Predicting Heart Attack using Fuzzy C Means Clustering Algorithm

Dr. G. Rasitha Banu
MCA., M.Phil., Ph.D.,
Assistant Professor,Dept of HIM&HIT,Jazan University,
Jazan, Saudi Arabia.

J.H.BOUSAL JAMALA
MCA.,M.Phil.,
Mother Teresa Women's University Research Scholar, Chennai, India

Abstract - Cardiovascular disease remains the biggest cause of deaths worldwide. The percentage of premature death from this disease ranges from 4% in high income countries and 42 % in low income countries. This shows the importance of predicting heart disease at the early stage. In this paper, a new unsupervised classification system is adopted for heart attack prediction at the early stage using the patient's medical record. The information in the patient record are preprocessed initially using data mining techniques and then the attributes are classified using a Fuzzy C means classifier. In the classification stage 13 attributes are given as input to the Fuzzy C Means (FCM) classifier to determine the risk of heart attack. FCM is an unsupervised clustering algorithm, which allows one piece of data to belong to two or more clusters. The proposed system will provide an aid for the physicians to diagnosis the disease in a more efficient way. The efficiency of the classifier is tested using the records collected from 270 patients, which gives a classification accuracy of 92%. The result shows that the proposed clustering algorithm can predict the likelihood of patients getting a heart attack in a more efficient and cost effective way than the other well known algorithms.

KEY TERMS- FUZZY C means clustering algorithm, cardiovascular disease, classification,pre-processing.

I. INTRODUCTION

Cardiovascular disease is a kind of serious health imperiling and frequent happening disease. The world health organization has estimated that 12 million deaths occur worldwide, every year due to the cardiovascular disease. Advances in the field of medicine over the past few decades enabled the identification of risk factors that may contribute toward the cardiovascular diseases.. The most common cause of heart disease is narrowing or blockage of the coronary arteries, the blood vessels that supply blood to the heart itself. This is called coronary artery disease and happens slowly over time. It's the major reason people have heart attacks. A blockage that is not treated within a few hours causes the affected heart muscle to die. During about 30 percent of all heart attacks, the patient experiences no symptoms. However, unmistakable signs of the attack remain in the bloodstream for days. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource. A system for automated medical diagnosis would enhance medical care and reduce costs. With so many factors to analyze for a diagnosis of heart attacks, physicians generally make a diagnosis by evaluating a patients current test results. Previous diagnoses made on other patients with the same results are also examined by physicians. These complex procedures are not easy. Therefore, a physician must be experienced and highly skilled to diagnose heart attacks in a patient. Thus the effort to utilize knowledge and experience of numerous specialists and clinical screening data of patients collected in databases to facilitate the diagnosis process is considered a valuable system that is the integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

II. METHODOLOGY

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System with Fuzzy C Means Clustering algorithm using historical heart disease databases to make intelligent clinical decisions which traditional decision support systems. Fuzzy c-means clustering algorithms for predicting the heart attack diseases. In these fuzzy clustering algorithms, the membership degree is associated to the values of the features in the clusters for the cluster centers instead of being associated to the patterns in each cluster.

2.1 Data Mining in the Heart Disease Prediction.

Data Mining is defined as extracting the information from the huge set of data. In other words we can say that data mining is mining the knowledge from data. Data mining in health care has become increasingly popular because in can improved our patient care by early detecting of disease supports helping care providers for treatment programs and reduces the cost of health care.

Prediction Method

Accuracy of classifier refers to the ability of classifier. It predict the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

Proposed system:

To overcome this problem using Fuzzy C Means Clustering algorithm. For finding the risk of heart attack of a patient using the profiles collected from the patients. With the proper adaptation of FCM classifies, the method can thus evolve an optimum number of clusters and finds the abnormal and normal cases efficiently. Initially pre-processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not.

Advantage:

This method used to diagnosis in early period.

Records acquirement:

In this module, heart attack dataset is obtained from UCI centre for machine knowledge and intelligent system. The data have been collected from 270 patients. The data base contains 73 attributes but we use 13 attributes. That attributes are age, sex, chest pain, Rest BP, cholesterol, sugar, ECG, Max heart range, Angina, Old peak, Sis lope, vessels, thal.

Pre-processing:

Data pre-processing is an important step in the data mining process. Data-collection methods are often insecurely controlled, resulting in out of range values, impractical data combinations (e.g., Sex: Male, Pregnant: Yes),missing values, etc. In this method, data's are extracted from UCI centre dataset And this module remove to all duplicate records and missing records.

Description of 13 input attributes in proposed system

Serial no	Attribute Name	Attribute value	Attribute Condition
1	Sex	Gender	Male/Female
2	Smoking	Yes No	Abnormal Normal
3	Sugar	>120	Abnormal
		<120	Normal
		NULL	NULL

4	Cholesterol	0-160	Abnormal
		160-410	Normal
		120-250	Abnormal
		NULL	NULL
5	Rest BP	0-90	Abnormal
		90-192	Normal
		192-300	Abnormal
		NULL	NULL
6	ECG	Normal	Normal
		Having ST-T way abnormality	Normal
		Showing probable or definite left ventricular hypertrophy	Abnormal
		NULL	NULL
7	THAL	Normal	Normal
		Fixed	Abnormal
		Reversible Defect	Abnormal
		NULL	NULL
8	Chest Pain	Typical Type 1 Angina	Abnormal
		Typical Type Angina	Abnormal
		Non Angina	Normal
		Asymptomatic	Abnormal
		NULL	NULL
9	Angina	Yes	Abnormal
		No	Normal
		NULL	NULL
10	St Slope	Up Sloping	Abnormal
		Flat	Normal
		Down Sloping	Abnormal
		NULL	NULL
11	Old Peak	Continous	Abnormal
		Non-Continous	Normal
		NULL	NULL
12	Vessels	0	Normal
		1	Abnormal
		2	Abnormal
		NULL	NULL
13	Heart Rate	<72	Abnormal
		72 >72	Normal Abnormal
		NULL	NULL

Fuzzy C Means (FCM) classifier

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely tojust one cluster. Thus, points on the edge of a cluster, may be *in the cluster* to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available.

Any point x has a set of coefficients giving the degree of being in the kth cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

The degree of belonging, $w_k(x)$, is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center.

Assign randomly to each point coefficients for being in the clusters.

- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than ε, the given sensitivity threshold) :
    - Compute the centroid for each cluster, using the formula above.
    - For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well, but has the same problems as *k*-means; the minimum is a local minimum, and the results depend on the initial choice of weights.

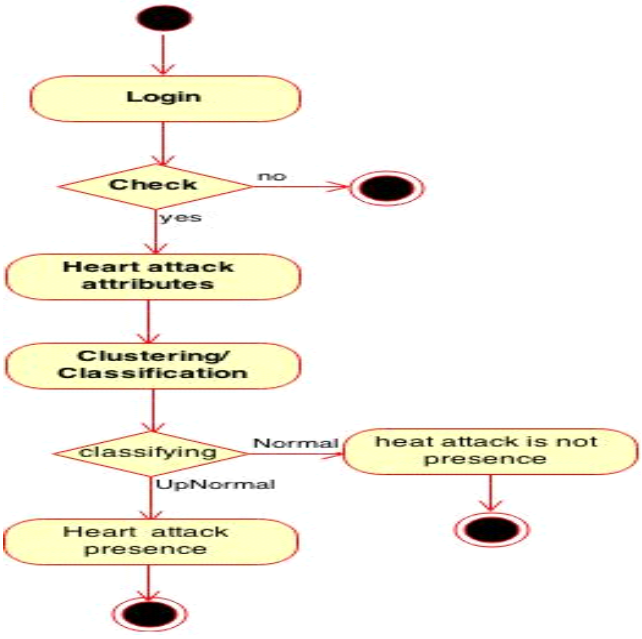
Using a mixture of Gaussians along with the EM Algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes.

III. Heart attack prediction using FCM

Fuzzy c means using clustering for assessing a risk level of the heart patient. The proposed system Fuzzy c means clinical decision support system has been implemented using VISUAL BASIC .NET. The .NET Framework is a new computing platform that simplifies application development in the highly distributed environment of the Internet.Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly.

The designed algorithm for clinical decision support system shown in figure contains" input and output levels }, where the two values are predicted as abnormal and normal. The inputs are related to the 13 attributes and output is related to the risk level presence or absence. In the proposed system, the biomarkers for cardiovascular diseases describe cholesterol level, smoking status, and Angina Heart rate, ECG that are mainly used to predict the risk level of heart patients. For better prediction of risk level, we make use of fuzzy C means clustering algorithm If condition to check Present or Not. Initially the fuzziness is set to a value between assign values in dataset.

CLUSTERING/CLASSIFICATION DIAGRAM



Clustering Data

Clustering is a process of partitioning or grouping a set of data objects into a number of clusters such that similar patterns are assigned to one cluster. It's main task of data mining. We use Fuzzy C-means Clustering. Data Clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. If items (data) in same class mean no presence of heart attack otherwise heart attack is presence

IV. CONCLUSION

In this paper, we have proposed an FCM clustering algorithm for finding the risk of heart attack of a patient using the profiles collected from the patients. With the proper adaptation of FCM classifies, the method can thus evolve an optimum number of clusters and finds the abnormal and normal cases efficiently. Initially pre-processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not. The results of classification experiment, preformed over a data sets obtained from 270 patients, shows that the classifier has achieved better accuracy than most of the existing algorithms. The performance of the proposed FCM is proved to be a well known approach in terms of accuracy**.**

REFERENCS

[1] Sellappan Palaniappan,Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques" IEEE Conference, 2008,pp 108-115.

[2] Carolas Ordonez "Assosiation 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] Shanthakumar B.Patil,Y.S,Kumaraswamy"Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". European Journal of Scientific Research Vol. 31, No. 04, 2009, 642-656 .

[4] Palaniappan, S., Awang, R., 2008. Intelligent heart disease prediction system using data mining techniques. International Journal of Computer Science and Network Security 8 (8), 108–115.

[5] Shanthakumar B.Patil,Y.S,Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificia Neural Network". European Journal of Scientific Research Vol. 31, No. 04, 2009, 642-656.

[6] Yang, M.-S., Hwang, P.-Y., Chen, D.-H., 2004. Fuzzy clustering algorithms for mixed feature variables. Fuzzy Sets Systems 141,301 – 317.[9].

[7] Ross, Timothy J., 2004. Fuzzy Logic with Engineering Applications, John Wiley & Sons, Second Edition.[10].

[8] Lanhai L. Comparison of conventional and fuzzy land classification and evaluation techniques in Oxford shire England. Int Agric Eng J 1998;7:1

[9] Asha Rajkumar, G.Sophia Reena, Diagnosis Of Heart Disease Using Datamining Algorithm, Global Journal of Computer Science and Technology 38 Vol. 10 Issue 10 Ver. 1.0 September 2010.