ISSN:0975-9646

Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve Bayes Algorithm

D Ratnam[1], P HimaBindu[2,] V.Mallik Sai[3], S.P.Rama Devi[4] , P.Raghavendra Rao[5]

[1]*Assistant Professor, Dept. of IT, P V P Siddhartha Inst. of Tech.*
[2, 3, 4, 5.]*IV/IV B.Tech, Dept. of IT, PVP Siddhartha Inst. of Tech*

***Abstract*—Our healthcare sector daily collects a huge data including clinical examination, vital parameters, investigation reports, treatment follow-up and drug decisions etc. But very unfortunately it is not analyzed and mined in an appropriate way. The Health care industry collects the huge amounts of health care data which unfortunately are not "mined" to discover hidden information for effective decision making for health care practitioners. Data mining refers to using a variety of techniques to identify suggest of information or decision making knowledge in database and extracting these in a way that they can put to use in areas such as decision support , Clustering ,Classification and Prediction. This paper has developed a Computer-Based Clinical Decision Support System for Prediction of Heart Diseases (CCDSS) using Naïve Bayes data mining algorithm. CCDSS can answer complex "what if" queries which traditional decision support systems cannot. Using medical profiles such as age, sex, spO_2,chest pain type, heart rate, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. CCDSS is Web-based, user-friendly, scalable, reliable and expandable. It is implemented on the PHPplatform.**

***Keywords*—Computer-Based Clinical Decision Support System(CCDSS), Heart disease, Data mining, Naïve Bayes.**

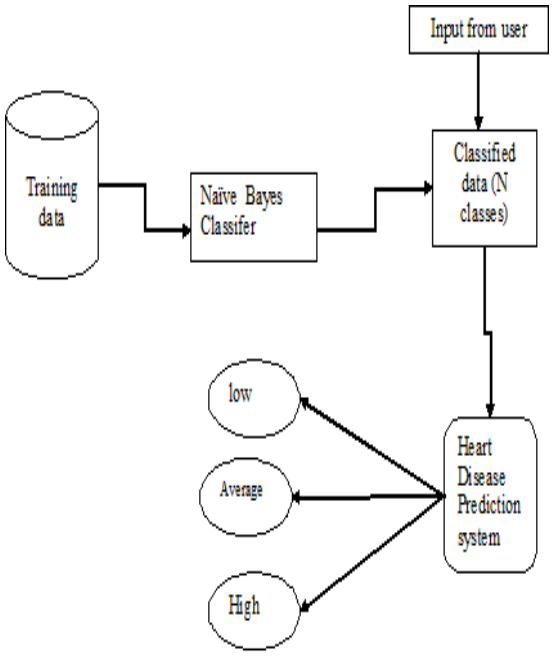
I. INTRODUCTION

In this fast moving world people want to live a very luxurious life so they work like a machine in order to earn lot of money and live a comfortable life therefore in this race they forget to take care of themselves, because of this there food habits change their entire lifestyle change, in this type of lifestyle they are most tensed they have blood pressure, sugar at very young age [10] and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of food, if they are sick they go for their own medication, as a result of all these small negligence it leads to a major threat that is heart disease[14]. It is a world known fact that heart is most essential vital organ in the human body; if that organ gets affected then it also affects the other vital parts of the body. Therefore it is very important for a people to go for heart disease diagnosis.

Now-a–days, in the world Heart Disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to heart diseases. In 2008, 17.3 million people died due to Heart Disease. The World Health Statistics 2012 reports enlighten the fact that one in three adults world-wide has raised blood pressure–a condition that causes around half of all deaths from stroke and heart disease. WHO estimated by 2030, almost 23.6 million will die due to Heart disease [21].Heart disease is also known as (CVD) cardiovascular disease, encloses a number of conditions that influence the heart- not just heart attacks. Heart diseases also include functional problems of heart such as infections in heart muscles like myocarditis (inflammatory heart diseases), heart-valve abnormalities or irregular heart rhythms etc these reasons can lead to heart failure.

II. DATA MINING

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenues, cuts costs, or both. Extracting the useful knowledge and providing scientific decision making for the diagnosis and the treatment of disease from the database increasingly becomes necessary. Data mining in medicine can deal with this problem. It can also improve the management level of hospital information [15] and promote the development of telemedicine and community medicine. Because the medical information is characteristic of redundancy, multi-attribution, incompletion and closely related with time, medical data mining differs from one another.



Fige 1. Using medical data for effective decisive making

Data mining uses two strategies: supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters where as in unsupervised learning no training data set is used [3].

Each data mining technique serves a different purpose depending on the modeling objective. The two most common modeling objectives are classification and prediction. Classification models predict categorical labels (discrete, unordered) while prediction models predict continuous-valued functions. Decision Trees, Bayes and Naïve Bayes, Neural Networks use classification algorithms while Regression, Association rules and Clustering use prediction algorithms.

Naïve or Bayes rule is the basis of many machine learning and data mining methods. This rule (algorithm) is used to create models with predictive capabilities.

III. Problem Statement

Prediction should be done to reduce the risk of Heart-Disease. Diagnosis is usually based on signs, symptoms and physical examination of patient. All most all the doctors are predicting heart diseases by learning and experience. The Diagnosis of disease is difficult and tedious task in medical field. Predicting heart disease from various factors and symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. Health care industry today generates [19] large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records, medical devices etc. The large amount data is a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Only human intelligence is alone is not enough for proper diagnosis. A number of difficulties will arrive during diagnosis such as, less accurate results, less experience, time dependent performance, knowledge up gradation is difficult listed in fig

3.1 Algorithm

Naïve classifier is a term dealing with simple probabilistic classifier based on applying Bayes theorem with strong independence assumptions. It assumes that the presence or absence of particular feature of class is unrelated to the presene or absence of any feature.

The Naïve Bayes algorithm is based on conditional probabilities. It uses Bayes theorem, a formula that calculates a probability by counting the frequency of values and combination of values in historical data. Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and A represents the prior event, Bayes theorem can be states as follows.

Pro(B given A) = Pro(A and B)/Pro(A)

To calculate the probability of B given A, the algorithm counts the number of cases where A and B occur together and divides it by number of cases where A occurs alone.

An advantage of Naïve Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of variables) necessary for classification. Since independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire. It can be used for both binary and multi class classification problems.

Naïve Bayes

1. Each data sample is represented by an n dimensional feature vector, X=(x1, x2…xn), depicting n measurements made on the sample from n attributes, respectively A1, A2…..An.

2. Suppose that there are m classes, C1, C2…Cm. Given an unknown data sample, X (i.e. having no class label), the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve probability assigns an unknown sample X to the class Ci.

if and only if :

$$P(Ci/X)>P(Cj/X) \text{ for all } 1<=j<=m \text{ and } j!=i$$

Thus we maximize P(Ci/X). The class Ci for which P(Ci/X) is maximized is called the maximum posteriori hypothesis. By Bayes theorem,
P(Ci/X)=(P(X/Ci)P(Ci)/P(X)

3. As P(X) is constant for all classes, only P(X/Ci)P(Ci) needed to be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, i.e P(C1)=P(C2)=...P(Cm), and we should therefore maximize P(X/Ci). Otherwise, we maximize P(X/Ci)P(Ci). Note that the class prior probabilities may be estimated by P(Ci)=Si/s where Si is the number of training samples of class Ci, and s is the total number of training samples.

In a simplified way the naïve bayes equationcan be written as

$$\text{posterior} = \frac{\text{prior} * \text{likelihood}}{\text{evidence}}$$

In this first phase of the project by using the naïve bayes algorithm prediction of probability of effectiveness of occurrence of heart disease by using the different attributes.

Two different classes are considered as C1 and C2 respectively.

C1-male records

C2-female records

and different functional attribute are considered for each class namely

Input Attributes:

TABLE 1

Attribute	Description
Gender	Male/Female Classes are divided based on this attribute
Blood Pressure	Hypertension
Fasting Blood Sugar	Whether the person is diabetic or not
Smoking	Habit of smoking
Alcoholic	Habit of drinking
Pulse rate	Rate of pulse
Resp rate	Rate of respiration per minute
Heart rate	Palpitation rate
spO2	Oxygen saturation rate in the heamoglobin
Resp problem	Whether the person is having the respiration problem are not
Chest pain	Problem of chest pain status
Age	Age of person

IV. RESULTS

Key attributes:

1. PatientID – Patient's identification number

Input attributes

1. Sex (value 1: Male; value 0 : Female)
2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)
3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)
4. Smoking(value 1:Yes, value 0:No)
5. Alcoholic(value 1:yes,value 0:No)
6. spO_2 oxygen saturation in hemoglobin(categoroal values)
7. Pulse Rate(categorical values)
8. Respiration Rate(categorical values)
9. Trest Blood Pressure (mm Hg on admission to the hospital)
10. Heart Rate(variable values)
11 Serum Cholesterol (mg/dl)
12. Age in Year

Prediction of risk factor for occurrence of heart disease

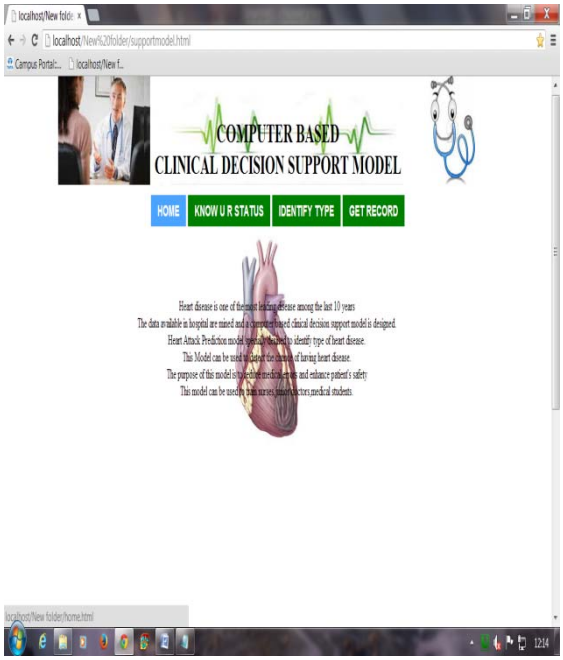


Fig 2. Home page for CCDSS

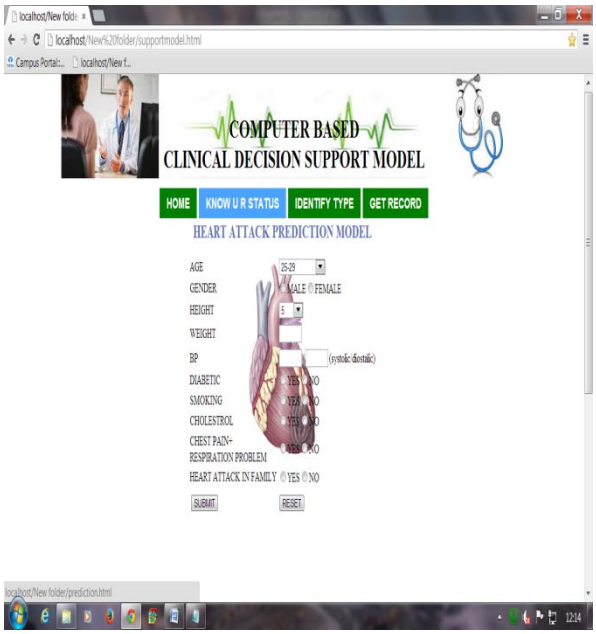


Fig 3. Naive Bayes classification

Fig 4. Result for occurance of heart disease

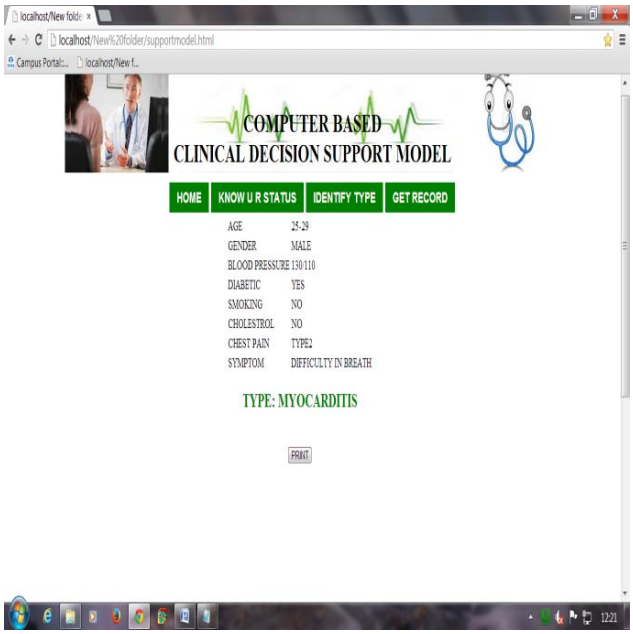


Fig 5. Identifying type of heart disease

V. CONCLUSION

Decision Support in Heart Disease Prediction System is developed using Naive Bayesian Classification technique. The system extracts hidden knowledge from a historical heart disease database. This is the most effective model to predict patients with heart disease. This model could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy. Computer-Based Clinical Decision Support Model can be further enhanced and expanded. For, example it can incorporate other medical attributes besides the above list like considering the thalach (maximum heart rate achieved), serum cholesterol etc and more number of records can be considered for more effective working of the system. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of just categorical data. Another area is to use Text Mining to mine the vast amount of unstructured data available in healthcare databases. Another challenge would be to integrate data mining and text mining.

AUTHORS ACKNOWLEDGMENT

The authors thanks to the Medical Director of KIMS Hospitals, Vijayawada, Krishna District in Andhra Pradesh for providing the opportunity for sharing case sheet information of different patients suffering from heart diseases.

REFERENCES

[1] Sellappan Palaniappan, Rafiah Awang Intelligent Heart Disease Prediction System Using Data Mining Techniques, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008.

[2] MachineLearningDatabases", http://mlearn.ics.uci.edu/databases/heart-disease/, 2004.

[3] Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0: Step by step data mining guide", SPSS, 1-78, 2000.

[4] Charly, K.: "Data Mining for the Enterprise", 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7,295-304, 1998.

[5] Fayyad, U: "Data Mining and Knowledge Discovery in Databases: Implications for scientific databases", Proc. of the 9th Int. Conf. on Scientific and Statistical Database Management, Olympia, Washington, USA, 2-11, 1997.

[6] Giudici, P.: "Applied Data Mining: Statistical Methods for Business and Industry", New York: John Wiley, 2003.

[7] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.

[8] Ho, T. J.: "Data Mining and Data Warehousing", Prentice Hall, 2005.

[9] Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.

[10] *Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Enginering Khaja Banda Nawaz College of Engineering, 'Prediction of Heart Disease using Naïve Bayes'

[11] Mehmed, K.: "Data mining: Concepts, Models, Methods and Algorithms", New Jersey: John Wiley, 2003.

[12] Mohd, H., Mohamed, S. H. S.: "Acceptance Model of Electronic Medical Record", Journal of Advancing Information and Management Studies. 2(1), 75-92, 2005.

[13] Microsoft Developer Network (MSDN). http://msdn2.microsoft.com/en-us/virtuallabs/aa740409.aspx,2007.

[14] Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[15] [15] Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

[16] Tang, Z. H., MacLennan, J.: "Data Mining with SQL Server 2005", Indianapolis: Wiley, 2005.

[17] [17] Thuraisingham, B.: "A Primer for Understanding and Applying Data Mining", IT Professional, 28-31, 2000.

[18] [18] Weiguo, F., Wallace, L., Rich, S., Zhongju, Z.: "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.

[19] Wu, R., Peters, W., Morgan, M.W.: "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management.16(4), 50-55, 2002.

[20] K.Srinivas B.Kavihta Rani Dr. A.Govrdhan Associate Professor, Dept. of CSE Principal andProfessor of CSE 'Applications of Data Mining Techniques in Healthcare and Prediction of HeartAttacks '-(IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255 [21].http://www.worldlifeexpectancy.com/life-expectancy-research