

Efficient Prediction of Heart Disease using Bagging

Merin Mary Kurian¹, Shamna Mol B², Chinnu Thomas³, Abhijith Kumar⁴, Raveendran R⁵
^{1,2,3,4}*B.Tech Student, Mount Zion College of Engineering, Kadammanitta, Kerala, India*
⁵*Assistant Professor, Department of Computer Science and Engineering, Mount Zion College of Engineering, Kadammanitta, Kerala, India*

Abstract- Heart Disease or Cardiovascular disorder is considered to be the most threatening condition today that can lead to death most often. Many scholars reveal that the number of people losing life due to heart disease is increasing dramatically. Therefore, it is necessary to predict the heart disease at the earliest. Data mining techniques play the best role in mining different data of the patient, thereby predicting the disease correctly.

Index Terms- Bagging, weka tool, classification.

I. INTRODUCTION

Data mining is defined as a process used to extract usable data from a larger set of any raw data. Heart is the most essential organ in human body. If this organ gets effected, then it also affects the other vital parts of the body. Data mining techniques are used to predict the heart disorder. Advanced data mining techniques can help remedial solutions for these problems. It is a web-based user-friendly system and can be used in hospitals. Quality service implies diagnosis patients correctly and administering treatments that are effective. Diagnosis is complicated and as well as an important task that needs to be executed accurately and efficiently. Automatic Medical diagnosis system would be exceedingly beneficial. Today, Heart disease is a broad term that can refer to any conditions that reflect the heart. These patients have symptoms like chest pain and fatigue, which occur when the heart is not receiving adequate oxygen quantity. A number of factors have been shown to increase the risk of developing heart disease. The different attributes such as age, cholesterol, type of chest pain, major vessels etc are considered while evaluating the heart disease.

The majority of research efforts in this area has involved the development of diagnostic tools that are

used to provide secured opinions for physicians in diagnosing. Thus, in order to identify the warning signs of heart disease in a patient, we propose the use of bagging algorithm to mine the heart disease database. Therefore, after analysing many scholastic articles and researches, we have come to a conclusion that bagging is the best technique because of its more accurate solution and less time complexity.

II. AIM

Objective of this project are:

- Efficiently predict the chance of heart disease by using an accurate method.
- Prove that the method which is used here is comparatively effective.

III. OVERVIEW OF RELATED SURVEY

Many papers have been released based on this topic. These studies apply different techniques of data mining. In recent years, heart disease issues have increased more than a limit. The main reason behind this is the late prediction of the issue.

A. Heart Disease Prediction Using Data Mining

Abhishek Rairakar proposed that KNN (k-nearest neighbour) has the best accuracy and reduced time complexity. He has compared it with other techniques like Decision tree and Naive Bayes.

B. Using Data Disease Forecasting System Using Methods

Clustering algorithms are mainly used to cluster the data to process it. The C4.5 technique was used to put forward the concept of Information Entropy.

C. Effective Diagnosis of Heart Disease

Chau Tu used the bagging algorithm to identify the symptoms of heart disease. The bagging algorithm is

compared with the decision tree algorithm and has proved to be the best technique.

There are many other papers which deal with same techniques but do have a number of limits and disadvantages.

IV. EXISTING SYSTEM

A. Decision Tree

Decision Tree is defined by Han and Kambr. According to the definition it is a flow chart like tree structure, it consists of nodes like internal node, leaf node and root node as well as branches. Decision tree technique is very powerful for predicting the heart disease. It is one of the non-parametric method based on supervised learning.

ID3, C4.5, CART and J48 are the popular decision tree algorithms for predicting heart disease.

ID3: Iterative Dichotomizer 3 is abbreviated as ID3. In this decision tree is constructed in top-down manner.

C4.5: It is a tool which helps for diagnose heart disease.

CART: It is a non-parametric technique for classification or regression.

J48: it is the implementation of ID3, and is developed by weka team.

B. REP Tree

It is the pruning technique based on machine learning. It helps the decision tree to reduce the size and complexity it leads to more accuracy.

It consists of 2 methods,

1. Post pruning: it is created after the construction of tree.
2. Online pruning: it is done during the creation.

D. Naïve Bayes Theorem

It is the classification method based on classifiers, it learns from the given set of training data. By Naïve Bayes rule we can compute the probability. After computing that it helps to predict the class with high probability. The technique that is based on Naïve Bayes algorithm is very simple and it is more efficient.

Advantages

1. We can implement it very easily.
2. Small amount of training data is required to determine the parameters.

3. It helps to obtain good results.

Disadvantages

1. It is not much accurate.

V. ATTRIBUTES

We can use more attributes for heart disease prediction. The database consists of 76 raw attributes, we use only 14 attributes for this.

1. Age: it takes age as input
2. Sex
3. Type of chest pain
4. Rest blood pressure
5. Chol: serum cholesterol in mg/Dl
6. Fasting blood sugar
7. Restecg
8. Thalach
9. Exang : it means exercise induced angina
10. Old peak
11. Slope
12. C A: the major vessels
13. Thal
14. Num: have 2 input values
Value 0: no heart disease
Value 1: has heart disease

VI. SYSTEM CONFIGURATION

A. System Overview

Here we didn't consider the requirement of any hardware. We use the Windows 7 which is an efficient and better operating system.

B. NetBeans 7.2.1

By using NetBeans, we can develop our application as different modules.

C. MySQL

VII. PROPOSED SYSTEM

Bagging is a classification technique which consist of a machine learning algorithm designed to improve the stability and accuracy of algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting. It is usually applied to decision tree method. In data mining, combining the outputs of different data models are more reliable in taking decisions.

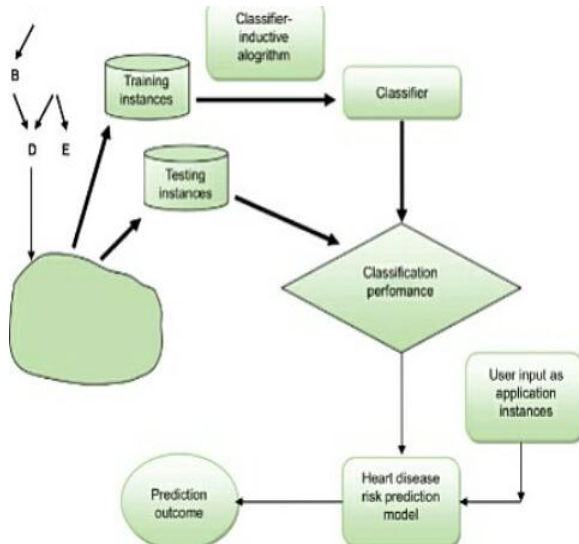


Fig 1: Design of Bagging system

A large data set is created and based on these data, the Bagging attempts to neutralize the instability of learning methods. In this paper, we mainly propose an efficient approach that can produce accurate results at reduced time complexity through different procedure. Other than Naïve Bayes, decision tree and other different types of decision tree like REP and J48, we have come to a conclusion that bagging is yet another important technique which is more accurate in its precision and recall value. Based on values of certain attributes like restecg, chol, fbps etc., heart attack probability can be predicted. There are several steps in predicting the heart disease probability for a patient. Some of the modules included are:

1. *Patient registration and login.*
The patient can register and login whenever they need to check their heart condition. Thereby, compare the previous values with the current ones.
2. *Data entering.*
The data is entered based on the values of each attributes. These attributes have got certain ranges.
3. *Predicting chance of Heart Disease*
Predicting the chances of heart disease using bagging technique. Bagging is a well-known classification technique known for its accuracy and less time complexity.
4. *My Bookings:*
After validating the probability of the disease, the user can enter into ‘My Bookings’ and can select any doctors according to their location, interest and the availability of the particular doctor.
5. *Doctor login page:*

Now, the doctor can get to know who has taken an appointment for him, when and at what time. This can be helpful for the doctor in his work and can maintain the time slot accurately.

VII. RESULT

The techniques were conducted as phases based on weka tool. We used the bagging technique with Different attributes to evaluate the accuracy of classification. Sensitivity, specificity and accuracy are measures used in medical data mining field. Sensitivity produce positive test results and specificity produce negative test results.

The measures are computed as the following equations:

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{\text{Correctly classified}}{\text{Total instances}}$$

This table shows that the bagging result is better than decision tree. The bagging technique is more effective than the decision tree technique.

Table 1: List of accuracy values for each technique

Type	Author/Year	Technique	Accuracy
Single	Cheung, 2001	Decision Tree	81.11%
		Naïve Bayes	81.48%
	Tu, et al., 2009	J4.8 Decision Tree	78.9%
		Bagging algorithm	81.41%

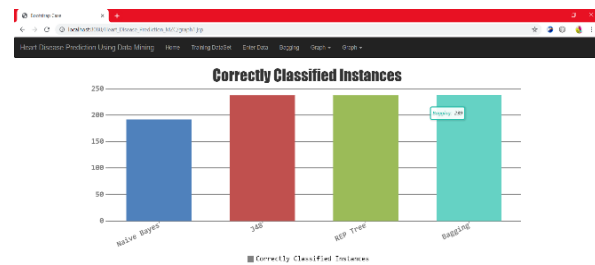


Fig 2: Graph of correctly classified instances for each technique.

VIII. CONCLUSION

In this paper, we have presented an effective heart disease prediction method using data mining. There

are different data mining techniques that can be used for the prediction of heart disease.

Data mining provides many kinds of classification techniques such as k-nearest neighbor (KNN), decision tree like CART, C4.5, J48, ID3 algorithm etc. But all these are weak classifiers. So, there is a need for introducing a technique called bagging to improve their performances. Bagging is a machine learning algorithm used to provide better accuracy and stability compared to other data mining techniques. We have developed a heart disease prediction system using bagging by evaluating a patient's clinical data. Our approach contains several steps. Firstly, we select a number of clinical data such as patient age, sex, cholesterol, trestbps etc. Finally, we predict the chance of occurring heart disease.

ACKNOWLEDGEMENT

We would like to thank KTU for giving us this opportunity to publish our paper. Lastly, we would like to thank Mount Zion College of Engineering. For their support and guidance.

REFERENCES

- [1] D. Resul, T. Ibrahim, S. Abdhulkadir, "Effective Diagnosis of Heart Disease Through Neural Network. Ensembles". Elsevier, 2008.
- [2] Anika Dewan, Meghna Sharma. "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification "(2014) 15, 13-24.
- [3] M. A Nishara Banul, B. Gomathy. "Disease Prediction System Using Data Mining Techniques "(2007)177, 3799-3821.
- [4] Sellappan Palaniappan, Rafiah Swing, "Intelligent Heart Disease Prediction System Using a Data Mining Techniques ", IJCSNS International Journal of Computer Science and Network Security, college. 8 no, August 2008.
- [5] J. Soni, U. Ansari, D. Sharma, S. Soni. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction ". International Journal of Computer Application (0975-8887) volume 17 -no. 8, March 2011.