A Closure Looks to Data Mining Techniques for Prediction of Heart Disease

Rajwant Kaur[1], Sukhpreet Kaur[2]

[1]M.Tech Student, Department of Computer Science & Engineering SGGSWU, Fatehgarh Sahib

[2]Assistant Professor, Department of Computer Science & Engineering SGGSWU, Fatehgarh Sahib

Abstract: *Data mining techniques have been widely used in medical field for prediction and diagnosis of various diseases. These techniques are very popular and effective to design clinical decision support systems because these have the ability to find out hidden patterns and medical data relations. One of most important application of such systems is diagnoses of Heart Diseases. All over the world, heart disease is one of the common disease leading causes of deaths. Many Medical systems are developed to predict disease on the basis of risk factors such as age, high cholesterol, diabetes, hypertension, family history, tobacco smoking, alcohol intake etc. Researchers have been helping to medical professionals to make Heart Disease prediction systems using data mining techniques. These data mining techniques are used to get accurate and better results. These techniques based systems have been helping medical professional to predict heart diseases, on the basis of risk factors, with good accuracy, high quality of service and without costly medical tests.*

Keywords: Heart disease, Neural Network, Risk factors, Genetic Algorithm, Medical system.

1. Introduction

Data mining is a knowledge discovery technique to analyze data and summarize it into useful information [8]. The analyzed data can be used for various applications. One of the most important applications is use in medical field. Nowadays as the population of world increases, medical industries generates huge amount of data related to patients and diseases diagnoses etc. and facing major challenges. A major challenge of medical industries is to diagnose disease accurately and high cost of medical tests. Data mining techniques are used to solve these challenges, with high quality of service. Data mining provides techniques to find out hidden patterns and relations of medical data [9] [10].

The mortality rate caused by heart disease has been increasing all over the world so there is need for development of heart disease prediction methods. According to one survey in 2008 approximately 17.3 million people died from Heart diseases, representing 30% of all global deaths. An estimated 7.3 million deaths were due to coronary heart disease and 6.2 million were due to stroke [9]. Recently, many researches in medical industry has been able to identify risk factors of heart diseases, but more contribution is necessary to use this knowledge to reduce causes of deaths [9]. Various data mining techniques has been used to make clinical decision support systems, to get accurate results on the basis of information collected by researches from study. These systems allow patients to calculate the heart disease risks [10]. IHDPS (intelligent heart disease prediction system) is capable to discover and extract hidden knowledge associated with heart disease on the basis of historical heart disease database [11].

a) **Risk factors of heart disease:**
- **Smoking**: Smoking is major reason of heart disease. The chemicals in tobacco smoke harm blood cells. They also can damage the function of heart and the structure and function of blood vessels. This causes to heart attack [5].
- **Obesity**: Extra weight increases risk for heart function. This condition is characterized by changes to the structure of the heart's main pumping chamber (left ventricle), which prevent it from filling sufficiently between beats. It causes to diseases like heart disease, stroke, diabetes and more [5].
- **High Blood Pressure**: High blood pressure also known as HBP or hypertension is a widely misunderstood medical condition. High blood pressure increase the risk of the walls of our blood vessels walls becoming overstretched and injured [5].
- **High blood cholesterol**: When there is high amount of cholesterol (a fat-like substance) in your blood, it builds up in the walls of arteries. This causes arteries become narrowed and blood flow to the heart is slowed down or blocked. The blood carries oxygen to the heart, and if enough blood and oxygen cannot reach your heart, it causes heart attack [5].
- **Poor Diet**: The foods chosen to consume on a daily basis can affect the health of heart. Choosing the right foods can make a difference in living a long and healthy life or helps to face with the diseases like a heart attack or heart disease [6].
- **Physical Inactivity:** Lack of exercise is a risk factor for developing coronary artery disease (CAD). Lack of physical exercise increases the risk of CAD, because it also increases the risk for diabetes and high blood pressure [5].

2. Literature Review

Amin et al [1] in (2013) have proposed a technique for prediction of heart disease using major risk factors. That technique was involved two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implemented with use of the global optimization advantage of genetic algorithm for initialization of neural network weights. The system was implemented to predict the risk of heart disease with an accuracy of 89%. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis.

Sudha et al [2] in (2012) have discussed that Stroke is a life threatening disease that has been ranked third leading cause of death in states and in developing countries. Number of work has been carried out for predicting various diseases by comparing the performance of predictive data mining. The classification algorithms like Decision Tree, Naive Bayes and Neural Network is used for predicting the presence of stroke disease with related number of attributes. In this paper, the principle component analysis (PCA) is applied to reduce the dimensions. It determined the attributes involved more towards the prediction of stroke disease and predicts whether the patient is suffering from stroke disease or not.

Abdullah et al [3] in (2012) have developed the data mining model by using Random Forest classifier to improve the prediction accuracy and to investigate various events related to CHD. That model could help the medical practitioners for predicting CHD with its various events and how it might be related with different segments of the population. The events investigated are Angina, Acute Myocardial Infarction (AMI), Percutaneous Coronary Intervention (PCI), and Coronary Artery Bypass Graft surgery (CABG). Experimental results had shown that classification using Random Forest Classification algorithm could successfully used in predicting the events and risk factors related to CHD.

Kaur et al [4] in (2014) have illustrated that the huge amounts of data generated for prediction of heart disease are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. By using data mining techniques it takes less time for the prediction of the disease with more accuracy.

3. Data Mining Techniques for Heart Disease

Data Mining contains various types of techniques that are used for heart diseases predictions. These techniques make the clinical support systems more efficient. Some of these classification techniques are described as follows:

a) Genetic Algorithm: -
Genetic algorithms (GA) visualize search method based on Evolutionary ideas of natural selection and genetics. They simulate natural process based on principles of Lamark and Darwin. The field of Genetic and Evolutionary computation was first developed by Turing. Turing Holland performed much of the fundamental work in GEC in 1960 and 1970. When an optimization problem is specified, Genetic algorithms are implemented using computer simulations. Genetic algorithms works in an iterative manner by generating new populations of strings from old ones .Every string is encoded in binary, real etc., versions of a candidate solution. An evolution function associates a fitness measures to every string indicating its fitness for the problem [7]. The three genetic operators are: -
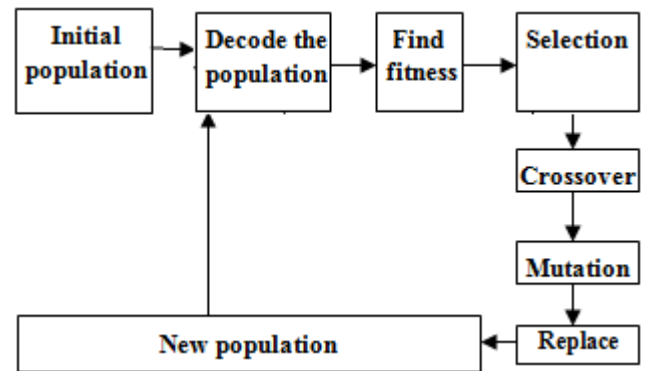
- Selection
- Crossover
- Mutation

Figure 1: Block diagram of genetic algorithm

b) Artificial neural network (ANN)
Artificial neural networks were originally developed in 1995, by researchers who were trying to mimic the neurophysiology of the human brain. With the combination of many simple computing elements (neurons or units) into a highly interconnected system, a complex phenomenon such as intelligence was tried to produce. In simple words, An Artificial Neural Network is the simulation of human brain and is being applied to an increasingly number of real world problems. Neural Networks as a tool used to mine knowledgeable data from data ware house. ANN is trained to recognize, store and retrieve patterns to solve combinatorial optimization problems. Pattern recognition and function Estimation abilities make ANN prevalent utility in data mining. Neural Networks consist of three layers- 1) input layer 2) hidden layer and 3) output layer [12, 13].

Table 2: Techniques used for heart disease prediction

S.No.	Algorithm Name	Description
1.	GeneticNeural Network[1]	This technique is used for prediction of heart disease using major risk factor. It involves two data mining tools: neural network, genetic algorithm. This hybrid system uses genetic algorithm, global optimization, for initialization of neural network weight. The advantage of this system is that it identified important risk factor for the prediction of heart disease and it does not require costly medical test.
2.	KNN+Genetic Algorithm[6]	This approach combines KNN with genetic algorithm to improve the classification accuracy of heart disease data set. It used genetic search as a goodness measure to prune redundant and irrelevant attributes, and to rank the attributes which contribute more towards classification. Least ranked attributes are removed, and classification algorithm is built based on evaluated attributes. This classifier is trained to classify heart disease data set as either healthy or sick.
3.	SVM-Decision Trees-Logistic Regression (SDL)[10]	A Rule based model is proposed to compare the accuracies of applying rules to the individual results of Support Vector Machine, Decision Tree, and Logistic Regression on the Cleveland Heart Disease Database in

		order to present an accurate model of predicting heart disease. Using the Cleveland Heart Disease database, this technique provides guidelines to train and test the system and thus attain the most efficient model of the multiple rule based combinations
4.	Genetic Algorithm[7]	A method for heart disease prediction is proposed using genetic algorithm. This proposed work will mine association rules from heart disease data more efficiently. The association rule mining problem is an NP hard problem because finding all frequent item sets having minimum support results in a search space of 2m which is exponential in m where m is number of item sets. Genetic algorithm is applied to optimize association rules. Thus fitness function is obtained for the task of optimizing the number of rules generated by efficiently pruning redundant Rules.

Table.1 describes the techniques used for prediction of heart diseases. In this table, some of techniques are discussed. Beside this, there are so many techniques that are used to diagnose heart diseases.

4. Conclusion

In this paper, various data mining classification techniques for heart disease are discussed. The study of these techniques shows that main issue in this field is accurate results of heart disease prediction on the basis of risk factors. Beside this high cost of medical tests is also the main issue. All these techniques overcome the problem of accurate results on the basis of information collected through risk factors. Now days, Neural Network is highly useable method for prediction of heart diseases. The study of various techniques results, neural network is best method as compare to others. By enhancing these techniques, better results can be achieved and cost of medical tests can also be reduced.

References

[1] Amin, Syed Umar, Kavita Agarwal, and Rizwan Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors", *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, pp. 1227-1231. IEEE, 2013.

[2] Sudha, A., P. Gayathri, and N. Jaisankar, "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods", *International Journal of Computer Applications* 43, no. 14, pp. 26-31, 2012

[3] Abdullah, A. S., and R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier", *International Conference in Recent Trends in Computational Methods, Communication and Controls*. 2012.

[4] Cinetha, K., and P. Uma Maheswari, "Decision Support System for Precluding Coronary Heart Disease (CHD)", *International Journal of Computer Science and Mobile Computing,* Issue.2, Vol.3, pp. 34-38, 2014

[5] Kaur, Beant, and Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", *International Journal on Recent and Innovation Trends in Computing and Communication,* Issue: 10, no.2, pp. 3004-3008, 2014

[6] Deekshatulu, B. L., and Priti Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm", *Procedia Technology* 10,pp. 85-94, 2013

[7] Jabbar, M. A., B. L. Deekshatulu, and Priti Chandra, "An Evolutionary Algorithm for Heart Disease Prediction", *Wireless Networks and Computational Intelligence*. Springer Berlin Heidelberg, pp. 378-389, 2012

[8] Masethe, Hlaudi Daniel, and Mosima Anna Maseth. "Prediction of Heart Disease using Classification Algorithms", *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 2. 2014.

[9] Waghulde, Nilakshi P., and Nilima P. Patil, "Genetic Neural Approach for Heart Disease Prediction", *International Journal of Advanced Computer Research* 4.3 pp. 778, 2014

[10] Mythili, T., Dev Mukherji, Nikita Padalia, and Abhiram Naidu, "A heart disease prediction model using svm-decision trees-logistic regression (sdl)", *International Journal of Computer Applications* 68, no. 16 pp.11-15, 2013

Paper ID: SUB156090

2931