Predictive Risk Factors of Heart Disease using an Efficient Classification based Approach

Sapna Sharma
Department of Computer Science & Technology
Sachdeva Institute of Technology Mathura

Pankaj Sharma, PhD
Professor
Department of Computer Science & Technology
Sachdeva Institute of Technology Mathura

ABSTRACT
Medical data mining used content, structure and methods to analyze the medical data. Data mining techniques and machine learning algorithms play a very important role in this area. Advanced data mining techniques can be used to discover hidden pattern in data. The term Heart disease encompasses the various diseases that affect the heart. Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad The researchers accelerating their research works to develop a software with the help machine learning algorithm which can help doctors to take decision regarding both prediction and diagnosing of heart disease. Several data mining data classification techniques like to Decision Tree, Artificial neural networks (ANNs), and Support Vector Machine (SVM) , Naïve Bayes, KNN and rule based classifier are used to classifies disease dataset. Performance of these techniques is compared through sensitivity, specificity, accuracy, error rate, True Positive Rate and False Positive Rate.IN this paper we proposed a new and more accurate techniques to classify heart disease. Proposed approach is based on dividing of dataset and used responsible symptoms based generations and combinations of these patterns to predict heart disease.

Keywords
Prediction, Disease, Patterns, diagnosis, Symptoms

1. INTRODUCTION
Data mining is the process of automatically extracting knowledgeable information from huge. The data generated by the health organizations is very vast and complex due to which it is difficult to analyze the data in order to make important decision regarding patient health. This data contains details regarding hospitals, patients, medical claims, treatment cost *etc.* So, there is a need to generate a powerful tool for analyzing and extracting important information from this complex data. The analysis of health data improves the healthcare by enhancing the performance of patient management tasks. Recent technologies are used in medical field to enhance the medical services in cost effective manner. Correctness of the classifier could be tested using test dataset. Classification is one of the most widely used methods of Data Mining in Healthcare organization. The analysis of health data improves the healthcare by enhancing the performance of patient management tasks. The outcome of Data Mining technologies are to provide benefits to healthcare organization for grouping the patients having similar type of diseases or health issues so that healthcare organization provides them effective treatments.[8]

2. LITERATURE SURVEY
In 2011 Jyoti Soni and Sunita Soni proposed "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers". They designed a GUI based Interface to enter the patient record and predict whether the patient is having Heart disease or not using Weighted Association rule based Classifier. The prediction is performed from mining the patient's historical data or data repository. In Weighted Associative Classifier (WAC), different weights are assigned to different attributes according to their predicting capability. They proved that the Associative Classifiers are performing well than traditional classifiers approaches such as decision tree and rule induction. Further from experimental results it has been found that WAC is providing improved accuracy as compare to other already existing Associative Classifiers[1].

In 2012 Chaitrali S. Dangare and Sulabha S. Apte, proposed "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques". They analyzed prediction systems for Heart disease using more number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. They added two more attributes i.e. obesity and smoking. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database. The performance of these techniques is compared, based on accuracy. As per our results accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively[2].

In 2013 M. A . Nishara Banu, B Gomathy proposed "Disease Predicting System Using Data Mining Techniques". Classification is an important problem in data mining. Given a database contain collection of records, each with a single class label, a classifier performs a brief and clear definition for each class that can be used to classify successive records. A number of popular classifiers construct decision trees to generate class models.. The heart disease database is clustered using the K-means clustering algorithm, which will remove the data applicable to heart attack from the database[3].

In 2014 Sathyabama Balasubramanian, Balaji Subramani proposed "Symptom's Based Diseases Prediction in Medical System by Using K-Means Algorithm". The**y** created and supported the historical knowledge. To compute the chance of prevalence of explicit unwellness from medical knowledge by using k-mean ,Large memory They reduces the multiple diseases showing the similar symptoms problem and it will increase the accuracy of such diagnosis. They used Hopfield network, LAMSTAR Network and K-Means algorithm to assist the doctors to perform differential diagnosis along with the possible implementation using SOA technique. By using

these techniques, it improves the overall speed and increase the accuracy of algorithm [4].

in 2015 Gadoya Komal, DR.Vipul Vekariya proposed "Novel Approach for Heart Disease Prediction" They developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees i.e. ID3.By using medical profiles of patients such as age, gender, blood pressure and blood sugar, chest pain, ECG graph etc, it can predict the likelihood of patients getting a heart disease or not. The proposed system is implemented in MATLAB as an application that takes parameters of medical test as an input. They shows as a training tool to train nurses ,medical students, and also for fresher in medical analysis to diagnose patients with heart disease. Basically, the learning of the decision tree from class labeled training tuples is called Decision tree induction[5].

In 2016 J. Vijayashree and .Ch. Sriman Narayana Iyengar proposed "Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review ". The objective of this review is to widespread about Heart related cardiovascular disease and to brief about existing decision support systems for the prediction and diagnosis of heart disease supported by data mining and hybrid intelligent techniques Many DSS exists to predict the heart disease with various techniques. The World life expectancy statistics implies that heart disease is prevailing more in number. So it is necessary to build an efficient intelligent trusted automated system which predicts the heart disease accurately based on the symptoms according to gender/age and domain knowledge of experts in the field at the lowest cost[6].

In 2017 P. Priyanga, Aparna N Venu, Latha Eshappa, M Archana, Navya Shree D P proposed "A Literature Survey on Mining they proposed efficient data mining techniques to select the best features with the lowest costs and shortest times and machine learning algorithms to achieve the accuracy. This technique will help to reduce the work load and cost for patients as well as health care unit. Different technologies give different precision depending on a number of attributes considered. These systems improve the quality of clinical evidence-based decisions and help reduce the financial and timing cost taken by patients. Hence, understanding the usefulness of data mining for early diagnosis of various heart diseases becomes important[7].

In 2018 Poornima V, Gladis D proposed "A novel approach for diagnosing heart disease with hybrid classifier" They proposed an Orthogonal Local Preserving Projection (OLPP) method to reduce the function dimension of the input high-dimensional data. The dimension reduction improves the prediction rate with the help of hybrid classifier. They used LM algorithm is used to solve the optimization problem and it determines the best network parameters such as weights and bias that minimizes the error. With experimental analysis it is clear that optimization the performance accuracy, sensitivity, and specificity. From the result, it is observed that hybrid optimization techniques increase the accuracy of the heart disease prediction system[9]

3. PROPOSED APPROACH

The proposed approach is based on dividing the data set into number of parts and used

responsible symptoms value for heart disease. We have taken 10 attribute which are responsible for the heart attack problem. Possible conditions for heart attack.

1. Age
2. BP
3. Cholesterol
4. FBS
5. Resting ECG
6. Thalach value
7. Beats/Minute
8. Old peak
9. Slope
10. Thal Value,

The values of these attribute for heart attack symptoms are taken by the expert.

Proposed Algorithm
Input:
D Heart patient database
F Minimum number of symptoms
Output:
Combination of symptoms satisfy responsible conditions for heart attack Method:

1. Scan the Heart patient database D and partition into two equal sizes.
2. Find minimum number attribute which satisfy symptoms responsible for heart attack disease
3. Those attribute which do not satisfy symptoms responsible for heart attach disease are with minimum number will be deleted.
4. To discover combination of two attribute Join $L1 \bowtie L2$ and perform logical AND. Those attribute which do not satisfy symptoms responsible for heart attach disease are with minimum number will be deleted are remaining are consider a higher level.
5. The algorithm iterates to find combination of higher attribute which satisfy symptoms responsible for heart attack up for n iterations
6. Each pair finds out n-attribute item sets. which satisfy symptoms responsible for heart attack up for n iterations
7. Same process is repeated for second partition
8. Intersect is used to find common the pair of attribute from each part to get global set of attribute which satisfy the symptoms.

4. ARCHITECTURE OF THE PROPOSED METHOD

First we convert data set into binary format according to the given conditions. In seconds step we divide the data set into two part apply the fitness value condition on each attribute Find pair for each attribute which satisfy the condition .we repeat the process for grouping the attribute until no more grouping is possible. At last we find the most common attribute in both the part and calculate how much percentage data is accurately classified. The working process of proposed model is show in the figure 3.3.

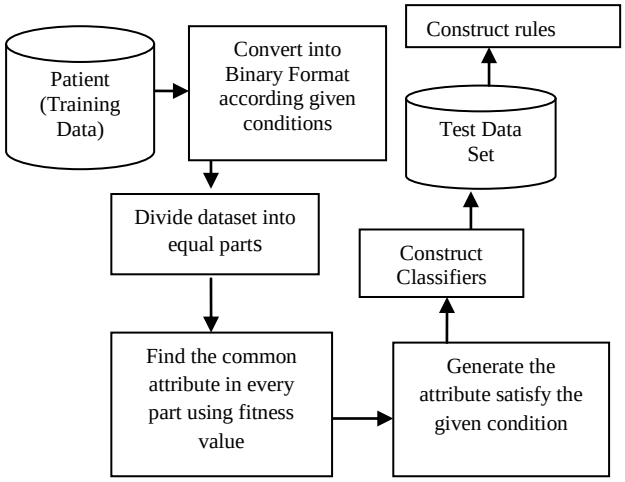

Figure 1 Outline of the proposed approach

5. ILLUSTRATE WITH EXAMPLE

Consider 10 attribute which re responsible for heart attack.

Table 1 Ten patient records with symptoms

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

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

One attribute satisfy the given condition

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

Two candidate attribute set which satisfy the given condition

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

Three candidate attribute set which satisfy the given condition

Divide the data set into two parts

Table 2 first part of the data set containing odd records

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 3 first part of the data set containing even records

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

SV(1,3,6)=4, SV(1,3,7)=4, SV(1,3,8)=4, SV(1,3,9)=4, SV(1,3,10)=4

Four candidate attribute set which satisfy the given condition

SV(1,3,6,7)=4, SV(1,3,6,8)=4, SV(1,3,6,9)=4, SV(1,3,6,10)=4
We repeat this process and last we got We repeat this process for three, four … attribute and finally we got
(1,3,6,7,8) (1,3,6,7,10), (1,3,6,8,10), (1,3, 7, 8, 10)……..
The attribute which are present in all the sets are {1,3,6,7,8,9,10}
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).

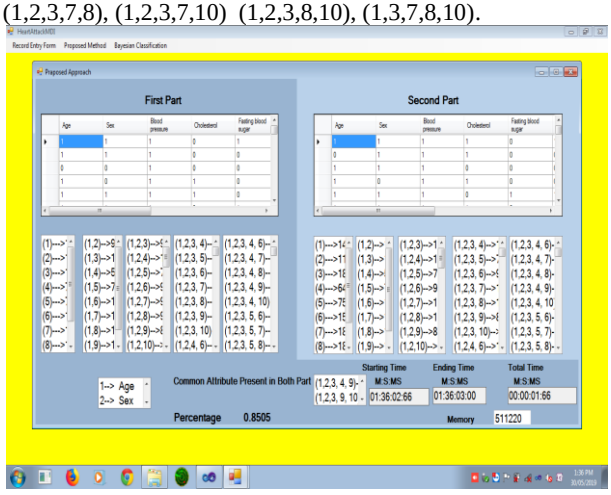

Figure 2 Output of proposed approached

6. IMPLEMETTION AND RESULTS

We have taken 10 attribute which are mainly responsible for heart attack condition. Age, Sex, Blood pressure, Cholesterol, Fasting blood sugar, Resting ECG, Thalach value, Old peak, Slope and Thal. We have taken real life data from a laboratory. We have taken data of 1000 patient.SQL Server R2 (2008) to store our database. For comparing the performance of the proposed algorithm with other methods we take records on the different number of records. In table 5.1 we have accuracy of the proposed method with other methods on 100, 500 and 1000 records of data. This comparison is based on execution time and number of objects.

Table 4 comparison using number of records and accuracy

Number of Records	Bayesian Classifiers	Weighted Classifiers	Proposed Method
100	0.0237	0.3682	0.7242
500	0.0364	0.3622	0.6234
1000	0.0282	0.3556	0.6021

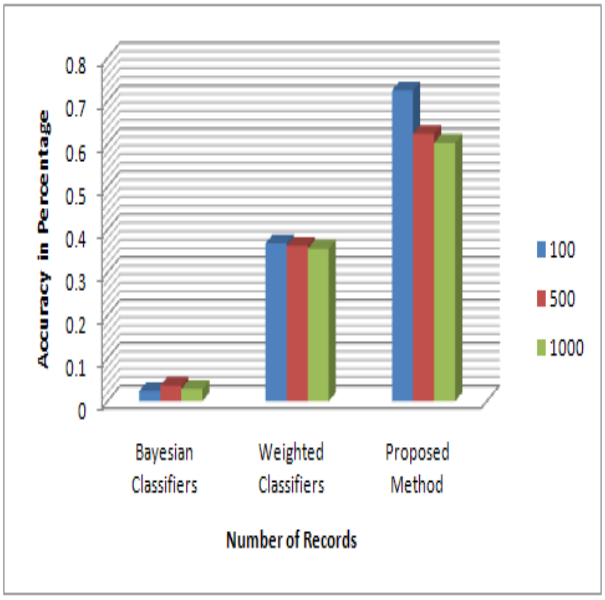

Figure 3 comparisons using number of records and accuracy

7. CONCLUSION

In this paper we proposed a new and efficient approach which predict heart disease problem at early stage. Proposed approach gives more accurate result by dividing data set into number of parts. By using most suitable symptoms value. Proposed approached used combination of symptoms value. By using minimum combination of attributes proposed approached predict heart diseases.

8. REFERENCES

[1] Jyoti Soni, Uzma Ansari, Dipesh Sharma "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers" Jyoti Soni et al. / International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 3 No. 6 June 2011 2385.

[2] Chaitrali S. Dangare " Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

[3] M.A.Nishara Banu, B Gomathy " Disease Predicting System Using Data Mining Techniques" International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 1, Issue 5 (Nov-Dec 2013), PP. 41-45

[4] Sathyabama Balasubramanian, Balaji Subramani, International Journal of Advances in Computer Science and TechnologyAvailable Online at al Journal of Advances in Computer.

[5] Gadoya Komal1, DR.Vipul Vekariya Novel Approach for Heart Disease Prediction Using Decision Tree Algorithm International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 11, November 2015

[6] J.Vijayashree and N. Ch. Sriman Narayana Iyengar "Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review" International Journal of Bio-Science and Bio-Technology Vol.8, No.4 (2016), pp. 139-148

[7] P. Priyanga, Aparna N Venu, Latha Eshappa2, M Archana2, Navya Shree D P2 A Literature Survey on Mining Health Data to Predict Heart Disease International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 5, Issue 5, May 2017.

[8] Shratik J. Mishra , Albar M. Vasi , Vinay S. Menon, K. Jayamalini "GDPS - General Disease Prediction System" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072.

[9] Poornima V, Gladis D "A novel approach for diagnosing heart disease with hybrid classifier" Biomedical Research 2018; 29 (11): 2274-2280 ISSN 0970-938X www.biomedres.info