A Study on the accessible techniques to classify and predict the risk of Cardio Vascular Disease

S. Sivagowry[#1], M. Durairaj[*2]

[#1] *Research Scholar*, [*2] *Assistant Professor, School of Computer Science, Engineering and Applications, Bharathidasan University, Trichy, India.*

Abstract--- *The health care environment is found to be loaded with information, but deprived in extracting knowledge from the information. This is because of the short of effectual Data Mining tool to determine concealed associations and trends in them. By applying the data mining techniques, important knowledge can be extracted from the health care system. Heart disease is a assemblage of circumstance affecting the arrangement and purpose of heart and has many root causes. Heart disease is the most important cause of fatality in the humankind over past ten years. Research has been made with many hybrid techniques for diagnosing heart disease. This paper deals with an overall appraisal of application of data mining in heart disease prediction.*

Keywords─ *Data Mining, Heart Disease, Classification, Prediction, Neural Network, Decision Tree, Naïve Bayes*
Introduction

I. INTRODUCTION

Data Mining is the exploration of large datasets to extract hidden and previously unknown patterns, relationships and knowledge that are difficult to detect with traditional statistics [1]. Data mining techniques are the result of a long process of research and product development. Data Mining is divided into two tasks such as Predictive Tasks and Descriptive Tasks. Predictive Tasks predict the value of a specific attribute based on other attribute. Classification, Regression and Deviation Deduction come under Predictive Tasks. Descriptive Tasks derive pattern that summarize the relationship between data. Clustering, Association Rule Mining and Sequential Pattern Discovery are coming under Descriptive Tasks. Data Mining involves few steps from raw data collection to some form of new knowledge. The iterative process consists of following steps like Data cleaning, Data Integration, Data Selection, Data transformation, Data Mining, Pattern Evaluation, and Knowledge Representation.

Medical Data Mining is a domain of challenge which involves lot of imprecision and uncertainty. Provision of quality services at affordable cost is the major challenge faced in the health care organization.

Poor clinical decision may lead to disastrous consequences. Health care data is massive. Clinical decisions are often made based on doctor's experience rather than on the knowledge hidden in the knowledge rich data base. This in some cases will result in errors, excessive medical cost which affects the quality of service to the patients [2]. Medical history data comprises of a number of tests essentials to diagnose a particular disease. It is possible to gain the advantage of Data mining in health care by employing it as an intelligent diagnostic tool. The researchers in the medical field identify and predict the disease with the aid of Data mining techniques [3].

II. HEART DISEASE

The initial diagnosis of a heart attack is made by a combination of clinical symptoms and characteristic electrocardiogram (ECG) changes. An ECG is a recording of the electrical activity of the heart. Confirmation of a heart attack can only be made hours later through detection of elevated creatinine phosphokinase (CPK) in the blood. CPK is a muscle protein enzyme which is released into the blood circulation by dying heart muscles when their surrounding dissolves [4].

World Health Organization in the year 2003 reported that 29.2% of total global deaths are due to Cardio Vascular Disease (CVD). By the end of this year, CVD is expected to be the leading cause for deaths in developing countries due to change in life style, work culture and food habits. Hence, more careful and efficient methods of cardiac diseases and periodic examination are of high importance [5].

III. DATA SETS

The Data set is taken from Data mining repository of University of California, Irvine (UCI). Data set from Cleveland data set, Hungary Data set, Switzerland Data set, long beach and Statlog data set are collected. Cleveland, Hungary, Switzerland and va long beach data set contains 76 attributes in all. But only 14 attributes are used. Among all those Cleveland data set and Statlog data set are the most commonly used data set. Because all the other data set have more missing values than Cleveland data set.

A. *Attributes Used*

TABLE I LIST OF ATTRIBUTES USED

No	Name	Description
1	age	Age in Years
2	sex	1=male, 0=female
3	cp	Chest pain type(1 = typical angina, 2 =atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
4	trestbps	Resting blood sugar(in mm Hg on admission to hospital)
5	chol	Serum cholesterol in mg/dl
6	fbs	Fasting blood sugar>120 mg/dl(1= true, 0=false)
7	restecg	Resting electrocardiographic results(0 = normal, 1 = having ST-T wave abnormality, 2 = left ventricular hypertrophy)
8	thalach	Maximum heart rate
9	exang	Exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	Slope of the peak exercise ST segment (1=upsloping, 2=flat, 3= downsloping)
12	ca	Number of major vessels colored by fluoroscopy
13	thal	3= normal, 6=fixed defect, 7= reversible defect
14	num	Class(0=healthy, 1=have heart disease)

IV. DATA MINING TECHNIQUES USED IN HEART DISEASE PREDICTION

Data Mining (DM) [6,7] classification techniques like Decision Tree, Naïve Bayes, Neural Network, Apriori Algorithm and MAFIA Algorithm[8] is used in Heart Disease prediction. Genetic Algorithm has been used to reduce the attributes number from 13 to 6. Decision Tree has outperformed Naïve Bayes and Classification via Clustering with 99.2%. In [9], ID3 technique gives a classification accuracy of 91.72% when compared with Naïve Bayes, k-NN.

Clustering Algorithms [10] like K-Means, k-NN and Entropy mean based clustering is tested. It was observed that time taken by Entropy based clustering was better and accuracy over clustering is 82.90% which supports the use of Entropy based clustering in the Heart Disease prediction. Tangara has been used as the classification tool. In [11] [12], MAFIA Algorithm is used to classify the data, where estimation for data is based on entropy based cross validation. For training the data, C4.5 Algorithm is used. K-Means Algorithm is suggested for clustering the data. The accuracy level while using K-Means based MAFIA with ID3 and C4.5 is 94% where as the precision and recall value is 0.82 and 0.94 which is better value when compared with K-Means based MAFIA with ID3 and K-means based MAFIA. AptaCDSS-E [13] combines four different classifiers like Support Vector Machine (SVM), Neural Network (NN), Decision Tree and Bayesian Network with ensembles. The system predict the level with CVD with an accuracy (>94%).

In [14], various data mining techniques have been studied for the diagnosis of HD. Data for the study was collected from PGI, Chandigarh which has a total of 15 attributes. A total of 7339 instances were trained. Standard metrics like accuracy, precision, recall, F-Measure were used for measuring the performance of Data Mining techniques. Of the 15, only 8 attributes were selected. The accuracy level was 95.56% with J48 where it was 92.42% with Naïve Bayes and 94.85 with NN. A survey on current techniques [15] on DM like Decision Tree (DT), Naïve Bayes, k-NN, Neural Network and Classification via Clustering (CVC) is done. Decision Tree outperforms all the techniques. Genetic Algorithm has been applied to pre process the data set. After pre processing the data set, the accuracy of DT and Naïve Bayes have been improved. Tangara tool has been used to classify the data and it is evaluated by using 10 fold cross validation. K- Means clustering [16] is a method of cluster analysis which aims to partition 'n' observation to 'k' clusters. Euclidean distance formula is used to minimize the sum of square of distance between data. When K-Means clustering is applied, Naïve Bayes was slow in predicting the HD where as NN takes a maximum number of iteration for the prediction.

In [17], a frequent feature selection method to predict the HD is proposed. Fuzzy measure and relevant non linear integral has been used. It improved the accuracy and reduced the computational time. Weka 3.6 is used for the prediction. In [18], a rule selection method for filtering a large number of extracted rules from the data set has been developed. Two stage rule selection is followed. In the first stage, selection is based on support for individual rule. In [19], a Graphical User Interface is designed to enter the patient record and to predict heart disease by using Weighted Associate Classifier (WAC). Each attribute is assigned weight from 0 to 1. After pre processing, Weighted Association Rule Mining (WARM) is applied by using weighted support and confidence. Rules obtained are stored in the rule base and used further.

Decision Support in Heart Disease Prediction System (DSHDPS) is proposed in [20]. After pre processing the data, the data is clustered by using K-means clustering algorithm with k=2. Accuracy in prediction is outperformed in DT and Naïve Bayes

(NB). In [21], Pruning Classification Association Rule (PCAR) is used for mining the Association Rules. PCAR comes from analyzing and considering the Apriori Algorithm. PCAR deletes minimum frequency item with minimum frequency sets. It deletes infrequent items from the item sets. Then classifies item sets based on frequency of item sets and discovers frequent item sets. Text mining is suggested to mine vast amount of unstructured data available in medical database. In [22], an algorithm called Heart Attack Prediction using Matrix (HAPBM) is used to discover the knowledge from the HD. The algorithm discretizes the medical data and converts it into the Boolean Matrix. Superhot item set Ci and transaction item support Ri are calculated. The column whose Ci < S is pruned. The performance of the algorithm is compared with Apriori algorithm. HAPDM takes less time to generate patterns and reduce the transactions at each stage and reduce the search space.

In [23], the performance of clustering and classification algorithms in DM is analysed. The performance of classifier is calculated using the cross validation test mode and clusters is by mode of classes. Data pre processing is done by attribute subset selection algorithm. Final output has shown the Naïve Bayes has more prediction accuracy. In [24], the performance of NB and WAC by using various performance measures is analyzed. In WAC, each attribute is assigned weight from 0 to 1 based on their importance. WARM algorithm is applied after pre processing to generate the interesting pattern. Weighted support and confidence framework is used and rules generated. The generated rules are stored in the rule base. Attribute weight is assigned by the doctor. Both the algorithms serve well as a practising tool for training the nurses or doctors.

Data Mining Classification [25], [26] is based on a supervised machine learning algorithm. Tanagra tool is used to classify the data and evaluated using 10 fold cross validation. Naïve Bayes, K-nn [27], Decision List Algorithm is taken and the performance of these algorithms is analyzed based on accuracy and time taken to build the model. Naïve bayes is considered to be better since it takes only lesser time to calculate accuracy than other algorithms. It also resulted in lower error rates. The Naïve Bayes algorithm gives 52.23% of accurate result. In [28], it is recommended that hill climbing rough set approach is inadequate to find optimum solution. It is not feasible for complete search even in medium-sized data set. The Particle Swarm Optimization interacts with individuals in the population to find the optimal region in complex search space. The main pro of PSO over GA is that it doesn't require cross over or mutation operation and also inexpensive in terms of both memory and run time. Experimental results show that PSO is efficient for rough set based feature selection.

In [29], a computer aided diagnosis system has been introduced for the heart valve disease by using binary PSO and Support Vector Machine (SVM) algorithm in conjunction with K-nearest neighbor and leave-one-out cross validation. The most weighted feature is selected by using binary PSO algorithm. SVM is used to classify the outcome into two classes as healthy and having heart valve disease. The proposed method helps to optimize the feature selection process. It can also be used an ideal preprocessing tool since it increase the classification accuracy.

A new feature selection based on Rough set Theory hybrided with Bee Colony Optimization is proposed [30]. The method is applied in the medical data set and minimal reduct set is found. The proposed method is compared with Quick Reduct, Entropy based Reduct and with Genetic Algorithm, Particle Swarm Optimization and Ant Colony Optimization hybrided with Rough Set. The solutions provided by Quick Reduct and Entropy based Reduct was close to minimal reduct set. But the solution is not optimal. The GenRSAR, AntRSAR, PSO-RSAR are also performing well but there is no consistency in the result. This is because these algorithms dealt with random parameters. Bee-RSAR exhibits constant and better performance on medical data set.

To offer better conditions on subsequent analysis and to reduce the complexity of data, the data is pre-processed. Rough Set Theory was applied in three pre-processing steps like [31] Discretization, Feature Selection and Instance Selection. Experimental results have shown that the rough Set Theory consider only feature with large dependency. K-NN classifier is used to validate the result. Similar objects are determined by using Expectation Maximization (EM) Clustering algorithm [32]. The features generated by this method are compared with the Fuzzy Rough Feature Selection and Tolerance based Feature Selection. The important aspect to reduce the irrelevant and redundant features is that they lead to slow learning and low accuracy. Rough set theory used data dependencies to reduce the input dimensionality. Function of the approximation is used to calculate the measure of dependency. The proposed method produces a smaller number of attributes compared with other Feature Selection algorithms. It also improved the average accuracy of classifiers like J48, JRIP and CART.

Quick Reduct [33] algorithm is used to reduce the number of genes from gene expression data. The Quick Reduct algorithm is used to obtain minimal data set. Rough Set Theory is used to pre-process the data to mine suitable rules. For post processing, Formal concept analysis is used for mining rules. The rules are mined to extract knowledge and the most important factors that affect the decision making. The aim of using Rough Set Theory is that it supports

prediction whereas Formal Concept Analysis is used for describing the data. This model is believed to be useful for decision making in the medical field.

PSO [34, 35, and 36] is used for Feature Reduction. Here, the patients are classified as diseased and non-diseased with the aid of Artificial Neural Network. The parameters such as Regression, Performance plot, Confusion Matrix and ROC Values are used to analyze the performance. The performance of the whole network is increased after applying PSO. PSO Algorithm with boosting approach [37] is proposed, for extraction of rules to predict the presence or absence of disease. The weight of the training example is reduced by boosting mechanism. The classification accuracy is 85.76% when compared with other methods.

A remote health monitoring platform was designed to support heart failure severity assessment based on Classification and Regression Tree (CART) and PSO [38]. CART takes more time and memory to produce the result whereas the PSO takes less time and memory. Hence forth PSO is best suit to detect the Heart Failure. MisClassification Analysis [39] is used for Data Cleaning. The Complementary Neural Network is used to enhance the performance of network classifier. Two techniques are used. Falsity NN is obtained by complementing the target output of training data. True NN and False NN are trained for membership values. In the first technique, new training data are obtained by eliminating all misclassification patterns. In the second technique, only the misclassification patterns are eliminated. The classification accuracy is improved after Data cleaning. Technique II showed much accuracy than Technique I.

ANN [40] algorithm is used for classifying the heart disease based on the input. Learning Vector Quantization (LVQ) is a prototype based supervised classification algorithm. It creates prototype to interpret for experts in the respective domain. The accuracy level obtained is around 80%. DT, Naïve Bayes and NN are taken for study in [41]. A total of 13 and 15 attributes are taken separately and studied. In both the cases, it was observed that the NN performed well. It has been suggested that Clustering, Time Series, Association rule can also be used to predict the HD. Text mining has been suggested to mine a huge amount of unstructured data. Cengiz colak and et al., [42] used ANN to train eight models. The training resulted in promising output. Parthiban and et.al,[43] used Co active Neuro Fuzzy Inference System (CANFIS) to design an Intelligent HD prediction system. Fuzzy and NN is combined with Genetic algorithm in the proposed method. It resulted in promising result.

Qeethara Kadhim Al. Shayea[44] used Feed Forward back propagation Neural Network as a classifier to distinguish between infected or non infected person. Neural Network toolbox from Matlab 7.9 is used to evaluate the performance of the proposed network. Levenburg-Marquardt back propagation algorithm is used to train the network. Neural Network can be used for identifying the infected person. Multi Layer Perceptron (MLP) with K-Means Clustering [45] is proposed for HD prediction. 15 attributes are taken for experimentation. When compared with DT and NB, NN has given result in less time complexity system.

An automatic HD diagnosis system was designed using MATLAB [46]. It was developed as two systems. The MLP was used to develop the first system and the second system is based on Adaptive Neuro Fuzzy Inference System (ANFIS). Each system has training and testing modules. While using training module, Neuro-Fuzzy system outperformed with 100% and 90.74% of accuracy. In training module, ANN outperforms with an accuracy of 80.74% and 75.93%.

Weighted Fuzzy rule based Clinical Decision Support System (CDSS) is proposed [47]. It consists of two phases. The first phase is automated approach for the generation of fuzzy rules and the second is developing a fuzzy rule based on Decision Support System. The CDSS is compared with Neural Network based system by Sensitivity, Specificity and Accuracy. Cleveland, Hungarian and Switzerland Data sets are used. The sensitivity of Neural Network (NN) and CDSS is 52.47% and 45.22%, Specificity is 52.46% and 68.75%, Accuracy is 53.86% and 57.85%.

Genetic Algorithm is used to final the relevant set by optimizing the fitness function by using the selection and cross over operators. Support sets [48] are used to frame the fuzzy rules. The classification results obtained using the Fuzzy Inference System was better than the other techniques. To predict Heart Disease, fuzzy based Soft computing Approach [49] is proposed. It is implemented on multiple parameters. The final output was 'very low', 'low', 'medium', 'high' and 'very high'. This proposed method correlate between medical issues related to disease. It can be used as an intelligent system by doctors.

Utilized Short Term Fourier Transform (STFT) algorithm [50] and their Riemann has been used to categorize the ECG signal. Fuzzy Logic System has been used to recognize the disease with best accuracy. Guassian function is selected for correlation. Mamadani fuzzy engine is used. A fuzzy rule based expert system is designed by Kantesh kumar and et. al. [51]. At the initial stage, the data was pre processed. Following that, the attributer number was reduced using the DM techniques. Then it is processed by using the Fuzzy Logic System by using MATLAB Programming. Mamdani inference system has been used for designing the system. The proposed method is compared with NN and J48 Decision Tree.

It was observed that the proposed method has outperformed the other two.

Adaptive Neuro Fuzzy Inference system (ANFIS) [52] is used to develop a method for the classifying the Heart Disease. Seven variables are used as input. K-fold cross validation is used to test the performance of the model. An accuracy level of 92.30% was obtained. The training and testing error was 0.01 and 0.15 respectively. Fuzzy Logic based HD diagnosis system [53] is designed based on parameters like cholesterol, blood pressure, diabetes, sex and age. Experimental results showed an accuracy of 92%.

Advanced Fuzzy Resolution Mechanism [54] is developed which predict value to diagnose the heart disease with five layers. Each layer has its own node. The accuracy level was 93.88% found to be the best accuracy when compared with other existing work. A Neuro-Fuzzy System [55] is designed with eight input fields and one output field to predict the heart disease. It detects the risk level of the output classified into four classes. The system was designed in such a way that the patient can refer it for personal use too. Decision Tree and NB [56] is combined with Fuzzy Logic for HD Diagnosis. The method has reduced the number of attributes so that the number of tests for the patients can be reduced. The proposed method is experimented with Weka. DT outperform with the prediction probability of 99.62%.

PSO is combined with C4.5 [57] Algorithm to develop a model to predict and analyse the events related to coronary HD. The proposed method has reduced the number of features and has improved the prediction accuracy. Two data sets are used, data set from UCI and real time data set. The PSO J48 has higher prediction accuracy with reduced number of features. CART [58] decision tree was used to extract rules from the Heart Disease database with a total of essential 14 attributes. PSO has been used for optimizing the fuzzy membership function. The results obtained from the fuzzy system are interpreted and it was observed that the accuracy was good. The accuracy of Cleveland and Switzerland data set was 92.2% and 86% before optimization. But after applying optimization it was 94.4% and 94% respectively.

Supervised Feature Selection [59] method which is based on hybridization of PSO, PSO based Relative Reduct (PSO-RR), and PSO based QR is proposed. The overall accuracy acquired was promising result when compared with traditional QR and RR algorithm. Rough set indiscernibility [60] method with Back propagation Neural Network (RS-BPNN) is proposed. It is splitted into two stages. The first stage is handling the missing values by using indiscernibility relation method. In the second stage, BPN is used to classify the selected reducts of data sets. The method has been tested with hepatitis, Wisersia breast cancer and Statlog HD dataset from UCI. The accuracy was 94.3%, 98.6% and 90.4% for hepatitis, Breast Cancer and HD respectively.

A feature selection model is proposed based on the Binary PSO and Genetic Algorithm for determining Coronary Artery disease by using SVM [61]. The results obtained proved that the method was successful in selecting the feature for determining the Coronary Artery Disease. Neural network and GA [62] has been combined together to develop a Heart Disease prediction system. Number of hidden nodes for Neural Network was calculated to train the network. For initializing the network, the GA was used which is a global optimization algorithm. Back propagation learning is used for training and learning the network. To calculate the error, they have used $Error = 1/2(output_{desired} - output_{actual})2$. MATLAB 2012A has been used to develop the system. Global optimization and Neural Network toolbox has been used. The accuracy obtained was 98%.

Genetic Algorithm and FL [63] are used to diagnose the Heart Disease. Doctors can use this system for Heart Disease diagnosis. To find out the optimal solution by using stochastic search GA is used. Gaussian Membership function was used for Fuzzification and Centroid method has been used to Defuzzify the output value. Performance metrics like accuracy, sensitivity, and specificity and Confusion matrix has been used to evaluate the performance of the system. The accuracy obtained was 86%. The accuracy was increased by 1.54% when compared with the existing work.

A Neuro-Fuzzy System [64] is designed with eight input fields. Mean Squared Error is used to evaluate the performance of the model. A mixture of back propagation and least square are hybrid with each other to develop the system. The validation of the system gives a value of 0.014463 for Mean Squared Error. The accuracy obtained from the experimental results is 90%. The comparison of all the techniques is summarized as given in Table 2

TABLE III COMPARISON OF ALL THE EXISTING TECHNIQUES FOR THE PREDICTION OF CVD

Data Mining Techniques	Accuracy (%)	References
Decision Tree	99.2	6,7
ID3	91.72	9
Entropy Based Clustering	82.90	10
K-Means based MAFIA	94	11,12
Apta CDSS-E	>94	13
J48	92.56	
NB	92.42	14

NN	94.85	
NB	52.23	27
PSO with Boosting approach	85.76	37
ANN	80	40
Neuro Fuzzy System ANN	100 – Training 90.74- Testing 80.74 – Training 75.93 - Testing	46
NN	53.86	
Weighted Fuzzy Rule Based Clinical DSS	57.85	47
FL	92	53
Advanced Fuzzy Resolution Mechanism	93.88	54
DT	99.62	55
RS-BPNN	90.4	60
NN & GA	98	62
GA & FL	86	63
Neuro Fuzzy System	90	64

V. CONCLUSION

This paper summarizes review of the techniques used for classifying and also for predicting the risk level of Heart Diseases. The techniques adopted produced good results. Humans can be saved from the Heart disease, in case of early diagnosis. This paper reviews the DM techniques applied for diagnosing Heart Disease. It is observed that the prediction results are strongly encouraging and would assist physicians to do early diagnosis and make more accurate designs. Nonetheless, the techniques do not achieve 100% accuracy for Heart Disease prediction and hence cannot be utilized solely for diagnosis. Future work focus on improving the prediction accuracy of CVD. Furthermore, Feature Reduction should be utilized to achieve better accuracy. This can be achieved by using or combining the available techniques.

REFERENCES

[1] Liangxiao. J, Harry.Z, Zhihua.C and Jiang.S "One Dependency Augmented Naïve Bayes", ADMA, pp 186-194, 2005.

[2] Siri Krishnan Wasan, Vasutha Bhatnagar and Harleen Kaur "The Impact of Data Mining techniques on medical diagnostics", Data Science Journal, Vol 5(19), pp 119-126, October 2006.

[3] Setiawan N.A, " Rule Selection for Coronary Artery Disease Diagnosis Based on Rough Set" ,International Journal of Recent Trends in Engineering, Vol 2(5), pp 198-202, Dec 2009.

[4] Asha Rajkumar and Mrs. Sophia Reena, " Diagnosis of Heart Disease using Data Mining Algorithms, Global Journal of Computer Science and Technology, vol. 10(10), 2010, pp 38-43.

[5] Anbarasi.M, Anupriya and Iyengar "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering and Technology, Vol 2(10), 2010, pp 5370-5376.

[6] Aditya Methaila, Prince Kansal, Himanshu Arya, Pankaj Kumar, " Early Heart Disease Detection Using Data Mining Techniques", *Computer Science and Information Technology, pp 53-59, 2014.*

[7] Fabricio Voznika, Leonardo Viana, " Data Mining Classification", Retrieved from: *http://courses.cs.washington.edu/courses/csep521/07wi/prj/leonardo_ fabricio.pdf",*

[8] Deepali Chadna, " Diagnosis of Heart Disease using Data Mining Algorithm", *International Journal of Computer Science and Information Technology,* Vol 5(2), pp 1678-1680, 2014.

[9] Angadi S.A, Mouna M.Naravani, " Predictiong Heart Attack using NBC, k-NN and ID3", *International Journal of Computer Science and Engineering,* Vol 2 (7), pp 6-12, 2014.

[10] Jaya Rama krishniah V,V, Chandra Sekar D,V, K.Ramachand H Rao, " Predicting the Heart Attack Symptoms by using Biomedical Data Mining Techniques", *The International Journal of computer Science and Applications,* Vol 1(3), pp 10-18, May 2012.

[11] Nishara Banu, Gomathy.B, "Disease Predicting System Using Data Mining Techniques", *International Journal of Technical Research and Applications,* Vol 1(5), pp 41-45, December 2013.

[12] Manikantan.V & Latha.S, " Predicting the Analysis of Heart Disease Symptoms by using Medical Data Mining", *International Journal on Advanced Computer Theory and Engineering,* Vol 2(2) pp 5-10, 2013.

[13] Jae Hong Eom, Sung-Chun Kim, Byoung-Tak Zhang, " AptaCDSS-E:A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction", *International Journal of Expert System with Application*, Vol 34(4),pp 2469-2479, May 2008.

[14] Abhishek Taneja, " Heart Disease Prediction System using Data Mining Techniques", *Oriental Journal of Computer Science and Technology,* Vol 6(4), pp 457-466, December 2013.

[15] Sunita Soni "Predictive Data Mining for Medical Diagnosis : An overview of Heart Disease Prediction", *International Journal of Computer Application,* Vol 17 (8), pp 43-48, March 2011.

[16] Bala Sundar.V, "Development of Data Clustering Algorithm for predicting Heart", *International Journal of Computer Applications,* vol 48 (7), pp 8-13, June 2012.

[17] Nithya N.S, " Assessment of the risk factors of Heart Attack using frequent feature selection method", *International Journal of Communications and Engineering,* Vol 1(1), pp 127-133, March 2012.

[18] Setiawan N.A, "Rule Selection for Coronary Artery Disease Diagnosis based on Rough Sets", *International Journal of Recent Trends in Engineering,* Vol 12(5), pp 198-202, December 2009.

[19] Jyothi Soni, " Intelligent and Effective Heart Disease Prediction System using Weighted Associate Classifier", *IJCSE,* Vol 3(6), pp 2385-2392, June 2011.

[20] Raghu.D," Probability based Heart Disease Prediction using Data Mining Techniques", *IJCST,* Vol 2(4), pp 66-68, December 2011.

[21] Deepika, " Association Rule for classification of Heart attack patients", *IJAEST,* Vol 11 (2), pp 253-257, 2011.

[22] Jabbar M.A, "Knowledge discovery from mining association rules for Heart Disease Prediction", *JATIT,* Vol 41 (2), pp 166-174, 2012.

[23] Santhi.P, " Improving the performance of Data Mining Algorithms in Health Care", *IJCST,* Vol 2 (3), 2011.

[24] Aditya Sundar, Pusphalatha, Rama Chandra, " Performance analysis of classification of Data Mining Techniques over heart disease database", *IJESAT,* Vol 2(3), pp 470-478, 2012.

[25] Rajkumar and Mrs. Sophia Reena, "Diagnosis of Heart Disease using Data Mining Algorithms", Global Journal of Computer Science and Technology, vol. 10(10), pp 38-43, 2010.

[26] Chaltrali S. Dangare and Sulabha, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", IJCA, Vol 47(10), pp 44-48, June 2012.

[27] R. Setthukkarase and Kannan "An Intelligent System for mining Temporal rules in Clinical database using Fuzzy neural network",European Journal of Scientific Research, ISSN 1450-216, Vol 70(3), pp 386-395, 2012.

[28] Xiangyang Wang, "Feature Selection based on Rough Sets and Particle Swarm Optimization", *Elsevier*, Volume 4(1), March 2007.

[29] Mona Nagy Elbedwedhy, Hossam M.Zawbaa, Naveen Ghali and About Ella Hassanien, " Detection of Heart Disease using Binary Particle Swarm Optimization ", *Proceedings of Federated Conference on Computer Science and Information System*, pp 177-182, © IEEE, 2012.

[30] N. Suguna and Dr. K. Thanuskodi, " A Novel Rough Set Reduct Algorithm for Medical Doman based on Bee Colony Optimization", Journal of Computing, Vol 2(6), pp 49-54, 2010.

[31] Mona Nagy Elbedwedhy, Hossam M.Zawbaa, Naveen Ghali and About Ella Hassanien, " Detection of Heart Disease using Binary Particle Swarm Optimization ", *Proceedings of Federated Conference on Computer Science and Information System*, pp 177-182, © IEEE, 2012.

[32] Faridah Fazayeli, Lipo Wang and Jacek Mandziuk, " Feature Selection based on Rough Set Theory and Expectation Maximization Clustering Algorithm", RSCTC 2008, PP 272-282, 2008.

[33] K. Anitha and Dr. P. Venkatesan, " Feature Selection by Rough-Reduct Algorithm", IJIRSET, Vol. 2 (8), 2013.

[34] Sivagowry.S, Dr. Durairaj.M. "A Pragmatic Approach of Preprocessing the Data Set for Heart Disease Prediction." *International Journal of Innovative Research in Computer and Communication Engineering* 2.11 (2014): 6457-6465.

[35] Sivagowry, S., M. Durairaj, and A. Persia. "An empirical study on applying data mining techniques for the analysis and prediction of heart disease." *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*. IEEE, 2013.

[36] Durairaj. M, Sivagowry. S,"Feature Diminution by Using Particle Swarm Optimization for Envisaging the Heart Syndrome", IIJITCS, vol.7, no.2, pp.35-43, 2015. DOI: 10.5815/ijitcs.2015.02.05

[37] Ghadri Hedeshi.N and Saniee Abadeh, " Coronary Artery Disease Prediction using a Fuzzy Boosting PSO Approach", *Computation Intelligence and Neuroscience,* 2014.

[38] Priyatharshini.R and Usha Nandhini.C, " Evaluating the clinical Domain with Data Mining using Classification and Regression Tree and Particle Swarm Optimization method", *International Journal of Innovative Research and Development,* Vol 2(37), pp 85-89, March 2013.

[39] Shouman.M, Turner.T and Stocker.R, "Applying K-Nearest Neighbour in diagnosing Heart Disease Patients", International Journal of Information and Education Technology, Vol 2(3), June 2012.

[40] Chen A.H, "HDPS: Heart Disease Prediction System", *Computing in Cardiology,* ISSN 0276-6574, pp 557-560, 2011.

[41] Chaitrali S. Dangaree and Sulabha, " Improved Study of Heart Disease Prediction System using Data Mining classification techniques", *International Journal of Computer Applications,* Vol 47(10), pp 44-48, June 2012.

[42] Cengiz Colak.M , " Predicting Coronary artery disease using different artificial neural network models", *CAD and Artificial Neural Network,* pp 249-254, 2008.

[43] Latha Parthiban and Subramanian.R, "Intelligent Heart Disease prediction system using CANFIS and Genetic Algorithm", *Journal of Biological and Life Sciences,* Vol 3(3), pp 157-160, 2007.

[44] Qeethara Kadhim Al.Shayea, "Artificial Neural Network in Medical Diagnosis", *IJCSI,* Vol 8(2), March 2011.

[45] Dimple and Dr. Rakul Rishi, " Heart Disease prediction using Multi Layer Perceptron", *International Journal in MultiDisciplinary and Academic Research,* Vol 2(4), July-August 2013.

[46] Mohd. A.M. Abushariah, Assal A.M Alquadah, Omar Y.Adwan, Rana M.M. Yousef, " Automatic Heart Disease Diagnosis system based on Artificial Neural Network and ANFIS approach", *Journal of Software Engineering and Application,* pp 1055-1064, 2014.

[47] Annoj P.K.," Clinical decision support system: Risk level prediction of heart disease using Data Mining Algorithms", *Journal of King Saud University- Computer and Information Sciences*, pp 27-40, 2012.

[48] Ephzibah E.P, SundaraPandian V, " Framing Fuzzy rules using support sets for effective Heart Disease diagnosis", *International Journal of Fuzzy Logic System,* Vol 2(1), pp 11-16, February 2012.

[49] Upasana Juneja, Deepti, " Multiparametric Approach using Fuzzification on Heart Disease analysis", *International Journal of Advances in Computer Science and Communication Engineering,* Vol 2(2), pp 1-7, June 2014.

[50] Mohd Reza Karimi Rad, Farahnaz Nouri, Masih Hashemi and Mostafa Sadeghi Moheseli, "Utilization of Fuzzy Logic for classification of Heart Disease", *International Conference on Bio-Informatics and Bio-Medical Technology,* Singapore, 2012.

[51] Kantesh Kumar oad, Xu Dezhi and Pinial Khan Butt, " A Fuzzy Rule based approach to predict risk level of Heart Disease", *Global Journal of Computer Science and Technology,* Vol 14(3), 2014.

[52] Negar Ziasabounchi, Iman Askerzade, "ANFIS based classification model for Heart Disease Prediction", *International Journal of Electrical and Computer Science,* Vol 14(2), 2014.

[53] Neeru Pathania and Ritika, " Implementation of Fuzzy Controller for Diagnose of Patient Heart Disease", *International Journal of Innovative Science, Engineering and Technology,* Vol 2(4), pp 694-698, April 2015.

[54] Senthil kumar A.V, "Diagnosis of Heart Disease using Advanced Fuzzy Resolution Mechansim", *International Journal of Science and Applied Information Technology,* Vol 2(2), pp 22-30, 2013.

[55] Obanijesu Opeyemi,Emuoyibofarhe O. Justice,"Development of Neuro-fuzzy System for Early Prediction of Heart Attack", IJITCS, vol.4, no.9, pp.22-28, 2012.

[56] Nidhi Bhatia and Kiran Jyothi, "A Novel approach for Heart Disease diagnosis using Data Mining and Fuzzy Logic", *International Journal of Computer Applications,* Vol 54(17), pp 16-21, September 2012.

[57] Sheik Abdullah.A, " A Data Mining model to predict and analysis the events related to coronary Heart Disease using Decision Tree with Particle Swarm Optimization for Feature

Selection", *International Journal of Computer Applications,* Vol 55(8), pp 49-55, October 2012.

[58] Persi Pamela, Gayathri.P and Jaisankar.N, " A Fuzzy Optimization technique for the prediction of Coronary Heart Disease using Decision Tree", *International Journal of Engineering and Technology,* Vol 5(3), pp 2506-2514, June-July 2013.

[59] Hannah Inbarani.H, Ahmad Taher Azan, Jothi.G, "Supervised hybrid feature selection based on Particle Swarm Optimization and Rough sets for medical diagnosis", *Computer Methods and Programs in Bio Medicine,* pp 175-185, Elsevier, 2014.

[60] Kindie Biredagn Nahato, Khanna Nehemiah Harichandran and Kannan Arputharaj, " Knowledge mining from clinical data sets using Rough Sets and Back Propagation Neural Network", *Computational and Mathematical methods in Medicine,* pp 1-14, 2015.

[61] Ismail bebarglu, Oguz Findik, Erkan Ulker, " A comparison of feature selection models utilizing binary PSO and Genetic Algorithm in determining coronary artery disease using Support Vector Machine", *Expert System with Application,* pp 3177-3183, 2013.

[62] Nilakshi P.Waghulde, Nilima P.Patil, "Genetic Neural approach for Heart Disease Prediction", *International Journal of Advanced Computer Research,* Vol 4(3), pp 778-784, September 2014.

[63] Santhanam.T and Ephzibah E.P, "Heart Disease prediction using Hybrid Genetic Fuzzy Model", *Indian Journal of Science and Technology,* Vol 8(9), pp 797-803, May 2015.

[64] Obanijesu opeyemi, Emnoyibafarhe O.Justice, "Development of Neuro-Fuzzy System for Early Prediction of Heart Attack", *International Journal of Information Technology and Computer Science,* Vol 9, pp 22-28, 2012.