iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Prediction of Heart Disease Based on Decision Trees

Lakshmishree J[1], K Paramesha[2]
[1]PG Scholar, [2]Associate Professor, Department of CSE
Vidyavardhaka College of Engineering, Mysuru, India

Abstract: ***The prediction of heart disease is most complicated task in the field of medical sciences which cannot be observed with a naked eye and comes instantly anywhere, anytime. So there arises a need to develop a decision support system for detecting heart disease. A heart disease prediction model using data mining technique, called decision tree algorithm which helps medical practitioners in detecting the disease based on the patient's clinical data. In this project, we propose an efficient decision tree algorithm technique for heart disease prediction. To achieve correct and cost effective treatment computer-based systems can be developed to make good decision. . Data mining is a powerful new technology for the extraction of hidden predictive and actionable information from large databases, the main objective of this project is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. It can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions.***
Keywords: data mining, data mining techniques, decision tree algorithm.

I. INTRODUCTION

Heart is the most important organ in living beings; if it does not function properly it will affect other parts of body. As per the estimation of World Health Organization (WHO) 12 million deaths occur due to heart diseases, and nearly 80% of death in the world will be because of heart disease .Now a days at many places clinical test results are often based on doctor's intuition and experiences rather than information available in many large databases. Hence this process leads to unintentional biases, errors and a huge medical cost which affects the quality of service provided to patients.

Today many hospitals have installed databases systems to manage their clinical data or patient data. These information systems typically generate large amounts of data which can be in any format like numbers, text, charts and images but unfortunately, this database that contains rich information is rarely used for clinical decision making. There is much information stored in repositories that can be used effectively to support decision making in healthcare. Data mining techniques is widely used in medical field for extracting data from database. In data mining decision tree is a method which is used extensively. Decision trees are non-parametric supervised learning method used for classification. The main aim is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The structure of the decision tree is in the form of tree and leaf nodes. Decision trees are most commonly used in operations research, mainly in decision analysis. Advantages are that they are easy to understand and interpret. They are robust, performed well with large datasets, able to handle both numerical and categorical data. By providing efficient treatments, it can help to reduce costs of treatment. Using data mining techniques it takes less time for the prediction of the disease with more accuracy.

A. The Risk Factors of Heart Disease are:

1) *Heredity:* Most people know that the heat disease can run in families. That if anybody has a family history of heart disease, he/she may be at greater risk for heart diseases.
2) *Smoking:* smoking is major cause of heart attack. Nearly 40% of all people who die from smoking tobacco do so due of heart and blood vessel diseases. A smoker's risk of heart attack reduces rapidly after only one year of not smoking. It damages the lining of arteries and builds up a fatty material called atheroma which narrows the arteries causing heart attacks.
3) *Cholesterol:* Cholesterol is a soft, waxy substance found among the lipids in the blood stream and in all the body's cells, abnormal levels of lipids (fats) in the blood are risk factor of heart diseases. High level of triglyceride (most common type of fat

in body) combined with high levels of LDL (low density lipoprotein) cholesterol speed up atherosclerosis increasing the risk of heart diseases. Increase in the fatty deposits (high cholesterol) does not allow sufficient blood to flow in through the arteries thus causing heart attacks.

4) *High Blood Pressure:* when the heart pumps blood, the force of the blood pushes against the walls of the arteries causing pressure. If the pressure rises and stays high over the time it is called high blood pressure or hypertension. This can harm the body in a way by increasing the risk of getting heart disease.
5) *Obesity:* The term obesity is used to describe the health condition of anyone significantly above his or her ideal healthy weight. Being obese puts anybody at a higher risk for health problem such as heart disease, stroke, high blood pressure, diabetes and more. Obese people are more likely to have high blood pressure, high cholesterol level and diabetes (increase in blood sugar level) which increases the risk of heart strokes in human body.
6) *Lack of Physical Exercise:* lack of exercise is a risk factor for developing various diseases. The increase in the blood pressure and increases cholesterol level in blood vessels increases the risk of heart diseases.

II. LITERATURE SURVEY

In 2004, Carlos Ordonez[1], surveyed set of information which enveloped restorative records of the general population of coronary illness with qualities for the hazard elements, estimations of heart perfusion and limited conduit. The three imperatives were acquainted with decrease the quantity of examples, they are as per the following: 1) the attributes need to show up on one side of the as it was. 2) The rule isolates the characteristics into the uninteresting gatherings. 3)The quantity of qualities from the rule is controlled by the medicinal records of the general population of coronary illness at long last.

Additionally falling the running time according to the trials the limitations of indicating standards have been amazingly diminished the number. The author, Carlos Ordonez, foresaw the nearness or the nonappearance of coronary illness in four particular heart supply routes into the two gatherings of guidelines. Information mining techniques may help the clinicians in the expectation of the survival of patients and in the adaption of the practices subsequently.

In 2004, Franck Le Duff et al [2] build an efficient decision tree for executing medical procedure. Data mining methods may aid the clinicians in the prediction of the survival of patients, the comparison of traditional analysis and data mining analysis which sorted the variables. This provided the importance of data and variables for building a decision tree.

In 2006, Kiyong Noh et al [3], has used classification method for extracting patterns from database. Assessing the heart rate variability from ECG of 670 people and grouped them into normal people and people with heart disease. This method is developed to identify people with heart disease based on the clinical data.

In 2006, Boleslaw Szymanski et al[4], "Using Efficient Supanova Kernel For Heart Disease Diagnosis" used heuristic method for computation of sparse kernel in SUPANOVA. It was applied to a benchmark Boston housing market dataset to detect the heart disease. The noninvasive measurement which was based on magnetic field generated by human heart. Support vector machine was used for obtaining results which was accurate.

In year 2008, Sellappan Palaniappan, et. al. [5] performed a work, “Intelligent Heart Disease Prediction System Using Data Mining Techniques". In this paper he created Intelligent Heart Disease Prediction System (IHDPS) utilizing data mining methods, i.e. Decision Trees, Naïve Bayes and Neural Network. Every strategy has its own particular energy to increase appropriate outcomes. The shrouded examples and connections among them have been utilized to build this framework. The IHDPS is easy to understand, web-based, scalable, versatile, and expandable.

In year 2012, Chaitrali S. Dangare, et. al. [6] performed a work, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”. This paper has examined expectation frameworks for heart disease utilizing more number of attributes. The framework utilizes terms, for example, cholesterol, gender, blood pressure, like credits to anticipate the likelihood of patient getting a coronary illness. In this exploration work included two more properties i.e. obesity and smoking. The data mining classification techniques, namely Decision Trees, Naïve Bayes, and Neural Networks are analyzed on heart disease database.

In 2012,Akhil Jabbar et al [7], proposes proficient arrangement calculation utilizing hereditary approach for coronary illness forecast. The fundamental inspiration for utilizing hereditary calculation in the revelation of abnormal state forecast tenets is that the found standards are exceedingly intelligible, having high predictive precision and of high intriguing quality qualities. This paper goes for investigating diverse information mining methods that has been presented as of late for Heart Disease Prediction framework by various authors.

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

In 2012, R.Bhuvaneshwari and K .Kalaiselvi [8], used Naïve Bayes classifier method, and back propagation neural network. these algorithms which calculates the probability of all objects from past experience. The posterior from prior is calculated by using Bayes rules based on the probability model.

In 2012,Nidhi Bhatla et al[9], used several data mining techniques namely Decision tree, Naïve Bayes and Neural Network. This helped in developing a prototype Intelligent Heart Disease Prediction system by using several attributes. The clinical data was obtained from Cleveland heart disease database for the research purpose. The analysis show that the provides results accurately.

In 2012 Shadab Adam Pattekari and Asma Parveen [10], used the naïve Bayesian classification technique for developing a decision support in heart disease prediction system .This system will identify the hidden knowledge from the database and solves complex queries. The interpretation and access to information is very accurate.

III. PROPOSED APPROACH

Proposed system is the medical sector application. This system helps to predict of heart disease, based on manually inserted inputs. Inputs are like patient's age, gender, chest pain, the resting blood pressure in mmHg, serum cholesterol in mg/d, hereditary, fasting blood pressure in mg/dl, thal and smoking. The database contains all the information of the user i.e. the user details and the admin uploaded files, these will be compared to produce an output. Output is consisting either patient has heart disease or not.

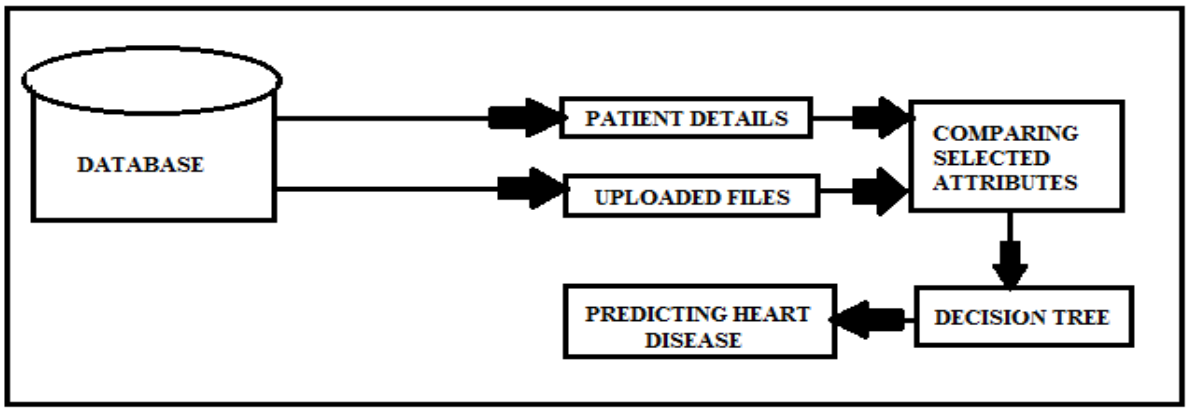


Fig1: Proposed system diagram

From the fig 1, the JSP java server pages will connect the user and admin to the database. Database consists of user details and admin uploaded files. Database is a storage device which has the capacity to hold huge data. In this proposed system it consists of patient details and uploaded files, and the attribute is selected. Then for selected the decision tree algorithm tree is applied to predict the heart disease.

A. Decision Tree

Decision tree is a simpler classifier which is easy to implement and which does not require any domain knowledge. The main advantage is decision tree method can be applied to huge data which is to interpret. Decision tree is a tree like structure which consists of arcs, nodes, and branches. The arcs connect from one node to another. The branch has attributes, an internal node has a test on which the attribute is used for, and leaf node consists of the classes which are predicted from decision tree.to make an appropriate decision the traversing starts from root node to leaf node.

B. Iterative Dichotomiser 3

Iterative Dichotomiser 3 is also known as ID3 algorithm which was introduced by J R Quinlan .this is mainly used in the field of data mining, one of the most important algorithms of decision tree algorithms. The main aspect of ID3 algorithm is to select the appropriate attribute to test at each node in a tree which is an iterative process. It uses top down approach which traverses from top node to leaf node at the bottom which completes the decision tree. This is the basic algorithm used for classification in data mining. In this project, we have considered clinical data which has clinical datasets which are in the form of symbolic and numeric.

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

The selected clinical data is obtained from Cleveland Heart Disease Data Set http://archive.ics.uci.edu/ml/datasets/Heart+Disease for experimental purpose.

Table 1 : Selected Cleveland Heart Disease Data Set

ATTRIBUTE	TYPE	DESCRIPTION
Age	Continuous	Age in years
Gender	Discrete	1 = male 0 = female
Chest Pain	Discrete	Chest pain type: 1 = less pain 2 = more pain
Resting Blood Pressure(Trbps)	Continuous	Resting blood pressure (in mm Hg)
Cholesterol	Continuous	Serum cholesterol in mg/dl
Hereditary	Continuous	1 = yes 0 = no
Fasting Blood Sugar(Fbs)	Discrete	Fasting blood sugar > 120 mg/dl: 1 = true 0 = false
Smoking	Continuous	1 = yes 0 = no
Thal	Discrete	3 = normal 6= fixed defect 7= reversible defect
Result	Discrete	0 =patient with no heart disease 1= patient with heart disease

Iterative Dichotomiser 3 algorithm main aim is to select the best attribute in dataset obtained. Here we use entropy which measures the purity of collection in a dataset. “S” is a collection containing both negative and positive examples, and the attribute can take “n” different values “p_i” is the probability that the sample example belongs to class i,base 2 is measure of the information available. Another term called Information gain is used to select the best attribute at every step this helps building decision trees.

$$\text{Entropy}(S) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Table 2 : Entropy calculated for Heart Disease Data Set

Sl No	Attributes	Description	Number Of Patients	Has Disease	No Disease	Entropy
1	Age	<=50 >50	32 68	8 48	24 20	0.811 0.873
2	Gender	Male Female	51 49	39 12	12 37	0.781 0.796
3	Chest Pain	Less Pain More Pain	33 67	10 50	23 17	0.884 0.817
4	Trbps	>120 <=120	71 29	21 12	50 17	0.876 0.978
5	Cholesterol	High Normal	57 43	24 23	33 20	0.981 0.996
6	Heredity	Yes No	57 43	23 14	34 29	0.972 0.910
7	Fbs	Yes No	33 67	27 15	6 52	0.684 0.752
8	Smoking	Yes No	22 78	17 21	5 57	0.773 0.840
9	Thal	Normal High	52 48	12 32	40 16	0.77 0.917

This dataset in table: 2 consists of 100 patient details of which 51 male and 49 female patients the chances of getting heart disease totally is 26 patients and 74 are predicted as no heart disease.

IV. CONCLUSION

In medical field datamining technique has been widely used in clinical decision support system .The user details and uploaded files will be stored in the database The users details will be compared with the previously stored dataset using decision tree algorithm. The clinical data considered which includes attributes such as age, gender, cholesterols level, chest pain, the resting blood pressure, hereditary, fasting blood pressure, thal and smoking. The decision tree algorithm is the simple and an efficient algorithm to select the attributes. Since clinical data will be very large, it should be mined very fast so that efficient result is obtained. This application is very simple and user friendly which helps the users to get to know about the occurrence of disease. In this way the prediction of heart disease can be a useful approach to improve the quality and accuracy of healthcare service while lowering the healthcare cost and diagnosis time.

REFERENCES

[1] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.
[2] Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 1256-9, 2004.
[3] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and KeunHoRyu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006.
[4] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel, Lijuan Zhu, "Using Efficient Supanova Kernel for Heart Disease

Volume 5 Issue V, May 2017
ISSN: 2321-9653

Diagnosis", proc. ANNIE 06,intelligent engineering systems through artificial neural networks, vol. 16, pp:305-310, 2006.

[5] Sellappan Palaniappan, Rafiah Awang, ―Intelligent Heart Disease Prediction System Using Data Mining Techniquesl; 978-1-4244-1968-5/08/$25.00©2008 IEEE.

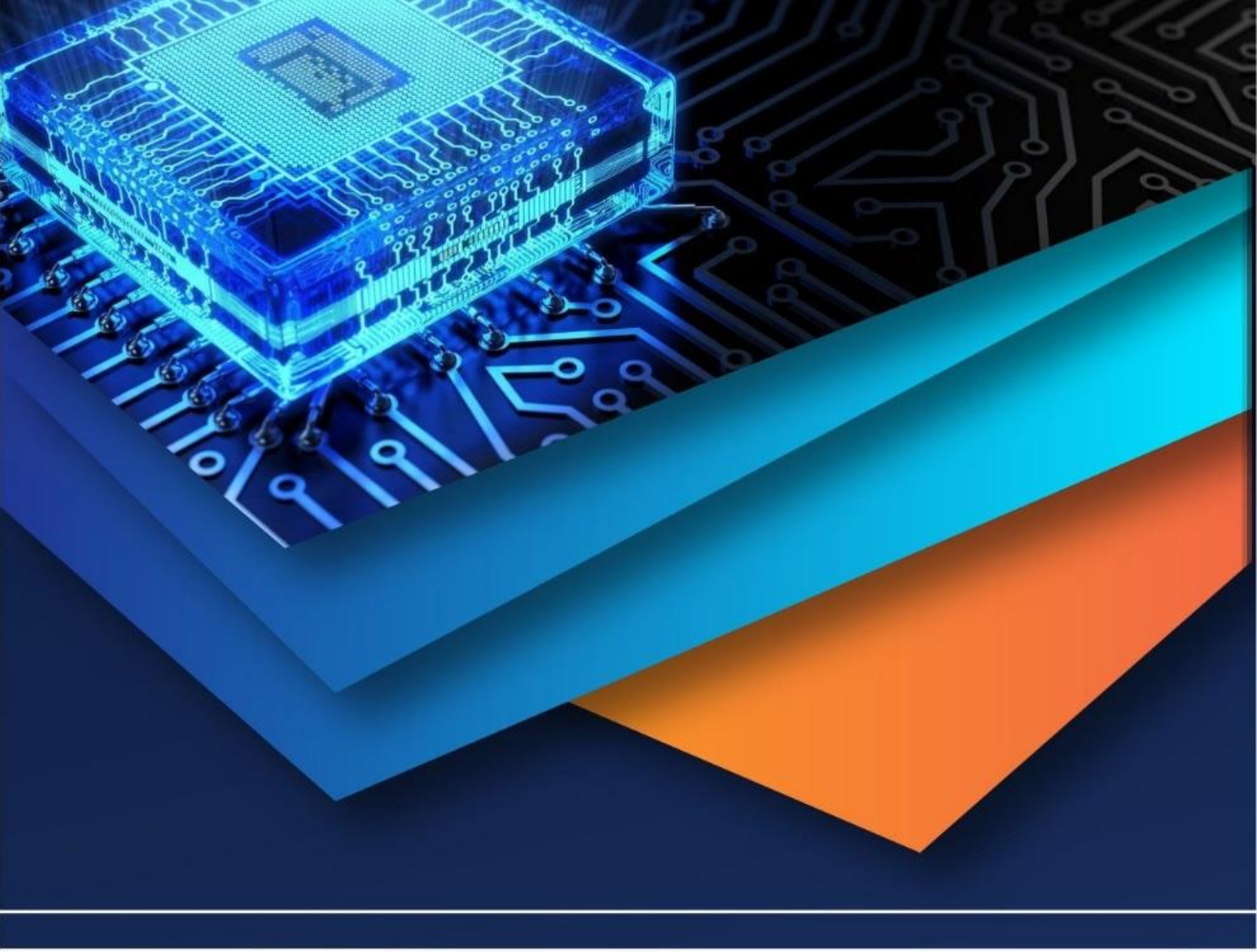
[6] Chaitrali S. Dangare et. al., "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", (IJCA) (0975 – 8887), Vol. 47, No. 10, June 2012, page no. 44-48.

[7] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, Heart Disease Prediction System using Associative Classification and Genetic Algorithm,International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012.

[8] R.Bhuvaneshwari and K .Kalaiselvi,"Naïve Bayesian Classification approach In Healthcare Applications", International Journal of Computer Science and Telecommunication ",vol 3,no 1,pp.106-112 , 2012.

[9] Nidhi Bhatla, Kiran Jyoti, " An Analysis of Heart Disease Prediction using Different Data Mining Techniques" International Journal of Engineering and Technology Vol.1 issue 8 2012.

[10] Shadab Adam Pattekari and Asma Parveen ,"Prediction System for Heart Disease Using Naïve Bayes", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624,Vol 3,Issue 3,2012,pp290-294.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY