

SURVEY OF HEART DISEASE PREDICTION BASED ON DATA MINING ALGORITHMS

S.B.BHALERAO¹, DR. B.L.GUNJAL²

¹ME Student, Department of Computer Engineering, Amrutvahini COE, Sangamner, Maharashtra, India

²Associate Professor, Department of Computer Engineering, Amrutvahini COE, Sangamner, Maharashtra, India

ABSTRACT

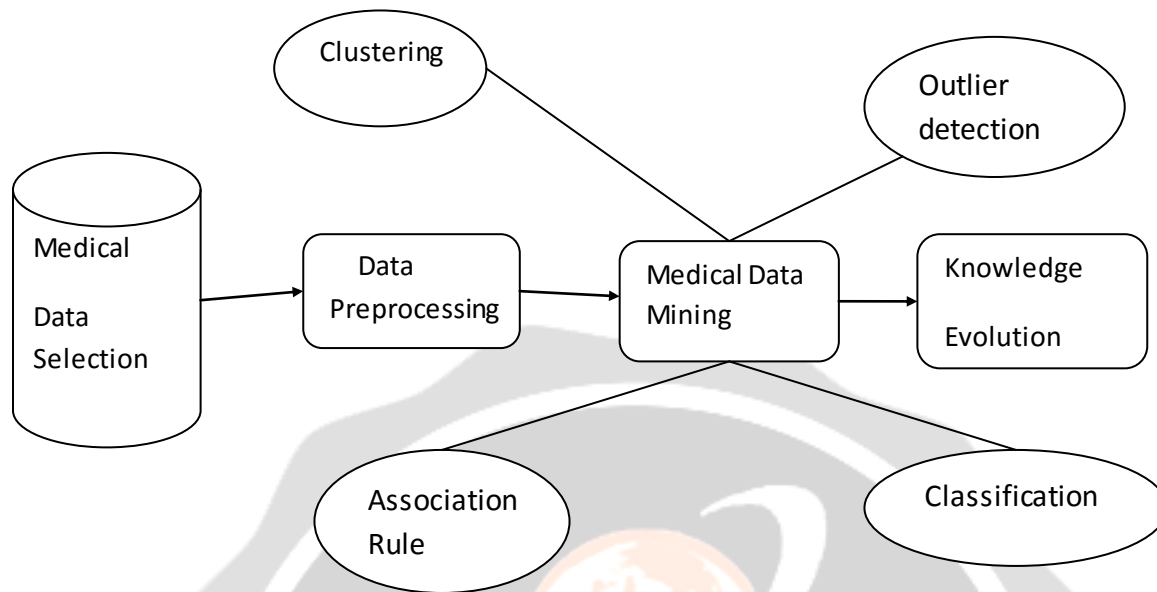
The medical field is dealing with huge amount of data regularly. Handling that large data by traditional way may affect the results. Advanced data mining techniques are used to find out facts in database and for medical research, particularly in Heart disease prediction. Heart disease is the leading cause of death all over the world. The massive amounts of data generated for prediction of heart disease which is too difficult and baggy to be processed and analyzed by conventional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. Use of data mining algorithms will result in quick prediction of disease with high accuracy. In this paper we survey different papers in which single or hybrid combination of data mining algorithms are used for the prediction of heart disease, so that we can identify the algorithms with high accuracy for further research.

Keyword: - Data mining; Heart disease prediction; Data mining techniques

1. Introduction

1.1 Data mining and medical industry

Data mining is technique of exacting implicit, previously unknown and future useful information from available data. In short it is process of analyzing data from different viewpoint and assembling the knowledge from it. To extract unknown patterns and relationships from huge databases, data mining mixes the statistical analysis, machine learning and database technology. The acquire knowledge can be useful for different application namely health care industry. Health Informatics is a rapidly growing field that is concerned with evolving Computer Science and Information Technology to medical and health data. Medical Data Mining is a domain of challenge which involves lot of misdiagnosis and uncertainty. A general framework proposed for medical data mining is shown in Fig.1 [2].



1.2 Heart disease

Heart is an important organ of all living individual, which plays an essential role of blood pumping to the rest of the organs through the blood vessels of the circulatory system. If circulation of blood in body is improper the organs like brain suffer and if heart stops working altogether and death occurs. Life is completely dependent on proper working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it.

Factors that increase the risk of Heart disease:

- Family history
- Age
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

In many cases, diagnosis is generally based on patient's current test report & doctor's experience. Thus the diagnosis is a complex task that requires high skill & much experience.

2. Literature Survey

Carlos Ordonez [3] did study on prediction of heart disease with the help of Association rules. They used a simple mapping algorithm. This algorithm constantly treats attributes as numerical or categorical. This is used to convert medical records to a transaction format. An improved algorithm is used to mine the constrained association rules. A mapping table is prepared and attribute values are mapped to items. The decision tree is used for mining data because they automatically split numerical values [3]. The split point chosen by the Decision tree are of little use only. Clustering is used to get a global understanding of data.

Deepika [4] have worked on Pruning-Classification Association Rule (PCAR), for mining Association Rule. PCAR based on analysis of Apriori algorithm. PCAR deals with minimum frequency item sets. It deletes irregular item from item sets. Then it classifies item sets based on frequency of item sets and find out the frequent item sets.

C. Ordonez [12] uses association rules a technique in data mining to get better heart disease prediction result .The author have worked on the limitation of association rule which is nothing but mining the entire data set without validation on an independent sample. The modified algorithm with search constraints was introduced to trim down the number of association rules and validated using train and test approach. They have studied two complementary tasks: predicting the absence and predicting the existence of heart disease.

P. Chandra, M.Jabbar [13] produced class association rules using feature subset selection to detect a heart disease. Association rule determines relations between attributes values and classification to predict the class in the patient dataset [13]. Feature selection measures like genetic search determines attributes which helps in predicting heart diseases.

Usha Rani [5] have proposed a system for predicting heart disease with the help of artificial neural network which is a combination of feed forward and back propagation algorithm. The experiment is carried out by considering single and multilayered neural network models. Parallelism is implemented to speed up the learning process at each neuron in all hidden and output layer.

In [6] R. Setthukkarasi, have developed a novel neuro fuzzy technique to diagnose the fact of the disease from the set of the patient report. A generalized database is constructed for decision making from the reduced attributes set which is output of genetic algorithm. A four layered fuzzy neural network for efficient modeling and reasoning with temporal dependencies under uncertainty is used.

In [8] Chaitrali Dangare has implemented system to predict heart disease three data mining classification techniques were applied that is Decision trees, Naive Bayes & Neural Networks. From results it has been seen that Neural Networks better than Decision trees & Naive Bayes.

M.Akhil jabbar, B.LDeekshatulu, Priti Chandra [7] propose a new algorithm which combines KNN with Genetic Algorithm for effective classification. To provide optimal solution genetic algorithms perform global search on complex large and multimodal dataset. From the results it is also observed that hybridizing GA with KNN performs well and give great accuracy.

Ankita Dewan [1], proposed efficient genetic algorithm mix with the back propagation technique approach for heart disease prediction. They have made conclusion that neural network is best among all the classification techniques for a non-linear data. BP algorithm is the best classifier of Artificial Neural Network which is a common method of training. In this the primary system output is compared to the desired output, and the system is adjusted until the difference between the two is minimized. But it has drawback of being stuck in a local minima.

Shadab Adam Pattekari and Asma Parveen [9] developed a Decision Support in Heart Disease Prediction System using Naive Bayesian Classification technique. The system discovers the hidden knowledge from a past heart disease database. This is the most effective model to predict patients with heart disease. This model could respond to complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy.

R. Bhuvaneswari et al., [10], use Naive Bayes classifier in medical applications. The author have used two well-known algorithms of data mining classification that is Back propagation Neural Network(BNN) and Nave Bayesian (NB) which calculate the probability of the object among all objects by studying the previous experience. Bayesian technique is developed on the probability concept. The posterior from the prior is calculated by bayes rules based on the precise nature of the probability model, Naive Bayes classifiers is used to trained very efficiently in a supervised learning setting.

Nilakshi P. Waghulde, Nilima P. Patil [11] did an experiment with Heart Disease dataset by taking into consideration of Multi-Layer Neural Network along with Back propagation Learning Algorithm for training the network. To optimize the initialization of neural network weights genetic algorithm is used. This work shows the

result of Genetic Neural Network for prediction of heart disease by improving the accuracy as 98% using optimize neural network architecture it predicts whether the patient is suffering from heart disease or not.

K. Srinivas [14] proposed Application of Data Mining Technique in the prediction of Heart Attacks. The author have studied classification based data mining techniques such as Rule based, Decision tree, Naive Bayes and Artificial Neural Network to the massive Volume of healthcare data. Tanagra data mining tool was used for examining data analysis, machine learning and statistical learning algorithms. They have used the training data set of 3000 instances with fourteen different attributes. The instances in the dataset are representing the outcome of different types of testing to predict the accuracy of heart disease.

Sudha et al. [15] studied the classification algorithm like Naive Bayes, Decision tree and Neural Network for detecting the stroke diseases. The classification technique like decision trees, Bayesian classifier and back propagation neural network were used in this study. The records with unrelated data were deleted from data warehouse before mining process occurs.

Dr. D. P. Shukla et. al. [16] performed a work of, designing a system which could recognize the probability of a coronary heart disease. He has separated all parameters into two levels based on the criticality of the parameter and assigned each level a separate weight age. Finally both the levels are taken into consideration to take a final decision. The author has proposed neuro-fuzzy integrated approach at two levels. In this the error rate is very low and work efficiency is high. The author have made conclusion that this same method could be used to perform the analysis on some other diseases also.

3. CONCLUSIONS

In this survey paper we have studied the various exiting technique for predicting heart disease with the support of data mining. From this survey we got the knowledge about how to apply data mining technique to predict the heart disease. Previously exiting system was design mostly with single algorithm which is not providing good accuracy, from some researchers work we have observed that by hybridization of two or more algorithm we can acquire more accuracy.



4. ACKNOWLEDGEMENT

It gives me an immense pleasure to express my sincere and heartiest gratitude towards my guide Dr.B.L.Gunjral for her guidance, encouragement, moral support and affection during the course of my work. I am also thankful to all who have contributed indirectly and materially in words and deeds for completion of this work.

5. REFERENCES

- [1]. Ankita Dewan, Meghna Sharma," Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706.
- [2]. R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, R. Boghrati, *et al.*, "Diagnosis of coronary arteries stenosis using data mining," *J Med Signals Sens*, vol. 2, pp. 153-9, Jul 2012.
- [3]. 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, 2001, pp: 433-440.
- [4]. Deepika. N, "Association Rule for Classification of Heart Attack patients", IJAEST, Vol 11(2), pp 253-257, 2011.
- [5]. Usha. K Dr, "Analysis of Heart Disease Dataset using neural network approach", IJDKP, Vol 1(5), Sep 2011.
- [6]. R. Setthukkarase, 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.
- [7]. M Akhil Jabbar, BL Deekshatulu, Priti Chandra," Heart disease classification using nearest neighbor classifier with feature subset selection", Anale. Seria Informatica, 11, 2013.
- [8]. Chaitrali S Dangare, Sulabha S Apte." Improved study of heart disease prediction system using data mining classification techniques", International Journal of Computer Applications, 47(10):44-48, 2012.
- [9]. Shadab Adam Pattekari and Asma Parveen," PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624, Vol 3, Issue 3, 2012, pp 290-294.

- [10]. R. Bhuvaneswari and K. Kalaiselvi, "Naïve Bayesian Classification Approach in Healthcare Applications", International Journal of computer Science and Telecommunication", vol. 3, no. 1, pp. 106-112, Jan 2012.
- [11]. Nilakshi P. Waghulde, Nilima P. Patil,"Genetic Neural Approach for Heart Disease Prediction", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ,Vol 4 Number-3 Issue-Sept 2014.
- [12]. C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction.," IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society, vol. 10, no. 2, pp. 334-43, Apr. 2006.
- [13]. P. Chandra, M. Jabbar, and B. Deekshatulu, "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection," in 12th International Conference on Intelligent Systems Design and Applications (ISDA), 2012, pp. 628-634.
- [14]. Srinivas K, Rani KB, Govrdhan A.," Application of data mining techniques in healthcare and prediction of heart attacks", International Journal on Computer Science and Engineering. 2011,2(2):2505.
- [15]. Sudha A, Gayathiri P, Jaisankar N," Effective analysis and predictive model of stroke disease using classification methods", International Journal of Computer Applications. 2012; 43(14):26-31.
- [16]. Ashish Kumar Sen, Shamsheer Bahadur Patel, Dr. D. P. Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671.

	<p>Miss Shubhada B. Bhalerao Perusing masters in Computer Engineering at AVCOE, Sangamner. She has received his Bachelors in Computer Engineering from Cummins college of Engineering, Pune. She is a member of Association of Computer Machinery (ACM).</p>
	<p>Dr. B. L. Gunjal is Associate Professor in Amrutvahini College of Engineering, Sangamner. Received PhD degree from University of Pune. Her Research interests Includes Image Processing, Data Mining. She is having 18 Years teaching experience In Computer Engineering. She is Having 25 international and national journals. And conferences publications. She is Awarded as "BEST TEACHER" by SPPU.</p>