Emperor Journal of Applied Scientific Research

ISSN: 2581-964X Mayas Publication® www.mayas.info

Volume- VII Issue- V May 2025

Heart Attack Prediction Using Fuzzy C-Means: A Comparative Study with Neural Networks

G. Sharmila

Assistant Professor,
PG Department of Computer Science,
Islamiah Women's Arts and Science college,
Vaniyambadi, Tamil Nadu, India. 635 601.

R. Sangeetha

Assistant Professor
PG Department of Computer Science,
Islamiah Women's Arts and Science College,
Vaniyambadi, Tamil Nadu, India. 635 601.

Abstract

Fuzzy logic has emerged as a powerful tool in the medical domain, offering effective solutions for complex diagnostic tasks. It has been widely used in detecting critical conditions such as breast cancer, lung cancer, prostate cancer, and heart disease. This study presents an unsupervised classification model for early prediction of heart attacks using the Fuzzy C-Means (FCM) algorithm. The system analyzes patient medical records, utilizing 13 key attributes as input to assess heart attack risk. A dataset comprising 297 patient records was used to evaluate the model's performance, resulting in a classification accuracy of 100%. When compared to traditional neural network models like back propagation and adaptive linear networks, the FCM-based approach demonstrated superior efficiency and cost-effectiveness. The model was developed using MATLAB's Fuzzy Logic Toolbox and aims to support physicians in making more accurate and timely diagnoses of heart-related conditions.

Keywords: Fuzzy C means (FCM), Heart Attack Prediction, Unsupervised Classification, Medical Data Mining, And Neutral Network Comparison.

I. INTRODUCTION

Cardiovascular disease (CVD) is a leading global health concern, responsible for millions of deaths each year. According to estimates by the World Health Organization, approximately 12 million people die annually due to various forms of cardiovascular conditions. One of the most common causes is the gradual narrowing or obstruction of the coronary arteries, which supply blood to the heart muscle. This condition, known as coronary artery disease, can lead to heart attacks if not identified and treated promptly. A significant challenge in managing heart attacks is that around 30% of patients may not experience any noticeable symptoms, making timely diagnosis more complex. However, biochemical markers that indicate a heart attack remain in the bloodstream for several days and can assist in post-event analysis. Accurate diagnosis of heart disease is a critical yet challenging task in clinical practice. It typically requires physicians to assess numerous medical parameters, interpret diagnostic reports, and draw on their clinical experience. Given the variation in expertise among healthcare providers and the complexity of medical data, automated decision support systems are becoming increasingly important. These systems can help reduce diagnostic errors, improve patient outcomes, and provide consistent and cost-effective healthcare solutions.

Artificial Intelligence (AI), particularly fuzzy logic, has shown promising applications in the medical field. Fuzzy logic is capable of dealing with uncertainties and vague data, which are common in medical diagnosis. Unlike traditional binary logic, which handles only two states (true or false), fuzzy logic allows reasoning across a range of values between 0 and 1, closely resembling human thought processes.

Originally conceptualized by Jan Lukasiewicz in the 1930s and later developed into a comprehensive mathematical theory by Lotfi A. Zadeh in 1965, fuzzy logic provides a way to model imprecise reasoning. It uses "degrees of membership" rather than fixed classifications, making it suitable for analyzing complex systems like human health conditions. As such, fuzzy logic has been successfully applied in diagnosing several diseases, including breast cancer, lung cancer, and heart disorders.

This research focuses on the use of the **Fuzzy C-Means (FCM)** clustering algorithm, an unsupervised learning technique, to classify and predict heart attack risks. FCM assigns a degree of membership to each data point in multiple clusters, offering flexibility in handling overlapping or uncertain data. In this study, 13 clinical attributes from patient records are analyzed using FCM to assess the likelihood of heart disease. The system is developed using MATLAB's Fuzzy Logic

Toolbox, which supports both graphical and command-line operations for fuzzification, inference, and defuzzification processes.

By incorporating fuzzy logic into the diagnostic process, the proposed system aims to enhance the accuracy and reliability of early heart attack detection. This tool can serve as a supportive aid for healthcare professionals, especially in settings where experienced specialists are limited, thereby contributing to better patient care and resource optimization.

Related Works

Several research efforts have explored the use of intelligent systems for the diagnosis of heart disease. Data mining techniques such as decision trees, Naïve Bayes, and neural networks have been used to predict heart conditions based on medical parameters. For instance, Palaniappan et al. developed a web-based heart disease prediction system that is user-friendly and scalable, effectively identifying patterns associated with cardiac risk.

Tsipouras et al. introduced a fuzzy rule-based decision support system for diagnosing coronary artery disease, optimizing the model parameters through a four-stage methodology involving decision trees and fuzzy logic. Similarly, Setiawan et al. designed a fuzzy system using rough set theory to extract and weigh rules based on medical data.

Clustering techniques have also been widely applied. Shanthakumar et al. employed K-means to identify relevant patterns, while Dan Li et al. enhanced Fuzzy C-Means (FCM) to handle incomplete data using nearest-neighbor intervals. These approaches effectively manage uncertainty and missing values in medical datasets. Bayesian networks and neural networks have also gained traction for modeling uncertainty in diagnosis. Yan et al. utilized a multilayer perceptron with over 90% accuracy in predicting multiple heart diseases, and Avci and Turkoglu applied PCA and ANFIS for heart valve diagnosis.

More recent work by Yang et al. introduced adaptive FCM algorithms for datasets with mixed symbolic and fuzzy attributes, showing improved clustering performance. These studies emphasize the growing role of unsupervised and semi-supervised learning methods in medical diagnosis, especially when labeled data is scarce.

Proposed System

Fuzzy C Means: Fuzzy clustering is a class of algorithms for cluster analysis in which the allocation of data points to clusters is not "hard" (all-or-nothing) but "fuzzy" in the same sense as fuzzy logic.

- Explanation of clustering
- Fuzzy c-means clustering

Explanation of 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. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used as in clustering include distance, connectivity, and intensity.

In hard clustering, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In **fuzzy clustering** (also referred to as **soft clustering**), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. One of the most widely used fuzzy clustering algorithms is the <u>Fuzzy C-Means</u> (FCM) Algorithm (Bezdek 1981).

The FCM algorithm attempts to partition a finite collection of n elements $X = \{x_1, ..., x_n\}$ into a collection of c fuzzy clusters with respect to some given criterion.

Given a finite set of data, the algorithm returns a list of c cluster centres $C = \{c_1, ..., c_c\}$ and a partition $W = w_{i,j} \in [0,1], \ i=1,...,n, \ j=1,...,c$, where each element w_{ij} tells the degree to which element x_i belongs to cluster c_j . Like the kmeans algorithm, the FCM aims to minimize an objective function. The standard function is:

$$w_k(x) = \frac{1}{\sum_j \left(\frac{d(\operatorname{center}_k, x)}{d(\operatorname{center}_j, x)}\right)^{2/(m-1)}}.$$

which differs from the k-means objective function by the addition of the membership values u_{ij} and the fuzzifier m. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships w_{ij} and hence, fuzzier clusters. In the limit m=1, the memberships w_{ij} converge to 0 or 1, which implies a crisp partitioning. In the absence of experimentation or domain knowledge, m is commonly set to 2. The basic FCM Algorithm, given n data points $(x1, \ldots, xn)$ to be clustered, a number of c clusters with $(c1, \ldots, cc)$ the center of the clusters, and m the level of cluster fuzziness with.

Fuzzy c-means clustering

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely too just 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.

The fuzzy c-means algorithm is very similar to the k-means algorithm

- Choose a number of clusters.
- 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 expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes.

Table 3.1-Sample Input attributes values for the dataset used for the prediction system

SLNo	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 cholestoral 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

Flow Chart

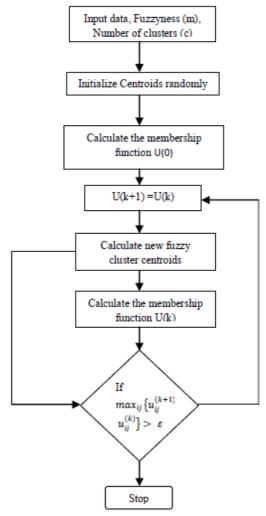


Fig 1: Flow diagram of FCM

displays that do not demand a backlight, making them thinner and more efficient than Display screens (which do require a white backlight).

Methodology

Performance Evaluation

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.

The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease. **SENSITIVITY=TP/TP+FN** (1)

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.

$$SPECIFICITY = TN/TN + FP$$
 (2)

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

$$ACCURACY = \frac{TN+TP}{TN+TP+FN+FP}$$
(3)

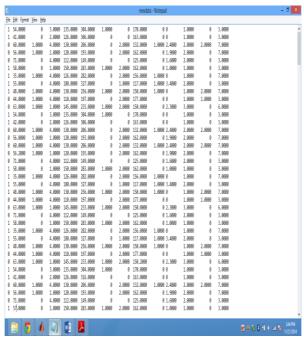
Table 4.1.1: Sample Data Given for Testing

AGE	SEX	CHEST PAIN	RES TBP	CHOLES TEROL	BLOO DSUG AR	EC G	MAX HEART RATE	ANGI NA	OLDP EAK	ST SLOP E	VESS ELS	TH AL
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

Mayas Publication

34

Data Set Taken



The above snap shot, shows the sample data taken with a total of 297 patients. The data is viewed through a notepad file. The datasets with 14 attributes taken as a class 0 or class 1. Class 1 represents the abnormal cases and 0 represents the normal cases.

Assumption:

The total patients taken as samples for input are 297, out of which 162 are abnormal patients and 135 are normal patients.

Proposed Work:

The above specified samples are given as input to the BPN, ADALINE and FCM classifiers to the results are tabulated below.

In this thesis FCM classifier results are compared with the other classifier results namely BPN and ADALINE in terms sensitivity, specificity and accuracy using the equation 1, 2 and 3 and the results are tabulated below

35

Table 4.1.2 Performance of different classifier

Classifier	INPUT(NUMBER OF THE PATIENTS)		OUTPUT(NUM THE PATIENT				
	Categories of the patient	No of Patients	Detected as abnormal	Detected as normal	Sensitivity (%)	Specificity (%)	Accuracy (%)
BPN	Abnormal	162	145	17	89.5 %	92.5%	90.9%
	Normal	135	10	125			
ADALINE	Abnormal	162	142	20	87.6%	88.8%	88.2%
	Normal	135	15	120			
FCM	Abnormal	162	162	0	100%	100%	100%
	Normal	135	0	132			

4.2 Graphical Comparison

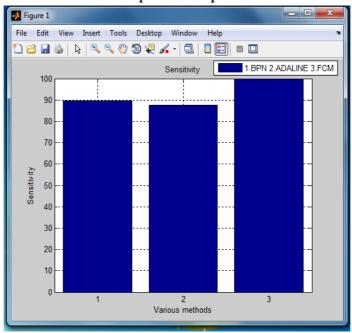


Figure 4.2.1 Sensitivity comparison for BPN, ADALINE and FCM

The above graph shows Sensitivity percentage for BPN, ADALINE and FCM. From the bar chart it is evident that FCM has the highest percentage (%) of sensitivity when compared with BPN and ADALINE.

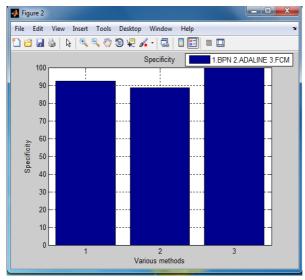


Figure 4.2.2 Specificity comparison for BPN, ADALINE and FCM

The above graph shows Specificity percentage for BPN, ADALINE and FCM, from the bar chart it is evident that FCM has the highest percentage (%) of specificity when compared with BPN and ADALINE.

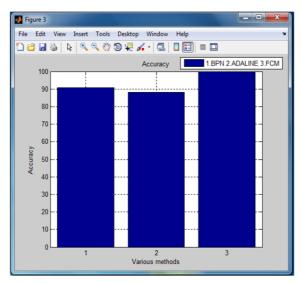


Figure 4.2.3 Accuracy comparison for BPN, ADALINE and FCM

The above graph shows accuracy percentage for BPN, ADALINE and FCM, from the bar chart it is evident that FCM has the highest percentage (%) of accuracy when compared with BPN and ADALINE.

Discussion

From the Table 4.1.2 it is evident that FCM clustering algorithm process to be 100% efficient in terms of sensitivity, specificity and accuracy when compared with BPN and ADALINE. In the classification stage 13 attributes are given as input to the Fuzzy C Means (FCM) classifier to determine the risk of heart attack. The efficiency of the classifier is tested using the records collected from 297 patients, which gives a classification accuracy of 100%. The proposed system is implemented using the features of fuzzy logic toolbox in matlab.

II. CONCLUSION

Thus the proposed work, FCM, ADALINE and BPN classifier are implemented 297 samples and the taken as input to the results were calculated table 5.2. Each person had got 13 sets of data with the last data as a label. Only in case of ADALINE, target will be supplied along with data sets. But in other cases the target will be given separately into the algorithm. Fuzzy C means based clustering outperformed and has accuracy increased and reached 100% at even randomized new data sets. Hence FCM saves time in training as well as testing. The FCM performs better and hence this work could be highly useful the various applications. When compared the fuzzy c means, process to more efficient and cost effective rather than the back propagation and adaptive linear network. The proposed system is implemented using the features of fuzzy logic toolbox in matlab. The proposed system will provide an aid for the physicians to diagnosis the disease in a more efficient way.

Future Enhancement

Future research could focus on expanding the dataset size and diversity to further validate the robustness of the FCM-based approach. Integration with real-time clinical data and exploration of hybrid models combining fuzzy logic with other machine learning techniques may improve diagnostic accuracy and reliability. Additionally, enhancing the system's user interface and integrating it with hospital management systems could increase its practical applicability in healthcare settings.

III. 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)`
- [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.