IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

Detection of Heart Disease using Machine Learning

Mrs. Priti Sudhakar Gund Pujari[1], Miss. Prajakta Avhad[2], Miss. Monika Rathod[3], Miss. Prachi Zine[4], Mr. Rohit Rathod[5]
Assistant Prof. Electronic & Telecommunication Engineering Department[1]
Students, Computer Engineering Department[2]
Student, Information & Technology Engineering Department[3]
Student, AIDS Engineering Department[4]
Student, Electrical Engineering Department[5]
Adsul's Technical Campus, Ahilyanagar, India

Abstract: *Heart disease is one of the most known and deadly diseases in the world, and many people lose their lives from this disease every year. Early detection of this disease is vital to save people's lives. Machine Learning (ML), an artificial intelligence technology, is one of the most convenient, fastest, and low-cost ways to detect disease. In this study, we aim to obtain an ML model that can predict heart disease with the highest possible performance using the Cleveland heart disease dataset. The features in the dataset used to train the model and the selection of the ML algorithm have a significant impact on the performance of the model. To avoid overfitting (due to the curse of dimensionality) due to the large number of features in the Cleveland dataset, the dataset was reduced to a lower dimensional subspace using the Jellyfish optimization algorithm. The Jellyfish algorithm has a high convergence speed and is flexible to find the best features. The models obtained by training the feature-selected dataset with different ML algorithms were tested, and their performances were compared. The highest performance was obtained for the SVM classifier model trained on the dataset with the Jellyfish algorithm, with Sensitivity, Specificity, Accuracy, and Area Under Curve of 98.56%, 98.37%, 98.47%, and 94.48%, respectively. The results show that the combination of the Jellyfish optimization algorithm and SVM classifier has the highest performance for use in heart disease prediction..*

Keywords: Heart Disease Diagnosis, Feature Selection, Jellyfish Optimization, Machine Learning, SVM.

I. INTRODUCTION

A smart According to the World Health Organization, despite significant advances in diagnosis and treatment, mortality from heart disease remains the leading cause of death worldwide, accounting for about one-third of annual deaths [1]. "Heart disease" is a general term used to describe a group of heart conditions and diseases, including Coronary Artery Disease, Arrhythmia, Heart Valve Disease, and Heart Failure, which cause the heart not to pump blood healthily.

The most common type of heart disease is Coronary Artery Disease. The disease is a medical condition in which the coronary arteries that supply blood to the heart muscle become narrowed or blocked due to plaque build-up on their inner walls. This can lead to serious complications such as a heart attack, heart failure, and arrhythmias, as it reduces blood flow to the heart muscle. In some cases, procedures such as angioplasty or bypass surgery may be necessary to improve blood flow to the heart.

The second common heart disease is Arrhythmia. Arrhythmia is caused by disturbances in the normal electrical activity of the heart. The normal beating rhythm of the heart is disrupted because the electrical impulses in the heart responsible for synchronizing the heartbeat are not working properly. As a result, the heartbeat may be faster, slower, or more irregular than normal [2,3]. Millions of people worldwide are affected by Arrhythmia. Symptoms can include a fast or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

irregular heartbeat, shortness of breath, dizziness or fainting, chest pain or discomfort, fatigue, and weakness. There are many different types of arrhythmias, and some types of arrhythmias are harmless, while others can be life-threatening. While many people may experience occasional episodes of mild arrhythmia in their lives, some people may struggle with more serious types of arrhythmias. For example, a type of Arrhythmia known as Atrial Fibrillation can occur in about 10% of adults over the age of 60 and can increase the risk of stroke. On the other hand, a serious type of Arrhythmia known as Ventricular Fibrillation is considered a cause of heart attacks and can be fatal. Some types of arrhythmias can be inherited, while others can be caused by lifestyle factors or other heart diseases. In most early-diagnosed cases, arrhythmias can be treated. Patients with these disorders are much less likely to die suddenly if they receive prompt, thorough diagnosis and medical care [4,5].

The main reasons for the significant increase in heart disease in recent years are people's lifestyle, lack of exercise, and consumption of various processed foods. Heart disease in its advanced stages can cause heart attacks and endanger the lives of patients, so it is necessary to detect the disease quickly and in its early stages with intelligent and therapeutic methods. One of the major challenges in the diagnosis of heart disease is the reluctance of patients to participate in clinical trials. On the other hand, the cost of these trials is high, and they take a lot of time, which is why they receive little attention. In contrast to clinical methods for diagnosing heart disease, some methods can be used to analyze the pattern of the disease by analyzing information from patients and healthy people [6].

In recent years, applications of artificial intelligence technology, especially Machine Learning (ML), in the field of auxiliary diagnosis have developed rapidly, and efficient progress has been made in automatic detection applications [7,8,9,10]. The advantage of ML methods is that they can diagnose diseases, such as heart disease, with low-cost and reasonable accuracy [11]. ML techniques for diagnosing heart disease do not require multiple clinical trials, most of which are invasive, and a set of information and features can help to diagnose the disease with high accuracy. It should be noted that although ML technology has made advances in the automatic diagnosis of heart disease, the approval of doctors is still a necessary link in diagnosis and treatment. It is also clear that ML-based disease diagnosis offers an opportunity to increase doctors' work efficiency and generate economic benefits. In the age of big data, with ever-expanding datasets and the development of new ML algorithms, it is expected that ML applications will undoubtedly have a major impact on automated heart disease prediction [12,13,14,15,16]. In the literature, there are research papers that try to predict heart disease with different datasets and different types of ML algorithms.

Problem Statement:

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyse the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

Objective of Project:

1. Data collection from different sources and pre-processing
2. To develop machine learning model to predict future possibility of heart disease by implementing Logistic Regression, KNN, XGboost, Decision Tree
3. To determine significant risk factors based on medical data set which may lead to heart disease.
4. To analyse feature selection methods and understand their working principle.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

II. LITERATURE SURVEY

[1] Kavitha Dubey A. K. et al. examined the performance of ML models such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), SVM with grid search (SVMG), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) for heart disease classification. Cleveland and Statlog datasets from the UCI Machine Learning repository were used for training and testing. The experimental results show that LR and SVM classifier models perform better on the Cleveland dataset with 89% accuracy, while LR performs better on the Statlog dataset with 93% accuracy [17].

Karthick K. et al. used SVM, Gaussian Naive Bayes (GNB), LR, LightGBM, XGBoost, and RF algorithms to build an ML model for heart disease risk prediction. In this study, the authors applied the Chi-square statistical test to select the best features from the Cleveland heart disease dataset. After feature selection, the RF classifier model obtained the highest classification accuracy rate of 88.5% [18].

In Ahmad G. N. et al.'s study, Cleveland, Hungarian, Switzerland, Statlog, and Long Beach VA datasets were combined to obtain a larger dataset compared to existing heart disease datasets. They compared the performances of LR, KNN, SVM, Nu-Support Vector Classifier (Nu-SVC), DT, RF, NB, ANN, AdaBoost, Gradient Boosting (GB), Linear Discriminants Analysis (LDA) and Quadratic Discriminant Analysis (QDA), algorithms for heart disease classification. In this study, the authors claimed that the best classification accuracy of 100% was achieved with the RF algorithm [24].

Veisi H. et al. developed various ML models such as DT, RF, SVM, XGBoost, and Multilayer Perceptron (MLP) using the Cleveland heart disease dataset to predict heart disease. Various preprocessing (outlier detection, normalization, etc.) and feature selection processes were applied to the dataset. Among the ML models evaluated, the highest accuracy of 94.6% was achieved using the MLP [19].

Sarra R. R. et al. proposed a new classification model based on SVM for better prediction of heart disease using the Cleveland and Statlog datasets from the UCI Machine Learning repository. The $\chi 2$ statistical optimal feature selection method was used to improve the prediction accuracy of the model. The performance of the proposed model is evaluated against traditional classifier models using various performance metrics, and the results showed that the accuracy improved from 85.29% to 89.7% by applying the proposed model [20].

Malavika G. et al. investigated the use of ML algorithms to predict heart disease. The heart disease dataset from the UCI repository was used for this study. They used various ML algorithms, including LR, KNN, SVM, NB, DT, and RF, to predict heart disease, and their performances were compared. The results showed that RF (91.80%) had the highest accuracy in predicting heart disease, followed by NB (88.52%) and SVM (88.52%). The authors concluded that ML algorithms could be a useful tool in predicting heart disease and could potentially help doctors diagnose and treat patients more accurately [21].

Sahoo G. K. et al. compared the performance of LR, KNN, SVM, NB, DT, RF, and XG Boost Machine Learning models for predicting heart disease. The Cleveland heart disease dataset from the UCI ML repository was used to train the models. Comparing the results of the tested ML algorithms, the RF algorithm performed the best, with a classification accuracy of 90.16% [22].

Ijaz Bo Jin, Chao Che et al. (2018) proposed a "Predicting the Risk of Heart Failure with EHR Sequential Data Modelling" model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analysing the results, we tend to reveal the importance of respecting the sequential nature of clinical records [1].

Aakash Chauhan et al. (2018) presented "Heart Disease Prediction using Evolutionary Rule Learning". This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. This will facilitate (help) in decreasing the number of services and shown that overwhelming majority of the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

rules helps within the best prediction of coronary sickness [2]. Ashir Javeed, Shijie Zhou et al. (2017) designed "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection". This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. Two forms of experiments are used for cardiovascular disease prediction. In the first form, only random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed. This methodology is efficient and less complex than conventional random forest model. Comparing to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection.

III. METHODOLOGY

A. System Architecture

The system architecture gives an overview of the working of the system. The working of this system is shown below:

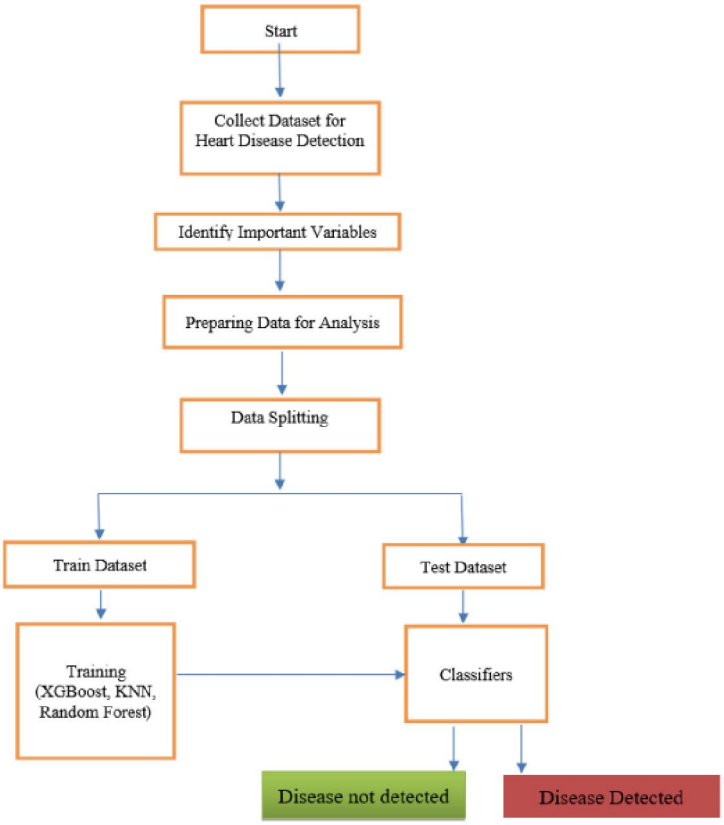

Fig. 1. Proposed Methodology

B. Dataset Details:

1. Age: age of the patient [years]
2. 2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 18, April 2026 Impact Factor: 8.2

4. RestingBP: resting blood pressure [mm Hg]
5. Cholesterol: serum cholesterol [mm/dl]
6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
7. RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
9. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
11. ST_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, down: downsloping]
12. HeartDisease: output class [1: heart disease, 0: Normal]

C. MACHINE LEARNING:

In machine learning, classification refers to a predictive modelling problem where a class label is predicted for a given example of input data

SUPERVISED MACHINE LEARNING

As its name suggests, Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset. The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y).

Categories of Supervised Machine Learning: Supervised machine learning can be classified into two types of problems, which are given below:

a) Classification

Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "Yes" or No, Male or Female, Red or Blue, etc. The classification algorithms predict the categories present in the dataset.

b) Regression

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc. Some popular Regression algorithms are given below:

- Simple Linear Regression Algorithm
- Multivariate Regression Algorithm
- Decision Tree Algorithm
- Lasso Regression

IV. IMPLEMENTATION

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre-processed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1. Data Collection
2. Data Pre-Processing
3. Feature Selection
4. Model Selection

It is the process to select one final algorithm for concerned purpose. It is decided by observing the accuracy by applying multiple algorithms. We can use logistic regression, XGBoost, KNN, random forest, etc. The final accuracy depends of the type of model we select. While selecting the algorithm, we have to compare the accuracies. Comparative analysis is performed among algorithms and the algorithm that gives the highest accuracy is used for heart disease prediction

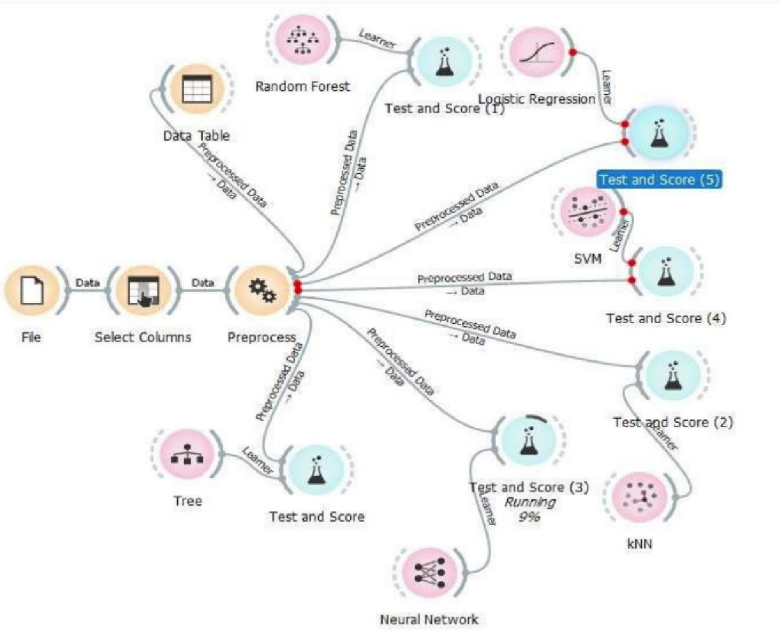

Fig. 2. Connection of Widgets in Orange

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

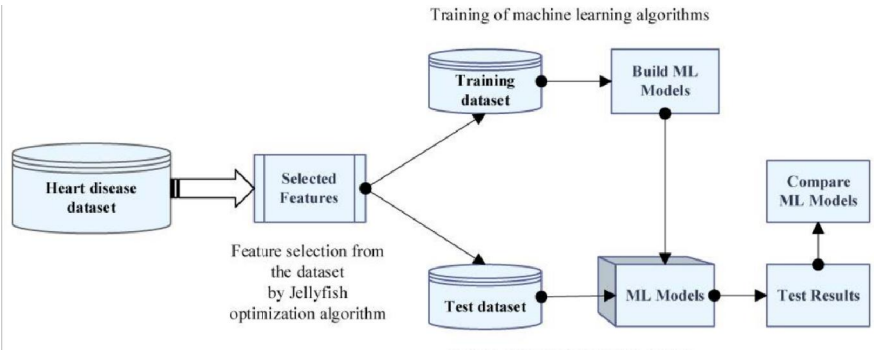

Fig. 3. Flowchart of Proposed Approached of Heart Disease Pediction

V. CONCLUSION

Cardiovascular disease (CVD) is one of the leading causes of deaths happening worldwide, making early detection and intervention crucial for improving patient outcomes. To address this need, a machine learning technique was used to develop a model using patient medical history data to predict the probability of fatal heart disease. The dataset includes variables such as chest pain, sugar levels, and blood pressure, which are important indicators of heart health. These classification algorithms - XGBoost, Random Forest Classifier, and KNN - were utilized to develop the model, which achieved an accuracy rate of over 95%. The accuracy of the model was further improved by increasing the size of the dataset, enabling the identification of more subtle patterns and risk factors. The application of machine learning techniques in medical diagnosis has several benefits, including increased speed and accuracy of diagnoses, reduced costs, and improved patient outcomes. By analysing large amounts of data and identifying complex patterns, machine learning algorithms can provide valuable insights into patient health that may not be immediately apparent to human clinicians. Compared to previous models, the accuracy of the developed model represents a significant improvement, with an accuracy rate of 98%. The XGBoost algorithm demonstrated the highest accuracy of 96% among the three algorithms used, indicating its effectiveness in predicting heart disease. The dataset used in this project indicates that 44% of individuals suffer from heart disease, highlighting the importance of early detection and intervention. The developed model offers a reliable and efficient method for identifying individuals who are at risk of heart disease, potentially benefiting both patients and healthcare providers.

FUTURE SCOPE

While the field of Educational Data Mining and Learning Analytics has already made great strides in improving education outcomes, there is still a lot of potential for further advancements. One area where this could be particularly impactful is in the detection and prevention of heart disease in students. Heart disease is a leading cause of death worldwide, and early detection is key to successful treatment. By analysing data from student health records, as well as their physical activity levels and other relevant metrics, we can identify those who may be at risk for developing heart disease. Machine learning algorithms can be trained to recognize patterns and risk factors, and can provide early warnings to health professionals. Furthermore, with the rise of wearable technology and other health monitoring devices, we can collect even more data on students' health and activity levels. This data can be integrated with educational data, such as quiz scores and attendance records, to provide a more comprehensive picture of each student's overall health and wellbeing. By using advanced analytics techniques to detect early warning signs of heart disease, educational institutions can play a vital role in promoting better health outcomes for their students. This will not only improve individual student outcomes, but can also have broader societal benefits by reducing healthcare costs and improving overall public health.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 18, April 2026
Impact Factor: 8.2

ACKNOWLEDGMENT

It gives us great pleasure in presenting the paper on "Detection of Heart Disease using Machine Learning". We would like to take this opportunity to thank our guide, Prof. Priti Sudhakar Gund Pujari, Professor, E&TC Department, Adsul's technical Campus, Ahlyanagar, for giving us all the help and guidance we needed. We are grateful to her for hers kind support, and valuable suggestions were very helpful.

REFERENCES

1. Bo Jin , Chao Che,Zhen Liu ,Shulong Zhang ,Xiaomeng Yin And Xiaopeng Wei "Predicting the Risk of Heart Failure with EHR Sequential Data Modelling". IEEE Access 2018.
2. Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", "International Conference on "Computational Intelligence and Communication Technology" (CICT 2018).
3. Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor. "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection". IEEE Access (Volume: 7) 2019
4. Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava. "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques". IEEE Access (Volume: 7) 2019.
5. K. Prasanna Lakshmi, Dr. C.R.K.Reddy. "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining". International Conference on Computer, Communication and Control (IC4) 2015
6. M.Satish, D Sridhar, "Prediction of Heart Disease in Data Mining Technique", International Journal of Computer Trends & Technology (IJCTT), 2015. 7. Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", IJCTA, International Press 2015.
7. Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February–2015.
8. Mamatha Alex P and Shaicy P Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique", International Conference on Communication and Signal Processing 2019.
9. Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44-8.
10. Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-8.
11. Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.