

Detection of Heart Disease using Classification Algorithm

Sumesh Harale

Department of Computer Engineering,
Atharva College of Engineering,
University of Mumbai,
Mumbai-400095, India.

Jay Nirmal

Department of Computer Engineering,
Atharva College of Engineering,
University of Mumbai,
Mumbai-400095, India.

Abhijeet Singh Dhillon

Department of Computer Engineering,
Atharva College of Engineering,
University of Mumbai,
Mumbai-400095, India.

Neha Kunte

Assistant Professor
Department of Computer Engineering,
Atharva College of Engineering,
University of Mumbai,
Mumbai-400095, India.

Abstract— Modern Lifestyle has resulted in a significant increase in cardiovascular disease. The diagnosis and treatment for this sum up to a huge amount. To overcome this problem many research have been made to reduce the cost and deliver quick results. The study of existing systems has resulted in the discovery of the best data mining classification algorithm among various other algorithms. The algorithms used for comparison were Naïve Bayes, REPTREE, J48 and SIMPLE CART in which SIMPLE CART proved to be comparatively accurate. This comparison was done on the basis of eleven attributes which were Patient Identification Number (replaced with dummy values), Gender, Cardiogram, Age, Chest Pain, Blood Pressure Level, Heart Rate, Cholesterol, Smoking, Alcohol consumption and Blood Sugar Level. Basically the existing systems have only done comparisons among classification of algorithms. Based on the previous researches done, this paper would use “SIMPLE CART” algorithm which will be tried to integrate in a web application for detection of cardiovascular diseases. The aim of this real time web application is to produce quick and accurate results at low cost. This web application will be developed on the Visual Studio Platform and SQL SERVER will be used for database management. The technology used will be Asp.Net using MVC framework.

Keywords— Cardiovascular disease, Data mining, SIMPLE CART, Web Application

I. INTRODUCTION

All the diseases or conditions that affect the heart are generally termed as heart diseases. Now days due to advancement in technology most of the hospitals store their patient’s information and medical issues in an information system. Huge amounts of data which taken in the form of numbers, text, and images are stored and managed by these systems. But, all this data is most of the time never used to support clinical decisions. This gives rise to an important question that how can this data be turned into useful content so that it can help the medical practitioners to take important decisions in the medical field or support these decisions in an effective way. So this is the main question which needs to be crated so that all this important data can be used instead of just consuming memory.

II. RELATED WORK

The research paper [1] performed a work, Problems with Mining Medical Data”. Data mining methods are expected to find interesting patterns from databases, thus beneficially using the stored information from the databases and making it useful for the medical practitioners to derive results. In this paper, we focus on the characteristics of medical data and discuss how data miners deal with medical. This paper [2] performed a work, “Medical Knowledge Acquisition through Data Mining”. Data mining has been widely considered as an effective tool for knowledge discovery. This paper [2] discusses the important role of medical experts for medical data mining, and presents a model of medical knowledge acquisition through data mining.

V.Chaurasia provides a paper which resolves cardiovascular disease dataset using different data mining algorithms, such as Support Vector Machine, Artificial neural networks (ANNs), Decision Tree, and RIPPER classifier. The author in this paper [3] has analyzed the performance of the given algorithms by referring several statistical analysis factors such as accuracy and error rate. Accuracy of RIPPER, Decision Tree, ANN and SVM are 81.02%, 79.08%, 80.01% and 84.16% respectively. While the results of error rates for ANN, SVM, RIPPER and Decision Tree are 0.2248, 0.1588, 2.756 and 0.2755 respectively. The most optimized algorithm among all the above compared is SVM as it produces the highest accuracy and the least error. The research based on the published paper [6] also proposed a new approach for association rule mining based on sequence number and clustering transactional data set for heart disease predictions. C programming requires less main memory and hence the proposed model was implemented in C programming language which resulted it to be scalable and efficient model as the time constraint was also minimized due to less memory requirement [6].

K. Sudhakar created class association rules using feature subset selection to predict a model for heart disease. Association rule determines relations amongst attributes values and classification predicts the class in the patient dataset [10]. Feature selection measures such as genetic search determines attributes which contribute towards the prediction of heart diseases. Waikato Environment for Knowledge Analysis (WEKA) has been useful in prediction due to its predicting patterns and analysis [9].

This paper [5] used pattern recognition and data mining methods in predicting models in the domain of cardiovascular diagnoses. An experiment was conducted using Neural Network, K-NN, Naïve Bayes and Decision Tree and results proved that Naïve Bayes technique outperformed other techniques that were used [5].

S. Vijayarani and S. Sudha in the published paper [7] which used association rules for representation of a technique in data mining to improve disease prediction. An algorithm with search constraints was also introduced to reduce the number of association rules and validated using train and test approach [7].

The research work in the paper [4] leads to the use Naive Bayes classifier in medical applications. Backpropagation Neural Network (BNN) and Naive Bayesian (NB) are the two of the well-known algorithms used in data mining classification that calculate the priors, the probability of the object using information based on the previous experience among all objects. Bayesian technique is constructed on concept of probability. The posterior from the prior is calculated by bayes rules. Depending on the precise nature of the probability model, Naive Bayes classifier is used to train efficiently in a supervised learning setting.

Two evolutionary data mining algorithms termed as GA-KM and MPSO-KM cluster are used in this paper [8] which deal with the cardiac disease data set and predict model accuracy. Momentum-type particle swarm optimization (MPSO) and K-means technique are combined by this hybrid method. C5, Naïve Bayes, K-means, Ga-KM and MPSO-KM were compared for evaluating the accuracy of the techniques. The results were more accurate when GA-KM was used [8].

III. METHADODOLOGY

The aim of prediction methodology is to design a model using Simple Cart Classification algorithm which will produce output or results based on the 11 attributes provided by the medical practitioner. The previous research paper applies following algorithms: - Naïve Bayes, REPTREE, J48 and SIMPLE CART in which SIMPLE CART proved to be comparatively accurate. Simple Cart method is CART (Classification and Regression Tree) analysis. CART is stated as Classification and Regression Tree algorithm. It was developed by Leo Breiman in the early 1980s. It is used as regression tree in classification methods which in order are used to construct decision trees using historical data. CART uses learning sample which is a set of historical data with pre-assigned classes for all observations for building decision tree. Simple Cart is a classification technique that generates the

binary decision tree. Since output is binary tree, it generates only two children. Entropy is used to choose the best attribute between the two. Simple Cart handles the missing data by ignoring that record. This algorithm is best for the training data. Classification and regression trees (CART) is a learning technique, which gives the results as either classification or regression trees, depending on categorical or numeric data set. Its methodology proposed by is perhaps best known and most widely used. Simple Cart uses cross-validation or a large independent test sample of data to select the best tree from the sequence of trees considered in the pruning process. In the implementation of CART, the dataset is divided into the two subgroups that are the most different with respect to the outcome. This procedure is continued on each subgroup until some minimum subgroup size is reached. Fig.1. shows working of a SIMPLE CART application.

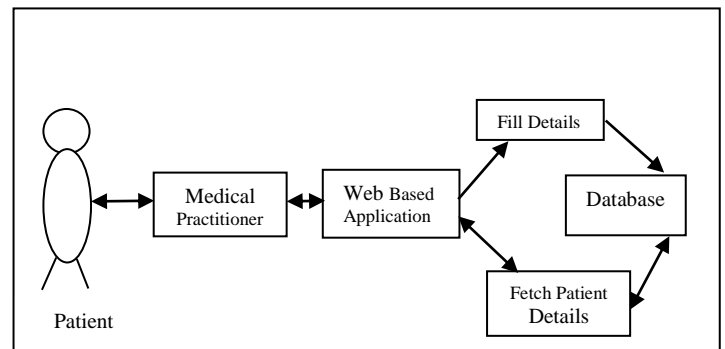


FIG. 1. Block Diagram

IV. DECISION TREE MODEL

Decision trees are represented by a set of questions which divides the learning sample into small parts. Fig.2. shows SIMPLE CART Decision tree. According to the fig. SIMPLE CART asks only yes/no questions. A possible question could be: “Is age greater than 50?” or “Is sex female?”. CART algorithm will search for all possible variables and all possible values in order to find the best split - the question then divides the data into two parts with maximum homogeneity. This process is then repeated for each of the resulting data fragments. Following is an example of a data set which is to be selected to detect heart disease in a person’s body.

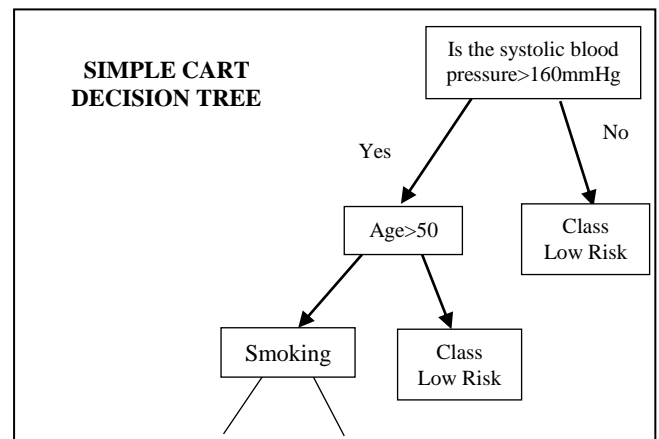


Fig. 2. Decision Tree

CART can easily handle both numerical and categorical variables. Cart methodology consist of three parts: Construction of maximum tree, Choice of the right tree size, Classification of new data using constructed tree.

V. CONCLUSION

The term Heart disease covers all diseases and disorders related to heart and blood vessels. Nowadays there is an increase in heart disease. The diagnosis and treatment for this sums up to a huge amount. In solution to this problem many research have been made to reduce the expense and to get quick result. This model would be able to develop an integrated web based application for detection of heart disease using classification algorithm.

VI. FUTURE SCOPE

This paper depicts the use of web based application integrated with classification algorithm. This is beneficial in field of medical science as it could reduce the load of medical practitioner and treat the patient more efficiently. Also, cardiovascular disease can be discovered at an early stage which reduce the heart attacks among patients. Treatment can be done according to the severity of the detected disease.

VII. REFERENCES

- [1] A. K. Sen, S. B. Patel, and 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*, vol. 2, no. 9, pp. 1663–1671, 2013.
- [2] S. Ishtake and S. Sanap, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," *International Journal of healthcare & biomedical Research*, vol. 1, no. 3, pp. 94–101, 2013.
- [3] V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," *Caribbean Journal of Science and Technology*, vol. 1, pp. 208–217, 2013.
- [4] D. S. Chaitrali and A. S. Sulabha, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks," *International Journal of Computer Engineering & Technology (IJCET)*, vol. 3, no.3, pp. 30–40, 2012.
- [5] M. Jabbar, P. Chandra, and B. Deekshatulu, "Cluster Based Association Rule Mining For," *Journal of Theoretical & Applied Information Technology*, vol. 32, no. 2, pp. 196–201, 2011.
- [6] R. Rao, "Survey On Prediction of Heart Morbidity Using Data Mining Techniques," *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, vol. 1, no. 3, pp. 14–34, 2011.
- [7] S. Vijayarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey," *International Journal of Computer Applications & Information Technology*, vol. II, no. I, pp. 17–21, 2013.
- [8] T. J. Peter and K. Somasundaram, "An Empirical Study On Prediction of Heart Disease Using Classification Data Mining Techniques," 2012.
- [9] K. Sudhakar, "Study of Heart Disease Prediction using Data Mining," vol. 4, no. 1, pp. 1157–1160, 2014.
- [10] A. Aziz, N. Ismail, and F. Ahmad, "Mining Students 'academic Performance,'" *Journal of Theoretical & Applied Information Technology*, vol. 53, no. 3, 2013.