

Data Mining Technique in the Field of Medical Science for Determining Heart Disease – Decision Tree

Bhoomika S ,4ai15cs024 ,
Pallavi C H ,4ai15cs069,
Pooja G H ,4ai15cs075,
Pooja S, 4ai15cs076,

Dr. Pushpa Ravi Kumar B
.E.,M.Tech.,
Ph.D,HOD & professor,CS &E Dept,AIT ,
Chikmagalur.

Priyanka N
B.E.,M.Tech,
Asst.Prof, CS & E dept , AIT College,
Chikmagalur.

Abstract:- Heart disease is the leading cause of death among all other diseases, even cancers. Most of the people facing heart disease is on a raise each year. This prompts for its early diagnosis & treatment. Due to lack of resources in the medical field, the prediction of heart disease periodically may be a problem. Utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity & patients. This issue can be resolved by adopting Data mining techniques. This paper intends to adopt Decision tree –a data mining techniques for the effective prediction of Heart disease. It compares the efficiency & accuracy of the two techniques to decide among them the best.

Keywords: Heart Disease, Decision tree, Classification.

I. INTRODUCTION

Today's era is witnessing an alarming ascent in the incidences of numerous life style disorders like Heart disease, Hypertension, Obesity, etc. Among these, Heart disease is on top of the list which is the leading cause of death in most of the countries. It includes diseases of heart muscles, valves, conduction system, heart attack & others. Myocardial infarction or heart attack is the major one among all other types of heart diseases. Heart diseases are seen in all the classes of people in recent times, in contrast to previous days when it was disease of high class people. Heart disease is even being highlighted as a hidden which leads to death of a person without obvious symptoms. This nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented to suit the present day health needs. Data mining techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data they can arrive at a conclusion. This technique can be very well adapted to the do the prediction

of heart disease. As the well-known quote says "Prevention is better than cure", early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

Heart disease

Heart disease is a term used for any type of disorder that affects the heart. Heart disease means the same as cardiac disease. Depending upon the pathology occurred, the heart disease are of various forms.

A. Coronary artery disease

Coronary Artery disease (CAD) is also called by the name Ischemic heart disease (IHD). It comes under cardiovascular diseases which involve the valves of the heart. It comprises of a group of diseases, angina, myocardial infarction & sudden cardiac arrest. The symptoms include, $\frac{3}{4}$ Chest pain / discomfort which radiates to shoulder, neck or back $\frac{3}{4}$ Heartburn $\frac{3}{4}$ Shortness of breath

B. Heart attack

Heart attack is the common term used for Myocardial infarction (MI). It is due to the interruption of blood supply to a part of heart leading to damage to the heart muscle. Chest pain is the common symptom which may pass on to shoulder, neck, back or jaw. It may present in the centre or left side of the chest.

C. Arrhythmia Cardiac

Arrhythmia is also known commonly as irregular heart beat is a group of conditions. Here, the heart beat will be too fast or slow or irregular. Symptoms may not be present many a times. Symptoms may include palpitations or feeling of gap between heart beats.

D. Heart failure

Heart failure is the inability of the heart to pump the blood for the body functioning. Coronary artery disease, heart attack, valve disease etc. can lead to heart failure.

Symptoms are difficulty in breathing, fatigue & leg swelling. Chest pain is not present always in heart failure.

II. LITERATURE SURVEY

The paper by Vanishree. K developed a system for diagnosis of congenital heart disease using decision support system. It used Back propagation Neural Network with MLP. It was based on the data – signs of the heart disease, symptoms & result of evaluation obtained from the patient. It showed 90 % accurate results [3]. The study done by Kharya highlighted the fact that artificial neural network is the frequently used technique for prediction in the medical field. The paper also demonstrated the merits & demerits of the machine learning techniques like Decision tree, Neural network & SVM [4]. C.D. Katsis et al study devised ways using Correlation Feature Selection (CFS) procedure & an Artificial Immune Recognition System (AIRS) classifier to diagnose breast cancer. Data for the study was collected from 53 patients among the 4726 cases. Biopsy was taken in all the patients & it was kept as standard parameter to validate the methodology. The features along with the biopsy result were used for the analysis in 53 patients. Adoption of SVM technique resulted in 70.00 + 6.33% accurate result [5].

III METHODOLOGY

This paper aims to predict the Heart disease based on the data set values stored by using the MSQl database as backend server. Using this training data set values it is possible to predict the presence or absence of Heart disease in that particular patient record. Here classification techniques like Decision tree algorithm is utilized. It is a web based application. In the front end .NET framework which acts as client.

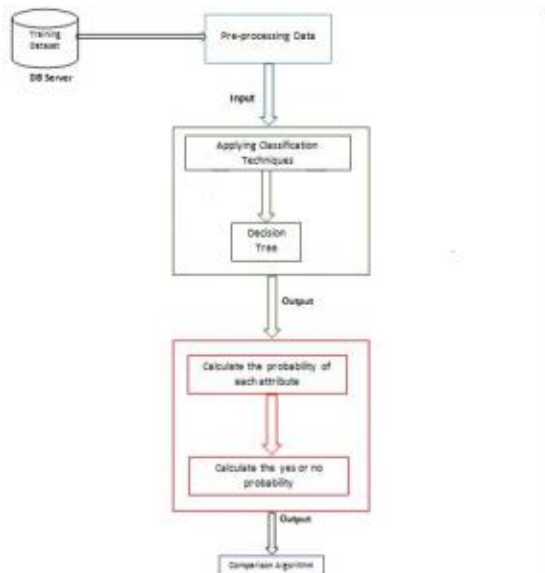


Fig1: Architecture of the proposed model The Fig 1 depicts the architecture of the proposed model used in the prediction of heart disease. It consists of 3 steps,

A.Data base server

The data set was collected from UCI laboratory. It contains 13 attributes which include sex, serum cholesterol level, resting ECG etc. This data set was stored using MSQl database. It acts as a DB server (backend server). Connection was established between DB server & client (front end). Queries are posed on the server using this DB & each individual data is retrieved.

B. Data pre-processing

In this step data is taken out from UCI repository in a recognized format. Missing fields are evacuated in this process & thus the data is transformed. Mean is entered in place of missing attributes.

C. Classification

In data mining mainly two methods are used, supervised & unsupervised learning. Supervised learning uses a training set in order to learn model parameters. But no such training set like k – means clustering is utilised in unsupervised learning. Data mining has got two most frequent modelling goals – classification & prediction. Classification model classifies discrete, unordered values or data. In this prediction process, the classification techniques utilized are, ¾ Decision tree.

Decision tree

In this type of classification the knowledge is represented in a tree diagram. Schematic representation will be in the form of tree to depict the decisions. Rules are framed through these decisions to classify the data. The data classification is in the form of root nodes to begin with, ending with the terminal node of the attributes. Nodes are having attribute names, positive values are the edges & different classes represent the leaves of the tree. This paper has used the popular ID3 algorithm. Fig 2 depicts the tree structure of prediction of heart disease which consists of root node & child node attributes of heart disease. Root node contains the maximum entropy value like old peak & the child node values like number of major vessels & chest real.

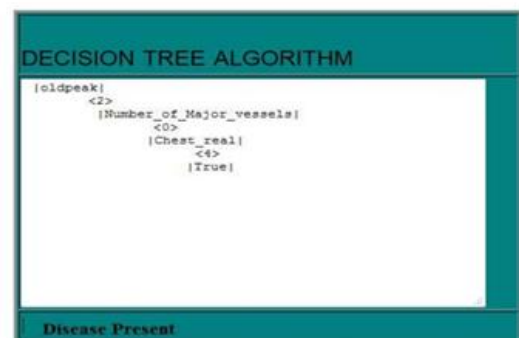


Fig 2: Tree structure for prediction of disease attributes

id	age	sex	chestpain	restingbp	cholesterol	fastingbs	maxslope	oldpeak	majorvessels	ca	result
1	70	1	4	200	202	0	2	0	0	0	0
2	61	1	3	170	160	0	1	0	0	0	0
3	69	1	4	200	202	0	2	0	0	0	0
4	71	1	4	190	190	0	2	0	0	0	0
5	71	1	3	190	190	0	2	0	0	0	0
6	71	1	4	190	190	0	2	0	0	0	0
7	71	1	4	190	190	0	2	0	0	0	0
8	71	1	4	190	190	0	2	0	0	0	0
9	71	1	4	190	190	0	2	0	0	0	0
10	71	1	4	190	190	0	2	0	0	0	0
11	71	1	4	190	190	0	2	0	0	0	0
12	71	1	4	190	190	0	2	0	0	0	0
13	71	1	4	190	190	0	2	0	0	0	0
14	71	1	4	190	190	0	2	0	0	0	0
15	71	1	4	190	190	0	2	0	0	0	0
16	71	1	4	190	190	0	2	0	0	0	0
17	71	1	4	190	190	0	2	0	0	0	0
18	71	1	4	190	190	0	2	0	0	0	0
19	71	1	4	190	190	0	2	0	0	0	0
20	71	1	4	190	190	0	2	0	0	0	0

Fig 4: Calculation of yes/no probability us

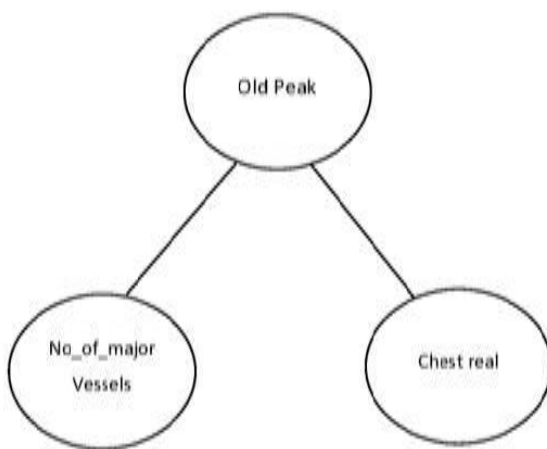


Fig 2: Tree structure for prediction of disease attributes Fig 3 depicts the performance of the decision tree algorithm and it predicts the presence of heart disease based on the 3 attributes like old peak, Number of major vessels and chest real.

IV. RESULT

Data Set:

id	age	sex	chestpain	restingbp	cholesterol	fastingbs	maxslope	oldpeak	majorvessels	ca	result
1	70	1	4	200	202	0	2	0	0	0	0
2	61	1	3	170	160	0	1	0	0	0	0
3	69	1	4	200	202	0	2	0	0	0	0
4	71	1	4	190	190	0	2	0	0	0	0
5	71	1	3	190	190	0	2	0	0	0	0
6	71	1	4	190	190	0	2	0	0	0	0
7	71	1	4	190	190	0	2	0	0	0	0
8	71	1	4	190	190	0	2	0	0	0	0
9	71	1	4	190	190	0	2	0	0	0	0
10	71	1	4	190	190	0	2	0	0	0	0
11	71	1	4	190	190	0	2	0	0	0	0
12	71	1	4	190	190	0	2	0	0	0	0
13	71	1	4	190	190	0	2	0	0	0	0
14	71	1	4	190	190	0	2	0	0	0	0
15	71	1	4	190	190	0	2	0	0	0	0
16	71	1	4	190	190	0	2	0	0	0	0
17	71	1	4	190	190	0	2	0	0	0	0
18	71	1	4	190	190	0	2	0	0	0	0
19	71	1	4	190	190	0	2	0	0	0	0
20	71	1	4	190	190	0	2	0	0	0	0

Fig 1 Data set of the Heart Disease Prediction System

The data set of the heart attack, which contain the 300records.

Architecture of the proposed model depicts in the prediction of heart disease. It consists of 3 steps, In this system, training data set is tested in predicting the heart disease. 13 attributes were taken for the prediction system of the disease. The algorithms selected for is – Decision tree. The most accurate & effective system was tested among the two. Study found that Decision tree was more precise in its prediction of heart disease. Various test cases included in the paper prove the above said fact in this regard.

V CONCLUSION

This paper was planned with an objective to make the clinical decision making in Opredicting the heart disease easier and quicker by using the Decision tree algorithm. Even though both of them were good enough in predicting the Heart disease using various parameters, Decision tree was found to be the best. It gave the most accurate result whether the patient had the possibility of the heart disease. This system can also be used in future projects to detect the specific type of heart disease in particular. Thereby the diagnosis & management of Heart disease can be made simpler as much as possible.

REFERENCES

- [1] M. Raihan¹, Saikat Mondal², Arun More³, Md. Omar Faruque Sagor⁴, Gopal Sikder⁵, Mahbub Arab Majumder⁵, Mohammad Abdullah Al Manjur⁵ and Kushal Ghosh⁵, "Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design", International Conference on Computer and Information Technology, December 18-20, 2016.
- [2] Vanishree K.,Jyothi Singaraju,"Decision Support System for congenital heart disease diagnosis based on signs and symptoms using neutral network"
- [3] S.Kharya, D. Dubey, and S. Soni - Predictive Machine Learning Techniques for Breast Cancer Detection, (IICSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (6), 2013, 1023-1028.
- [4] C.D. Katsis, I. Gkogkou, C.A. Papadopoulos, Y.Goletsis, P. V. Boufounou, G. Stylios "Using artificial immune recognition systems in order to detect early breast cancer." International Journal of Intelligent Systems and Applications 5.2 (2013): 34.