ISSN:0975-9646

An Efficient Method for Predicting Heart Disease Problem Using Fitness Value

Prachi Paliwal , Mahesh Malviya
Computer Science & Engineering Department, JIT Borawan, RGPV University
Khargone, Madhya Pradesh, India

***Abstract*— In the field of Medical Science there are huge amount of data. Data mining techniques are be used to discover hidden pattern form these data. Advance data mining techniques have been developed now days. The efficiency of these techniques are compared through sensitivity, specificity, accuracy and error rate. Some well known Data mining classification techniques, Decision Tree, Artificial neural networks, and Support Vector Machine and Naïve Bayes Classifier. In this paper we introduce a new method based on the fitness value of the attribute to predict the heart disease problem. We use 10 attribute for our proposed method and use simple calculation.**

***Keyword:* - fitness value, heart disease, attribute, data mining, classifier.**

I. INTRODUCTION

In our everyday life there are several example exit where we have to analyze the historical data for example a bank loans officer needs analysis of her data in order to learn which loan applicants are "safe" and which are "risky" for the bank[1,2].

Similarly for a medical researcher it is necessary to analyze breast cancer data in order to predict specific treatments for a patient. These are some examples where the data analysis task required before taking any decision. Classification is a data analysis process, where a classifier is constructed to predict class, for bank loan example prediction class is "yes" or "no" Similarly for a medical researcher prediction class is "treatment A," "treatment B," or "treatment C" for the medical data[3,4].

Classification process can be divided into two parts

(1)Learning: Training data are analyzed by a classification algorithm. Here, the class label attribute is loan decision, and the learned model or classifier is represented in the form of classification rules.

(2) Classification: Test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples

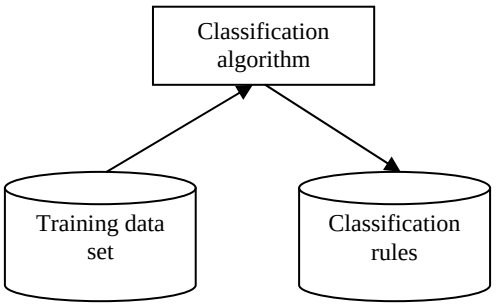

Figure 1.1 Rules construction with training dataset

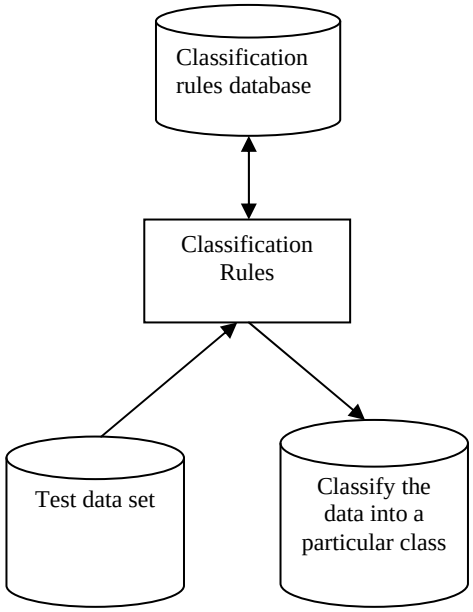

Figure 1.2 Rules uses for classifying unknown tuples

II. LITERATURE REVIEW

In 2010 Sunita Soni and O.P. Vyas proposed "Associative Classifiers for Predictive Analysis in Health Care Data Mining". They used a combined approach that integrates association rule mining and classification rule mining. The integration is done by focusing on mining a special subset of association rules and then classification is being performed using these rules. Given the readability of the associative classifiers, they are especially fit to applications were the model may assist domain experts in their decisions. They also introduce that combining the advanced association rule mining with classifiers gives a new type of associative classifiers with small refinement in the definition of support and confidence that satisfies the validation of downward closure property[4,5].

In 2011 Mrs .G. Subbalakshmi and Mr. K. Ramesh Proposed "Decision Support in Heart Disease Prediction System using Naive Bayes ". They proposed a Decision Support in Heart Disease Prediction System (DSHDPS) using data mining modeling technique, namely, Naïve Bayes. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. They implement the system by using web based questionnaire application. This system helps to train nurses and medical students to diagnose patients with heart disease. 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. DSHDPS can be further enhanced and expanded[6,15].

In 2011 Mai Shouman, Tim Turner, Rob Stocker proposed "Using Decision Tree for Diagnosing Heart Disease Patients ". Heart disease is the leading cause of death in the world over the past 10 years. Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. Decision Tree is one of the successful data mining techniques used. However, most research has applied J4.8 Decision Tree, based on Gain Ratio and binary discretization. Gini Index and Information Gain are two other successful types of Decision Trees that are less used in the diagnosis of heart disease. Also other discretization techniques, voting method, and reduced error pruning are known to produce more accurate Decision Trees. This research investigates applying a range of techniques to different types of Decision Trees seeking better performance in heart disease diagnosis. [7,12]

In 2012 M. Akhil jabbar , Dr. Priti Chandra , Dr. B. L. Deekshatuluc Heart Disease Prediction System using Associative Classification and Genetic Algorithm Associative classification is a recent and rewarding technique which integrates association rule mining and classification to a model for prediction and achieves maximum accuracy. Associative classifiers are especially fit to applications where maximum accuracy is desired to a model for prediction. There are many domains such as medical where the maximum accuracy of the model is desired. Heart disease is a single largest cause of death in. They proposed an efficient associative classification algorithm using genetic approach for heart disease prediction. The main motivation for using genetic algorithm in the discovery of high level prediction rules is that the discovered rules are highly comprehensible, having high predictive accuracy and of high interestingness values. In this paper, we proposed a system for heart disease prediction using data mining techniques. In our feature work we plan to reduce no. of attributes and to determine the attribute which contribute towards the diagnosis of disease using genetic algorithm[8,13]

In 2013 V. Krishnaiah , Dr. G. Narsimha, Dr. N. Subhash Chandra Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques Cancer is the most important cause of death for both men and women. The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. A disease that is commonly misdiagnosed is lung cancer. Earlier diagnosis of Lung Cancer saves enormous lives, failing which may lead to other severe problems causing sudden fatal end. Its cure rate and prediction depends mainly on the early detection and diagnosis of the disease. One of the most common forms of medical malpractices globally is an error in diagnosis. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, we briefly examine the potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to massive volume of healthcare data. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information. For data preprocessing and effective decision making One Dependency Augmented Naïve Bayes classifier (ODANB) and naive creedal classifier 2 (NCC2) are used[9,14].

III. PROBLEM STATEMENT

The heart is very important part of human body. Which pumps blood into the entire 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[9,10,11].

Some of the risk factors for heart disease are

1. Smoking: Smokers risk a heart attack twice as much as non smokers.

2. Cholesterol: A diet low in cholesterol and saturated Tran's fat will help lower cholesterol Levels and reduce the risk of heart disease.

3. Blood pressure: High BP leads to heart Attack

4. Diabetes: Diabetes if not controlled can lead to significant heart damage including heart attack and death.

5. Sedentary life style: Simple leisure time activities like gardening and walking can lower our risk of heart disease.

6. Eating Habits: Heart healthy diet, low in salt, saturated fat, Trans fat, cholesterol and refined sugars will lower our chances of getting heart disease.

7. Stress: Poorly controlled stress an danger can lead to heart attacks and strokes.

IV. PROPOSED METHOD

In the proposed method we use the concepts of fitness number. In the proposed approach we convert the given data set into binary format as per the given condition for heart attack. We divide the large data set into number of parts. We perform simple intersection calculation for every part of data set and finally we consider only those item set which satisfy the given minimum support fitness value Consider a simple example:-

Table 1 Simple patient's database

S.NO	1	2	3	4	5	6	7	8	9	10
1	63	1	145	233	1	2	150	2	3	6
2	67	1	160	286	0	2	108	1	2	3
3	67	1	129	229	0	2	129	2	2	7
4	37	1	130	250	0	0	187	3	3	3
5	40	0	130	204	0	2	172	1	1	3
6	56	1	126	236	0	0	178	0	1	3
7	62	0	140	268	0	2	160	3	3	3
8	57	0	120	354	0	0	163	0	1	3
9	63	1	130	254	0	2	147	1	2	7
10	53	1	140	203	1	2	155	3	3	7

Table 2 Description of attribute for patient's database

S.NO	Attribute
1	Age
2	Sex
3	Blood pressure
4	Cholesterol
5	Fasting blood sugar
6	Resting ECG
7	Thalach value
8	Old peak
9	Slope
10	Thal

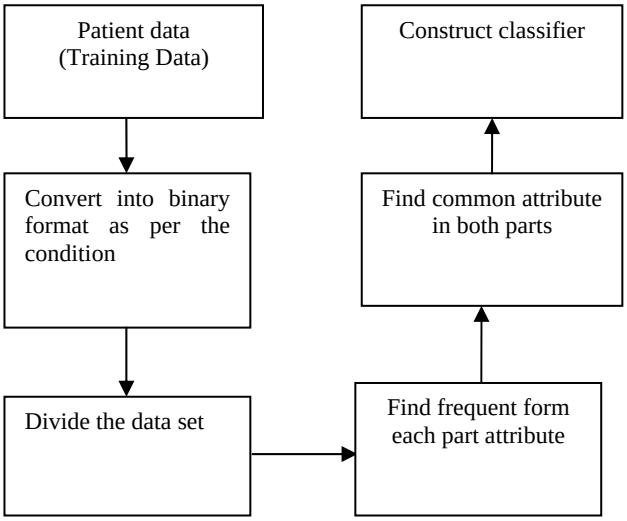

Figure 3 working process of proposed method

We have taken 10 attribute which are responsible for the heart attack problem. We convert the given test data set into binary format with a possible conditions for heart attack.

Possible conditions for heart attack

Age>45, BP>120, Cholesterol range>240, FBS>120, Resting ECG>1, Thalach value>100 Beats/Minute, Old peak>0, Slope>=2, Thal Value>3,

As per the given conditions we convert the test data set into binary format. 1 represent for condition true and 0 for condition false.

Table 3 binary conversion as per the given condition

S.NO	1	2	3	4	5	6	7	8	9	10
1	1	1	1	0	0	1	1	1	1	1
2	1	1	1	1	0	1	1	1	1	1
3	1	1	1	0	0	1	1	1	1	1
4	0	1	1	1	0	0	1	1	1	1
5	0	0	1	0	0	1	1	1	0	1
6	1	1	1	0	0	0	1	1	0	1
7	1	0	1	1	0	1	1	1	1	1
8	1	0	0	1	0	0	1	0	0	1
9	1	1	1	1	0	1	1	1	1	1
10	1	1	1	0	0	1	1	1	1	1

Divide the data set into two part.

Table 4 first part of the data base

S.NO	1	2	3	4	5	6	7	8	9	10
1	1	1	1	0	0	1	1	1	1	1
3	1	1	1	0	0	1	1	1	1	1
5	0	0	1	0	0	1	1	1	0	1
7	1	0	1	1	0	1	1	1	1	1
9	1	1	1	1	0	1	1	1	1	1

Table 5 second t part of the data base

S.NO	1	2	3	4	5	6	7	8	9	10
2	1	1	1	1	0	1	1	1	1	1
4	0	1	1	1	0	0	1	1	1	1
6	1	1	1	0	0	0	1	1	0	1
8	1	0	0	1	0	0	1	0	0	1
10	1	1	1	0	0	1	1	1	1	1

Taking the first part

FV(1)=4,FV(2)=3 ,FV(3)=5 ,FV(4)=2 FV(5)=0 FV(6)=5, FV(7)=5, FV(8)=5, FV(9)=4 , FV(10)=5

One attribute satisfy the given condition

FV(1)=4,FV(2)=3 ,FV(3)=5 , FV(6)=5, FV(7)=5, FV(8)=5, FV(9)=4 , FV(10)=5

Two candidate attribute set which satisfy the given condition

FV(1,2)=3, FV(1,3)=4, FV(1,6)=4, FV(1,7)=4, FV(1,8)=4, FV(1,9)=4, FV(1,10)=4, FV(2,3)=3, FV(2,6)=3, FV(2,7)=3, FV(2,8)=3, FV(2,9)=3, FV(2,10)=4 FV(3,6)=5, FV(3,7)=5, FV(3,8)=5, FV(3,9)=4, FV(3,10)=4 FV(6,7)=5, FV(6,8)=5, FV(6,9)=4, FV(6,10)=5 FV(7,8)=5, FV(7,9)=4, FV(7,10)=5 FV (8,9)=4, FV(8,10)=5, FV(9,10)=4

We repeat this process for three, four … attribute and finally we got

(1,3,6,7,8), (1,3,7,8,10),(1,3,6,7,9),(1,3, 7, 8, 9)

We repeat this procedure for seconds' part and we got (1,2,3,7,8), (1,2,3,7,10) (1,2,3,8,10), (1,3,7,8,10)

Table 6 second part of the data base

S.NO	1	2	3	4	5	6	7	8	9	10
2	1	1	1	1	0	1	1	1	1	1
4	0	1	1	1	0	0	1	1	1	1
6	1	1	1	0	0	0	1	1	0	1
8	1	0	0	1	0	0	1	0	0	1
10	1	1	1	0	0	1	1	1	1	1

From the result of both parts we find (1, 3, 7, 8, 10) are common attribute satisfying the condition Age>45, Blood pressure > 120, Thalach Value>100, Old Peak>0, Thal Value>3 mean heart Attack possibilities.

V. EXPERIMENTAL EVALUATION

We use VB dot net 2010 as front end and SQL server as back end for data base. .All the experiments were performed on a i3 4M Cache, 2.50 GHz Intel PC machine with 2 gigabyte main memory, running Microsoft Windows 7. To evaluate the performance Real life dataset is used.

We have implemented three algorithms first one Bayesian classification algorithm, second is weight associated classifiers and third is our proposed method. We have taken 10 attribute which are mainly responsible for heart attack problem. We have taken the data from a pathology laboratory. We perform experiments using 1000, 2,000 and 5000 records.

VI. COMAPARISION AND GRAPHS

We have taken number of record and number of record classify accurately. From the graph it clearly show that the percentage of classify correct records by proposed method is more as compared to the previous methods.

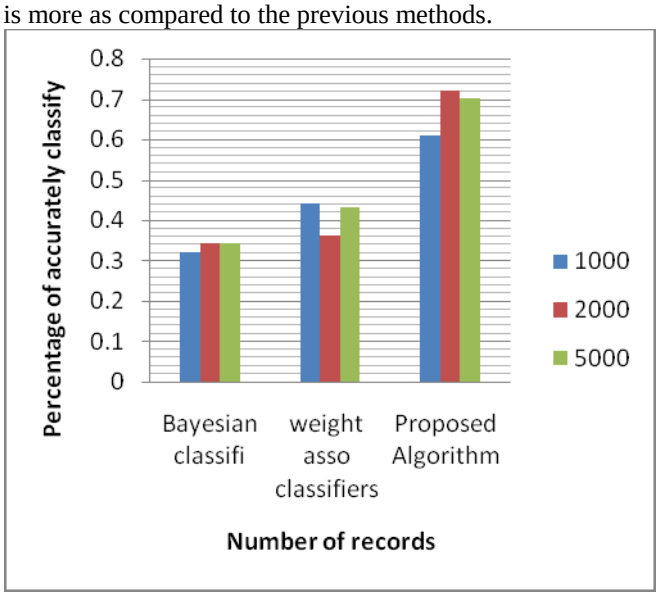

Figure 4 percentage of records classify correctly

VII. CONCLUSION AND FUTURE WORK

From the experiment it clear that proposed method is more accurately classify the recodes as compared to pervious method. Proposed method considers all attribute given to heart attack condition. Proposed method is also simple to under stands and calculation is also easy. We have taken only ten attribute which are mainly responsible for heart attack , in future we have consider more then ten attribute which are also responsible for heart attack.

REFERENCES

[1] S. Sandhiya P. Pavithra and A. Vidhya "Novel Approach for Heart Disease verdict Using Data Mining Technique" International Journal of Modern Engineering Research (IJMER) www.ijmer.com ISSN: 2249-6645

[2] Divya Tomar and Sonali Agarwal A survey on Data Mining approaches for Healthcare International Journal of Bio-Science and Bio-Technology Vol.5, No.5 (2013), pp. 241-266

[3] N. Suneetha CH.V. M. K. Hari and V. Sunil Kumar Modified Gini Index Classification: A Case Study Of Heart Disease Dataset (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 06, 2010, 1959-1965

[4] Sunita Soni O. P. Vyas Using Associative Classifiers for Predictive Analysis in Health Care Data Mining International Journal of Computer Applications (0975 – 8887) Volume 4 No.5, July 2010

[5] Anushya & A. Pethalakshmi A Comparative Study of Fuzzy Classifiers with Genetic On Heart Data. International Conference on Advancement in Engineering Studies & Technology, ISBN 978-93-81693-72-8, 15th JULY, 2012, Puducherry

[6] Mariammal. D, Jayanthi. S, Dr. P. S. K. Patra Major Disease Diagnosis and Treatment Suggestion System Using Data Mining Techniques International Journal of Advanced Research in Computer Science & Technology IJARCST All Rights Reserved 338 Vol. 2 Issue Special 1 Jan-March 2014 ISSN: 2347 - 8446 (Online) ISSN: 2347 - 9817

[7] V. Manikantan & S. Latha Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods International Journal on Advanced Computer Theory and Engineering (IJACTE) ISSN 2319 – 2526, Volume-2, Issue-2, 2013.

[8] Milan Kumari and Sunila Godara Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction IJCST Vol. 2, Issue 2, June 2011 ISSN 2229 - 4333

[9] Nidhi Bhatla and Kiran Jyoti A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic International Journal of Computer Applications (0975 – 8887) Volume 54– No.17, September 2012

[10] V.Jaya Rama krishniah D.V.Chandra Sekar and Dr. K. Ramchand H Rao Predicting the Heart Attack Symptoms using Biomedical Data Mining Techniques V Volume 1, No. 3, May 2012 ISSN 2278-1080 The International Journal of Computer Science & Applications (TIJCSA)

[11] Classification of Data for Heart Disease Prediction System Using MLP International Journal in Multidisciplinary and Academic Research (SSIJMAR) Vol. 2, No. 5, August – September 2013 (ISSN 2278 – 5973)

[12] Mai Shouman, Tim Turner, Rob Stocker Using Decision Tree for Diagnosing Heart Disease Patients Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia Copyright © 2011, Australian Computer Society

[13] M. Akhil jabbar, Dr B.L Deekshatulu, Dr. Priti Chandra "Heart Disease Classification Using Nearest Neighbor Classifier With Feature Subset Selection Computer Science and Telecommunications 2013|No.3(39) ISSN 1512-1232.

[14] Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques V. Krishnaiah (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 – 45

[15] Mrs. G. Subbalakshmi and Mr. K. Ramesh Decision Support in Heart Disease Prediction System using Naive Bayes Journal of Computer Science and Engineering (IJCSE) ISSN : 0976-5166 Vol. 2 No. 2 Apr-May 2011