

Heart Attack Prediction System Using Fuzzy C Means Classifier

R.Chitra¹, Dr.V.Seenivasagam²

¹Associate Professor, Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kanyakumari District, India.

²Dr.V.Seenivasagam, Professor, Department of Computer Science and Engineering, National Engineering College, Kovilpatti, 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.

Keywords: Cardiovascular disease, Clustering, Fuzzy C Means, Heart attack prediction,

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 patient's 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.

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 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, naïve bayes and neural network was proposed by sellappan palaniappan et al.[4]. The result illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives. It facilitated the establishment of vital knowledge eg. Patterns connected with heart disease. It subsist well being web-based, user- friendly, scalable, reliable and expandable. Markos G. Tsipouras et al. [6] presented a fuzzy rule based decision support system for the diagnosis of coronary artery disease. Fuzzy model was used to optimize the parameters. Shanthakumar et al. [2] introduced K-means clustering algorithm 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 weightage. Carlos Ordóñez [3] used association rules to improve heart disease prediction. Association rules were plied on a real data set contacting medical records of patient with heart disease and the

risk factors were identified. K.Srinivas et al [5] applied data mining techniques to predict heart attack. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart attack. Based on the calculated significant weightage the frequent pattern having value greater than a predefined threshold were chosen for the valuable prediction of heart attack.

Bayesian networks are a probability-based inference model, increasingly used in the medical domain as a method of knowledge representation for reasoning under uncertainty for a wide range of applications, including disease diagnosis [11]. Neural networks have also been applied to the medical and diagnosis fields, most actively as the basis of a soft computing method to render the complex and fuzzy cognitive process of diagnosis. Recently, multilayer perceptrons (MLP), one of the most popular ANN models, has been applied to build an MDSS for five different heart diseases diagnoses (Yan et al., 2006). The three-layered MLP with 40 categorical input variables and modified learning method achieved a diagnosis accuracy of over 90% [12]. For the diagnosis of coronary artery disease (CAD), Tsipouras et al. (2008) have proposed a fuzzy rule-based decision support system (DSS). Using a four stage methodology: (1) induction of a decision tree from the data; (2) extraction of a set of rules from the decision tree, in disjunctive normal form and formulation of a crisp model; (3) transformation of the crisp set of rules into a fuzzy model; and (4) optimization of the parameters of the fuzzy model, the system was automatically generated from an initial annotated dataset [13]. Setiawan et al. (2009) [16] have developed a fuzzy decision support system for the diagnosis of heart disease. By using rules extraction method based on rough set theory, the knowledge base of fuzzy decision support system was taken. Based on information from the discretization of numerical attributes, the rules then were selected and fuzzified. Using the information from the support of extracted rules, fuzzy rules weights were proposed.

A prototype intelligent heart disease prediction system (IHDPS) has been developed by Palaniappan and Awang (2008) using data mining techniques, namely, decision trees, naive Bayes and neural network. Results exposed that in realizing the objectives of the defined mining goals, each technique has its exclusive strength. IHDPS can respond to complex “what if” queries whereas the traditional decision support system is unable to answer. It can foretell the possibility of patients getting a heart disease, using medical profiles such as age, sex, blood pressure and blood sugar. It facilitates significant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be recognized. IHDPS is Web-based, user-friendly, scalable, reliable and expandable [14]. Engin Avci and Ibrahim Turkoglu study an intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases [15]. Clustering-analysis-based methods can provide a nonparametric, unsupervised approach to the analysis of data. Dan Li et al. [7] adopted the idea of nearest-neighbor rule, a novel fuzzy c-means algorithm for incomplete data based on nearest-neighbor intervals is proposed. Medical profiles are subjected to the uncertainty of missing attributes, these attributes are represented by nearest-neighbor intervals [7] based on the nearest neighbor information. Secondly, the clustering problem can be thus viewed as clustering for interval-valued data, which will result in interval cluster prototypes rather than point prototypes. In the last two decades, the use of artificial intelligence tools has become widely accepted in medical applications to support patient diagnosis more effectively. Especially, the application of various machine learning approaches such as decision trees (DTs), artificial neural networks (ANNs), Bayesian networks (BNs), and support vector machines (SVMs) have been actively tried for meeting clinical support requirements. Consequently, medical diagnosis systems using different machine learning approaches have shown great potential, and many machine learning methods have been tried for a wide variety of clinical and medical applications.

In the actual clustering application, such as data analysis and image segmentation, some prior knowledge is always incidental, but the number of that is generally little. If the supervised method was adopted, the learning effect would be poor due to fewer training samples. Thus, the unsupervised algorithm is a better choice to solve this problem. Yang et al. (2004) [8] presented fuzzy clustering algorithms for mixed features of symbolic and fuzzy data. This paper introduces adaptive 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. The detailed steps involved in the proposed system are explained in Fig 1.

II. Materials and Methods

Database Description and Pre-Processing

Heart attack dataset is obtained from UCI (University of California, Irvine C.A) centre for machine learning and intelligent systems. The data have been collected from 270 patients are used for proposed work. This database contains 76 attributes, for the experimental purpose 13 attributes are used. The attributes such as age, sex, chest pain type, resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar, resting electrocardiographic results, and maximum heart rate achieved, exercise induced angina, ST depression, and

slope of the peak exercise ST segment, number of major vessels, thal and diagnosis of heart disease are used. The digitized data has 150 normal and 120 abnormal cases.

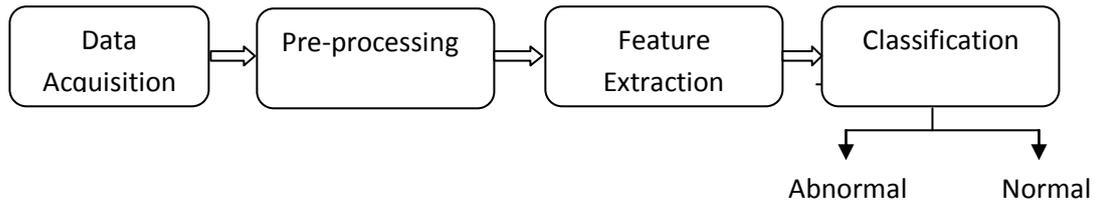


Figure 1. Block Diagram of the Proposed Intelligent System

Data preprocessing is done to extract relevant data and then these data should be converted into the format necessary for the prediction of risk level. Due to the irrelevant information in the heart disease datasets, the original raw data cannot be directly used in the prediction procedure, hence in data preprocessing phase, raw data need to be cleaned, analyzed and transformed for further steps. Cleaning and filtering of the data set is done to remove duplicate records, normalize the values, accounting for missing data and removing irrelevant data items. There are two output classes for the diagnosis of heart attack. In the selected dataset, class 0 specifies the no presence of heart attack and class 1 specifies the presence of heart disease. Normally, direct support clinical decision-making is the intention behind the design of a clinical decision support system and it presents patient-specific assessments or recommendations produced using the characteristics of individual patients to clinicians for consideration [1].

In recent years, clinical decision support system based on computer aided diagnosis methodologies have been proposed in the literature by which evaluating the data obtained by some of the methods or other sources (i.e., laboratory examinations, demographic and/or history data, etc.) from a computer-based application leads to a computer-aided diagnosis. The data analysis methods used in most of the proposed methods cannot provide clear and direct explanation for the decisions made to examine the risk factors for cardiovascular diseases as they are based on neural networks. Hence, a method based on easily obtained features capable of calculating the risk level of computer-aided diagnosis and providing explanation for the decisions made would be of immense clinical value So, the soft computing technique in particular the fuzzy logic technique could be used for assessing the risk level of heart patients in developing the clinical decision support system of heart disease diagnosis.

Table 1. Description of 13 Input Attributes used in the Proposed System

Sl.No	Attribute Name	Attribute Description	Attribute Values
1.	AGE	Age in years	25-75 years
2.	SEX	Male/Female	value 1: Male; value 0 : Female
3.	CHESTPAIN	Chest Pain Type	value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic
4.	RESTBP	resting blood pressure	90-192
5.	CHOLESTEROL	serum cholestorol in mg/dl	160-410
6.	BLOODSUGAR	fasting blood sugar > 120 mg/dl	value 1: > 120 mg/dl; value 0: < 120 mg/dl
7.	ECG	resting electrocardiographic results	value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy
8.	MAXHEARTRATE	maximum heart rate achieved	71-202
9.	ANGINA	exercise induced angina	value 1: yes; value 0: no
10.	OLDPEAK	ST depression induced by exercise relative to rest	Continuous
11.	STSLOPE	the slope of the peak exercise ST segment	value 1: unsloping; value 2: flat; value 3: downsloping)
12.	VESSELS	number of major vessels (0-3) colored by flourosopy	value 0 – 3
13.	THAL:	thalac	value 3: normal; value 6: fixed defect; value 7: reversible defect

Fuzzy C-means Clustering

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. Fuzzy clustering is a class of algorithm in cluster analysis wherein the allocation of data points to clusters is not ‘hard’ but ‘fuzzy’ in the same sense as fuzzy logic. Fuzzy logic is a multi-valued logic derived from fuzzy set theory, proposed by Lofti Zadeh to deal with reasoning that is approximate rather than precise [9].

Fuzzy C means clustering (FCM) was developed by Dunn in 1973 and improved by Bezdek in 1981 is frequently used in pattern recognition. Fuzzy c-means algorithm permits a data point’s belonging to one or more clusters utilizing membership value concept. Therefore, elements of partition matrix consist of membership values varying within the interval [0, 1], and a data point can partially belong to a cluster. Basically, fuzzy c-means algorithm calculates fuzzy partition matrix to group some of data points into c clusters. Therefore, the aim of this algorithm is to cluster centers (centroids) that minimize dissimilarity function (J_m), which is given by [10]. FCM is an overlapping data clustering technique wherein each data point, $X = \{x_1, x_2, \dots, x_k\}$ belongs to a cluster i to some degree specified by a membership grade, u_{ij} .

$$J_m = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty \quad (1)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d -dimensional measured data, c_j is the d -dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left\{ \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right\}^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (2)$$

$$\max_{ij} \{u_{ij}^{(k+1)} - u_{ij}^{(k)}\} < \varepsilon$$

This iteration will stop when, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m . The algorithm is composed of the following steps:

Step1 : Choose m, c and ε , where $\varepsilon > 0$ is a small positive constant; then initialize the partition matrix $U^{(0)}, U = [u_{ij}]$

Step2: When the iteration index is k ($k = 1, 2, \dots$), calculate the matrix of cluster prototypes $C^{(k)}$ using c_j and $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

Step3: Update the values of $U^{(k)}, U^{(k+1)}$.

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left\{ \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right\}^{\frac{2}{m-1}}} \quad (4)$$

Step 4: If $\forall i; j : \max_{ij} \{u_{ij}^{(k+1)} - u_{ij}^{(k)}\} < \varepsilon$ then stop and get the partition matrix U and the matrix of cluster prototypes C

Otherwise set $k = k + 1$ and return to Step 2.

III. Heart Attack Prediction using FCM

This section describes the developing of clinical decision support system using fuzzy C means clustering for assessing the risk level of the heart patient. The proposed fuzzy C means clinical decision support system has been implemented using MATLAB 7. The designed algorithm for clinical decision support system shown in figure 2 contains ‘ X_i ’ input vector of the form $X_i = \{x_1, x_2, \dots, x_{13}\}$, and two output levels Y_i of the form $Y_i = \{y_1, y_2\}$, 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 described in the literature are age, sex, total cholesterol level age, smoking status, and hypertension 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. Initially the fuzziness is set to a value between 0 and 1. The

output data set is mapped to three clusters. In the proposed work patient predicted as heart attack comes under cluster 1 and patents predicted without heart attack comes under cluster 0 and cluster 2 contains mispredicted values. Euclidean distances (EUCDIS) between cluster centers in the three-dimensional space are determined to check the relevance of the clusters. We expect that the spacing between the cluster centers in the three-dimensional space should reflect the spacing in the higher dimensional space.

$$EUCDIS(0,2) = 0.75$$

$$EUCDIS(0,1) = 0.68$$

In the 3-dimensional space the Euclidean distances of cluster 0 and cluster 2 is greater than the distance between cluster 0 and cluster 1. This shows that there is a similarity between the cluster 0 and cluster 1, but these groups are considered to be misclassified clusters.

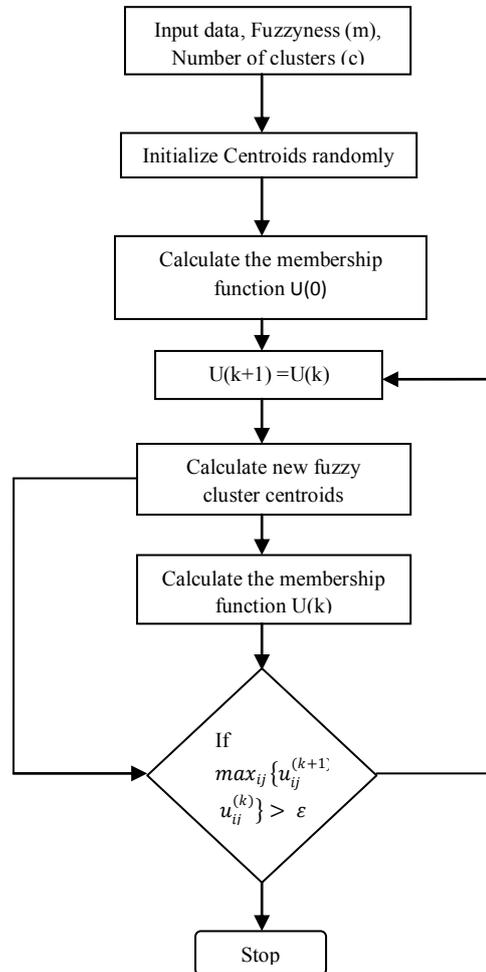


Figure 2. Fuzzy C means Clustering Process

IV. Performance Evaluation and Discussion

The performance of FCM is evaluated by statistical measures like sensitivity, specificity and accuracy to illustrate the medical diagnostic test. These metrics also enumerate how the test was good and consistent. Sensitivity evaluates the diagnostic test correctly at detecting a positive disease. Specificity measures how the proportion of patients without disease can be correctly ruled out. The performance of the proposed heart attack prediction system is also analyzed by calculating the sensitivity and specificity. The usefulness of diagnostic tests, that is their ability to detect a person with disease or exclude a person without disease, is usually described by terms such as sensitivity, specificity, positive predictive value and negative predictive value. The basic idea of heart attack interpretation is to calculate the probability a patient has a disease under consideration given a certain test result. In other words, a perfect test is never positive in a patient who is disease free and is never negative in a patient who is in fact diseased. Most clinical tests fall short of this ideal. The words "sensitivity" and "specificity" have their origins in screening tests for diseases.

When a single test is performed, the person may in fact have the disease or the person may be disease free. The test result may be positive, indicating the presence of disease, or the test result may be negative,

indicating the absence of the disease. The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease.

$$sensitivity = \frac{TP}{TP+FN} \quad (5)$$

Where TP is True Positive i.e the patient has the disease and the test is positive and FN is false negative i.e the patient has the disease but the test is negative

A high sensitivity is clearly important where the test is used to identify a serious but treatable disease. The specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease.

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

Where TN true negative i.e is the patient does not have the disease and the test is negative, FP is false positive i.e the patient does not have the disease but the test is positive. Therefore, a test with 100% specificity correctly identifies all patients without the disease. A test with 80% specificity correctly reports 80% of patients without the disease as test negative (true negatives) but 20% patients without the disease are incorrectly identified as test positive (false positives).

Accuracy measures correctly figured out the diagnostic test by eliminating a given condition and it is defined as

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (7)$$

The two modeling tasks were considered:

1. The evaluation criteria is the measurement of how many of the actual patients with abnormality (possibility of heart attack) are identified while limiting the number of false alarms (False Positive rate) to a range between 0.2 and 0.3 per record.
2. Designing a model, so that there are no false negatives.

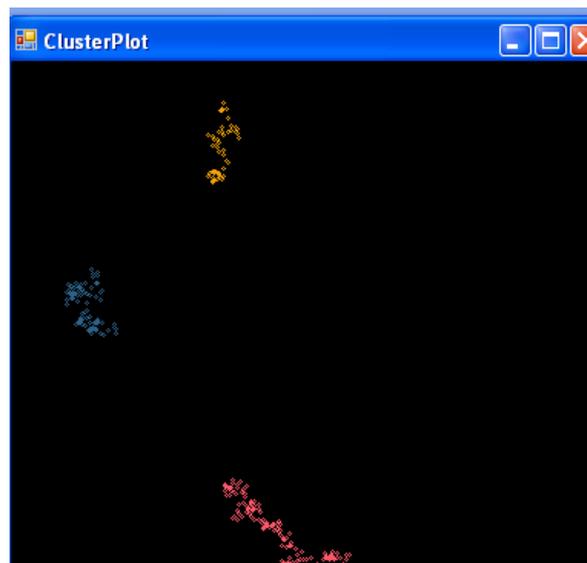


Figure 3. Cluster Distribution of Heart Attack Dataset with 3 Cluster Centers

A comparative analysis has been made with a number of model classes including Artificial Neural Network (ANN), K-Means Clustering (KM) and Fuzzy C Means clustering (FCM). Figure 3 shows the clustering process by finding the Euclidean distances and grouping into one of the 3 clusters. The cluster of the dataset of abnormal data (cluster 0) shown at the bottom, normal data (cluster 1) shown in the middle and data that does not belong to both (cluster 2) shown at the top.

Table 2. Sample Input attribute values for the Dataset used for the Prediction System

AGE	SEX	CHEST PAIN	RES TBP	CHOLESTEROL	BLOOD SUGAR	ECG	MAX HEART RATE	ANGINA	OLDP EAK	ST SLOPE	VESSELS	THAL
60.0	1.0	4.0	130.0	206.0	0.0	2.0	132.0	1.0	2.4	2.0	2.0	7.0
56.0	1.0	1.0	120.0	193.0	0.0	2.0	162.0	0.0	1.9	2.0	0.0	7.0
71.0	0.0	4.0	112.0	149.0	0.0	0.0	125.0	0.0	1.6	2.0	0.0	3.0
58.0	0.0	1.0	150.0	283.0	1.0	2.0	162.0	0.0	1.0	1.0	0.0	3.0
35.0	1.0	4.0	126.0	282.0	0.0	2.0	156.0	1.0	0.0	1.0	0.0	7.0
55.0	0.0	4.0	180.0	327.0	0.0	1.0	117.0	1.0	3.4	2.0	0.0	3.0
48.0	1.0	4.0	130.0	256.0	1.0	2.0	150.0	1.0	0.0	1.0	2.0	7.0
44.0	1.0	4.0	110.0	197.0	0.0	2.0	177.0	0.0	0.0	1.0	1.0	3.0
63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0
54.0	0.0	3.0	135.0	304.0	1.0	0.0	170.0	0.0	0.0	1.0	0.0	3.0
41.0	0.0	2.0	126.0	306.0	0.0	0.0	163.0	0.0	0.0	1.0	0.0	3.0

The performance of the proposed classifier is compared with the other classifier using the performance metrics specifies in equ 5, 6 and 7. The number of abnormal and normal records in the dataset is 120 and 150 respectively. For the purpose of performance evaluation the dataset is separated with a cutoff point of 30 and hence further evaluation is restricted to 30 records.

Table 3. Performance Evaluation for Different Classifier Models

Classifier	Type	No. of Patients	Detected as abnormal	Detected as Normal	Sensitivity (%)	Specificity (%)	Accuracy (%)
ANN	Abnormal	120	100	20	83.33	86.67	85
	Normal	150	20	130			
K-means	Abnormal	120	105	15	86.07	89.86	88
	Normal	150	17	133			
FCM	Abnormal	120	108	12	91.53	92.10	92
	Normal	150	10	140			

Table 4. shows the results based on this cutoff point and each instance is assigned one of the two classes either abnormal (positive) or normal (negative). These classes form the confusion matrix and for the purpose of comparison the dataset from 1-30, 91-120 and 151- 180 records are taken. In the first case (1-30 records), the sensitivity and the specificity achieved is 100%, and hence this proves that the classifier has predicted the abnormal and normal cases with high accuracy. The case (91-120 records), the sensitivity is achieved to be 93.33 and specificity 86.67 and hence the accuracy is 90%. Though, the accuracy is less in this case, its high when compared to ANN and K-means which has an accuracy of 86.67% and 83.37% respectively. The case (151-180 records), accuracy is achieved to be 93.33% for proposed FCM algorithm. The case (91-120 records), the sensitivity is achieved to be 93.33 and specificity 86.67 and hence the accuracy is 90%. The sensitivity achieved for the FCM classifier is 91.53 with an average false positive of 0.9 per 30 records. The achieved accuracy is 92%, which is better than the performance of neural network based classifier and K-means clustering algorithm.

Table 4. Evaluation measure for Classifier Performance

Classifier	Type	1-30		31-60		61-90		91-120		121-150		151-180		181-210		211-240		241-270	
ANN	Abnormal	13	2	10	4	15	2	12	1	10	2	11	3	8	3	10	2	11	1
	Normal	1	14	3	13	2	11	2	15	3	15	1	15	3	16	2	16	3	15
K Means	Abnormal	8	2	12	1	13	3	10	1	14	2	13	1	10	2	13	1	12	2
	Normal	1	19	2	15	2	12	1	18	2	12	1	15	3	15	2	14	3	13
Fuzzy C Means	Abnormal	12	0	11	1	14	2	12	2	12	2	10	2	12	1	13	2	12	0
	Normal	0	18	1	17	1	13	1	15	2	14	0	18	2	15	1	14	2	16

Figure 4 - 6 shows the comparison of the sensitivity, specificity and accuracy rates of the three classifier models. The x-axis is taken as the different data classes of various records from 1-270 which is categorized into 9 records with a cutoff value of 30. The comparative analysis has proved that the proposed FCM algorithm dominates the other approached in terms of accuracy, specificity and sensitivity rates for 270 patients.

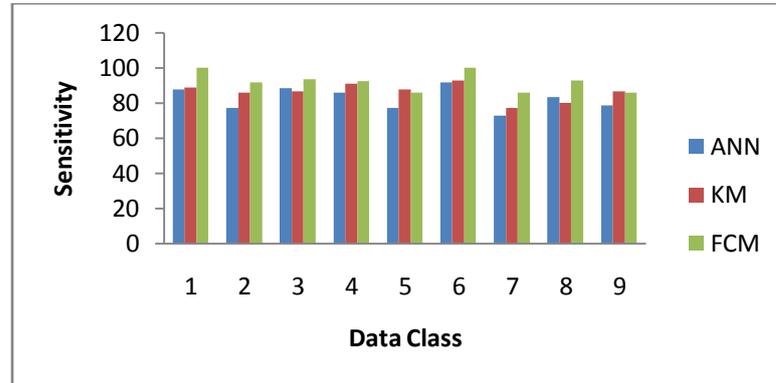


Figure 4. Sensitivity Comparison for ANN, KM and FCM

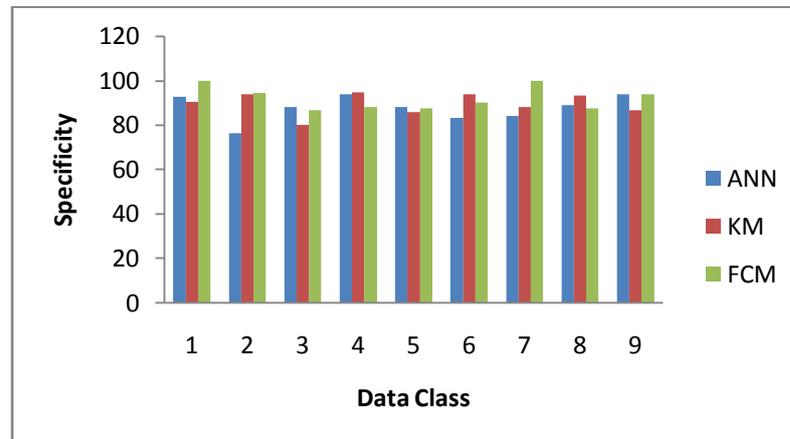


Figure 5. Specificity Comparison for ANN, KM and FCM

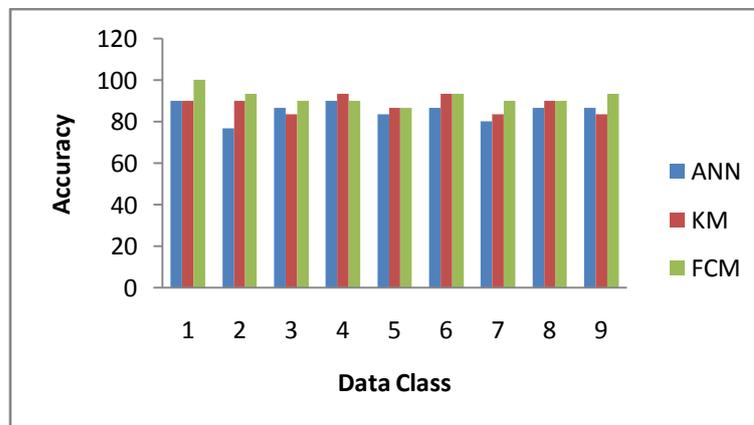


Figure 6. Accuracy Comparison for ANN, KM and FCM

The increase in the performance of the FCM clustering algorithm is because the weights of data attributes are set to adjust original samples to the uniform distribution, which could be suitable for the character of FCM calculation so as to improve the accuracy. To accelerate the convergent speed, the appropriate initial clustering centers are selected from experts' knowledge, which also eliminates the influence on how to choose the prior samples.

V. 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, performed 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 .

References

- [1]. Kawamoto, K., Houlihan, C.A., Balas, E.A., Lobach, D.F., 2005. "Improving clinical practice using clinical decision support systems: Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules 39 a systematic review of trials to identify features critical to success". *BMJ* 330, 765.
- [2]. 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 .
- [3]. 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.
- [4]. Sellappan Palaniappan,Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques" *IEEE Conference*, 2008,pp 108-115.
- [5]. K.Srinivas ,B.Kavihta Rani ,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, 2010, 250-255.
- [6]. Markos G. Tsipouras, Themis P. Exarchos, Dimitrios I. Fotiadis, Anna P. Kotsia, Konstantinos V. Vakalis,Katerina K. Naka, and Lampros K. Michalis" Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling " *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, No. 4, July 2008 .
- [7]. Dan Li, Hong Gu, Liyong Zhang," A fuzzy c-means clustering algorithm based on nearest-neighbor intervals for incomplete data", *Expert Systems with Applications* 37 (2010) 6942–6947
- [8]. Yang, M.-S., Hwang, P.-Y., Chen, D.-H., 2004. Fuzzy clustering algorithms for mixed feature variables. *Fuzzy Sets Systems* 141, 301–317.
- [9]. Ross, Timothy J., 2004. *Fuzzy Logic with Engineering Applications*, John Wiley & Sons, Second Edition.
- [10]. Lanhai L. Comparison of conventional and fuzzy land classification and evaluation techniques in Oxfordshire England. *Int Agric Eng J* 1998;7:1–12.
- [11]. Sadeghi, S., Barzi, A., Sadeghi, N., & King, B. (2006). A Bayesian model for triage decision support. *International Journal of Medical Informatics*, 75(5), 403–411.
- [12]. Yan, H.-M., Jiang, Y.-T., Zheng, J., Peng, C.-L., & Li, Q.-H. (2006). A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Systems with Applications*, 30(2), 272–281.
- [13]. Tsipouras, M.G., Exarchos, T.P., Fotiadis, D.I., Kotsia, A.P., Vakalis, K.V., Naka, K.K., Michalis, L.K., 2008. Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling. *IEEE Transactions on Information Technology in Biomedicine* 12 (4), 447–458.
- [14]. 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.
- [15]. Avci, E.; Turkoglu, I., "An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases". *Journal of Expert Systems with Application*, Vol. 2, No. 1, 2009, pp. 2873-2878.
- [16]. Setiawan, N.A., Venkatachalam, P.A., Hani, A.F.M., 2009. Diagnosis of coronary artery disease using artificial intelligence based decision support system. In: *Proceedings of the International Conference on Man-Machine Systems*, Batu Ferringhi, Penang, 11–13 October, 2009.