Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 959–968

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Experimental Analysis of Fuzzy Clustering Techniques for Outlier Detection

Harsh Yadav[a], Jaspreeti Singh[b], Anjana Gosain[c]

[a]*Department of Information Technology, USICT, GGS Indraprastha University, Sector 16C, Dwarka, Delhi-110075, India*
[b]*Assistant Professor, USICT, GGS Indraprastha University, Sector 16C, Dwarka, Delhi-110075, India*
[c]*Professor, USICT, GGS Indraprastha University, Sector 16C, Dwarka, Delhi-110075, India*

Abstract

Fuzzy clustering is an effective clustering approach that divides the dataset into fuzzy segments. Fuzzy clustering and outlier detection are two interconnected processes. Outlier detection is essential because it discloses hidden patterns and crucial information about a dataset and is beneficial in a wide range of applications like fraud detection, military surveillance, identifying computer network intrusions, image processing, insurance or health care, etc. Many fuzzy clustering techniques are robust to anomalies since they limit the outlier's impact on the cluster's centroid.

In this paper, we have compared four fuzzy clustering techniques namely Fuzzy C-means (FCM), Noise Clustering (NC), Credibilistic Fuzzy C-means (CFCM), and Density Oriented Fuzzy C - Means (DOFCM) on the basis of some crucial properties that are essential for efficient outlier detection. To better evaluate the feasibility of all these techniques, we experimentally evaluated these techniques on six datasets (two Real and four Synthetic) in the presence of noise and outliers. The results of the comparative analysis show that DOFCM outperforms all other techniques in terms of outlier detection and cluster formation. This research will surely benefit any researcher willing to detect outliers in their study using fuzzy clustering or either implicitly or explicitly working on fuzzy clusters.

Keywords: Fuzzy Clustering; Outlier Detection; FCM; CFCM; NC; DOFCM

1. Introduction

Fuzzy clustering is a kind of unsupervised classification technique, that is extremely useful when working with data fuzziness, vagueness, and uncertainty. Many applications have used fuzzy clustering to do outlier detection [1], data analysis, market research, business intelligence, image pattern recognition, web search, recommender systems, fault diagnosis, resource optimization, etc [2]. Outlier detection concentrates on a tiny fraction of data items that are

10.1016/j.procs.2023.01.076

frequently ignored as noise. Outliers or anomalies can be more interesting than frequently occurring normal points in some applications (e.g., fraud detection). Hawkins [3] provided the most popular definition of an outlier in 1980. According to [3], the term outlier refers to data points that contradict or deviate significantly from the general behavior of the dataset as if it were produced by using some distinct method. Outlier detection can discover system flaws and fraud before they have disastrous implications. So, this anomalous data must be analyzed since it has the potential to be turned into valuable data, that can be used in a variety of applications.

In 1973, Dunn [4] proposed FCM, which is among the most eminent method for creating fuzzy clusters, and in 1981, Bezdek [5] improved it. The FCM method has proven effective in a variety of technological fields, including pattern identification and image analysis, etc. [6, 7]. It performs well with noiseless data but drastically failed to recognize noise & outliers as a result its centroid is attracted to outliers rather than the actual cluster center. So far, A lot of researchers have attempted various improvements to FCM algorithm to overcome its shortcomings. In 1993 Custem and Gath [8] proposed a fuzzy-clustering technique by employing 2 hypotheses to determine the existence of outliers. However, it failed to consider the potential of several clusters of outliers. Krishnapuram and Keller [9] in 1993, developed the possibilistic c-means (PCM) by relaxing the column sum constraint of the FCM membership matrix. It depicts clustering like a possibilistic segmentation. It was concluded that it could be utilized to tackle various issues, including large data clustering [10,11], boundary detection and surface approximation [12,13], Image segmentation [14,15,16], and outlier detection [17] but it is extremely sensitive to good initialization & required a solid scale approximation to function appropriately [18]. Nikhil et al. [19] highlighted another shortcoming of the PCM algorithm that it could produce coincident clusters if each row's initialization is not distinguishable. To address these shortcomings, in 2005 [19] proposed Possibilistic Fuzzy C-Means (PFCM) by combining the FCM's fuzzy values and PCM's typicality values to generate a superior clustering model. However, PFCM also failed to provide the desired outcome since its centroid is highly influenced by noise and outliers and it also failed to work for unequally sized clusters.

Further, Dave [20][21] proposed noise clustering (NC) based on an extra centroid (for all outliers) and noise distance. It produced excellent results with noisy datasets but does not detect outliers situated between normal clusters. Frank Rehm [22] emphasized that this method failed to detect outliers when the no. of partitions on same dataset increases because the mean difference among data items & good partitions shrinks while noise distance stays unchanged [22]. By introducing a new parameter, credibility, Krishna K. Chintalapudi et al. [23] introduced CFCM to minimize the influence of outliers. It outperforms fuzzy c-means, PCM, and PFCM, but it frequently allocates some outlier points to several clusters [24] and is sensitive to good initialization of prototype. Chen and Wang [25] bring forward another shortcoming of CFCM that it has been using unreliable instances to calculate the credibility, these unreliable instances are experiencing the convergence process & outcome can be influenced by changing the initial values of these instances. In contrast with NC, CFCM does not detect anomalies; instead, it focuses on lowering their impact on the resulting clusters. Many research papers [26-28] have mentioned this problem, including Gosain A. et al [23], which compares the performance of various fuzzy clustering techniques on the D12 data set. In 2011, Gosain et al. [29] introduced Density oriented FCM to address the shortcomings of FCM, PCM, PFCM, CFCM, and NC algorithms. DOFCM approach of finding outliers prior to clustering yielded excellent results. It was found to be beneficial for image segmentation as well as for enhancing image quality from noisy versions [30] but it fails to deal with clusters that are not linearly separable [31].

This paper provides a comparative analysis of FCM, CFCM, NC, and DOFCM algorithms to find out which fuzzy clustering algorithm is best for outlier detection in the presence of noise and outliers.

This paper is organized as follows: Section 2 briefly reviewed FCM, NC, CFCM & DOFCM algorithms. Section 3 discusses dataset used. Section 4 presents experimental results of algorithms used in the form of figures and tables. Section 5 depicts comparative analysis, and Section 6 is the conclusion of this research.

2. An Overview

This section contains a detailed analysis of various fuzzy clustering methods for outlier detection namely: FCM, CFCM, NC and DOFCM.

2.1. Fuzzy C-Means (FCM)[4,5]

FCM utilizes the distance among all the data items & the center of the cluster for determining the membership of all the data items with respect to each centroid. It begins with the assumption that we knew the value of c ("number of clusters") & attempts to reduce the objective function (J_{fcm}) to a bare minimum. Membership (U_{ik}) and cluster centers (V_K) are modified after each iteration using the following formula:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ki}}{d_{ji}}\right)^{\frac{2}{m-1}}} \ \forall k, i \qquad v_k = \frac{\sum_{i=1}^{n} u_{ik}^{m} x_i}{\sum_{i=1}^{n} u_{ik}^{m}} \qquad J_{fcm} = \sum_{k=1}^{c}\sum_{i=0}^{n} u_{ik}^{m}\, d_{ik}^{2} \quad (1,2,3)$$

It performs well with noiseless data but drastically failed to recognize noise & outliers as a result its centroid is attracted to outliers rather than the cluster center.

2.2. Noise Clustering (NC)[20,21]

Dave [20] proposed noise clustering (NC) based on additional centroid (for all the outliers) and noise distance. The constant, noise distance suggested by him is defined as the gap separating the centroid of the noise cluster & data items. It's a key factor in noise clustering efficiency. As a result, the statistical average is utilized to determine $\boldsymbol{\delta^2}$. The membership of data points in a noise cluster is defined as: $u_{*i} = 1 - \sum_{k=1}^{c} u_{ki}$ (4)

Modified FCM objective function [30]: $J(U,V) = \sum_{k=1}^{c+1}\sum_{i=1}^{N}(u_{ki})^m\,(d_{ki})^2$ (5)

Where k = n = c+1.

and Membership eq. of Normal data points: $u_{ij} = \left[\sum_{k=1}^{c}\left[\frac{d_{ij}^2}{d_{kj}^2}\right]^{\frac{1}{m-1}} + \left[\frac{d_{ij}^2}{\delta^2}\right]^{\frac{1}{m-1}}\right]^{-1} \qquad \delta^2 = \lambda\left[\frac{\sum_{k=1}^{c}\sum_{i=1}^{N}(d_{ki})^2}{Nc}\right]$ (6,7)

Noise clustering is robust due to the influence of noise distance in the process but still does not provide efficient clusters as it does not detect outliers situated between normal clusters.

2.3. Credibilistic Fuzzy C-Means (CFCM)[23]

By introducing a new parameter, credibility, Krishna K. Chintalapudi et al. [23] introduced CFCM to minimize the influence of outliers. Outliers have a low value, but good points have a high value in credibility. Credibility is defined by CFCM as [23]: $\Psi_k = 1 - \frac{(1-\theta)\alpha_k}{max_{j=1..n}(\alpha_j)}, 0 \le \theta \le 1 \quad Where\ \alpha_k = min_{i=1..c}(d_{ik})$ (8)

Here, α_k represents the distance from any data item x_k to its closest centroid [10]. Θ tries to maintain the credibility as low as possible to ensure, the noisiest vector receives credibility equivalent to Θ. On setting Θ equivalent to one, CFCM will work as FCM, and on setting Θ =0, the noisiest vector will be assigned zero memberships. It outperforms fuzzy c-means, PCM, and PFCM, but it frequently allocates some outlier points to several clusters. Hence reliable clustering is not possible.

2.4. Density Oriented Fuzzy C - Means (DOFCM)[29]

Kaur et al. [29] introduced DOFCM, for efficient outlier detection. Just like the Noise Clustering technique it also produces "n+1" clusters, including an additional cluster for all the outliers. But unlike the NC, which creates an extra cluster of outliers during clustering, DOFCM eliminates outlier points before clustering. In order to identify outliers, the DOFCM algorithm employs a neighborhood membership. The neighborhood membership (or density factor) can be calculated by using the equation

$$M_{neighborhood}^{i}(X) = \frac{\eta_{neighborhood}^{i}}{\eta_{max}} \qquad M_{neighborhood}^{i} = \begin{cases} < \alpha; Outlier \\ \ge \alpha; Non-outlier \end{cases} \quad (9,10)$$

Here, $\eta^i_{neighborhood}$ denotes No. of other points present in the surrounding of that data item.

η_{max} denotes the highest no. of points present in the surrounding of any data item in the dataset.

In DOFCM Outlier detection relies heavily on the threshold value α. The threshold value "α" is chosen by the user, depending mostly on the data item's range of density factors. The Fuzzy C-Means method bears the restriction of not allowing the membership value to go zero [29]. By assigning outliers a membership value of zero, the DOFCM method eliminates the influence of outliers [29]. Hence, DOFCM accurately identifies outliers and picks more appropriate cluster centroids, as a result improving the clustering accuracy.

3. Dataset Used

We have used 6 datasets out of which 2 are real datasets (Iris and Wine) and 4 are synthetic datasets (D12, D14, Pearl, and Spirals Square) to evaluate the performance of FCM, NC, CFCM, and DOFCM algorithms.

3.1. Real Datasets

Iris dataset [32] is having 150 four-dimensional feature vectors divided into two clusters as well as 13 outliers. Figure 1(a) shows the original Iris dataset. Wine dataset [33] is another real-world dataset with 178 four-dimensional feature vectors separated into two clusters and 17 outliers. Figure 2(a) shows the original Wine dataset.

3.2. Synthetic Datasets

D12 [1] is often referred to as the Diamond data set. It is a small dataset with 2 clusters and a single outlier. Figure 3(a) shows the original D12 dataset. D14 [35] dataset is another synthetic dataset consists of two clusters and 3 outliers. Figure 4(a) shows the original D14 dataset. Pearl dataset [34] is a massive dataset, having1000 two-dimensional feature vectors separated into two clusters and 12 outliers. Figure 5(a) shows the original Pearl dataset. Spirals Square dataset [34] is an enormous, complicated, and noisy dataset. It consists of 1500 two-dimensional feature vectors split across six groups and 27 outliers. Figure 6(a) shows the original Spirals Square dataset.

4. Simulation and Results

In this section of the research paper, FCM, NC, CFCM, and DOFCM algorithms have been evaluated on two real and four synthetic datasets in the presence of noise and outliers. Tables and figures have been used to present experimental data in order to give a comparative view. For testing, simulation is performed on a PC running MATLAB Version R2021a 9.10 on AMD Ryzen 7 4800H CPU running at 2.90GHz with 8 GB of RAM. The following symbols are used to illustrate clustering results: Red/Green/Yellow/Purple/Blue "+, x or o" for distinct clusters, bold black '+ or o' for centroids and outlier with Grey 'o'.

Iris Dataset

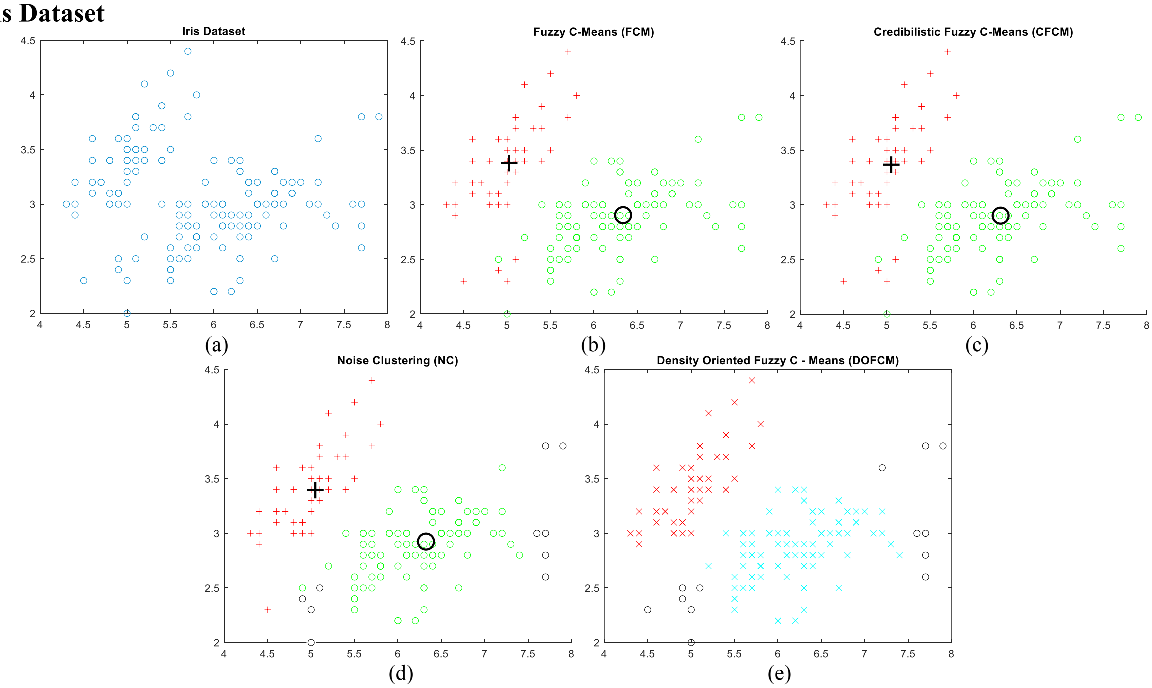


Fig 1. - Performance of (a) Dataset, b) FCM, c) CFCM, d) NC, e) DOFCM with Iris Dataset

Table 1: Performance b) FCM, c) CFCM, d) NC, e) DOFCM with Iris Dataset.

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids	
			C1	C2
FCM	2	0	5.02328, 3.38108	6.33543, 2.90528
CFCM	2	0	5.04909, 3.36756	6.31072, 2.90083
NC	2	10	5.04837,3.39595	6.32287, 2.92429
DOFCM	2	13	5.07347, 3.32595	6.35576, 2.92222

Wine Dataset

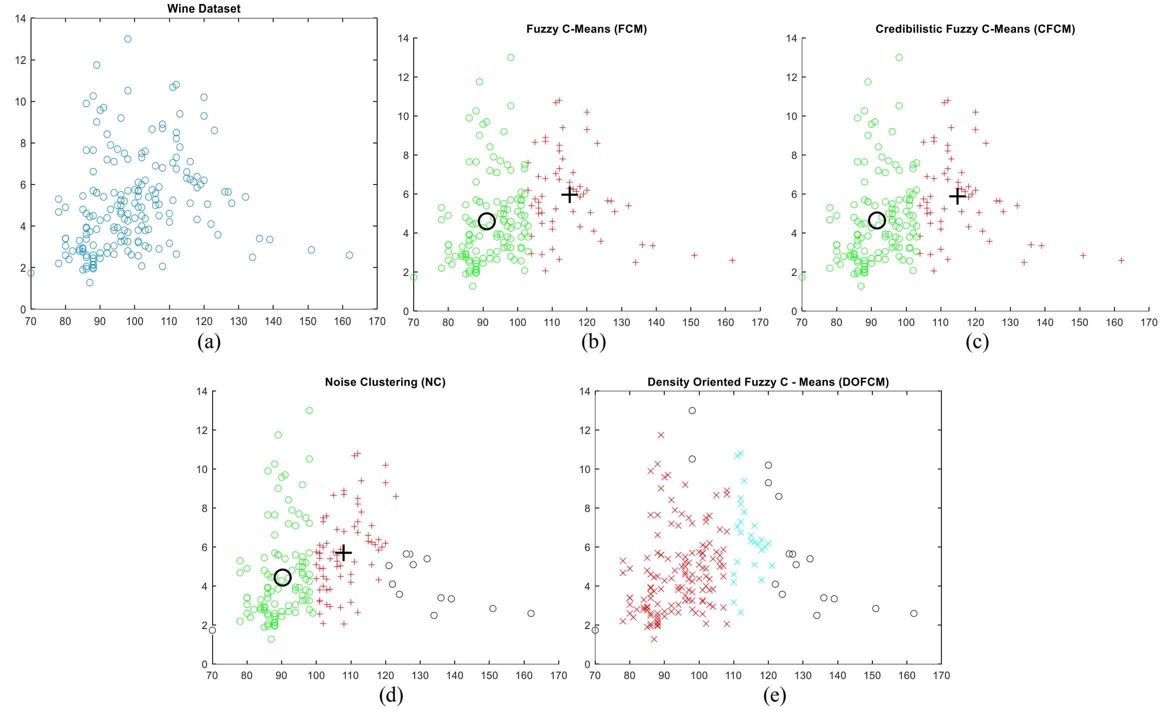
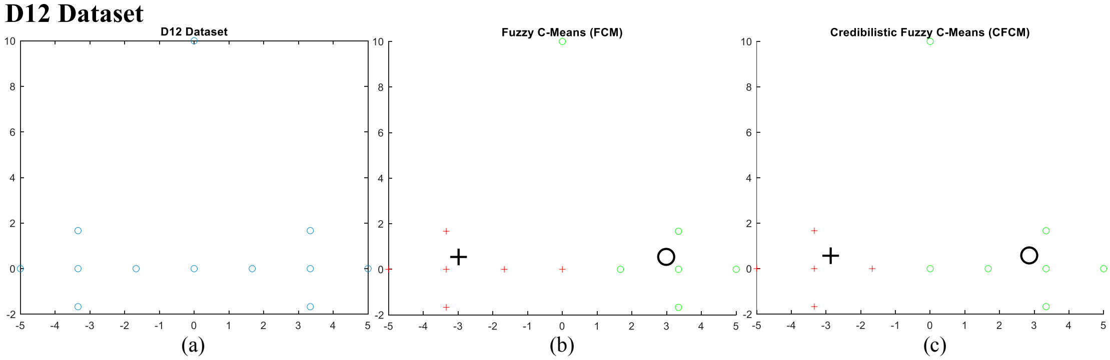


Fig 2. - Performance of (a) Dataset b) FCM, c) CFCM, d) NC, e) DOFCM with Wine Dataset

Table 2: Performance of b) FCM, c) CFCM, d) NC, e) DOFCM with Wine Dataset.

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids	
			C1	C2
FCM	2	0	91.0795, 4.60342	115.011, 5.96634
CFCM	2	0	91.6082, 4.64602	114.801, 5.88305
NC	2	13	90.2878, 4.43768	107.883, 5.70861
DOFCM	2	17	90.598, 4.4067	107.208, 5.6562

D12 Dataset



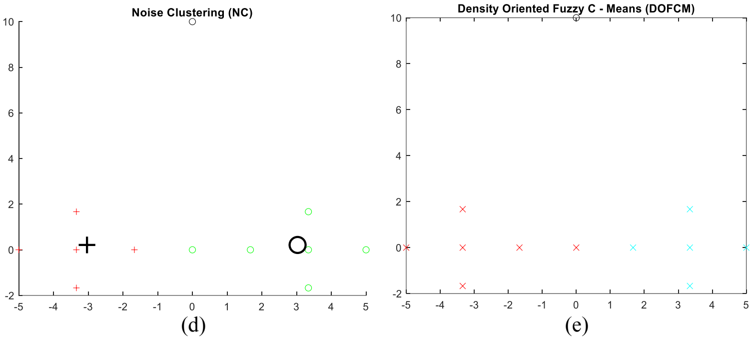


Fig 3. - Performance of (a) Dataset, b) FCM, c) CFCM, d) NC, e) DOFCM with D12 Dataset

Table 3: Performance of b) FCM, c) CFCM, d) NC, e) DOFCM with D12 Dataset.

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids C1	C2
FCM	2	0	-2.98465, 0.542144	2.98596, 0.544983
CFCM	2	0	-2.86696, 0.569816	2.85603, 0.582472
NC	2	1	-3.03465, 0.21291	3.02895, 0.212238
DOFCM	2	1	-3.0, 0.2	3.0, 0.19

D14 Dataset

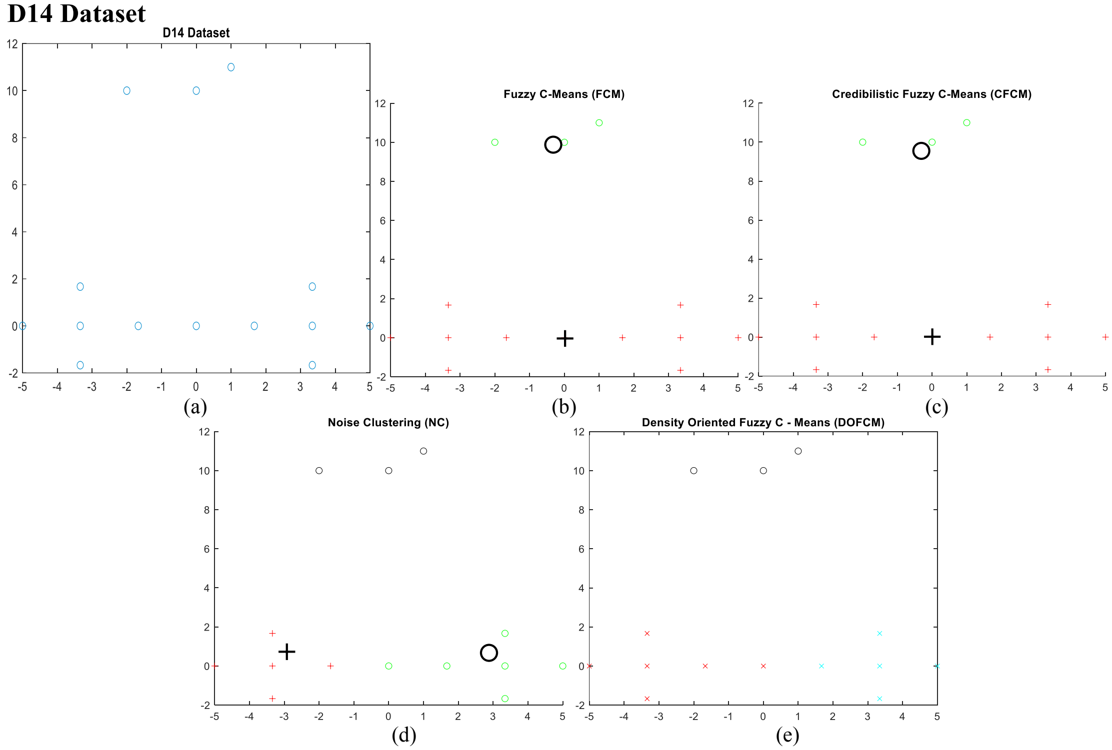


Fig 4. - Performance of (a) Dataset , b) FCM, c) CFCM, d) NC, e) DOFCM with D14 Dataset

Table 4: Performance of b) FCM, c) CFCM, d) NC, e) DOFCM with D14 Dataset.

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids	
			C1	C2
FCM	2	0	-0.0407631, 0.0193162	-0.319903, 9.87747
CFCM	2	0	0.00713688, 0.01852	-0.309922, 9.55166
NC	2	0	-2.92452, 0.729625	2.88099, 0.6729
DOFCM	2	3	-2.92, 0.2	2.9, 0.2

Pearl Dataset

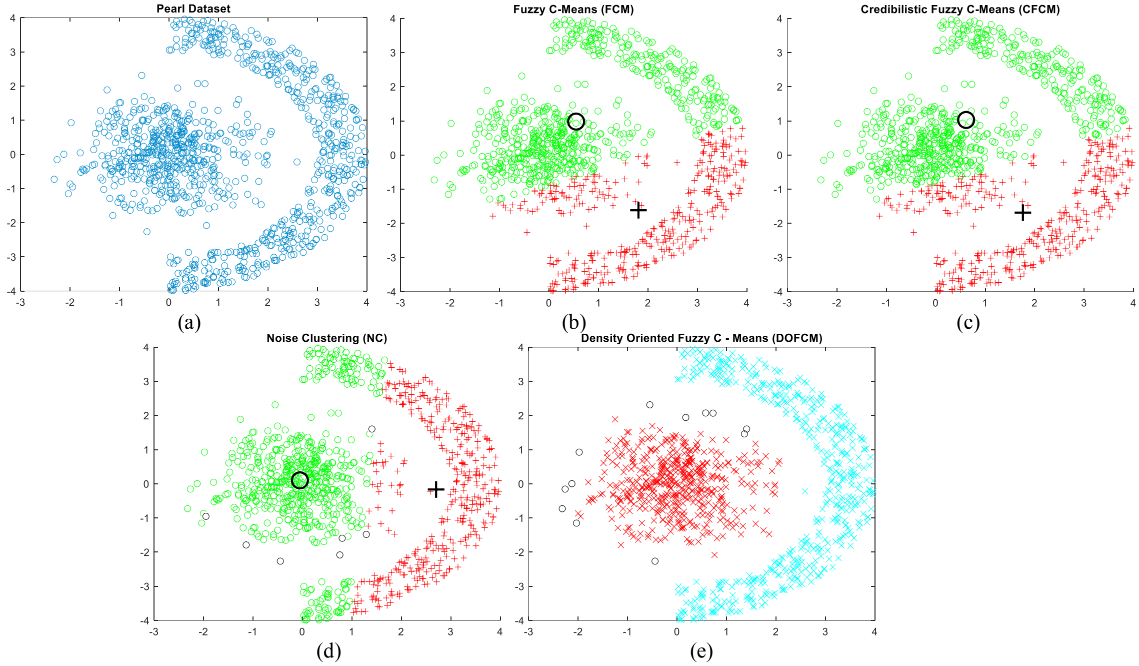
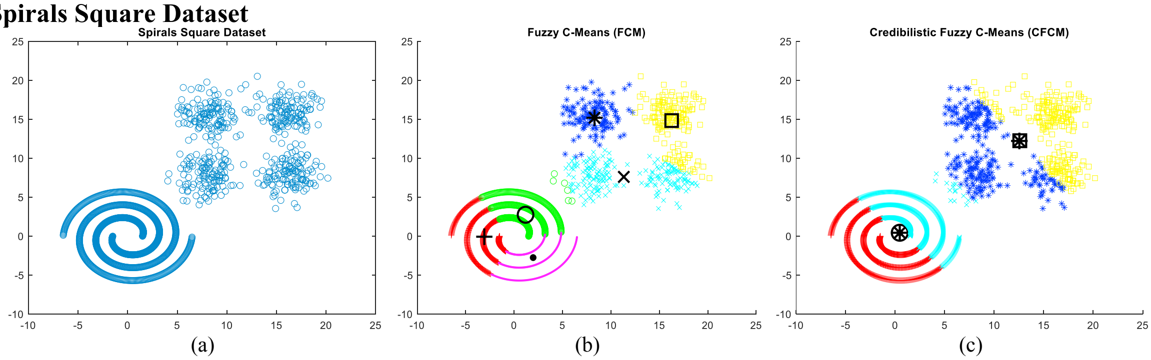


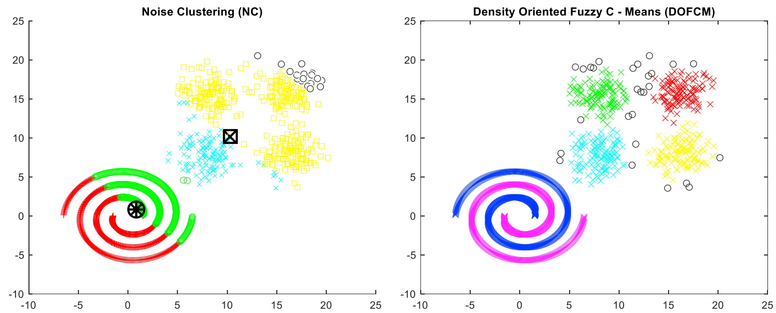
Fig 5. - Performance of (a) Dataset, b) FCM, c) CFCM, d) NC, e) DOFCM with Pearl Dataset.

Table 5: Performance of b) FCM, c) CFCM, d) NC, e) DOFCM with Pearl Dataset

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids	
			C1	C2
FCM	2	0	0.546413, 0.978649	1.80632, -1.61666
CFCM	2	0	0.613199, 1.02623	1.76062, -1.68595
NC	2	7	-0.050322, 0.100359	2.70286, -0.166638
DOFCM	2	12	-0.05034, 0.10036	3.52006, 0.0746348

Spirals Square Dataset





(d) (e)

Fig 6. - Performance of (a) Dataset, b) FCM, c) CFCM, d) NC, e) DOFCM with Spirals Square Dataset.

Table 6: Performance of b) FCM, c) CFCM, d) NC, e) DOFCM with Spirals Square Dataset.

TECHNIQUE/ ALGORITHM	NO. OF CLUSTERS FORMED	NO. OF OUTLIERS DETECTED	Centroids					
			C1	C2	C3	C4	C5	C6
FCM	6	0	-3.07818,	1.9663,	1.18118,	11.3282,	8.3009,	16.2701,
CFCM	4	0	0.46146,	0.46146,	12.582,	12.582,	-------	-------
NC	4	18	0.83485,	0.83485,	10.3211,	10.3211,	-------	-------
DOFCM	6	27	-3.07818,	1.6435,	8.3009,	16.4077,	16.6141,	8.22301,

As we can see in Fig 1-6, FCM and CFCM algorithms drastically failed to recognize noise & outliers. CFCM allocated some outliers to both partitions (Fig 2,5,6) and outliers were included in the final results. Hence, both algorithms failed to provide efficient clusters. However, NC & DOFCM showed no attraction towards outliers and produced better results with the noisy datasets. NC and DOFCM maintained low membership values of outliers. As a result, the DOFCM method provided more exact centroids than all other methods. Whereas Figure 6 shows that increasing the number of clusters & outliers had a negative impact on the performance of the CFCM and NC algorithm. CFCM and NC both identified 4 incorrect partitions. We observed that only NC and DOFCM algorithms identified outliers but compare to DOFCM, NC failed to detect all the outliers, especially outliers situated between normal clusters (Fig 1,5,6). whereas DOFCM accurately identified all the outliers since it determines outliers based on data-set density and is independent of the number of clusters and outliers. Moreover, It has successfully detected the original clusters from noisy datasets. However, when we modified the values of α and neighbourhood, we noticed that DOFCM generated inefficient clusters.

5. Comparative Analysis of Different Techniques

These algorithms have been compared based on some crucial properties, that are essential for efficient outlier detection. These properties are- i) Its centroid should not be attracted to outliers or noise [23]. ii) All outliers in all clusters must be assigned lower memberships [23]. iii) It must be independent of the number of clusters for the same data set [22]. iv) It should be independent of any number of outliers [30]. v) It should be independent of the location of outliers in the data sets [30]. vi) It should be able to detect all the outliers accurately so that it can successfully detect original clusters from noisy datasets [9].

Table 7: Comparative Analysis

PARAMETERS	FCM	CFCM	NC	DOFCM
APPROACH	DISTANCE BASED	DISTANCE BASED	DISTANCE BASED	DENSITY BASED
FEATURE	FOUNDATION OF THE FUZZY CLUSTERING APPROACH	USE OF CREDIBILITY OF DATA ITEMS	USE OF NOISE CLUSTER & NOISE DISTANCE	IDENTIFYING OUTLIERS BEFORE THE CLUSTERING PROCESS
IMPLEMENTATION	EASY	EASY	COMPLICATED	HARD
ASSIGN LOWER MEMBERSHIP TO OUTLIERS	NO	YES	YES	YES
CENTROID ATTRACTED TOWARDS OUTLIERS	YES	NO	NO	NO
OUTLIERS INCLUDED IN FINAL CLUSTERS	YES	YES	YES	NO
INDEPENDENT OF NUMBER OF CLUSTERS	NO	NO	NO	YES
INDEPENDENT OF NUMBER OF OUTLIERS	NO	YES	YES	YES
INDEPENDENT OF LOCATION OF OUTLIERS	NO	NO	NO	YES
PRODUCE ACCURATE CENTROIDS	NO	NO	YES	YES
ACCURATELY DETECTED ALL THE OUTLIERS	NO	NO	NO	YES
SUCCESSFULLY DETECTS ORIGINAL CLUSTERS FROM NOISY DATASET	NO	NO	NO	YES
DRAWBACKS	1) DRASTICALLY FAILED TO RECOGNIZE NOISE & OUTLIERS. 2) ITS CENTROID IS ATTRACTED TO OUTLIERS RATHER THAN THE CLUSTER CENTER	1) IT FREQUENTLY ALLOCATES SOME OUTLIER POINTS TO SEVERAL CLUSTERS	1) FAILS TO DETECT OUTLIERS LOCATED BETWEEN THE NORMAL CLUSTERS. 2) IF NO. OF CLUSTERS INCREASED, NC DOES NOT DETECT OUTLIERS	1) ITS SUCCESS IS STRONGLY RELIANT ON PICKING THE CORRECT VALUES OF α & $r_{neighborhood}$.

Table 7 clearly illustrates that DOFCM is the most effective fuzzy clustering technique for outlier detection. DOFCM possesses all the qualities required for effective outlier detection and effectively satisfies all the above-mentioned properties.

6. Conclusion

In this paper, we have compared four fuzzy clustering techniques namely Fuzzy C-means (FCM), Noise Clustering (NC), Credibilistic Fuzzy C-means (CFCM), and Density Oriented Fuzzy C - Means (DOFCM) on the basis of some crucial properties that are essential for efficient outlier detection. We have also empirically evaluated these fuzzy clustering techniques using six datasets out of which two are real datasets (Iris and Wine) and four synthetic datasets (D12, D14, Pearl, and Spirals Square) in the presence of noise and outliers. It was observed that DOFCM outperformed all other techniques. It accurately identified all the outliers, selected the more desirable cluster centers, and generated the original clusters from noisy datasets. Thus, the DOFCM approach is extremely robust regardless of the number of clusters, outliers, or the position of outliers; however, its success is strongly reliant on picking the correct values of α & neighbourhood. In the future, we will try to come up with an algorithm that optimizes existing clustering results.

References

[1] Gosain, A., & Dahiya, S. (2016). Performance analysis of various fuzzy clustering algorithms: a review. *Procedia Computer Science*, *79*, 100-111.
[2] Han, J., Kamber, M., & Pei, J. (2006). Data Mining Concepts and Techniques Book. Southeast Asia Edition: Morgan Kaufmann.
[3] Hawkins, D. M. (1980). Identification of Outliers Chapman & Hall London 1.
[4] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
[5] Bezdek, J.C. (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York
[6] Kesavaraj, G., & Sukumaran, S. (2013, July). A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1-7). IEEE.
[7] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, *31*(3), 264-323.
[8] Van Cutsem, B., & Gath, I. (1993). Detection of outliers and robust estimation using fuzzy clustering. *Computational statistics & data analysis*, *15*(1), 47-61.
[9] Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE transactions on fuzzy systems*, *1*(2), 98-110.
[10] Zhang, C., Gao, R., Qin, H., & Feng, X. (2021). Three-way clustering method for incomplete information system based on set-pair analysis. *Granular Computing*, *6*(2), 389-398.
[11] Zhang, Q., Yang, L. T., Castiglione, A., Chen, Z., & Li, P. (2019). Secure weighted possibilistic c-means algorithm on cloud for clustering big data. *Information Sciences*, *479*, 515-525.
[12] Krishnapuram, R., Frigui, H., & Nasraoui, O. (1995a). Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation. i. *IEEE Transactions on Fuzzy Systems*, *3*(1), 29-43.
[13] Krishnapuram, R., Frigui, H., & Nasraoui, O. (1995b). Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation. II. *IEEE transactions on Fuzzy Systems*, *3*(1), 44-60.
[14] Kalist, V., Ganesan, P., Sathish, B. S., & Jenitha, J. M. M. (2015). Possiblistic-fuzzy C-means clustering approach for the segmentation of satellite images in HSL color space. *Procedia Computer Science*, *57*, 49-56.
[15] Zhang, Y., Huang, D., Ji, M., & Xie, F. (2011). Image segmentation using PSO and PCM with Mahalanobis distance. *Expert systems with applications*, *38*(7), 9036-9040.
[16] Xie, Z., Wang, S., Zhang, D., & Chung, F. L. K. (2007). Image segmentation using the enhanced possibilistic clustering method. *Information Technology Journal*, *6*(4), 541-546.
[17] Filippone, M., Masulli, F., & Rovetta, S. (2010). Applying the possibilistic c-means algorithm in kernel-induced spaces. *IEEE Transactions on Fuzzy Systems*, *18*(3), 572-584.
[18] Krishnapuram, R., & Keller, J. M. (1996). The possibilistic c-means algorithm: insights and recommendations. *IEEE transactions on Fuzzy Systems*, *4*(3), 385-393.
[19] Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE transactions on fuzzy systems*, *13*(4), 517-530.
[20] Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters*, *12*(11), 657-664.
[21] Dave, Rajesh N. "Robust fuzzy clustering algorithms." In Fuzzy Systems, 1993., Second IEEE International Conference on, pp. 1281-1286. IEEE, 1993.
[22] Rehm, F., Klawonn, F., & Kruse, R. (2007). A novel approach to noise clustering for outlier detection. *Soft Computing*, *11*(5), 489-494.
[23] Chintalapudi, K. K., & Kam, M. (1998, May). A noise-resistant fuzzy c means algorithm for clustering. *IEEE international conference on fuzzy systems proceedings,* May, Vol. 2, pp. 1458-1463.
[24] Kaur, P. (2016). Improved version of kernelized fuzzy C means using credibility. Int J Comput Sci Netw, 5(1), 50-54.
[25] Chen, J. L., & Wang, J. H. (1999, October). A new robust clustering algorithm-density-weighted fuzzy c-means. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* (Vol. 3, pp. 90-94). IEEE.
[26] Gosain, A., & Dahiya, S. (2020). A new robust fuzzy clustering approach: DBKIFCM. *Neural Processing Letters*, *52*(3), 2189-2210.
[27] Dahiya, S., Gosain, A., & Mann, S. (2021). Experimental analysis of fuzzy clustering algorithms. In *Intelligent data engineering and analytics* (pp. 311-320). Springer, Singapore.
[28] Dahiya, S., Gosain, A., & Gupta, S. (2020, April). RKT2FCM: RBF Kernel-Based Type-2 Fuzzy Clustering. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.
[29] Prabhjot, K., Lamba, I. M. S., & Anjana, G. (2011). DOFCM: a robust clustering technique based upon density. International Journal of Engineering and Technology, 3(3), 297.
[30] Kaur, P., & Gosain, A. (2011). A density oriented fuzzy c-means clustering algorithm for recognising original cluster shapes from noisy data. *International Journal of Innovative Computing and Applications*, *3*(2), 77-87.
[31] Gosain, A., & Dahiya, S. (2021). An effective fuzzy clustering algorithm with outlier identification feature. Journal of Intelligent & Fuzzy Systems, 41(1), 2417-2428.
[32] Dua, D., & Graff, C. (2019). The UCI machine learning repository-iris data set. Source: https://archive. ics. uci. edu/ml/datasets/Iris.
[33] Dua, D., & Graff, C. (2019). The UCI machine learning repository-iris data set. Source:https://archive.ics.uci.edu/ml/datasets/wine
[34] Yadav Harsh. (2022). Synthetic Datasets. Source: https://github.com/Harshyadav14/SyntheticDataset
[35] Sharma, S., Goel, M., & Kaur, P. (2013). Performance comparison of various robust data clustering algorithms. *IJ Intelligent Systems and Applications*, *7*, 63-71.