Impact Factor (SJIF): 2.125

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444
Volume 3, Issue 1, January-2016

Pathological detection of the Coronary heart diseases through the technique of Data Mining

Dr. Rituraj Amrawat[1], Mr. Lavesh Gupta[2]

[1]Senior Demonstrator, PBM Hospital, Bikaner, Rajasthan, India
[2]Electronics & Communication Engineer, Bikaner, Rajasthan, India

Abstract

In the modern times, heart diseases have become prevalent all over globe. Many people die every year because of the wrong/misdiagnosis of these heart diseases. Therefore, a need to develop the correct diagnosis of the heart related diseases, based on the data from the previous case scenarios is felt. So, to aid the early and precise diagnosis of Coronary Heart Diseases, fewer data mining techniques are devised using the large amount of data available to medical hospitals. Many algorithms devised are used to assist physicians in diagnosis of the heart diseases, based on the different patient parameters like fasting blood sugar, previous heart events, age, etc. This research paper analyses the various techniques which include decision tree algorithm, the Naive Bayes algorithm and neural networks algorithm, and lists the good and bad in each and every one of them.

Keywords-*Coronary Heart Disease, Decision trees, Naive Bayes and Neural Networks.*

I. INTRODUCTION

In this research paper, we are dealing with three data mining algorithms- decision tree, Naïve Bayes and the neural network, that will be used to classify the data sets based on the patient history and assist the physicians in diagnosing coronary heart diseases. Therefore, mining the patient data and using it for the prediction of possible heart diseases in forthcoming data can save many lives.

- In the *decision tree algorithm,* historical data has been used to form the conditional tree as a machine learning method. Afterwards, algorithm will start from root node each time and proceed stepwise based on set of patient parameters entered, till it reaches any of the end node. This end node shows probability of person suffering from a specific coronary heart disease.
- *The Naïve Bayes* algorithm uses different patient parameters and finds probability of a patient having a coronary heart disease, based on individual parameters. These individual probabilities later on combined by some formulae is used to find probability of patient suffering from the coronary heart disease, based on all parameters combined.
- *Neural network algorithm* help us to train the large amount of patient data, even when input symptoms are less. These Neural Networks give more importance to the inputs with the higher weights, which makes more prominent parameters of the greater value.

II. ALGORITHMS

In the given section, we proceed further to explain each of the data mining techniques:

2.1.1 DECISION TREE:

These trees are developed by the breakdown of large data sets into subset. But at same time, they also look after relation between the data sets. They are traversed in the top-down approach. In Learning phase of algorithm, a set of the historical data is taken. Using this data, a tree is formed that satisfies all constraints given in data.

Afterwards, the algorithm is used to obtain probability of a patient suffering from specific heart condition by means of a number of steps. The patient parameters such as age, sex, fasting blood sugar level, drinking and smoking habits, and a set of other symptoms(which are predefined in decision tree) are taken during the patient check-up. This collected data is then entered into algorithm, which then starts from root node and proceeds further to each corresponding node, based on the values of a few parameters. This process continues until patient parameter set is exhausted and the leaf node is reached. Along each step, algorithm gets closer to the final diagnosis and its probability is shown, by the value at the leaf node.

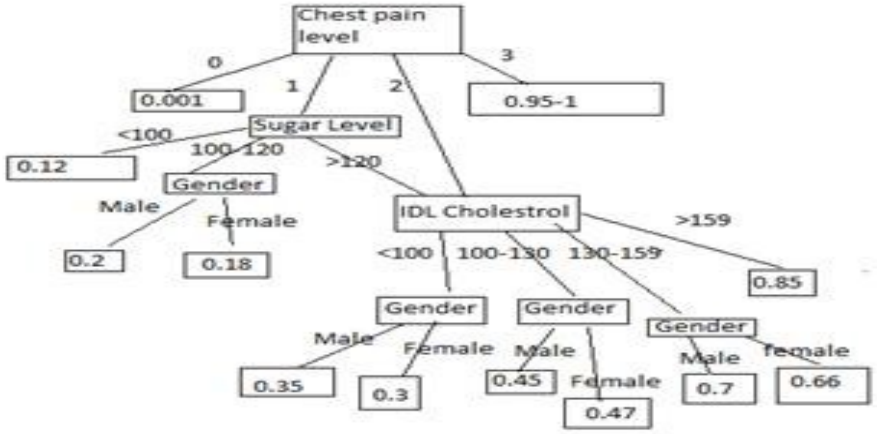

FIGURE 1 DECISION TREE

2.2 NAIVE BAYES:

The given algorithm is used to determine separate probability of a person, suffering from a specific coronary heart disease, based on a few individual parameters. The patient parameters are entered into system. After this, algorithm takes the every single parameter and attributes a probability to it, based on information extracted from historical data, available over a system.

The calculation of probability of a heart disease based on any of the parameter is completely independent of probability of the same for any other parameters. The final probability will be a cumulative function of the average of all values of the weightage times probability. Once each individual probability is determined, we can now devise a certain formula to derive probability of there, being a coronary heart disease based on all parameters combined. Each parameter is given a different weightage depending on importance of that parameter, and hence final probability would be a cumulative function of the average of all the values of weightage time's probability.

Cumulative probability=$\sum[(n-N)*P(W-N)]/\sum[n-N]$

N=0 to d

Where, W-0, W-1, W-2…..W-d are different parameters and n-0, n-1….n-d is the weightage, given to each of the parameters respectively.

2.3 NEURAL NETWORKS:

Neural Networks are able to use their hidden units to learn from the derived features of the task, that is, classification of the patient data and the forthcoming inputs.

This method is always preferred because of the ability to learn from the examples. The inputs to Neural Network would be parameters such as sex, cholesterol, age, etc. After training for the numerous iterations until the error is minimized, we get actual classification. The fresh data which would be fed will be classified with help of previously fed training data, which we got from many hospitals. This process will work in much iteration. Each of the input is initially assigned a random weightage. After that, output is computed. Then, output is checked with real set of desired or the ideal output. The difference between two is called as error. This algorithm aims at removing and minimizing this error with each of the iteration, by adjusting weights of each of the parameter, by the means of some mathematical formulae. Thus, neural networking takes time to compute and would arrive at ideal conclusion, even with the fewer input conditions.

III. RESULTS:

- The algorithms discussed in this paper were- decision trees, Naïve Bayes and Neural networks. As reliability of the decision trees depends on precision of input data from the beginning, this might prove difficult when dealing with the larger data sets.

Though the data used will be from reliable sources, this might prove to be the disadvantage. On other hand, decision tree has the advantage of having comprehensive data.

- Naïve Bayes must be combined with the other statistical techniques to classify best features, but it easily runs on the larger sets of data because it needs learned probability of a heart, even with respect to the every individual parameter and then uses a formula to combine the conditional probabilities to find overall probability of the coronary heart disease, occurring in a person with a specific set of parameters. The Naive Bayes algorithm is fastest of the three discussed in this research paper, but it needs a moderate sized input set.

- Neural networks are used, when input sets are small. This method has higher time complexity but it is the only one method, which will give close ideal results with a limited input data set. The scope for any error is minimal, as almost all errors are removed by numerous numbers of iterations.

Model Type	Prediction Attributes	No. of Cases	Prediction
Decision Tree	+WHD,+PHD	150	Correct
	-WHD,+PHD	23	Incorrect
	-WHD,-PHD	221	Correct
	+WHD,-PHD	60	Incorrect
Naïve Bayes	+WHD,+PHD	181	Correct
	-WHD,+PHD	34	Incorrect
	-WHD,-PHD	210	Correct
	+WHD,-PHD	29	Incorrect
Neural Network	+WHD,+PHD	180	Correct
	-WHD,+PHD	33	Incorrect
	-WHD,-PHD	210	Correct
	+WHD,-PHD	31	Incorrect

+WHD stands for- Patients with heart disease

–WHD stands for-Patients with no heart disease

+PHD stands for- Patients predicted as having heart disease

–PHD stands for- Patients predicted as having no heart disease

Figure-2: Result Table

REFERENCES

[1] Studies on application of Support Vector Machine in diagnose of coronary heart disease -Yan Zhang, Fugui Liu, Zhigang Zhao, Dandan Li, Xiaoyan Zhou and Jingyuan Wang.

[2] Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees- Minas A. Karaolis, Member, IEEE, Joseph A.Moutiris, Demetra Hadjipanayi, and Constantinos S. Pattichis, Senior Member, IEEE.

[3] Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems Eman Abu Khousa, Piers Campbell- Faculty of Information Technology.

[4] Intelligent Heart Disease Prediction System Using Data Mining Techniques- Sellappan Palaniappan and Rafiah Awang, Department of Information Technology, Malaysia University of Science and Technology, Selangor, Malaysia.

[5] AKAMAS: Mining Association Rules Using a New Algorithm for the Assessment of the Risk of Coronary Heart Events- M. Karaolis, Student Member, IEEE, I.A.Moutiris, FESC, L. Papaconstantinou,C.S. Pattichis, Senior Member, IEEE.

[6] A Study on Classification Techniques in Data Mining – G. Kesavaraj, Assistant Professor Department of Computer Applications,Vivekanandha College Of Arts And Sciences For Women, Dr. S. Sukumaran Associate Professor, Department of Computer Science, Erode Arts and Science College, Erode.

[7] Using Data Mining Techniques in Heart Disease Diagnosis and Treatment- Mai Shouman, Tim Turner, Rob Stocker School of Engineering and Information Technology, University of New South Wales at the Australian Defence Force Academy

[8] S.B. Patil and Y.S. Kumaraswamy, Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction. International Journal of Computer Science and Network Security (IJCSNS), VOL.9, No.2, 2009.

[9] Kim, J., et al., A Novel Data Mining Approach to the Identification of Effective Drugs or Combinations for Targeted Endpoints— Application to Chronic Heart Failure as a New Form of Evidence based Medicine, Cardiovascular Drugs and Therapy, Springer 2005, pp. 483–489, 2012.

[10] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman,"Decision trees: An overview and their use in medicine,"J. Med. Syst., vol. 26, no. 5, pp. 445–463, 2002.

[11] Harita Jagad, Jehan Kandawalla and Sindhu Nair, "Detection of Coronary Heart Diseases using Data Mining Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, pp. 6090-6092, VOL.3, NO. 11, 2015.

[12] Tan, G. & Cbye H., "Data mining applications in healthcare," Journal of Healthcare Information Management, Vol. 19, No. 2, 2004.

Mr. Lavesh Gupta is a reputed Electronics & Communication Engineer. In 2013, he Completed his B.tech (Hons.) with distinction. At present, he is a research scholar in the field of Digital Communication. He is a proud author of the research papers, which have already been published in the two International Journals. His research areas include Biomedical Instrumentation, Data Mining, Antennas, Antenna Array & Designing, etc.

Dr. Rituraj Amrawat is a reliable and reputated name in the field of Medicine and is currently serving as a Senior Demonstrator in PBM Hospital, Bikaner, Rajasthan. He is affiliated to S.P. Medical College in Bikaner. Known for his dynamic personality, he is having a vast experience in medical field and Pathology. He has also served the Indian Army as a Major for five years.