

A Survey of Heart Disease Prediction Using Classification Techniques

C. Sowmiya¹, Dr. P. Sumitra²

¹Ph.D Research Scholar, Department of Computer Science,

²Professor, Department of Computer Science,

^{1,2}Vivekananda College of Arts and Sciences for Women (Autonomous), Elayampalayam.

sowmiyac83@gmail.com

ABSTRACT

Heart disease is a number one problem for in the world. Every year more than people death for heart disease. This disease for attack not only India all country people affected by heart disease. Some of the people occur deaths for during the first heart attack. Now day's computer field is very high and particular work is finished. So used for medicine area diagnosis. Every year Researchers have been applied data mining techniques and algorithms for diagnosing heart disease. Heart is a very important part of each and every people. Over the past few years more than people deaths, reason from disease is easily attack for patients. Not proper treatment for many hospitals. Mostly affected heart attack for men's. But Women are rare for affected heart attack.

Keywords: Data Mining, Heart Disease, Neural Network, K-Nearest Neighbor Algorithms, Support Vector Machine (SVM), Naive Bayes Algorithm, Decision Tree Method.

I. INTRODUCTION

Now' a day's data mining is very useful for some places and more then details. All the applications of using in world level. Data mining is huge amount of several data base. Data mining is an intelligent creative process. Data mining is used various fields. Now days data mining many places using. This data mining techniques are many advantages and efficient that can be heart disease prediction. Different types of data mining techniques available are classification, cluster, feature selection, association rule can be analyzing the heart disease prediction. . Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific, medical research. Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years due to the wide availability of huge amounts of dataData mining is also called as a KDD Process.KDD process means Knowledge Discovery Process.

Now day's data mining using of more than important application fields.

Medical, Insurance, Telecommunication, Finance, Utilities, Data service providers, Transport, Consumer good , IT , Railway, ship, banking.

II. HEART DISEASE

Now day's heart disease is a main reason for death in the world. Heart is very useful part of our human body. This disease is number one problem for in the world. But not only heart attack some of the disease attacked form valve, ventricle, lung cancer, breast cancer ,diabetes etc.....A common heart disease is nothing, but coronary heart disease or cardiovascular disease is a very dangers disease. Most people attack for heart disease from the world. Coronary heart disease blood vessels around it. This disease causes disability as damage to brain so resulting in death. That ranges of the age group 25 to 69 have 25% risk of having heart disease. Some types of

disease occur from heart. There are several kinds of causes, reason, factor which increase the risk factor of heart disease. There are considered as important reason of heart disease. Most of hospitals admitted in heart disease patient. This disease mostly affected in male because smoking habits. This paper analyzes the different kinds of heart disease using the classification techniques.

Types of disease considered are:

Coronary heart disease, angina pectoris, congestive heart failure, cardiomyopathy, congenital heart disease, arrhythmias, myocarditis, heart attack; heart cancer etc.

There are considered some important reasons of heart disease:

Age, Smoking, Sugar, Obesity, Depression, Hypertension, High blood cholesterol, poor diet, Family history, Physical inactivity.

III. DATA MINING TECHNIQUES

1. Classification
2. Cluster
3. Association rule

Classification

Classification is one of the well known problems under data mining. To classify the data/objects into different classes or groups. For example data can be broken down according to topical content file type, average file size, gigabytes, and megabytes. Classification is the process of learning a function that cans data objects to a subset of a given class set. Some types goals of classification, first finding a good general that can predict the class of but far unknown data objects with high accuracy. Second to find a compact and easy Understandable class model for each other classes.

Cluster

Cluster is a group of objects. For example data elements into different groups of similarity between in a single group cluster partitions the data set in to cluster classes. Each and every near object is neighborhood object. There are two goals of cluster. First one is a inter class

second is an intra class. Inter class cluster means cluster distance is maximized. Intra cluster means cluster distances are minimized.

Association rule

Association rule mining is a very import rule of data mining techniques. Association rule is identifying of association huge data base and their values. In this pattern creative techniques which does not serve to solve classification problems and predict problems. This paper focus on study of exiting heart disease prediction task by a using data mining techniques and different issues in exiting of heart disease prediction.

IV. LITERATURE REVIEW

k.srinivas etal(2011) “presented application of data mining techniques in healthcare prediction of heart attacks”. The powerful use of classification various data mining techniques using such as a decision tree,k-nearest neighgor naïve bayes. Huge volume of health care data using in the data mining tool Tanagra used for conducted data analysis for the learning purpose. This paper consider of 3000 instances for training data set it is various type 14 attributes. The instances data set is providing the result of various types of testing to say the accuracy of heart disease. The performance of the classifiers set of value and results are analyzed. So to the attributes dataset is divided into two parts. Using training data set is 70% and testing dataset is 30%. The comparison of some classification algorithms using. This paper considered the best algorithm naïve bayes.this model using the naïve bayes taken a time to run data for best result achieved. comparision of various classifications algorithm hence better performance accuracy in naïve bayes.

Algorithm	Accuracy	Time Taken
Naïve bayes	52.33%	609 ms
Decision tree	52%	719 ms
k-nearest neighbor	45.67%	1000 ms

A1-Milli N, (2013) “Back Propagation algorithm method. To improve the multilayer neural networks in a supervised manner. Back propagation algorithm based on error correction learning. This algorithm based through the various layer of network. From the forward and backward class. Compare to oher algorithm better result for back propagation algorithm .this algorithm research to compare classification techniques. The author efficiency and deliver high accuracy from the heart disease prediction.

Anuradha, srinivasaraghavan, vincyjoseph (2016) “Comparative analysis of accuracy on heart disease prediction using classification method “This paper research method algorithm is naïve bayes and SVM (Support Vector Machine), logistic regression.

Naïve bayes : Naïve bayes algorithm is a good tool in medical diagnosis. The classifier processes each attributes probability in a class. Naïve bayes is a simple good and efficient performance in classification due to good accuracy for used in medical diagnosis.

Support Vector Machine: SVM is a one of the classification method in this type used to recognize patters and data in a regression and classification analysis. SVM is two classes classified recognizes and separates. Similar data by finding the best hyper plane that separates all data points of one class from other class. Mathematical functions are involved in SVM. Its performance magnify with number of attributes.

Logistic Regression: Logistic Regression is mainly used for prediction besides it can also be used calculating the probability of success. Basically logistic regression involves fitting and equation. It is a type of statistical regression analysis method used for approximation and prediction of result of a dependent attributes. Dependent means it can take only some set of values for example good, or bad, true, or false, on, or off.

These papers consider various classification algorithm using. But high accuracy of performance of heart disease prediction in SVM algorithm is better result.

Algorithm	Accuracy
Naïve bayes	75%
SVM(Support vector machine)	80%
Logistic Regression	79%

Mai Shouman, Tim Turner, Rob Stocker(2011) “Using Decision Tree for Diagnosing Heart Disease Patients”. Particular researches have been investigating the application of the various classification techniques. To help health care management with improved accuracy in the diagnosis of heart disease. Andreev a used c4.5 decision tree in the diagnosis of heart disease. Accuracy of 75.73% (Andreev 2006) this paper comparison of 3 classification algorithm c4.5 and j4.8 and bagging algorithm. Considered the best performance algorithm bagging algorithm. When using the UCI repository heart disease data set.

Performance comparison algorithm

Algorithms	Accuracy
J4.8 Decision tree	78.9%
Bagging Algorithm	81.41%
C4.5	75.73%

Neural Network

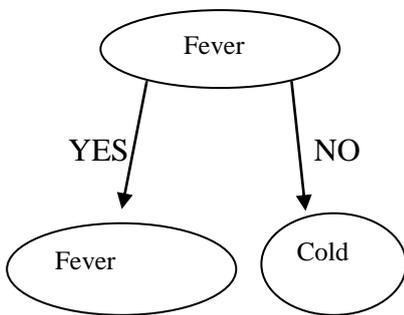
Neural Network are basically used in biological inspired large inter connected cells that simulate the human brain. The perception is the very simplest architecture. This has one neuron and learning method .many sophisticated architecture is mult-layer neural network. Which are one (or) more

neurons connected various layers. Neural network is based on the two types of training method. 1. supervised training method; another method is a 2. un supervised training method.

Decision Tree

Decision tree is a like tree structure format. Decision trees are powerful classification algorithms. It's include the below element.

1. Root node
2. Branch Node
3. Leaf node



Decision tree advantages:

1. Easy to understand and interpret.
2. It not required the domain knowledge to learn

V. FEATURE WORK

A. Naive Bayesian Algorithm

1. Start the process
2. Load the input user value
3. Initial redail velocity for blood prushur e level (Ex; 100 Pations).
- 3.1 Identify the disease occurrence chance with relative input parameters
- 3.2 Check the Train and test each layer under Neuro- functions
4. Assign a separate weight age to each layer
5. Check the deduction data value (normal & abnormal).
6. Print the result
7. Stop the process....

B. PROGRAM CODE

```

Clear;close all;clc;

load 100_ECG_0_20
s1=ECG_1;s2=smooth(s1,150);ecgsmooth=s1-s2;

[C,L]=wavedec(ecgsmooth,8,'db4');
[d1,d2,d3,d4,d5,d6,d7,d8]=detcoef(C,L,[1,2,3,4,5,6,7,8])
;

[thr,sorh,keepapp]=ddencmp('den','wv',ecgsmooth);
cleanecg=wdencmp('gbl',C,L,'db4',8,thr,sorh,keepapp);
max_value=max(cleanecg);
mean_value=mean(cleanecg);
threshold=(max_value-mean_value)/2;

a5=appcoef(C,L,'db4',5);
C1=[a5;d5;d4;d3];
L1=[length(a5);length(d5);length(d4);length(d3);length(
cleanecg)];
R_detect_signal=waverec(C1,L1,'db4');
R_detect_squared=R_detect_signal.^2;

for a=1:length(R_detect_squared)
    if R_detect_squared(a)>threshold
        R_detect_new(a)=R_detect_squared(a);

    else

        R_detect_new(a)=0;

    end

end

mean_R_detect=5*mean(R_detect_new);

for q=1:length( R_detect_new)-1
    if R_detect_new(q)< mean_R_detect
        R_detect_new(q)=0;

    end

end

d=0;

for b=1:length( R_detect_new)-1
    if ( R_detect_new(b)==0) & (
R_detect_new(b+1)~=0)
        d=d+1;
        indext(d)= b+1;

    end

end
end
    
```

```

fs_R_deetect=length(R_detect_new)/20;
time=indext.*1/fs_R_deetect;
ind=0;

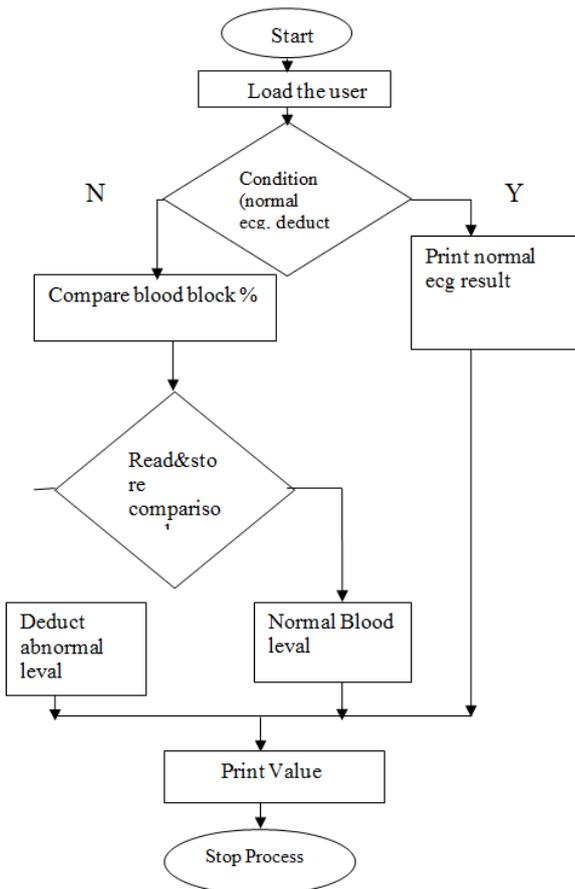
for z=1:length(time)-1
    ind=ind+1;
    time_diff(ind)=time(z+1)-time(z);

end

av_time=mean(time_diff);
Square_Number=av_time/.2;
beat_Rate=300/Square_Number;
high=max(R_detect_new);

subplot(411);plot(s1);title('Orginal Signal');
subplot(412);plot(s1-s2);title('Baseline drift
Elimination');
subplot(413);plot(cleanecg);title('Main Signal');
subplot(414);plot(R_detect_new);title('R detected
Signal');
text(length(R_detect_new)/2,high,['BeatRate=',num2str(
fix(beat_Rate))],'EdgeColor','red'
    
```

C. FLOWCHART



VI. CONCLUSION

In the world every day heart disease problem in attack from people. But every year more than people for death. But not good accuracy presentation in comparisons of various classification algorithm and cluster feature selection method using. This paper considered result. So High accuracy provided from medical field every year death people decrease. The result of the various classification algorithm compared to the Bagging algorithm are very encouraging. The difference in the accuracy is noticeable. This paper the high accuracy and best performance of using Bagging Algorithm 81.41% use the heart disease prediction. It has proven that classification based techniques proposed algorithm – Naïve Bayies contribute high effectiveness and obtain high accuracy compare than the previous methods. Proposed work it can experiment for algorithm Naïve Bayies algorithm

VI. REFERENCES

- [1] Han.j.and m.kamber (2006) Data Mining concepts and techniques, Morgan Kaufmann publishers.
- [2] K .Viswanathan 1, Dr. K .Mayilvahanan 2 , R. Christy Pushpaleela(2016) Prediction in data minng comparative study of heart disease Elysium journal Engineering Research and Management. Volume-3, Issue-4.
- [3] Detrano, R.; Steinbrunn, W.; Pfisterer, M., "International application of a new probability algorithm for the diagnosis of coronary artery disease". American Journal of Cardiology, Vol. 64, No. 3, 1987, pp. 304-310.
- [4] Anuradhasrinivasaraghavan, vincyjoseh(2016) "Comparative analysis of accuracy on heart disease prediction using classification method international Journal of applied information systems volume 11-No 2. July 2016.
- [5] Chau, M.; Shin, D., "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms". Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing 2009, pp.183-187.

- [6] Palaniappan, S.; Awang, R., "Intelligent Heart Disease Prediction System Using Data Mining Techniques". Proceedings of IEEE/ACS International Conference on Computer Systems and Applications 2008, pp. 108-115.
- [7] Keyue Ding and Kent R Bailey et al, " Genotype-informed estimation of risk of coronary heart disease based on genome-wide association data linked to the electronic medical record", International journal of BMC cardiovascular Disorders, Vol.11,2011
- [8] Daisy JA Janssen and Emiel FM Wouters et al, "Self-perceived symptoms and care needs of patients with severe to very severe chronic obstructive pulmonary disease, congestive heart failure or chronic renal failure and its consequences for their closest relatives: the research protocol", Journal of BMC palliative care, Vol.7, 2008
- [9] Shou-En Lu and Gloria L Beckles et al, "Evaluation of risk equations for prediction of short-term coronary heart disease events in patients with long-standing type 2 diabetes: the Translating Research into Action for Diabetes", International Journal of BMC Endocrine Disorders, Vol.12, 2012.
- [10] Eman AbuKhoua, Piers Campbell, " Predictive Data Mining to Support Clinical Decisions: An Overview of Heart Disease Prediction Systems", IEEE, International Conference on Innovations in Information Technology, pp.267-272, 2012.
- [11] Lamia Abed Noor Muhammed, " Using Data Mining technique to diagnosis heart disease", IEEE, International conference on statistics in science, Business and Engineering, pp.1-3, 2012.