



HAL
open science

A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches

Yogita Solanki, Sanjiv Sharma

► **To cite this version:**

Yogita Solanki, Sanjiv Sharma. A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches. *International Journal of Information Engineering and Electronic Business*, 2019, 11 (4), pp.43-51. 10.5815/ijieeb.2019.04.05 . hal-02188879

HAL Id: hal-02188879

<https://hal.science/hal-02188879>

Submitted on 18 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches

Yogita Solanki

Department of Computer Science and Engineering, Madhav Institute of Technology and Science,
Gwalior, Madhya Pradesh, India
Email: solankiyogita53@gmail.com

Sanjiv Sharma

Department of Computer Science and Engineering, Madhav Institute of Technology and Science,
Gwalior, Madhya Pradesh, India
Email: dr.sanjiv@mitsgwalior.in

Received: 23 March 2019; Accepted: 22 April 2019; Published: 08 July 2019

Abstract—This document presents the required layout of articles to Medical data mining has become one of the prominent issues in the field of data mining due to the delicate lifestyle opted by the people which are leading them towards various chronic health diseases. Heart disease is one of the conspicuous public health concern worldwide issues. Since clinical data is growing rapidly owing to deficient health awareness, various techniques and scientific methods are opted for analyzing this huge data. Several data mining techniques such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision tree, Naïve Bayes and Artificial Neural Network (ANN) are introduced for the prediction of health disease. These techniques help to mine the relevant and useful amount of data, form the medical dataset which helps to provide beneficial information to the medical institutions. This study presents various issues related to healthcare and various machines learning algorithms which have to withstand to provide the best possible output. A comprehensive review of the literature has been summarized to put lights on the previous work done in this field.

Index Terms—Medical Data Mining, Machine Learning algorithm, Heart Disease Prediction, Heart Disease, Comprehensive Review.

I. INTRODUCTION

Among numerous life-threatening diseases, heart disease has received an enormous deal of attention in the medical field. Heart disease is considered one of the deadliest & cardiovascular diseases. Heart-attack is also a result of heart disease & when the heart becomes very weak to circulate blood then breathlessness can occur. Many complications occur if heart disease remains untreated and unidentified. There are numerous risk factors that cause heart disease such as obesity, cholesterol, blood pressure, eating an unhealthy diet, age,

etc. There are various types of heart disease. Some kind of heart disease is introduced as follows.

- 1) *Coronary Heart Disease*: It is a situation in which plaque deposits block the coronary blood vessels, that provides less supply of oxygen & blood to the heart.
- 2) *Myocarditis*: It is an irritation of the heart muscles that are usually caused by a viral, fungal, and bacterial infection that affects the heart.
- 3) *Cardiomyopathy*: This is a change in muscle structure due to the weakening of heart muscles or inadequate heart pumping.
- 4) *Congestive heart failure*: It is a situation where the heart can't pump adequate blood to the rest of the body.
- 5) *Arrhythmias*: It is associated with a disorder in the recurring movement of the heartbeat. Abnormal heartbeat is due to a short circuit in the heart's electrical system [1].
- 6) *Stroke*: Inhibiting the supply of blood causes damage to the brain. Cell death [2] occurs due to the bad flow of blood in the brain. The two major kinds of stroke are ischemic (because of lack of blood flow) & hemorrhagic (because of bleeding).

Diagnosis of heart-disease is very time consuming & complex process for a doctor, and wrong clinical decisions due to negligence would cause the death of a patient that can't be afforded by any hospital. Therefore prediction of heart disease at an early stage is necessary to diminish the risk of heart failure.

In the present era, technology plays an imperative role in every industry as well as in our daily life. Medical technology is a broad area where innovation plays an essential role in maintaining health. The Electronic Medical Record (EMR) software system was created to fully adjust all aspects of medical workflow with retrieval, storage & modification of patient records digitally, clinical annotation, ordering laboratory, prescription

writing, viewing test & imaging tests outcomes. The clinical database contains a huge quantity of records of patients and their medical conditions.

Analyzing these qualitative & quantitative data for discovering the relationships among a huge number of samples can be possible by using data mining techniques. Health Care Big Data [3] have numerous challenges as the complexity of data is too high for the reason that everyone cannot examine or process the medical report & genomic data i.e. imaging data & unstructured data. EMR is also limited to data acquisition than analytics. Clinical records are hard to understand in the right manner as data of the clinical reports are in the distinct format & the result of the medical test also varies based on the distinct symptoms of diseases. Big data analysis offer the opportunities to the health care industry in order to aggregate the huge amount of patients care data to understand, classify & make some learning methods that are utilized as an alternative treatment.

II. PREDICTION OF HEART DISEASE

Prediction of heart disease [4] is identifying the presence of heart disease by analyzing the risk factor of heart disease patient's data. Data mining techniques are utilized for various applications. In the health care industry, data mining has played a significant role in predicting heart disease. Doctors have to do a number of tests to identify any disease. But heart disease prediction reduced the time and providing accurate results. The basic prediction model is shown below in Fig.1.

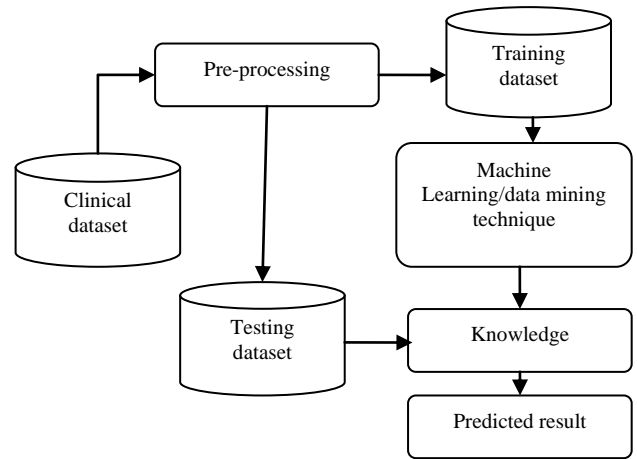


Fig.1. Basic Prediction Modal

III. TECHNIQUES USED FOR HEART DISEASE PREDICTION

There are numerous Data Mining & machine learning techniques that are mainly utilized for predicting heart disease. The objectives of mining heart disease[5] dataset are:-

- Identify the characteristics of patients having heart disease.
- Decide the attribute values that differentiate between patients having heart disease & patients with no heart disease.
- Predict people whose heart disease is likely to be diagnosed.
- Identify the impact of medical attributes.

Some techniques for heart disease prediction have been introduced which consist of clustering, regression, classification, etc as shown in Fig.2.

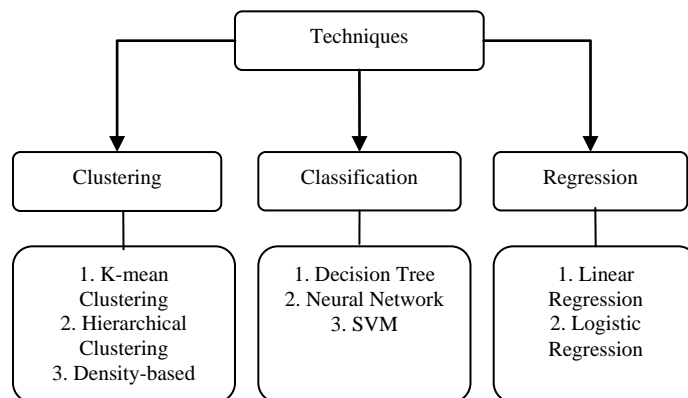


Fig.2. Techniques used for Heart Disease Prediction

A. Clustering:

Clustering is the procedure of combining data objects that are similar to each other contained by the cluster. It splits the large dataset into groups as per the similarity.

1. K-mean clustering:

K-means[6] is a clustering algorithm that is popular for cluster analysis in data mining. It is simple to implement the learning algorithm. Simplicity & speed are the advantages of the K- Means clustering algo.

Steps:-

- Choose k-centre (randomly)
- Split the data into k-clusters by combining the points.
- To find a new center, calculate the mean of k cluster.
- Until centers don't change, repeat steps 2 & 3

Heart disease prediction model mainly used K- mean clustering for grouping the attributes. The working of K-mean clustering is represented as shown in Fig.3.

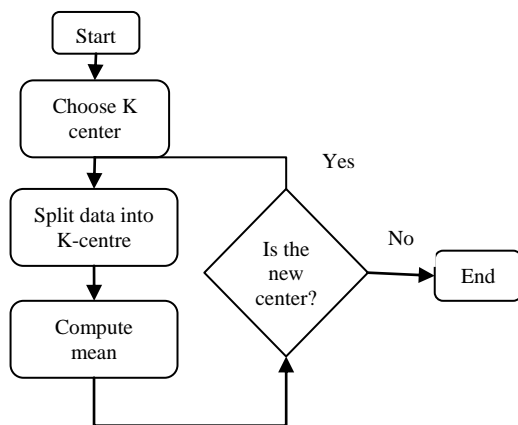


Fig.3. K-mean Clustering

2. Hierarchical Clustering:

It is an agglomerative (bottom-up) clustering technique begins with one factor (singleton) and repeatedly adjoins two or more appropriate clusters. It gets stop when certain (k) number of clusters is attained. In this technique, the prior information of the clusters is not required and hence it becomes easy to use.

3. Density-based Clustering:

Density-based clustering[7] consists of two different concepts of density i.e. reachability and connectivity. One of the benefits of the DBSCAN is that unlike Hierarchical clustering, it does no longer feel the necessity for the information about the number of clusters in the data set.

B. Classification:

In this technique, a series of data is classified to achieve more precise predictions of the target class for all case of data.

1. Decision Tree:

The Decision Tree [8] is a classifier that is easily implemented. The main benefit of the decision tree is that it can be applied to enormous data which is to be interpreted. A decision tree is a tree-like structure that contains branches, nodes, and arcs. The arcs connect from one node to another. The branch has attributes. For making a suitable decision the traversing starts from the root node to the leaf node. There are various decision tree for example CART(Classification and Regression tree), ID3 (iterative dichotomiser 3), C4.5, random forest, etc.

1.1. Random Tree:

It is a supervised classifier that builds a decision tree that assumes randomly selected components to say K at every node of the tree without reducing. The random tree[9] methodology could work both with regression & classification difficulties. It is a bagging tree that highlights the ability of multiple varied analyses and ensemble learning to provide deep data understanding.

1.2. ID3 (Iterative Dichotomiser 3)

ID3 [10] developed by Ross Quinlan is the easiest decision tree learning method. The fundamental reason behind ID3 algo is to develop the decision tree by introducing a greedy search, top-down through the provided sets of data for cross-checking every characteristic at each and every tree node for choosing the most useful attribute for classifying a given set.

1.3. C4.5:

This methodology is the advanced version of ID3 algo that was developed by Ross Quinlan. C4.5[11] takes both a particular & continuous attribute to build a decision tree so as to cope with continuous attributes. It has some extra features like coping with missing values, categorizing the continuous attributes, and sorting of decision trees, rule derivation, and others.

1.4. CART:

Classification and Regression Trees(CART) [12] is one of the decision trees. In the classification trees, the tree is utilized to classify the "class" within which there is a possibility of a target variable & the target variable is categorical. In regression trees, the target variable is continuous & a tree is utilized to expect its value.

2. Neural Network:

The neural network[13] depicts a biological nervous system as a computational model with the adaptive human brain. It can analyze the huge amount of data at very high speed for producing meaningful information such as in prediction and classification. However, ANN has proposed different models, the Feed-forward Neural Network (FNN) is the most well-known & widely used in many applications. ANN has shown in Fig. 4.

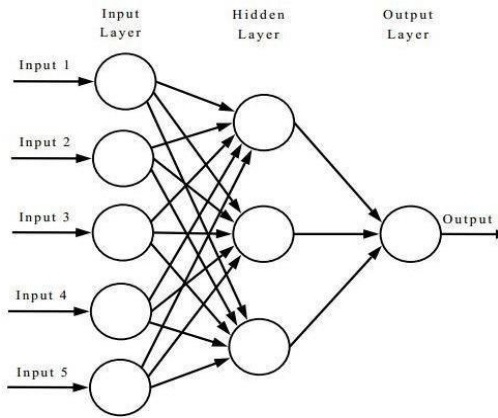


Fig.4. Artificial Neural Network

3. SVM (Support Vector Machine):

Support Vector Machine is a differentiate classification technique that is analyzed by partitioning a hyperplane. SVM[14] make a hyperplane for classification and regression techniques. It determines the closest data vectors known as support vectors (SV), for the decision detention in the training set & a known new test vector could be separated by using only the specified closest data vectors. The graphical representation of SVM is shown in Fig.5.

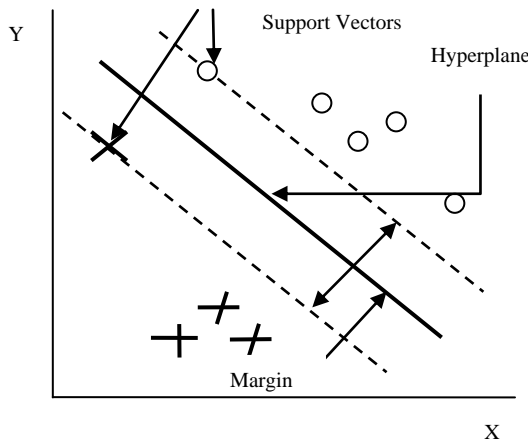


Fig.5. Support Vector Machine

C. Regression:

Regression [15] is the technique that establishes the relationship between the dependent and independent variable. The dependent variable is called as Target variable and other than target variable, all variable are called Independent variable.

1. Linear regression:

Linear regression is a correlation among the predictors & the output could be approximated with a straight line. It has only one dependent and one independent variable. The Graphical representation of linear regression is shown in Fig.6.

$$Y = \beta_2X + \beta_1 \tag{1}$$

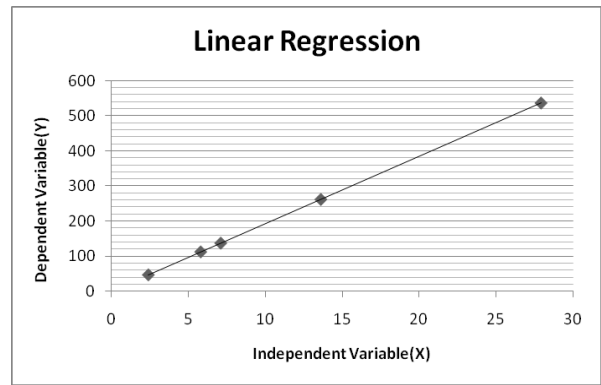


Fig.6. Linear Regression

2. Logistic regression:

Logistic Regression is a classification algorithm. It is utilized to predict a binary outcome. It has one dependent and one & more independent variable. General representation of logistic regression is

$$Y = \beta_3X_1 + \beta_2X_2 + \beta_1 \tag{2}$$

where

- Y= Dependent variable
- X=Independent variable
- β_2, β_3 =Regression coefficient
- β_1 =constant.

The graphical representation of logistic regression is shown below in Fig.7.

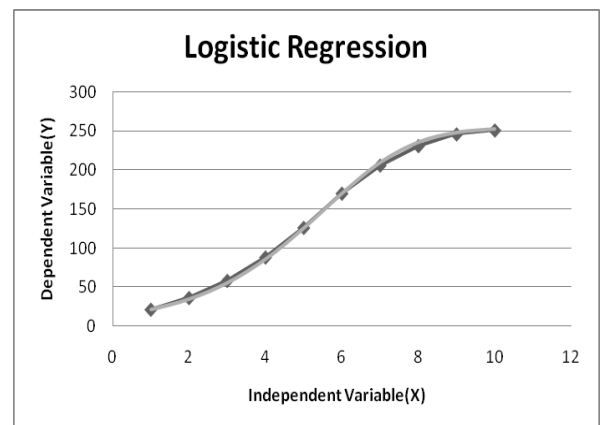


Fig.7. Logistic Regression

IV. RELATED WORK

Xiao Liu et al. [16] suggested a hybrid classification scheme based on the ReliefF & Rough Set (RFRS) technique by a diagnosis of heart disease. It consists of 2 subsystems, the first system made by the RFRS feature selection model & other is a classification scheme with

all together classifier. As per the jackknife cross-validation method, the highest accuracy of 92.59% was obtained. FEN MIAO et al. [17] made an all-inclusive risk model by an improved random survival forest (iRSF) through a high level of accuracy for predicting heart failure rate. Thus, the model separated the non-survivors & 1-year survivors with better accuracy compared to earlier heart prediction models. This achieved an accuracy of 82.1%.

R. Chitra et al. [18] designed a new intelligent method to establish information on the possibility of Heart Disease via Optimized Neural Network. By Genetic Algorithm for prediction, a Feed Forward Neural Network (FFNN) optimized. The accuracy is 85.3%, sensitivity is 83.5%, & specificity is 86.4% for GA-ANN model of prediction. Mohammad A.M. Abushariah et al. [19] design & build an automatic heart disease diagnosis scheme in MATLAB. The Cleveland dataset develops the training and testing dataset for heart diseases prediction. Two schemes were developed for train & testing the Cleveland dataset. The primary scheme based on the Multilayer Perceptron (MLP) structure on the Artificial Neural Network (ANN), while the following scheme based on the Adaptive Neuro-Fuzzy Inference Systems (ANFIS) method. The ANN scheme improves the Neuro-Fuzzy scheme with the greatest accuracy i.e. 87.04%.

Sheetal Sonawane et al. [20] represents an MDSS for cardiovascular disease classification in the support vector machine (SVM) based on sequential minimal optimization (SMO) technique. It trained SVM in the Cleveland Heart Disease database by SMO method in UCI Machine Learning Repository Data. The accuracy of the proposed work is 84.12%. Nilakshi P. Waghulde et al. [21] proposed a cardiovascular Prediction scheme by Genetic Algorithm & Neural Network. This scheme computes the various hidden nodes for a neural network that train the network with the appropriate selection of neural network architecture & utilizes the worldwide optimization of a genetic algorithm for initializing the neural network. As a result, it has been found that the genetic neural approach predicts cardiovascular disease up to 98% accuracy.

Ritika Chadha et al. [22] attempts to provide the methodology & implementation of the techniques for example Decision Tree, Artificial Neural Networks (ANN) & Naive Bayes. Certainly, the observations expose that Artificial ANN performs better than Decision Tree & Naive Bayes. ANN predicted the disease with an accuracy of 99% as per this research. Hidayat TAKCI et al. [23] proposed the best feature selection algo & the best machine learning methodology for predicting heart attacks. By the optimal parameters & numerous feature selection methods, numerous machine learning techniques were utilized. As per the outcome, the SVM algorithm with the linear kernel is the best machine learning methodology, whereas the relief technique is the best feature selection algorithm. It achieved the highest accuracy value of 84.81%.

Rajesh Jangade et al. [24] studied various classification techniques that could be used for the prediction of heart

associated difficulties. This work focuses to determine an appropriate technique that can assist future decision making. The comparison is analyzed among the classifier for detecting that tends to be more efficient for the dataset. The decision tree presents the accuracy of 75.10%. E. K. Hashi et al. [25] offered a proficient clinical decision to maintain a scheme for predicting disease by classification methods. The paper is based on WEKA software and percentage ratio method for train and test dataset using KNN & C4.5. They give 76.96% & 90.43% accuracy correspondingly. C4.5 Decision Tree gives better accuracy compared to KNN and helps for the clinical decision support system.

M. Sultana et al. [26] proposed a prediction of heart disease by J48, SMO, KStar, Multilayer perceptron, Bayes Net by WEKA tool. On the basis of performance from different factor SMO (89% of accuracy) and Bayes Net (87% of accuracy) achieve optimum performance than Multilayer perceptron, J48 and KStar techniques using k-fold cross-validation. The accuracy performance achieved by those algorithms is still not satisfactory. So that if the performance of accuracy is improved more to give better decision to diagnosis disease. A. Davari Dolatabadi et al. [27] describe automatic examination of coronary artery disease (CAD) patients by improved SVM, in this parameters of SVM are optimized to improve the accuracy of prediction, which gives 99.2% accuracy by k-fold cross-validation. The paper helps to diagnosis disease at an early stage and to reduce the cost. The accuracy obtained is good to predict if the individual has heart disease or not.

Kaan Uyar et al. [28] suggested an algorithm based on trained recurrent fuzzy neural networks (RFNN) i.e Genetic Algorithm (GA) for predicting heart diseases. 97.78% accuracy was achieved in the testing dataset. They used the probability of the misclassification error, root means a square error, sensitivity, F-score, precision & specificity are computed for knowing the accuracy. Ashok Kumar Dwivedi [29] predicted heart disease by evaluating the 6 machine learning techniques. The 85% classification accuracy was obtained by logistic regression with the specificity of 81% & sensitivity of 89% correspondingly.

R. Kannan et al. [30] perform four distinct machine learning techniques, for example, stochastic gradient boosting, logistic regression, SVM & Random Forest are compared by the obtained accuracy in R language. Compare the accurateness of the model predictions presents that logistic regression & the area under the ROC provides the greatest 87% accuracy. Sarabi H. Mujawar et al. [31] objective is to detect more precisely the occurrence of heart disease. Instead of going for a no. of tests, the fewer no. of attributes is a challenging job for Data Mining. Two data classification methods were performed that are Naïve Bayes & modified K-means. The modified K-mean clustering technique obtained an accuracy of 91%. In this research, a modified K-means technique is suggested that remove one of the main limitations of vital K-means technique that involve a number of clusters as input.

Amita Malav et al. [32] offer an effective prediction method to conclude & take out the unknown information about heart disease by a hybrid of artificial neural network & K-means clustering technique. For performing the combination of numerous attributes it utilizes k-means & it utilizes the Back-propagation method in neural networks for prediction. This hybrid model achieved an accuracy of 97%. Purushottam et al. [33] proposed an effective prediction scheme by data mining. This scheme helps the medical practitioner to make effective decision producing that based on the definite

parameter. In the testing phase, it provides 86.3% accuracy and in training phase 87.3% accuracy.

Vikas Chaurasia et al. [34] deliberated & the experiments different classifier to predict the patients those have heart disease. Three classifiers, for example, DT, CART, and ID3 were utilized for analysis of patients having heart diseases. Based on the patient's data, the best algorithm is the CART classification, whose accuracy is 83.49.

The comparative analysis of the literature survey is mention below in Table 1.

Table 1. A Comparative Analysis of Numerous Algorithms

Author	Year	Proposed	Technique	Accuracy
R. Chitra et al.[18]	2013	An FFNN optimized by Genetic Algorithm	GA-ANN	85.3%
Sheetal Sonawane et al.[20]	2013	It trained SVM by SMO	Trained SVM by SMO	84.12%
Vikas Chaurasia et al.[34]	2013	Compare Different classifier	CART	83.49%
Mohammad A. M. Abushariah et al.[19]	2014	The 2 system is performed first MLP structure on the ANN and another system by ANFIS	The system based on MLP structure on the ANN	87.04%
Nilakshi P. Waghulde et al.[21]	2014	Hybrid of genetic algorithm and neural network	A genetic- neural approach	98%
Sairabi H. Mujawar et al.[31]	2015	Apply classification technique	Modified k-mean	91%
Prof. (Dr.) Kanak Saxena et al. [33]	2016	Build an effective Heart Disease Prediction scheme by classified rules.	Decision classifier	86.3% (testing phase) & 87.3% (training phase)
Ritika Chadha et al.[22]	2016	Comparison of ANN, Decision Tree, Naïve Bayes	ANN	99%
M. Sultana et al.[26]	2016	Prediction by WEKA the tool with 10-Fold cross-validation	SMO	89%
Ashok Kumar Dwivedi et al.[29]	2016	Evaluated the potential of six machine learning techniques	Logistic Regression(LR)	85%
Xiao Liu et al.[16]	2017	Proposed the Hybrid classification scheme based on the ReliefF & Rough Set (RFRS) method.	C4.5 with cross-validation and ReliefF and Rough Set (RFRS) method	92.57%
A. Davari Dolatabadi et al.[27]	2017	SVM is optimized by K-fold cross-validation(CV)	SVM with K fold cross-validation	99.2%
E. K. Hashi et al.[25]	2017	Compare KNN and C4.5	C4.5	90.43%
Kaan Uyar et al.[28]	2017	Apply genetic algorithm(GA)	GA	97.78%
Amita Malav et al.[32]	2017	Hybrid Model	K-mean and ANN	97%
FEN MIAO et al.[17]	2018	Heart failure prediction by iRSF	iRSF	82.1%
Hidayat TAKCI et al.[23]	2018	Determine the best machine learning technique with the most excellent feature selection algorithm for predicting heart attacks.	SVM algorithm with the linear kernel(LK) with best characteristic selection technique is the reliefF.	84.81%
Rajesh Jangade et al.[24]	2018	The comparison is analyzed among the classifier.	Decision Tree	75.10%
R. Kannan et al.[30]	2018	Comparison of logistic regression, Random Forest, stochastic gradient boosting, & SVM in R language.	Logistic Regression(LR)	87%

Nowadays advanced technologies are utilized in the health domain but traditional clinical equipment & physical resource aren't compatible to work with them like- hospitals, medical reports, and imaging sensor devices. Existing technology involves human interference more towards manual, so for overcoming this problem, a technology requires, that getting the computers to learn, act like humans & enhance their learning in excess of time in a sovereign manner, via providing them data in the form of observations & real-world interactions. This technology is known as machine learning that trains

machines so that it can handle novel conditions via self-training, analysis, experience & observation. By developing a medical diagnosis system that is based on machine learning for prediction of heart disease will provide more accurate consequences than traditional way & will also cut down the expenditure of treatment. Heart disease has now become the leading cause of mortality in India. Therefore it is very important to research in this field. To save the lives of people, we need to meet an accurate result.

Deep learning provides accurate result than the machine learning and other data mining algorithm. Deep learning has not yet been used in heart disease prediction. The limited dataset has been used till now by the researchers which were obtained from the same source. This is the key issue in the prediction of heart disease. That's why more research is required in this field. Using different data sets from other sources and a huge amount of data may be useful for improving the performance of the prediction system.

V. EXPERIMENTAL ANALYSIS

In most of the research of heart disease prediction, the Cleveland and Statlog dataset is mainly used. The steps of the experiment are shown in Fig. 8.

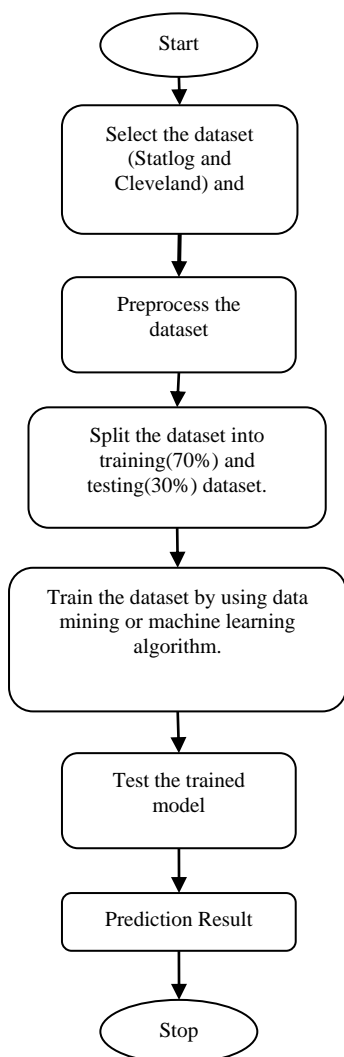


Fig.8. Flowchart of the Experimental setup

For this experiment, RStudio is utilized with the R programming language. The overall comparison of accuracy is visualized as below Fig.9.

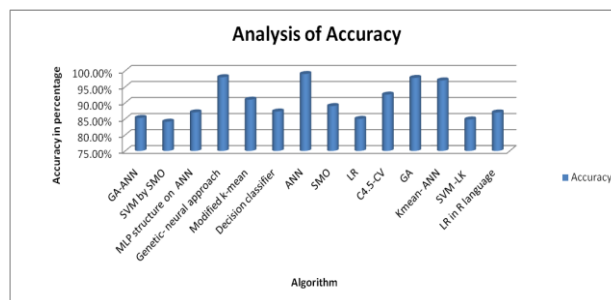


Fig.9. Accuracy Analysis

VI. CONCLUSION

Heart disease is a chronic disease due to which a number of people are suffering which has become a great deal of attention. In this study, numerous data mining & machine learning methods have discussed and examined for predicting heart disease. It analyses and concludes that the data mining algorithm i.e. ANN and SVM that used UCI Heart disease is a chronic disease due to which a number of people are suffering which has become a great deal of attention. In this study, numerous data mining & machine learning methods have discussed and examined for predicting heart disease. It analyses and concludes that the data mining algorithm i.e. ANN and SVM that used UCI Repository dataset perform better for heart disease prediction over than the remaining algorithms.

REFERENCES

- [1] Susmitha K, B. Senthikumar, "A Survey on Data Mining Techniques for Prediction of Heart Diseases", *IOSR Journal of Engineering (IOSRJEN)*, Vol. 08, Issue 9 (September, 2018), pp. 22-27.
- [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* ISSN 0973-6107 Volume 10, Number 7 (2017), pp. 2137-2159
- [3] Abhishek Bajpai, and Dr. Sanjiv Sharma. (2018). "BIG DATA ANALYSIS IN HEALTH CARE DOMAIN: A SYSTEMATIC REVIEW." *International Journal of Engineering Technologies and Management Research*, 5(2: SE), pp. 1-8. DOI: 10.5281/zenodo.1195065.
- [4] K.Gomathi, Dr. D. Shanmuga Priyaa, "A Survey On Heart Disease Prediction Using Data Mining Techniques", *International Journal Of Multidisciplinary Educational Research* May 2016, pp. 56-62.
- [5] Amruta Powar, Prof. Dr. Vijay Ghorpade, "Survey On Heart Disease Prediction System Using Data Mining Technique", *International Journal Of Innovations In Engineering Research And Technology [IJIERT]* Oct.-2017, pp. 25-28.
- [6] Rucha Shinde, Sandhya Arjun, Priyanka Patil, Prof. Jaishree Waghmare, "An Intelligent Heart Disease

- Prediction System Using K-Means Clustering and Naïve Bayes Algorithm”, *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 6 (1), 2015, pp. 637-639.
- [7] Mirpouya Mirmozaffari, Alireza Alinezhad, and Azadeh Gilanpour, “Heart Disease Prediction with Data Mining Clustering Algorithms”, *Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE)* Vol. 4, Issue 1 2017 pp. 16-19. DOI: <https://doi.org/10.15242/IJCCIE.DIR1116009>.
- [8] Lakshmeshree J, K Paramesha,” Prediction of Heart Disease Based on Decision Trees”, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* Volume 5 Issue V, May 2017 pp. 943-948.
- [9] Ajay Kumar Mishra, Bikram Kesari Ratha, "Study of the random tree and random Forest data mining algorithms for microarray data analysis, *International journal on advanced electrical and computer engineering (IJAECE)*, ISSN (Print): 2349-932X Volume-3, Issue-4, 2016, pp. 5-7.
- [10] Ms. Shikha Sharma, Mr. Chetan Chauhan, et al., "Improve the Accuracy of ID3 Classifier for Heart Disease Prediction”, *International Journal for Rapid Research in Engineering Technology & Applied Science* ISSN (Online): 2455-4723 Vol. 3 Issue 5 May 2017.
- [11] Wiharto, W., Kusnanto, H., & Herianto, H., “Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease”, *Healthcare Informatics Research*, (2016) Vol. 22, pp. 186-195. DOI: 10.4258/hir.2016.22.3.186
- [12] 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* ISSN 0973-6107 Volume 10, Number 7 (2017), pp. 2137-2159
- [13] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, “A Data Mining Approach For Prediction Of Heart Disease Using Neural Network”, *International Journal of Computer Engineering and Technology (IJCET)* Volume 3, Issue 3, October - December (2012), pp. 30-40.
- [14] Amandeep Kaur, Jyoti Arora, “Heart Disease Prediction Using Data Mining Techniques: A Survey”, *International Journal of Advanced Research in Computer Science*, Volume 9, No. 2, March-April 2018, pp. 569-572. DOI: <http://dx.doi.org/10.26483/ijarcs.v9i2.5872>.
- [15] Paria Soleimani, Arezoo Neshati, “Applying the Regression Technique for Prediction of the Acute Heart Attack”, *International Journal of Biomedical and Biological Engineering* Vol:9, No:11, 2015, pp 767-771.
- [16] Xiao Liu, “A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method”, *Hindawi Computational and Mathematical Methods in Medicine* Volume 2017. DOI: <https://doi.org/10.1155/2017/8272091>.
- [17] Fen Miao, Yun-Peng Cai, Yu-Xiao Zhang, Xiao-Mao, "Predictive Modeling of Hospital Mortality for Patients with Heart Failure by Using an Improved Random Survival Forest”, *IEEE* Volume 6, 2018, pp. 7244-7253. DOI: 10.1109/ACCESS.2018.2789898
- [18] R. Chitra, V. Seenivasagam, "Knowledge Discovery from Heart Disease Dataset Using Optimized Neural Network”, *Springer International Publishing Switzerland* 2013, pp. 423-431. DOI: https://doi.org/10.1007/978-3-319-03844-5_43.
- [19] Mohammad A. M. Abushariah, Ali Alqudah, Omar Adwan, “Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches”, *Journal of Software Engineering and Applications*, 2014, pp. 1055-1064. DOI: 10.4236/jsea.2014.712093.
- [20] Sheetal Sonawane, Deepti Vadicherla,” Decision Support System For Heart Disease Based On Sequential Minimal Optimization In Support Vector Machine”, *International Journal of Engineering Sciences & Emerging Technologies*, Feb. 2013, Volume 4, Issue 2, pp. 19-26.
- [21] Nilakshi P. Waghulde, Nilima P. Patil., “Genetic Neural Approach for Heart Disease Prediction”, *International Journal of Advanced Computer Research*, Volume-4 Number-3 Issue-16, September 2014, pp. 778-783.
- [22] Ritika, Chadha, Mayank, Shubhankar, “Prediction of heart disease using data mining techniques”, Springer 2016. DOI: <https://doi.org/10.1007/s40012-016-0121-0>.
- [23] Hidayat TAKCI," Improvement of heart attack prediction by the feature selection methods", *Turkish Journal of Electrical Engineering & Computer Sciences* 2018, pp. 1-10. DOI:10.3906/elk-1611-235.
- [24] Rajesh Jangade, Ritu Chauhan, Ruchita Rekapally,” Classification Model for Prediction of Heart Disease”, *Springer Nature Singapore* Pte Ltd. 2018, pp. 707-714. DOI: https://doi.org/10.1007/978-981-10-5699-4_67.
- [25] E. K. Hashi, M. S. U. Zaman, and M. R. Hasan, "An expert clinical decision support system to predict disease using classification techniques,” 2017 Int. Conf. Electr. Comput. Commun. Eng., pp. 396–400, 2017.
- [26] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction, "2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2016.
- [27] A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease(CAD) patients using optimized SVM," *Comput. Methods Programs Biomed.*, vol. 138, pp. 117–126, 2017. DOI: 10.1155/2018/2016282.
- [28] Kaan Uyar, Ahmet İlhan, "Diagnosis of heart disease using a genetic algorithm based trained recurrent fuzzy neural networks”, *9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW, Budapest, Hungary*, Volume 120, August 2017, pp. 588-593. DOI: <https://doi.org/10.1016/j.procs.2017.11.283>.
- [29] Ashok Kumar Dwivedi, “Performance evaluation of different machine learning techniques for prediction of heart disease”, *Springer, The Natural Computing Applications Forum* 17 September 2016, Volume 29, Issue 10, pp. 685–693. DOI: <https://doi.org/10.1007/s00521-016-2604-1>.
- [30] R. Kannan, V. Vasanthi, “Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease”, *Springer Briefs in Forensic and Medical Bioinformatics*, 14 June 2018, pp 63-72. DOI: https://doi.org/10.1007/978-981-13-0059-2_8.
- [31] Sairabi H. Mujawar1, P. R. Devale, "Prediction of Heart Disease using Modified K-means and by using Naive Bayes”, *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)* Vol. 3, Issue 10, October 2015, pp. 10265-10273. DOI: 10.15680/IJIRCCE.2015.0310123.
- [32] Amita Malav, Kalyani Kadam, Pooja Kamat, “Prediction

Of Heart Disease Using K-Means And Artificial Neural Network as Hybrid Approach to Improve Accuracy”, *International Journal of Engineering and Technology (IJET)*, Vol. 9 August 2017, pp. 3081-3085. DOI: 10.21817/ijet/2017/v9i4/170904101.

- [33] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, “Efficient Heart Disease Prediction System”, 2016, pp.962-969. DOI: <https://doi.org/10.1016/j.procs.2016.05.288>.
- [34] Vikas Chaurasia, Saurabh Pal, “Early Prediction of Heart Diseases Using Data Mining Techniques”, *Caribbean Journal of Science and Technology*, Vol. 1 December 2013, pp. 208-217.



Sanjiv Sharma, he works as an assistant professor in the Department of Computer science Engineering and Information Technology in Madhav Institute of Technology and Science, Gwalior. He has 12 years of teaching and research experience. He received his B.E from Pt. Ravishankar University, Raipur, and M.Tech. from RGPV, Bhopal, and Ph. D for Bansthali Vidhya Peeth, Jaipur. His area of research is Network security, Data Mining, and Social Network Analysis.

Authors' Profiles



Yogita Solanki was born on June 5, 1994. She received the B.E degree in Computer Science and Engineering from Indian Institute of Technology and Management (ITM), Gwalior, Madhya Pradesh, India in the year of 2017. She is pursuing M. Tech in Computer Science and Engineering from Madhav Institute of Technology and Science (MITS), Gwalior, Madhya Pradesh, India.

How to cite this paper: Yogita Solanki, Sanjiv Sharma, "A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches", *International Journal of Information Engineering and Electronic Business (IJIEEB)*, Vol.11, No.4, pp. 43-51, 2019. DOI: 10.5815/ijieeb.2019.04.05