

Estimation of Cardiac Disease Using Data Mining

Ms. V.Geetha¹, Dr.S.Rajalakshmi², T.V. Bhaskar Sastry³, Ch. Sasi Kumar⁴

¹Assistant Professor, DEPT OF CSE

²Professor, DEPT OF CSE,

^{3,4} DEPT OF CSE

^{1,2,3,4}Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Deemed to be University Kanchipuram, Tamil Nadu, India

Abstract- *The main objective of this research is to develop an Intelligent System using data mining modeling techniques. It is implemented as web based application in this user answers the predefined questions. It retrieves hidden data from stored database and compares the user values with trained data set. It can answer complex queries for diagnosing cardiac disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs. Data Mining is non trivial extraction of implicit data, previously not known, and imaginably useful information from data. Using data mining ,it can evaluate patterns which can be used in future to take intelligent decisions and that can present the knowledge extracted in better way. Data mining refers to using a variety of techniques to identify information or decision making knowledge in the database and extracting these in a way that they can put to use in areas such as decision making, predictions, for valuable forecasting and computation. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not “mined” to discover hidden information, to take decisions effectively, to discover the relations that connect parameters in a database are the subject of data mining. Using advanced data mining techniques to excavate valuable information has been considered as an activist approach to improve the quality and accuracy of healthcare service while lowering the healthcare cost and diagnosis time. Using this technique presence of heart disease can be predicted accurately. Using more input attributes such as controllable and uncontrollable risk factors, more accurate results could be achieved.*

Keywords- Data modeling Technique, Decision Making, cardiac Disease, Prediction Model

I. INTRODUCTION

The rising high-performance computing has benefited numerous disciplines in finding realistic solutions to their problems. Our health care is no exception to this. Data mining tools have been developed for effective study of medical information, in order to assist clinicians in making better diagnosis for treatment purposes.

In heart disease research, data mining technique have performed a significant role. From the different interpretation between the healthy persons and the heart diseased persons in the already existing medical data is an appreciable and great approach in the study of heart related disease classification to find the conceal medical information. Heart disease categorization contributes the dangerous basis for the treatment of patients. To foresee the prominence of heart disease grounded on the expression of the medical data, the two key approaches employed are Statistics and machine learning.

Data Mining (DM) is the important stage of information discovery in databases (KDD), which is an extraction of implicit, unique, and potentially useful information from data. The difference between mining of data and discovering knowledge is that the latter is the utilization of different intelligent algorithms to excerpt patterns from the data whereas information discovery is the complete process that is involved in discovering knowledge from data. The ultimate objective is to abstract high-level information from low-level data.

II. RELATED WORK

Data mining is one of the essential areas of research that is more popular in health organization. Data mining plays an effective role for uncovering new trends in healthcare organization which is helpful for all the parties associated with this field. Heart disease is the leading cause of death in the world over the past 10 years. Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the irregular health condition that directly affects the heart and all its parts. The healthcare industry gathers enormous amount of heart disease data which are not “mined” to discover hidden information for effective decision making. Data mining techniques are useful for analyzing the data from many different dimensions and for identifying relationships. This paper explores the utility of various decision tree algorithms in classify and predict the disease. Every human being can not be equally skillful and so as doctors. All doctors cannot be equally skilled in every sub specialty and at many places we don't have skilled and

specialist doctors available easily. An automated system in medical diagnosis would enhance medical care and it can also reduce costs. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the user's requirement. The performance of the system is evaluated in terms of classification accuracy and the results shows that the system has great potential in predicting the heart disease risk level more accurately.

III. PROPOSED METHODOLOGY

Data mining is defined as “a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database” or as “a process of selection, exploration and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of database”.

A. Bayesian Classifiers

Using Bayesian classifiers, the system can discover the concealed knowledge associated with diseases from historical records of the patients having heart disease. Bayesian classifiers predicts the class membership probabilities, in a way that the probability of a given sample belongs to a particular class statistically. Bayesian classifier is based on Bayes' theorem. We can use Bayes theorem to determine the probability that a proposed diagnosis is correct, given the observation. A simple probabilistic, the naive Bayes classifier is used for classification based on which is based on Bayes' theorem. When the dimension of the inputs is high and more efficient result is expected, the chief Naïve Bayes Classifier technique⁵⁻⁷ is applicable. Naïve Bayes model identifies the physical characteristics and features of patients suffering from heart disease. For each input it gives the possibility of attribute for the expectable state. The Figure 1 shows the implementation of Naive Bayes algorithm on patient data.

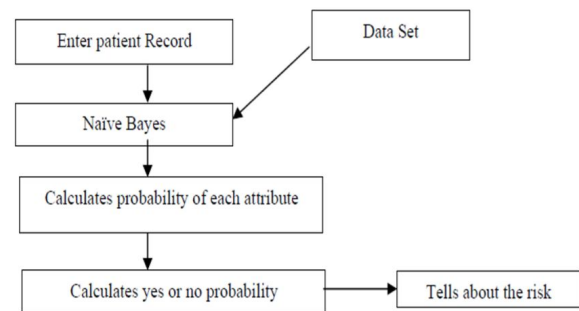


Figure 1. Implementation of Naïve Bayes algorithm on patient data..

B. Support Vector Machine

Support Vector Machines proved themselves to be very effective in a variety of pattern classification tasks and thus received a great deal of attention recently. Support vector machine is a supervised machine learning technique. The SVM algorithm predicts the occurrence of heart disease by plotting the disease predicting attributes in multidimensional hyperplane and classifies the classes optimally by creating the margin between two data clusters. This algorithm attains high accuracy by the usage of nonlinear functions called kernels

IV. MODULE DESCRIPTION

Medical Misdiagnoses are a serious risk to our healthcare profession. If they continue, then people will fear going to the hospital for treatment. We can put an end to medical misdiagnosis by informing the public and filing claims and suits against the medical practitioners at fault. The National Patient Safety Foundation cites that 42% of medical patients feel they have had experienced a medical error or missed diagnosis. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Thus we proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. It is implemented as web based questionnaire application. Based on the user answers, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database.

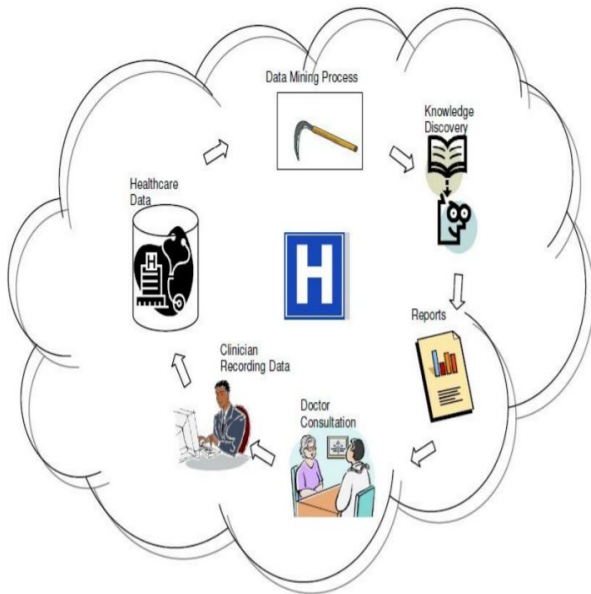


Fig: 2 illustrating the working process of the system

The main role of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network.

So its providing effective treatments, it also helps to reduce treatment costs. The figure 2 explains the working process of the proposed system effectively.

Modules for Heart Attack Prediction Are

- User module
- Admin module
- Classification module

I. User Module

- With the datasets already acquired from repository, We can now proceed with collecting the details of the patients.
- Data like the patients name, Blood group, Address, Phone number, sex, weight, Height are collected from the patients themselves to help with the creation of an individual user account.
- Already registered user can directly start accessing the system with the help of the user id and password provided.

II. Admin Module

- Admin views all patient details and their Medical History.

- Admin with the help of the medical records, punch in the value of the attributes which is fed for classification purposes.
- The result is a screen that displays the chance of a person contracting heart attack
- The system then produces a list of suggestions for the person with the given attributes

III. Classification Module

- Naïve bayes and decision making tree which is apt for classifying medical data. The data given by the doctor and the patient has been Fed into then algorithm. It is now the job of the classification algorithm.
- Data classification is process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible.

V. RESULTS AND DISCUSSIONS

The heart is important organ of human body part. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it. To analyze the results the user has to login in to the website, if user is not register then register first and the login into the system. After that user has to enter the values to attributes displayed on the screen. After entering the values, user can analyze their health condition for each and every attribute results , that refers to the heart disease chances are high or low. The figure 3: shows the result models.

STRUCTURED DATA RISK		
RISK BASED ON (GENDER, AGE, NAME)		
ID	6	--
AGE	41.0	LOW LEVEL
GENDER	0.0	LOW LEVEL
NAME	XXX	--

UNSTRUCTURED DATA RISK		
IN RISK BASED ON (CP, TRESTPDS, CHOL, FBS, RESTECG, THALACH, EXANG, OLDPEAK, SLOPE, CA, THRO, STENO)		
CP	2.0	HIGH LEVEL
TRESTPDS	150.0	HIGH LEVEL
CHOL	204.0	LOW LEVEL
FBS	0.0	LOW LEVEL
RESTECG	2.0	HIGH LEVEL
THALACH	172.0	NORMAL LEVEL
EXANG	0.0	LOW LEVEL
OLDPEAK	1.4	HIGH LEVEL
SLOPE	1.0	HIGH LEVEL

Fig 3: Model of predicting the results

Admin of the website also can check the all users prediction results and they may send alerts to the patients . Figure 4 shows the overall results of all users.

ID	AGE	GENDER	CP	TRESTBPS	CHOL	FBS	RESTECG	THALACH	EXANG	OLDPEAK	SLOPE	CA	THAL	NUM	NAME	RISK PREDICTION
1	63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0	0.0	XXX	(HIGH RISK)
2	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2.0	XXX	(HIGH RISK)
3	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1.0	XXX	(HIGH RISK)
4	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0.0	XXX	(HIGH RISK)
6	41.0	0.0	2.0	130.0	304.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0.0	XXX	(HIGH RISK)
7	56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	0.0	XXX	(HIGH RISK)
8	62.0	0.0	4.0	140.0	268.0	0.0	2.0	160.0	0.0	3.6	3.0	2.0	3.0	3.0	XXX	(HIGH RISK)
9	57.0	0.0	4.0	120.0	354.0	0.0	0.0	163.0	1.0	0.6	1.0	0.0	3.0	0.0	XXX	(LOW RISK)
10	63.0	1.0	4.0	130.0	254.0	0.0	2.0	147.0	0.0	1.4	2.0	1.0	7.0	2.0	XXX	(HIGH RISK)

FIG 4:. Shows the overall analysis of users

VI. CONCLUSION

Application of Data mining in analyzing the medical data is a good method for considering the existing relationships between variables. From our proposed approach we have shown that mining helps to retrieve useful correlation even from attributes which are not direct indicators of the class we are trying to predict. In our work we have tried to predict the chances of getting a heart disease using attributes from diabetic’s diagnosis. This can be extended to predict other type of ailments which arise from diabetes, such as visual impairment in future. Further, the data analysis results can be used for further research in enhancing the accuracy of the prediction system in future.

REFERENCES

[1] Shadab Adam Pattekari and Asma Parveen:”PREDICTION SYSTEM FOR HEART DISEASE USING NAÏVE BAYES” on International Journal of Advance computer and Mathematical Science ,March,2012.

[2] Dhamashree S. Medhekar, Mayur P. Bote, Shruthi D. Deshmukh: ”HEART DISEASE PREDICTION SYSTEM USING NAÏVE BAYES” on International Journal of Enhanced Research in science Technology and Engineering, March,2013.

[3] Ms.Rupalin R.Patil:”HEART DISEASE PREDICTION SYSTEM USING NAÏVE BAYES AND JELINKE-MERCER SMOOTHING” on International Journal of Advanced Research in Computer and Communication Engineering,May 2014.

[4] Vincy Cherian, Bindhu M.S:”HEART DISEASE PREDICTION USING NAÏVE BAYES ALGORITHM AND LAPLACE SMOOTHING TECHNIQUE” on

International Journal of Computer Science Trends and Technology, March 2017.

[5] N. Adithya, P. Puspa Latha, M. Rama Chandra:” PERFORMANCE ANALYSIS OF CLASSIFICATION DATA MINING TECHNIQUE OVER HEART DIESASE DATA BASE” on International Journal of Engineering Science and Advanced Technology,May 2013

[6] Devendra Ratnaparkhi, Thshar Mahajan, Visual Jadhay:”HEART DISEASE PREDICTION SYSTEM USING DATA MINING TECHNIQUE” on International Research Journal of Engineering and Technology, November 2015

[7] Abhishek Taneja:”Heart Disease Prediction System Using Data Mining Techniques” on ORIENTAL JOURNAL OF COMPUTER SCIENCE &TECHNOLOGY, December 2013

[8] S. Kiruthika Devi*, S. Krishnapriya and DristiponaKalita:”Prediction of Heart Disease using Data Mining Techniques” on Indian Journal of Science and Technology, October 2016

[9] Amit Ganatra:”A SURVEY ON NAÏVE BAYES BASED PREDICTIONOF HEART DISEASE USING RISK FACTOR” on Charota University of Science and Technology, January 2016

[10]J.Thomas, R.TheresaPrincy:”HUMAN HEART DISEASE PREDICTION SYSTEM USING DATA MINING TECHNIQUES: on International Conference on March ,2016

[11]JagdeepSingh, AmitKamra,Harbhag Singh:”PREDICTION OF HEART DISEASES USING ASSOCIATIVE CLASSIFICATION” on 5th International Conference on,October 2016

[12]Keerthana. T.K:”HEART DISEASE PREDICTION SYSTEM USING DATA MINIG METHOD” on International Journal of Engineering Trends and Technology, May 2017

[13]M.AkhilJabbar, Dr. PritiChandra , Dr.B.L.Deekshatulu:” HEART DISEASE PREDICTION SYSTEM USING ASSOCIATIVE CLASSIFICATION AND GENETIC ALGORITHM” on International Conference on Emerging Trends in Electrical,Electronics and Communication,2012