

## A Study for Predicting Heart Disease using Machine Learning

Suriya Begum<sup>a</sup>, Farooq Ahmed Siddique<sup>b</sup>, Rajesh Tiwari<sup>c\*</sup>

<sup>a</sup>CSE Department BIET Telangana, India

<sup>b</sup>CSE Department GIT, Karnataka, India

<sup>c</sup>CSE Department BIET Telangana, India

\*suriyabegumstore@gmail.com, <sup>c</sup>dr Rajesh Tiwari20@gmail.com

**Article History:** Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

**Abstract:** Due to heart disease in India almost one person dies every day. A technique should be developed to detect the heart disease to reduce the number of deaths which is handy and at the same time reliable also. In the health care sector, Machine Learning plays an important role in the health care Industry. This paper deals with exploring and investigating different Machine Learning Algorithms. Also, it deals with applying multiple Algorithms on Heart Disease Dataset. In this study, from UCI the Dataset is taken. Six models were trained and tested, which are Logistic Regression, Random Forest Classifier, XGBoost Classifier, Support Vector Machine Classifier, Artificial Neural Network Classifier, K Nearest Neighbors Classifier. The Machine Learning algorithm Random Forest Classifier has proven to be the most accurate and reliable algorithm and hence used in the proposed system.

**Keywords:** Machine Learning, Heart Disease, Logistic Regression, Random Forest Classifier, XGBoost Classifier, Support Vector Machine Classifier, Artificial Neural Network Classifier, K Neighbors Classifier

### 1. Introduction

One of the human body's most vital organs is the heart. Heart attacks are the most common heart condition in India. The heart, through the body's circulatory system, pumps blood. Oxygen is distributed through the circulatory system of the body in the blood, and if the heart does not function correctly, the entire circulatory system of the body will fail. So if the heart doesn't work properly, it could even lead to death.

The types of heart disease include cardio-vascular disease (CVD) or heart disease, including the human body's blood and heart. Myocardial infarction (as a heart attack) is part of the CVD as well. Another form of heart disease is called coronary heart disease (CHD). In this type of disease, the coronary arteries develop a substance called plaque. Over the course of time, plaque growth will block the vessel completely. Heart Attack symptoms are:

**Chest pain:** One of the signs of a heart attack is chest pain. This occurs mostly because of the blockage of the plaque of the coronary artery of the body.

**Arm pain:** The pain generally begins in the chest and mostly travels towards the left arm.

**Low oxygen:** The level of oxygen decreases in the body because of the plaque which induces dizziness and loss of balance.

**Tiredness:** This cause of fatigue suggests that it becomes difficult to perform basic tasks.

**Excessive sweating:** Sweating is another common symptom.

**Diabetics:** In this case, patients have a heart rate of 100 bpm and even a heart rate of 130 bpm rarely.

**Bradycardia:** The patient may have a slower pulse of 60 bpm in this process.

**Cerebrovascular disease:** The patient will normally have a high heart rate of 200 bpm above average and may cause a heart attack higher than this [1].

**Hypertension:** The heart rate of the patient typically varies from 100-200 bpm in this situation.

Worldwide, due to CVD alone nearly 17.5 million deaths takes place. In middle-income and low-income nations, more than 75% of cardiovascular disease fatalities occur. 80% of the deaths caused by CVDs are also due to stroke and heart attack. India is also adding a rising figure of CVD patients per year. In India, nearly 30 million people are suffering from heart disease. Per year, more than two open heart surgeries are conducted in India. In recent years, the number of patients needing coronary intervention has risen from 20 percent to 30 percent, a matter of increasing concern [2].

## 2. Literature Survey

A lot of work has been carried out using the UCI Machine Learning dataset to predict heart disease. Using different data mining methods, various levels of accuracy have been achieved. Typically, heart is unable to push the necessary amount of blood to other areas of the body in order to satisfy the normal functioning of the body in this disease, and because of this, heart failure eventually occurs. The prevalence of heart disease is very high in the United States. Symptoms of heart disease include shortness of breath, physical body fatigue, swollen feet, and tiredness with associated signs, such as increased jugular venous pressure and peripheral edoema due to functional or non-functional cardiac irregularities. The early-stage investigation approaches used to detect heart disease have been difficult, and the resulting difficulty is one of the key factors affecting the standard of living. Diagnosis and treatment of heart disease is very difficult, especially in developing countries, owing to the rare availability of diagnostic instruments and the shortage of doctors and other services affecting the proper prediction of heart disease. The precise and correct detection of heart disease is important to reduce the associated risk of serious heart complications and to improve heart safety. Approximately 3 percent of the health care financial budget is impacted by the costs of heart disease management [3]. His invasive methods for the diagnosis of heart disease are based on the medical history of the patient, the physical examination study, and the medical experts' interpretation of the symptoms concerned. It is more costly and complex in terms of computation and takes time to evaluate [4]. A non-invasive medical decision support system focused on machine learning predictive models such as support vector machine (SVM), artificial neural network (ANN), k- nearest neighbour (K-NN), logistic regression (LR), decision tree (DT), Naive Bayes, AdaBoost, fuzzy logic and rough set [5] have been introduced to solve these complications in invasive-based heart disease diagnosis [6]. Heart disease dataset of Cleveland is accessible online in the data mining library which was used by numerous researchers [7] [8].

Detrano et.al. [7] Introduced a decision support method based on a logistic regression classifier for the classification of heart disease and obtained a 77 percent classification accuracy. With global evolutionary methods, the Cleveland dataset used and achieved high prediction efficiency in accuracy. For the collection of functions, the analysis used feature selection methods. Gudadhe et.al. [9] used SVM and MLP for cardiac disease classification. They suggested a method of classification and received 80.41 percent precision. A classification system for heart disease was developed by Kahramanli and Allahverdi [10], using a hybrid technique in which a neural network combines a fuzzy neural network and an artificial neural network. And a classification precision of 87.4 percent was achieved by the suggested classification system. An expert medical diagnostic heart disease system was developed by Palaniappan and Awang [11]. The predictive model of Naive Bayes obtained 86.12 percent output accuracy. ANN, which obtained an accuracy of 88.12 percent, was the second best predictive model, and the decision tree method reached 80.4 percent with the right forecast. Olaniyi et.al. [12] suggested a three-phase model to diagnose angina heart disease based on the ANN and obtained an accuracy of 88.89 percent in the classification. In addition, in healthcare information systems, the new system could be easily deployed. Das et.al. [13] adopted statistical analysis system and achieved 89.01 percent precision. Jabbar et.al. [14] developed a heart disease diagnostic system by using MLP. In order to detect heart disease, the authors have developed an integrated decision support medical system based on Fuzzy Logic. An accuracy of 91.10 percent [1] was achieved by their proposed classification scheme. Avinash Golande et.al. are researching various different ML algorithms that can be used for heart disease classification. Analysis was carried out to review the algorithms Decision Tree, KNN and K-Means that can be used for classification and compare their accuracy [15]. This study concludes that the accuracy obtained by Decision Tree was the highest, and it was concluded that the combination of different techniques and parameter tuning could make it successful. A system that deployed data mining techniques along with the Map Reduce algorithm was suggested by T.Nagamani, et al. for the 45 test instances set, the accuracy obtained according to this paper was greater than the accuracy obtained using traditional fuzzy artificial neural networks [16]. Here, due to the use of dynamic schema and linear scaling, the accuracy of the algorithm used has been enhanced. An ML model comparing five different algorithms has been developed by Fahd Saleh Alotaibi [17]. Compared to the Matlab and Weka tools, the Rapid Miner method was used to result in greater accuracy. The accuracy of the classification algorithms for Decision Tree, Logistic Regression, Random Forest, Naive Bayes and SVM is compared in this analysis. The tree decision algorithm had the highest precision. Anjan Nikhil Repaka, et al. [18] suggested a method using Naïve Bayesian technique and Advanced Encryption Standard stable data transfer technique for disease prediction. Theresa Princy explained various classification algorithms used to predict heart disease was carried out by R, et al. Naive Bayes, KNN (K- Nearest Neighbour), Decision tree, neural network and classifier accuracy were analysed for the different number of attributes [19] in the classification techniques used.

Nagaraj M Lutimath, et.al. applied Naive bayes and SVM to predict heart disease. Mean Absolute Error, Sum of Squared Error and Root Mean Squared Error are the performance indicators used in the analysis. It is known that SVM has emerged as a superior algorithm in terms of accuracy over Naive Bayes [20] [21] [22]. To predict the heart disease, RBF is applied by Shaikh Abdul Hannan et. al. [23]. A number of RBF units (nh) and biases

comprise the hidden layer (bk). A Gaussian function is commonly the most often used RBF. Random sub-set collection, k-means clustering and others are the different methods of choosing the centres. In MATLAB, the technique was introduced. The results obtained show that the radial base feature can be used successfully to prescribe medicines for heart disease (with an accuracy of 90 to 97%). AH Chen et al. [24] adopted a method to predict heart disease that can allow doctors to predict the status of heart disease based on patients' clinical data. Thirteen significant clinical characteristics have been selected, such as age, sex, type of chest pain. Based on Heart Disease Diagnosis and Prediction using Machine Learning and Data, an artificial neural network algorithm was used. Data was gathered from the UCI machine learning repository. Three layers were used in the artificial neural network model, i.e. the input layer, hidden layer, and output layer with 13 neurons, 6 neurons, and 2 neurons respectively. In this experiment, Learning Vector Quantization (LVQ) was used. LVQ is a special case of an artificial neural network that applies a supervised classification algorithm based on a prototype. The language of C programming was used as a method for classifying and predicting heart disease via an artificial neural network. The framework was built in the environment of C and C. The accuracy of the system for prediction proposed is close to 80%. Mrudula Gudadhe et. al. [25] proposed a decision support method for the classification of heart disease. The two key methods used in this framework are Support Vector Machine (SVM) and Artificial Neural Network (ANN). For the diagnosis of heart disease, to build a decision support system, a multilayer perceptron neural network (MLPNN) with three layers was used. Training for the multilayer perceptron neural network was given by a computer-efficient method of back-propagation algorithm. Results have shown that MLPNN can be successfully used to diagnose heart disease using a back-propagation technique. A prediction framework for heart disease based on Structural Equation Modeling (SEM) and Fuzzy Cognitive Map [26] was suggested by Manpreet Singh et al (FCM). They used a dataset from the 2012 Canadian Community Health Survey (CCHS). Twenty important attributes have been included here. The weight matrix for the FCM model is developed by SEM, which then predicts the probability of cardiovascular diseases. With a correlation between 20 attributes and CCC 121, a SEM model is specified. In order to establish FCM, a weight matrix must be first constructed. Previously used SEM is now used as the FCM, although the necessary ingredients have been achieved. For training SEM model, 80 percent of the data set was used and the remaining 20 percent for testing the FCM model. The accuracy achieved was 74 percent using this model. Using the concept of train and test on a heart disease prediction dataset, Carlos Ordonez [27] has tested the mining association rule. Generally, on the entire data collection, association rules are often mined without validating an independent sample [28]. To overcome this, an algorithm is developed that uses search constraints to reduce the number of rules. With motivation, confidence and elevation, the medical value of the discovered rules is then evaluated. Big Data was used by Prajakta Ghadge et. al. [29] to work on an effective method of heart attack prediction. Heart attack must be diagnosed in a timely and effective way due to its high prevalence. A record collection of 13 characteristics was obtained from the web-based Cleveland Heart Database (age, gender, serum cholesterol, fasting blood sugar, etc.). Three techniques are used to extract the patterns, neural network, Naïve Bayes and Decision tree. Asha Rajkumar et. al. [30] used the Tanagra tool for classification of data, 10 fold cross validation is used for evaluation of the data, and finally, the results are compared. The dataset is divided into two parts: training set used 80% of the data and testing set used 20% of the data for analysis. Naïve Bayes shows lower error ratios and takes the less time, when compare to the other three methods. G Purusothaman et. al. [31] done a survey on various classification algorithms for prediction of heart disease and compared them. The authors concentrate on working on hybrid models. The performance of Single models such as Decision Tree, Artificial Neural Network and Naïve Bayes are 76%, 85% and 69%, respectively. An accuracy of 96 percent is shown by Hybrid methods. Therefore, Hybrid models are accurate and efficient classifiers for better accuracy in prediction of heart disease [2].

### 3 Proposed Model

Six models were trained and tested, for heart disease prediction in the proposed work by applying six classification algorithms and also analysis on the performance is carried out. The main goal of this study is to predict whether a patient is suffering from heart disease or not by developing an efficient Model. Fig. 1 shows the Model for prediction of Heart Disease.

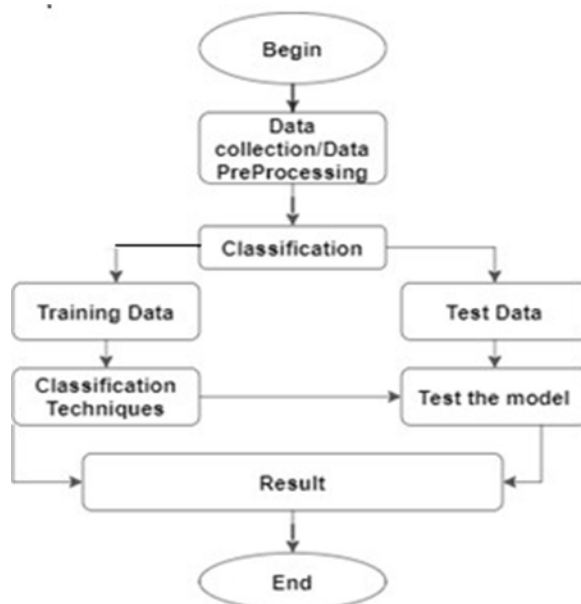


Fig. 1. Heart Disease Predicting Model.

**A. Collection and Preprocessing of Data**

The dataset is taken from UCI repository. This dataset consists of a total of 15 features. Dataset from UCI repository is used for our analysis. 13 attributes are used in the proposed work and they are described in Table I.

**B. Classification**

As an input to the various ML algorithms such as Logistic Regression, Random Forest, ANN, XGBoost Classifier, SVM and K Neighbors Classifier classification techniques, the attributes listed in Table 1 are given. 70 percent of the training dataset is divided into the input dataset and the remaining 30 percent into the evaluation dataset. The training dataset is the dataset used for the training of a model. The test dataset is used to verify the efficiency of the model being educated. Performance is measured and evaluated for each of the algorithms based on various metrics used, such as precision, accuracy, and recall and F-measure scores, as mentioned below. The numerous algorithms discussed are described as below:

Random Forest Classifier For regression and classification, Random Forest algorithms are widely used. It builds a data tree and makes predictions on that basis. On large datasets, the Random Forest algorithm can be used and missing values are also taken care by this classifier You can save the samples created from the decision tree so that it can be used on other data. Two main steps in the creation of random forests are : random forest construction and then predicting a random forest classifier created in the first step.

**Table I** Features Selected From Dataset

Sl.No.	Description of Attributes	Distinct Attribute Values
1.	Age : Represent a person's age in years	Several values from 29 to 77
2.	Sex: Describe a person's gender (0- Female, 1-Male)	0 and 1
3.	Chest-pain-type: With values 1, 2 and 3, people are at a high risk to have heart disease when compare tothe people with a value 0.	0,1,2 and 3
4.	Resting-pressure-blood: It reflects the BP of the patient.	Several values from 94 to180

5.	Serum-cholesterol-mg-per-dl:It indicates the patient's cholesterol amount.	Several values from 126 to 564
6.	Resting-ekg-results: Displays the ECG results	0,1 and 2
7.	Max-heart-rate-reached: reflects the patient's max heartbeat	Several values from 96 to 202
8.	Exercise-induced-angina-used: to determine whether angina is induced by exercise. If yes=1 or otherwise=0	0 and 1
9.	Oldpeak-eq-st-depression: Patient condition during peak exercise is defined by Slope of Peak Exercise St Section. It is divided into three parts (Unsloping, Flat, Down sloping)	Several values from 0 to 6.2.
10.	Slope of peak exercise st segment: Patient condition during peak exercise is defined by Slope of peak exercise segment st . It is divided into three parts of the dataset (Unsloping, Level, Down Sloping). It's Colum's class or name. This dataset has a binary classification, 0 and 1. There is less risk of heart failure in class '0'.	1,2 and 3
11.	Num-major-vessels:Fluoroscopy Effect.	0,1,2 and 3
12.	Thal: test is required for patients with chest pain or trouble breathing. There are four types of values to indicate the Thallium test..	0,1,2 and 3
13.	Heart disease present: It is the dataset aim column. This is Colum's class or name. In the dataset, this reflects the number of groups. This dataset has a binary classification, 0 and 1. There is less risk of heart attack in class '0'.	0 and 1

**XGBoost Classifier** These days, it is the most popular algorithm for machine learning. It is well known to have better solutions than other ML algorithms irrespective of the data form (regression or classification). Extreme Gradient Boosting (XGBoost) is similar, but more effective, to the gradient boosting system.

**Logistic Regression** Mostly used for binary classification problems, it is a classification algorithm. The logistic regression algorithm uses the logistic function in logistic regression, instead of fitting a straight line or hyper plane, to squeeze the output of a linear equation between 0 and 1. There are 13 independent variables that make classification good for logistic regression.

**Support Vector Machine** Support Vector Machine (SVM) is a technique of supervised learning which classifies data over a hyper plane into two classes. Except that it does not use Decision trees at all. To reduce any possibility of misclassification, SVM seeks to optimize the margin (distance between the hyper plane and the two closest data points from each respective class). Scikit-learn, MATLAB and LIBSVM are some common implementations of support vector machinery.

**Artificial Neural Network** A computer model focused on functions and structure of biological neural networks is the Artificial Neural Network (ANN). The structure of the artificial neural network is influenced by knowledge that passes through the network. ANN's are known to be nonlinear statistical data processing tools. ANNs have interconnected layers where the dynamic relationships between inputs and outputs are modelled or patterns are identified. In order to improve current data processing systems, artificial neural networks are fairly simple mathematical models.

**K-Nearest Neighbors** It is one of the supervised ML algorithm that can be used for both predictive problems of classification and regression. However, predictive problems in industry are primarily used for classification. It uses 'feature similarity' to predict the values of new data points, which further suggests that a value will be assigned to the new data point based on how closely the points in the training set are matched.

**C. Methodology**

Our approach to solve this problem is to make Multiple Regression Models and then choosing the Model with the highest accuracy and tuning the hyper-parameters of that model to obtain maximum accuracy.

Techniques Used For Feature Selection

- Correlation
- Feature Importance

- Missing Values
- Domain Knowledge

Techniques Used For Dropping Features

Correlation (Highly correlated features are dropped)

Feature Importance(Features contributing 0% are dropped)

Missing Values (Features having 60% missing values are dropped)

#### 4. Results And Analysis

The data obtained is cleaned, supervised and categorical data. The dependent variable is heart disease present, the raw data obtained was having 180 rows and 15 columns. To analyse this, Correlation between independent features and even with respect to target variable is used, along with this Pandas Profiling is done to the entire dataset to understand each and every feature. Label Encoding and Scaling of the dataset has been done. The features has been finalized based on Correlation between variables and feature importances of the model.

After Exploratory data analysis, the finalized features are:

- Slope-of-peak-exercise-st-segment
- Thal
- Resting-blood-pressure
- Chest-pain-type
- Num-major-vessels
- Serum-cholesterol-mg-per-dl
- Oldpeak-eq-st-depression
- Sex
- Age
- Max-heart-rate-achieved
- Exercise-induced-angina

Logistic Regression, Random Forest Classifier, Artificial Neural Network Classifier, XGBoost Classifier, Support Vector Machine Classifier, and K Neighbors Classifier. K- Fold cross validation technique is also applied to the model.

This section demonstrates the outcomes obtained through the application of Logistic Regression, Random Forest Classifier, Artificial Neural Network Classifier, XGBoost Classifier, Support Vector Machine Classifier, and K Neighbors Classifier. Accuracy ranking, Accuracy, Recall and F- measure are the metrics used to conduct performance analysis of the algorithm. The metric of precision equation (1) provides the proper measure of positive analysis. The measure of actual positives that are right is defined by recall equation (2). The F-measurement equation (3) measures precision. Accuracy measures correct predictions over the output size equation (4).

$$Precision = \frac{TruePositive(TP)}{TruePositive(TP) + FalsePositive(FP)} \quad (1)$$

$$Recall = \frac{TruePositive(TP)}{TruePositive(TP) + FalseNegative(FN)} \quad (2)$$

$$F-Measure = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (3)$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

- TP: the patient has the disease and the test is positive.
- FP: the patient does not have the disease but the test is positive.
- TN: the patient does not have the disease and the test is negative.
- FN: the patient has the disease but the test is negative.

The pre-processed dataset is used to conduct the experiment. Exploration of the algorithms have been carried out and finally applied. The success metrics discussed above are obtained using the uncertainty matrix. The model’s efficiency is described by the Confusion Matrix. Table II shows the confusion matrix for the propped model for various algorithms. The accuracy score obtained for the classification techniques for Logistic Regression, Random Forest Classifier, Artificial Neural Network Classifier, XGBoost Classifier, Support Vector Machine Classifier and K Neighbors is shown in Table III.

**TABLE II** Confusion Matrix

Sr. No.	Algorithm	True Positive	False Positive	False Negative	True Negative
1.	Random Forest	21	0	32	0
2.	XGBoost	14	2	30	7
3.	Logistic Regression	24	3	35	9
4.	Artificial Neural Network	14	2	30	7
5.	Support Vector Machine	13	5	27	8

**TABLE III** Analysis Of Machine Learning Algorithms

Algorithm	Training Accuracy	Testing Accuracy
Random Forest	100%	100%
XGBoost	92.60%	83%
Logistic Regression	80%	83%
Artificial Neural Network	86.99%	83%
Support Vector Machine	92.68%	79.56%
K Neighbors Classifier	Accuracy score is 71.69% with 12-neighbors.	

**5. Conclusion**

Due to heart disease, there is an increased in the number of deaths, day by day. The implementation of a method to efficiently and reliably predict heart diseases has become compulsory. The main motivation of this study is to find a powerful ML algorithm for detection of heart disease. This study uses Logistic Regression, Random Forest Classifier, Artificial Neural Network Classifier, XGBoost Classifier, Support Vector Machine Classifier, and K Neighbors algorithms to predict heart disease. The outcome of this analysis shows that the Random Forest algorithm is the most powerful algorithm for heart disease prediction, with an accuracy score of 100%. The study can be strengthened in the future by taking Indian dataset from the well-known hospitals to efficiently predict heart disease.

**References**

1. S.Nandhini, Monojit Debnath, Anurag Sharma , Pushkar, “Heart Disease. Prediction using Machine Learning”, International Journal of Recent Engineering Research and Development (IJRERD), Volume 03, Issue 10, October 2018, PP. 39-46, ISSN: 2455-8761.
2. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, “Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review”, Advances in Computational Sciences and Technology, Volume 10, Number 7 (2017) pp. 2137-2159, ISSN 0973-6107.
3. J. Lo’pez-Sendo’n, ”The heart failure epidemic,” Medicographia, vol. 33, pp. 363-369, 2011.
4. K. Vanisree and J. Singaraju, ”Decision support system for congenital heart disease diagnosis

- based on signs and symptoms using neural networks,” *International Journal of Computer Applications*, vol. 19, no. 6, pp. 6-12, 2011.
5. S. Nazir, S. Shahzad, and L. Septem Riza, ”Birthmark-based software classification using rough sets,” *Arabian Journal for Science and Engineering*, vol. 42, no. 2, pp. 859-871, 2017.
    - A. Methaila, P. Kansal, H. Arya, and P. Kumar, ”Early heart disease prediction using data mining techniques,” in *Proceedings of Computer Science Information Technology (CCSIT-2014)*, vol. 24, pp. 53-59, Sydney, NSW, Australia, 2014.
  6. R. Detrano, A. Janosi, and W. Steinbrunn, ”International application of a new probability algorithm for the diagnosis of coronary artery disease,” *American Journal of Cardiology*, vol. 64, no. 5, pp. 304-310, 1989.
    - A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, ”A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms,” *Mobile Information Systems*, vol. 2018, p. 3860146, Dec. 2018, doi: 10.1155/2018/3860146.
  7. M. Gudadhe, K. Wankhade, and S. Dongre, ”Decision support system for heart disease based on support vector machine and artificial neural network,” in *Proceedings of International Conference on Computer and Communication Technology (ICCCCT)*, pp. 741-745, Allahabad, India, September 2010.
  8. H. Kahramanli and N. Allahverdi, ”Design of a hybrid system for the diabetes and heart diseases,” *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82-89, 2008.
  9. S. Palaniappan and R. Awang, ”Intelligent heart disease prediction system using data mining techniques,” in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications (AICCSA 2008)*, pp. 108-115, Doha, Qatar, March-April 2008.
  10. E. O. Olaniyi and O. K. Oyedotun, ”Heart diseases diagnosis using neural networks arbitration,” *International Journal of Intelligent Systems and Applications*, vol. 7, no. 12, pp. 75-82, 2015.
  11. R. Das, I. Turkoglu, and A. Sengur, ”Effective diagnosis of heart disease through neural networks ensembles,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675-7680, 2009.
  12. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, ”Classification of heart disease using artificial neural network and feature subset selection,” *Global Journal of Computer Science and Technology Neural Artificial Intelligence*, vol. 13, no. 11, 2013.
  13. Rajesh Tiwari, Manisha Sharma, Kamal K. Mehta and Mohan Awasthy, ”Dynamic Load Distribution to Improve Speedup of Multi-core System using MPI with Virtualization”, *International Journal of Advanced Science and Technology*, Vol. 29, Issue 12s, 2020, pp 931 – 940, ISSN: 2005 – 4238.
  14. T.Nagamani, S.Logeswari, B.Gomathy,” Heart Disease Prediction using Data Mining with Mapreduce Algorithm”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278- 3075, Volume-8 Issue-3, January 2019.
  15. Fahd Saleh Alotaibi,” Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 6, 2019.
  16. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, ”Design And Implementation Heart Disease Prediction Using Naives Bayesian”, *International Conference on Trends in Electronics and Information(ICOEI 2019)*.
  17. Theresa Princy R,J. Thomas,’Human heart Disease Prediction System using Data Mining Techniques’, *International Conference on Circuit Power and Computing Technologies*,Bangalore,2016.
  18. Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,’Prediction Of Heart Disease using Machine Learning’, *International journal Of Recent Technology and Engineering*,8,(2S10), pp 474-477, 2019.
  19. Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT)* Volume 09, Issue 04 (April 2020), ISSN (Online) : 2278-0181.



20. P. M. Awantika and Rajesh Tiwari, "A Novel Based AI approach for Real Time Driver Drowsiness Identification System using Viola Jones Algorithm in MATLAB platform", *Solid State Technology*, Vol. 63, Issue 05, 2020, pp 3293 – 3303, ISSN: 0038-111X.
21. Shaikh Abdul Hannan, A.V. Mane, R. R. Manza, and R. J. Ramteke, Dec 2010, "Prediction of Heart Disease Medical Prescription using Radial Basis Function", *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, DOI: 10.1109/IC-CIC.2010.5705900 ,28-29 .
22. AH Chen, SY Huang, PS Hong, CH Cheng, and EJ Lin, 2011, "HDPS: Heart Disease Prediction System", *Computing in Cardiology*, ISSN: 0276-6574, pp.557- 560.
23. Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010, "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network", *International Conference on Computer and Communication Technology (IC-CCT)*, DOI:10.1109/ICCCT.2010.5640377, 17-19.
24. Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago, 2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model Fuzzy Cognitive Map", *IEEE International Conference on Fuzzy Systems (FUZZ)*, pp. 1377-1382.
25. Carlos Ordonez, 2006, "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction", *IEEE Transactions on Information Technology in Biomedicine (TITB)*, pp. 334-343, vol. 10, no. 2.
26. Rajesh Tiwari, Manisha Sharma and Kamal K. Mehta "IoT based Parallel Framework for Measurement of Heat Distribution in Metallic Sheets", *Solid State Technology*, Vol. 63, Issue 06, 2020, pp 7294 – 7302, ISSN: 0038-111X.
27. Suriya Begum, Farooq Ahmed Siddique, "A study to predict home loan defaulter using machine learning", *SSIM* April 2021 .
28. Prajakta Ghadge, Vrushali Girme, Kajal Kokane, and Prajakta Deshmukh, 2016, "Intelligent Heart Attack Prediction System Using Big Data", *International Journal of Recent Research in Mathematics Computer Science and Information Technology*, Vol. 2, Issue 2, pp.73- 77, October 2015–March.
29. Asha Rajkumar, and Mrs G. Sophia Reena, 2010, "Diagnosis of Heart Disease using Data Mining Algorithms", *Global Journal of Computer Science and Technology*, Vol. 10, Issue 10, pp.38-43, September.
30. Purusothaman, and P. Krishnakumari, June 2015, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", *Indian Journal of Science and Technology*, Vol. 8(12), DOI:10.17485/ijst/2015/v8i12/58385, pp. 1-5