



Review on Hybrid Data Mining Techniques for The Diagnosis of Heart Diseases in Medical Ground

KEYWORDS

Data mining, Hybrid Data mining Algorithm, Neural Network, Artificial intelligence, K-map

Rishi Dubey

School of Information Technology, MATS University, Raipur, Chhattisgarh, INDIA

Santosh chandrakar

School of Information Technology, MATS University, Raipur, Chhattisgarh, INDIA

ABSTRACT This paper, presents a systematic review of the application of Data Mining methods in healthcare domain, with a focus on the application and the techniques used which will optimize the results. These methods are new approaches to solve the problems in healthcare domain. Data mining applications can have tremendous potential and usefulness. However, the success of data mining techniques on the availability of clean data. In this respect, it is critical that the industry look into how data can be better captured, stored, prepared and examined. Whenever we are using hybrid data mining techniques in the real data we get more and more accurate data for the system.

1. Introduction

Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases is given as follows: —Data mining is the non-trivial extraction of implicit previously unknown and potentially useful information about data [17]. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. The discovered knowledge can also be used by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives. Applications of data mining techniques that can be used in health care management. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Health care data is massive. It includes patient centric data, resource management data and transformed data. Health care organizations must have ability to analyze data. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several problems.

The availability of integrated information via the huge patient repositories, there is a shift in the perception of clinicians, patients and payers from qualitative visualization of clinical data by demanding a more quantitative assessment of information with the supporting of all clinical and imaging data.[12]

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a

process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data mining is a logical process that is used to search through large amount of data in order to find useful data.

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.

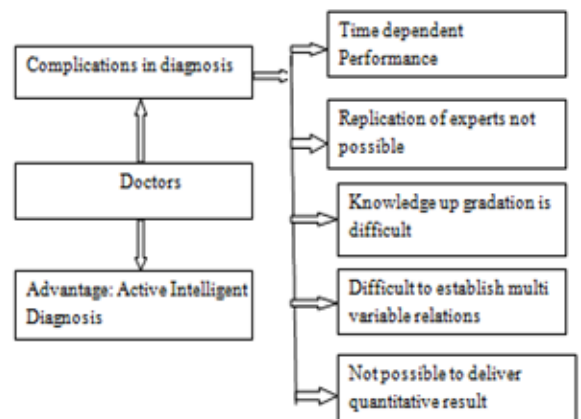


Fig:Complexity in Diagnosis with Doctor

DATA MINING TECHNIQUES TO FIND OUT HEART DISEASES: AN OVERVIEW

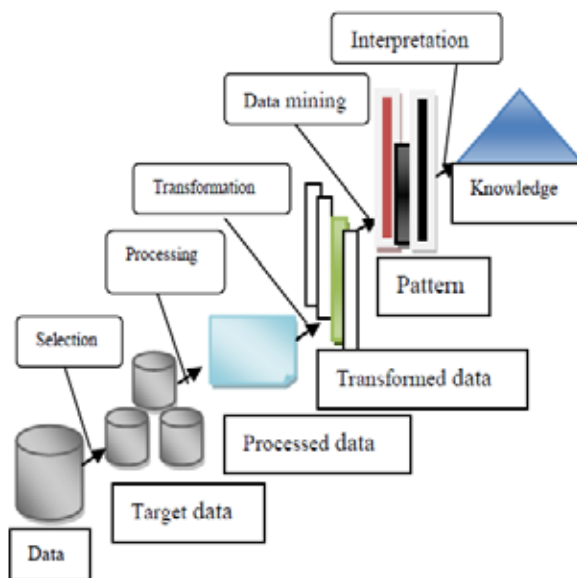
Once these patterns are found they can further be used to make certain decisions for development of their businesses. Different steps are

1. Exploration
2. Pattern identification
3. Deployment

Exploration: In the first step of data exploration data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.

Pattern Identification: Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction. **Deployment:** Patterns are deployed for desired outcome.

Data mining with intelligent algorithms can be used to tackle the said problem of prediction in medical dataset involving multiple inputs. Now a day's Artificial neural network has been used for complex and difficult tasks. The neural network is trained from the historical data with the hope 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.



DATA MINING ALGORITHMS AND TECHNIQUES

Various algorithms and techniques present in data mining for finding out heart diseases like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

Classification learning:- The learning algorithm takes a set of classified examples (training set) and use it for training the algorithms. With the trained algorithms, classification of the test data takes place based on the patterns and rules extracted from the training set. Types of classification models:

Classification by decision tree induction

- i. Bayesian Classification
- ii. Neural Networks
- iii. Support Vector Machines (SVM)
- iv. Classification Based on Associations

Numeric prediction:- This is a variant of classification learning with the exception that instead of predicting the discrete class the outcome is a numeric value.

Association rule mining:- The association and patterns between the various attributes are extracted and from these attributes rules are created. The rules and patterns are used predicting the categories or classification of the test data.

Clustering:- The grouping of similar instances in to clusters takes place. The challenges or drawbacks considering this type of machine learning are that we have to first identify clusters and assign a new instance to these clusters.

HEART DISEASE PREDICTION USING DATAMINING

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 infinding out efficient methods of medical diagnosis for various diseases.

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 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 report and etc. The supervised network has been focused 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.

Mrs. G. Subbalakshmi, Mr. K. Ramesh and Mr. M. Chinnarao [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.

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. Improving Heart Disease Prediction Using Constrained Association Rule by Carlos Ordóñez proposed the problem for heart disease prediction by identifying constrained association [4]. The data set focused 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.

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 of the risk of cardiovascular 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.

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 Heart Disease Prediction System (IHDP) 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.

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.

HEART DISEASE WITH HYBRID DATA MINING TECHNIQUES

Motivated by the world-wide increasing mortality of heart disease patients each year and the availability of huge amounts of data, researchers are using data mining techniques in the diagnosis of heart disease. Although applying data mining techniques to help health care professionals in the diagnosis of heart disease is having some success, the use of data mining techniques to identify a suitable treatment for heart disease patients has received less attention.

Applying hybrid data mining techniques we can show 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 and Genetic Algorithm to optimize the system to get better results and more accuracy than the normal neural network.

Some of the objectives are:-

- 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 (hybrid system).
- 3) Hybrid system will be used for optimization.
- 4) To develop better and more accurate proposed system.

Neural Networks is now a days the most promising area of interest. It is believed that for all the biomedical problems Neural Networks will go 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. The diagnosis process consists of two succeeding steps i.e. training and testing.

This survey defines the proposal of building a hybrid methodology, combining data mining techniques such as association rules and classification trees. The methodology is applied to data collected from a hospital and is evaluated by comparing with other techniques. The methodology is expected to help physicians to make faster and more accurate decisions. In our future work, we will be training the neural network by other data mining techniques and back propagation method to optimize the system to get better results and more accuracy than the normal neural network.

CONCLUSION

This survey presents a systematic review of the application of Data Mining methods in the healthcare domain, with a focus on the application and the techniques used which will optimize the results. These methods are new approaches to solve the problems in healthcare domain. In this literature survey we present an overview of the current research being carried out using the data mining techniques for the diagnosis and prognosis of different heart diseases. In our work 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 diseases.

Data mining applications can have tremendous potential and usefulness. However, the success of data mining techniques on the availability of clean data. In this respect, it is critical that the industry look into how data can be better captured, stored, prepared and examined. Whenever we are using hybrid data mining techniques in the real data we get more and more accurate data for the system. The advantages of such system are remarkable. By using accurate data in the health care system, people can be checked for heart diseases quickly and painlessly and thus detecting the disease at an early stage.

REFERENCE

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. JyotiSoni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications, Vol.17, No.8, March 2011. | 6. NidhiBhatlaKiranJyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), 2012. | 7. ShantakumarB.Patil, Dr.Y.S. Kumaraswamy, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", (IJCSNS) International Journal of Computer Science and Network Security, 2009. | 8. Abhishektaneja, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Scientific Publishing Co., India, 2013. | 9. M. Anbarasi, E. Anupriya, N.ch.s.n.lyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology, 2010. | 10. [http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies palace/datamining.htm](http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies%20palace/datamining.htm) | 11. Silwattanusarn, Tipawan, and KulthidaTuamsuk. "Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012." arXiv preprint arXiv:1210.2872 (2012) | 12. Shouman, M.; Turner, T.; Stocker, R., "Using data mining techniques in heart disease diagnosis and treatment," Electronics, Communications and Computers (JEC-ECC), 2012 Japan-Egypt Conference on , vol., no., pp.173,177, -9 March 2012 | 13. Robu, R.; Hora, C., "Medical data mining with extended WEKA," Intelligent Engineering Systems (INES), 012 IEEE 16th International Conference on , vol., no., pp.347,350,13-15 June 2012 | 14. Stewart, A.; Herder, E.; Smith, M.; Nejd, W., "A user study on public health events detected within the medical ecosystem," Digital Ecosystems and Technologies Conference (DEST), 2011 Proceedings of the 5th IEEE International Conference on , vol., no., pp.127,132, May 31 2011- June 3 2011 | 15. Shen, Z., Clarke, M. Jones, R. Alberti, T. —A neural network approach to the detection of coronary artery disease, IEEE Transaction on Computers in Cardiology, 5-8 Sep 1993. | 16. Karaolis, M.A. Moutiris, J.A. Hadjipanayi, D. Pattichis, C.S., — Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees IEEE Transactions on Information Technology in Biomedicine, pp 559 -566, May 2010. | 17. Frawley and Piatetsky-Shapiro, 1996. Knowledge Discovery in Databases: An Overview. The AAAI/MIT Press, Menlo Park, C.A. | 18. Miller, A., B. Blott and T. Hames, 1992. Review of neural network applications in medical imaging and signal processing. Med. Biol. Engg. Comp., 30: 449-464 | 19. Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers. In Proc. of UAI-99, pp.101-108, 1999. | 20. Shaikh Abdul Hannan, V.D. Bhagile R. R. Manza, R. J. Ramteke, — Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation technique , International Journal on Computer Science and Engineering, pp 2150-2159, 2010. | |