI	Body Mass Index
13	Heart Rate	Rate of Heart
14	Glucose	Glucose level
15	Education	Education of person
16	Ten Year CHD	Target-10 Year CHD Risk 0: No Risk of CHD 1: Risk of CHD

3. METHOD

In this work various machine learning (ML) and deep learning algorithms were utilized for predicting the heart disease. Both machine learning and a deep learning (MLP) algorithms were applied on two different datasets. Various machine learning algorithms taken were decision tree (DT), random forest (RF), K nearest neighbors (KNN) and support vector machine (SVM). The algorithms were optimized using Hyperparameter tuning. Random forest (RF) and decision tree (DT) algorithms were also utilized for generating the feature importance score. In deep learning algorithm, Multilayer Perceptron was used. It is a kind of Neural Network which is feed forward in nature consisting of input layer for receiving the input, hidden layer and output layer for displaying the output. The datasets used in this work were UCI heart disease dataset and Framingham heart study dataset.

- UCI heart disease: In UCI dataset, the preprocessing of data was performed. Null and missing values were handled. As the dataset was already balanced so data balancing was not required.
- Framingham dataset: In Framingham dataset also, first the preprocessing of data was done in which the handling of null and missing values was taken care of. The detection of outliers was performed using box plot and outliers were removed. The balancing of dataset was done as the dataset was highly imbalanced.

Both the datasets were divided into training data and testing data. The data used for training was 80% and for testing was 20%. Figure 1 shows the proposed framework for predicting the heart disease.

3.1. Machine learning algorithms used

3.1.1. Decision tree (DT)

This classifier is mostly utilized for solving the categorization problem. It is easier to use. It can be used both for categorization as well as regression. The structure of decision tree classifier is similar to tree where the features of a dataset represent internal nodes, decision rules are represented by branches and decisions are represented by leaf nodes.

3.1.2. Random forest (RF)

It consists of decision trees build by taking different subset of dataset. The final output is predicted based on majority of votes for prediction. If the number of trees is more the accuracy of the model increases and the problem of overfitting also reduces.

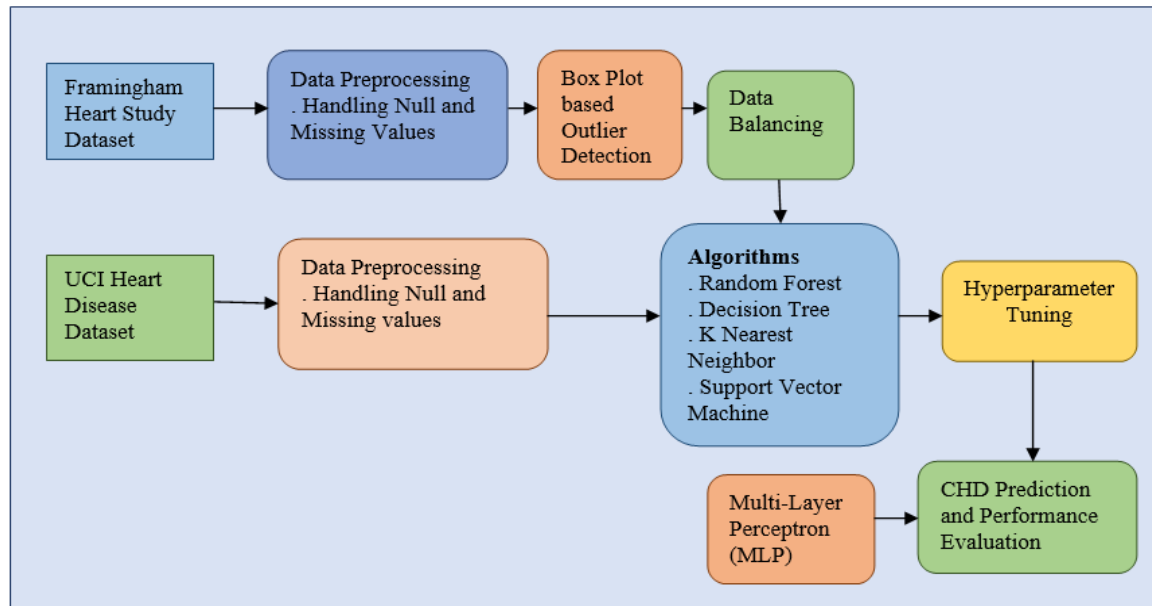


Figure 1. Proposed framework for heart disease prediction

3.1.3. K-nearest neighbors (KNN)

In KNN classifier, all the data available is stored and data point which is new is classified on the similarity measure. It compares the unclassified data with classified data by calculating the distance between data points using Euclidean distance, Manhattan distance. Here we have to select the 'K' number of neighbors.

3.1.4. Support vector machine (SVM)

This classifier can be employed to Regression as well as for Classifying. In SVM a hyperplane is used which is acting as a decision boundary between the classes. Hyperplane should be maximal margin hyperplane. It uses labelled data for training.

3.2. Hyperparameter tuning

Hyperparameter tuning of algorithms was performed. Various parameters for respective algorithms were applied on both the datasets separately. The best performing parameter was taken for each dataset.

For Decision tree classifier, best performance of 83.61% was shown on taking max_features=11 for UCI dataset while for Framingham dataset best accuracy of 92.64% was observed by taking the parameter of max_features=6. The performance of Random Forest was compared by taking different number of estimators, 5, 10, 100, 200, 500 estimators. The best performance was obtained at 200 estimators with UCI heart study dataset giving accuracy of 85.25% and for Framingham dataset, the accuracy of 97.13% was achieved at 500 estimators.

The performance of Support Vector Classifier was compared by considering four different types of kernels, [Kernel = 'linear', 'poly', 'sigmoid', 'rbf']. For UCI dataset the best accuracy of 86.89% was achieved with Kernel = 'linear'. While for Framingham dataset the best results of 69.22% were observed on Kernel = 'rbf'. The performance of K Nearest Neighbors was calculated by considering number of neighbors from 1 to 20. For UCI dataset the best prediction accuracy of 75.4% was observed at n_neighbors = 11. In Framingham dataset the best results of 93.2% was observed on n_neighbors = 1.

3.3. Multi-layer perceptron (MLP)

Multilayer Perceptron is a type of deep learning algorithm. It is a kind of neural network which is feed forward in nature consisting of input layer for getting the input, hidden layer and output layer for giving the output. Each layer consists of different number of neurons. An activation function is used at the hidden layer which transforms input into output. In this work, for MLP hidden layer size of 512*512 and maximum iterations i.e., max_iter=350 were taken for both the datasets.

3.4. Feature importance

Different datasets have different number and types of attributes, although desirable end result is the same i.e., whether the heart disease is present or not. All the attributes do not contribute equally for prediction of heart disease. Significance of some features is more making them more useful in predicting heart disease. Machine learning has been used in analysing feature importance in various areas. The importance of each feature was estimated using machine learning algorithms. The importance score was found for random forest and decision tree algorithms but KNN, SVM and MLP do not produce any feature score. The features were ranked according to their feature scores.

4. RESULTS AND DISCUSSION

In this work, various machine learning (ML) classifiers, decision tree (DT), random forest (RF), K nearest neighbor (KNN), support vector machine (SVM) and a deep learning algorithm (multi-layer perceptron) were applied on two different datasets for prediction of heart disease. These algorithms were compared based on various performance metrics like precision, recall, f1-score and accuracy. Comparison of classifiers in UCI heart disease dataset for performance measures as shown in Figure 2. The results for UCI heart disease dataset are summarized in Table 3 and Figures 2(a) and (b).

Table 3. Performance measures of classifiers for UCI heart disease dataset

Algorithms	Precision	Recall	F1-Score	Accuracy
Decision Tree	0.89	0.78	0.83	83.61
Random Forest	0.85	0.88	0.86	85.25
KNN	0.74	0.81	0.78	75.4
MLP	0.85	0.91	0.88	86.89
SVM	0.88	0.88	0.88	86.89

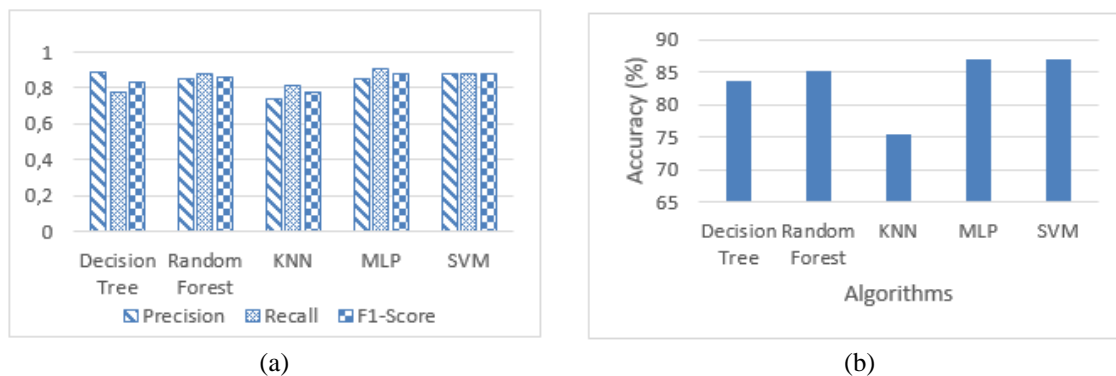


Figure 2. Comparison of classifiers in UCI heart disease dataset for performance measures (a) precision, recall, f1-score and (b) accuracy

It can be observed that the multi-layer perceptron and support vector machine has shown the best performance. Decision tree and random forest have also performed well for predicting the heart disease using UCI heart disease dataset. K nearest neighbor has shown poor performance. Comparison of classifiers in Framingham heart study dataset for performance measures as shown in Figure 3. The results for Framingham heart study dataset are summarized in Table 4 and Figures 3(a) and (b)

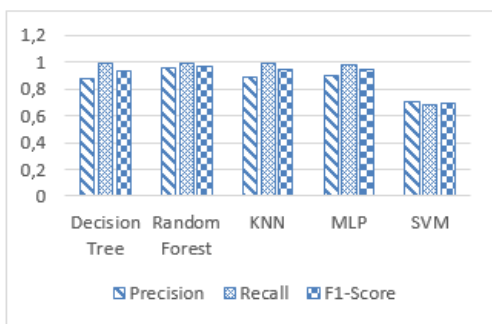
It can be examined that random forest has shown the best prediction accuracy. MLP has shown good performance. In a similar manner decision tree and KNN also performed well for heart disease prediction. support vector machine (SVM) has shown poor performance for Framingham heart study dataset.

It can be observed that the performance of the classifiers has improved on Framingham heart study dataset except SVM. The attributes used in Framingham dataset can be easily obtained at home or near-by clinic whereas in case of UCI heart disease dataset the attributes are difficult to obtain and may require intensive testing. Moreover, the number of records in Framingham dataset is quite large consisting of 4240 records and 15 attributes. But in case of UCI dataset, the dataset consists of only 303 records and 13 attributes which is very low as compared to Framingham dataset. Larger number of records in the dataset not only helps in better training of the model but also increases the prediction accuracy of the model.

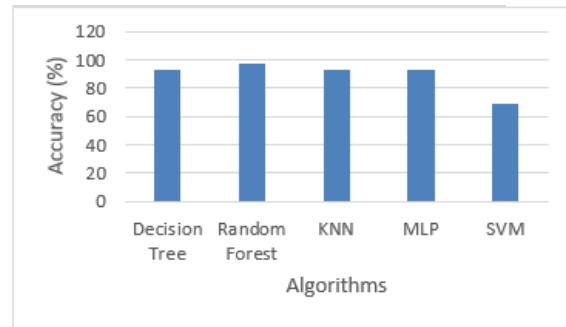
Graphical representation of feature importance for UCI dataset as shown in Figure 4. Table 5 and Figures 4(a) and (b) shows the feature importance and coefficient value for random forest and decision tree algorithms for UCI dataset but MLP, KNN and SVM algorithms do not produce feature importance and coefficient value. Table 6 shows the top five features which are more significant according to feature importance value. Out of five, these four attributes i.e., ca, cp, oldpeak and thalach are common for both the algorithms and are more significant and important for predicting the heart disease.

Table 4. Performance measures of classifiers for Framingham dataset

Algorithms	Precision	Recall	F1-Score	Accuracy
Decision Tree	0.88	0.99	0.93	92.64
Random Forest	0.96	0.99	0.97	97.13
KNN	0.89	0.99	0.94	93.2
MLP	0.90	0.98	0.94	93.45
SVM	0.70	0.68	0.69	69.22

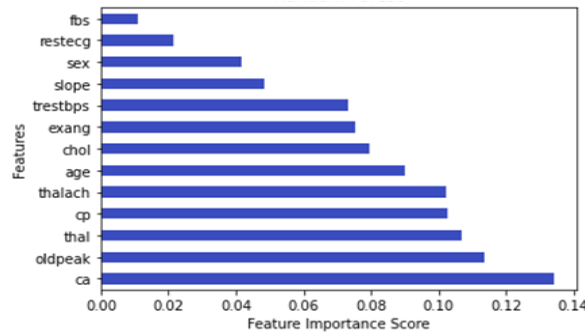


(a)

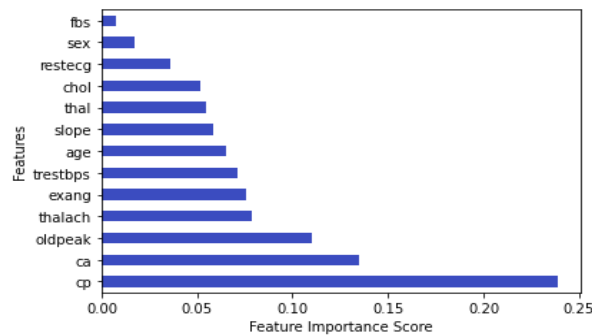


(b)

Figure 3. Comparison of classifiers in Framingham heart study dataset for performance measures: (a) precision, recall, f1-score and (b) accuracy



(a)



(b)

Figure 4. Graphical representation of feature importance for UCI dataset (a) random forest and (b) decision tree

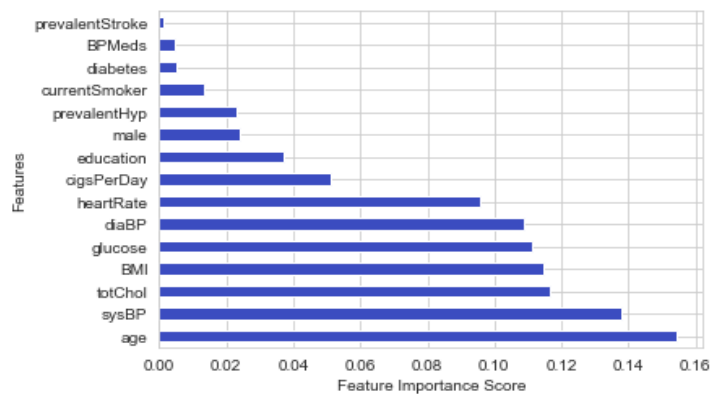
Table 5. Feature importance scores of algorithms for UCI dataset

Feature name	Random Forest	Decision Tree
ca	0.134151	0.135048
oldpeak	0.113460	0.110050
thal	0.106975	0.054416
cp	0.102704	0.238788
thalach	0.102250	0.078316
age	0.090120	0.065282
chol	0.079398	0.051418
exang	0.075407	0.075421
trestbps	0.073286	0.071445
slope	0.048336	0.058499
sex	0.041542	0.017528
restecg	0.021301	0.036235
fbs	0.011070	0.007553

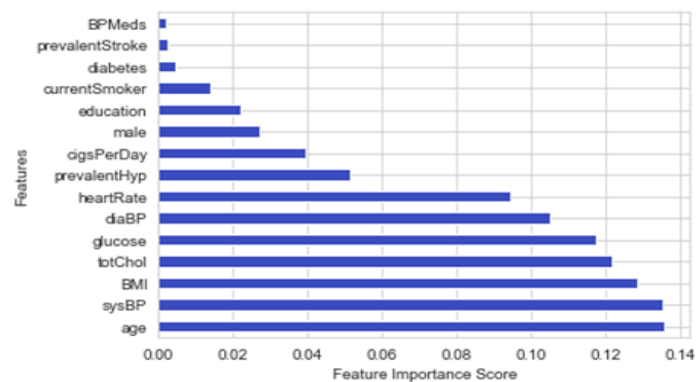
Table 6. Top five features using UCI dataset for heart disease prediction

Feature Ranking	Random Forest	Decision Tree
1 st Feature	ca	cp
2 nd Feature	oldpeak	ca
3 rd Feature	thal	Oldpeak
4 th Feature	cp	thalach
5 th Feature	thalsch	exang

Graphical representation of feature importance for Framingham dataset as shown in Figure 5. Table 7 and Figure 5(a) and (b) shows the feature importance and coefficient value for random forest and decision tree algorithms for Framingham dataset. Table 8 shows the top five features which are more significant according to feature importance value. All the top five attributes are common for both the algorithms. Out of these, age and sysBP are most important, while the rest three i.e., totChol, BMI and glucose are also significant and important factors for predicting the heart disease.



(a)



(b)

Figure 5. Graphical representation of feature importance for Framingham dataset (a) random forest and (b) decision tree

Table 7. Feature importance scores of algorithms for Framingham dataset

Feature name	Random Forest	Decision Tree
age	0.154541	0.135764
sysBP	0.138082	0.135400
totChol	0.116697	0.121400
BMI	0.114776	0.128448
glucose	0.111344	0.117517
diaBP	0.109023	0.104977
heartRate	0.095775	0.094206
cigsPerDay	0.051267	0.039562
education	0.037237	0.022033
male	0.024034	0.027005
prevalentHyp	0.022807	0.051354
currentSmoker	0.013426	0.013742
diabetes	0.004931	0.004443
BPMeds	0.004693	0.001772
prevalentStroke	0.001368	0.002376

Table 8. Top five features using Framingham dataset for heart disease prediction

Feature Ranking	Random Forest	Decision Tree
1 st Feature	age	age
2 nd Feature	sysBP	sysBP
3 rd Feature	totChol	BMI
4 th Feature	BMI	totChol
5 th Feature	glucose	glicose

5. CONCLUSION

Heart or cardiovascular disease is main cause of mortality. In this work, two different datasets were used i.e., UCI heart disease dataset and Framingham heart study dataset for heart disease prediction. It was observed that the performance of the classifiers improved on Framingham dataset as compared to UCI dataset except the support vector classifier. In Framingham dataset, random forest has shown the best performance with an accuracy of 97.13% while for UCI heart disease dataset MLP and SVM has shown the best results and achieved an accuracy of 86.89%. MLP has also shown good performance for Framingham dataset. In addition, feature importance scores were estimated for each feature using machine learning algorithms. The ranking of the features was given based on their scores, finding those features which gives higher predictions.




REFERENCES

- [1] S. S. Reddy, N. Pilli, P. Voosala, and S. R. Chigurupati, "A comparative study to predict breast cancer using machine learning techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 171-180, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.
- [2] K. Yothapakdee, S. Charoenkhum, and T. Boonnuk, "Improving the efficiency of machine learning models for predicting blood glucose levels and diabetes risk," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 555-562, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.
- [3] P. Singh, S. Singh, and G. S. Pandi-Jain, "Effective heart disease prediction system using data mining techniques," *International Journal of Nanomedicine*, vol. 13, pp. 121-124, Mar. 2018, doi: 10.2147/IJN.S124998.
- [4] N. M. Lutimah, C. Chethan, and S. P. Basavraj, "Prediction of heart disease using machine learning," *Int J. of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2S10, pp. 474-477, Sep. 2019, doi:10.35940/ijrte.B1081.0982S1019.
- [5] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, pp. 1-9, 2019, doi: 10.1016/j.imu.2019.100203.
- [6] B. K. Sarkar, "Hybrid model for prediction of heart disease," *Soft Computing*, vol. 24, no. 3, pp. 1903-1925, 2020, doi: 10.1007/s00500-019-04022-2.
- [7] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
- [8] S. Shylaja and R. Muralidharan, "Comparative analysis of various classification and clustering algorithms for heart disease prediction system," *CIIT International Journal of Biometrics and Bioinformatics*, vol. 10, no. 4, pp. 74-77, April 2018.
- [9] R. Kannan and V. Vasanthi, "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease," in *Soft Comp. and Medical Bioinform.*, Springer, Singapore, pp. 63-72, 2019, doi: 10.1007/978-981-13-0059-2_8.
- [10] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.3001149.
- [11] A. Rajdhan, A. Agarwal, M. Sai, and P. G. D. Ravi, "Heart disease prediction using machine learning," *International Journal of Engineering Research*, vol. 9, no. 04, pp. 659-662, May 2020, doi: 10.17577/IJERTV9IS040614.
- [12] U. S. B. "Effective heart disease prediction model through voting technique," *International Journal of Engineering Technology and Management Sciences*, vol. 4, no. 5, pp. 10-13, Sep. 2020, doi: 10.46647/ijetms.2020.v04i05.003.
- [13] F. Tasnim and S.U. Habiba, "A comparative study on heart disease prediction using data mining techniques and feature selection," in *2021 2nd IEEE International Conference on Robotics, Electrical and Signal Processing techniques (ICREST)*, Jan. 2021, pp. 338-341, doi: 10.1109/ICREST51555.2021.9331158.




- [14] A. Ishaq *et al.*, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.
- [15] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, pp. 1-10, 2021, doi: 10.1016/j.compbiomed.2021.104672.
- [16] K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing coronary heart disease using ensemble machine learning," *Int. J. of Advanced Computer Science and Applications*, vol. 7, pp. 30–39, 2016, doi: 10.14569/IJACSA.2016.071004.
- [17] S. I. Ayon, Md. M. Islam, and Md. R. Hossain, "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques," *IETE J. of Res.*, pp. 1–20, Jan. 2020, doi: 10.1080/03772063.2020.1713916.
- [18] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDP: an effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.
- [19] A. K. Garate-Escamila, A. H. E. Hassani, and E. Andrés, "Classification models for heart disease prediction using feature selection and PCA," *Informatics in Medicine Unlocked*, vol. 19, 2020, doi: 10.1016/j.imu.2020.100330.
- [20] P. Rani, R. Kumar, N. M. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *Journal of Reliable Intelligent Environments*, 2021, doi: 10.1007/s40860-021-00133-6.
- [21] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.
- [22] S. Bhojar, N. Wagholikar, K. Bakshi, and S. Chaudhari, "Real-time heart disease prediction system using multilayer perceptron," in *2021 2nd Int. Conf. Emerg. Technol. (INCET)*, May 2021, pp. 1–4, doi: 10.1109/INCET51464.2021.9456389.
- [23] Md. Nahiduzzaman, Md. J. Nayeem, Md. T. Ahmed, and Md. S. U. Zaman, "Prediction of heart disease using multi-layer perceptron neural network and support vector machine," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, Dec. 2019, pp. 1–6, doi: 10.1109/EICT48899.2019.9068755.
- [24] Heart Disease Data Set: UCI repository of machine learning databases having Cleveland Dataset: [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease> -.
- [25] Framingham heart study dataset; Kaggle: [Online]. Available: <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset>.

BIOGRAPHIES OF AUTHORS



Priyanka Gupta    is awarded B. Tech Degree in Computer Science and Engineering from Northern India Engineering College, Delhi affiliated to Guru Gobind Singh Indraprastha University, Delhi in 2010 and M. Tech Degree in CSE from Ideal Institute of Technology, Ghaziabad affiliated to Uttar Pradesh Technical University, Lucknow in 2014. Currently working as an Assistant Professor in Computer Science and Engineering Department in SRM Institute of Science and Technology, Delhi-NCR Campus, Ghaziabad, India. She can be contacted on email: priyanka6gupta@gmail.com.



Dambarudhar Seth    he is Senior Member IEEE is currently working as Professor in Computer Science and Engineering in SRM Institute of Science and Technology, Delhi-NCR Campus. Dr. Seth is PhD from Birla Institute of Technology, Mesra and M. Tech from Indian Institute of Technology, Allahabad, India. He has more than 28 years of experience in teaching and research at both graduate and post graduate level in reputed Institutions such as BIT Mesra, KIIT University Bhubaneswar, and SRM University. His field of research includes Cognitive Radio Network, Software defined Network, Applications of AI/ML in wireless Networks. He can be contacted on email: ddgayatri@gmail.com.