

A Pragmatic Approach of Preprocessing the Data Set for Heart Disease Prediction

Dr. Durairaj.M, Sivagowry.S

Assistant Professor, Department of Computer Science, Engineering and Technology, Bharathidasan University, India.

Research Scholar, Department of Computer Science, Engineering and Technology, Bharathidasan University, India.

ABSTRACT: Medical Ecosystem is originated with rich information database, but inadequate in techniques to extract the information from the database. This is because of the lack of effective analysis tool to discover hidden relationships and trends in them. By applying the data mining techniques, valuable knowledge can be extracted from the health care system. Extracted knowledge can be applied for the accurate diagnosis of disease and proper treatment. Heart disease is a group of condition affecting the structure and functions of the heart and has many root causes. Heart disease is the leading cause of death in all over the world over past ten years. Researchers have developed many hybrid data mining techniques for diagnosing heart disease. This paper describes a preprocessing technique and analyzes the accuracy for prediction after preprocessing the noisy data. It is also observed that the accuracy has been increased to 91% after preprocessing. Swarm Intelligence techniques hybridized with Rough Set Algorithm are to be taken as future work for exact reduction of relevant features for prediction.

KEYWORDS: Data Mining, Artificial Neural Network, Multilayer Perceptron, Radial Basis Function, Root Mean Squared Error, Prediction Accuracy, Redunt Features.

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 [17]. Data mining techniques are the result of a long process of research and product development [5]. Data Mining involves few steps from raw data collection to some form of new knowledge. The iterative process consists of the following steps like Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Representation. The Figure 1 shows that the Data Mining is the core of the various processes involved in Knowledge Discovery Process. Knowledge Discovery is a process of getting high level knowledge from low level data [38].

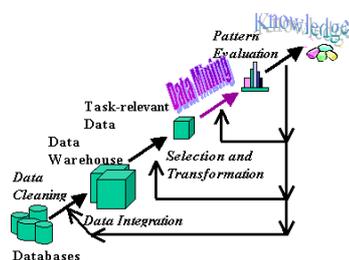


Figure 1: Data Mining- the core of Knowledge discovery process

Medical Data Mining is a domain of challenge which involves a lot of imprecision and uncertainty. Provision of quality services at affordable cost is the major challenge faced in the health care organization. The poor clinical decision may lead to disastrous consequences. Health care data is massive. Clinical decisions are often made based on the doctor's experience rather than on the knowledge rich data hidden in the database. This in some cases results in errors, excessive medical cost which affects the quality of service to the patients [33]. Medical history data comprise of a number of tests essentials to diagnose a particular disease. It is possible to gain the advantage of Data mining [21], [25], [26] in health care by employing it as an intelligent diagnostic tool [34]. Accuracy of the prediction can also be improved by using Data Clustering Algorithms [4]. The researchers in the medical field have succeeded in identifying



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

and predicting the disease with the aid of Data mining techniques [29]. Association rules of Data Mining have been significantly used [7], [11], and [12].

The paper is organized as follows: Section II gives description about heart disease and its impact on the society. Section III is about the data set collected for experimentation. Section IV describes the preprocessing techniques and training of the data set using Multilayer Perceptron Networks and Radial Basis Function. Section V concludes the paper.

II. HEART DISEASE

The rise of health care cost is one of the world's most important problems [21]. Heart attack happens when there is irregularity in the flow of blood and heart muscle is injured because of inadequate oxygen supply [47]. World Health Organization in the year 2008 reported that 30% of total global deaths are due to Cardio Vascular Disease (CVD). By 2030, almost 25 million people will die from CVDs, mainly from heart disease and stroke [10], [14], [49]. These are projected to remain the single leading cause of death. CVD is expected to be the leading cause of deaths in developing countries due to changes in lifestyle, work culture and food habits. Hence, more careful and efficient methods of cardiac diseases and periodic examination are of high importance [27], [28].

III. RELATED WORK

Genetic Algorithm [1], [16] is used to determine the attributes for the diagnosis of heart disease. Feature extraction is done with the aid of Genetic Algorithm (GA). The attribute number is reduced to 6 by using GA. Naïve Bayes [35], Classification by Clustering and Decision Tree, are the classifiers which are used for testing the reduced data set. It is observed that the Decision Tree outperforms but takes more time to build the model. Naïve Bayes has performed consistently before and after the reduction of attributes. Classification via Clustering is poor in performance. Weka tool is used for evaluation.

Weighted Fuzzy rule based Clinical Decision Support System (CDSS) is proposed [2], [13]. 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%.

Data Mining Classification [3], [9] 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 [32], 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.

Intelligent Heart Disease Prediction System (IHDPS) [25] is developed using Data Mining Techniques namely Decision Tree, Naïve Bayes and Neural Network. Each technique has its own strength in realizing the objectives of Data Mining. DMX Query language is used which answers complex "What if" queries where Decision Support System can't. Five Data mining rules are defined and evaluated using the three models. Naïve Bayes [26], [16] is found to be the most effective in Heart Disease diagnosis.

MisClassification Analysis [41] 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. And 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.

Ripper Incremental Pruning to Produce Error Reduction (RIPPER) [19], [20], Support Vector Machine (SVM), Decision Tree and Artificial Neural Network are compared. The performances of the algorithms are compared with



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

each other based on Sensitivity, Specificity, Accuracy, Error rate, True Positive rate and False Positive rate. SVM predicts with least error rate and highest accuracy.

Subhagata Chattopadhyay [48] has mined some important pre-disposing factors of heart attack. 300 real world cases have been taken for study. 12 factors are taken. Divisive Hierarchical Clustering (DHC) techniques has been used to cluster the sample as 'single', 'average' and 'complete' linkage. It has been also observed that male with age group of 48-60 are prone to suffer severe and moderate heart attack, where women over 50 years are affected mostly with mild attacks.

IV. DATA SET

The Data set used for experimentation is reserved from Data mining repository of the 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. Among all the 76 attributes, 14 attributes are taken for experimentation. Cleveland data set and Statlog data set are the most commonly used data set for testing purpose by the researchers in the medical domain. This is because all the other data set has more number of missing values than Cleveland data set [46].

A. Attributes Used

The Table 1 shows the attributes used for the purpose of the heart disease prediction.

Table 1: Attributes collected for heart disease prediction

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 ventricularhypertrophy)
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)

The data set gathered for mining will contain either numeric attributes or nominal attributes. The data set collected for Heart disease encompasses both the numeric and nominal attributes. From the above 14 attributes, the listed features such as age, trestbps, Chol, thalach and oldpeak are numeric attributes and the remaining 9 comes under nominal. The relative importance of the variables in predicting the heart disease is shown in the Figure 2.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

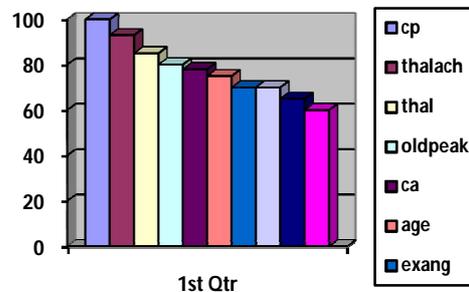


Figure 2: Relative Importance of Variables

V. PROPOSED WORK

The proposed work methodology for preprocessing is as follows:

- As a preliminary stage, the data set is preprocessed by using Numeric to Nominal and Replace Missing Value techniques.
- After cleaning, the data set is trained for accuracy.
- The next stage is the extraction of redundant feature for prediction.
- This is to be effected by using the Swarm Intelligence Techniques hybridized with Rough set Algorithm.
- The data set is validated to acquire the optimal Redundant feature for prediction. The Figure 3 describes the proposed work methodology.

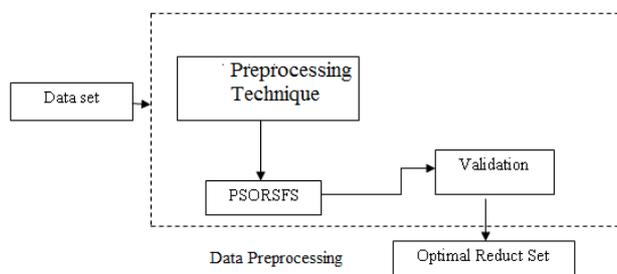


Figure 3: The proposed work Methodology

VI THE DATA SET PREPROCESSING AND TRAINING

An Artificial Neural Network is a mathematical model inspired by biological Neural Networks [8], [23], and [24]. A Neural Network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation [10]. A neural network is considered to be an adaptive system that changes its structures during its learning phase [6]. Neural Networks are used to model complex relationships between inputs and outputs or to find patterns in data. Weights of the interconnections are adjusted to produce the desired output [38].

Multilayer Perceptron (MLP) is a Neural Network which is based on Supervised Learning method and the network is trained by using the Back Propagation Algorithm [15]. Back propagation algorithm is the most popularly used Neural Network Algorithm. Feed Forward Neural Network or Multilayer Perceptron is the most widely studied network algorithms for classification purpose MLP uses the non-linear activation function. The hidden neurons make the network active for highly complex tasks [30], [31].

The Figure 4 gives the architecture of the MLP network. One of the most important characteristics of a Perceptron network is the number of neurons in the hidden layer(s). If an inadequate number of neurons are used, the network will be unable to model complex data, and the resulting fit will be poor. If too many neurons are used, the training time may become excessively long, and worse, the network may overfit the data. When over fitting occurs, the network will

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data. Validation must be used to test for this.

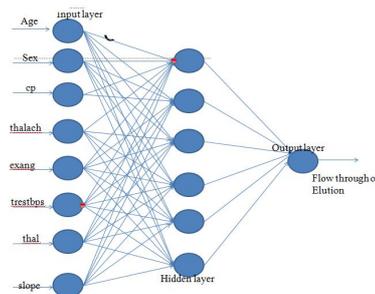


Figure 4: Multilayer Perceptron Network

The Radial Basis Function Network is an ANN which works based on the Radial Basis Function as activation functions. Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. The input can be modeled as a vector of real numbers $x \in R^n$. The output of the network is then a scalar function of the input vector, $\varphi : R \rightarrow R$ and is given by

$$\varphi(x) = \sum_{i=1}^N a_i p(\|x - c_i\|)$$

Where N is the number of neurons in the hidden layer, c_i is the centre vector for neuron i , and a_i is the weight of neuron i in the linear output neuron. Functions that depend only on the distance from a centre vector are radially symmetric about that vector, hence the name radial basis function. In the basic form all inputs are connected to each hidden neuron.

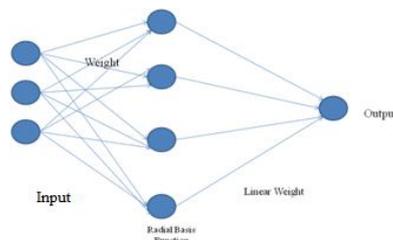


Figure 5: Architecture of RBF network

Figure 5 shows the architecture of Radial Basis Function Network. The main features of RBF are they are two layered feed-forward networks. The hidden node implements a set of Radial Basis Function. The output node implements a set of linear summation functions as in MLP. The network training is divided into two stages. In the first stage, the weight from the input to the hidden layer is determined. And in the second stage, the weight from the hidden to the output layer is determined. The networks are very good at interpolation.

Data Preprocessing [45] plays a significant role in Data Mining. The training phase in the Data Mining during Knowledge Discovery will be very difficult if the data contains irrelevant or redundant information or more noisy and unreliable data. The medical data contain many missing values. So preprocess is an obligatory step before training the medical data. A total of 303 instances are trained before preprocessing.

The Table 2 compares the performance of both MLP and RBF before preprocessing the data. The RMSE values obtained are evaluated. It is initiated that the RMSE and Correlation Coefficient values are much lower. Correlation Coefficient is a measure of statistical correlation between predicted and actual values. If the Correlation Coefficient is 1, it is a perfect statistical Correlation and there is no correlation if it is 0. The Correlation value of the RBF is nearer to

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

0. Hence, training a network with MLP will yield better results than RBF. The Table 2 compares the performance of both the networks in terms of Correlation, Mean Absolute Error, RMSE, RAE and Root relative squared error.

Table 2: Comparing the Performance of MLP and RBF before preprocessing

Parameters	MLP	RBF
Time taken	0.79	0.02
Correlation co-efficient	0.8784	0.3719
Mean Absolute Error	0.3975	0.9017
Relative Mean Squared Error	0.6023	1.1385
Relative Absolute Error (%)	39.1771	88.8686
Root Relative Absolute Error (%)	49.1035	92.8285

The heart disease data set has both numeric and nominal data sets. The primary step of preprocessing involves the conversion of numeric attribute to nominal attribute. The Numeric to Nominal conversion is used for renovating the attributes as Nominal. The Table 1 shows the result before preprocessing the data set. Before preprocess the data set has no such categorization as absence or presence of disease. The num attribute only depicts the output. But after converting the numeric attributes to nominal, the presence and absence of disease is easily analyzed. So nominal conversion of attributes is found to an effective preprocessing technique. The MLP and RBF networks are trained after preprocessing. The results obtained are also having promising results.

The Data is again subjected to Classification by using both MLP and RBF. The prediction accuracy is calculated in both the cases. It is witnessed that MLP outperforms RBF in accuracy. The Kappa Statistic value is higher in MLP. If the Kappa Statistic value is 0.7 or greater than 0.7, then it is said to good statistic correlation. The correlation is found to be better in the case of high Kappa value. Table 3 compares both the network for performance after converting numeric attributes to nominal.

Table 3: Comparing the Performance of RBF with MLP after converting attributes to Nominal

Measures	RBF	MLP
Prediction Accuracy	95%	63%
Kappa statistic	0.9225	0.3
Mean Absolute Error	0.0321	0.1433
Root Mean Squared Error	0.127	0.3569
Relative Absolute Error(%)	12.4102	55.3535
Root Relative Squared Error (%)	35.3664	99.3946
True Positive Rate	0.950	0.637
False Positive Rate	0.039	0.305
Precision	0.950	0.648
Recall	0.950	0.637
F-measure	0.950	0.576
ROC	0.989	0.893

From the table 3, it is inferred that both MLP and RBF perform better. But MLP outperforms Radial Basis Function network in accuracy and also in Relative Absolute Error. It is also observed that the RAE value is better after preprocessing the data set. Replace Missing value is another preprocessing technique used in the analysis. After replacing the value, the network is again trained to evaluate its performance. Multilayer Perceptron network outperforms Radial Basis Function network. It is revealed from the following figures.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

```

Time taken to build model: 0.03 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correlation coefficient           0.4107
Mean absolute error              0.4128
Root mean squared error          0.4543
Relative absolute error          83.1312 %
Root relative squared error      91.1763 %
Total Number of Instances       303
    
```

Figure 6 (a)

```

Time taken to build model: 0.58 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correlation coefficient           0.8779
Mean absolute error              0.1521
Root mean squared error          0.2496
Relative absolute error          30.6207 %
Root relative squared error      50.0898 %
Total Number of Instances       303
    
```

Figure 6(b)

Figure 6(a) & (b): Training the RBF and MLP network after ReplaceMissing Values

VII CONCLUSION

Heart disease prediction is a major challenge in the health care industry. Selecting less number of attributes without affecting the accuracy of diagnosis is a challenging task in Data Mining. Removing and correcting all the noisy data and extracting information from the medical data would help medical practitioners in many ways. Apart from the removal of noisy data, feature extraction is a significant task for prediction of Cardio Vascular Disease. It is observed from the experiments that the preprocessing of data yields promising results. The preprocessing of data enhances the diagnosing and prediction accuracy and it was nearly 91%. The MLP network prediction is has high accuracy with low error rates when compared with RBF. Extracting relevant features and proper training of the network will result in more promising diagnosis. The future direction in the research is the extraction of relevant Redunt feature which would further improve the prediction accuracy.

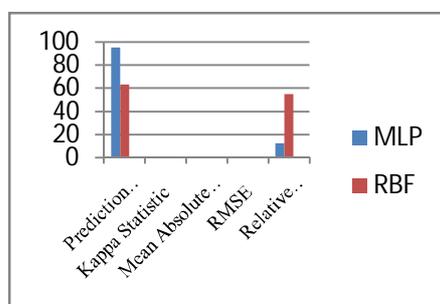


Figure 7: MLP's Performance over RBF

REFERENCES

1. 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), pp 5370-5376, 2010.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

2. 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.
3. AshaRajkumar 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.
4. BalaSundar V, "Development of Data Clustering Algorithm for predicting Heart", IJCA, Vol 48(7), pp 8-13, June 2012.
5. BhagyashreeAmbulkar and VaishaliBorkar "Data Mining in Cloud Computing",MPGINMC, Recent Trends in Computing, ISSN 0975-8887, pp 23-26, June 2012.
6. Bhuvanewari. R, "Naïve Bayesian Classification Approach in Health Care Application", International Journal of Computer Science and Telecommunication, Volume 3(1), pp 106-112, Jan 2012.
7. Carlos Ordonez, Edward Omincenski and Levien de Braal "Mining Constraint Association Rules to Predict Heart Disease", Proceeding of 2001, IEEE International Conference of Data Mining, IEEE Computer Society, ISBN-0-7695-1119-8, pp: 433-440,2001.
8. Cengizcolak.M ,Cemizcolak and HasanKocaturk "Predictingcoronary artery disease using different artificial neural network models", CAD and Artificial neural network, pp 249-254, 2008.
9. 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.
10. Chen A.H., "HDPS: Heart Disease Prediction System", Computing in Cardiology, ISSN 0276-6574, pp 557-560, 2011.
11. Deepika. N, "AssociationRule for Classification of Heart Attack patients", IJAEST, Vol 11(2), pp 253-257, 2011.
12. Durairaj.M, and Meena.K" A Hybrid Prediction System using Rough Sets and Artificial Neural Network", International Journal of Innovative Technology and Creative Engineering, Vol 1(7), July 2011.
13. Myocardia_infraction:en.wikipedia.org/wiki/myocardial_infarction.
14. UCI machine learning repository: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>: Last visited 18th March, 2014
15. Artificial Neural Network: <http://en.wikipedia.org/wiki/Artificial-neural-network>
16. DTREG- Predictive Modeling Software: <http://www.dtrek.com/mlfn.htm>
17. Jabbar M.A., "Knowledge discovery from mining association rules for Heart disease Prediction", JATIT, Vol 41(2), pp 166-174, 2012.
18. JyothiSoni, Uzmaansari and Dipesh Ansari "Intelligent and Effective Heart Disease Prediction System using Weighted Associate Classifier", IJCSE, Vol 3(6), pp 2385-2392, June 2011.
19. Jyothi. S, Ujma.A, Dipesh. S and Sunita. S "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", IJCA, Vol 17(8), pp 43-48, March 2011.
20. K.Rajeswari, "Predictionof Risk Score for Heart Disease in India using Machine Intelligence",IPCSIT, Vol 4, 2011.
21. Kavitha K.S, "Modeling and designing of evolutionary neural network for heart disease prediction", IJCSI, Vol 7(5), pp 272-283, September 2010.
22. LathaParthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Sciences, Vol 3(3), pp157-160, 2007.
23. Liangxiao. J, Harry.Z, Zhihua.C and Jiang.S "One Dependency Augmented Naïve Bayes", ADMA, pp 186-194, 2005.
24. Mia Shouman, "Usingdata mining techniques in heart disease diagnosis and treatment", 978-1-4673-0483-2, Japan-Egypt Conference on Electronics, Communications and Computers, pp 189-193, 2012.
25. Milan Kumari and SunilaGodara, "Comparative Study of Data Mining Classification Methods in Cardio-Vascular Disease Prediction", IJCST, Vol 2(2), June 2011.
26. Milan Kumari and SunilaGodara, "Reviewof Data Mining Classification Model in Cardio Vascular Disease diagnosis", IJCA, 2011.
27. N. Suguna, Dr. K.Dhansushkodi, " A Novel Rough set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization", Journal of Computing, Vol 2(6), pp 49-54, June 2010.
28. Nidhi Bhatia and KiranJyothi, "A Novel Approach for heart disease diagnosis using Data Mining and Fuzzy logic", IJCA, Vol 54(17), pp 16-21, September 2012.
29. Nithya N.S, Sarumathi. S and Dr. Duraisamy. K " Assessment of the risk factors of Heart Attack using frequent feature Selection Method", International Journal of Communications and Enggineering, Vol 1(1), ISSN 0988-0382, pp 127-133, March 2012.
30. PiyqwqkJeatrakul, KokWai Wong and Chun Che Fung, "Using MisClassification Analysis for Data Cleaning", International Workshop on Advanced Computational, Intelligence and Intelligent Informatics, Tokyo, 2009.
31. QeetharaKadhim Al. Shayea, "Artificialneural networkin Medical Diagnosis", IJCSI, Vol 3(2), March 2011.
32. R. Sethukkarase 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.
33. RafiahAwang and Palaniappan. S "Intelligent Heart Disease Prediction System Using Data Mining techniques", IJCSNS, Vol 8(8), pp 343-350, Aug 2008.
34. RafiahAwang and Palaniappan. S "Web based Heart Disease Decision Support System using Data Mining Classification Modeling techniques", Proceedings of IIWAS, pp 177-187, 2007.
35. Raghu. D.Dr, "Probability Based Heart Disease Prediction using Data Mining Techniques", IJCST, Vol 2(4), pp 66-68, Dec 2011.
36. Santhi. P, "Improving the performance of Data Mining Algorithm in Health Care data", IJCST, Vol 2(3), 2011.
37. 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.
38. ShantakumarB.Patil, "Intelligent and Effective Heart Attack Prediction System using Data Mining and Artificial Neural Network", European Journal of Scientific Research, Vol 31(4), pp 642-656, 2009.
39. Shanthakumar B. Patil, "Extraction of Significant patterns from Heart Disease Ware Houses for Heart Attack Prediction", IJCSNS, Vol 9(2), pp 228-235, Feb 2009.
40. Sheik Abdullah, " A Data Mining Model to predict and analyse the events related to Coronary Heart Disease using Decision Tree with Particle Swarm Optimization for Feature Selection", IJCA, Vol 55(8), pp 49-55, october 2012.
41. 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.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2014

42. Siri Krishnan Wasan, VasuthaBhatnagar and HarleenKaur "The Impact of Data Mining techniques on medical diagnostics", Data Science Journal, Vol 5(19), pp 119-126, October 2006.
43. 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.
44. Srinivas, Kavitha Rani and Dr. Govarthan, "Application of Data Mining Techniques in Health Care and Prediction of Heart Attack", IJCSE, Vol 2(2), pp 250-255, 2010.
45. Subbulakshmi, Ramesh and ChinnaRao "Decision Support in Heart Disease Prediction System using Naïve Bayes", IJCSE, ISSN 0976-5166, Vol 2(2), May 2011.
46. SubhagataChatropadhyay, "Mining the risk of heart attack: A comprehensive study", International Journal of BioMedical Engineering and Technology, Vol 1(4), 2013.
47. Sudha.A, Gayathri.p and Jaishankar. N "Utilization of Data Mining Approaches for prediction of life Threatening Disease Survivability", IJAC (0975-8887), Vol 14(17), March 2012.
48. Usha. K Dr, "Analysis of Heart Disease Dataset using Neural network approach", IJDKP, Vol 1(5), Sep 2011.
49. Sivagowry.S, Dr. Durairaj.M, " PSO-An intellectual Technique for Feature Reduction on Heart Malady Anticipation Data", UARCSSE, Vol 4(10), September 2014.

BIOGRAPHY

Dr. Durairaj.M is currently working as a Assistant Professor, Department of Computer Science, Engineering & Applications, Bharathidasan University, Trichy, Tamilnadu, India. He completed his Ph.D. in Computer Science as a full time research scholar at Bharathidasan University on April, 2011. Prior to that, he received master degree (M.C.A.) in 1997 and bachelor degree (B.Sc. in Computer Science) in 1993 from Bharathidasan University. Prior to this assignment of Assistant Professor in Computer Science at Bharathidasan University, he was working as a Computer Programmer at National Research Centre on Rapeseed-Mustard (Indian Council of Agricultural Research), Rajasthan, India, and as a Technical Officer (Computer Science) at the National Institute of Animal Nutrition and Physiology (ICAR), Bangalore for 12 years. He has published 20 research papers in national and International journals.

Sivagowry.Sis currently a Ph.D Research Scholar, Department of Computer Science, Engineering and Application, Bharathidasan University, Trichy, Tamilnadu, India. Her research comforts are Network Security, Data Mining, Soft Computing and Swarm Intelligence. She has circulated 6 research articles in Transnational Conferences and in Journals.