

# Impact Of Classification Algorithms For The Prediction Of Heart Disease: A Survey

N. Jagadeesan<sup>1</sup> and T. Velmurugan<sup>2\*</sup>

<sup>1</sup>) Department of Computer Applications (B.C.A), D.G.Vaishnav College,  
Arumbakkam, Chennai - 600106, India.

<sup>2</sup>) PG and Research Department of Computer Science, D.G.Vaishnav College,  
Arumbakkam, Chennai - 600106, India.

**Abstract.** Nowadays, modern technology creates a high impact in all fields of human life. Particularly, a number of recent improvements in the area of medical diagnosis utilized for the prediction of diseases in the available medical records. Heart disease is the number one disease, which happens frequently in all kinds of humans. The researchers suggest many methods to predict and classify heart disease in their research work. The role of computer techniques for the prediction of heart disease is very high and more research work done by many researchers. Different methods of data mining techniques applied for the identification of various diseases. This survey work is carried out to analyze such techniques used for the identification of heart disease which is carried out by different persons in their work. In particular, this work will explore the use of classification algorithms for the prediction and identification of heart disease. At the end of this work, the best classification method is identified from different researchers' perspectives in healthcare diligence. Also, it is suggested the best method for diagnosing heart disease in the future.

**Keywords;** Heart Disease Prediction, Data Mining Techniques, Classification Algorithms, Healthcare Data Analysis

## 1. Introduction:

In the digital era, an enormous volume of data is produced from numerous sources. Data may be present in a structured or unstructured form. Data mining is the method of separating important examples and information from a colossal measure of information in the database, data warehouses, web, or data stored in different information depositories. It is indispensable in many fields of studies to ascertain hidden information from gigantic datasets that help investors to understand and salvage their data within a short period. Different data mining techniques are used to classify, predict and cluster data to make specific or accurate decision-making in various officialdoms.

---

\* <sup>1</sup>) Assistant Professor, jaga.dgvit@gmail.com; <sup>2</sup>) Associate Professor, velmurugan\_dgvc@yahoo.co.in

Received: Feb. 8, 2021; Revised: Feb. 26, 2021; Accepted: Mar. 31, 2021

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2021. Journal of Industrial Information Technology and Application (JIITA)

Classification is one of the most important subroutines in machine learning and data mining tasks. In the healthcare field (hospitals, or other clinical centers) using data mining techniques helps to treat patients to identify if the individual has diseases or not and for early automatic diagnosis of patients from their diseases within the result recover in a short time[1].

Heart disease is the prime cause of death nowadays. Blood Pressure (BP), cholesterol, pulse rate are the key reason for heart disease. Some non-modifiable factors are also there. Smoking and drinking is the reasons for a heart ailment in the human body [2]. The heart is an operating system of our human body. If the heart not functioning correctly means, it will affect other parts of the human body. Some risk factors of heart disease are Cholesterol, High BP, Age, Family history, Poor diet, and Smoking. When blood vessels are overstretched, the threat level of the blood vessels is increased. BP is typically measured in terms of systolic and diastolic. Systolic specify the pressure in the arteries when the heart muscle contracts and diastolic specifies the pressure in the arteries when the heart muscle is in a resting state [3]. Increased level of lipid or fat in the blood leads to heart disease. The lipids are in the arteries hence the arteries become narrow and blood flow becomes slow. Age is the non-modifiable risk aspect which also a reason for heart disease. Smoking is the reason for 40% of the death of heart diseases. Because it limits the oxygen level in the blood then it mutilation and shrinks the blood vessels.

**Signs of a Heart Attack:** Symptoms of an attack square measure Distress, pressure, heaviness, or then again torment inside the chest, arm, or beneath the sternum. Nervousness consuming at back, mandible, pharynx, or arm. Totality, manifestation, or stifling inclination may desire heartburn). A number of the common symptoms square measure sweating, nausea, vomiting, or symptom that conjointly includes extreme weakness, anxiety, or shortness of breath, speedy or asymmetrical heartbeat [4].

**Various types of cardiovascular diseases:** Heart disease encompasses a good vary of vessel issues. Many diseases and conditions constitute the umbrella of cardiovascular disease. Some sorts of Heart diseases square measure Coronary Artery Disease (CAD): CAD is the most common type of heart disease across the world. It is a condition in which plaque deposits block the coronary blood vessels leading to a reduced supply of blood and oxygen to the heart. Arrhythmia: Arrhythmia is a heart rhythm irregularity. The heartbeat can be slow, fast, or irregular. The unusual heartbeats are resulted by a short circuit in the heart's electrical system. Congenital heart disease: It also known as congenital heart anomaly is a defect in the structure of the heart or great vessels that is present at birth Congestive heart failure: It is an ailment where the heart can't siphon adequate blood to the remainder of the body. It is generally known as heart failure. Cardiomyopathy: It is the wearying of the heart muscle or an alteration in the structure of the muscle due to poor heart pumping. Persistent of cardiomyopathy are hypertension, alcohol consumption, viral infections, and genetic defects.

Classification is a supervised Machine Learning technique that consigns labels or classes to not identical objects or clusters. Classification is a two-step method First step

is a model building which is well-defined as the examination of the training records of a database. The second step is the model used in the built model is used for classification. The classification accurateness is assessed by the percentage of test samples or records that are acceptably classified. The Classification has been successfully applied to a wide range of application areas, such as scientific experiments, medical diagnosis, meteorological conditions prediction, credit approval, consumer segmentation, text mining, web mining, target marketing, and swindle exposure.

The organization of this paper is structured as follows. Section 2 discusses the Methods for the Prediction of Cardio Vascular Disease. Analysis Data Mining Techniques for the Healthcare Data are explored in section 3. Section 4 elaborated the Role of Classification Algorithms in Heart Disease Prediction finally; section 5 concludes the research work.

## 2. Methods for the Prediction of Cardio Vascular Disease

To diagnosis heart disease, several tests are used. The procedure for diagnosis will begin with the taking of the patient's Medicinal antiquity, documenting the present and previous symptoms, laboratory testing and an electrocardiogram. Based on the examination and test results, the patient may be asked to go for Supplementary checks. Some of these tests are non-invasive, which indicates the insertion of no structures into the body. Other tests are invasive and involve structures to be inserted into the body.

**Laboratory Tests:** Laboratory tests include blood tests to monitor the risk of heart failure and to determine other parts of the body that can affect the cardiac health of a patient.

**Blood Tests for Heart Disease [5]:** Total cholesterol is the entire amount of cholesterol in the blood. Low-density lipoprotein (LDL) cholesterol is occasionally called "bad" cholesterol. Too much of it reasons fat to build up in arteries, which reduces blood flow. This condition may lead to a heart attack or stroke. High-density lipoprotein (HDL) cholesterol is occasionally called "good" cholesterol. It supports reducing LDL cholesterol and clear arteries. C-receptive protein (CRP) liver produces C-receptive protein (CRP) as a feature of the human body's reaction to harm or virus. Infection assumes a predominant part throughout atherosclerosis, wherein greasy stores stop up in corridors. CRP test results pooled with other blood test results and danger factors for coronary illness. Homocysteine is used by the human body to make protein and to grow and sustain tissue. Whatever it may be, an excess of homocysteine may increase the risk of heart disease and stroke. Homocysteine is commonly recommended for individuals who are at high risk of growing heart disease or have a known ancient history of heart disease.

**Non-Invasive Tests:** Electrocardiogram (ECG) An ECG is a realistic proportion of the electrical movement in your heart. There are exact plans on the ECG that the clinician searches for to characterize whether there are distortions, for example, atrial fibrillation (an anomalous beat), or new or past coronary episode. Over the span of the test, an electrocardiograph instrument records the data. The patient will be joined to the

electrocardiograph by stickers on the chest that are associated with wires prompting the instrument. Stress Echocardiogram Stress tests are executed to perceive how the heart is working under actual pressure. The Cardiovascular can be pushed with an exercise on a treadmill or in a couple of activities, a bike. On the off chance that a patient can't exercise on a treadmill or bike, meds can be utilized to expand the pulse misleadingly, pretending ordinary reactions of the heart to exercise. During the pressure test, the patient will wear EKG leads and wires while working out so the electrical indications of the heart can be archived simultaneously.

The patient's pulse is seen throughout the test. Carotid Ultrasound is done to ascertain the danger of stroke. The sonographer pinions the transducer delicately against the sides of the patient's neck, which sends photos of your corridors to a PC screen for the expert to see. The expert notices the patient's blood course through the carotid veins on the two sides of the neck to investigate for stenosis. Nuclear Stress Test Nuclear pressure tests have two segments to them: a treadmill (or synthetic) stress test and an investigation of the heart after immunization of a radionuclide substance. This substance has been utilized securely for quite a while to close the measure of blood the heart muscle is getting during suspension and stress. The testing is finished with an atomic camera. An echocardiogram ("echo") "echo" is an ultrasound of the heart. A little test like a mouthpiece, called a transducer, is situated on the chest in a few spots. The ultrasound waves sent by the transducer spring back the different pieces of the heart. A PC in the instrument controls the time it takes for the sound wave to get back to the transducer and produces a picture with the data.

Holter Monitor a Holter screen is a minuscule, advantageous instrument that patients can wear for 24 to 48 hours. It engages relentless account of ECG as the patient is doing his day by day exercises. The patient is will be mentioned to keep a journal log of his/her exercises and manifestations. This screen can see arrhythmias that probably won't appear on a resting ECG that solitary register for a couple of moments. Event Recorder An event recorder (loop recorder) is a tiny, handy transtelephonic monitor that may be worn for many weeks. This type of recorder is worthy for patients whose signs are infrequent. The monitor 'loops' two- to five-minute footage into its memory which is constantly overwritten. When the patient experiences the signs he must press the 'record' button on the monitor, which stores an associating strip of ECG.

**Invasive Tests:** Cardiac Catheterization and Coronary Angiography Cardiac catheterization is a typical system that can help examine coronary illness. In certain conditions, catheterization is likewise used to fix coronary illness by opening blocked veins with swell angioplasty and stent settlement. Electrophysiology (EP) Study an EP study is a recording of the electrical motion of the heart. This test helps your clinician define the cause of heart rhythm disruption (arrhythmia) and the best treatment. During the test, the physician may safely replicate the arrhythmia, and then give certain prescriptions to see which one controls it best. An EP study is executed in the Electrophysiology Laboratory. Recording the heart's electrical signals, pacing the heart to bring on certain abnormal rhythms, and ablation are three important stages in an EP study.

### 3. Data Mining Techniques for the Healthcare Data Analysis

Data mining methods such as association, classification, and clustering are utilized by medical services association to expand their capacity for building suitable ends with respect to quiet wellbeing [6]. Clustering is totally different from classification; it doesn't have predefined categories. oversized info is divided into a variety of tiny subgroups known as clusters. It divides the info supported similarities it has. Agglomeration algorithms discover collections of the info specified objects within the same cluster are additional a twin of alternative one another [8]. Tapia et al. examined the organic phenomenon knowledge with the support of the gradable agglomeration approach by employing a genetic formula [9].

Association likewise has an incredible effect in the medical services industry to find the connections between infections, condition of human wellbeing, and the manifestations of the illness. Ji et al. utilized relationship to learn unprecedented easygoing connections in the Electronic wellbeing information base [9]. An incorporated methodology of utilizing Association and Classification procedures likewise improved the capacities of Data Mining Soni et al., et al., have utilized this coordinated methodology of affiliation and characterization for considering medical care information. This incorporated methodology is helpful for deciding guidelines in the information base and afterward by utilizing these standards, a compelling classifier is raised. The examination made an analysis on the information of heart patients and produces rules by weighted acquainted classifier [10]. In this way, Association likewise has a sufficient impact in the medical care field to distinguish the connections among different sicknesses, condition of human wellbeing, and the indications of the illness

Classification is the uncovering of an extrapolative intelligence work that classifies an information thing into one of the various predefined classes [11]. The related work in Classification will be thought in the resulting areas. The Mahalanobis-Taguchi System (MTS) method is by and large generally applied in multivariable factual examination. The Mahalanobis distance (MD) is utilized to construct measurable choices to separate one gathering from another and the Mahalanobis Space (MS) is utilized to address the measure of anomaly of understandings from the notable reference assortment. In the measurable classifiers, the researchers Su et al. [12], have utilized the Mahalanobis Taguchi System (MTS) to build up the gauge model for pressure bruises. The class imbalance complexities are a lot of transcendent in the medical care datasets. Utilization of the information mining strategies is regularly adorned with abnormal spreading when utilizing warped or imbalanced informational indexes. This issue often prompts the affinity of building exceptionally extrapolative order rightness over the lion's share class and denied accuracy over the minority class. Having such a nature to separate the measure of abnormality of translations, this procedure would be a decent strategy to do the appraisal on the genuine informational index pressure ulcers.

Linear discriminant examination (LDA) is broadly utilized in the discriminant investigation to imagine the class dependent on a given arrangement of measurements on new unlabeled explanations [11]. Armañanzas et al. [13] and Jen et al [14] have utilized the straight discriminant investigation in their relating work. Jen et al [14] had

the calculation in foreseeing the cruelty execution of Parkinson's infection patients utilizing scores of non-motor manifestations. Their investigation is expected to quantitatively examine the internal connections between both engine and non-motor indications. The direct discriminant investigation is the contingent likelihood thickness capacity of the indicators follow a typical circulation dependent on the given class esteem. The calculation's capacity to catch factual conditions among the indicator factors shows that this calculation would be appropriate to investigate the direct imperative of this examination to find the collaboration among engine and nonmotor manifestations. The proposed model got an exactness assessment of 69% contrasted with different calculations since the calculation's exhibition increments fundamentally when the conditions are in straight structure. In light of a similar sort of the calculation the researchers Armañanzas et al. [13], utilized the calculation to assess the order exactness to look for the most generous danger factor and build up the underlying arrangement of significant danger factors for ongoing sickness early admonition. From the consequences of the two works, we can securely say that the calculation has great outcomes and it is reasonable to be used to distinguish critical exactness if the connections of the medical care information are in straight structure.

Researchers Armañanzas et al. [13] and Bandyopadhyay et al. [24] have utilized the Bayesian classifier strategy in their particular prescient model. The Bayesian classifiers is notable for its computational productive and capacity to deal with missing information normally and proficiently. Having this favorable position both the researchers have recorded great expectation precision from the models planned separately. By having the models executed the Bayesian classifier likewise demonstrates that the model is appropriate since the averaging approach has prompted improved expectation precision and permits creators to extricate more highlights from the information without being over fitting. This technique would be a decent methodology if there informational collections are experiencing missing information.

The Support Vector Machine (SVM) is demonstrated to be invaluable in taking care of classification errands with great speculation execution. The strategy looks to limit the upper bound of the speculation mistake dependent on the underlying danger minimization guideline. The SVM preparing is comparable to tackle a direct compelled quadratic programming issue [25]. The technique is regularly utilized in clinical determination. Researchers García-Laencina et al. [21], Zheng et al. [25], Kang et al. [26], and Su et al. [12] have utilized the strategy in their model in clinical judgments. A portion of the researchers have utilized the SVM strategy for similar examination purposes. The SVM strategy speculation capacity is constrained by two unique factors, that is the preparation mistake and the limit of the learning machine estimated. The preparation blunder rate can be constrained by changing the highlights in the classifiers. The outcomes got from the investigations, obviously show that the SVM demonstrated more noteworthy execution since it maps the highlights to higher dimensional space.

The researchers Yeh et al. [18], Fei [19], and Abdi and Giveki [20] have utilized the multitude insight strategy to plan their conclusion model. The algorithm particle swarm optimization (PSO) can proficiently locate the ideal or close to ideal arrangements in huge hunt spaces. All the three researchers attempted to determine optimization issue

which regularly includes highlights in the arrangement issues. The order cycle will be quicker and more precise if less number of highlights are utilized. From the work considered, the PSO based methodology demonstrates to improve the general grouping results since PSO is being utilized to choose reasonable boundaries in the elaborate classifiers. A few examinations have investigated the Decision Tree (DT) technique to dissect clinical information. The researchers Sharma and Om [15], Wang et al. [16] and Zolbanin et al. [17] have utilized the DT algorithm in their separate work. Having the nature to look at information and make the tree and its standards are utilized to make an expectation. All the three works have utilized the DT to the informational index to improve the prognostic presentation, as far as exactness. The idea of the informational index utilized in this examination is somewhat adjusted arrangement of informational index. From the near of the works, we infer that DT as can't be utilized in proposing prognostic choice to tackle imbalanced issues on the grounds that the DT recursively separate perceptions into branches to develop a tree.

Researchers García-Laencina et al. [21], Armañanzas et al.[13], Jen et al.[22], Bagui et al.[22], and aşahan et al. [23] have utilized the k-Nearest Neighbor in their separate prescient models. The k-Nearest Neighbor is an occurrence based classifier strategy. The boundary units comprise of tests that are utilized in the strategy and this algorithm at that point accepts that all cases identify with the focuses in the N-dimensional space RN. The algorithm is practical as the data in the preparation information is rarely lost. Notwithstanding, this calculation would be appropriate if the preparation informational index is enormous as this calculation is exceptionally tedious when every one of the example in preparing set is handled while ordering another information and this interaction requires a more extended arrangement time. From the work by the referenced creators, the grouping exactness is the thing that they might want to achieve rather than order time as the characterization precision is more significant in the clinical analysis.

### 3.1 Summary for Diagnosis in Healthcare

In terms of prediction and decision making, Data mining techniques have substantial expansion in medical industry with respect to various diseases like diabetes, heart disease, liver diseases, cancer and others. Table 1 summarizes the medical data mining, its techniques used and for the related diseases.

Table 1: Summary of medical data Analysis

Article Ref. No.	Data Mining Techniques used	Disease Predicted
27	Prediction models using DT Algorithms such as C4.5, C5, and CART	Coronary heart disease
28,29	Distinguish disease subtypes using Ensemble approach	Lymphoma Disease and Lung Cancer
30	Predicate the probability of a psychiatric patient on the basis detected symptoms using BBN Bayesian networks	Psychiatric Diseases
31	Identify frequency of diseases in particular Geographical area using Apriori algorithm	Frequent Disease

31	Diagnosing Cardio Vascular Diseases using Classification Algorithm	Cardio Vascular Diseases
32	Classification using Bayesian Ying Yang (BYY)	Liver diseases.
7,33	Categorization of skin disease using integrated tree model with neural network classification methods	Skin Disease
34	Classification of Medical Data using Genetic Algorithm	Diabetes
35	Constructed model using Artificial Neural Network	Chest Disease
36	Classification of Disease using k-Nearest Neighbor	Diabetes, Cancer
37,42	Improving classification accuracy using Naive	Coronary Heart Disease
38	Prediction of Diseases Using Apriori Algorithm	Chronic Disease
39	Disease classification using Support Vector Machine	Diabetes
40	Accurate Classification of medical data using K- Self- Organizing Map(SOP) and Naïve Bayes	Breast Cancer
42	Familiarized an adaptive Fuzzy K-NN approach for diagnosing the disease	Parkinson Disease

#### 4. Role of Classification Algorithms in Heart Disease Prediction

T. S. Brisimi et al. [43] articulated his elucidation for heart disease forecast using SVM, sparse logistic regression, and random forests. In order to afford a clarification for the identified problem, the researchers suggested two novel methods: K-LRT, a likelihood-ratio-based method, and Joint Clustering and Classification (JCC) method to classify the unknown patient clusters. Each of these approaches adapts different classifiers to each recognized cluster. After an investigational study, in terms of average forecast precision, the random forests technique (81.62%) was predicted as the finest technique followed by the ACC method (77.06%). J. Zhang et al [44] projected a system that makes use of Fast Fourier Transformation and an ensemble model. The correctness of the projected model was 87.00%, 93.00%, and 94.83% for the three-feature set, six-feature set, and eight-feature set correspondingly. This result was equaled with three techniques: Neural Networks, Least Square-Support Vector Machine (LS-SVM), and Naive Bayes which showed less accuracy when unrushed against the suggested system

R. G. Saboji [45] proposed a mountable elucidation for heart ailment forecast using random forest technique. The projected elucidation produced the prediction precision rates as 88%, 96%, and 98% for the sum of records 200, 400, and 600 individually. The result was related to the Naive Bayes technique which produced 44%, 55%, and 64% forecast precision rates for 200, 400, and 600 records correspondingly. From the appraisal result, it is investigated that, though random forest technique outclassed naive Bayes.

S. Pouriyeh et al. [46] created a meticulous examination and comparison on data processing and machine learning techniques that are used for predicting the likelihood of heart diseases. the various varieties of data processing and machine learning techniques used for comparison includes call Tree (DT), Naive mathematician (NB), Multi-Layer Perceptron (MLP), K-Nearest Neighbor (K-NN), Single Conjunctive Rule Learner (SCRL), Radial Basis perform (RBF) and Support Vector Machine (SVM). Among that the DT achieved 77.55% accuracy, NB 83.49%, K-NN 83.16%, MLP 82.83%, RBF 83.82%, SCRL 69.96% and SVM 84.16%. From the results, it absolutely was clear that SVM achieved the best forecast preciseness rate. As a terminal experiment, the researchers applied the pillar technique on the uniting of fixed



classifiers, which ends up in that SVM + MLP fashioned the most effective preciseness rate of 84.15%.

Tulay Karayilan and Ozkan Kilic [47] established a heart disease forecast system that uses Artificial Neural Network's Backpropagation Procedure in which the Principal Component Analysis (PCA) method was executed with reduced the number of neurons at the input layer from 13 to 8 neurons. The generated results showed that the proposed technique outperformed the other techniques with a 95.56% prediction accuracy rate. K. Pahwa and R. Kumar [48] suggested the hybrid approach using Naive Bayes and Random Forest techniques to forecast the incidence of heart disease. The projected framework had chosen the finest features to forecast heart disease by applying the SVM-Recursive Feature Elimination (SVM-RFE) procedure and gain ratio. In the random method, 10 features were nominated for correctness prediction using both the Naive Bayes and Random Forest methods. The results exhibited that forecast precision was 78.22% for Naive Bayes and 76.90% for Random Forest. In the amalgam method, prediction correctness is determined for various numbers of features in the order of 12, 10, 8, and 6. As a consequence, Naive Bayes reached the uppermost forecast correctness of 84.16% for ten best features, whereas the Random Forest method reached only 82.84% for the same ten feature set. Amusingly, the correctness forecast was uppermost in the Random Forest technique with 84.16% for 12 best features compared to that of Naive Bayes with 83.83%.

Abhishek Rairikar et al. [49] engaged three methods namely k -nearest neighbors algorithm (KNN), Decision Trees (DT), and Naive Bayes to examine the medical dataset. The researchers anticipated that KNN was very exact and quicker related to the other two methods but they do not postulate the statistical values of correctness. Shahid Anzarus Sabab et al.[50] applied three controlled learning methods: C4.5 DT, SVM, and Naive Bayes. Initially, their prediction accuracy was relatively poor when compared to preceding works in the akin field. So the researchers applied a ranker algorithm for feature selection to confiscate the inessential attributes. The execution outcomes exposed that the forecast correctness was enriched by 3.63% with eight finest traits for Naive Bayes, 4.95% for SVM with nine finest traits, and 5.94% for DT with five 5 finest traits. Thus the leading exactness was given by SVM (87.79%), followed by Naive Bayes (86.60%) and DT (79.87%) correspondingly.

M. A. Jabbar and S. Samreen [51] make use of the Hidden Naive Bayes (HNB) model for extrapolation of heart disease. The researchers applied pre-processing filter discretization and Inter-Quartile Range (IQR) filter to split the data set into quartiles. The suggested HNB + IQR model attained 100% accurateness in forecasting heart disease related to other associated works that used Naive Bayes. J. Singh et al. (2016)[52] established a structure for the forecast of heart ailments using an amalgam method that encompasses classifiers such as Naive Bayes, J48 Decision Tree, ZeroR, OneR, and k-Nearest Neighbour (IBk). Investigational study displays that the forecast of correctness is 99.19% for IBk, which is the uppermost rate amongst the other classifiers while using Apriori association on heart ailments dataset S. Radhimeenakshi [53] projected a model to envisage the early identification of heart maladies using SVM and ANN methods. The performance assessment of SVM has shown 84.7%. M.A.Jabbar et al. [54] established a smart heart ailment forecasting method using random forest and evolutionary tactics. The researchers make use of chi-square and

genetic algorithm as feature selection procedures on heart ailment datasets the suggested system produced 100% forecast accurateness on heart ailment

Zriqat et al. [55] made a benchmarking study on five sorts of classification algorithms: Naive Bayes, DT, Discriminant Analysis, Random Forest, and SVM to forecast the manifestation of heart ailments. After investigational scrutiny, the researchers presented that decision trees produced the uppermost forecast correctness of 99%. S. Bashir et al. [56] anticipated a resolution support structure for intellectual heart ailment analysis using the popular vote based novel classifier collaborative based on assorted classifiers namely Naive Bayes, DT based on Gini Index (DT-Gini), and SVM. The investigational study exhibited that the anticipated collaborative structure achieved the finest precision of 81.82%. A comparative analysis of heart disease prediction methods are elaborately discussed in the table 2. In which it is given a in depth analysis of using some of the classification algorithms for the prediction of heart disease via its accuracy of the algorithms.

Table 2: Comparative Analysis of Heart Disease Prediction.

Paper Ref. No.	Name of the Authors	Proposed Methods	Prediction Precision Rate
43	T. S. Brisimi et al.	Random Forest	81.62%
44	J. Zhang et al.	Ensemble Model	94.83%
45	R. G. Saboji	Random Forest	98%
46	S. Pouriyeh et al.	SVM + MLP	84.15%
47	T. Karayilan and O. Kilic	Neural Network	95.56%
48	K. Pahwa and R. Kumar	NB+Random Forest	84.16%
49	Abhishek Rairikar et al.	KNN	NA*
50	S. A.Sabab et al.	SVM	87.79%,
52	J. Singh et al.	IBk (k Nearest Neighbor)	99.19%
53	S. Radhimeenakshi	SVM	84.7%
54	M.A..Jabbar et al.	Random Forest	100%
55	Zriqat et al.	DT	99%
56	S. Bashir et al.	Ensemble Model	81.82%

From the observed survey based analysis of this research work, it is ascertained that the many methods are prove their efficiency in order to predict the disease. Particularly, this work confirmed that the Random forest method yields better result via its accuracy and predict the disease in an efficient way.

## 5. Conclusion

The innovation of automatic diagnosis in heart disease prediction requires improvement and transformation. Numerous tactics were established to envisage heart disease and assist the cardiologist as well as patent to make a decision. But, still no distinctive technique was established to predict heart disease accurately and as much as earlier in healthcare. This research work scrutinized the different approaches anticipated by various researchers in heart disease prediction. Each category shows its performance luminously. From this survey oriented research work, it is recognized that Random Forest Classification technique yields virtuous precision in order to envisage

heart disease. Hence, any classification technique or hybrid classification technique uses minimal attributes performs well in envisage heart disease compared with existing methods.

## References

- [1] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction” *International Journal of Computer Applications* Vol.17 No.8, 2011, PP. 0975 – 8887, DOI: 10.5120/2237-2860.
- [2] Mirmozaffari, Mirpouya, Alireza Alinezhad, and Azadeh Gilanpour "Data Mining Apriori Algorithm for Heart Disease Prediction." *International Journal of Computing, Communications & Instrumentation Eng. (IJCCIE)*, Vol. 4, Issue 1, 2017 DOI: 10.15242/IJCCIE.DIR1116010.
- [3] Susmitha K, B. Senthil Kumar “A Survey on Data Mining Techniques for Prediction of Heart Diseases”, *IOSR Journal of Engineering (IOSRJEN)*, Vol. 08, Issue 9, 2018, PP.22-27.
- [4] Khaing, Hnin Wint. "Data Mining Based Fragmentation and Prediction of Medical Data", *IEEE International Conference on Computer Research and Development (ICCRD)*, Vol. 2. 2011, pp. 480-485, DOI: 10.1109/ICCRD.2011.5764179.
- [5] Gayathri, P, N. Jaisankar “Comprehensive Study of Heart Disease Diagnosis Using Data Mining and Soft Computing Techniques”, *International Journal of Engineering and Technology (IJET)*, Vol. 5 No 3, 2013, PP.2947-2958.
- [6] Sheetal L. Patil, “Survey of Data Mining Techniques in Healthcare”, *International Research Journal of Innovative Engineering*, Vol. 01 Issue: 09, 2015, PP. 53-60, DOI: 10.5121/ijist.2016.6206.
- [7] C. Hattice & K. Metin, “A Diagnostic Software Tool for Skin Diseases with Basic and Weighted K-NN”, *Journal of Innovations in Intelligent Systems and Applications (INISTA)*, 2012, DOI: 10.1109/INISTA.2012.6246999.
- [8] K.Sharmila & Dr.S.A.Vethamanickam, “Survey on Data Mining Algorithm and Its Application in Healthcare Sector Using Hadoop Platform”, *International Journal of Emerging Technology and Advanced Engineering*, Vol. 05, Issue: 01, 2015, PP. 12-21.
- [9] J.Yanqing & H.Ying & J.Tran & P.Dews & A.Mansour & R.Michael Massanari, “Mining Infrequent Causal Associations in Electronic Health Databases”, 11th IEEE International Conference on Data Mining Workshops, 2011, DOI: 10.5120/21307-4126.
- [10] S. Soni & O. P. Vyas, “Using Associative Classifiers for Predictive Analysis in Health Care Data Mining”, *International Journal of Computer Applications*, Vol. 04, No. 05, 2010, DOI: 10.5120/821-1163.
- [11] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “Knowledge Discovery and Data Mining: Towards a Unifying Framework”. In *Proc. KDD-96, Second International Conference on Knowledge Discovery & Data Mining*, 1996.
- [12] C. T. Su, P. C. Wang, Y. C. Chen, and L. F. Chen, “Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients.” *Journal of Medical Systems.*, Vol. 36, No. 4, 2012, pp. 2387–2399, DOI:10.1016/j.procs.2015.12.145.
- [13] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, “Unveiling Relevant Non-Motor Parkinson’s Disease Severity Symptoms Using A Machine Learning Approach,” *Artificial Intelligence in Medicine*. Vol. 58, No. 3, 2013, PP 195–202, DOI:10.1016/j.artmed.2013.04.002.
- [14] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, and M.-S. Chen, “Application of Classification Techniques on Development an Early-Warning System for Chronic Illnesses”, *Journal of Expert Systems with Applications*, Vol. 39, No. 10, 2012, PP. 8852-8858, DOI:10.1016/j.eswa.2012.02.004.
- [15] N. Sharma and H. Om, “Data Mining Models for Predicting Oral Cancer Survivability” *International Journal of Network Modeling Analysis in Health Informatics and Bioinformatics*, Vol. 2, No. 4, 2013, PP, 285–295, DOI: 10.1016/j.procs.2015.12.145.
- [16] K.-J. Wang, B. Makond, and K.-M. Wang, “An Improved Survivability Prognosis of Breast Cancer by Using Sampling and Feature Selection Technique to Solve Imbalanced Patient Classification Data.”

- International Journal of BMC Medical Informatics and Decision Making, Vol. 13, 2013, DOI: 10.1186/1472-6947-13-124.
- [17] H. M. Zolbanin, D. Delen, and A. Hassan Zadeh, "Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach", *Journal of Decision Support System*, Vol. 74, 2015, PP. 150-161, DOI:10.1016/j.dss.2015.04.003.
- [18] W.-C. Yeh, W.-W. Chang, and Y. Y. Chung, "A New Hybrid Approach For Mining Breast Cancer Pattern Using Discrete Particle Swarm Optimization And Statistical Method", *Journal of Expert Systems with Applications*, Vol. 36, No. 4, 2009, PP. 8204-8211, DOI: 10.1016/j.eswa.2008.10.004.
- [19] S. W. Fei, "Diagnostic Study on Arrhythmia Cordis Based On Particle Swarm Optimization - Based Support Vector Machine", *Journal of Expert Systems with Applications*, Vol. 37, No. 10, 2010, PP. 6748-6752, DOI: 10.1016/j.eswa.2010.02.126.
- [20] M. J. Abdi and D. Giveki, "Automatic Detection Of Erythematous-Squamous Diseases Using PSO-SVM Based on Association Rules," *International Journal of Engineering Applications on Artificial Intelligence*, Vol. 26, No. 1, 2013, DOI: 10.1016/j.eswa.2010.02.
- [21] P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonso, "Missing Data Imputation on the 5-Year Survival Prediction of Breast Cancer Patients with Unknown Discrete Values", *Computational. Biological Medicine*, Vol. 59, 2015, DOI: 10.1016/j.combiomed.2015.02.006.
- [22] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast Cancer Detection Using Rank Nearest Neighbor Classification Rules," Vol. 36, 2003, PP. 306-313, DOI: 10.1016/j.procs.2015.12.145.
- [23] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "A New Hybrid Method Based on Fuzzy- Artificial Immune System And K-NN Algorithm for Breast Cancer Diagnosis", *Comput. Biol. Med.*, Vol. 37, No. 3, 2007, PP. 415-423, DOI: 10.1016/j.combiomed.2006.05.003.
- [24] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrissi, P. E. Johnson, and P.J.O'Connor, "Data Mining for Censored Time-To-Event Data: A Bayesian Network Model for Predicting Cardiovascular Risk From Electronic Health Record Data". 2014, DOI: 10.1007/s10618-014-0386-6.
- [25] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast Cancer Diagnosis Based on Feature Extraction Using a Hybrid of K-Means and Support Vector Machine Algorithms", *International Journal of Expert Systems with Applications*, Vol. 41, No. 4, Part 1, 2014, PP. 1476-1482.
- [26] S. Kang, P. Kang, T. Ko, S. Cho, S. Rhee, and K.-S. Yu, "An Efficient and Effective Ensemble of Support Vector Machines for Anti-Diabetic Drug Failure Prediction", *International Journal of Expert Systems with Applications*, Vol. 42, No. 9, 2015, DOI: 10.1016/j.eswa.2015.01.042.
- [27] Doron Shalvi & Nicholas DeClaris, "An Unsupervised Neural Network Approach to Medical Data Mining Techniques", *IEEE*, 1998.
- [28] Gustavo Santos-Garcia & Gonzalo Varela & Nuria Novoa & Marcelo F. Jimenez, "Prediction of Postoperative Morbidity After Lung Resection Using an Artificial Neural Network Ensemble", *International Journal of Artificial Intelligence in Medicine*, Vol. 30, 2004, PP.61-69.
- [29] Hojin Moon & Hongshik Ahn & Ralph Kodell & Songjoon Baek & Chien- Ju Lin & James Chen, "Ensemble Methods for Classification of Patients for Personalized Medicine With High-Dimensional Data". *Artificial Intelligence in Medicine*, Vol. 41, 2007, PP. 197-207, DOI: 10.1016/j.artmed.2007.07.003.
- [30] I. Curiac & G. Vasile & O. Baniias & C. Volosencu & A. Albu, "Bayesian Network Model for Diagnosis of Psychiatric Diseases", *Proceedings of the ITI 2009, 31st Int. Conf. on Information Technology Interfaces*, 2009, DOI: 10.1109/ITI.2009.5196055.
- [31] Ilayaraja & T. Meyyappan, "Mining Medical Data to Identify Frequent Diseases Using Apriori Algorithm", *Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, 2013, DOI: 10.1109/ICPRIME.2013.6496471.
- [32] Jeong-Yon Shim & Lei Xu, "Medical Data Mining Model for Oriental Medicine via BYY Binary Independent Factor Analysis", *Proceedings of the 2003 International Symposium on Circuits and Systems*, 2003. ISCAS '03, 2003, PP. 5-5, DOI: 10.1109/ISCAS.2003.1206413.
- [33] L.Chang & C.H.Chen, "Applying Decision Tree and Neural Network to Increase Quality of Dermatologic Diagnosis", *International Journal of Expert Systems with Applications- Elsevier*, Vol. 36, 2009, PP. 4035-4041.

- [34] Markus Brameier & Wolfgang Banzhaf, "A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining", IEEE Conference on Transactions on Evolutionary Computation, 2001, PP. 17-26, DOI: 10.1109/4235.910462.
- [35] O.Er & N. Yumusak & F. Temurtas, "Chest Diseases Diagnosis Using Artificial Neural Networks", Expert Systems with Applications- Elsevier, Vol. 37, 2010, PP. 7648-7655.
- [36] Ping-Hung Tang & Ming-Hseng Tseng, "Medical Data Mining Using BGA and RGA for Weighting of Features in Fuzzy K-NN Classification", IEEE International Conference on Machine Learning and Cybernetics, PP. 3070-3075, 2009, DOI: 10.1109/ICMLC.2009.5212633.
- [37] Ranjit Abraham & Jay B.Simha & Iyengar, "A Comparative Analysis of Discretization Methods for Medical Data Mining with Naïve Bayesian Classifier", IEEE International Conference on Information Technology (ICIT'06), 2006.
- [38] R.Karthiyayini & J.Jayaprakash, "Association Technique on Prediction of Chronic Diseases Using Apriori Algorithm", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 04, Special Issue 06, 2015.
- [39] Sarojini Balakrishnan & Ramaraj Narayanaswamy, "Feature Selection Using FCBF in Type II Diabetes Databases", Special Issue of the International Journal of the Computer, the Internet and Management, Vol. 17 No. Special Issue 1, 2009.
- [40] Syed Zahid Hassan & Brijesh Verma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases". IEEE Seventh International Conference on Intelligent Systems Design and Applications (ISDA 2007). 2007, PP.503-510, DOI: 10.1109/ISDA.2007.48.
- [41] Tsang-Hsiang Cheng & Chih-Ping Wei & Vincent S. Tseng, "Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches", CBMS '06: Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, PP.165-170, 2006, DOI: 10.1109/CBMS.2006.87.
- [42] W.L.Zuoa & Z.Y.Wanga & T.Liua & H.L.Chenc, "Effective Detection Of Parkinson's Disease Using an Adaptive Fuzzy K-Nearest Neighbour Approach", International Journal of Expert Systems with Applications- Elsevier, 2013, PP. 263-271 DOI: 10.1016/j.eswa.2012.07.014.
- [43] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," Proceedings of the IEEE. Institute of Electrical and Electronics Engineers, Vol. 106, No. 4, 2018. PP. 690-707.
- [44] J. Zhang et al., "Coupling a Fast Fourier Transformation with a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment," in IEEE Access, Vol. 5, 2017, PP.99-112
- [45] R. G. Saboji, "A Scalable Solution for Heart Disease Prediction Using Classification Mining Technique," International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017, PP. 1780-1785.
- [46] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A Comprehensive Investigation and Comparison of Machine Learning Techniques in The Domain of Heart Disease", IEEE Symposium on Computers and Communications (ISCC), 2017, PP. 204-207.
- [47] T. Karayilan and Ö. Kılıç, "Prediction of Heart Disease Using Neural Network," International Conference on Computer Science and Engineering (UBMK), 2017, PP. 719-723, DOI: 10.1109/UBMK.2017.8093512.
- [48] K. Pahwa and R. Kumar, "Prediction Of Heart Disease Using Hybrid Technique for Selecting Features", 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017, PP. 500-504, DOI: 10.1109/UPCON.2017.8251100.
- [49] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale and A. Lamgunde, "Heart Disease Prediction Using Data Mining Techniques", International Conference on Intelligent Computing and Control (I2C2), 2017, PP. 1-8, DOI: 10.1109/I2C2.2017.8321771.
- [50] S. A. Sabab, M. A. R. Munshi, A. I. Pritom and Shihabuzzaman, "Cardiovascular Disease Prognosis Using Effective Classification and Feature Selection Technique", International Conference on Medical Engineering, Health Informatics and Technology (MediTec), 2016.
- [51] M. A. Jabbar and S. Samreen, "Heart Disease Prediction System Based on Hidden Naive Bayes Classifier" International Conference on Circuits, Controls, Communications and Computing (I4C), 2016, PP. 1-5, DOI: 10.1109/CIMCA.2016.8053261.

- [52] J. Singh, A. Kamra and H. Singh, "Prediction of Heart Diseases Using Associative Classification," 5th International Conference on Wireless Networks and Embedded Systems (WECON), 2016, PP. 1-7, DOI: 10.1109/WECON.2016.7993480.
- [53] S. Radhimeenakshi, "Classification and Prediction of Heart Disease Risk Using Data Mining Techniques of Support Vector Machine and Artificial Neural Network", 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, PP. 3107-3111.
- [54] M.A.Jabbar, B.L.Deekshatulu and Priti Chandra, "Intelligent Heart Disease Prediction System Using Random Forest and Evolutionary Approach", Journal of Network and Innovative Computing, Vol.4, 2016, PP. 175-184.
- [55] Isra'a Ahmed Zriqat, Ahmad Mousa Altamimi, Mohammad Azzeh "A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods", International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 12, 2016.
- [56] S. Bashir, U. Qamar and M. Younus Javed, "An Ensemble Based Decision Support Framework for Intelligent Heart Disease Diagnosis", International Conference on Information Society, 2014, PP. 259-264, DOI: 10.1109/i-Society.2014.7009056.