

Analytical Study of Heart Disease Diagnosis Using Classification Techniques

C.Sowmiya

[#]Ph.D Research Scholar[Full-Time]

PG and Research Department of Computer Science
Vivekanandha College of Arts and Sciences for Women
(Autonomous)

Elayampalayam, Tiruchengode - 637205
Namakkal (DT), TamilNadu, India
sowmiyac83@gmail.com

Dr.P.Sumitra

Assistant Professor

PG and Research Department of Computer Science
Vivekanandha College of Arts and Sciences for Women
(Autonomous)

Elayampalayam, Tiruchengode - 637205
Namakkal (DT), TamilNadu, India
sumitravaradharajan@gmail.com

Abstract— Heart disease is the number one problem for world. Heart disease more than people deaths occur during the first heart attack .But not only for heart attack have some problems attacked for breast cancer, lung cancer, ventricle. Valve, etc... It is essential to have a frame work that can effectually recognize the prevalence of heart disease in thousands of samples instantaneously. In this paper the potential of nine (9) classification techniques was evaluated of prediction of heart disease. Namely decision tree, naive Bayesian neural network, SVM, ANN, KNN. My proposed algorithm of Apriori algorithm and SVM (support vector machine) in heart disease prediction. Using medical profiles such as a age, sex, blood pressure, chest pain type, fasting blood sugar. It can predict like of patients getting heart disease Based on this, medical society takes part interest in detecting and preventing the heart disease. From the analysis it have proved that classification based techniques contribute high effectiveness and obtain high accuracy compare than the previous methods.

Keywords— Heart disease, Disease diagnosis, Prediction, Classification techniques, Apriori Algorithm.

I.INTRODUCTION

Now day's Heart disease is main reason for death in the world. Heart disease is a number one problem from the world. Heart disease is leading cause of death in the world. Over the past few years. 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.

II.HEART DISEASE

Heart is important part of our human body. More than country affected for heart disease every year some of the million people for death from heart disease. Life is itself dependent on efficient working a heart. as brain, circulation of blood in body is inefficient the organs like brain suffer and if then heart is not properly within it. More than disease attack for heart. Now days many hospitals not proper treatment. But increasing the payment of bill. Some of hospitals average treatment for patients so result is better. Heart disease is a group of condition affecting the structure and function of heart and has more root causes. Heart disease is the leading cause of every year death in the world. Some types of disease occurs attack for heart. Types of disease considered are coronary heart disease, angina pectoris, congestive heart failure, cardiomyopathy, congenital heart disease, arrhythmias, myocarditis, heart attack; heart cancer etc. in this disease is particularly specific very dangers disease to cardiovascular disease or coronary heart disease. There are considered some important reasons of heart disease.

Age Smoking Sugar Obesity Depression Hyper tension
High blood cholesterol poor diet Family history Physical inactivity

Types of heart diseases

A. Angina

It can be referred to as angina pectoris. It occurs when an area of the heart muscle does not get enough oxygen. The patient well experiences in chest discomfort, tightness or pain. It is a symptom of coronary artery disease. Due to lack oxygen in the heart muscle is usually caused by the narrowing of the coronary arteries because of plaque accumulation.

B. Arrhythmia

Arrhythmia is an irregular heartbeat. They caused problems with heart-rhythm. It happens when the heartbeats do not work properly. To make the heart beat in

a better way it should not, either move too fast, slow or erratically.

C. Fibrillation

Fibrillation occurs when the heartbeat is irregular. We experience irregular heartbeats. We feel like a fluttering or a racing heart. Precaution has to be taken when they veer too far from normal heartbeat. Irregular heartbeats can become fatal.

D. Congenital heart disease

It refers to born with it. In the country UK it is surveyed that every 1,000 babies are born with some kind of congenital heart disease.

E. Coronary artery disease

It causes disease or damaged because of cholesterol-containing deposits. Plaque accumulation narrows the coronary arteries and the heart gets less oxygen.

F. Myocardial infarction

The other name is known as heart attack, cardiac infarction and coronary thrombosis. Interrupted blood flow damages or destroys part of the heart muscle. It is usually caused due to blood clot. It can also occur if an artery suddenly becomes narrows.

G. Heart failure

The other name called as congestive heart failure. It does not pump blood around the body efficiently. Left side or Right side or both side of the body might be affected. Coronary artery disease can make the heart stiff or weak to fill and pump properly.

III. DATA MINING

Data mining is huge amount of several data base. Data mining is an intelligent creative process. Data mining is used various fields. Now day's 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 analysing 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.

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

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

Data mining basically using the 4 Techniques.

A. 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.

B. 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 neighbourhood 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.

C. Feature selection

Feature selection also known as variable selection attribute selection and variable subset selection in this process of selecting a subset of relevant features for use in model construction.

D. 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. EXISTING SURVEY OF HEART DISEASE DIAGNOSIS

K.C. Tan and E.J. Teoh et al, [1] proposed a new hybrid approach called the Genetic Algorithms (GAs) and Support Vector Machines (SVMs). From the evolutionary process, they have proved that genetic algorithm is best suited for the collection of attribute data set. It attained a high average accuracy of 76.20%..

Result Das and Ibrahim Turkoglu, et al, [2] have proposed a new effective medical decision support system. The neural network ensemble model is applied. From the experimental analysis the heart disease diagnosis obtains classification accuracy is better.

Akin Ozcift and ArifGulten, [3] have created is heart disease in compare to better accuracy. Many problems accrue from heart. More than classification algorithm

compares. Particular same disease data compare to 60 algorithms. Average accuracy for diabetes. They attained from the proposed RF the accuracy of 74.47%, 80.49% and 87.13% respectively.

Jesmin Nahar and Tasadduq Imam[4], have implemented the heart disease diagnosis by applying computational intelligent techniques. For the major data set the accuracy was seen to be improved by applying the technique called Motivated Feature Selection (MFS). The experimental results proved mainly on the performance of accuracy for most of the classifiers and the majority of the datasets. Finally, they concluded that the proposed system named MFS with Computer Feature Selection (CFS) places a major role in heart disease diagnosis.

Chen A.H [6], have compared several data mining techniques which are mainly in Decision Tree, Naïve Bayes and Neural Network. Five Data mining rules were applied and evaluated by the three models. Among the three models Naïve Bayes performs better in Heart Disease diagnosis when compared with other models.

Eman AbuKhoua, Piers Campbell[7] have experimented the Naïve Bayes technique to build the predictive model.

RaghunathNambiar, and AdhiraajSethi et al [8] evaluated the data mining techniques such as CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and Decision Table (DT). Different classifiers and research are conducted to find the best classifier for calculating the patients Diagnosis. Among the entire algorithm CART gives more accuracy.

K.Srinivas et al(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-nearest neighbour naïve bayes. Huge volume of health care data using in the data mining tool Tanagra used for conducted data analysis for the learning purpose. 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 analysed. 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. comparison of various classifications algorithm hence better performance accuracy in naïve bayes.

V. FEATURE WORK

A. Apriori Algorithm

Step 1: load heartdisease.mat;

Load heartdisease.mat file which contains heart Disease data and its label separately.

Step 2: Randomizing and dividing data into 1:1 ratio for Training and testing.

Step 3: Computing Euclidean distance for each test data Euclidean distance.

For each observation in test data, we compute the Euclidean distance from each observation in training data.

Step 4: compute k nearest neighbors and store them in an array.

Step 5: We apply the label for which distance is minimum.

Step 5.1: In case of a tie, we randomly label the class.

(Nearestdistances (1) ==nearestdistances (2) &&nearestdistances (1) <nearestdistances (3))

Step 6: Return the class label.

Step 7: Compute confusion matrix.

B. Program Code

```
clear all;
```

```
clc;
```

```
load heartdisease.mat;
```

```
k=input('Enter the number of nearest neighbors: ');
```

```
split=0;
```

```
count=0;
```

```
while(count~=1)
```

```
    numofobs=length(heartdiseasedata);
```

```
    rearrangement= randperm(numofobs);
```

```
newheartdiseasedata=heartdiseasedata(rearrangement,:);
```

```
newheartdiseaselabel=heartdiseaselabel(rearrangement);
```

```
    split = ceil(numofobs/2);
```

```
    count=count+1;
```

```
end
```

```
heartdiseasetrainingdata = newheartdiseasedata(1:split,:);
```

```
heartdiseasetraininglabel = newheartdiseaselabel(1:split);
```

```
heartdiseasetestdata = newheartdiseasedata(split+1:end,:);
```

```
originallabel = newheartdiseaselabel(split+1:end);
```

```
numoftestdata = size(heartdiseasetestdata,1);
```

```
numoftrainingdata = size(heartdiseasetrainingdata,1);
```

```
for sample=1:numoftestdata
```

```
    %Step 3: Computing euclidean distance for each test data
```

```

EuclideanDistance=
sum((repmat(heartdiseasetestdata(sample,:),numoftraining
data,1)-heartdiseasetrainingdata).^2,2);
[dist position] =
sort(euclideanDistance,'ascend');
Nearest neighbors=position(1:k);
Nearest distances=dist(1:k);
if(nearest distances(1)<nearest distances(2))

heartdiseasetestlabel(sample)=heartdiseasetraininglabel(p
osition(1));
end

if
(nearestdistances(1)==nearestdistances(2)&&nearestdista
nces(1)<nearestdistances(3))
index=randi(2,1,1);

heartdiseasetestlabel(sample)=heartdiseasetraininglabel(p
osition(index));
end

if
(nearestdistances(1)==nearestdistances(2)&&nearestdista
nces(2)==nearestdistances(3)&&nearestdistances(3)<near
estdistances(4))
index=randi(3,1,1);

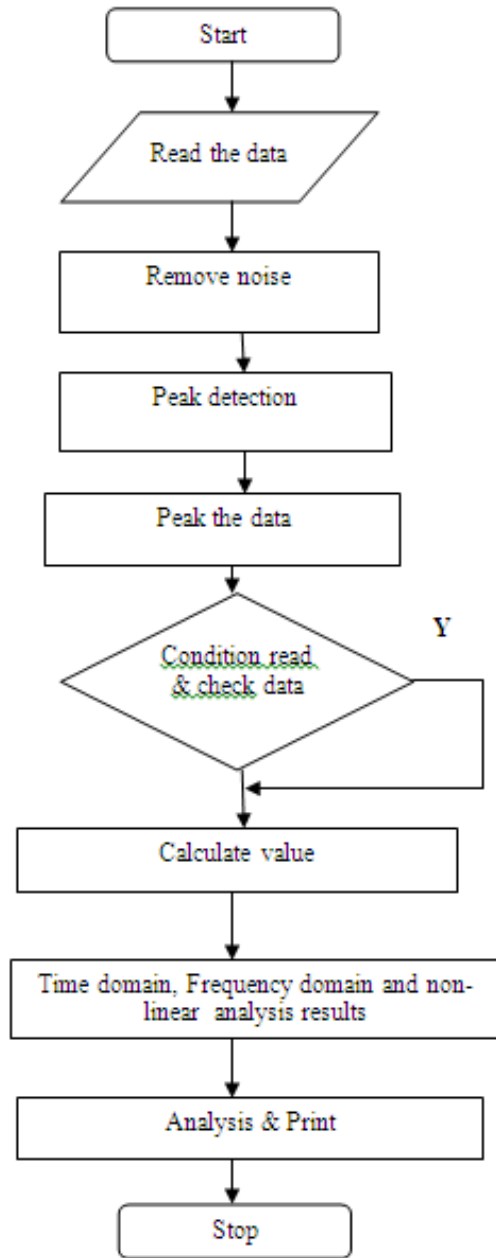
heartdiseasetestlabel(sample)=heartdiseasetraininglabel(p
osition(index));
end

if
(nearestdistances(1)==nearestdistances(2)&&nearestdista
nces(2)==nearestdistances(3)&&nearestdistances(3)==nea
restdistances(4))
index=randi(4,1,1);

heartdiseasetestlabel(sample)=heartdiseasetraininglabel(p
osition(index));
else
index=randi(k,1,1);
heartdiseasetestlabel(sample)=heartdiseasetraininglabel(p
osition(index));
end
end
originallabel=originallabel';
[confmat
order]=confusionmat(originallabel,heartdiseasetestlabel)
accuracy=(sum(diag(confmat)))/
sum(sum(confmat));
gscatter(heartdiseasedata(:,3),heartdiseasedata(:,7),heartdi
seaselabel));
end

```

C. FLOWCHART



VI. CONCLUSION

In this paper, we have analysed several classification techniques that are very useful in data mining for detecting the heart disease. From the above existing system we have analysed all types of heart disease diagnosis. We have classified several techniques that are useful in data mining for detecting heart disease. It has proven that classification based techniques contribute high effectiveness and obtain high accuracy compare than the previous methods. Future work it can experiment for another algorithm for a priori algorithm.

REFERENCES

- [1]. 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
- [2]. 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
- [3]. Shou-En Lu and Gloria L Beckles *et al*, "Evaluation of risk equations for prediction of short-term coronary heart disease events inpatients with long-standing type 2 diabetes: the Translating Research into Action for Diabetes", International Journal of BMC Endocrine Disorders, Vol.12, 2012
- [4]. Nazri Mohd Nawi and Rozaida Ghazali *et al*, "The Development of Improved Back-Propagation Neural Networks Algorithm for Predicting Patients with Heart Disease", In proceedings of the first international conference ICICA, Vol.6377, pp.317-324, 2010
- [5]. Jae-Hong Eom and Sung-Chun Kim, *et al*, "AptCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction", Journal of Expert Systems with Applications, Vol. 34, pp 2465-2479, 2008
- [6]. S. Paredes *et al*, "Long term cardiovascular risk models' combination", Journal of computer methods and programs in biomedicine, Vol.101, pp 231-242, 2011
- [7]. Swati Shilaskar *et al*, Feature selection for medical diagnosis: Evaluation for cardiovascular diseases", Journal of Expert System with Application, Vol.40, pp 4146-4153, 2013
- [8]. Raghunath Nambiar, Adhiraaj Sethi, Ruchie Bhardwaj, Rajesh Vargheese, " A Look at Challenges and Opportunities of Big Data Analytics in Healthcare", 2013 IEEE International Conference on Big Data.
- [9]. 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.
- [10]. Lamia Abed Noor Muhammed, " Using Data Mining technique to diagnosis heart disease", IEEE, International conference on statistics in science, Buiseness and Engineering, pp.1-3, 2012.
- [11]. M.Akhil Jabbar ,Dr.Priti Chandra, Dr.B.L Deekshatulu, " Heart Disease Prediction System Using Associative Classification and Genetic Algorithm", ICECIT, 2012.
- [12]. Vikas Chaurasia, Saurabh Pal, "Early Prediction of Heart Diseases Using Data Mining Techniques", Carib.j.SciTech, 2013, Vol.1, 208-217.