International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 5- 8
ISSN: 2278-2419

An Ill-identified Classification to Predict Cardiac Disease Using Data Clustering

C.Kiruthika[1], S.Nirmala Sugirtha Rajini[2]
[1]Research Scholar, Department of Computer Applications, Dr. M.G.R. Educational and Research Institute University, Chennai
[2] Professor, Department of Computer Applications, Dr. M.G.R. Educational and Research Institute University, Chennai
E-mail: kiruthisenan@yahoo.co.in

Abstract-The health care industry contains large amount of health care data with hidden information. This information is useful for making effective decision. For getting appropriate result from the hidden information computer based data mining techniques are used. Previously Neural Network (NN) is widely used for predicting cardiac disease. In this paper, a Cardiac Disease Prediction System (CDPS) is developed by using data clustering. The CDPS system uses 15 parameters to predict the disease, for example BP, Obesity, cholesterol, etc. This 15 attributes like sex, age, weight are given as the input. In this paper by using the patient's medical record, an ill-defined classification is used at the early stage of the patient to diagnose the cardiac disease. Based on the result the patients are advised to keep the sensor to predict them.

I. INTRODUCTION

Cardiac disease is a disease that affects on the operation of heart. There are number of factors which increases risk of Cardiac disease. They are

- Family history of Cardiac disease
- Smoking
- Cholesterol
- High blood pressure
- Obesity
- Lack of physical exercise

Nowadays, in the world, Cardiac disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the cardiac diseases. WHO estimated by 2030, almost 23.6 million people will die due to cardiac disease. [1]. Predication should be done to reduce risk of cardiac disease. Diagnosis is usually based on signs, symptoms and physical examination of a patient. Almost all the doctors are predicting heart disease by learning and experience. The diagnosis of disease is a difficult and tedious task in medical field. Predicting cardiac disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. Healthcare industry today generates large amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc. Only human intelligence alone is not enough for proper diagnosis. A number of difficulties will arrive during diagnosis, such as less accurate results, less experience, time dependent performance, knowledge up gradation is difficult. The data mining techniques are used to pre-process the information from the patient's medical record and data clustering is used to classify the attributes.

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System with K means Clustering algorithm using historical heart disease databases to make intelligent clinical decisions which traditional decision support systems cannot. Several computer aided diagnosis methodologies have been proposed in the literature for the diagnosis of heart attacks. An intelligent heart disease prediction system built with the aid of data mining technique like decision trees, naïvebayes and neural network was proposed. The result illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives.[2]. It facilitated the establishment of vital knowledge e.g. Patterns connected with heart disease. It subsist well being web-based, user- friendly, scalable, reliable and expandable. K-means clustering algorithm is introduced to extract the data appropriate to heart attack from the warehouse. In addition the pattern vital to heart attack were selected on basis of the computer significant weight age. Association rules are used to improve heart disease prediction. Association rules were applied on a real data set contacting medical records of patient with heart disease. [3]. after diagnosing, the doctors advised the patients to keep the sensor device with them to check the body condition monthly once to maintain the normal level. If the condition is abnormal the alert message is send to the doctors and the care takers to take immediate action.

II. MATERIALS AND METHODS

A. Data Base

For this proposed work the data's are collected from various numbers of patients and it will be stored in the data base. The data base contains so many attributes, for experiment only 15 attributes are used. The attributes are such as age, weight, BP, pulse rate, sugar, and cholestral.etc.

B. Data Pre-Processing

Data pre-processing is used to extract the relevant data. This is done because if the irrelevant information present in the data set the original data will not be directly used for the prediction method. In data-pre-processing the raw-data will be cleaned, analysed and transformed. Cleaning is done to remove the

International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 5- 8
ISSN: 2278-2419

duplicate records. There are two phases are used to analyse whether the person is having cardiac disease are not. If the phase is 0-no heart disease and 1-presence of heart disease.

Table 1: Attributes of Heart Disease Data Sets

1. Age in year
2. Sex (value 1: Male;
value 0 :Female)
3. Chest pain type
(value 1:typical type 1 angina;
value 2 : typical type angina;
value 3: non-angina pain;
value 4 : asymptomatic)
4. Resting blood pressure
(mm Hg on admission to the hospital)
5. Serum Cholesterol in mg/dl
6. Fasting blood sugar
(value 1: > 120 mg/dl;
value 0 :< 120 mg/dl)
7. Resting Electrocardiographic results
(values 0:normal;
value1: 1 having ST-T wave abnormality;
value 2:showing probable or
definite left ventricular hypertrophy)
8. Maximum heart rate achieved
9. Exercise induced angina
(value 1:yes; value 0 : no)
10. Old peak = ST
(depression induced by exercise relative to rest)
11. The slope of the peak exercise ST segment
(value 1: unsloping;
value 2 : flat;
value 3 :down sloping)
12.Number of major vessels colour
by fluoroscopy(value 0-3)
13. Thal(value3 = normal;
value 6 = fixed defect;
value 7 = reversible defect)
14. Smoking habit – number of years
15. Body weight(obesity)

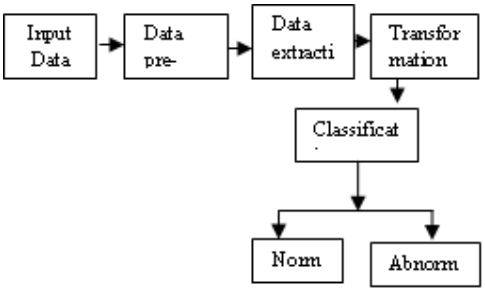

Fig.1: Block diagram for data pre-processing

C. Data Clustering

The data clustering is a process of portioning or grouping a set of data objects into a number of clusters such that similar patterns are assigned to one clusterThe methods that are used to diagnose the cardiac disease previously cannot provide a clear explanation for the decision to examine the risk factors for cardiac disease. The fuzzy logic with data clustering technique is used to assess the risk-level of cardiac patient. But it is an assumption based technique by using the fuzzy c means the efficiency of the output is an assumption value. By using the k-means algorithm the efficiency of the output will be quiet higher.

III. CLUSTERING ALGORITHM

A. The *k*-means algorithm

The k-means algorithm is a simple iterative method to partition a given dataset into a user specified number of clusters, *k*. This algorithm has been discovered by several researchers across different disciplines.

The steps included in the algorithm

Step 1: Data are assigned in two data sets. One is the user data set and another one is the Pre-defined data set

Step 2: Analyse the data from the two data sets by partitioning method

Step 3: Match the user data with the pre-defined data according to the age group and set the symptom

Step 4: if the symptoms are normal no alert

Step 5: if the symptoms are abnormal message will send ,

Step 6: end

B. Evaluation

The performance of the algorithm will be evaluated by taking a sample data attributes from the data set.

Based on the inputs the k-means algorithm is used to diagnose the cardiac with more accuracy.

International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 5- 8
ISSN: 2278-2419

Table 2.Sample input data

Age	Sex	Chest Pain	Res TBP	Cholesterol	Blood sugar	ECG	Heart Rate	Angina	Old peak	ST Slope	Vessel	thal
60	1.0	4.0	130.0	203.0	0.0	2.0	132.0	1.0	2.4	2.0	2.0	7.0
52	1.0	1.0	120.0	193.0	0.0	2.0	162.0	0.0	1.9	2.0	0.0	7.0
72	0.0	4.0	112.0	149.0	0.0	0.0	125.0	0.0	1.6	2.0	0.0	3.0
35	1.0	4.0	126.0	282.0	0.0	0.0	156.0	1.0	0.0	1.0	0.0	7.0
43	0.0	2.0	126.0	306.0	0.0	2.0	163.0	0.0	0.0	1.0	0.0	3.0

C. Performance evaluation with different methods

The performance of the proposed algorithm is compared and evaluated.

Table 3. Performance evaluation

Algorithm	**Accuracy**
K-Means	80%
FCM	78%

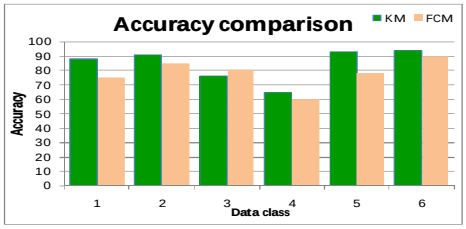

Fig.2. Comparison of K-means and FCM algorithm

The above graph shows that the k means having the more accuracy than the FCM.

IV. CARDIAC PREDICTION USING SENSOR DEVICE

An Android based health care smart phone is used as prediction device to monitor the cardiac disease. This device sets a medical sensor power to detect the health tests which combined With related applications and cloud based services. It will be work by keeping a finger to the sensor Device that measures track and analyse the medical measurements to take immediate action. This test includes ECG, BP, pulse rate, cholesterol, etc. This collected data is automatically saved to the remote server and it can be retrieved from cloud in anytime and anywhere .This data will be shared with doctors, family members and others to take immediate action by using mobile phones.

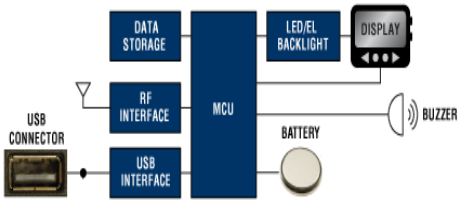

Fig.3.Main diagram

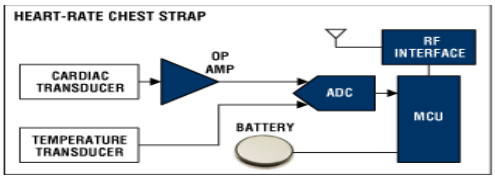

Fig.3.2.Working of sensor device

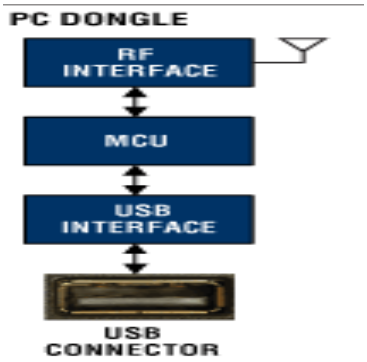

Fig.3.2.server side working

The reading will taken by using a kit and the data are stored ,that stored data's are send to the connector by using RF interface sender, in other side the RF interface receiver receives the data and display the result.

A. Medical specifications

Some of the medical specifications are given below

ECG	ECG dynamic range : 0.15 – 5 mV Frequency response : 0.5 – 40 Hz Sampling : 12 bit, 250 samples/sec Test time : 35 secs
Pulse rate	Range : 30 – 250 bpm QRS detection sensitivity > 98% QRS detection predictability > 98% Result time: 20 secs

International Journal of Data Mining Techniques and Applications
Volume: 03 Issue: 01 June 2014, Page No. 5- 8
ISSN: 2278-2419

V. CONCLUSION

In this paper, the K-means clustering algorithm is used for diagnose the risk of cardiac disease. The fuzzy logic is assumption based technique by using that efficiency of the output is an assumption value. So that we are using the k-means algorithm the efficiency of the output will be quiet higher. Data pre-processing is done to remove all the duplicate records. The k –means clustering is used to group data objects. A sensor based smart phone is used to detect the cardiac disease and take immediate action. In future more medical specifications are added to diagnose the cardiac disease effectively to reduce the risk factor.

REFERENCES

[1] World Health Organization, "Global status report on non communicable diseases", 2010.
Vanisree K, Jyothi Singaraju, Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks, International Journal of Computer Applications (0975 – 8887) Volume 19– No.6, April 2011.
[2] Carolas Ordonez, "Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction" IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, April 2006.
[3] Sellappan Palaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques" IEEE, Pp. 108-115; 2008.
[4] K.Srinivas, B.KavithaRani and Dr.A. Govrdhan "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" K. Srinivas et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 250-255, 2010
[5] S.Vijayarani, M.Divya, an Efficient Algorithm for Generating Classification Rules, ISSN: 0976-8491 (Online) | ISSN: 2229-4333 (Print), IJCST Vol. 2, Issue 4, 2011
[6] N. Deepika and K. Chandra shekar, "Association rule for classification of Heart Attack Patients", International Journal of Advanced Engineering Science and Technologies, Vol. 11, No. 2, pp. 253 – 257, 2011.
[7] K. Srinivas, B. Kavitha Rani and Dr. A. Govrdhan, "Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering, Vol. 02, No. 02, pp. 250 - 255, 2011.
[8] Asha Rajkumar and B. Sophia Reena, "Diagnosis Of Heart Disease Using Data mining Algorithm" , Global Journal of Computer Science and Technology, Vol. 10, No. 10, pp.
[9] 38 - 43, 2010
[10] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Science, Vol. 15, pp. 157 - 160, 2007.
[11] Mrs.G.Subbalakshmi, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering.