Comparative Analysis and Evaluation of Machine Learning Algorithms for Early Detection of Cardiac Disease

Ritika[1], Rajender Singh[2], Sandeep Dalal[3]

1Deptt of Computer Science and Applications, Maharshi Dayanand University, Rohtak, Haryana India
2Deptt of Computer Science and Applications, Maharshi Dayanand University, Rohtak, Haryana India
3Deptt of Computer Science and Applications, Maharshi Dayanand University, Rohtak, Haryana India

Abstract
The diagnosis of Cardiac disease at the heart attack or stroke stage is very costly with lesser chances of survival. The online and automated healthcare system can be adapted to diagnose cardiac disease at an earlier stage. Various machine learning models are methods integrated within the health care systems to optimize the performance of cardiac disease prediction. In this paper, six supervised learning algorithms are implemented on an authenticated cardiac disease dataset. Bayesian network, decision tree, random forest, SVM, Neural network and Radial Basis Function classifiers are applied in this work on a filtered and preprocessed dataset. The analytical results are obtained in terms of accuracy, precision, recall, TP rate, FP rate and F-score parameters. The analysis results identified that the SVM and Naive Bayes classifiers achieved the most effective results. SVM with gaussian kernel achieved the maximum accuracy rate, TP rate and least FP rate.

Keywords:
Machine Learning; Supervised Learning; Cardiac Disease; Feature Selection; Classifiers

1. Introduction

In the recent years, the Cardiovascular Disease (CVD) has become one of most critical disease with millions of deaths per year. Based on a study, 31% of total deaths are because of heart disease. Any kind of abnormal condition or behavior of blood vessels can cause cardiac disease or attack. This abnormal blood pressure, blood flow can build up plague that narrows up the veins and arteries. It disturbs the blood flow from the heart that increases the risk of heart stroke or failure. Various reasons or symptoms of heart disease include high blood pressure, high cholesterol level, genetic mutations, physical inactivity, poor nutrition and obesity. Once the heart attack or heart failure situation arises, it is very costly and risky for a patient. The chances of survival are also very lesser in such situations. But if the heart disease symptoms are detected at an earlier state, then precautionary treatment can be taken that can increase the survival chances [1].

Manuscript received April 5, 2026
Manuscript revised April 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.4.12

Various machine learning, mathematical and deep learning-based automated models are investigated and implemented in numerous applications to predict heart disease. The effectiveness of these models depends on the number of available or collected features and the data collection methods. Data collection methods are also improved in recent times. The real-time and IoT-based methods are integrated to capture the real situation of a patient. Various health organizations are collecting this kind of information to take decisive results accurately. The machine learning methods use these features or data and extract the pattern that can identify healthy and heart patients. The supervised, semi-supervised and unsupervised learning methods were used by the researchers in different models to correlate the features and generate the patterns. The size and reliability of these patterns are reflected in the accuracy and performance of the disease prediction and classification models [1]. A prototype model for IoT and ML[22][23][24] based heart disease prediction is provided in Figure 1. The figure shows that the functional server is responsible to process the adequate feature collected from patients. These general features such as age, gender, pain type etc. can be collected from patient directly. The actual heart associated features such as heart rate, glucose level, blood pressure can be captured through various IoT devices. The adequate model can be applied on these composite features and the predictive result can be stored in the available dataset[3][25][26][27].

In this paper, an analytical and comparative study is provided on various machine learning methods. The supervised learning methods are implemented on cardiac disease dataset. In this section, the criticality and features of cardiac disease system are described. The standard prototype used that captures the data from different sources and process the model is also defined. In section 2, various machine learning methods and model defined by the earlier researchers are discussed with their associated features and capabilities. In section 3, functional behavior and features of various supervised machine learning methods are discussed. The adaptation of classification algorithm is described for cardiac health care system. In section 4, the description of the dataset and the results obtained for different classifiers are provided. In section 5, the conclusion and future scope of the work are included.

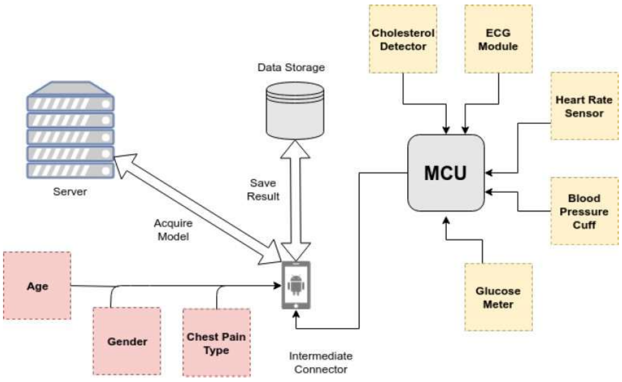

Figure 1: Prototype for Cardiac Disease Prediction [3]

2. Related Works

The severity of cardiac disease attracts researchers to predict the occurrence of heart disease. Various heart disease prediction and classification methods and models were proposed by the researchers to predict the symptoms of cardiac disease accurately. Author used various hybrid method, optimization methods and feature selection methods to improve the capabilities of machine learning models. Some of the work provided by the researchers for machine learning methods and models is provided in this section.

Li et al.[5] proposed a conditional mutual information based feature selection method to optimize the heart disease classification. Author improved the existing Support Vector Machine (SVM) for optimizing the performance cardiac disease. The fuzzy system based ensemble learning model was proposed for accurate diagnosis of heart disease. Fitriyani et al.[6] proposed a heard disease based clinical decision support system to identify the cardiac disease at the earlier stage. Author integrated the density based spatial clustering of application with noise (DBSCAN) clustering method to avoid the outlier. The filtered dataset was processed under hybrid synthetic minority over sampling technique with edited nearest neighbor (SMOTE-ENN) method with XGBoost classifier for heart disease prediction. The model was applied on Statlog and Cleveland datasets and achieved 95,90% and 98.40% accuracy respectively. A fuzzy and genetic algorithm based hybrid model [7] was proposed to optimize the cardiac disease prediction. Author defined the rough sets for effective feature selection and fuzzy rule within the classification stage. An adaptive genetic algorithm was applied on filtered feature set for predicting the heart disease. The method was applied on publicly available UCI heart disease dataset and achieved higher accuracy than existing methods.

Amin et al.[8] provided an analytical study on various feature selection and prediction methods to identify the most effective classification technique. The author applied the model on KNN, Naive Bayes, Decision Tree, SVM, Logistic Regression and Neural network methods. The analysis results identified that the voting based ensemble learning method achieved 87.4% accuracy method. Haq et al.[11] proposed a hybrid machine learning model to predict the cardiac disease accurately. Author implemented the seven machine learning algorithms with feature selection stage to optimize the predictive results. Author implemented the preprocessing, feature selection, machine learning and validation algorithms to evaluate the performance results. The health care system is defined to optimize the diagnosis of cardiac disease. Latha et al.[12] combined weak predictive algorithms and proposed an ensemble learning algorithm to optimize the performance of cardiac disease prediction. Author combined the multiple classifiers and compared it with existing classifiers and ensemble methods. The proposed ensemble classifier improved the accuracy up to 7% over the weak classifiers. Gupta et al.[13] proposed an intelligent machine learning framework by processing the factor analysis of mixed data (FAMD). This method identified the effective features of the dataset and applied different predictive models. The proposed model improved the accuracy and performance over baseline methods. Singh et al.[16] provided a study and analytical work for predicting the performance of various machine learning algorithms. The author analyzed the decision tree, KNN, linear regression, and support vector machine methods for predicting their performances. The author identified the KNN as the most effective method with maximum accuracy.

Rani et al.[19] defined a hybrid feature selection method to minimize the dataset features. The author used the single scalar-based multiple classifiers to optimize the work. This hybrid model achieved 86.6% accuracy. Garate-Escamila et al.[20] used the chi square and principal component analysis-based hybrid filtration method and combined it with a random forest classifier. The proposed method achieved 98.7% accuracy for Cleveland and 99% for Hungarian datasets. The method achieved the performance in comparison with existing methods. Muhammad et al.[21] proposed a computational predictive system for the prediction of cardiac disease. The study was defined on various feature filtration and classification method. The author identified that the proposed model achieved effective results for Chi-square and P-value filtration methods.

3. ML Models/Approaches

Cardiovascular disease can lead to death. It has the highest death rate overall diseases. The early prediction of heart disease can save the life of an individual. Various predictive methods and classification models were investigated by the researchers. In this section, some of the most effective machine learning models are discussed. These models were used by researchers for the prediction of various medical diseases [18]. Some researchers have used these machine-learning models for predicting the chances of cardiac disease. In this paper, we have defined a cardiac disease prediction model which is further integrated with multiple machine learning models. Figure 2 shows the model used in this work for implementing and analyzing different machine learning methods. In this section, the detailed description of the model and functional description of each of the machine learning methods is provided.

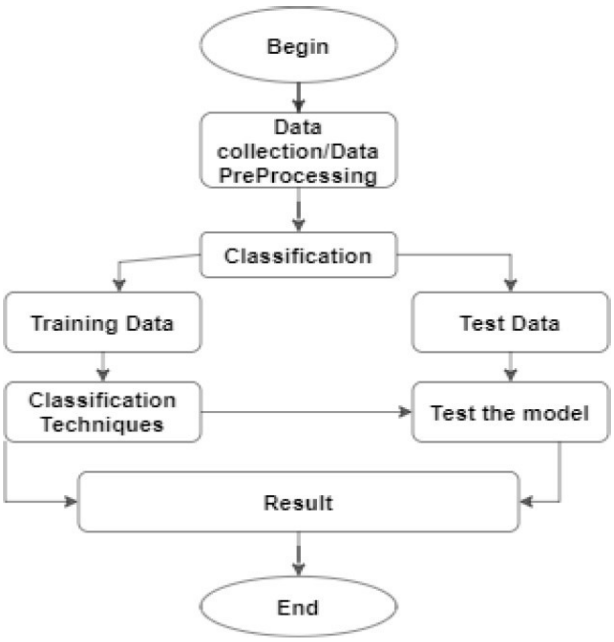

Figure 2: Prototype for ML Model on Cardiac Disease

Figure 2 shows the cardiac disease prediction model that is used in this research. All the machine learning algorithms defined in this paper are implemented in the classification stage of this model. In the first step of this model, the cardiac disease dataset is collected. In this work, the dataset is collected from authenticated external source and its description is provided in section 4. Once the dataset is obtained and the classification model is applied over it. According to this model, the dataset is divided into two subsets called training and testing set. While performing the implementation and experimentation, 90% of data is taken as training set and 10% is considered as test set. The classification algorithms applied in this work are Decision Tree, Neural network, Bayesian network, SVM, and RBF networks. The model is first applied to the training set and generates the classification. Later, the test set is applied to this model to predict cardiac disease. The analysis results are obtained on the parameters described in section 4.2. In this section, the functional description of each of the classification methods used in this paper are provided.

Decision Tree

Decision tree is an effective tree based classifier that performs a level specific analysis on individual feature to identify the instance class. It is a recursive process that analyzes each feature to take the decision over the information space. The node level split is performed over the dataset with specification of certain function. On each node, the attribute value based decision is performed and this process is repeated on each feature till the leaf node not occurs. Each of the leaf node is represented by a particular class. There are multiple forms of decision tree with specific integrated method [14][15]. In this work, we have applied the decision tree with three integrated steps.

Compute the Gini Index for each individual feature

Apply the node based decision to split the dataset. Identify the optimal value for this split ion.

Apply this process recursively to generate the complete decision tree. The process is pleated till the terminal node not occur.

Represent each of the leaf node with specific class.

Random Forest

Random forest is an integrated ensemble learning method that has provided good results for many classification applications. It is the combination of multiple decision trees and generate a forecast to perform effective prediction. It uses a likelihood estimation based analysis method to identify the prediction class. It uses a voting method on each decision tree to identify the respective instance class[14][15]. The functional process of random forest algorithm is shown in Figure 3[17]. The steps followed by random forest algorithm are given below:

Take a sub-sample set from training set.

Create a decision tree on this sub-sample set

Identify the interior node for each sub tree and identify the best split for that part

Evaluate the values for each tree separately and define the expected classes for each leaf

Apply majority voting method while combining the trees and take the final decision of the relative instance class.

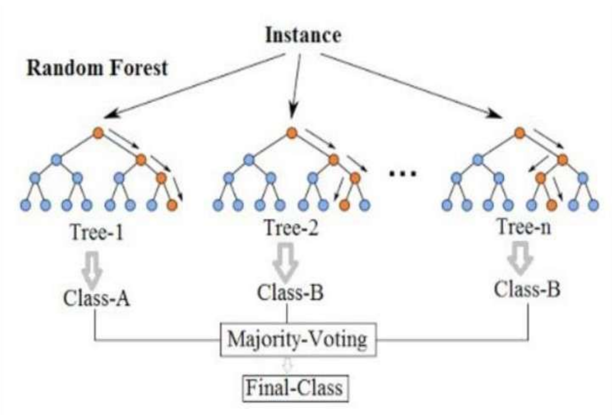

Figure 3: Functional Process of Random Forest

Naive Bayes is a probabilistic learning model that uses the Bayes rule on independent features. The subsequent probability computation is performed on different features and the class is allocated to the highest probability value. The gaussian function is implied within the method with prior probability for computing the data class. The conditional probability is computed for dependent features[4]. The gaussian function based priority probability P(Xf) i used for feature class computation is provided in equation (1) and (2). It is one of simplest and faster method that can predict best results when assumption and independence of features are known[9][10]. Algorithm-1 defines the algorithmic process followed by Naïve bayes classifier.

$$P(Xf1, Xf2, \ldots Xfn|c) = \prod^{n} P(Xf_i|c) \quad (1)$$

$$P(Xf_i|c) = \frac{(c_i|Xf)\ (Xf)}{} P(c_i) c \ \epsilon\ \{cardic\ patient, ealthy\} \quad (2)$$

Algorithm-1

1. Define the dataset D with n features called X(x1,x2....xn) and m is number of classes C(c1,c2
 cm)
2. Define the Bayes theorem for mapping the features to the

classes $P(X|C_i) = \frac{P(X|C_i) * P(C_i)}{P(X)}$

3. Compute probabilistic estimation for each feature P(X) and respective to each class $P(X|C_i)$ such that $P(C_i)$ is maximized
4. Compute the class driven conditional independence

$$P(X|C_i) = P(x1|C_i) * P(x2|C_2) \ldots P(xm|C_i)$$

5. $P(X|C_i)\ P(C_i)$ is evaluated for obtaining the class of X and provide it as the final identified class.

Support Vector Machine (SVM) is an effective classifier that establishes the best relation between the features and the relative class. The method is effective to minimize the classification error. Sequential Minimal Optimization is the algorithm used to identify the optimum class for a particular feature set. The effectiveness of this algorithm depends on the kernel used in the algorithm. Two of the most common kernel functions we have compared in this work are Linear and Gaussian Kernels. Figure 4 shows the graphs for Gaussian and Linear Kernels. As shown in figure, a Linear Kernel provides a straight-line division to separate the classes. Whereas, the gaussian kernel can perform a Gaussian kernel-based interpretation to separate the classes. It uses a curved or non-planer separation as shown in Figure 4. Equation (3) shows the linear kernel and equation (4) shows the gaussian kernel specification. Linear kernel is basically used when features are more than available samples and if the samples are lesser than Gaussian kernel is used. The dimension of data affects the selection of kernel[2].

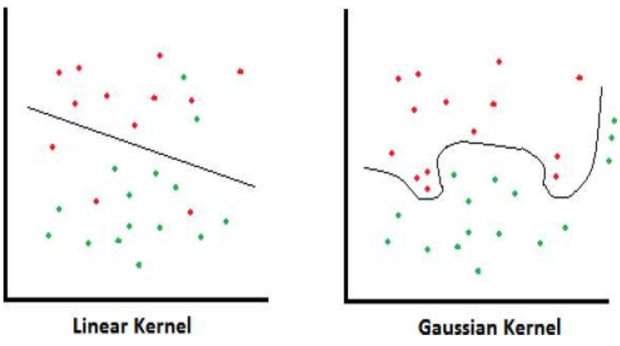

Figure 4: SVM Kernels

Neural network uses the brain's neuron cell structure to solve a problem. It is an intelligent machine learning algorithm that uses the functionality similar to a human brain. It is a layered architecture that is provided in Figure 5. This architecture consist of three layers called Input, Hidden and output layers. Neural network is a non-linear and statistical architecture that can solve complex problems. In first layer of this layer, the featured input is provided. In the hidden layer, the weights are applied on the input to map the input to specific class. In this stage, the feature driven weights are applied to improve the performance and accuracy of the model. Once the data is processed over the model, the particular disease class is obtained as the final output.

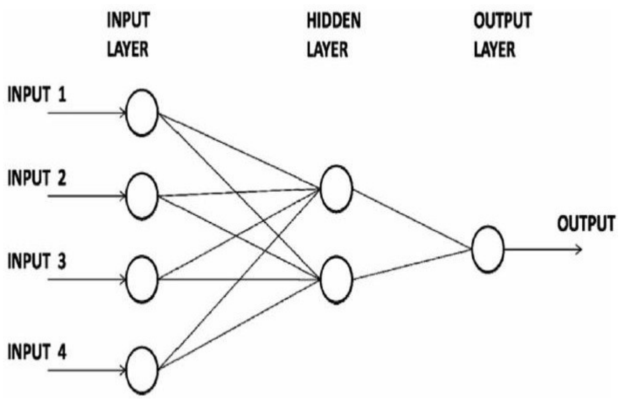

Figure 5: Neural Network Model

The effectiveness of the model depends on the behavior of the neurons and the modulation of these neurons can be improved within the hidden layer. The hidden layer is the actual processing layer that can identify the pattern from the input and perform the computation over it. A neural network can have multiple hidden layers. In this work, the hidden layers are processed with back propagation and it is called Multi-layer perception. After processing the pattern and performing the computations, the particular disease class is obtained by the model and presented as output through the output layer. Another effective element of this classifier is the activation function. linear, sigmoid, logistic, etc. are different activation function that controls the flow of this classifier. In this work, we have used the sigmoid activation function for optimizing the result. The implementation and comparative results of this classifier are provided in the next section.

Radial Basis Function

The radial basis function is a neural network that performs the interpolation on data points within the multidimensional space. The functional link in the network form is provided in Figure 6. It defines a regulation with a basis function. The associated stabilizer is defined for achieving radial symmetry. A multivariate continuous function is defined with domain specific modeling to achieve higher accuracy. The regularization function is defined to minimize the error in the system. It is a three-layer network in which the radial basis function is defined in the hidden layer. The nonlinear activation function is defined to map the input to the output with better accuracy.

It is an optimized function in which the activation function is applied at the hidden layer. The activation function of the classifier is defined as $\emptyset_0 = 1$. In this optimization function the main objective is tominimize the mean square error (MSE). The weights are adjusted in the hidden layer for generating the optimized results. The radial function is here defined with Gaussian function. The output function is given as

4. Results

In this paper, the defined standard model for cardiac disease prediction is defined and experimented. The experimental results are provided and discussed in this section.

Dataset

The experiments and implements of the proposed model and classification methods are implemented to an authenticated dataset called Statlog Heart disease dataset [28]. This dataset is taken from UCI repository. The dataset contains 14 features including six numeric and eight categorical features. The dataset has two main classescalled healthy and cardiac disease. The analysis measures, experimentation details and results are provided in thissection.

Performance Measures

The performance measures are used to identify the effectiveness and reliability of each classification method. Thekey measures used for identification of prediction class are True Positive (TP) defines correctly identified, False Positive (FP) defines correctly rejected, True Negative (TN) defined Incorrectly Identified and False Negative (FN) defines incorrectly rejected. Other derived measures to analyze the performance of the classifiers are given below:

Accuracy

Accuracy defines the ratio of correctly classified instances over the available instances. The accuracy measure isprovided in equation (6).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Precision is the ratio of True positive and total positive instances identified over the sample space. Equation (7) shows the computation of precision measure.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

Recall rate defines the ratio of actual positives identified for the test set. Equation (8) shows the computation ofrecall measure.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

This measure is defined as the harmonic mean of precision and recall values. F-score can be computed usingequation (9)

$$\text{F-Score} = 2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall} \quad (9)$$

The comparative evaluation on various machine learning methods for cardiac attack prediction is provided in this section. The analytical analysis parameters are already described in section 4.2. The experimentation is conductedby using 10 fold method. Figure 7 shows the accuracy based analysis results. The figure shows that the accuracyobtained for decision tree and random forest methods are not effective with lesser accuracy rate of 77.78% and 76.54%. Naïve Bayes and SVM methods achieved the most effective results with 86.42% accuracy. The analysisresults show that the SVM and Bayes network can predict the cardiac disease more accurately. The neural and RBF neural network based methods also provided the significant results with accuracy higher than 80%.

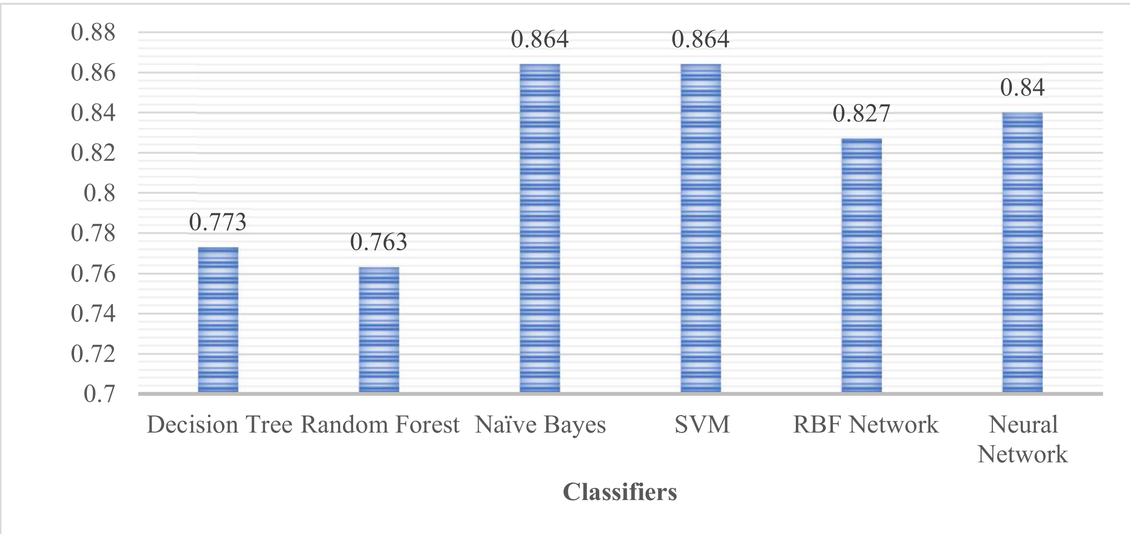

Figure 6: F-Score Measure Analysis

Another effective accuracy based analysis measure is F-Score. The analysis results based on F-score parameter is provided in Figure 6. The figure shows that the F-score analysis results for decision tree and random forest methods least effective with 0.773 and 0.763. Naïve Bayes and SVM methods achieved the most effective resultswith least lease F-score of 0.864. The analysis results show that the SVM and Bayes network can predict the cardiac disease with most adaptive and effective F-Score rate. The neural and RBF neural network based methodsalso provided the effective F score rate.

5. Conclusion

The diagnosis of cardiac disease at early stage can reduce the chances of death. Various machine learning methodsand models are available to predict the cardiac disease. In this paper, a standard model of cardiac disease predictionis defined. Various classifiers are integrated in this model to predict the cardiac disease accurately. The classifiers used in this research work are Bayesian network, decision tree, neural network, radial basis network and SVM. The analysis results shows that the SVM and Bayesian network classification based machine learning model achieved the maximum accuracy rate of 86.42. The method also achieved effective results under true positive, true negative, precision and recalls measures.

References

[1] Princy, R. Jane Preetha, Saravanan Parthasarathy, P. Subha Hency Jose, Arun Raj Lakshminarayanan, and Selvaprabu Jeganathan. "Prediction of cardiac disease using supervised machine learning algorithms." In 2020 4th international conference on intelligent computing and control systems (ICICCS), pp. 570-575. IEEE, 2020.

[2] Yadav, Samir S., Shivajirao M. Jadhav, Rohit G. Bonde, and Shivani T. Chaudhari. "Automated cardiacdisease diagnosis using support vector machine." In 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), pp. 56-61. IEEE, 2020.

[3] Tabassum, Shaira, Md Ittahad Uz Zaman, Md Sabbir Ullah, Ashikur Rahaman, Samsun Nahar, and AKM Muzahidul Islam. "The cardiac disease predictor: IoT and ML driven healthcare system." In 2019 4th International Conference on Electrical Information and Communication Technology (EICT), pp. 1-6. IEEE, 2019.

[4] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart diseaseprediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.

[5] Li, Jian Ping, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. "Heart disease identification method using machine learning classification in e-healthcare." IEEE Access 8 (2020): 107562-107582.

[6] Fitriyani, Norma Latif, Muhammad Syafrudin, Ganjar Alfian, and Jongtae Rhee. "HDPM: an effective heart disease prediction model for a clinical decision support system." IEEE Access 8 (2020): 133034- 133050.

[7] Reddy, G. Thippa, M. Reddy, Kuruva Lakshmanna, Dharmendra Singh Rajput, Rajesh Kaluri, and Gautam Srivastava. "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis." Evolutionary Intelligence 13, no. 2 (2020): 185-196.

[8] Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan,Identification of significant features and data mining techniques in predicting heart disease,Telematics and Informatics,Volume 36,2019,Pages 82-93

[9] Rajdhan, Apurb, Avi Agarwal, Milan Sai, Dundigalla Ravi, and Poonam Ghuli. "Heart disease predictionusing machine learning." International Journal of Research and Technology 9, no. 04 (2020): 659-662.

[10] Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naives Bayesian." In 2019 3rd International conference on trends in electronicsand informatics (ICOEI), pp. 292-297. IEEE, 2019.

[11] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms." Mobile Information Systems 2018 (2018).

[12] C. Beulah Christalin Latha, S. Carolin Jeeva, Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, Informatics in Medicine Unlocked, Volume 16, 2019, 100203

[13] Gupta, Ankur, Rahul Kumar, Harkirat Singh Arora, and Balasubramanian Raman. "MIFH: A machine intelligence framework for heart disease diagnosis." IEEE access 8 (2019): 14659-14674.

[14] Pencina, Michael J., Ann Marie Navar, Daniel Wojdyla, Robert J. Sanchez, Irfan Khan, Joseph Elassal,Ralph B. D'Agostino Sr, Eric D. Peterson, and Allan D. Sniderman. "Quantifying importance of major risk factors for coronary heart disease." Circulation 139, no. 13 (2019): 1603-1611.

[15] Ayon, Safial Islam, Md Milon Islam, and Md Rahat Hossain. "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques." IETE Journal of Research 68, no. 4 (2022): 2488-2507.

[16] Singh, Archana, and Rakesh Kumar. "Heart disease prediction using machine learning algorithms." In 2020 international conference on electrical and electronics engineering (ICE3), pp. 452-457. IEEE, 2020.

[17] Ramalingam, V. V., Ayantan Dandapath, and M. Karthik Raja. "Heart disease prediction using machine learning techniques: a survey." International Journal of Engineering & Technology 7, no. 2.8 (2018): 684-687.

[18] Katarya, Rahul, and Sunit Kumar Meena. "Machine learning techniques for heart disease prediction: a comparative study and analysis." Health and Technology 11, no. 1 (2021): 87-97.
[19] Rani, Pooja, Rajneesh Kumar, Nada MO Ahmed, and Anurag Jain. "A decision support system for heartdisease prediction based upon machine learning." Journal of Reliable Intelligent Environments 7, no. 3 (2021): 263-275.
[20] Gárate-Escamila, Anna Karen, Amir Hajjam El Hassani, and Emmanuel Andrès. "Classification models for heart disease prediction using feature selection and PCA." Informatics in Medicine Unlocked 19 (2020): 100330.
[21] Muhammad, Yar, Muhammad Tahir, Maqsood Hayat, and Kil To Chong. "Early and accurate detectionand diagnosis of heart disease using intelligent computational model." Scientific reports 10, no. 1 (2020):1-17.
[22] Buettner, Ricardo, and Marc Schunter. "Efficient machine learning based detection of heart disease." In2019 IEEE international conference on E-health networking, application & services (HealthCom), pp. 1-6. IEEE, 2019.
[23] Jindal, Harshit, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath. "Heart disease prediction using machine learning algorithms." In IOP conference series: materials science and engineering, vol. 1022, no. 1, p. 012072. IOP Publishing, 2021.
[24] Krishnan, Santhana, and S. Geetha. "Prediction of Heart Disease Using Machine Learning Algorithms."In 2019 1st international conference on innovations in information and communication technology (ICIICT), pp. 1-5. IEEE, 2019.
[25] Ahmed, Hager, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A. Ali. "Heart disease identification from patients' social posts, machine learning solution on Spark." Future Generation Computer Systems 111 (2020): 714-722.
[26] Gavhane, Aditi, Gouthami Kokkula, Isha Pandya, and Kailas Devadkar. "Prediction of heart disease using machine learning." In 2018 second international conference on electronics, communication and aerospace technology (ICECA), pp. 1275-1278. IEEE, 2018.
[27] Tougui, Ilias, Abdelilah Jilbab, and Jamal El Mhamdi. "Heart disease classification using data mining tools and machine learning techniques." Health and Technology 10, no. 5 (2020): 1137-1144.
[28] https://archive.ics.uci.edu/ml/datasets/statlog+(heart)