Artificial Neural Networks with Enhanced Nearest Neighbor algorithm for Heart Disease Prediction

J.Jayashree; J.Vijayashree, Vishal Goar* & N.Ch.S.N.Iyengar#
School of Computer Science and Engineering, VIT, Vellore, Tamilnadu
*Government Engineering College, Bikaner 334001, India
#Department of Information Technology, SNIST, Yamnampet, Ghatkesar, Hyderabad (TS),

Abstract
Heart failure is the root cause of the increased mortality rate happening now in the current scenario. Many health care services have been investing in the research forefronts to improvise the predictive methodology of cardio arrests and cardio related diseases to avoid this major health hazard. The improvised predictive methodology by the research forefronts with timely response helps to provide proper medical aid before the condition gets severe. The forecast methods use artificial intelligence which includes the neural networks and fuzzy techniques. The neural network algorithm makes use of the datasets and predicts the future datasets under various conditions. The fuzzy technique implements the data sets performed by the neural algorithms provides automated training to the system which in turn makes the system reliable to device any decisions on their own under any critical circumstances. To contribute in providing the solution, this paper analyzes the invisible medical information of medications in-taken for various medical issues and their effects on the patients along with the patient's medical history are taken as dataset inputs to the classification with association along with the nearest neighbor which process the possibilities and facts that bares the key reason for cardiac arrest. To avoid critical circumstances in the last minute of cardiac arrest, the neural algorithms forecast the cardiac related hazards on time which could help them to survive even in critical condition with proper medication and intensive care. There are 14 Attributes such as gender, cardiac acute type, pressure level, body glucose level, age factor, cholesterol and body fat contents, BMI, medications are the key factors taken as the primary source for the input datasets. The interpreted data sets are analyzed and visualized as patterns. The nearest neighbor algorithm is enhanced for recognizing the patterns and grouping the similarities and dissimilarities which later formed as data subsets for the forecast. To improvise the efficiency of results 14 attributes are taken as input datasets and by the process of neural algorithm, it is then reduced to 6 attributes results 99.1% accuracy with 0.9% of noise figures.

Keyword: Neural Algorithms, Nearest Neighbor, classification with association and back propagation, Cardiac hazards

I. INTRODUCTION

The. Cardiac diseases found to be the Key issue in current medical industry. Cardiac diseases are caused by various reason includes intake food, environment, medications and other Physical & biological impacts. In general, doctor finds the severity of the cardiac diseases by means of various test datasets and also by means of experience in handling the cardiac related issues. A timely detection of cardiac issues resolves the problem of increased mortality rate and also eases the problem for acute level of sufferer by intensive medications. In general, Health hazards related to cardio vascular system arises due to improper health

maintenance and awareness which causes obesity and other physical changes in the body. Obesity is due to improper diet balance and improper exercise plan as it is nowadays a big concern. Heart disease of various types which can be classified into many categories depends upon the personal behavior, Food intake, Physic Fitness, environmental conditions and medications. The severity of cardiac issues is chest acute with migraine, breath discomfort, Excessive sweat, improper pulse & beat rate. The factors of personal behavior include Smoking, Alcohol intake, and un-prescribed medicine intake without proper consultation with the practitioners. The factors of food intake include high cholesterol meals, excessive starch contents, and excessive meat intake. Excessive Cholesterol leads to the formation of sediments on the heart valves which then cause improper contraction and relaxation. This results in improper breathing and cardiac arrest at later stages. Physical fitness includes improper diet balance and poor exercise leads the body somber and impedes obesity. The common cardiac issues are Angina pectoris, infractions on myocardial, congenital heart defects, tachycardia, hypercholesterolemia, Atheroma occurs due to the various factors mentioned above. Classification with association and neural network algorithm utilizes the possibilities, factors and cardiac issues criteria's and process the datasets to match the issues and outputs perfect timely results with the forecast data of how the issue affects the cardio vascular system. Many researchers have worked with many algorithms and various techniques which are examined and described in the literature survey. Section II. The classification with association algorithm along with nearest neighbor is evaluated in the section III and section IV.

II. LITERATURE REVIEW

Nabeer et.al [1] evaluated backpropogation algorithm to forecast the cardiac arrest with 11 attributes and found the system to be accurate by performing the matlab evaluation along with training plot box test. They had performed 11 consecutive runs on the test and each found to be a close to each other between the testing and training patterns.

Veera &Anchana.et.al [2] evaluated the multilayer perceptron along with backpropogation learning algorithm to classify the cardiac diseases with 13 attributes minimized to 8 attributes after algorithmic process and feature extraction found to 1% higher than the previous evaluation and 0.8% higher in training the datasets.

Jagdeep & Amit.et.al [3] evaluated the various algorithms and their classifications for training the system to forecast the cardiac related diseases and classified the results under different circumstances and the accuracy on each algorithm and found the classification association rule found to be accurate with 99% accuracy.

Jabbar et.al [4] evaluated the predictive methodology using nearest neighbor method along with swarm optimization. Selecting the future data set and extracting the feature found to be difficult since it contain the noisy data sets which reduces the accuracy of determination. The elimination of Noise factors includes optimization technique and found 100% accuracy in the prediction of cardiac related issues.

Amin haq et.al [11] proposed an intelligent framework for predicting heart disease using KNN,ANN,SVM,DT and RF. He also used feature selection techniques like relief, mRMR and lasso for reducing the number of features. K-fold validation method was used. The performance of SVM along with mRMR feature selection technique was best compared to other techniques.

Beulah and carolin [12] investigated ensemble technique which combines dofferent classifiers inorder to enhance the accuracy of the classifier. Ensemble techniques such as bagging and boosting improves the accuracy of the classifiers and also predicts the presence of the disease in an efficient manner. In this work a maximum of 7% accuracy was improved using ensemble technique.

Chandrasegar and Aman [13] proposed Fine tune prediction model for identifying most important features which predicts the presence of heart disease. Accuracy of algorithms such as SVM, RF and DT are increased using fine tune prediction model.

Shashikant et.al [14] proposed SVM sequential minimization optimization technique for disease diagnosis. The accuracy of SVM algorithm is less compared to SVM algorithm with optimization technique. The presence and absence of heart disease is predicted accurately using optimization techniques.

Jaymin et.al [15] compared different algorithms of decision tree classification for heart disease prediction. Author compared J48 algorithm, Logistic model tree algorithm and Random Forest algorithm. Author concluded thatJ48 tree technique turned out to be best classifier for heart disease prediction because it contains more accuracy and least total time to build.

Senthilkumar et.al. [16] aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease using the hybrid random forest with a linear model (HRFLM). This model achieves an accuracy of 88.7%.

Sellappan and rafiuq [10] has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naive Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. IHDPS can answer complex ";what if"; queries which traditional decision support systems cannot [17].

III. ANN ASSOCIATION ALGORITHMS ALONG WITH NEAREST NEIGHBOR

Neural Networks are bio-inspired concepts, implemented to design the intelligent process and perform simulation to provide future data subsets along with various statistical results. It acquires the insights between the original and the resultant linkage in the datasets and train

the system with experience by means of various conditional processing. Intelligence System train them with the datasets provided and it determines their own state of output at critical conditions. The Neural network algorithms are classified with their original state and hybrid systems. The neural algorithms are Levenber-marquard, Quasi newton's conjugate gradients, backpropogation with momentum, multilayer perceptron, feed forward system of propagation and various statistical methodology like chi-squared, roc analysis, root cause analysis are used for evaluation.

In Neural algorithms, the association is to find the bond among the datasets and their patterns. In most cases the Associative functions with the support and confidence terms in which the original data set and trained data sets gives perfect results for the system to identify the behavioral nature. The other term of classification associates is attributes. Attributes designates the type of data. Classification, defines the similarity level of the datasets or separation among the group of datasets which are essential for training or identifying the class nature of each and every datasets. Classification defines various methods for the statistical analysis via decision tree, Programming and neural systems and in which the rule system for classification is derived from the association. The associative rule defines the characteristic nature but not in vice versa.

The classification system of neural networks narrows the datasets, which in turn results a fine tuned datasets with the proper classifier. In this the classification with associative is combined with Nearest Neighbor system in which the classification with association defines the fine-tuned datasets and the Nearest Neighbor algorithms trains the datasets with the system to device its intelligence also the various algorithm techniques like quasi newton and conjugate gradients are also compared with the proposed system.

3.1 Input Datasets

There are 14 attributes used for the forecast methodology to analyze with classification with association along with nearest neighbor algorithm, Quasi Newton, Conjugate gradients and multilayered feed forward systems. These 14 attributes are considered to be in both critical and primary state of attributes. As these attributes increases/ decreases it level which turns out to be critical concern. The data sets include the body physiological conditions, food intake, medications and behavioral attributes. The data sets are taken from National Cardiovascular Disorder survey data.

Table 1 - Input datasets for the neural algorithm

Attribute	Description	Data type	Value
Age	Defines the age of the experimenter	Integer, Numerical, Continuous	Age in Years
Sex	Defines Male/Female	String	0 for Female and 1 for male
Glucose Level	Defines the blood sucrose & fructose level. Normally	Float, Numerical	Before meal-70to99mg/dl and after meal 140mg/dl

	classified into After fasting and Before Fasting		and if the level exceeds 11.1mmol/L then it is a risk of cardio vascular disorders
Body Weight	Defines the weight of the experimenter.	Float, Numerical	Body weight should be always less than the normal height of human
BMI	Defines Body Mass Index, which calculates the Height to weight ratio	Float, Numerical	If BMI is <18 results less weight, 18<BMI<24 results normalcy and 25<BMI<29 results excessive weight
Obesity	Defines the bad cholesterol in the body	Float, Numerical	If BMI>30 results obesity which in turn increases the risk of cardiac disorders
Hypertension	The excessive tension along with the brain granial enlargement.	Integer, Numerical	Over 140mm/hg results cardiac arrest
Smoking	Behavioral classification defines the mucus formation and wall damage detection of blood vessels	-	Excessive smoking results blood valves destruction, increase risk factors
Alcohol	Behavioral classification defines the abnormal blood flow and wall damage detection of blood vessels	-	Abnormal blood flow in blood valves results oxygen deficit in turn increases cardiac arrest
Blood Pressure	Enlarged Blood vessels and multiplied blood flow with hypertensive nature	Integer, Numerical	Normal Pressure level is 120/80 mm/Hg and for cardiac disorder patients it is found to be between 130/95-100 mm/hg
Heart Rate	Pulse rate per minute	Integer, Numerical	>80bpm results cardiac disorder [18]
Medications Dosage	Regular medication intake dosage level	Double, Numerical	Limited dosage level of 650mg on regular

			medication
Cardiograph Results	Results the pulse wave pattern	Float, graphical and Numerical	93% when n=365 &7% when n=27% results risk of cardiac arrest with 5.3 percent accuracy
Cardio Acute type	Acute may be migraine or stagnant.	Numerical	Class 1,2,3,4

The data sets are taken from the National Cardiovascular disorder survey data conducted by the government of India. The cardio acute types are classified with various classes and stage which are taken into consideration by the input data sets. This helps in predicting the risk assessment predictive methodology to identify the severity of the cardiac disorder.

a. If the Acute type is in class 1 indicates the that beginning of cardiac disorder results no symptoms but there a short span acute in the systolic medium
b. If the Acute type is class 2 indicates the fatigue and lazy after the excessive workout and even during the normal conditions. The symptoms of class are chest palpitation and shortness of pulse rate
c. If the Acute type is class 3 indicates the discomfort during the work and discrete acute in systolic and diastolic sides of cardio system, Shortness of blood flow and increased risk of arrest
d. If the Acute type is class 4 indicates the high priority risk results shortness of breath during the work, continuous discomfort and acute in the Ventricular fibrillation, high risk of cardiac arrest and increased risk of fatigue.

The information's are taken as input datasets to the classification with association along with nearest neighbor system and also processed with multi layered feed forward training algorithm and Nearest Neighbor to classify the accuracy based on the collected datasets.

3.2 Nearest Neighbor Neural Algorithm

Nearest Neighbor Algorithm is efficient for pattern recognition from the processed datasets. It is termed as un- parameterized data classifier system and has certain limitation on the datasets. The datasets should be of continuous type and it should be larger in value of k to eliminate the noise signals. The input datasets are classified and it should be close to the training samples. The sample from the dataset is identified by forming patterns to the close neighbor. The close neighbor is determined by means of statistical and weight estimation. This technique has high efficient processing time and confluence of simplicity and speed.

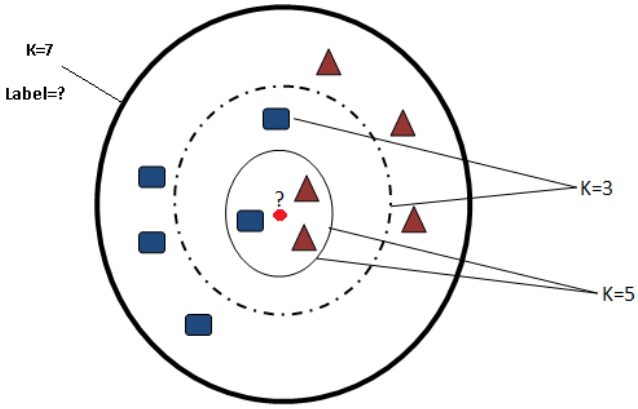

Figure 1- Pattern formation

The model depicts two different patterns formed by the datasets after processing the initial stages. For different values of k, the future classification datasets can be either triangle or rectangle.

a. To determine the value of *k* which is closes to the class instance *d*
b. Accuracy, specificity are defined by the value of k and evaluates the samples by means of statistical analysis and its Euclidean distance between the patterns.

3.2.1.Procedural steps for Nearest Neighbor Neural Algorithm

Identify the Parameter k for finding possibilities of the nearest neighbors. Evaluate the distance among the sample datasets. Extract the minimum distance value by various statistical methods and find the nearest sample. Extract all the nearest samples in the datasets. Categorize the datasets and form patterns with their Euclidean distance.

3.3 Association with classification

Association with classification defines the rules and limitations for the associative attributes in the consecutive datasets. The rules produced are to build the classifier models. The primary evaluation of this technique is to move the constraints by the noise elimination techniques by the post processing techniques.

The Association with Classification algorithm is an extended version of Apriori algorithm is to evaluate the rules of the datasets {class_set, x} where class_set is a input dataset where x belongs to X which is the label. The count of associative rule is to identify the instances of consecutive datasets “class_set” and is named as X class labels which areclass_set X.

The key attributes imbalance results severe cardiac disorders which results various acute type and thus it has been utilized in the predictive methodology. Figure 2 depicts attribute contribution for the forecast methodology.

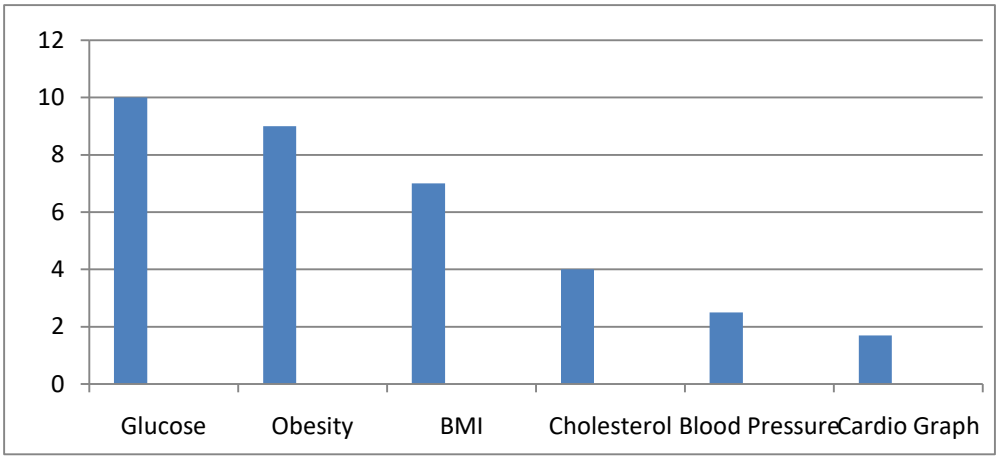

Figure-2 Attribute contribution for the forecast methodology

IV. Results and Discussion

The algorithm is iterated for N values and found to be more accurate when the nth order is low. This is due to the minimum datasets which has the possibility of resulting the accurate values when it is reduced from 14 attributes to 6 attributes. The reduction in attributes is due to the consideration of key elements which are maximum concern for the cardiac disorder as per the datasets obtained from National cardiac disorder survey data. The datasets are analyzed using the Matlab tool and their accuracy levels are defined at the various stages of iteration. A classification study was done with associative, nearest neighbor, feed forward multilayered perceptron and backpropogation algorithms to define the best accuracy results with the smallest iteration conditions.

Table 2: Algorithm and its iterative results

Algorithm	K=1	K=2	K=3	K=4	K=5	K=6	K=7
Association with Classification	75.16	77.15	78.27	83.26	92.34	95.14	99.1
Nearest neighbor	77.2	78.24	81.22	85.14	95.4	98.52	100
Backpropogation	79.3	79.55	81.65	84.69	92.83	96.23	98.17
ML Perceptron	78.4	81.23	82.54	84.87	88.37	90.14	94

The results shows that at K=7, the classification of algorithm possess maximum accuracy and efficiency of results and it is represented by the graphical representation in the below Figure-2

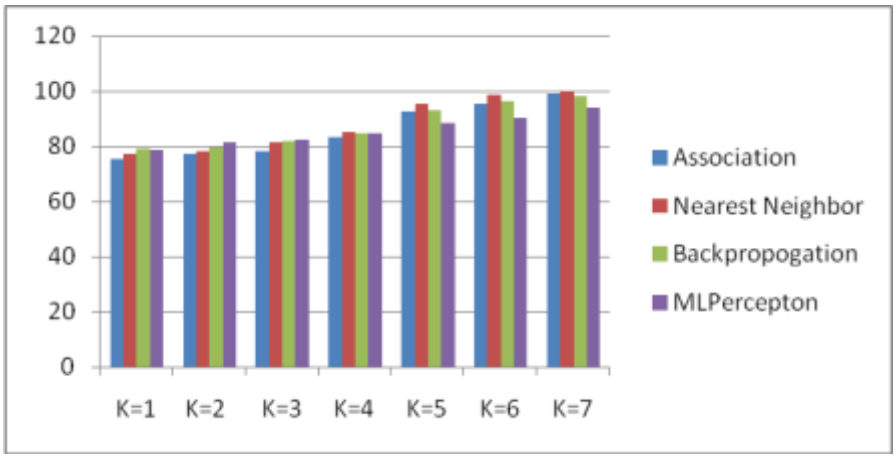

Figure-3 Accuracy of Classifier

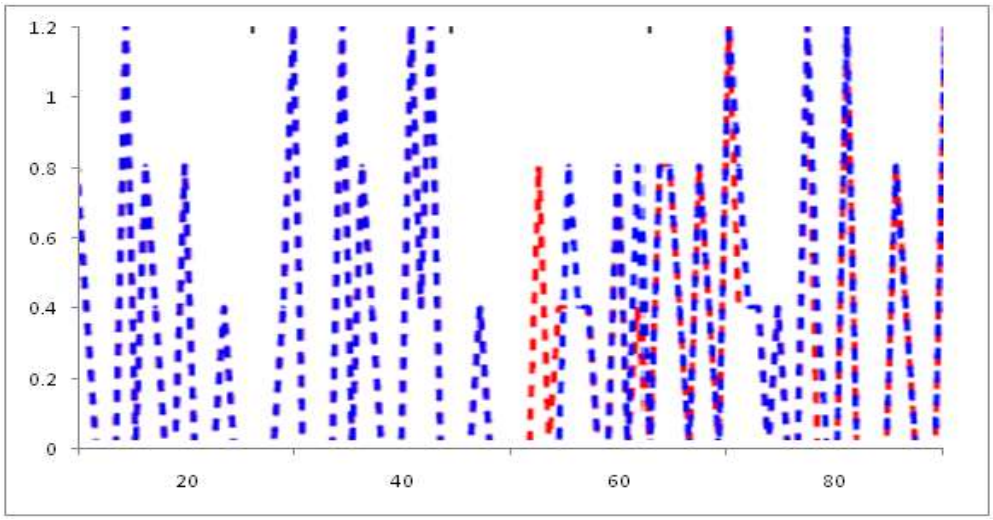

Figure 4- Original datasets along with forecast output for Training the System

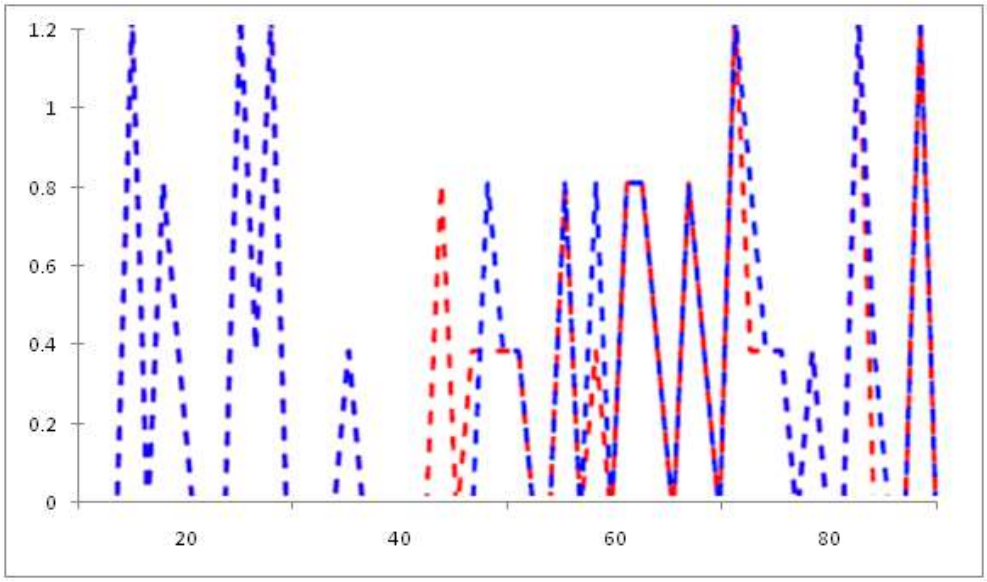

Figure 5- Original datasets along with forecast output for Testing the System

V. Conclusion

In this System of association with classification neural algorithmic technique, 14 attributes are reduced to 6 attributes. With the nearest neighbor technique, the un-recognized noise patterns are transformed into useful data to fine tune the results and thus resulted 100% accurate system for forecasting the cardiac disorders. The key attributes are highly concerned for the disorders related to cardiac which requires care during the period of time. Sudden change in the values of Key attributes requires Intensive care alert. The feed forward multilayered perceptron and back propagation system results are analyzed and compared along with the association with classification and nearest neighbor algorithms to find the accuracy and specificity of the forecasting methodology and found effective results in the association with classification and nearest neighbor techniques. Noise elimination is the key for the accuracy. Fine tuning noise datasets results in efficient pattern recognition and enhancement in efficiency level. Thus the association with classification neural algorithm results 99.1% accuracy with 0.9% noise figure.

REFERENCES

[1] Nabeer Al-Milli, Backpropogation Neural Network for Prediction of Heart Disease, Journal of Theoretical and Applied Information Technology, Vol.56, 2013.
[2] AnchanaKhemphila and VeeraBoonjing, Heart Disease classification using Neural Network and Feature Selection, International conference on systems engineering,2011
[3]Jagdeep Singh, AmitKamra, Harbhang Singh, Prediction of Heart disease using associative classification, IEEE, 2016
[4]Jabbar, Prediction of Heart disease using K-Nearest neighbor and particle swarm optimization, biomedical research, Vol.28, Issye.9, 2017
[5] Dilip Roy Chowdhury, MridulaChatterjee R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011.

[6] Anil Dahiya, NavinRajpal, Niti Guru, Decision Support System for Heart Disease Diagnosis Using Neural Network, Business review, Vol. 8, , 2007.
[7] Milan Kumari, SunilaGodara, Comparative Study of Data Mining Classification Methods in
Cardiovascular Disease Prediction, IJCST Vol. (2), Issue (2), June 2011.
[8] ShantakumarB.Patil, Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research, 2009.
[9] Vanisri K, JyothiSingaraju, Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks, International Journal of Computer Applications , 2011.
[10] SellappanPalaniappan, RafiahAwang, Intelligent Heart Disease Prediction System Using Data Mining Technique, IEEE, 2008.
[11] Haq, A.U., Li, J.P., Memon, M.H., Nazir, S. and Sun, R., 2018. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.
[12] Latha, C.B.C. and Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, p.100203.
[13] Chandrasegar, T. and Choudhary, A., 2019, March. Heart Disease Diagnosis using a Machine Learning Algorithm. In 2019 Innovations in Power and Advanced Computing Technologies (i-PACT) (Vol. 1, pp. 1-4). IEEE.
[14] Ghumbre, S.U. and Ghatol, A.A., 2012. Heart disease diagnosis using machine learning algorithm. In Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012 (pp. 217-225). Springer, Berlin, Heidelberg.
[15] Jaymin, P. TejalUpadhyay, Sami, P., 2016. Heart Disease Prediction Using Machine learning and Data Mining Technique. International journal of computer science and communication. 7(1) pp. 129-137
[16] Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, pp.81542-81554.
[17] Basha, S. M., Balaji, H., Iyengar, N. C. S., & Caytiles, R. D. (2017). A soft computing approach to provide recommendation on PIMA diabetes. Heart, 106.

[18] Basha, S. M., Rajput, D. S., Iyengar, N. C. S. N., & Caytiles, D. R. (2018). A novel approach to perform analysis and prediction on breast cancer dataset using R. International Journal of Grid and Distributed Computing, 11(2), 41-54.