International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

Different Fuzzy C Means Clustering Algorithm Variations on Heart Beat Datasets

Moksud Alam Mallik[1], Hafsa Yasmeen[2], Nousheen Begum[3] ,Mirza Younus Ali Baig[4]

[1]Dean R&D and Associate Professor,Department of CSE-Data Science,Lords Institute of Engineering and Technology,Hyderabad,India

[2]UG Student's,Department of CSE-Data Science, Lords Institute of Engineering and Technology, Hyderabad, India

[3]UG Student's,Department of CSE-Data Science, Lords Institute of Engineering and Technology, Hyderabad, India

[4]Assistant Professor,Department of CSE-Data Science, Lords Institute of Engineering and Technology, Hyderabad, India

***Abstract*— Since cardiovascular illnesses are among the world's top causes of death, techniques for accurate and timely detection are essential. ECG signal-derived heartbeat datasets contain intricate, noisy, and non-linear patterns that necessitate the use of clever clustering techniques for efficient analysis. One soft-clustering approach that works well for medical signals with ill-defined borders is fuzzy C-Means (FCM) clustering, which divides data points into several clusters with different levels of membership. Nevertheless, the conventional FCM algorithm frequently has drawbacks such dependence on initial cluster centers, delayed convergence, and sensitivity to noise. A number of better FCM variations, such as Possibilistic FCM, Kernel-based FCM, Weighted FCM, and Spatial FCM, have been proposed to address these problems. To improve resilience, accuracy, and computational efficiency when processing ECG heartbeat data, these improved methods use spatial constraints, kernel techniques, adaptive weighting, and membership regularization. By using these techniques on heartbeat datasets, aberrant patterns like arrhythmias may be properly identified, which can be very helpful for patient monitoring and clinical decision-making. An overview of these FCM changes is provided in this abstract, along with information on how their alterations enhance performance, particularly on heartbeat datasets. According to the study, enhanced FCM variants are a useful tool in biomedical signal analysis since they offer superior clustering and classification capabilities when compared to the conventional FCM approach.**

***Keywords*—Clustering, Different fuzzy clustering, Heart beat signal, ECG signals, Machine Learning**

I. INTRODUCTION

Computer-based data classification, also known as clustering, is a multivariate statistical analysis technique that separates a data set into classes or clusters based on a set of criteria and maximizes the similarity of the data object rather than the difference within the same cluster. Each of us has an algorithm bias when working with text large data, and we employ different algorithms to address various issues[1][2].

Dividing a data collection into c homogenous fuzzy

clusters is the aim of fuzzy clustering. The Fuzzy C-Means (FCM) technique is the most popular fuzzy clustering algorithm. The However, the technique requires a user-specified number of clusters. Clustering partitions vary depending on the initial number of clusters chosen. As a result, after each fuzzy c-partition is identified, it must be validated. Cluster validity is the term for this assessment. Bezdek's partition coefficient (PC), partition entropy (PE), and Xie-Beni's index are among the many fuzzy cluster validity indices that have been suggested in the literature and are commonly employed in contemporary research[3][4].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

As civilization advances and people's quality of life continues to rise, the high demands of mental and physical labor cause individuals to pay more attention to their health. As one of the four vital indicators of the human body, heart rate is a fundamental indicator of health. In order to determine a subject's emotional state, stress level, sleep quality, and other medical and psychological conditions, heartbeat detection and analysis can efficiently assess the sympathetic and parasympathetic nerve activity status of the individual. The non-invasive, easy, and affordable nature of ballistocardiogram (BCG)-assisted heart rate monitoring has attracted a lot of interest from both academics and industry. In order to record the force of blood on blood vessels at each heartbeat, the BCG system often includes piezoelectric, acceleration, and other non-contact sensors.[5][6]

Cardiovascular diseases (CVDs) are the world's largest cause of death, according to the World Health Organization1. In order to provide tools, methods, and other best practices for lowering the occurrences of first and recurring cardiovascular events, a number of programs and policies have been put into place in more varied communities in recent years. In order to accomplish this, the electrocardiogram (ECG) has emerged as the most widely used biosignal for the timely identification of cardiovascular diseases. A graphical depiction of cardiac electrical activity, the ECG is used to detect a number of heart conditions and anomalies2. For more than 70 years, medical professionals have used ECG signals to identify heart conditions like arrhythmias and myocardial infarctions. The P, QRS complex, and T waves make up an ECG signal3,4,5[7][8]

In order to create a content recommendation system tailored to DHH, this work focuses on learning style computation as the foundation for individualized suggestions.[9] The scientific study of statistical models and methods that allow computers to carry out particular tasks without explicit programming is known as machine learning. ML systems are especially good at identifying patterns and forecasting outcomes because they learn from data rather than predetermined rules.[10]. By facilitating data-driven decision-making and automation, machine learning (ML) has profoundly changed sectors like healthcare, finance, robotics, and autonomous systems[11].

The structure of this paper is organized as follows:

Section 1 is Introduction,Literature Review is there in section 2 follwed by Methodology at section 3.Section 4 is Results and Discussion,final section Conclusion and Future Work.

II. LITERATURE REVIEW

1. Possibilistic C-Means (PCM) and fuzzy C-Means (FCM) Dr. Ashok Sharma, Dr. Parveen Singh, and Sarneet Kaur are authors.
These clustering methods can manage uncertainty and show soft clustering capacity. They can distinguish between noise and real data pieces more successfully and are more resilient to noise. They do, however, need preset clusters and are susceptible to outliers. Because predetermined cluster numbers are required, their implementation is more complicated.

2.Hierarchical Clustering and K-Means Clustering Himanshu Sharma, Anas Akhtar, and Aastha Gupta are the authors.
K-Means and Hierarchical Clustering are straightforward, simple to comprehend, and simple to use. They divide into smaller groups, which facilitates comprehension and makes analysis easier. One of their benefits is that dynamic selection is possible because the number of clusters does not have to be predetermined. However, only when K is small are they computationally quicker. Furthermore, cluster splitting or merging is an irreversible process that makes it challenging to undo mistakes later.

3. Fuzzy C-Means (FCM) and fuzzy adaptive resonance theory (ART) Choukri Djellali et al. wrote this.
Fuzzy ART is an extension of Binary Adaptive Resonance Theory that handles real-valued data using fuzzy set operations. An input layer and a competitive layer make up its two layers. It uses unsupervised learning that is founded on interactions through both top-down and bottom-up long-term memory processes and a competitive mechanism. Compared to the standard initialization-based fuzzy ART, the random initialization-based fuzzy ART has slower training and a lower recognition accuracy. Comparing random initialization-based FCM to usual initialization-based FCM, the former also has a lower recognition accuracy. For these strategies to work well, the clustering models need to be carefully designed and adjusted.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

III. METHODOLOGY

1. Heartbeats
One full cycle of the heart's contraction (systole) and relaxation (diastole) is called a heartbeat. The pumping activity is responsible for the body's blood circulation.
Heart Rate Normal
Adults' resting heart rates range between 60 and 100 beats per minute (bpm).
Athletes: As low as 40–60 beats per minute
Children: Depending on age, faster, between 70 and 120 bpm

2. ◆ Heart Rate Types:
Description of Type
Heart Rate at Rest Heart rate at total rest
Heart Rate Active While exercising or engaging in physical activities
Maximum Heart Rate (approximate formula) = 220 - age
For healthy exercise, the target heart rate should be between 50 and 85% of maximum heart rate.
◆ Oddities
Bradycardia: Slow heartbeat (less than 60 beats per minute)
Fast heart rate (>100 bpm) is known as tachycardia.
Arrhythmia: Missed beats or irregular rhythm
Instruments for Heartbeat Monitoring
Stethoscope
ECG (Electrocardiogram) pulse oximeter

3. Heartbeat Datasets
The following available datasets can be used to study ECG or heartbeat data:
Well-liked Datasets
Description of the Dataset Source: MIT-BIH Arrhythmia Dataset ECG recordings annotated with various cardiac disorders 290 participants and 549 ECG recordings are available in the PhysioNet PTB Diagnostic ECG Database. PhysioNet Heart Condition Medical characteristics associated with heart health that are frequently utilized in ML UCI Repository MIMIC-III includes clinical information, such as ECG, from ICU patients. PhysioNet

4. The Jupyter Notebook
You can create and execute code in real-time, together with text and visualizations, using the open-source interactive web application Jupyter Notebook.
Crucial Elements
supports a variety of languages, including R, Julia, Python, and others.
Code execution that is interactive
Charts and inline plots
Support for Markdown and LaTeX
Excellent for data visualization, machine learning, and data science
Fundamental Structure
There are two primary sorts of cells in every notebook:
Code cells are where code is written and executed.
Markdown cells: Used to write explanations and styled text

5. ◆ Sample Python Code to Analyze Heart Rate

```
import pandas as pd

import matplotlib.pyplot as plt

# Load sample heart rate data

df = pd.read_csv('heart_rate_data.csv')

# replace with real path

# Basic info

print(df.head())

print(df.describe())

# Plot heart rate

plt.plot(df['Time'], df['HeartRate'])

plt.title('Heart Rate Over Time')

plt.xlabel('Time')

plt.ylabel('Heart Rate (bpm)')
```

```
plt.grid(True)

plt.show()
```

Two crucial libraries are imported at the start of this Python script: matplotlib.pyplot for visualization creation and pandas for data processing. The data is then stored in a DataFrame called df after the script uses pandas.read_csv() to read a CSV file called "heart_rate_data.csv." After providing a brief overview of the data by displaying the first five rows of the dataset using df.head(), it uses df.describe() to print summary statistics for each numeric column, including mean, standard deviation, and range, which aid in understanding the distribution and central tendencies of the data.

The script then uses plt.plot() to plot the Time column on the x-axis and the HeartRate column on the y-axis in order to visualize the heart rate data. To make the graph more informative, it adds labels for the x-axis ('Time') and y-axis ('Heart Rate (bpm)') as well as a title ('Heart Rate Over Time'). For improved reading, plt.grid(True) is used to add a grid to the background. The graph window is then shown to the user using plt.show(), enabling them to see how heart rate varies over time.

There are numerous reliable datasets available for the study and analysis of heart rate and associated conditions. These include the MIMIC-III dataset, which contains comprehensive clinical data from intensive care unit patients, including ECG readings; the PTB Diagnostic ECG Database, which includes detailed diagnostic recordings from hundreds of people; and the MIT-BIH Arrhythmia Dataset, which includes annotated ECG recordings of various cardiac disorders. Additionally, the UCI repository has heart health datasets that are frequently utilized for machine learning.

Data scientists use Jupyter Notebook, a well-liked and robust open-source web tool, to analyze such data. It is appropriate for research projects, data visualization, and machine learning since it enables users to create and execute code in an interactive style. It has two primary cell types: code cells, which are used for writing and running code, and markdown cells, which are used for written text, explanations, and LaTeX equations. It supports a variety of programming languages, including Python, R, and Julia.

Python's pandas and matplotlib.pyplot packages are used in a straightforward example of heart rate analysis. A CSV file containing heart rate data is loaded into a DataFrame by the script, which then uses df.head() to display the first few rows and df.describe() to summarize statistical data. After that, it uses plt.plot() to plot the heart rate over time, labeling the axes and including a title and grid for improved viewing. The plot is finally rendered using plt.show(), enabling users to see patterns and trends in the heart rate data. Understanding cardiovascular behavior requires this type of analysis, which may also be used to create predictive health models.

- Well known datasets heartbeat datasets

Dataset Name	Source
MIT-BIH Arrhythmia Dataset	Physio Net
PTB Diagnostic ECG Database	Physio Net
Heart Condition Data	UCI Machine Learning Repository
MIMIC-III	Physio Net

IV. RESULT

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

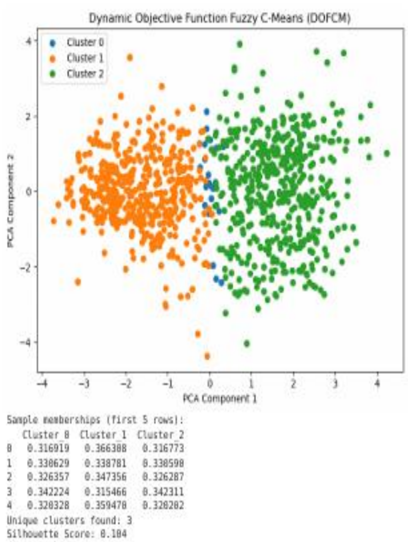

Fig 1 "DOFCM Clustering Results (PCA Projection)"

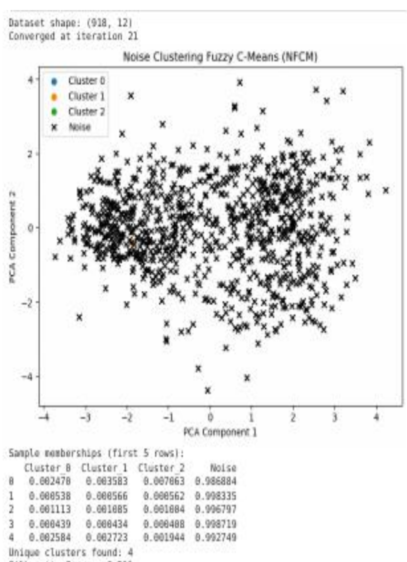

Fig 2 "Noise Clustering Fuzzy C-Means Results"

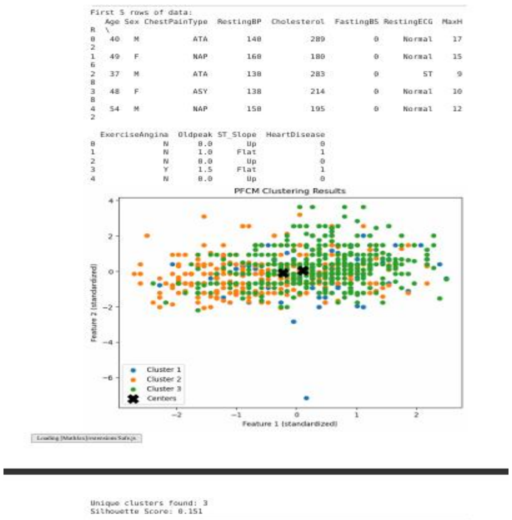

Fig 3 "PFCM Clustering Results"

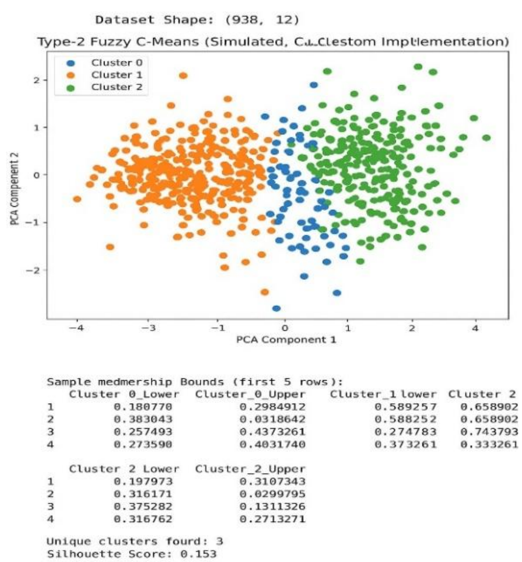

Fig 4 "Type-2 Fuzzy C-Means Clustering Results (PCA Projection)"

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

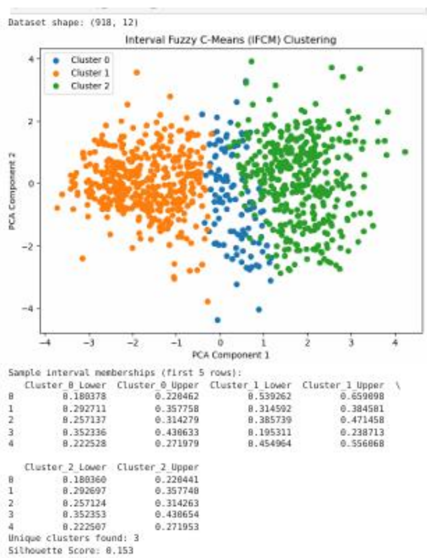

Fig 5 "Interval Fuzzy C-Means (IFCM) Clustering Results"

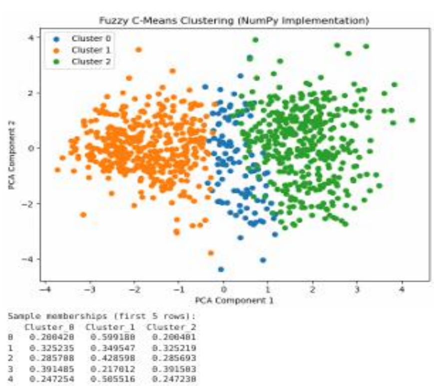

Fig 6 "Fuzzy C-Means (FCM) Clustering Results"

V. CONCLUSION AND FUTURE ENHANCEMENTS

Analysis of heartbeat datasets using various variants of the fuzzy C-Means (FCM) clustering algorithm—FCM, PFCM, T2FCM, IFCM, CFCM, NCFCM, and DOFCM—shows that each algorithm provides special improvements designed to address particular biomedical signal processing issues like noise, uncertainty, and overlapping data.
Basic fuzzy clustering is provided by standard FCM, however it has trouble processing noisy or non-linear heartbeat data.
By combining fuzzy and possibilistic memberships, PFCM (Possibilistic FCM) increases robustness and decreases sensitivity to outliers.
T2FCM (Type-2 FCM) improves its performance in noisy, real-world cardiac signals by introducing uncertainty modeling with Type-2 fuzzy logic.
By incorporating changes like weighted distances or spatial limitations, IFCM (Improved FCM) improves segmentation accuracy and convergence.
By adding domain-specific constraints, CFCM (Constrained FCM) improves clustering and makes clinical data easier to comprehend.

By focusing on datasets with high noise levels, NCFCM (Noise Clustering FCM) provides more robust heartbeat segmentation.
Distance-Optimized FCM, or DOFCM, maximizes distance measurements to enhance cluster separation, particularly in intricate heartbeat patterns.
Since noise, uncertainty, and non-linearity are crucial aspects of physiological signal analysis, advanced FCM variants such as T2FCM, NCFCM, and DOFCM perform better overall on heartbeat datasets. The particular dataset's properties and the clinical needs should guide the algorithm selection.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 -6435 (Online)), Volume 15, Issue 3, March 2026)

VI. References

[1] Chen, Z.L. (2022) 'Research and Application of Clustering Algorithm for Text Big Data', *Computational Intelligence and Neuroscience*. Edited by L. Sun, 2022, pp. 1–8. Available at: https://doi.org/ /2022/7042778

[2] Qin, J., Ma, Q., Shi, Y., & Wang, L. (2020). Recent Advances in Consensus of Multi-Agent Systems: A Brief survey. *IEEE Transactions on Industrial Electronics*, *64*(6), 4972–4983. https://doi.org/10.1109/tie.2020.2636810

[3] Zhang, N. M., Zhang, N. W., Sicotte, H., & Yang, N. P. (2020). A new validity measure for a correlation-based fuzzy c-means clustering algorithm. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3865–3868. https://doi.org/10.1109/iembs.2020.5332582

[4] Demirhan, H. (2024). Mixed fuzzy C-means clustering. *Information Sciences*, 121528. https://doi.org/10.1016/j.ins.2024.121528

[5] Geng, D., Yin, Y., Fu, Z., Pang, G., Xu, G., Geng, Y., & Wang, A. (2024). Heart rate detection method based on Ballistocardiogram signal of wearable device : Algorithm development and validation. *Heliyon*, *10*(5), e27369. https://doi.org/10.1016/j.heliyon.2024.e27369

[6] Namazi, H., Baleanu, D., & Krejcar, O. (2021). AGE-BASED ANALYSIS OF HEART RATE VARIABILITY (HRV) FOR PATIENTS WITH CONGESTIVE HEART FAILURE. *Fractals*, *29*(03), 2150135. https://doi.org/10.1142/s0218348x21501358

[7] Aziz, S., Ahmed, S., & Alouini, M. (2021). ECG-based machine-learning algorithms for heartbeat classification. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-97118-5

[8] Shen, C., Liu, X., Zhang, H., Zhao, H., and Qi, T. (2024). investigation into hybrid residual network-based ECG signal categorization. 11202 in Applied Sciences, 14(23). The article https://doi.org/10.3390/app142311202

[9] A. Poly, P. K. N. Banu, N. Althuniyan, A. T. Azar, and N. A. Kamal, "Fuzzy Logic Approach to Cold-Start Challenges in Deaf and Hard of Hearing Recommender Systems", *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 3, pp. 23449–23460, Jun. 2025.

[10] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: a survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

[11] Dodampegama, S., Hou, L., Asadi, E., Zhang, G., & Setunge, S. (2023). Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications. *Resources Conservation and Recycling*, *202*, 107375. https://doi.org/10.1016/j.resconrec.2023.107375