Applied Soft Computing Journal 161 (2024) 111712

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Feature-weight and cluster-weight learning in fuzzy *c*-means method for semi-supervised clustering

Amin Golzari Oskouei [a,b,*], Negin Samadi [c], Jafar Tanha [c]

[a] Faculty of IT and Computer Engineering, Urmia University of Technology, Urmia, Iran
[b] Department of Software Engineering, Faculty of Engineering and Natural Science, Istinye University, Istanbul, Turkey
[c] Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Iran

HIGHLIGHTS

- This paper proposes a novel Semi-Supervised Fuzzy C-means approach based on Feature-Weight and Cluster-Weight learning.
- An adaptive weight for each feature in each cluster applies based on its importance in forming the clusters.
- The weight of the clusters is calculated dynamically to decrease cluster center initialization sensitivity.
- Considering the conjunction of feature and cluster weighting helps form an optimal clustering structure.

ARTICLE INFO

Keywords:
Semi-Supervised Clustering
Fuzzy *c*-means
Feature weighting

ABSTRACT

Semi-supervised clustering aims to guide the clustering by utilizing auxiliary information about the class labels. Among the semi-supervised clustering categories, the constraint-based approach uses the available pairwise constraints in some steps of the clustering procedure, usually by adding new terms to the objective function. Considering this category, Semi-supervised FCM (SSFCM) is a semi-supervised version of the fuzzy *c*-means algorithm, which takes advantage of fuzzy logic and auxiliary class distribution knowledge. Despite the performance enhancement caused by incorporating this extra knowledge in the clustering process, semi-supervised fuzzy approaches still suffer from some problems. All the data attributes in the feature space are assumed to have equal importance in the cluster formation, while some features may be more informative than others. Thus the feature importance issue is not addressed in the semi-supervised category. This paper proposes a novel Semi-Supervised Fuzzy *c*-means approach, which is designed based on Feature-Weight, and Cluster-Weight learning, named SSFCM-FWCW. Inspired by the SSFCM, a fuzzy objective function is presented, which is composed of (1) a semi-supervised term representing the external class knowledge; (2) a feature weighting; and (3) a cluster weighting. Both feature weights and cluster weights are determined adaptively during the clustering. Considering these two techniques leads to insensitivity to the initial center selection, insensitivity to noise, and consequently helps to form an optimal clustering structure. Experimental comparisons are carried out on several benchmark datasets to evaluate the proposed approach's performance, and promising results are achieved. The Matlab implementation source code of the proposed method is publicly accessible at https://github.com/Amin-Golzari-Oskouei/SSFCM-FWCW.

Code metadata

Permanent link to reproducible Capsule: https://doi.org/10.24433/CO.1422782.v1

1. Introduction

Data clustering aims to partition the samples into dense groups to maximize the intra-cluster similarity and minimize the inter-cluster similarity [1]. Data clustering provides insight into the distribution of the data and is also used as a vital tool in a wide range of scientific

* Correspondence to: Urmia University of Technology (UUT), Band Road, Golshahr 2, Urmia 57155-419, West Azerbaijan, Iran.
E-mail address: a.golzari@tabrizu.ac.ir (A. Golzari Oskouei).

https://doi.org/10.1016/j.asoc.2024.111712
Received 29 April 2023; Received in revised form 19 April 2024; Accepted 3 May 2024
Available online 7 May 2024

applications, such as recommender systems, market segmentation, social network analysis, image processing, and biological data analysis [2–10].

Among the diverse clustering categories, Centroid-based clustering is one of the well-known clustering categories in which each cluster is represented by a central point vector, which is not necessarily among the provided data points [11,12]. Over the years, many algorithms have been proposed in this field, and methods such as fuzzy c-means and k-means are the most widely used [13]. The standard procedure followed in these approaches is the iterative calculation of cluster centers and the assignment of samples to the clusters.

Considering the restriction degree for assigning the samples to the clusters, approaches are divided into two general categories [14]: a) hard clustering; and b) soft/fuzzy clustering. In the hard clustering approach, cluster assignment is in a manner that each sample is restricted to be a member of only one cluster. In contrast, the soft clustering approach considers a membership degree in the range $[0,1]$ for each sample concerning each cluster, which represents the degree of the partial membership of the sample in that cluster. In cases where the data distribution requires overlapping cluster structures or the boundary between the clusters is ambiguous, a fuzzy point-of-view in cluster assignment results in a better clustering structure [15].

Despite the superiority of fuzzy clustering approaches in comparison with hard clustering, they still act in an unsupervised manner. In some real-world applications, some kinds of prior knowledge about the class distribution of the data are provided by experts [16,17]. Consequently, semi-supervised clustering, which attempts to enhance the unsupervised clustering performance by utilizing this prior knowledge about the labels of data points [18], has attracted the attention of many researchers. Based on how this external knowledge is included in the clustering process, the semi-supervised algorithms are categorized into three groups [19,20]: a) Distance-based approach, b) Constraint-based approach, and c) Hybrid approach. The distance-based category attempts to design distance/similarity metrics to satisfy the class label constraints, e.g., studies in [21–23]. The constraint-based approach utilizes the available pairwise constraints to guide the clustering process, usually by adding new terms to the objective function, e.g., studies in [24,25]. The third group represents the studies that consider a combination of distance-based and constraint-based solutions, such as the methods proposed in [26,27].

Also, there exist some very recent fuzzy approaches that yield in the semi-supervised clustering category, such as S3FCM [28], CS3FCM [29], LHC-S3FCM [30], and AS3FCM [31], TS3FCM [32], TS3MFCM [33]. Most of these proposed approaches are inspired by the SSFCM algorithm, which is a pioneering semi-supervised fuzzy clustering approach [25]. SSFCM provides a reformulated fuzzy objective function embedded with the external class label information in a way that considers the difference between the predicted class and the actual class of the instances as a penalty term [34]. Despite the better performance of these approaches compared to unsupervised methods, they still suffer from some basic assumptions that prevent them from determining the optimal clusters. None of the investigated semi-supervised fuzzy clustering approaches distinguish between the diverse attributes in the feature space. All features are assumed to have equal importance in forming the clusters, while some features may be more informative than others. This assumption prevents the formation of the optimal cluster structure. Although this challenge is partially solved in unsupervised clustering [35,36], there still is a lack of enough research on using feature weighting techniques in semi-supervised clustering.

In this paper, to overcome the abovementioned problems, a novel semi-supervised fuzzy c-means clustering approach, named SSFCM-FWCW, is proposed. Inspired by the SSFCM [25], a novel objective function is designed based on Feature-weight and Cluster-weight learning. Embedding the class label information directly to the objective function enhances the algorithm's knowledge about the initial number of clusters, better approximation of initial center vectors, and assigning the labeled instances to the correct cluster representing the same label. Embedding a feature weighting parameter to the objective function bolds the effect of important features in constructing optimal clusters. Furthermore, considering different weights for the clusters leads to insensitivity to the initial centers. Moreover, using a non-Euclidean distance metric in the objective function prevents the side effects of noisy and outlier data points in the performance, which results in a noise-insensitive approach. Finally, the main contributions of the proposed method can be concluded as follows:

- By taking advantage of fuzzy logic and the prior class distribution knowledge, a novel semi-supervised fuzzy clustering objective function is designed, along with the presentation of the optimization steps.
- A feature weighting technique is carried out during the clustering iterations, which applies an adaptive weight for each feature based on its importance in forming the clusters.
- In addition to feature weighting, a weighting scheme is also considered for the clusters, which is set adaptively for each cluster, leading to an insensitivity to initial centers selection.
- Considering the conjunction of feature weighting and cluster weighting simultaneously investigates the importance of each feature concerning each cluster, consequently ensuring an optimal clustering structure.

The performance of the proposed method is compared to several state-of-the-art approaches over various benchmark datasets in terms of clustering accuracy. The results from the experiments indicate the superiority of the proposed method compared to the other methods. Besides, multiple detailed analyses are performed to investigate the effect of cluster weighting, feature weighting, and the labeled data ratio on the performance.

The remainder of the paper is organized as follows: Section 2 briefly discusses the related works; the preliminaries and the contribution of this paper are presented in Sections 3 and 4, respectively; the experimental evaluation details and the results are provided in Section 5. Section 6 briefly discusses the real-world applications of the proposed method; Finally, the conclusion of the paper and future works are summarized in Section 6.

2. Related work

In this section, state-of-the-art fuzzy clustering approaches are reviewed. Sections 2.1 and 2.2 are dedicated to unsupervised and semi-supervised fuzzy clustering categories.

2.1. Unsupervised fuzzy clustering

Fuzzy c-means (FCM) [37] as a premier fuzzy clustering algorithm was proposed by Bezdek in 1981. To omit the forcible assignment of samples to only one cluster, a membership degree matrix is considered with entries in the range $[0,1]$ which indicates the degree of membership of each sample to each cluster. Hence, the samples located in the boundaries of multiple classes can partially be a member of each cluster. In FCM, an iterative assignment procedure is adopted for minimizing the objective function, which is designed based on the membership degrees of samples to the cluster centers and their relative distance. The original FCM algorithm is sensitive to the initial location of cluster centers [38], and the use of relative distance metric makes it sensitive to noise [39]

Over the years, many methods have been proposed to deal with these drawbacks. To enhance the physical interpretation of the objective function of FCM clustering, EFCM [40] is proposed by Li and Mukaidono, in which the fuzzifier m is replaced with the entropy coefficient of membership degrees. PCM [39], as a probabilistic version of FCM, solves the noisy data sensitivity of FCM by relaxing the constraint of "sum of membership degrees of each sample to all the clusters should be equal to

1". FPCM [41] combines both fuzzy and probabilistic point of views followed in FCM and PCM. Appending the typicality metric based on the absolute distance to the objective function of FCM solved the problem of emerging coincident clusters, which was a challenge in PCM. PFCM [42] relaxed the constraint "sum of the typicality values of all data samples to each cluster must be equal to 1," which is the main challenge of FPCM encountering big datasets.

Dave and Krishnapuram [43] proposed an NC clustering approach to handle noisy data samples in the FCM clustering procedure. NC assumes a separate cluster center representing the noise/outlier cluster and sets a constant distance value for all the samples concerning this center. BCFCM [44] is a modification of FCM for image clustering designed specifically to segment MRI medical images. Neighborhood information of samples is embedded as a new term in the FCM with the aim of robustness against noisy points.

Robust Fuzzy *c*-means (FCM-σ) [45] replaces the standard distance metric of FCM with a normalized distance metric to handle non-hyperspherical clusters and clusters with uneven density distributions. The same authors also proposed the Kernel Fuzzy *c*-means (KFCM) [45] approach, which is a kernelized version of FCM. As FCM applies to linearly separable data, KFCM is presented to deal with non-linear data structures by transforming the feature space to a higher dimension. Robust Kernel Fuzzy *c*-means (KFCM-σ) [46] is an improvement for KFCM, which is suitable for non-hyperspherical data structures by normalizing the distance metric of KFCM.

REFCM [47] developed by Zarinbal et al., enhanced the FCM by considering a new term of the relative entropy, which aims to minimize the intra-cluster similarity. Size-insensitive integrity-based Fuzzy *c*-means (siibFCM) [48] is proposed to solve the problem of pulling centers of trivial clusters by the large cluster, which was a challenge in FCM. siibFCM is formulated in a manner that can cope with unequal cluster sizes.

In 2019, FWCW-FCM [35] algorithm was developed in which a feature weighting technique is applied to the objective function of FCM, along with cluster weight learning. Weights are adaptively determined during the iterative clustering process. This modified objective function solves the initialization sensitivity challenges and the features' equal importance assumption. Other version of this algorithm was proposed in [12,38,49–52]

Salar Askari [53] developed a revised version of FCM, namely RFCM, to cope with non-uniform data distribution, noisy data points, and disparate clusters. RFCM consists of two subsequent optimization steps. The objective function of the first step results in the initial cluster centers, which are detected in a size-insensitive manner. The second step is fed with the central points chosen by the first step and outputs final cluster centers in a noise-resistant form.

In 2021, a unified form of FCM was designed, which supports distributed implementation as a vital advantage to apply to big data cases [13]. LPFCM [54] is an extension of FCM designed for high-dimensional feature spaces. By retaining the locality of structures, a projection term is embedded directly in the objective function of FCM, which ensures the adaptation of LPFCM to high dimensional data.

2.2. Semi-supervised fuzzy clustering

Semi-supervised clustering brings up the idea of utilizing the external class label information, available for a small subset of instances, to guide the clustering process [55]. In constraint-based semi-supervised approaches, the external label information can guide the clustering in two ways [56]: a) Using the labeled instances only for initializing the cluster centers, b) Injecting the class label information to the expectation-maximization steps of the clustering, e.g., designing a semi-supervised objective function [30]. SSFCM [25], as a baseline semi-supervised fuzzy clustering approach, reformulates the FCM object function by adding a semi-supervised term, which considers the consonance rate between the actual label and the obtained membership degree.

Pedrycz and Vukovich [57] propose a semi-supervised version of FCM that attempts to attune the number of clusters, which is adaptively equal to or greater than the number of class labels. In another work, Bouchachia and Pezdrycz [58] redesigned the SSFCM loss function, which deals with cases where the number of classes is less than the number of clusters.

Active Fuzzy Constrained Clustering (AFCC) [59] applies the violation costs of pairwise must-link and connot-link constraints to the lost function, along with a competitive agglomeration term which helps to detect the appropriate number of clusters. Yasunori et al. propose two semi-supervised fuzzy clustering approaches, namely sSFCM and eSFCM [60]. The former adds a supervised membership degree as a new parameter to the FCM, and the latter reformulates the EFCM objective function by using this supervised parameter.

SCAPC [61] is developed as a semi-supervised fuzzy clustering algorithm with pairwise constraints which is an extension of the AFCC algorithm. SPASC appends a new cost term to the objective function of AFCC, considering the sum of weighted distances of the must-link and cannot-link constraints as a penalty.

SSFCM-HPR [56] is a weighted semi-supervised version of FCM that embeds the prior label information as a constraint of membership degrees. Also, samples are distinguished based on their weights which indicates their representativeness in each cluster based on the distance criterion. SSFCM-HPR objective function parameters are optimized using the Hestense–Powell multiplier method.

Salehi et al. developed the SFC-ER [62] algorithm by adding a semi-supervised term (augmented regularization function) to the unsupervised EFCM objective function. Relative entropy divergence metric is adopted instead of the geometric distance measure in the semi-supervised term. In the same study, a kernelized version of SFC-ER is also proposed, named the SMKFC-ER algorithm.

Recently, several extensions for the SSFM approach have been proposed to embed a label consistency term objective function. In 2018, LHC-S3FCM [30] was presented, which considers local homogeneous consistency as a criterion. The newly added term results in the convergence of clustering predictions of each labeled sample with its homogeneous local neighbors. Moreover, the Gaussian kernel similarity measure is used to specify the neighboring relationship between the samples.

S3FCM [28] is proposed with partial distrust of the external class label knowledge. At each iteration, S3FCM predictions are compared with the unsupervised predictions of FCM, which are available as ground-truth information. Any discrepancy in the predictions of samples is considered a penalty. CS3FCM [29] is a confidence-weighted S3FCM that enhances the LHC-S3FCM efficiency by applying a weight factor to the labeled instances that represents their risk/confidence level. The adopted weighting scheme works based on the agreement between the label predictions and the FCM membership values. AS3FCM [31] is an improvement for CS3FCM, which calculates the safety level of labeled instances adaptively along the clustering process instead of using FCM hints.

To lower the computational time, and deal with a noisy data point, TS3FCM [32] proposed a Trusted S3SFCM approach which redesigned a new FCM objective function based on the local neighborhood of the labeled samples. The clustering ground truth information of the labeled samples and the initial centers are obtained based on this low-cost step. Accordingly, the ground-truth of the unlabeled instances are determined based on only one FCM iteration. Tuan et al. [33] propose a new version of TS3FCM based on the multiple fuzzifiers concept, namely TS3MFCM. The semi-supervised objective function is reformulated by considering a different fuzzifier for each sample, which is determined based on the sample and its local neighborhood. On the other hand, MCSSFC-C [63] is proposed as another improvement for TS3FCM, which considers different fuzzifier for each cluster instead of samples. The fuzzifier value of each cluster is determined based on the FCM membership values of its samples, which indicates that the number of uncertain samples in a cluster represents its uncertainty.

3. Preliminaries

In this section, the prerequisite concepts of this paper are discussed briefly. Sections 3.1 and 3.2 cover the SSFCM, and FWCW-FCM baseline approaches, respectively.

3.1. SSFCM

SSFCM [25] was proposed by Pedrycz and Waletzky in 1997 as a semi-supervised extension of FCM, which enhances the clustering procedure by injecting the class labels' prior knowledge into the FCM. In SSFCM, the data is considered a union of N_l labeled instances, and $N_u = N - N_l$ unlabeled instances, where $N_l \ll N_u$, and N refers to the count of all data instances. Adopting the same principles as FCM, a membership degree matrix $\boldsymbol{U} = [u_{nk}]_{K\times N}$ is considered, which indicates the membership degree of the nth sample to the kth cluster, and obeys the same membership constraints as FCM ($u_{nk} \in [0\ ,\ 1]$, $\sum_{k=1}^{K} u_{nk} = 1$). Accordingly, the objective function of the SSFCM approach is defined as:

$$J_{SSFCM} = \sum_{n=1}^{N}\sum_{k=1}^{K} {u_{nk}}^2 d_{nk}^2 + \alpha \sum_{n=1}^{N}\sum_{k=1}^{K} (u_{nk} - b_n f_{nk})^2 d_{nk}^2$$

$$u_{nk} \in [0,1],\ \sum\nolimits_{k=1}^{K} u_{nk} = 1,\ \text{where } 1 \le n \le N \text{ and } 1 \le k \le K \tag{1}$$

In Eq. (1), the second term represents the prior supervised knowledge, in which, $\boldsymbol{B} = [b_n]$ is an N-dimensional label indicator vector, with $b_n = 1$ for labeled samples, and $b_n = 0$ for unlabeled samples. $\boldsymbol{F} = [f_{nk}]$ corresponds to the fuzzy degrees of the labeled samples in which $f_{nk} = 1$ if sample n belongs to the class k, and 0 otherwise. Finally, $\alpha \ge 0$ is the regularization parameter.

Adopting the Lagrange multipliers approach, the membership degree of sample n in cluster k (u_{nk}) is computed as follows.

$$u_{nk} = \frac{1}{1+\alpha}\left[\frac{1 + \alpha\left(1 - b_n \sum\limits_{h=1}^{K} f_{nh}\right)}{\sum\limits_{h=1}^{K} d_{nk}^2 / d_{nh}^2}\right] + \alpha b_n f_{nk}, \tag{2}$$

$\forall n = 1, \dots, N$ and $\forall k = 1, \dots, K$

For cluster k, the cluster center c_k is computed according to Eq. (3).

$$c_k = \frac{\sum\limits_{n=1}^{N} u_{nk}^2 x_n + \alpha \sum\limits_{n=1}^{N} (u_{nk} - b_n f_{nk})^2 x_n}{\sum\limits_{n=1}^{N} u_{nk}^2 + \alpha \sum\limits_{n=1}^{N} (u_{nk} - b_n f_{nk})^2}, \quad \forall k = 1, \dots, K \tag{3}$$

3.2. FWCW-FCM

FWCW-FCM [35] is proposed as an improved version of FCM which distinguishes between different data features concerning the clusters. Specifically, an adaptive feature-weighting and cluster-weighting mechanism are adopted to remedy the initial center sensitivity problem of FCM. Thus, the objective function of FCM is reformulated as Eq. (4):

$$J_{FWCW-FCM} = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M} {u_{nk}}^{\sigma} w_{km}^{q} z_k^{p} d^2(x_{nm} - c_{km}) \tag{4}$$

For the proposed objective function, constraints are presented in Eq. (5).

$$\begin{aligned} &u_{nk} \in [0,1],\quad \sum\nolimits_{k=1}^{K} u_{nk} = 1,\ \text{where } 1 \le n \le N \text{ and } 1 \le k \le K; \\ &w_{km} \in [0,1],\quad \sum\nolimits_{m=1}^{M} w_{km} = 1,\ \text{where } 1 \le k \le K \text{ and } 1 \le m \le M; \\ &z_k \in [0,1],\quad \sum\nolimits_{k=1}^{K} z_k = 1,\ \text{where } 1 \le k \le K \text{ and } 0 \le p < 1. \end{aligned} \tag{5}$$

In Eq. (4), N, M, and K indicate the data sample count, feature space dimension, and cluster count, respectively. $\boldsymbol{U} = [u_{nk}]_{K\times N}$ is a membership matrix in which entry u_{nk} corresponds to the membership degree of the nth sample to the kth cluster center. In the $\boldsymbol{C} = [c_{km}]_{K\times M}$ matrix, c_{km} refers to the mth feature in the kth cluster. Moreover, in the $\boldsymbol{W} = [w_{km}]_{K\times M}$ matrix, w_{km} indicates the weight of the mth feature in the kth cluster, and in $\boldsymbol{z} = [z_k]_{1\times K}$ vector, entry z_k indicates the kth cluster weight. Also, the term $d^2(x_{nm} - c_{km})$ indicates a non-Euclidean distance metric.

Based on the closed-form solution obtained for Eq. (4), formulas for u_{nk}, c_{km}, w_{km}, and z_k are presented in Eqs. (6, 7, 8, and 9), respectively.

$$u_{nk} = \frac{1}{\sum\limits_{l=1}^{K}\left[\frac{z_k^p \sum_{m=1}^{M} w_{km}^q d^2(x_{nm} - c_{km})}{z_l^p \sum_{m=1}^{M} w_{lm}^q d^2(x_{nm} - c_{lm})}\right]^{\frac{1}{(\sigma-1)}}} \tag{6}$$

$$c_{km} = \frac{\sum_{n=1}^{N} {u_{nk}}^{\sigma}\left(\exp\left(-\gamma_m (x_{nm} - c_{km})^2\right)\right)(x_{nm})}{\sum_{n=1}^{N} {u_{nk}}^{\sigma}\left(\exp\left(-\gamma_m (x_{nm} - c_{km})^2\right)\right)} \tag{7}$$

$$w_{km} = \frac{1}{\sum\limits_{s=1}^{M}\left[\frac{\sum_{n=1}^{N} {u_{nk}}^{\sigma} d^2(x_{nm} - c_{km})}{\sum_{n=1}^{N} {u_{nk}}^{\sigma} d^2(x_{ns} - c_{ks})}\right]^{\frac{1}{q-1}}} \tag{8}$$

$$z_k = \frac{1}{\sum\limits_{l=1}^{K}\left[\frac{\sum\limits_{n=1}^{N}\sum\limits_{m=1}^{M} {u_{nk}}^{\sigma} w_{km}^q d^2(x_{nm} - c_{km})}{\sum\limits_{n=1}^{N}\sum\limits_{m=1}^{M} {u_{nl}}^{\sigma} w_{lm}^q d^2(x_{nm} - c_{lm})}\right]^{\frac{1}{p-1}}} \tag{9}$$

4. Proposed SSFCM-FWCW approach

Due to the difficulties described for the SSFCM method, we apply cluster weighting and feature weighting techniques in the suggested approach to increase the efficiency of the SSFCM algorithm. The more significant features gain greater weight with the local weighting of the features. Weighting the clusters, on the other hand, causes balanced clusters during the clustering. The proposed method's objective function is as follows:

$$\begin{aligned} F(\mathbf{U}, \mathbf{C}, \mathbf{W}, \mathbf{z}) = &\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M} {u_{nk}}^2 w_{km}^q z_k^p d^2(x_{nm} - c_{km}) + \alpha \sum_{n=1}^{N}\sum_{k=1}^{K} \\ &\times \sum_{m=1}^{M} (u_{nk} - b_n f_{nk})^2 w_{km}^q z_k^p d^2(x_{nm} - c_{km}) \end{aligned} \tag{10}$$

In Eq. (10), N, M, and K indicate the data sample count, feature space dimension, and cluster count, respectively. $\boldsymbol{U} = [u_{nk}]_{K\times N}$ is a membership matrix in which entry u_{nk} corresponds to the membership degree of the nth sample to the kth cluster center. In the $\boldsymbol{C} = [c_{km}]_{K\times M}$ matrix, c_{km} refers to the mth feature in the kth cluster. Moreover, in the $\boldsymbol{W} = [w_{km}]_{K\times M}$ matrix, w_{km} indicates the weight of the mth feature in the kth cluster, and in $\boldsymbol{z} = [z_k]_{1\times K}$ vector, entry z_k indicates the kth cluster weight. $\alpha \ge 0$ is a regularization parameter for the second part of the objective function that exploits class information, $\boldsymbol{B} = [b_n]$ is a label indicator vector with the length of N, where $b_n = 1$ if x_{nm} is labeled and $b_n = 0$, otherwise; $\boldsymbol{F} = [f_{nk}]$ denotes the fuzzy degrees of the labeled samples where $f_{nk} = 1$ if sample n belongs to class k and $f_{nk} = 0$ otherwise. The parameter q is in the range $q < 0$ and $q > 1$, and the parameter p is within the range $0 \le p < 1$. The values of the p and q parameters are determined by the user. Also, the term $d^2(x_{nm} - c_{km})$ indicates a non-Euclidean distance metric and is defined as follows:

$$1 - \exp(-\gamma_m (x_{nm} - c_{km})^2), \quad (11)$$

where γ_m shows the inverse of the variance of the mth feature of x dataset:

$$\gamma_m = \frac{I}{\mathrm{var}_m},\ \mathrm{var}_m = \sum_{n=1}^{N} \frac{(x_{nm} - \overline{x}_m)^2}{N},\ \overline{x}_m = \sum\nolimits_{n=1}^{N} \frac{x_{nm}}{N},\ I = (0, 1]. \quad (12)$$

For Eq. (10), the following constraints should be addressed:

$$\begin{aligned} &u_{nk} \in [0,1], \quad \sum\nolimits_{k=1}^{K} u_{nk} = 1, \text{ where } 1 \le n \le N \text{ and } 1 \le k \le K; \\ &w_{km} \in [0,1], \quad \sum\nolimits_{m=1}^{M} w_{km} = 1, \text{ where } 1 \le k \le K \text{ and } 1 \le m \le M; \\ &z_k \in [0,1], \quad \sum\nolimits_{k=1}^{K} z_k = 1, \text{ where } 1 \le k \le K \text{ and } 0 \le p < 1. \end{aligned} \quad (13)$$

Based on the closed-form solution obtained for Eq. (10), formulas for u_{nk}, c_{km}, w_{km}, and z_k are presented in Eqs. (14, 15, 16, and 17), respectively (see Appendix A for more details about updating formulas):

$$u_{nk} = \frac{1}{1+\alpha}\left[\frac{1+\alpha(1-b_n)}{\sum\limits_{l=1}^{K}\frac{d_{nk}}{d_{nl}}} + \alpha b_n f_{nk}\right] \quad (14)$$

$$c_{km} = \frac{\sum_{n=1}^{N}(x_{nm})\left(\exp\left(-\gamma_m (x_{nm} - c_{km})^2\right)\right)\left(u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right)}{\sum_{n=1}^{N}\left(\exp\left(-\gamma_m (x_{nm} - c_{km})^2\right)\right)\left(u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right)} \quad (15)$$

The following principle should be considered when weighting features in each cluster: give a higher weight to a feature with a lower variance and a lower weight to a feature with a higher variance in the related cluster. In this principle, the meaning of variance is not its conventional and standard meaning, but the meaning of variance is the sum of intra-cluster distances. In Eq. (16), $\sum_{n=1}^{N}\left[u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right] d^2(x_{nm} - c_{km})$ is the variance (sum of intra-cluster distances) of the mth feature in the kth cluster.

$$w_{km} = \frac{1}{\sum\limits_{s=1}^{M}\left[\frac{\sum_{n=1}^{N}\left[u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right] d^2(x_{nm} - c_{km})}{\sum_{n=1}^{N}\left[u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right] d^2(x_{ns} - c_{ks})}\right]^{\frac{1}{q-1}}} \quad (16)$$

The following principle should be considered when weighting the clusters: assign a larger weight to a cluster with a larger SIWD (the sum of the intra-cluster weighted distances) and a smaller weight to a cluster with a smaller SIWD. In Eq. (17), $\sum_{n=1}^{N}\sum_{k=1}^{K} w_{km}^q\, d^2(x_{nm} - c_{km})\left[u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right]$ shows the SIWD (the sum of the intra-cluster weighted distances) for the kth cluster.

$$z_k = \frac{1}{\sum\limits_{l=1}^{K}\left[\frac{\sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} w_{km}^q d^2(x_{nm} - c_{km})\left[u_{nk}{}^2 + \alpha(u_{nk} - b_n f_{nk})^2\right]}{\sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} w_{lm}^q d^2(x_{nm} - c_{lm})\left[u_{nl}{}^2 + \alpha(u_{nl} - b_n f_{nl})^2\right]}\right]^{\frac{1}{p-1}}} \quad (17)$$

The purpose of weighting the clusters is summarized as follows:

1) Preventing the creation of unbalanced clusters during the clustering according to the definition of SIWD.

2) Taking into account the previous goal, reducing the algorithm's sensitivity to the selection of primary centers and preventing the creation of empty clusters.

4.1. Pseudo-code of the SSFCM-FWCW Algorithm

Algorithm 1. (see Appendix B) depicts the SSFCM-FWCW method's pseudo-code. To discover the suitable value of p, we use an iteration-based strategy with three parameters: $p_{\max}$, p_{step}, and p_{init}, similar to the method provided in [35]. We begin the clustering procedure by setting p to p_{init}. In each iteration, we increase p by p_{step} until we reach the $p_{\max}$. If an empty cluster or a cluster with only one sample appears, we reduce p by p_{step}. We select the u_{nk}, w_{km}, and z_{km} that correspond to the p in the previous iteration. The procedure is repeated until the difference between the objective function values is smaller than the threshold value (ε) in two consecutive iterations or the number of iterations approaches the $t_{\max}$.

We add a memory effect parameter to the weights to increase the algorithm's stability. Smoother transitions of values between iterations are achieved by utilizing this option. In other words, the weights from the prior iteration have a limited influence on the current update. The parameter β controls this transition (see Eqs. (18 and 19)).

$$\boldsymbol{W}^{(\mathrm{t})} = \beta \boldsymbol{W}^{(\mathrm{t}-1)} + (1-\beta)\boldsymbol{W}^{(\mathrm{t})} \quad (18)$$

$$\boldsymbol{z}^{(\mathrm{t})} = \beta \boldsymbol{z}^{(\mathrm{t}-1)} + (1-\beta)\boldsymbol{z}^{(\mathrm{t})} \quad (19)$$

where the parameter β is in the range [0,1).

The Matlab implementation source code of SSFCM-FWCW is now publicly accessible at https://github.com/Amin-Golzari-Oskouei/SSFCM-FWCW.

4.2. Analysis on the noise robustness of SSFCM-FWCW Algorithm

In this subsection, we examine the robustness property of SSFCM-FWCW. Let $[x_1, x_2, \ldots, x_n]$ denote an observed dataset, and θ is an unknown parameter to be estimated. According to robust statistical theory, the M-estimate of θ can be obtained by minimizing $\sum_{n=1}^{N}\rho(x_n, \theta)$, where ρ is an arbitrary function measuring the loss between x_i and θ. In a location estimate minimizing $\sum_{n=1}^{N}\rho(x_n, \theta)$, the M-estimator is derived by solving the following equation:

$$\sum_{n=1}^{N} \varphi(x_n, \theta) \quad (20)$$

where $\varphi(x_n, \theta) = \frac{\partial\rho(x_n, \theta)}{\partial\theta}$. In the M-estimator, the influence of individual observations on the estimate closely aligns with the character of the function φ. If φ is unbounded, outliers and noise exert a greater influence on the estimate, resulting in reduced accuracy. Conversely, if φ is bounded, the influence of outliers and noise is diminished, leading to a relatively accurate estimate.

For instance, if we set $\rho(x, \theta) = 1 - \exp\left(-\gamma(x, \theta)^2\right)$, then the corresponding $\varphi(x, \theta) = \frac{2\gamma(x, \theta)}{\exp\left(\gamma(x, \theta)^2\right)}$ and $\lim\limits_{x \to -\infty}\varphi(x, \theta) = \lim\limits_{x \to \infty}\varphi(x, \theta) = 0$. Hence, φ is bounded, indicating that the estimate of θ is robust when using this ρ function. Conversely, if we take $\rho(x, \theta) = (x - \theta)^2$, Then the corresponding $\varphi(x, \theta) = -2(x - \theta)$ is unbounded, and the Euclidean distance often leads to a noise-sensitive algorithm.

The objective function of SSFCM-FWCW can be reformulated as:

$$F(\boldsymbol{U}, \boldsymbol{C}, \boldsymbol{W}, \boldsymbol{z}) = \sum_{k=1}^{K}\sum_{m=1}^{M} w_{km}^q z_k^p \sum_{n=1}^{N} u_{nk}{}^2 d^2(x_{nm} - c_{km}) + \alpha \sum_{k=1}^{K}\sum_{m=1}^{M} w_{km}^q z_k^p \times \sum_{n=1}^{N}(u_{nk} - b_n f_{nk})^2 d^2(x_{nm} - c_{km}) \quad (21)$$

By considering $d^2(x_{nm} - c_{km}) = 1 - \exp(-\gamma_m (x_{nm} - c_{km})^2)$, the terms $\sum_{n=1}^{N} u_{nk}{}^2 d^2(x_{nm} - c_{km})$ and $\sum_{n=1}^{N}(u_{nk} - b_n f_{nk})^2 d^2(x_{nm} - c_{km})$ of the above equation can be viewed as an M-estimate of the kth clustering centroid using the kth cluster as the observed dataset, with $1 - \exp\left(-\gamma_m (x_{nm} - c_{km})^2\right)$ serving as the loss function in the mth dimension subspace.

Based on the analysis above, we conclude that this estimate is relatively robust against noise. Therefore, from the perspective of robust statistical theory, the proposed method constitutes a robust clustering algorithm.

4.3. Analysis of SSFCM-FWCW Algorithm on bad initialization

As mentioned in the introduction, k-partitioning algorithms are highly sensitive to initialization. This sensitivity applies to algorithms with and without feature weighting. While feature weighting is indeed necessary, it alone may not suffice. Following a poor initialization, clusters with a high sum of intra-cluster weighted-feature distances (SIWDs) may merge, while those with low SIWDs may fragment into smaller clusters. Consequently, even if the dataset initially contains clusters with balanced SIWDs, the algorithm's execution may lead to the formation of clusters with unbalanced SIWDs. This scenario commonly occurs in k-means and fuzzy c-means-based algorithms.

Cluster weighting significantly mitigates this issue. The underlying principle of cluster weighting is straightforward: allocate greater weight to clusters with larger SIWDs, and lesser weight to clusters with smaller SIWDs. This principle is termed the "*cluster weighting principle*". By assigning a higher weight to clusters with substantial SIWDs, the formation of clusters with imbalanced SIWDs can be deterred.

Cluster weighting proves most effective when applied to datasets featuring clusters with nearly balanced SIWDs. This effectiveness stems from the emergence of clusters with balanced SIWDs during algorithmic restarts.

5. Experiments

The effectiveness of the SSFCM-FWCW is assessed in this section. The outcomes are contrasted with the most recent and baseline algorithms listed below:

- SSFCM [25]
- FCM [37]
- Seeded- k-means [24]
- SFC-ER [62]
- SMKFC-ER [62]
- S3FCM [28]
- CS3FCM [29]
- AS3FCM [31]
- LHC-S3FCM [30]
- TS3FCM [32]
- FWCW-FCM [35]

Parameter ε and $t_{\max}$ in the implemented methods are set to 10^{-5} and 100, respectively. In the SSFCM-FWCW, the secondary parameters are set as follows: $p_{\max} = 0.5$, $p_{\text{init}} = 0$, $p_{\text{step}} = 0.01$.

5.1. The used datasets

We use 13 benchmark datasets to thoroughly examine the effects of the strategies presented in our method and assess how the proposed method stacks up against other methods. Table 1 contains the statistical details. Like similar methods in [28–31], we randomly choose 20% of each dataset as the labeled subset and the remaining 80% as the unlabeled subset.

Like the method proposed in [35], the parameter β is selected for each dataset from set {0.1, 0.2, 0.3}, and the parameter q is chosen for each dataset from group {-10, −8, −6, −4, 2, 4, 6, 10}. Also, for *Breast Cancer Wisconsin (Diagnostic)* and *Balance* datasets, the parameter α is set to 4, and for the rest, it is set to 1 or 2.

Table 1
Characteristics of the real-world dataset.

Dataset	Number of data points	Number of dimensions	Number of Classes
Australian Credit Approval (statlog)	690	14	2
Balance	625	4	3
Breast Cancer Wisconsin (Diagnostic)	569	32	2
Bupa (Liver Disorders Dataset)	345	7	2
Dermatology	366	33	6
Ecoli	336	8	8
Glass	214	10	6
Ionosphere	351	34	2
Iris	150	4	3
Pima Indians Diabetes	768	8	2
Statlog (heart)	270	13	2
Waveform	5000	21	3
wine	178	13	3

5.2. Evaluation criteria

Like the similar methods presented in [28–31,64], in the experiments, we use clustering accuracy (*ACC*) and performance improvement (*PI*) to show how well various approaches perform. To examine the performance of the proposed algorithm, the ground truth labels of the samples are compared with the labels assigned by the clustering algorithm. In other words, after running the clustering algorithm, each sample is assigned to a cluster with a corresponding class label. *ACC* measure is applied to the clustering results to figure out the proportion of samples that are correctly classified by the clustering algorithm. Moreover, the *PI* measure is employed to determine the extent to which the proposed algorithm outperforms other algorithms in accurately clustering the samples considering their actual class label.

5.2.1. Accuracy (ACC)

The predicted label for x_n and its ground truth is denoted by $\widetilde{y}_k$ and y_n, respectively. The clustering methods' accuracy is calculated as follows:

$$ACC = \frac{\sum_{n=1}^{N} \delta(y_n, \text{map}(\widetilde{y}_k))}{N} \tag{22}$$

where the delta function $\delta(x, y)$ has an output of 1 if $x = y$ and 0 otherwise. The Kuhn-Munkres algorithm's permutation mapping function maps the label $\widetilde{y}_k$ to its equivalent by finding the best one-to-one mapping of the predicted label and ground truth label [65].

5.2.2. Performance improvement (PI)

This criterion represents the percentage of improvement obtained by the proposed method on each dataset compared to the other algorithms' best result. The performance improvement criterion is calculated, using the following equation:

$$PI = \frac{ACC_{\text{our}} - ACC_{\text{other}}}{ACC_{\text{other}}} \times 100 \tag{23}$$

where, ACC_{our} is the accuracy of the proposed method, and ACC_{other} is the other algorithms' best result.

5.3. Experiment 1: analyse the feature weighting technique

In this experiment, we analyze the impact of the feature weighting schema on the final clustering outcomes. We are interested in determining whether the SSFCM-FWCW correctly assigns weights to features. The produced weights by our proposed approach are compared to the baseline FWCW-FCM method. To analyze the importance of each feature, we depict the *Iris* dataset in Fig. 1.

As illustrated in Fig. 1, all features in Cluster1 are considered equally important. However, features 3 and 4 significantly impact the formation of optimal clusters. As shown in part (f) of Fig. 1, these two features are discriminative. In Cluster1, with the help of these two features, intra-cluster variance decreases, and all samples of this cluster are well separated from other samples. Furthermore, the variance of samples in Feature3 is lower than that of samples in Feature4. So among these two features, Featue3 is more important than Feature4.

In Cluster2, Feature1 and Feature2 don't have helpful information. Cluster2 and Cluster3 samples are indistinguishable when considering these two features, as demonstrated in Fiqure2 part (a). In Cluster2, Features 3 and 4 are almost equally significant. This is evident in Fig. 1, part (f). The variance of samples in Feature3 is nearly identical to that of samples in Feature4.

Like Cluster2, Cluster3 lacks useful information in Features 1 and 2 (see part (a) of Fig. 2). In contrast to Cluster2, in Cluster3, Feature3 is more significant than Feature4. Samples on Feature3 have a lower variance than samples on Feature4 for this cluster.

We assess the weights of the features obtained by each method when the clustering process is complete. The values of the weights generated for each feature in various clusters using two methods are compared in Fig. 2. For our approach, the weight differences for the different features are consistent with the assumption based on Fig. 1 and the preceding paragraphs.

According to Fig. 2, the proposed algorithm gives the third and fourth features in Cluster1 more weight than the FWCW-FCM algorithm while providing the first and second features less weight. In Cluster2, the proposed algorithm, in contrast to the FWCW-FCM algorithm, provides the third and fourth features equal weights. The proposed algorithm gives the third feature in Cluster3 more weight when compared to the FWCW-FCM algorithm. This demonstrates unequivocally that our algorithm correctly applies local feature weighting. Although FWCW-FCM can distinguish between important and non-important features, it performs poorly in weighting features. The third and fourth features in

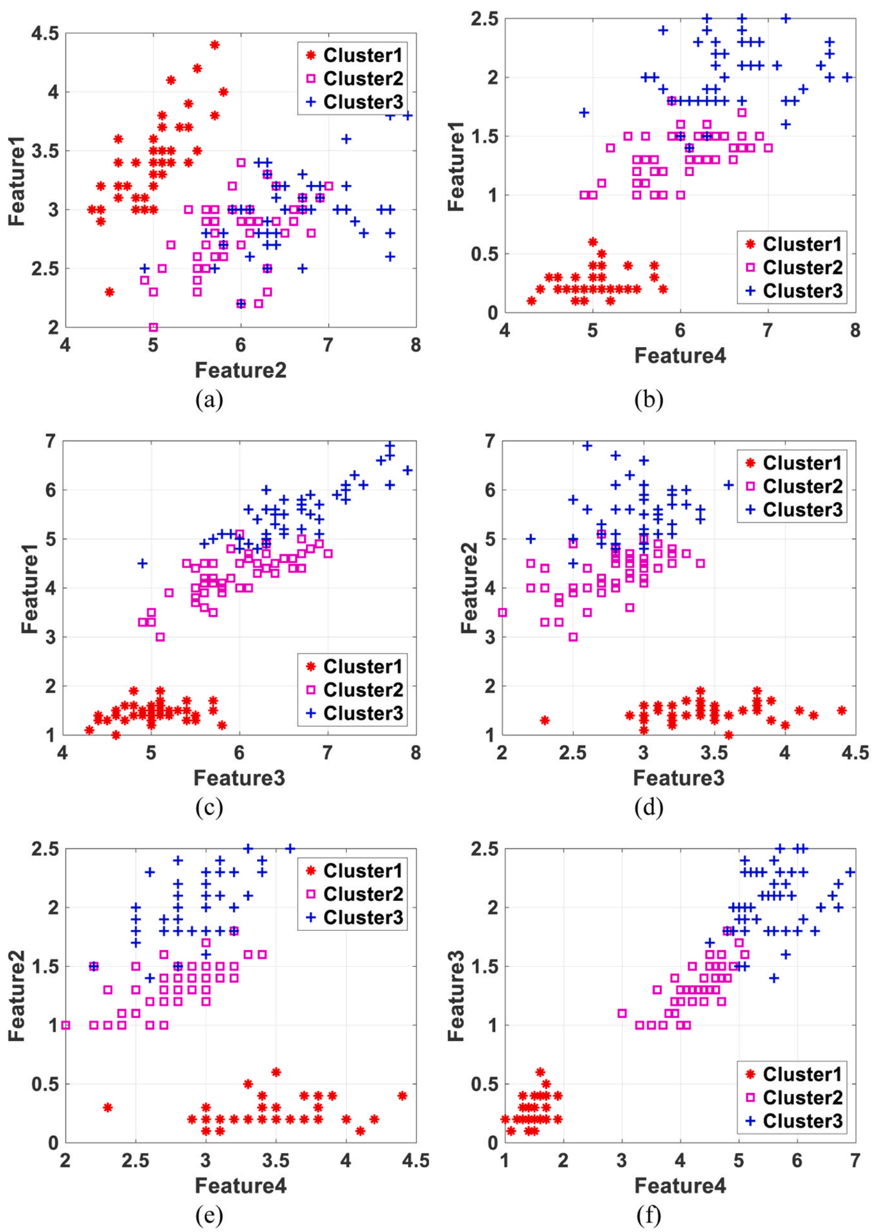

Fig. 1. Visualization *Iris* dataset.

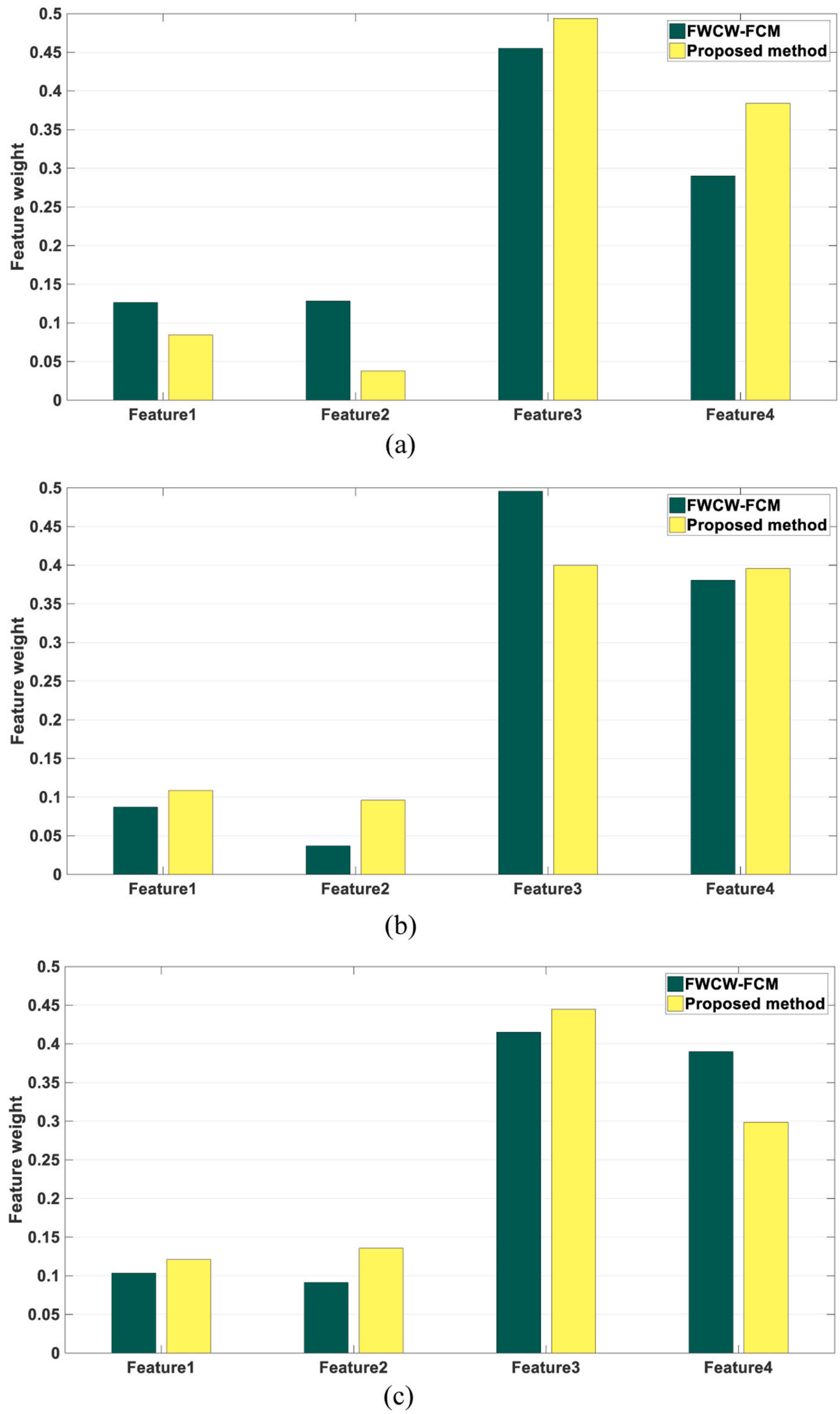

Fig. 2. Weights of the features assigned in the *Iris* dataset for (a) Cluster1, (b) Cluster2, and (c) Cluster3.

cluster2 are nearly equally significant. However, the third feature is given more weight by the FWCW-FCM method. This is because the significance of the features in this cluster was improperly understood. As a result, in FWCW-FCM, the feature weight does not accurately reflect the importance of each feature in each cluster.

5.4. Experiment 2: effect of feature weighting

In this experiment, we compare the performance of the proposed algorithm twice—once with and once without feature weighting—to investigate the impact of feature weighting on the clustering of real-world datasets. The average outcomes of 10 iterations of the algorithms on each dataset are shown in Fig. 3.

Fig. 3 shows how the proposed algorithm works better using the feature weighting approach. In all testing datasets, as demonstrated in this figure, the accuracy rate of the proposed method is enhanced by an average of 11% when using a weighting scheme compared to the mode without feature weighting. This is made more evident because different features in various datasets, such as *Iris*, *Ecoli*, *Dermatology*, and *Glass*, have relatively different levels of importance.

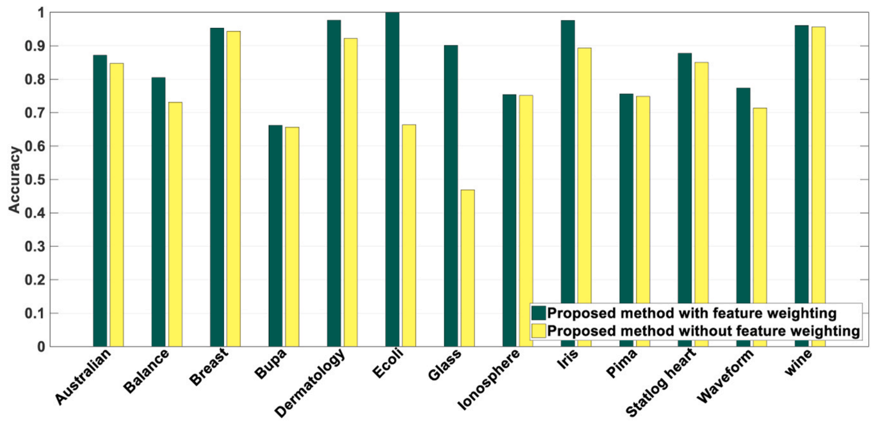

Fig. 3. The impact of feature weighting on the clustering quality.

5.5. Experiment 3: effect of cluster weighting

In this experiment, we compare the performance of the proposed algorithm twice—once with and once without cluster weighting—to investigate the impact of cluster weighting on the final clustering of real-world datasets. The average outcomes of 10 iterations of the algorithms on each dataset are shown in Fig. 4.

Fig. 4 shows how the proposed algorithm works better using the cluster weighting approach. In all testing datasets, as demonstrated in this figure, the accuracy rate of the proposed method is enhanced by an average of 5% when using a weighting scheme compared to the mode without cluster weighting. This is made more evident by the fact that different clusters in various datasets, such as *Balance*, *Ecoli*, *Dermatology*, and *Glass*, have relatively different levels of importance.

The results of Figs. 3 and 4 demonstrate that the weighing of features rather than the weighting of clusters significantly influences the creation of optimal clusters. Although the average results are improved by employing cluster weighting schemes, the performance of the suggested method is similar to the without cluster weighting mode for several datasets, such as *Breast* and *Australian*. This is so balanced clusters can be created using the cluster weighting technique. As a result, this method generates better clusters when the dataset contains balanced natural groups. However, when unbalanced natural groupings are present in the dataset, which is a regular occurrence in practice, this strategy may prove troublesome because it will prevent the clustering process from revealing the natural structure of the data.

5.6. Experiment 4: effect of labeled data

To investigate the behavior of the proposed algorithm for the percentage of labeled data, we test the proposed method for different ratios in this experiment. For this purpose, we try the proposed method with the rate of labeled data from 5% to 50% with step 5%.

Fig. 5 compares the effectiveness of our suggested technique for raising the percentage of labeled data in terms of accuracy. With more labeled data, the semi-supervised proposed model performs better. The degree of hesitation diminishes as label information is added, and as a result, the proposed model tends to perform with almost the same accuracy for high percentages.

As the number of labeled data increases, cluster centers more precisely point to the class's more real centroids. As a result, the proportion of labeled data greatly influences the clustering process. The sample distribution of classes in the *Balance* dataset is very unbalanced, with class 1 having a distribution of 6% and classes 2 and 3 having a distribution of 47%. The selected labeled data may not cover all classes because samples were randomly chosen. As a result, an increase in the number of labeled data does not entirely increase the accuracy of this dataset.

5.7. Experiment 5: SSFCM-FWCW vs. State-of-the-art methods

This experiment compares the SSFCM-FWCW's performance with other clustering techniques. Table 2 shows the results of various methods. The accuracy rate for each technique is provided in the table. In this table, the performance improvement percentage of the obtained results by the SSFCM-FWCW, compared to the best result of the competing semi-supervised algorithms, is also reported. The outcomes of alternative techniques have been directly cited from the pertinent publications.

The performance of SSFCM-FWCW is the best among all datasets, as

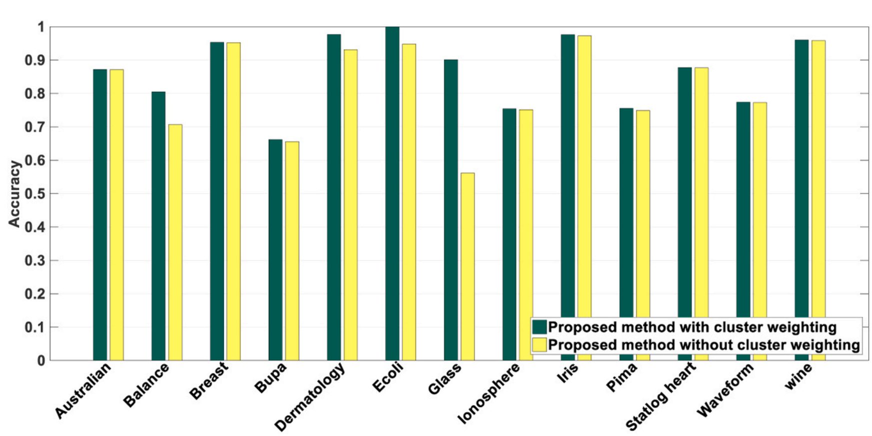

Fig. 4. The impact of cluster weighting on the clustering quality.

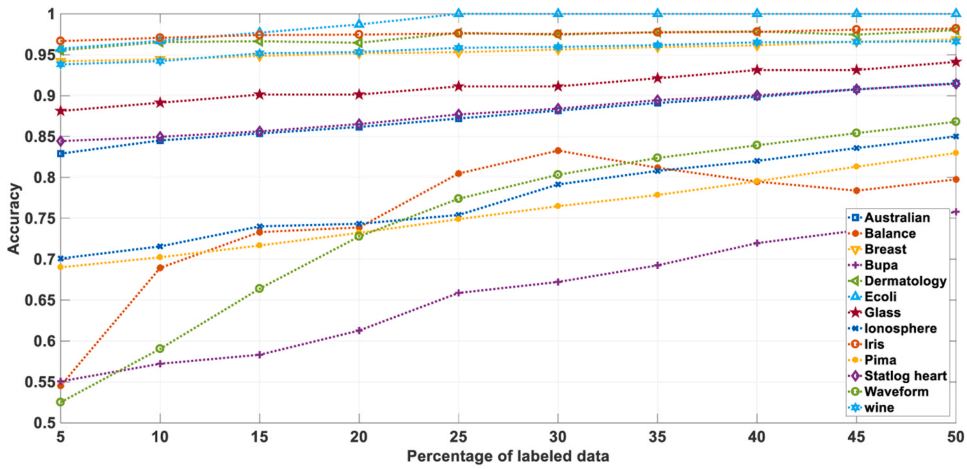

Fig. 5. The behavior of the proposed algorithm for the percentage of labeled data.

Table 2
SSFCM-FWCW vs. State-of-the-art methods.

	Australian	Balance	Breast Cancer	Bupa	Dermatology	Ecoli	Glass	Ionosphere	Iris	Pima	Statlog (heart)	Waveform	Wine
[37]	56	51	85	52	78	84.6	68.5	70.9	89.3	68	79.7	49	67
[24]	56.5	53	85	46	94	-	-	-	89.1	67.8	79	50	68
[25]	59.7	53.4	85.2	52.1	91	93.7	80.8	71.2	90.5	69.7	81	54	68
[62]	-	-	86.1	52.2	60	96.4	83.2	-	93.3	71.8	82.5	77	-
[62]	-	-	-	-	-	97	87.8	-	96.7	-	80.1	-	-
[28]	60.9	-	85.2	-	-	-	-	-	90	-	80.1	54	-
[29]	63.5	52.6	86.2	56	91	-	-	54	94.8	72	81.5	76	69
[31]	-	74	86.8	57.2	95.5	-	-	-	-	71.9	83	76.1	-
[30]	63	79	86.2	56.5	94	-	-	72	95	72.5	81.8	76.5	70
[32]	-	56.8	-	-	66.9	-	-	-	76.4	-	77.2	-	62.1
Ours	**87.2**	**80.48**	**95.3**	**66.2**	**97.66**	**100**	**90.13**	**75.41**	**97.6**	**75.59**	**87.77**	**77.39**	**96.06**
PI	**37.3**	**1.87**	**9.79**	**15.7**	**2.26**	**3.09**	**2.65**	**4.73**	**0.93**	**4.26**	**5.74**	**0.5**	**37.22**

shown in Table 2. The findings show that SSFCM-FWCW performs better than other cutting-edge techniques in this area. LHC-S3FCM [30] and CS3FCM [29] methods perform reasonably well after SSFCM-FWCW, respectively. The improvement percentages are high for some datasets, including *Breast Cancer*, *Wine*, *Bupa*, and *Australian*. This demonstrates the significant impact of feature and cluster weighting on creating better clusters.

For a thorough analysis of the proposed method with baseline algorithms, Fig. 6 displays the outcomes for various combinations of the proposed method and the two baseline methods, SSFCM [25] and FWCW-FCM [35].

This figure illustrates how SSFCMFWCW performs better than baseline techniques across all tested datasets. This demonstrates the superior ability of SSFCM-FWCW to improve semi-supervised clustering and further supports the findings of our experiment. In the following, all approaches were compared with each other to evaluate each algorithm thoroughly.

5.7.1. SSFCM-FWCW vs. SSFCM

As shown in Fig. 6, all versions of SSFCM-FWCW produce better results than SSFCM [25] for all datasets, except for *Ecoli* and *glass*. The average accuracy of SSFCM-FWCW is 86.68%, while SSFCM is 73.10%. This demonstrates how the proposed algorithm improves the performance of the base SSFCM algorithm. The proposed method outperforms

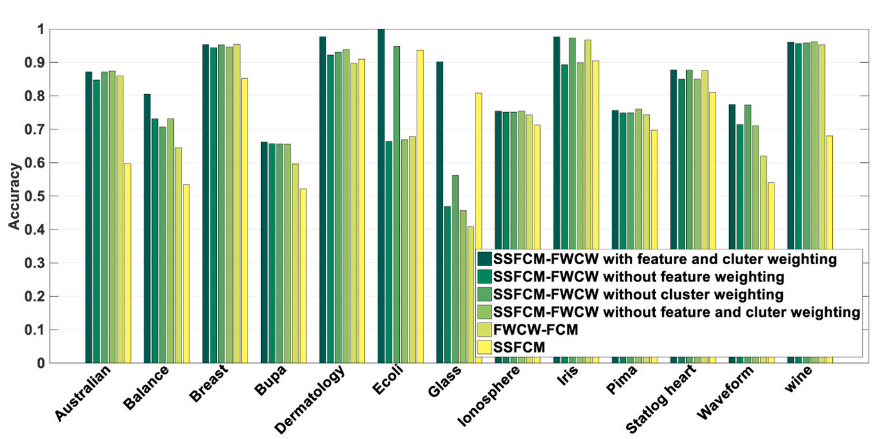

Fig. 6. SSFCM-FWCW vs. baseline methods.

SSFCM even when cluster and feature weighting are not used. This is because the proposed method uses a non-linear distance criterion.

5.7.2. SSFCM-FWCW vs. FWCW-FCM

Almost all versions of SSFCM-FWCW produce better results than FWCW-FCM [35] for all datasets. The average accuracy of SSFCM-FWCW is 86.68%, while FWCW-FCM is 76.43%. This demonstrates how the proposed algorithm outperforms the FWCW-FCM algorithm's baseline performance. It is easy to see why the proposed method is better than FWCW-FCM. The proposed method is based on semi-supervised learning, while method FWCW-FCM is entirely unsupervised.

5.7.3. SSFCM vs. FWCW-FCM

Comparing these two methods is very important. Method SSFCM is based on semi-supervised learning, and the approach FWCW-FCM is entirely unsupervised. Method SSFCM is expected to produce better results because it has information from some labels in the dataset. But this method is, on average, 3% lower than method FWCW-FCM in terms of accuracy. This observation shows the importance of weighting features and clusters. Although method FWCW-FCM does not have a mechanism to manage labeled samples, it has produced better results with the help of weighting features and clusters.

6. Real-world applications of proposed method

Proposed Semi-supervised clustering, which combines labeled and unlabeled data, has several potential real-world applications across various domains. Here are some use cases where this method can be beneficial:

Anomaly Detection: In cybersecurity, proposed semi-supervised clustering can help detect unusual patterns or anomalies in network traffic by leveraging both labeled normal data and unlabeled potentially malicious data. This approach can improve the accuracy of anomaly detection systems and reduce false positives.

Customer Segmentation: In marketing and customer relationship management, proposed semi-supervised clustering can be used to segment customers based on their behavior, preferences, and interactions with a company. By incorporating both labeled data (e.g., customer demographics) and unlabeled data (e.g., browsing history), businesses can gain deeper insights into customer segments and tailor their marketing strategies accordingly.

Medical Image Analysis: In healthcare, proposed semi-supervised clustering can assist in the analysis of medical images for tasks such as tumor detection or disease classification. By combining labeled images with a large set of unlabeled images, this method can help improve the accuracy of automated image analysis systems and aid healthcare professionals in making more accurate diagnoses.

Fraud Detection: In the financial sector, proposed semi-supervised clustering can be applied to detect fraudulent activities in transactions by utilizing both labeled fraudulent transactions and a large volume of unlabeled transaction data. This approach can help financial institutions identify potential fraud patterns and take proactive measures to prevent financial losses.

Text Clustering: In natural language processing, proposed semi-supervised clustering can be used to group similar documents or texts based on their content. By leveraging a combination of labeled text samples (e.g., categorized documents) and unlabeled text data, this method can assist in tasks such as document categorization, sentiment analysis, and information retrieval.

Social Network Analysis: In social media and network analysis, proposed semi-supervised clustering can be used to identify communities or clusters of users with similar interests or behaviors. By utilizing labeled user profiles or connections along with a vast amount of unlabeled social network data, this method can help uncover hidden patterns and relationships within social networks, enabling targeted marketing strategies or content recommendations.

Genomics and Bioinformatics: In biological data analysis, proposed semi-supervised clustering can be used to classify gene expression patterns or DNA sequences into distinct groups. By combining labeled gene expression data with unlabeled genomic data, this method can help identify genetic markers associated with specific diseases or traits, leading to advancements in personalized medicine and bioinformatics research.

Recommendation Systems: In e-commerce and personalized recommendations, proposed semi-supervised clustering can enhance the performance of recommendation systems by grouping users or items based on their preferences or purchase history. By utilizing labeled user ratings or item attributes along with unlabeled interaction data, this method can improve the accuracy of product recommendations and enhance the user experience on e-commerce platforms.

Overall, the flexibility and adaptability of semi-supervised clustering make it a valuable tool for various applications where labeled and unlabeled data are available, enabling more accurate and efficient data analysis and decision-making processes.

7. Conclusion

Semi-supervised clustering has become a hot research topic, and numerous types of research have been conducted in this area. Despite the performance enhancement caused by utilizing auxiliary class knowledge during the clustering, some issues are still not addressed thoroughly. Feature weight and cluster weight learning are neglected in most of the semi-supervised fuzzy clustering approaches, which disrupt the forming of an optimal clustering structure.

This paper's novel semi-supervised fuzzy objective function is formulated based on prior class knowledge, feature weight, and cluster weight learning. This approach assigns a different weight for each feature concerning each cluster adaptively during the clustering procedure. Various experimental analyses were performed to compare the performance of the proposed approach with the state-of-the-art methods in several benchmark datasets in terms of clustering accuracy. Besides, the effect of feature weighting and cluster weighting on the algorithm's efficiency was examined.

This is the first attempt to append feature weight and cluster weight learning in semi-supervised fuzzy clustering; there are some open rooms for future research in this era as follows. (1) Extending this objective function for the environments containing wrongly labeled instances among the provided prior knowledge. (2) Adopting a new kernel-based formulation will improve the proposed approach to be applicable to the non-linearly separable data.

CRediT authorship contribution statement

Amin Golzari Oskouei: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Data curation, Conceptualization. **Jafar Tanha:** Writing – review & editing, Validation, Supervision. **Negin Samadi:** Writing – review & editing, Methodology, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A

In this section, we describe how to obtain the updated equations for u_{nk}, c_{km}, w_{km}, and z_k. First, we write the Lagrange function of Eq. (10):

$$\widetilde{F}=\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M}{u_{nk}}^{2}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})+\alpha\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M}(u_{nk}-b_{n}f_{nk})^{2}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})-\delta\left(\sum\nolimits_{k=1}^{K}u_{nk}-1\right)-\psi\left(\sum\nolimits_{m=1}^{M}w_{km}-1\right)-\omega\left(\sum\nolimits_{k=1}^{K}z_{k}-1\right), \tag{A.1}$$

where δ، ψ و ω are the parameters of the Lagrange multiplier.

Theorem 1. If $\boldsymbol{C}$, $\boldsymbol{W}$, and $\boldsymbol{Z}$ are assumed to be fixed, $\boldsymbol{U}$ is a strict local minimum of $\boldsymbol{F}(\boldsymbol{U})$ if and only if $\boldsymbol{U}$ is calculated using Eq. (14).

Proof. Considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{U})$ to u_{nk} and set equal to 0, we obtain:

$$\begin{aligned}
&\frac{\partial\widetilde{\mathrm{F}}}{\partial u_{nk}}=0\\
&\sum_{m=1}^{M}2u_{nk}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})+\alpha\sum_{m=1}^{M}(2u_{nk}-2b_{n}f_{nk})w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})-\delta=0\\
&2u_{nk}\sum_{m=1}^{M}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})\;+\;\alpha2(u_{nk}-b_{n}f_{nk})\sum_{m=1}^{M}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})\;-\;\delta=0\\
&(2u_{nk}+2\alpha u_{nk}-2\alpha b_{n}f_{nk})\sum_{m=1}^{M}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})=\delta\\
&((2+2\alpha)u_{nk})=\frac{\delta}{\sum\limits_{m=1}^{M}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})}\;+\;2b_{n}f_{nk}\\
&u_{nk}=\frac{\dfrac{\delta}{\sum_{m=1}^{M}w_{km}^{q}z_{k}^{p}d^{2}(x_{nm}-c_{km})}+\;2\alpha b_{n}f_{nk}}{2+2\alpha}
\end{aligned} \tag{A.2}$$

now, considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{U})$ to u_{nl} and set equal to 0, we obtain δ:

$$\begin{aligned}
&\frac{\partial\widetilde{\mathrm{F}}}{\partial u_{nl}}=0\\
&\sum_{m=1}^{M}2u_{nl}w_{lm}^{q}z_{l}^{p}d^{2}(x_{nm}-c_{lm})+\alpha\sum_{m=1}^{M}2(u_{nl}-b_{n}f_{nl})\,w_{lm}^{q}\,z_{l}^{p}d^{2}(x_{nm}-c_{lm})-\delta=0\\
&\delta=\sum_{m=1}^{M}2u_{nl}\,w_{lm}^{q}\,z_{l}^{p}\,d^{2}(x_{nm}-c_{lm})+\alpha\sum_{m=1}^{M}2(u_{nl}-b_{n}f_{nl})\,w_{lm}^{q}\,z_{l}^{p}\,d^{2}(x_{nm}-c_{lm})\\
&\delta=2u_{nl}\sum_{m=1}^{M}w_{lm}^{q}z_{l}^{p}d^{2}(x_{nm}-c_{lm})+2\alpha(u_{nl}-b_{n}f_{nl})\sum_{m=1}^{M}w_{lm}^{q}z_{l}^{p}d^{2}(x_{nm}-c_{lm})\\
&\delta=((2+2\alpha)u_{nl}-2\alpha b_{n}f_{nl})\sum\nolimits_{m=1}^{M}w_{lm}^{q}z_{l}^{p}d^{2}(x_{nm}-c_{lm})
\end{aligned} \tag{A.3}$$

It follows from Eqs. (A.2) and (A.3) that:

$$u_{nk} = \frac{\dfrac{((2+2\alpha)u_{nl} - 2\alpha b_n f_{nl}) \sum_{m=1}^{M} w_{lm}^{q} \quad z_l^p d^2(x_{nm} - c_{lm})}{\sum_{m=1}^{M} w_{km}^{q} \quad z_k^p d^2(x_{nm} - c_{km})} + 2\alpha b_n f_{nk}}{2+2\alpha}$$

$$(2+2\alpha)u_{nk} = ((2+2\alpha)u_{nl} - 2\alpha b_n f_{nl}) \left(\frac{\sum_{m=1}^{M} w_{lm}^{q} z_l^p d^2(x_{nm} - c_{lm})}{\sum_{m=1}^{M} w_{km}^{q} z_k^p d^2(x_{nm} - c_{km})} \right) + 2\alpha b_n f_{nk}$$

$$(2+2\alpha)u_{nl} = \frac{(2+2\alpha) \quad u_{nk} - 2\alpha b_n f_{nk}}{\left(\dfrac{\sum_{m=1}^{M} \quad w_{lm}^{q} \quad z_l^p d^2(x_{nm} - c_{lm})}{\sum_{m=1}^{M} \quad w_{km}^{q} \quad z_k^p d^2(x_{nm} - c_{km})} \right)} + 2\alpha b_n f_{nl} \tag{A.4}$$

$$u_{nl} = \frac{\dfrac{(2+2\alpha)u_{nk} - 2\alpha b_n f_{nk}}{\left(\frac{\sum_{m=1}^{M} \quad w_{lm}^{q} \quad z_l^p d^2(x_{nm} - c_{lm})}{\sum_{m=1}^{M} \quad w_{km}^{q} \quad z_k^p d^2(x_{nm} - c_{km})} \right)} + 2\alpha b_n f_{nl}}{(2+2\alpha)}$$

considering $d_{nl} = \sum_{m=1}^{M} \quad w_{lm}^{q} \quad z_l^p d^2(x_{nm} - c_{lm})$, $\sum_{l=1}^{K} u_{nl} = 1$, and $d_{nk} = \sum_{m=1}^{M} \quad w_{km}^{q} \quad z_k^p d^2(x_{nm} - c_{km})$, Eq. (A.5) is obtained:

$$\sum_{l=1}^{K} u_{nl} = \frac{1}{(2+2\alpha)} \sum_{l=1}^{K} \frac{(2+2\alpha)u_{nk} - 2\alpha b_n f_{nk}}{\frac{d_{nl}}{d_{nk}}} + 2\alpha b_n f_{nl}$$

$$(2+2\alpha) = \sum_{l=1}^{K} \frac{(2+2\alpha)u_{nk} - 2\alpha b_n f_{nk}}{\frac{d_{nl}}{d_{nk}}} + 2\alpha b_n f_{nl}$$

$$(2+2\alpha) = \sum_{l=1}^{K} \frac{(2+2\alpha)u_{nk} - 2\alpha b_n f_{nk}}{\frac{d_{nl}}{d_{nk}}} + \sum_{l=1}^{K} 2\alpha b_n f_{nl}$$

$$(2+2\alpha) = \sum_{l=1}^{K} \frac{(2+2\alpha)u_{nk}}{\frac{d_{nl}}{d_{nk}}} - \sum_{l=1}^{K} \frac{2\alpha b_n f_{nk}}{\frac{d_{nl}}{d_{nk}}} + \sum_{l=1}^{K} 2\alpha b_n f_{nl}$$

$$(2+2\alpha) = (2+2\alpha)u_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}} - 2\alpha b_n f_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}} + 2\alpha b_n \sum_{l=1}^{K} f_{nl}$$

$$1 = 1 u_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}} - \frac{1}{1+\alpha} \alpha b_n f_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}} + \frac{1}{1+\alpha} \alpha b_n \sum_{l=1}^{K} f_{nl}$$

$$u_{nk} = \frac{1 + \frac{1}{1+\alpha} \alpha b_n f_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}} - \frac{1}{1+\alpha} \alpha b_n \sum_{l=1}^{K} f_{nl}}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}}$$

$$u_{nk} = \frac{1}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \frac{\frac{1}{1+\alpha} \alpha b_n f_{nk} \sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} - \frac{\frac{1}{1+\alpha} \alpha b_n \sum_{l=1}^{K} f_{nl}}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}}$$

$$u_{nk} = \frac{1 - \left[\frac{1}{1+\alpha} \alpha b_n \sum_{l=1}^{K} f_{nl} \right]}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \frac{1}{1+\alpha} \alpha b_n f_{nk}$$

$\sum_{l=1}^{K} f_{nl}$ always equals 1, so it doesn't play any specific roles here.

$$u_{nk} = \frac{1 - \left[\frac{1}{1+\alpha}\alpha b_n\right]}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \frac{1}{1+\alpha}\alpha b_n f_{nk}$$

$$u_{nk} = \frac{\frac{1+\alpha-\alpha b_n}{1+\alpha}}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \frac{1}{1+\alpha}\alpha b_n f_{nk}$$

$$u_{nk} = \frac{1+\alpha-\alpha b_n}{(1+\alpha)\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \frac{1}{1+\alpha}\alpha b_n f_{nk} \tag{A.5}$$

$$u_{nk} = \frac{1}{1+\alpha}\left(\frac{1+\alpha(1-b_n)}{\sum_{l=1}^{K} \frac{d_{nk}}{d_{nl}}} + \alpha b_n f_{nk}\right)$$

As a result, the membership updating function is obtained.

Theorem 2. If $\boldsymbol{U}$, $\boldsymbol{W}$, and $\boldsymbol{Z}$ are assumed to be fixed, $\boldsymbol{C}$ is a strict local minimum of $\boldsymbol{F}(\boldsymbol{C})$ if and only if $\boldsymbol{C}$ is calculated using Eq. (15).

Proof. Considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{C})$ to c_{km} and set equal to 0, we obtain:

$$\frac{\partial\widetilde{\mathrm{F}}}{\partial c_{km}} = 0$$

$$\sum_{n=1}^{N} 2{u_{nk}}^2 w_{km}^q z_k^p(-\gamma_m)(x_{nm}-c_{km})\left(\exp\left(-\gamma_m(x_{nm}-c_{km})^2\right)\right) + \alpha\sum_{n=1}^{N} 2(u_{nk}-b_n f_{nk})^2 w_{km}^q z_k^p(-\gamma_m)(x_{nm}-c_{km})\left(\exp\left(-\gamma_m(x_{nm}-c_{km})^2\right)\right) = 0 \tag{A.6}$$

By applying some mathematical simplifications, the center updating function is obtained as Eq. (A.7):

$$c_{km} = \frac{\sum_{n=1}^{N}(x_{nm})\left(\exp\left(-\gamma_m(x_{nm}-c_{km})^2\right)\right)\left({u_{nk}}^2 + \alpha(u_{nk}-b_n f_{nk})^2\right)}{\sum_{n=1}^{N}\left(\exp\left(-\gamma_m(x_{nm}-c_{km})^2\right)\right)\left({u_{nk}}^2 + \alpha(u_{nk}-b_n f_{nk})^2\right)} \tag{A.7}$$

Theorem 3. If $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{Z}$ are assumed to be fixed, $\boldsymbol{W}$ is a strict local minimum of $\boldsymbol{F}(\boldsymbol{W})$ if and only if $\boldsymbol{W}$ is calculated using Eq. (16).

Proof. Considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{W})$ to w_{km} and set equal to 0, we obtain:

$$\frac{\partial\widetilde{\mathrm{F}}}{\partial w_{km}} = 0$$

$$q\, w_{km}^{q-1}\, z_k^p \sum_{n=1}^{N} {u_{nk}}^2\, d^2(x_{nm}-c_{km}) + \alpha q\, w_{km}^{q-1}\, z_k^p \sum_{n=1}^{N} (u_{nk}-b_n f_{nk})^2\, d^2(x_{nm}-c_{km}) - \psi = 0 \tag{A.8}$$

$$w_{km} = \left[\frac{\psi}{q z_k^p\left[\sum_{n=1}^{N} {u_{nk}}^2\, d^2(x_{nm}-c_{km}) + \alpha\sum_{n=1}^{N}(u_{nk}-b_n f_{nk})^2\, d^2(x_{nm}-c_{km})\right]}\right]^{\frac{1}{q-1}}$$

now, Considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{W})$ to w_{ks} and set equal to 0, we get ψ:

$$\frac{\partial\widetilde{\mathrm{F}}}{\partial w_{ks}} = 0$$

$$\psi = q\, w_{ks}^{q-1}\, z_k^p\left[\sum_{n=1}^{N} {u_{nk}}^2 d^2(x_{ns}-c_{ks}) + \alpha\sum_{n=1}^{N}(u_{nk}-b_n f_{nk})^2\, d^2(x_{ns}-c_{ks})\right] \tag{A.9}$$

We can rewrite from Eqs. (A.8) and (A.9):

$$\sum_{s=1}^{M} w_{ks} = \sum_{s=1}^{M} w_{km}\left[\frac{\sum_{n=1}^{N}\left[{u_{nk}}^2 + \alpha(u_{nk}-b_n f_{nk})^2\right] d^2(x_{nm}-c_{km})}{\sum_{n=1}^{N}\left[{u_{nk}}^2 + \alpha(u_{nk}-b_n f_{nk})^2\right] d^2(x_{ns}-c_{ks})}\right]^{\frac{1}{q-1}} \tag{A.10}$$

Considering $\sum_{s=1}^{M} w_{ks} = 1$, Eq. (A.11) is obtained:

$$1=\sum_{s=1}^{M} w_{km}\left[\frac{\sum_{n=1}^{N}\left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]\, d^2(x_{nm}-c_{km})}{\sum_{n=1}^{N}\left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]\, d^2(x_{ns}-c_{ks})}\right]^{\frac{1}{q-1}}$$

$$w_{km}=\frac{1}{\sum_{s=1}^{M}\left[\frac{\sum_{n=1}^{N}\left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]\, d^2(x_{nm}-c_{km})}{\sum_{n=1}^{N}\left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]\, d^2(x_{ns}-c_{ks})}\right]^{\frac{1}{q-1}}} \tag{A.11}$$

Thus, the updating function of the feature weighting is obtained.

Theorem 3. If $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{W}$ are assumed to be fixed, $\boldsymbol{Z}$ is a strict local minimum of $\boldsymbol{F}(\boldsymbol{Z})$ if and only if $\boldsymbol{Z}$ is calculated using Eq. (17).

Proof. Considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{Z})$ to z_k and set equal to 0, we obtain:

$$\frac{\partial\widetilde{\mathrm{F}}}{\partial z_k}=0$$

$$\sum_{n=1}^{N}\sum_{m=1}^{M}{u_{nk}}^2w_{km}^q\mathrm{p}z_k^{p-1}d^2(x_{nm}-c_{km})+\alpha\sum_{n=1}^{N}\sum_{m=1}^{M}(u_{nk}-b_nf_{nk})^2w_{km}^q\mathrm{p}z_k^{p-1}d^2(x_{nm}-c_{km})-\omega=0 \tag{A.12}$$

now, considering the gradient of $\widetilde{\mathrm{F}}(\boldsymbol{Z})$ to z_l and set equal to 0, we get ω:

$$\frac{\partial\widetilde{\mathrm{F}}}{\partial z_l}=0$$

$$\omega=\mathrm{p}z_l^{p-1}\sum_{n=1}^{N}\sum_{k=1}^{K}{u_{nl}}^mw_{lm}^qd^2(x_{nm}-c_{lm})+\alpha\mathrm{p}z_l^{p-1}\sum_{n=1}^{N}\sum_{m=1}^{M}(u_{nl}-b_nf_{nl})^2w_{lm}^qd^2(x_{nm}-c_{lm}) \tag{A.13}$$

$$\omega=\mathrm{p}z_l^{p-1}\left[\sum_{n=1}^{N}\sum_{k=1}^{K}{u_{nl}}^mw_{lm}^qd^2(x_{nm}-c_{lm})+\alpha\sum_{n=1}^{N}\sum_{m=1}^{M}(u_{nl}-b_nf_{nl})^2w_{lm}^qd^2(x_{nm}-c_{lm})\right]$$

We can deduct from Eqs. (A.12) and (A.13) that:

$$z_l=z_k\left[\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{km}^qd^2(x_{nm}-c_{km})\left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]}{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{lm}^qd^2(x_{nm}-c_{lm})\left[{u_{nl}}^2+\alpha(u_{nl}-b_nf_{nl})^2\right]}\right]^{\frac{1}{p-1}}$$

$$\sum_{l=1}^{K}z_l=\sum_{l=1}^{K}z_k\left[\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{km}^qd^2(x_{nm}-c_{km})\ \left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]}{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{lm}^qd^2(x_{nm}-c_{lm})\ \left[{u_{nl}}^2+\alpha(u_{nl}-b_nf_{nl})^2\right]}\right]^{\frac{1}{p-1}} \tag{A.14}$$

Considering that $\sum_{l=1}^{K}z_l=1$, Eq. (A.15) is obtained as the updating function of the cluster weighting:

$$z_k=\frac{1}{\sum_{l=1}^{K}\left[\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{km}^qd^2(x_{nm}-c_{km})\ \left[{u_{nk}}^2+\alpha(u_{nk}-b_nf_{nk})^2\right]}{\sum_{n=1}^{N}\sum_{k=1}^{K}w_{lm}^qd^2(x_{nm}-c_{lm})\ \left[{u_{nl}}^2+\alpha(u_{nl}-b_nf_{nl})^2\right]}\right]^{\frac{1}{p-1}}} \tag{A.15}$$

Appendix B

Pseudo-code of the SSFCM-FWCW is provided in Algorithm 1.

Algorithm 1. . Pseudo-code of the SSFCM-FWCW.

Input: Dataset $\boldsymbol{\chi} = \{x_n\}_{n=1}^{N}$, Number of cluster K, Parameters ε, p_{init}, p_{step}, p_{step}, $t_{\max}$, q, β, and α.
Output: Membership matrix $\boldsymbol{U}$;
set $t = 0$
set $p_{\text{init}} = 0$
set $z_k^{(0)} = \frac{1}{K}$, $\forall k = 1 \dots K$
set $w_{km}^{(0)} = \frac{1}{M}$, $\forall k = 1 \dots K$
set empty= **FALSE** //No empty or singleton clusters yet detected
$p = p_{\text{init}}$
repeat
$t = t + 1$
Update the cluster assignments $\boldsymbol{U}$ by Eq. (14)
If empty or singleton clusters have occurred at a time t **then** //reduce p.
  empty=**TRUE**
  $p = p - p_{\text{step}}$
  if $p < p_{\text{init}}$ **then**
    return NULL
  end if
  //Revert to the assignments and weights corresponding to the reduce p.
  $u_{nk}^{(t)} = \left[\mathbf{U\_history}^{(p)}\right]_{nk}$, $\forall k = 1 \dots K$, $\forall n = 1 \dots N$
  $z_k^{(t-1)} = \left[\boldsymbol{Z}\mathbf{\_history}^{(p)}\right]_{k}$, $\forall k = 1 \dots K$
  $w_{km}^{(t-1)} = \left[\boldsymbol{w}\mathbf{\_history}^{(p)}\right]_{km}$, $\forall m = 1 \dots M$, $\forall k = 1 \dots K$
end if
Update the cluster center $\boldsymbol{C}$ by Eq. (15)
if $p < p_{\max}$ **and** empty=**FALSE** then //increase p.
  $\mathbf{U\_history}^{(p)} = [u_{nk}^{(t)}]$ //store the current $\boldsymbol{U}$ in $\mathbf{U\_history}^{(\boldsymbol{p})}$.
  $\boldsymbol{W}\mathbf{\_history}^{(p)} = [w_{km}^{(t-1)}]$ //store the previous $\boldsymbol{W}$ in $\boldsymbol{W}\mathbf{\_history}^{(\boldsymbol{p})}$.
  $\boldsymbol{Z}\mathbf{\_history}^{(p)} = [z_k^{(t-1)}]$ //store the previous $\boldsymbol{Z}$ in $\boldsymbol{Z}\mathbf{\_history}^{(\boldsymbol{p})}$.
  $p = p + p_{\text{step}}$
end if
Update the feature weight $\boldsymbol{W}$ by Eq. (16)
Update the cluster weight $\boldsymbol{Z}$ by Eq. (17)
until $\left|F^{(t)} - F^{(t-1)}\right| < \varepsilon$ ***or*** $t \geq t_{\max}$
return $\boldsymbol{U}$

References

[1] A.E. Ezugwu, et al., A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects, Eng. Appl. Artif. Intell. vol. 110 (2022) 104743.
[2] X. Zhang, Y. Zheng, X. Ye, Q. Peng, W. Wang, S. Li, Clustering with implicit constraints: a novel approach to housing market segmentation, Trans. GIS vol. 26 (2) (2022) 585–608.
[3] H. Zhang, H. Li, N. Chen, S. Chen, J. Liu, Novel fuzzy clustering algorithm with variable multi-pixel fitting spatial information for image segmentation, Pattern Recognit. vol. 121 (2022) 108201.
[4] J. Cai, J. Hao, H. Yang, X. Zhao, Y. Yang, A review on semi-supervised clustering, Inf. Sci. (2023).
[5] T. Chaira, A novel intuitionistic fuzzy *c*-means clustering algorithm and its application to medical images, Appl. Soft Comput. vol. 11 (2) (2011) 1711–1717.
[6] A. Omidi, A. Mohammadshahi, N. Gianchandani, R. King, L. Leijser, R. SouzaUnsupervised Domain Adaptation of MRI Skull-Stripping Trained on Adult Data to Newborns," in Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision , 2024, , 7718–7727.
[7] F. Arefinia, M. Aria, R. Rabiei, A. Hosseini, A. Ghaemian, A. Roshanpoor, Non-invasive fractional flow reserve estimation using deep learning on intermediate left anterior descending coronary artery lesion angiography images, 2024/01/20, Sci. Rep. vol. 14 (1) (2024) 1818, https://doi.org/10.1038/s41598-024-52360-5.
[8] S. Zarean Shahraki, et al., Time-related survival prediction in molecular subtypes of breast cancer using time-to-event deep-learning-based models, Front. Oncol. vol. 13 (2023) 1147604.
[9] K. Berahmand, F. Daneshfar, M. Dorosti, and M.J. Aghajani, "An Improved Deep Text Clustering via Local Manifold of an Autoencoder Embedding," 2022.
[10] J.N. Khiarak, et al., KartalOl: a new deep neural network framework based on transfer learning for iris segmentation and localization task—new dataset for iris segmentation, 2023/04/13, Iran. J. Comput. Sci. (2023), https://doi.org/10.1007/s42044-023-00141-0.
[11] A.M. Ikotun, A.E. Ezugwu, L. Abualigah, B. Abuhaija, J. Heming, *k*-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data, Inf. Sci. (2022).
[12] A.G. Oskouei, M.A. Balafar, T. Akan, A Brain MRI Segmentation Method Using Feature Weighting and a Combination of Efficient Visual Features. in Applied Computer Vision and Soft Computing with Interpretable AI, Chapman and Hall/CRC, 2024, pp. 15–34.
[13] I.-D. Borlea, R.-E. Precup, A.-B. Borlea, D. Iercan, A unified form of fuzzy *c-means* and *k*-means algorithms and its partitional implementation, Knowl. -Based Syst. vol. 214 (2021) 106731.
[14] M.B. Ferraro, P. Giordani, Soft clustering, Wiley Interdiscip. Rev.: Comput. Stat. vol. 12 (1) (2020) e1480.
[15] S. Sivarathri, A. Govardhan, Experiments on hypothesis" fuzzy *k*-means is better than *k*-means for clustering, Int. J. Data Min. Knowl. Manag. Process (IJDKP) Vol. vol. 4 (2014).
[16] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. vol. 109 (2) (2020) 373–440.

[17] J. Tanha, N. Samadi, Y. Abdi, N. Razzaghi-Asl, CPSSDS: Conformal prediction for semi-supervised classification on data streams, Inf. Sci. vol. 584 (2022) 212–234.
[18] Y. Reddy, P. Viswanath, B.E. Reddy, Semi-supervised learning: a brief review, Int. J. Eng. Technol. vol. 7 (1.8) (2018) 81.
[19] Y. Qin, S. Ding, L. Wang, Y. Wang, Research progress on semi-supervised clustering, Cogn. Comput. vol. 11 (2019) 599–612.
[20] G. González-Almagro, J. Luengo, J.-R. Cano, S. García, Enhancing instance-level constrained clustering through differential evolution, Appl. Soft Comput. vol. 108 (2021) 107435.
[21] X. Yin, S. Chen, E. Hu, D. Zhang, Semi-supervised clustering with metric learning: an adaptive kernel method, Pattern Recognit. vol. 43 (4) (2010) 1320–1333.
[22] X. Yin, T. Shu, Q. Huang, Semi-supervised fuzzy clustering with metric learning and entropy regularization, Knowl. -Based Syst. vol. 35 (2012) 304–311.
[23] R.C. De Amorim, B. Mirkin, Minkowski metric, feature weighting and anomalous cluster initializing in *k*-means clustering, Pattern Recognit. vol. 45 (3) (2012) 1061–1075.
[24] S. Basu, Semi-supervised clustering by seeding, Proc. ICML-2002 (2002).
[25] W. Pedrycz, J. Waletzky, Fuzzy clustering with partial supervision, IEEE Trans. Syst. Man, Cybern. Part B vol. 27 (5) (1997) 787–795.
[26] M. Bilenko, S. Basu, R.J. MooneyIntegrating constraints and metric learning in semi-supervised clustering," in Proceedings of twenty-first international conference on Machine learning , 2004, , 11.
[27] S. Wei, Z. Li, C. Zhang, Combined constraint-based with metric-based in semi-supervised clustering ensemble, Int. J. Mach. Learn. Cybern. vol. 9 (2018) 1085–1100.
[28] H. Gan, Safe semi-supervised fuzzy $ $-means clustering, IEEE Access vol. 7 (2019) 95659–95664.
[29] H. Gan, Y. Fan, Z. Luo, R. Huang, Z. Yang, Confidence-weighted safe semi-supervised clustering, Eng. Appl. Artif. Intell. vol. 81 (2019) 107–116.
[30] H. Gan, Y. Fan, Z. Luo, Q. Zhang, Local homogeneous consistent safe semi-supervised clustering, Expert Syst. Appl. vol. 97 (2018) 384–393.
[31] H. Gan, Z. Yang, R. Zhou, Adaptive safety-aware semi-supervised clustering, Expert Syst. Appl. vol. 212 (2023) 118751.
[32] P.T. Huan, et al., TS3FCM: trusted safe semi-supervised fuzzy clustering method for data partition with high confidence, Multimed. Tools Appl. vol. 81 (9) (2022) 12567–12598.
[33] T.M. Tuan, P.H. Thong, T.T. Ngan, An improvement of trusted safe semi-supervised fuzzy clustering method with multiple fuzzifiers, J. Comput. Sci. Cybern. vol. 38 (1) (2022) 47–61.
[34] G. Casalino, G. Castellano, C. Mencar, Data stream classification by dynamic incremental semi-supervised fuzzy clustering, Int. J. Artif. Intell. Tools vol. 28 (08) (2019) 1960009.
[35] M. Hashemzadeh, A.G. Oskouei, N. Farajzadeh, New fuzzy *c*-means clustering method based on feature-weight and cluster-weight learning, Appl. Soft Comput. vol. 78 (2019) 324–345.
[36] H. Li, M. Wei, Fuzzy clustering based on feature weights for multivariate time series, Knowl. -Based Syst. vol. 197 (2020) 105907.
[37] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy *c*-means clustering algorithm, Comput. Geosci. vol. 10 (2-3) (1984) 191–203.
[38] A.G. Oskouei, M. Hashemzadeh, B. Asheghi, M.A. Balafar, CGFFCM: cluster-weight and Group-local Feature-weight learning in Fuzzy *c*-means clustering algorithm for color image segmentation, Appl. Soft Comput. vol. 113 (2021) 108005.
[39] R. Suganya, R. Shanthi, Fuzzy *c*-means algorithm-a review, Int. J. Sci. Res. Publ. vol. 2 (11) (2012) 1.
[40] R.-P. Li, M. MukaidonoA maximum-entropy approach to fuzzy clustering IEEE , vol. 4 in Proceedings of 1995 IEEE International Conference on Fuzzy Systems.1995, in Proceedings of International Conference on Fuzzy Systems., IEEE19952227–2232.
[41] N.R. Pal, K. Pal, J.C. BezdekA mixed c-means clustering model IEEE , vol. 1 in Proceedings of 6th international fuzzy systems conference1997, , 11–21.
[42] N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek, A possibilistic fuzzy *c*-means clustering algorithm, IEEE Trans. Fuzzy Syst. vol. 13 (4) (2005) 517–530.
[43] R.N. Davé, R. Krishnapuram, Robust clustering methods: a unified view, IEEE Trans. Fuzzy Syst. vol. 5 (2) (1997) 270–293.
[44] M.N. Ahmed, S.M. Yamany, N. Mohamed, A.A. Farag, T. Moriarty, A modified fuzzy *c*-means algorithm for bias field estimation and segmentation of MRI data, IEEE Trans. Med. Imaging vol. 21 (3) (2002) 193–199.
[45] D.-M. Tsai, C.-C. Lin, Fuzzy *c*-means based clustering for linearly and nonlinearly separable data, Pattern Recognit. vol. 44 (8) (2011) 1750–1760.
[46] A. Gosain, S. Dahiya, Performance analysis of various fuzzy clustering algorithms: a review, Procedia Comput. Sci. vol. 79 (2016) 100–111.
[47] M. Zarinbal, M.F. Zarandi, I. Turksen, Relative entropy fuzzy *c*-means clustering, Inf. Sci. vol. 260 (2014) 74–97.
[48] P.-L. Lin, P.-W. Huang, C.-H. Kuo, Y. Lai, A size-insensitive integrity-based fuzzy *c*-means method for data clustering, Pattern Recognit. vol. 47 (5) (2014) 2042–2056.
[49] A. Golzari Oskouei, M.A. Balafar, C. Motamed, RDEIC-LFW-DSS: ResNet-based deep embedded image clustering using local feature weighting and dynamic sample selection mechanism, 2023/10/01/, Inf. Sci. vol. 646 (2023) 119374, https://doi.org/10.1016/j.ins.2023.119374.
[50] A. Golzari Oskouei, M.A. Balafar, C. Motamed, EDCWRN: efficient deep clustering with the weight of representations and the help of neighbors, 2023/03/01, Appl. Intell. vol. 53 (5) (2023) 5845–5867, https://doi.org/10.1007/s10489-022-03895-5.
[51] A. Golzari Oskouei, M. Hashemzadeh, CGFFCM: a color image segmentation method based on cluster-weight and feature-weight learning, 2022/02/01/, Softw. Impacts vol. 11 (2022) 100228, https://doi.org/10.1016/j.simpa.2022.100228.
[52] A. Golzari Oskouei, M.A. Balafar, C. Motamed, FKMAWCW: categorical fuzzy k-modes clustering with automated attribute-weight and cluster-weight learning, 2021/12/01/, Chaos Solitons Fractals vol. 153 (2021) 111494, https://doi.org/10.1016/j.chaos.2021.111494.
[53] S. Askari, Fuzzy *c*-means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: review and development, Expert Syst. Appl. vol. 165 (2021) 113856.
[54] J. Zhou, W. Pedrycz, X. Yue, C. Gao, Z. Lai, J. Wan, Projected fuzzy *c*-means clustering with locality preservation, Pattern Recognit. vol. 113 (2021) 107748.
[55] K. Taha, Semi-supervised and un-supervised clustering: a review and experimental evaluation, Inf. Syst. (2023) 102178.
[56] S. Zeng, X. Tong, N. Sang, R. Huang, A study on semi-supervised FCM algorithm, Knowl. Inf. Syst. vol. 35 (2013) 585–612.
[57] W. Pedrycz, G. Vukovich, Fuzzy clustering with supervision, Pattern Recognit. vol. 37 (7) (2004) 1339–1349.
[58] A. Bouchachia, W. Pedrycz, Data clustering with partial supervision, Data Min. Knowl. Discov. vol. 12 (2006) 47–78.
[59] N. Grira, M. Crucianu, N. Boujemaa, Active semi-supervised fuzzy clustering, Pattern Recognit. vol. 41 (5) (2008) 1834–1844.
[60] E. Yasunori, H. Yukihiro, Y. Makito, M. Sadaaki, On semi-supervised fuzzy *c*-means clustering. 2009 IEEE International Conference on Fuzzy Systems, IEEE, 2009, pp. 1119–1124.
[61] C.-F. Gao, X.-J. Wu, A new semi-supervised clustering algorithm with pairwise constraints by competitive agglomeration, Appl. Soft Comput. vol. 11 (8) (2011) 5281–5291.
[62] F. Salehi, M.R. Keyvanpour, A. Sharifi, SMKFC-ER: semi-supervised multiple kernel fuzzy clustering based on entropy and relative entropy, Inf. Sci. vol. 547 (2021) 667–688.
[63] T.M. Tuan, et al., A new approach for semi-supervised fuzzy clustering with multiple fuzzifiers, Int. J. Fuzzy Syst. vol. 24 (8) (2022) 3688–3701.
[64] M. Aria, E. Nourani, A. Golzari Oskouei, ADA-COVID: adversarial deep domain adaptation-based diagnosis of COVID-19 from lung CT scans using triplet embeddings, 2022/02/08, Comput. Intell. Neurosci. vol. 2022 (2022) 2564022, https://doi.org/10.1155/2022/2564022.
[65] L. Lovász, M.D. Plummer, Matching Theory, American Mathematical Soc, 2009.