International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Review on Heart Disease Diagnosis Based on Data Mining Techniques

Punam Bajaj[1], Preeti Gupta[2]

[1]Assistant Professor, Department of CSE, Chandigarh Engineering College, Mohali, Punjab, India

[2]Research Scholar, Department of CSE, Chandigarh Engineering College, Mohali, Punjab, India

Abstract: ***The availability of huge amount of data leads to the need of powerful data analysis tool to extract useful knowledge. To manage data analysis on large data sets researchers have long been concerned with statistical and data mining tools. Disease diagnosis is one of the major applications where data mining tools are showing successful results. Heart disease is the major cause of the death all over the world in the last few years. Many researchers are using different data mining tools to help professionals in the diagnosis of heart disease. Using single data mining technique in heart disease diagnosis has been showing great levels of accuracy. In the past few time researchers have been investigating the effect of hybridizing more than one technique showing better result in the diagnosis of heart disease. In this paper the neural network is trained with selected significant pattern for the diagnosis of heart disease and Genetic Algorithm has been used and applied for optimizing the neural network.***

Keywords: Data Mining, Diagnosis, Heart Diseases, Genetic Algorithm, Neural Networks.

1. Introduction

Heart is the significant part of our body. Life is fully dependent on efficient working of heart. If functioning of heart is not proper, it will also influence the other body parts such as brain, kidney, etc. Heart Diseases are the major cause of deaths in the world. Several factors which increase risk of Heart Diseases such as cholesterol, high blood pressure, lack of physical exercise, smoking and obesity.

The World Health Organization has estimated that 12 million deaths occur due to the cardio vascular diseases. Half of the deaths in the United States and other many developed countries occur due to cardiovascular diseases [5]. It is also the major reasons of deaths in many other developing countries. On the whole it is considered as the primary reason behind deaths in adults.

In order to reduce the risk of Heart Disease, prediction should be done. Discovering of Heart Disease is usually based on symptoms, physical examinations and signs of patient body. Normally, doctors are predicting heart disease by knowledge and experience. Discovering and predicting diseases is a difficult task in medical environment.

Discovering Heart Disease from several factors is a multi layered problem which may lead to negative presumptions and unpredictable effects. As a result, Healthcare industry today creates a large amount of complex data about patients, hospital resources, disease diagnosis, system patient records, medical devices etc. The large amount of data is a primary resource to be processed and analyzed for knowledge extraction that enables support for cost- savings and decision making. Figure 1, illustrates the difficulties that will arrive during diagnosis which leads to negative presumptions and unpredictable effects.

Now a day's Artificial Neural Network has been used widely for complex and difficult tasks. The Neural Network is generally trained from the past data with the expectation that it will discover hidden dependencies and that it will be able to use them for predicting. Feed-forward neural networks trained by back- propagation have become a standard technique for classification and prediction tasks. Neural Network has its own advantages and it has proved its worth in the medical field with great potentiality. Hence Neural Network is widely used in health care industry.

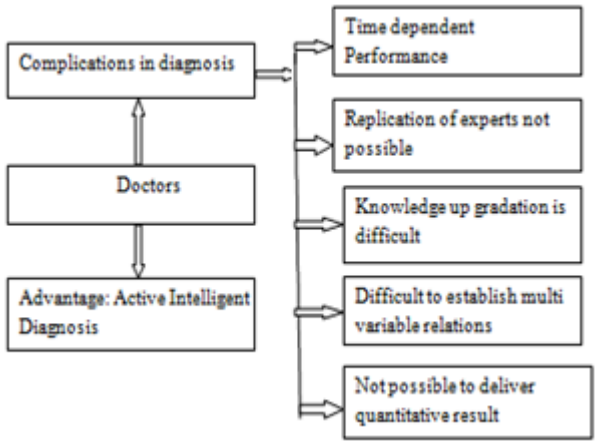

Figure 1: Complexity in Diagnosis with Doctor

2. Literature Review

Numerous works have been done related to heart disease diagnosis using different data mining techniques. The dataset, algorithms, methods used by the authors and the observed results along with the future work is carried out in finding out efficient methods of medical diagnosis for various diseases. Here is a brief discussion about the work which has already carried out in the past few years.

2.1 Heart disease diagnosis using Classification Methods

Milan Kumari [7] diagnose cardiovascular disease by using different data mining algorithms such as: Support Vector Machine, Artificial Neural Network (ANN), Decision Tree and RIPPER classifier. The authors analyze the performance of these algorithms through several analysis factors such as: Sensitivity, Accuracy, Error Rate, True Positive Rate and

False Positive rate. Accuracy of RIPPER, Decision Tree, ANN and SVM are 81.08%, 79.05% 80.06% and 84.12% respectively. While the results of error rates for RIPPER, Decision Tree, ANN AND SVM are 2.756, 0.2755, 0.2248 and 0.1588 respectively. The result shows that from these four classification models SVM predicts cardiovascular disease with least error rate and high accuracy. A prototype Intelligent Herat Disease Prediction System (IHDPS) based on data mining techniques is proposed by Sellappan Palaniappan [11]. The techniques used are Decision Trees, Naive Bayes and Neural Network. The implementation has been done on .NET platform. Dataset have several attributes such as age, sex, blood pressure and blood sugar which are used to predict the probability of patients getting a heart disease.

2.2 Heart Attack Prediction System Using Clustering

Shantakumar B. Patil [12] applied efficient methodology for the extraction of significant patterns from the heart disease warehouses for heart attack prediction. In this firstly the data warehouse is pre-processed in order to make it suitable for the mining process and secondly the K-mean clustering algorithm has been applied for clustering the heart disease warehouse. Hence the recurrent patterns applicable to heart disease are mined with the MAFIA algorithm from the data extracted. In addition, the patterns necessary to heart attack prediction are selected on the basis of computed significant weightage. The neural network is trained with the selected important patterns for the effective prediction of Heart Attack.

2.3 Heart Disease Prediction System Using Feature Selection

The prediction of Heart Disease, Blood Pressure and Sugar with the help of neural network was proposed by Niti Guru [9]. Tests were carried out on a specimen database of patient records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and etc. The supervised network has been focussed for diagnosis of heart diseases. Training of the data is done with the help of back propagation algorithm. Whenever unknown data was fed by the doctor, the system identified the unknown data from comparisons with the trained data and generate list of probable diseases that the patients may prone to. Swati Shilaskar [13] has proposed a technique to predict the presence of cardiovascular disease accurately with reduced number of attributes. They investigated the intelligent system to construct feature subset with enhancement in diagnostic performance. They proposed a hybrid forward selection model to diagnose cardiovascular disease. Their experiment demonstrated that their technique found smaller subsets and enhanced the accuracy of diagnosis contrast to forward inclusion and back elimination models.

2.4 Heart Disease Prediction System Using Naive Bayes

Mrs. G. Subbalakshmi, Mr. K. Ramesh and Mr. M. Chinna Rao [8] developed a Decision Support in Heart Disease Prediction System using Naive Bayes Data Mining technique to discover relations that connects variable in a database. Using medical attributes such as age, sex, blood sugar and blood pressure it can predict the probability of patients getting heart diseases. This model could answer complex queries and is resulted out as the most effective model in prediction of heart diseases.

2.5 Heart Disease Diagnosis Using Fuzzy Logic Approach

P.K. Anooj [10] has proposed a weighted fuzzy rule based CDSS for the diagnosis of heart disease. It automatically obtains the knowledge from the patient clinical data. The proposed CDSS for risk of heart patients consists of two phases. First is an computerized approach for generation of weighted fuzzy rules and decision tree rules and the second is creating a fuzzy rule based decision support system. The performance of the proposed CDSS improved the risk prediction when compared with the neural network based clinical support system. Latha Parthiban [6] formulated an approach for the prediction of heart disease on the basis of coactive neuro- fuzzy inference system (CANFIS). The CANFIS model combined neural network capabilities and fuzzy logic approach which is then integrated with genetic algorithm to diagnose the presence of heart disease and the results showed that the proposed CANFIS model has great potential in prediction of heart disease.

2.6 Heart Disease Prediction Using Association Rule

Improving Heart Disease Prediction Using Constrained Association Rule by Carlos Ordonez proposed the problem for heart disease prediction by identifying constrained association [4]. The data set focussed on medical records with attributes for risk factors of people having heart disease. Three limitations were introduced to decrease the number of patterns. They are as follows:

- The attributes have to appear on only one side of the rule.
- Separate the attributes into groups i.e. uninteresting groups.
- In a rule, there should be limited number of attributes.

The result of this is two groups, determines the presence or absence of heart disease.

2.7 Heart Disease Prediction System using Hybrid System

Cardio Vascular Disease prediction System using Neural Network and Genetic Algorithm is proposed by **Amma, N.G.B.** In this Genetic based Neural Network is used for training the system. The neural network final weights are stored in the weight base and are used for diagnosis the risk of cardio vascular disease. The accuracy of classification obtained using this approach is 94.17% [1]. Genetic Neural Network based data mining in prediction of Heart Disease using risk factors is proposed by Syed Umar Amin, Kavita Aggarwal and R. Beg [14]. In this paper two data mining techniques are used Genetic Algorithm and Neural Network to predict the risk of heart disease with an accuracy of 89%. The hybrid system is implemented using the optimization advantage of genetic algorithm and has been proved better than back propagation in terms of stability and accuracy.

2.8 Heart Disease Diagnosis Using Regression techniques

Decision Support System for Medical Diagnosis Using Data Mining is proposed by D. Senthil Kumar, G. Sathyadevi and S. Sivanesh [3]. In this paper C4.5, ID3 and CART algorithm are used to classify the diseases and compare the effectiveness and correction rate among them. After obtaining the comparison results CART performs better result in terms of accuracy and complexity. Early Diagnosis of Heart Disease using Classification and Regression technique is proposed by Amiri, A.M [2]. The goal of this research is to implement heart sounds diagnostic system with the reduced number of echocardiograms and prevent the release of newborns that are affected by heart disease. When CART technique is applied on the data set following results are obtained: 99.14% accuracy, 100% sensitivity and 98.28% specificity.

3. Review of Proposed Methodology

Neural Networks is now a days the most promising area of interest. It is believed that for all the biomedical problems Neural Networks will going to be the great solution in the coming years. Already it has been applied to various domains of medicine such as biochemical analysis, diagnostic system, drug development and image analysis. Neural Networks is a current research area at the moment. It will never replace human experts but they can help in screening and can be used by the experts to confirm their diagnosis. The advantages of such system are remarkable. People can be checked for heart diseases quickly and painlessly and thus detecting the disease at an early stage.

In our research work we are going to use hybrid data mining techniques in the diagnosis of heart disease patients. It investigates if integrating Neural Network with Genetic Algorithm can provide better performance than the normal neural network in the diagnosis of heart disease patients.

The diagnosis process consists of two succeeding steps i.e. training and testing. When performing analysis with a set of existing data, split the data set into 70 % – 30 % for training and testing purpose respectively. The objective of this study is to measure the accuracy of this neural network and compare it with accuracy of the optimized neural network which is optimized using swarm intelligence algorithm. The Figure 2 shows the flow of work that will going to be carried out during our research.

Steps which will be followed in our research are as follows:

1) To develop a system and to optimize it to get more accuracy after testing.
2) To compare results of normal neural network and optimized neural network.
3) Swarm intelligence algorithm i.e. Genetic Algorithm will be used for optimization.
4) To develop better and more accurate proposed system.

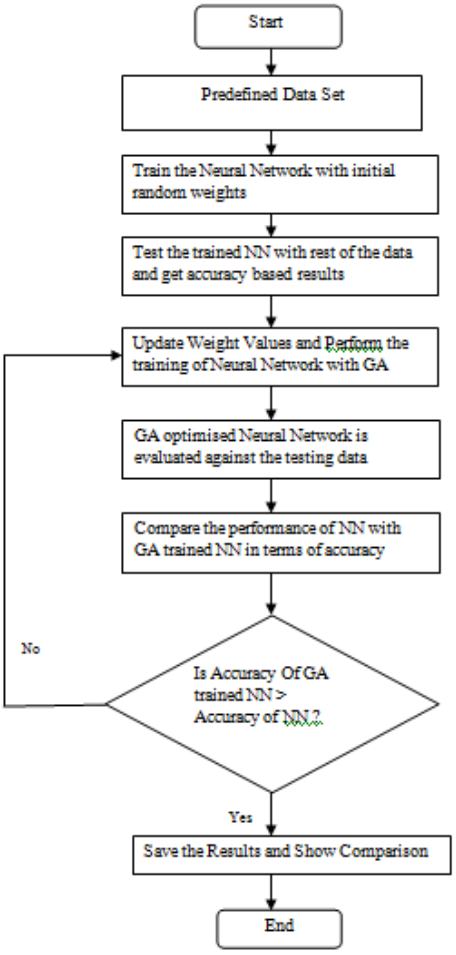

Figure 2: Basic flowchart depicting training of Neural Network with GA

4. Conclusion

In this paper, we have discussed that how different types of data mining techniques are used for diagnosis of heart diseases and also studied that how these techniques have performed better results when applied on different data sets. Each technique is unique in its own way, which might be suitable for different applications. Also applying hybrid data mining techniques has shown promising results in the diagnosis of heart disease, so applying hybrid data mining techniques in selecting suitable treatment for heart disease patients needs further investigation. Hence neural network can be effectively used in Heart Diseases Diagnosis. In our future work, we will be training the neural network by using Swarm Intelligence Optimization i.e. Genetic Algorithm to optimize the system to get better results and more accuracy than the normal neural network.

References

[1] Amma, N.G.B, "Cardio Vascular Disease Prediction System using Genetic Algorithm and Neural Network", IEEE International Conference on Computing Communication and Applications 2012.
[2] Amiri, A.M., "Early Diagnosis of Heart Disease using Classification and Regression Trees", International Joint Conference on Neural Networks, PP 1-4, August 2013.

[3] D.Senthil Kumar, G.Sathyadevi and S.Sivanesh, "Decision Support System for Medical Diagnosis using Data Mining", International Journal of Computer Science Issues, Vol.8 No.1, May 2011.
[4] Carlos Ordonez, "Improving Heart Disease Prediction using Constrained Association Rules", Seminar Presentation at University of Tokyo, 2004.
[5] Jyoti Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Herat Disease Prediction", International Journal of Computer Applications, Vol.17, No.8, March 2011.
[6] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological Biomedical and Medical Sciences, Vol.3 No.3, 2008.
[7] Milan Kumari, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction", International Journal of Computer Science and Telecommunications, Vol.2, Issue 2, June 2011.
[8] Mrs.G.Subbalakshmi, "Decision Support in Heart Disease Prediction System using Naive Bayes", ISSN:0976-5166, Vol.2 No.2 (Apr-May 2011).
[9] Niti Guru, Anil Dhaiya, Navin Rajpal, "Decision Support System for Heart Disease using Neural Network", Delhi Business Review, Vol.8, No.1 (January-June 2007).
[10] P.K. Anooj, "Clinical Decision Support System: Risk Level Prediction of Heart Disease using Weighted Fuzzy Rules", Journal of Computer Sciences, Vol.24, PP.27-40, 2012.
[11] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008.
[12] Shantakumar B. Patil, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack prediction", International Journal of Computer Science and Network Security, Vol.9, No.2, February 2009.
[13] Swati Shilaskar, "Feature Selection for Medical Diagnosis: Evaluation for Cardiovascular Diseases", Journal of Expert System with Application, Vol.40, PP.4146-4153, 2013.
[14] Syed Umar Amin, Kavita Aggarwal, Dr. Rizwan Beg, "Data Mining in Clinical Decision Support and Treatment of Heart Disease", International Journal of Advanced Research in Computer Science and Technology, Vol.2 Issue.1, January 2013.

Author Profile

Punam Bajaj received M. Tech and M. Phil degree in year 2007 and 2008 respectively. Currently she is an Assistant Professor in Chandigarh Engineering College, India

Preeti Gupta is presently working as a Research Scholar in Chandigarh Engineering College. She received her B. Tech degree in year 2011 and pursuing M. Tech from Chandigarh Engineering College, India