electronics

Article

Unsupervised Multiview Fuzzy C-Means Clustering Algorithm

Ishtiaq Hussain, Kristina P. Sinaga and Miin-Shen Yang *

Department of Applied Mathematics, Chung Yuan Christian University, Chung-Li, Taoyuan 32023, Taiwan; g10701102@cycu.edu.tw or ishtiaqhucane@gmail.com (I.H.); kristina@math.cycu.edu.tw (K.P.S.)
* Correspondence: msyang@math.cycu.edu.tw

Abstract: The rapid development in information technology makes it easier to collect vast numbers of data through the cloud, internet and other sources of information. Multiview clustering is a significant way for clustering multiview data that may come from multiple ways. The fuzzy c-means (FCM) algorithm for clustering (single-view) datasets was extended to process multiview datasets in the literature, called the multiview FCM (MV-FCM). However, most of the MV-FCM clustering algorithms and their extensions in the literature need prior information about the number of clusters and are also highly influenced by initializations. In this paper, we propose a novel MV-FCM clustering algorithm with an unsupervised learning framework, called the unsupervised MV-FCM (U-MV-FCM), such that it can search an optimal number of clusters during the iteration process of the algorithm without giving the number of clusters a priori. It is also free of initializations and parameter selection. We then use three synthetic and six benchmark datasets to make comparisons between the proposed U-MV-FCM and other existing algorithms and to highlight its practical implications. The experimental results show that our proposed U-MV-FCM algorithm is superior and more useful for clustering multiview datasets.

Keywords: clustering; fuzzy c-means (FCM); multiview FCM (MV-FCM); unsupervised multiview FCM (U-MV-FCM); number of clusters

check for updates

Citation: Hussain, I.; Sinaga, K.P.; Yang, M.-S. Unsupervised Multiview Fuzzy C-Means Clustering Algorithm. *Electronics* **2023**, *12*, 4467. https://doi.org/10.3390/electronics12214467

Academic Editors: Juan M. Corchado, Carlos A. Iglesias, Byung-Gyu Kim, Rashid Mehmood, Fuji Ren and In Lee

Received: 4 October 2023
Revised: 25 October 2023
Accepted: 29 October 2023
Published: 30 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Clustering is one of the fundamental techniques used to partition a dataset into clusters such that the data points in the same cluster are the most similar as each other and the data points in the different clusters are the most dissimilar [1,2]. It had been widely used in many areas [3]. Clustering may be categorized into two groups, named the parametric and non-parametric approaches. In 1993, model-based clustering was initially proposed by Banified and Raftrey [4] with its applications [5–7]. In the non-parametric methods, one of the renowned and pioneering partitional method is k-means [8,9]. Since Zadeh [10] first proposed fuzzy sets in 1965 where he introduced partial memberships, Ruspini [11] coined a fuzzy approach in the form of fuzzy c-partition for clustering by extending indicator membership function to allow fuzzy memberships in the interval [0, 1]. Based on fuzzy c-partitions, Dunn [12] introduced fuzzy c-means clustering as an advancement of k-means. Various extensions of FCM and its applications can be found in the literature, such as [13–17].

Nowadays, social media, social networks, and IoT rapidly grow, and so collected data are increasing each day. These produced data become more complex with a multiple-view way. That is, with these information and communication technologies via internet, massive numbers of data with multiview representation are generated each day. However, FCM and its extensions were used to handle (single-view) datasets. Usually, particular views capture specific aspects of information, and the evidence obtained from particular views is mutually supportive. Some researchers may extract information from different angles and then combined it. In general, the goal of multiview learning is to obtain more filtering and higher level of information. In 2001, Dhillon [18] was the first to propose co-clustering for 2-view data, and then Bickel and Scheffer [19] introduced multiview clustering for dealing

multiview data. With advances in clustering for multiple-view data, many researchers have contributed to the literature, such as [20–24]. Recently, Huang et al. [25] used multiview for deep matrix decomposition. Chen et al. [26] used it for graph regularized least square regression. Tan et al. [27] constructed low-rank subspace multiview clustering through squeezing integrated information from cross view and each view. Yang and Sinaga [28] developed FCM clustering for multiview data using a collaborative way of feature weights. Yang and Hussain [29] proposed unsupervised multiview k-means clustering algorithm. Papakostas et al. [30] introduced an augmented reality spatial ability training application based on fuzzy weights. Papakostas et al. [31] proposed customizing spatial aptitude training in an augmented reality system using fuzzy logic. Lengyel, and Botta-Dukát [32] introduced evaluating clustering efficiency with generalized mean-based silhouette width: a flexible approach. Yang et al. [33] proposed active sensing in the categorization of visual patterns and Xu et al. [34] offered reviews on determining the number of clusters.

Multiview clustering is a technique used in data analysis and machine learning to group data points into clusters when multiple perspectives or "views" of data are available. In the context of the evolving data landscape, it holds significant relevance for several reasons, such as data complexity and diversity, improved accuracy, robustness and stability, discovery of hidden patterns, and adaptability of new data, etc. However, the aforementioned traditional multiview clustering schemes are always affected by initials and necessity to assign a prior cluster number. It is sensitive to the choice of initial condition, and it is highly dependent on the initial guess for cluster centers. These sensitivity and dependence will let algorithms be more complicated and have cost increasing. In this article, we first formulate an unsupervised schema for multiview FCM (MV-FCM) clustering. We then propose the unsupervised MV-FCM (MV-FCM) such that it can find an optimal number of clusters without assigning a prior number of clusters and can also be free of initializations and parameter selection.

The proposed U-MV-FCM algorithm presents several key innovations including the ability to seamlessly integrate information from these multiple views to improve clustering accuracy and in an unsupervised learning framework so that it does not rely on prior knowledge of the number of clusters. Another notable innovation in the U-MV-FCM algorithm is its ability to automatically determine the number of clusters. The initialization-free nature and the capacity of automatically finding the number of clusters in the U-MV-FCM algorithm make it a user-friendly and robust algorithm for clustering multiview data. These advantages reduce the burden on users in terms of parameter tuning and initializations and make it more accessible and effective in various real-world applications. Some of the notations used in the paper are shown in Table 1. The reminder of this paper is as follows. Section 2 demonstrates related works. In Section 3, we introduce an unsupervised-regularization structure for MV-FCM clustering. More parameter estimations are discussed, and then the unsupervised MV-FCM (U-MV-FCM) clustering scheme is proposed. Section 4 presents the experimental results and comparisons, with some existing algorithms. Finally, Section 5 offers some conclusions.

Table 1. Notations used in the paper.

n	Number of data points
d	Number of dimensions
a_k	Cluster centers
μ_{ik}	Memberships
d_h	Number of dimensions in the hth view
v_h	The hth view weight
m	fuzziness index
a_k^h	Cluster centers in the hth view

2. Related Works

Before delving into our proposed U-MV-FCM algorithm, it is essential to explore these related existing MV-FCM clustering algorithms. This review offers a crucial foundation for our study, as it helps us understand the state of the art in FCM and MV-FCM algorithms in the realm of single-view as well as multiview clustering, which would be also used in our experimental comparisons. By examining these prior approaches, we aim to identify their strengths, limitations, and common trends. We also review cluster validity indices for MV-FCM clustering that can be used to find the number of clusters. Let $x_1, \ldots, x_n$ be n data points in the Euclidean space $\mathbb{R}^d$ and let $a_1, \ldots, a_c$ represents c cluster centers. Let μ_{ik} be the membership degree of the ith data point in the kth cluster center with $\mu_{ik} \in [0,1]$ and $\sum_{k=1}^{c} \mu_{ik} = 1 \forall i$. The well-known FCM [12,13] with its objective function is given by:

$$J(U\ ,\ A) = \sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik}^{m}\ \|x_i - a_k\|^2 \tag{1}$$

where $\|x_i - a_k\|$ is the Euclidean distance between ith data and kth cluster, m is the fuzziness index, n denotes the data points, and c is the cluster number. The FCM algorithm is iterated through the updating equations $a_{kj} = \sum_{i=1}^{n} \mu_{ik}^{m} x_{ij} / \sum_{i=1}^{n} \mu_{ik}^{m}$ and $\mu_{ik} = \left(\sum_{k\prime=1}^{c} (d_{ik}/d_{ik\prime})^{\frac{2}{m-1}}\right)^{-1}$ with $d_{ik} = \|x_i - a_k\|$ for minimizing the objective function $J(U, A)$ of FCM.

Since FCM and its extensions need to give the prior number of clusters, and always depend on initializations and some parameters like fuzziness index m, Yang and Nataliani [35] gave a novel FCM algorithm, called robust-learning FCM (RLFCM), to the robust initializations and also free to the parameter selection with finding an optimal number of clusters. The objective function is as follows:

$$\begin{aligned} J_{RLFCM}(U, \alpha, A^h) = [&\sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik} d_{ik}^2 - \eta_1 \sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik} \ln \alpha_k \\ &+ \eta_2 \sum_{i=1}^{n}\sum_{k=1}^{c} \mu_{ik} \ln \mu_{ik} - \eta_3 \sum_{i=1}^{n}\sum_{k=1}^{c} \alpha_k \ln \alpha_k] \end{aligned} \tag{2}$$

where $\eta_1, \eta_2, \eta_3 \geq 0$ and $d_{ik}^2 = \left\|x_i^h - a_k^h\right\|^2 = \sqrt{\sum_{j=}^{d} \left(x_i^h - a_k^h\right)^2}$. The updated equations of RLFCM are as follows: For cluster center a_{kj}, $a_{kj} = \sum_{i=1}^{n} \mu_{ik} x_{ij} / \sum_{i=1}^{n} \mu_{ik}$, for membership μ_{ik}, $\mu_{ik} = \exp\left(\frac{-d_{ik}^2 + \eta_1 \ln \alpha_k}{\eta_2}\right) / \sum_{t=1}^{c} \exp\left(\frac{-d_{it}^2 + \eta_1 \ln \alpha_t}{\eta_2}\right)$, and for mixing proportion α_k, $\alpha_k^{(new)} = \frac{1}{n}\sum_{i=1}^{n} \mu_{ik} + \frac{\eta_3}{\eta_1} \alpha_k^{(old)} \left(\ln \alpha_k^{(old)} - \sum_{t=1}^{c} \alpha_t^{(old)} \ln \alpha_t^{(old)}\right)$, where $\eta_1 = e^{-\frac{t}{150}}$, $\eta_2 = e^{-\frac{t}{400}}$. In this study, we borrow the idea of Yang and Nataliani [35] to propose the unsupervised multiview FCM clustering algorithm.

Although Bickel and Scheffer [19] proposed multiview clustering for the first time to handle multiview data, Cleuziou et al. [20] was a pioneer in developing an advance version of FCM, known as multiview FCM by augmenting more information among different views, and they also created Co-FKM. Let $\mathbf{X} = \{x_1, \ldots, x_n\}$ be a multiview dataset with $x_i = \left\{x_i^h\right\}_{h=1}^{s}, x_i^h \in \mathbb{R}^{d_h}$, and $x_i^h = \left\{x_{ij}^h\right\}_{j=1}^{d_h}$. The Co-FKM algorithm [20] used the collaborative idea of Pedrycz [36] with a sequencing strategy. To handle the multiview data, the Co-FKM combines two strategies. In the first strategy, the average disagreement term between any pair of the views, which are given as $\Delta = \frac{1}{s-1}\left(\sum_{h\prime=1, h\prime\neq h}^{s} \sum_{i=1}^{n}\sum_{k=1}^{c} \left(\left(u_{ik}^{h\prime}\right)^m - \left(u_{ik}^h\right)^m\right)\left(d_{ik}^h\right)^2\right)$, is integrated into the following objective function:

$$J_{CoFKM}(U\ ,\ A) = \sum_{h=1}^{s}\sum_{i=1}^{n}\sum_{k=1}^{c} \left(\mu_{ik}^h\right)^m \left(d_{ik}^h\right)^2 + \eta\Delta \tag{3}$$

where $d_{ik}^h = \left\| x_i^h - a_k^h \right\| = \sqrt{\sum_{j=1}^{d_h} \left(x_{ij}^h - a_{kj}^h \right)^2}$, and η is used to control the weight of the disagreement term. Thus, by minimizing the summation of the FCM objective function for each view and the pair disagreement term Δ, the membership of each view is achieved. The second strategy is then applied, in which the final consensus fuzzy membership is generated by calculating the geometric mean of memberships of all views with $\mu_{ik} = \sqrt[s]{\prod_{h=1}^{s} \mu_{ik}^h}$.

Another extension to multiview FCM clustering with weighted views was proposed by Jiang et al. [21], called WV-Co-FCM, where they considered different weights per view, by considering the parameters to regulate the distribution of each view weights. The objective function of the WV-Co-FCM is given as

$$J_{WVCoFCM}(U, A, V) = \sum_{h=1}^{s} v_h \left[\sum_{i=1}^{n} \sum_{k=1}^{c} \left(\mu_{ik}^h \right)^m \left(d_{ik}^h \right)^2 + \Delta_h \right] + \lambda \sum_{h=1}^{s} v_h \ln v_h \tag{4}$$

where v_h is the hth view weight with $\sum_{h=1}^{s} v_h = 1$, $v_h \in [0,1]$,$\sum_{k=1}^{c} \mu_{ik} = 1 \mu_{ik} \in [0,1]$. $\lambda > 0$ is used to regulate the view weights, $d_{ik}^h = \sqrt{\sum_{j=1}^{d_h} \left(x_{ij}^h - a_{kj}^h \right)^2}$ and $\Delta_h = \sum_{i=1}^{n} \alpha_{ik}^h \sum_{k=1}^{c} \mu_{ik}^h \left(1 - \left(\mu_{ik}^h \right)^{m-1} \right) - \sum_{i=1}^{n} \beta_{ik}^h \sum_{k=1}^{c} \mu_{ik}^h \left(1 - \left(\mu_{ik}^h \right)^{m-1} \right)$, where $\alpha_{ik}^h = \eta \left(d_{ik}^h \right)^2$, and β_{ik}^h has four cases including $\beta_{ik}^h = \eta \frac{1}{s-1} \sum_{h\prime=1, h\prime \neq h}^{s} \left(d_{ik}^{h\prime} \right)^2$, $\beta_{ik}^h = \frac{\eta}{s} \sum_{h=1}^{s} \left(d_{ik}^h \right)^2$, $\beta_{ik}^h = \eta \min_{h\prime \neq h} \left\{ \left(d_{ik}^{h\prime} \right)^2 \right\}$, and $\beta_{ik}^h = \eta \sqrt[s-1]{\prod_{h\prime \neq h} \left(d_{ik}^{h\prime} \right)^2}$, and $0 < \eta < 1$ is a parameter used to control the penalty related to the disagreement. The term, Δ_h, aims at reducing the disagreement between organizations on different views. The clustering process of the WV-Co-FCM is based on minimizing the objective function that highlights the fuzzy partition for a dataset using the collaborative clustering technique, and the weight of each view that can be achieved by adding the entropy regularization term. The WV-Co-FCM is used to tackle the multiview data and the weights are utilized in the last step. That is, the membership for each object and cluster center is updated independently with the influence of weights.

The multiview FCM clustering algorithm without the collaboration step that considered different weights for its view is called MinMax-FCM, and it was proposed by Wang and Chen [22]. The MinMax-FCM is constructed based on the single-view FCM, which measures the minimum distance between membership matrix U^* and the cluster centers in each view A^h. Here, the maximum value within views V^h impacts the minimum separation between U^* and A^h. In this sense, the unanimous clustering results in the MinMax-FCM are produced by the MinMax optimization, in which the dissimilarities of the diverse views are reduced. The MinMax-FCM [22] is formulated as:

$$\min_{U^*, \{A^h\}_{h=1}^{s}} \max_{\{v_h\}_{h=1}^{s}} \sum_{h=1}^{s} (v_h)^\beta Q^h \tag{5}$$

where $Q^h = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(u_{ik}^* \right)^m \left\| x_i^h - a_k^h \right\|^2$ subject to $\sum_{k=1}^{c} u_{ik}^* = 1, 1 \leq i \leq n, u_{ik}^* \geq 0 \forall i, k$, $\sum_{h=1}^{s} v_h = 1$ and $v_h \geq 0, 1 \leq h \leq s$. The expression $\left(u_{ik}^* \right)^m \left\| x_i^h - a_k^h \right\|^2$ is considered as the cost of the hth view, which is the FCM objective function. v_h is the weight of hth view. The parameter $\beta \in (0,1)$ controls the distribution of weights v_h for different views, and $m > 1$ is the fuzzifier for fuzzy clustering, which controls the fuzziness of the membership.

In Co-FKM, there are two strategies. The first strategy is as follows: the average disagreement between any pairs of the views is integrated into the objective function. The second strategy is applied in which the final consensus fuzzy membership is generated by calculating the geometric mean of the membership of all views. Based on similar strategies

as applied in the Co-FKM, the WV-Co-FCM was proposed to handle multiview data. Like the Co-FKM, the fuzzy membership for each object in each view is first calculated in the WV-Co-FCM. An additional step is needed to calculate the final consensus membership. There are mainly three differences between the aforementioned two approaches. First, instead of using standard FCM, the WV-Co-FCM is based on GIFP-FCM (Zhu et al. [37]) in which the entropy is added to enhance the fuzzy membership. Second, the weight for each view is considered in the WV-Co-FCM and the entropy regularization term of the weight is introduced into the objective function. Third, instead of using the geometric mean in Co-FKM, the final consensus membership is generated based on the weight of each view. In the MinMax-FCM, the consensus clustering results are generated based on the MinMax optimization, in which the maximum disagreements of the different weighted views are minimized. According to Wang and Chen [22], the MinMax-FCM generates harmonic consensus clustering results integrating the heterogeneous views of data on the consensus memberships.

In order to make comparisons with the state of the art in multiview clustering algorithms, the algorithms discussed above have been used. The key relationships among them are summarized as follows: the Co-FKM serves as the foundational algorithm, and the WV-Co-FCM is a variation with additional features, while the MinMax-FCM is a separate algorithm that focuses on minimizing disagreements to achieve consensus clustering. Each algorithm has its unique characteristics and strategies for handling multiview data. In the multiview clustering algorithms, they always depend on initializations to achieve good results. All three multiview algorithms of Co-FKM, WV-Co-FCM, and MinMax-FCM try to calculate the consensus membership in different ways. The three multiview algorithms required a number of clusters a priori with the parameter selection. In fact, we need to have an algorithm that can tackle these issues and automatically find the number of clusters. This is our goal in this research.

In general, cluster validity indices (CVIs) are used to assess the suitability of the partitions shaped by clustering algorithms and are commonly used to determine an appropriate clusters number for the datasets [12]. The most used fuzzy CVIs linked to FCM are the partition coefficient (PC) [38] and the partition entropy (PE) [39], which both only use fuzzy memberships obtained by the FCM algorithm. Since the geometry of data cannot be taken into consideration by these indices, the fuzzy hypervolume (FHV) was introduced by Gath and Geva [40]. The XB validity index introduced by Xie and Beni [41] is based on fuzzy memberships and the geometrical structure of data with its objective function, and Wu et al. [42] gave the robust-type CVIs. However, the FHV and XB validity indices are not fitted for multiview FCM clustering algorithms even though we modify them by considering the multiview FCM objective functions. In this sense, for the multiview FCM clustering algorithms, we use PC, PE, and MPC as the validity indices.

The PC [38] is calculated by taking the overall summation of squared fuzzy memberships μ_{ik}^2. The total data points n is embedded to the controlling weights. Thus, PC associated with fuzzy c-memberships is defined as

$$PC(c) = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}^2 \tag{6}$$

It is known that $1/c \leq PC(c) \leq 1$. An optimal number c is determined by solving $\max_{2\leq c\leq n-1} PC(c)$. To normalize $PC(c)$ such that it is between 0 and 1, Roubens [43] modified $PC(c)$ as

$$MPC(c) = 1 - \frac{c}{c-1}(1 - PC(c)) \tag{7}$$

On the other hand, PE [39] is estimated by taking the entropy of fuzzy c-memberships μ_{ik}. The PE associated with fuzzy c-memberships is expressed as

$$PE(c) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{ik}\log_2\mu_{ik} \tag{8}$$

Generally, we determine an optimal c by simplifying $\min_{2\le c\le n-1} PE(c)$ to obtain an optimal cluster number.

3. The Proposed Unsupervised Multiview Fuzzy C-Means Clustering Algorithm

The existing clustering algorithms of k-means and FCM are the most popular among single-view data. It is a common fact that FCM is a fuzzy extension of the k-means clustering algorithm. However, most multiview clustering algorithms need prior information such as a number of clusters and the parameter selection. Therefore, there is a need for an algorithm that can handle all these issues and automatically find an optimal number of clusters in clustering the multiview data. For filling this important research gap, we introduce a novel U-MV-FCM algorithm that has the capacity to address the drawbacks of these existing algorithms. That has been demonstrated as follows.

Let $X = \{x_1, \ldots, x_n\}$ be a multiview dataset with $x_i = \left\{x_i^h\right\}_{h=1}^{s}$, $x_i^h \in \mathbb{R}^{d_h}$, and $x_i^h = \left\{x_{ij}^h\right\}_{j=1}^{d_h}$, $i = 1, \ldots, n$. Let $V = [v_h]_{1\times s}$, where v_h is the hth view weight and $A^h = \left\{a_1^h, \ldots, a_c^h\right\}$ is the set of the c cluster centers in the hth view with $a_k^h = \left\{a_{kj}^h\right\}$, $j = 1, \ldots, d_h$. Let $U = [\mu_{ik}]_{n\times c}$ be the fuzzy c-membership matrix with $\mu_{ik} \in [0,1]$ and $\sum_{k=1}^{c} \mu_{ik} = 1 \ \forall\, i$ where μ_{ik} is interpreted as the agreed membership of ith data point in the cluster centers k shared across different view h. As the prime objective of our proposed unsupervised-regularization structure is to determine an optimal number of clusters for multiview FCM, we use the concept of robust-learning FCM proposed by Yang and Nataliani [35]. We first design an adequate membership architecture across different views to the entire multiview dataset, and then add more penalty terms to construct the unsupervised-regularization structure. Thus, we construct the following unsupervised multiview FCM (U-MV-FCM) objective function:

$$\begin{aligned} J_{U-MVFCM}(U, A^h, V, \alpha) = [&\sum_{h=1}^{s} v_h^{\beta} \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^h - a_k^h \right\|^2 - \eta_1 \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \ln \alpha_k \\ &+ \eta_2 \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \ln \mu_{ik} - \eta_3 \sum_{i=1}^{n} \sum_{k=1}^{c} \alpha_k \ln \alpha_k] \end{aligned} \tag{9}$$

subject to $\sum_{k=1}^{c} \mu_{ik} = 1, \forall i$, $\mu_{ik} \in [0,1]$, $\sum_{h=1}^{s} v_h = 1$, $v_h \in [0,1]$, and $\sum_{k=1}^{c} \alpha_k = 1$, $\alpha_k \in [0,1]$, where $\left\| x_i^h - a_k^h \right\|^2 = \sum_{j=1}^{d_h} \left(x_{ij}^h - a_{kj}^h \right)^2$, $\eta_1, \eta_2, \eta_3 > 0$, and β is a view-weight exponent. We mention that α_k is used to represent the probability of the kth cluster in the c clusters that will be utilized to determine an optimal number of clusters. The first term of $J_{U-MVFCM}(U, A^h, V, \alpha)$ is the only simple objective function of the multiview FCM, and the other three penalty terms are considered as unsupervised-regularization by using the parameters η_1, η_2, and η_3 to adjust these penalties. We will give the estimates of η_1, η_2, η_3 and β later.

In general, it is challenging to simplify the variables μ_{ik}, v_h, a_k^h, and α_k in Equation (9) directly, as Equation (9) is nonconvex. Yet we observe that the objective function is convex with respect to μ_{ik}, v_h, α_k and is concave with respect to a_k^h and therefore alternative optimization (AO) is utilized for the optimization problem by fixing other variables while solving the one.

Theorem 1. *The updating equations to find the necessary conditions for minimization of the U-MV-FCM objective function $J_{U-MVFCM}(U, A^h, V, \alpha)$ of Equation (9) are*

$$\mu_{ik} = \frac{e^{\frac{1}{\eta_2}\left(\eta_1 \ln \alpha_k - \sum_{h=1}^{s} v_h^{\beta} \left\| x_i^h - a_k^h \right\|^2\right)}}{\sum_{k\prime=1}^{c} e^{\frac{1}{\eta_2}\left(\eta_1 \ln \alpha_{k\prime} - \sum_{h=1}^{s} v_h^{\beta} \left\| x_i^h - a_{k\prime}^h \right\|^2\right)}} \tag{10}$$

$$v_h = \left(\sum_{h'=1}^{s} \left(\frac{\sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^h - a_k^h \right\|^2}{\sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^{h'} - a_k^{h'} \right\|^2} \right)^{\frac{1}{\beta-1}} \right)^{-1} \tag{11}$$

$$a_{kj}^h = \sum_{i=1}^{n} \mu_{ik} x_{ij}^h / \sum_{i=1}^{n} \mu_{ik} \tag{12}$$

$$\alpha_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \mu_{ik} + \frac{\eta_3}{\eta_1} \alpha_k^{(t)} \left(\ln \alpha_k^{(t)} - \sum_{k\prime=1}^{c} \alpha_{k\prime}^{(t)} \ln \alpha_{k\prime}^{(t)} \right) \tag{13}$$

Proof. The Lagrangian Multiplier technique is employed to resolve the optimization problem regarding the proposed U-MV-FCM w.r.t. $\mu_{ik}, v_h, a_k^{(h)}$ and α_k. The Lagrangian for $J_{U-MVFCM}(U, A^h, V, \alpha)$ is expressed as $\widetilde{J}(V, A^h, U, \alpha) = J_{U-MVFCM} - \lambda_1(\sum_{k=1}^{c} \mu_{ik} - 1) - \lambda_2(\sum_{k=1}^{c} \alpha_k - 1) - \lambda_3(\sum_{h=1}^{s} v_h - 1)$.

By taking the partial derivative of the Lagrangian $\widetilde{J}(V, A^h, U, \alpha)$ w.r.t. μ_{ik} equal to zero, we obtain $\frac{\partial J}{\partial \mu_{ik}} = \sum_{h=1}^{s} v_h^\beta \left\| x_i^h - a_k^h \right\|^2 - \eta_1 \ln \alpha_k + \eta_2(\ln \mu_{ik} + 1) - \lambda_1 = 0$. Then, we have $\ln \mu_{ik} = \frac{1}{\eta_2} \left(\eta_1 \ln \alpha_k - \sum_{h=1}^{s} v_h^\beta \left\| x_i^h - a_k^h \right\|^2 + \lambda_1 - \eta_2 \right)$. Thus, we have $\mu_{ik} = e^{\frac{1}{\eta_2}(\eta_1 \ln \alpha_k - \sum_{h=1}^{s} v_h^\beta \| x_i^h - a_k^h \|^2)} e^{\frac{\lambda_1 - \eta_2}{\eta_2}}$. Since $\sum_{k=1}^{c} \mu_{ik} = 1$, we obtain $e^{\frac{\lambda_1 - \eta_2}{\eta_2}} = 1 / \sum_{k=1}^{c} e^{\frac{1}{\eta_2}(\eta_1 \ln \alpha_{k\prime} - \sum_{h=1}^{s} v_h^\beta \| x_i^h - a_{k\prime}^h \|^2)}$. Hence, the updating equation for μ_{ik} can be obtained as $\mu_{ik} = e^{\frac{1}{\eta_2}(\eta_1 \ln \alpha_k - \sum_{h=1}^{s} v_h^\beta \| x_i^h - a_k^h \|^2)} / \sum_{k=1}^{c} e^{\frac{1}{\eta_2}(\eta_1 \ln \alpha_{k\prime} - \sum_{h=1}^{s} v_h^\beta \| x_i^h - a_{k\prime}^h \|^2)}$. and the updating Equation (10) for μ_{ik} is obtained. Similarly, the partial derivative of the Lagrangian $\widetilde{J}\left(V, U, A^h, \alpha\right)$ w.r.t. v_h and if we make it equal to zero, we can obtain the equation $\frac{\partial \widetilde{J}}{\partial v_h} = \beta v_h^{\beta-1} \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^h - a_k^h \right\|^2 - \lambda_3 = 0$. Thus, we have $v_h = (\lambda_3)^{\frac{1}{\beta-1}} \left(\beta \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^h - a_k^h \right\|^2 \right)^{-\frac{1}{\beta-1}}$. Since $\sum_{h'=1}^{s} v_h = 1,\ v_h \in [0,1]$, we obtain $(\lambda_3)^{\frac{1}{\beta-1}} = \left(\left(\beta \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} \left\| x_i^{h'} - a_k^{h'} \right\|^2 \right)^{-\frac{1}{\beta-1}} \right)^{-1}$. Thus, the updating Equation (11) for v_h is obtained. The derivative of the U-MV-FCM objective function $J_{UR-MVFCM}$ w.r.t. a_k^h and equal to zero, we can obtain the equation $\frac{\partial J_{U-MVFCM}}{\partial a_{kj}^h} = -2 v_h^\beta \sum_{i=1}^{n} \mu_{ik} \left(x_{ij}^h - a_{kj}^h \right) = 0$, and then $\sum_{i=1}^{n} \mu_{ik} x_{ij}^h = \sum_{i=1}^{n} \mu_{ik} a_{kj}^h$. Thus, the updating Equation (12) for a_k^h can be obtained as $a_{kj}^h = \sum_{i=1}^{n} \mu_{ik} x_{ij}^h / \sum_{i=1}^{n} \mu_{ik}$. We next take the partial derivative of Lagrangian $\widetilde{J}\left(V, A^h, U, \alpha\right)$ w.r.t. α_k and make them equal to zero. We obtain the equation $\frac{\partial J}{\partial \alpha_k} = -\eta_1 \sum_{i=1}^{n} \mu_{ik} \left(\frac{1}{\alpha_k} \right) - \eta_3 n(\ln \alpha_k + 1) - \lambda_2 = 0$, and then $-\eta_1 \sum_{i=1}^{n} \mu_{ik} - n\eta_3 \alpha_k \ln \alpha_k - n\eta_3 \alpha_k - \lambda_2 \alpha_k = 0$. Thus, we have $-\eta_1 \sum_{i=1}^{n} \sum_{k=1}^{c} \mu_{ik} - n\eta_3 \sum_{k=1}^{c} \alpha_k \ln \alpha_k - n\eta_3 \sum_{k=1}^{c} \alpha_k - \lambda_2 \sum_{k=1}^{c} \alpha_k = 0$. with $\lambda_2 = -n\eta_1 - n\eta_3 \sum_{k=1}^{c} \alpha_k \ln \alpha_k - n\eta_3$. We obtain $-\eta_1 \sum_{i=1}^{n} \mu_{ik} - n\eta_3 \alpha_k \ln \alpha_k - n\eta_3 \alpha_k - \alpha_k \left(-n\eta_1 - n\eta_3 \sum_{k'=1}^{c} \alpha_{k'} \ln \alpha_{k'} - n\eta_3 \right) = 0$ and then we find the updating Equation (13) for

α_k as $\alpha_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\mu_{ik} + \frac{\eta_3}{\eta_1}\alpha_k^{(t)}\left(\ln\alpha_k^{(t)} - \sum_{k'=1}^{c}\alpha_{k'}^{(t)}\ln\alpha_{k'}^{(t)}\right)$, where t denotes the iteration number in the algorithm. □

The data are commonly used as a potential source for reaching an educated decision. In a multiview data scenario, the high-level detail information is used to capture the correlation between the data features in the hth view. It is known that determining an optimal number of clusters for the multiview FCM clustering algorithm is quite difficult, and no work has been written yet in the literature. This process usually requires much more effort to understand the data, especially in real world application. To make proper assumptions, it is essential to have domain expertise in a specific application. In practice, to assign a specific value to the number c of clusters is needed on the base of proprietary data insights. However, different datasets may result in different c values. It becomes difficult in real applications, like biological applications with Prokaryotic data analysis. This variability becomes particularly prominent. Thus, the regularization of the proposed U-MV-FCM clustering algorithm is designed to accommodate varying preferences and conceptual frameworks (Figure 1). It is known that researchers had investigated the Prokaryotic dataset with its actual number c as 4. Based on these insights, the users will specify c = 4 in the implementation of most MV-FCM clustering algorithms for Prokaryotic data. However, without giving the number c of clusters, U-MV-FCM can potentially obtain the optimal number of clusters with c = 4.

In the U-MV-FCM objective function $J_{U-MVFCM}(U, A^h, V, \alpha)$, we have four parameters η_1, η_2, η_3 and β. We now use two artificial datasets to analyze the behaviors of these parameters and then give their estimations.

Artificial Data 1. *A two-view numerical dataset with two clusters & two feature components is considered. Gaussian mixture model (GMM) is used to generate a two-component two-variate data point for each view where their probability proportions are* $\alpha_1^{(1)} = \alpha_1^{(2)} = 0.7$ *and* $\alpha_2^{(1)} = \alpha_2^{(2)} = 0.3$. *The means* $\mu_k^{(1)}$ *for the first view are (5, 6) and (12, 6). The means* $\mu_k^{(2)}$ *for the second view are (6, 12) & (6, 5). For the two views, covariance matrices are* $\sum_1^{(1)} = \sum_1^{(2)} = \sum_2^{(1)} = \sum_2^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. *The* $x_1^{(1)}$ *and* $x_2^{(1)}$ *are the coordinates for the first view,* $x_1^{(2)}$ *&* $x_2^{(2)}$ *denotes the coordinates for the second view, as displayed in Figure 2a and Figure 2b, respectively.*

Artificial Data 2. *A three-view numerical dataset with five clusters and two feature components is considered. The data points in each view are generated from a two-component two-variate Gaussian mixture model (GMM) with their mixing proportions* $\alpha_1^{(1)} = \alpha_1^{(2)} = \alpha_1^{(3)} = 0.1, \alpha_2^{(1)} = \alpha_3^{(1)} = \alpha_2^{(2)} = \alpha_3^{(2)} = \alpha_2^{(3)} = \alpha_3^{(3)} = 0.25$ *and* $\alpha_4^{(1)} = \alpha_5^{(1)} = \alpha_4^{(2)} = \alpha_5^{(2)} = \alpha_4^{(3)} = \alpha_5^{(3)} = 0.2$. *The means* $\mu_k^{(1)}$ *for the first view are* $(4.580 \quad 3.344), (3.8730 \quad 2.637), (1.752 \quad 1.930), (1.752 \quad 3.344), (1.044 \quad 2.637)$. *The means* $\mu_k^{(3)}$ *for the third view are* $(2 \quad 5), (0.65 \quad 0.67), (3.10 \quad 2.48), (2 \quad 3.5), (1 \quad 2.48)$. *These* $x_1^{(1)}$ *&* $x_2^{(1)}$ *are the coordinates for the first view,* $x_1^{(2)}$ *&* $x_2^{(2)}$ *denotes the second view coordinates, and* $x_1^{(3)}$ *&* $x_2^{(3)}$ *denotes the third view coordinates, as presented in Figure 3a–c, respectively. As can be seen, the two views of the* $x_1^{(1)}$ *&* $x_2^{(1)}$ *and* $x_1^{(2)}$ *&* $x_2^{(2)}$ *are mirror images of each other. So it is clear that the data points* $x_1^{(2)}$ *&* $x_2^{(2)}$ *for the second view are generated on the basis of inverse data points of* $x_1^{(1)}$ *&* $x_2^{(1)}$.

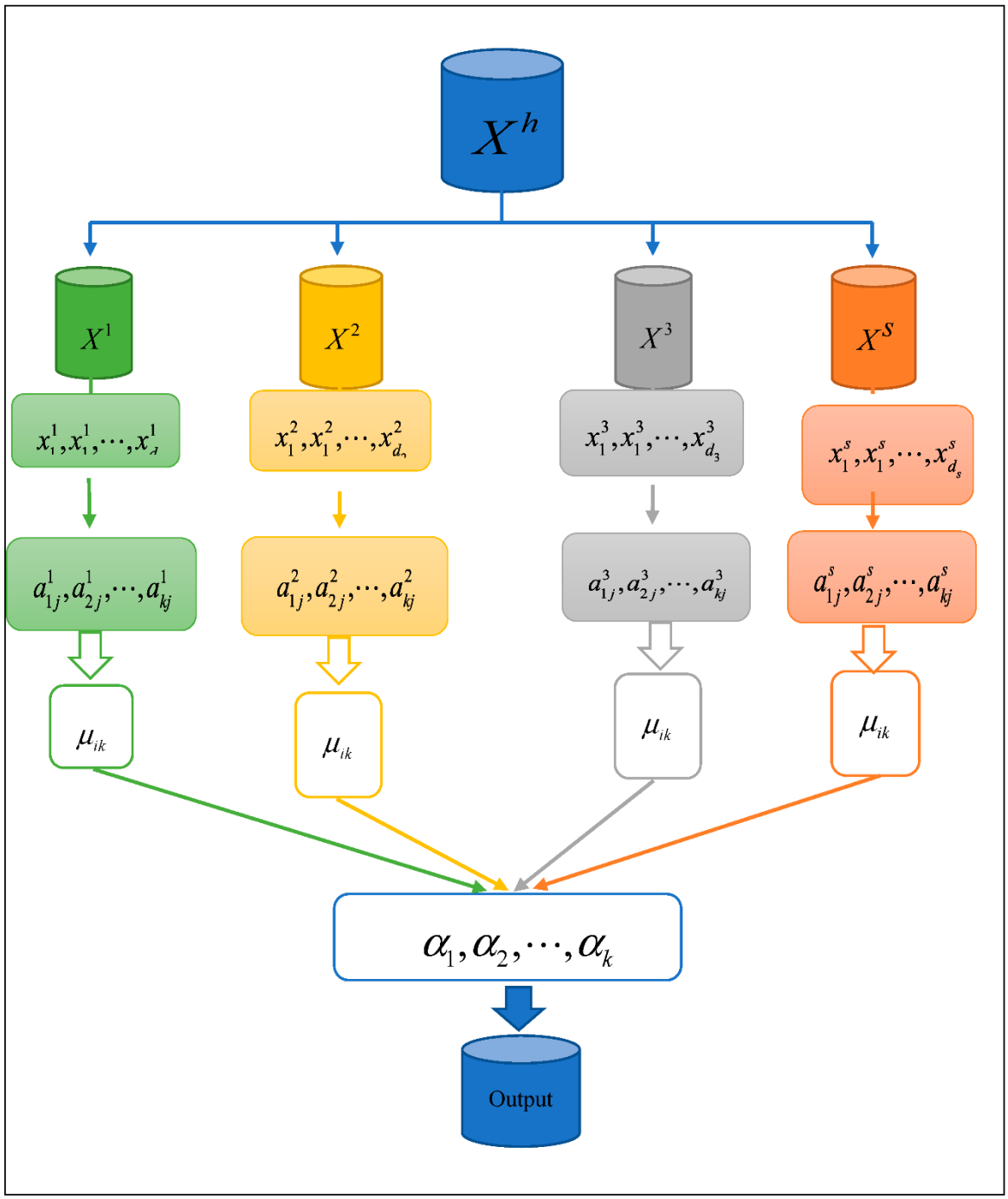
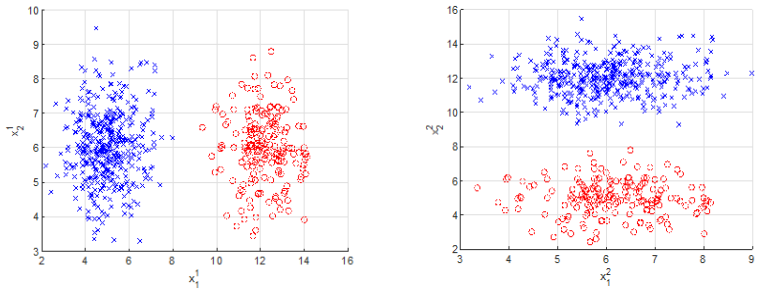
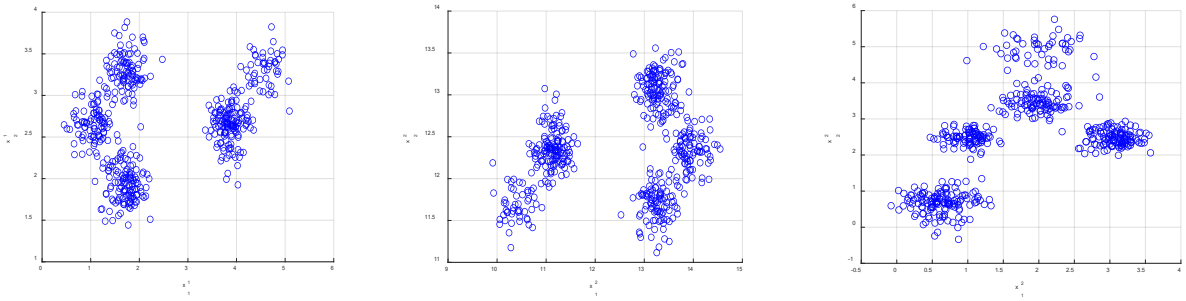


Figure 1. Data Processing of the proposed U-MV-FCM framework.



(**a**) (**b**)

Figure 2. The two-view two-cluster dataset for (**a**) first view; (**b**) second view.



(**a**) (**b**) (**c**)

Figure 3. The three-view-five-cluster dataset for (**a**) first view; (**b**) second view; (**c**) third view.

Parameter β: We know that the parameter β is a view-weight exponent. Our main idea is to find an optimal clusters number and also to estimate the importance of different views in multiview datasets. When we simulate with different values of β for the Artificial Datasets 1 and 2, we find that it makes a difference for the proposed U-MV-FCM. That is, if the values are too small or too large, then the overall accuracy rates will be decreasing. Intuitively, a small or a large β can represent an imbalanced importance for one view. Thus, the issue is: how to estimate β so that it can represent the importance of different views. This question can be answered by designing an appropriate balancing estimator for β. In this sense, a balance movement of β values can control the distribution of one view in the multiview scenario for simultaneously finding a good number of clusters. To test this hypothesis, we implement our balancing estimator of β by using the above Artificial Datasets 1 and 2. This idea can be addressed by considering the minimum weights of the probability proportion k into account. As at every iteration the values of mixing proportion k is changed, the necessary condition estimating β is to have a sufficient number that generated from $\mu_{ik}, a_k^{(t)}$ and $a_{k'}^{(t)}$. In machine learning, it is a common knowledge that π can be used to measure the identifiability of the data. Therefore, the view-weight exponent β will algebraically consider the minimal solution of $\alpha_k^{(t-1)}$ and then construct it with π value. In this sense, the infinite series of π is suitable to adjust the dimensions of each view into well-structured shapes. The values of β based on the movement of the probability proportion kth shared across different views combined with π value can be presented in the form

$$\beta^{(t)} = 2\pi \arg\min_{1\le k\le c}[\alpha_k^{(t-1)}] \tag{14}$$

where $\arg\min_{1\le k\le c}[\ ]$ stands for the argument for which the minimum of $\alpha_k^{(t-1)}$ is attained, and that is a real and positive number. To be noted $\alpha_k^{(t+1)}$ it must be derived based on the constant n, η_1, η_3, μ_{ik}, $\alpha_k^{(t)}$, and $\alpha_{k'}^{(t)}$. In this sense, the value of β not only depends on the input of the k and k' current probability proportion $\alpha_k^{(t)}, \alpha_{k'}^{(t)}$, but also depends on the

k and k' value of the probability proportion at the previous time $\alpha_k^{(t-1)}$, $\alpha_{k'}^{(t-1)}$. Using π and the minimal solution of α to estimate β enables the proposed U-MV-FCM algorithm to simultaneously approximate and reproduce the qualifying number of clusters such that the Euclidean distances $d_{ik} = \|x_i - a_k\|$ of hth components are minimized while $\alpha_k^{(t)}$ is maximized. In this sense, the data points will be well-structured and easy to be classified.

Parameters η_1, η_2: One of the motivations for this simulation of balancing the parameters is to tune the level in determining the number of clusters in each iteration. A good combination value between η_1 and η_2 needs to be in general stable, scalable, and fit to any types of multiview data. Usually, the functions of e^{-t}, $e^{-t/150}$, $e^{-t/250}$,$e^{-t/400}$,$e^{-t/600}$ and $e^{-t/1000}$ are used as the learning functions in which $y = e^{-t}$ decreases faster, but $e^{-t/1000}$ decreases slower at each iteration t. According to the penalty term $\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{ik}\ln\alpha_k$, it has effects on μ_{ik} and α_k. Thus, the learning behavior for η_1 should not decrease too slowly or too fast. Further, the penalty term $\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{ik}\ln\mu_{ik}$ is the entropy on μ_{ik}, and so the learning function for η_2 should decrease fast. Thus, the decreasing learning functions for η_1 and η_2 are chosen as follows:

$$\eta_1 = e^{-\frac{t}{150}} \tag{15}$$

$$\eta_2 = e^{-\frac{t}{400}} \tag{16}$$

The terms in (15) and (16) can accelerate the process in finding the number of clusters during iterations in the algorithm. To address this issue, taking the square root of the iteration t enables the U-MV-FCM algorithm to reduce the number of clusters in each iteration. To assist the smooth movement, we use the following η_1 and η_2 for the U-MV-FCM algorithm:

$$\eta_1 = e^{-\frac{t}{150}}\frac{1}{\sqrt{t}} \tag{17}$$

$$\eta_2 = e^{-\frac{t}{400}}\frac{1}{\sqrt{t}} \tag{18}$$

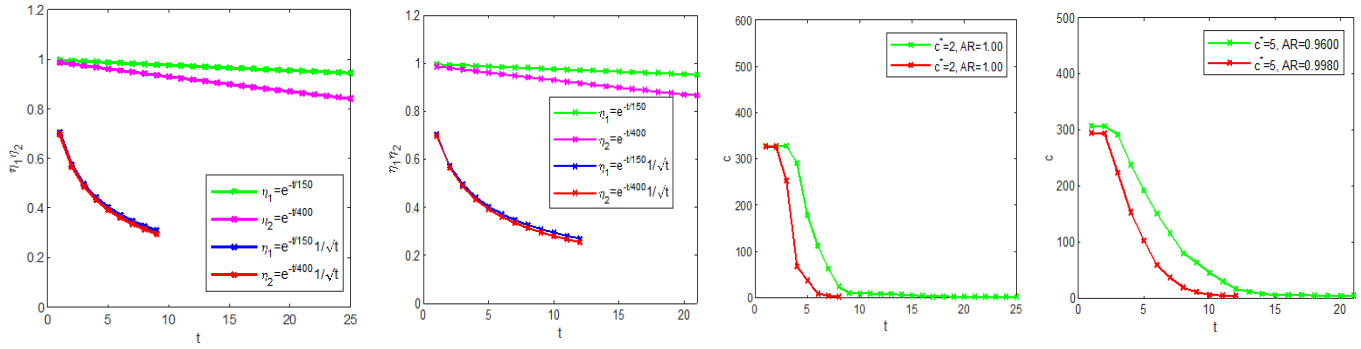
After updating the number of clusters c, the remaining mixing proportion $\alpha_{k\prime}$ and corresponding $\mu_{ik\prime}$ need to be re-normalized by

$$\alpha_{k\prime} = \frac{\alpha_{k\prime}}{\sum_{\updownarrow=1}^{c^{t+1}}\alpha_{\updownarrow}} \tag{19}$$

$$\mu_{ik\prime} = \frac{\mu_{ik\prime}}{\sum_{\updownarrow=1}^{c^{t+1}}\mu_{i\updownarrow}} \tag{20}$$

For the two parameters η_1 and η_2 in Figure 4, we consider four decreasing learning rates mentioned in Equations (15)–(18). From Figure 4a,b, we can see that, as the starting steps, the decreasing rates based on Equations (15) and (16) are slower than that based on Equations (17) and (18). However, the decreasing rates based on Equations (15) and (16) are faster than that based on Equations (17) and (18) when the algorithm becomes stable. In this sense, we suppose that the learning rates of Equations (17) and (18) should be better than the learning rates of Equations (15) and (16) for our proposed U-MV-FCM algorithm. For further observations, we represent the processes of the newly proposed U-MV-FCM in determining the optimal number of clusters c^* by using the learning rates of Equations (17) and (18) as presented in Figure 4c,d, in which Figure 4c is for Artificial Dataset 1 and Figure 4d is for Artificial Data 2. It is clear that Artificial Data 1 initially uses n data points as the number of clusters c, i.e., $c = n = 600$. After the second iteration, the number of clusters reduces from 600 to 327 and then it stops at the iteration $t = 9$ with the optimal number of clusters $c^* = 2$, as portrayed in Figure 4c. Given the result for Artificial Data 2 with initial $c = n = 500$, it successfully shows that the proposed U-MV-FCM can produce the optimal number $c^* = 5$ of clusters at the iteration $t = 12$, as shown in Figure 4d. These results

reflect the parameters η_1 and η_2 with Equations (17) and (18) can obtain the correct optimal number of clusters for Artificial Data 1 and 2. In general, the parameters η_1 and η_2 with Equations (17) and (18) as the learning rates are more efficient than those with Equations (15) and (16) for the proposed U-MV-FCM. In this sense, we use Equations (17) and (18) as the learning rates in our proposed U-MV-FCM algorithm.



(**a**) (**b**) (**c**) (**d**)

Figure 4. Parameter sensitivity of η_1, η_2 with Equations (15) and (18) for (**a**) Artificial Data 1; (**b**) Artificial Data 2; Further, the iterative c^* when processing the U-MV-FCM in determining the optimal number of clusters c^* with the learning rates of Equations (17) and (18) for (**c**) Artificial Data 1; (**d**) Artificial Data 2.

Parameter η_3: Furthermore, the Parameter η_3 is the balancing parameter to control the competition. Here, we directly use the restriction of $-e^{-1} \leq \alpha_k \ln \alpha_k < 0$. First, if $0 < \alpha_k \leq 1 \forall\, k$ and $E = \sum_{l=1}^{c} \alpha_l \ln \alpha_l < 0$, then we have $\alpha_k E = \alpha_k \sum_{l=1}^{c} \alpha_l \ln \alpha_l < 0$. Second, using $-e^{-1} \leq \alpha_k \ln \alpha_k < 0$ and $\alpha_k E = \alpha_k \sum_{l=1}^{c} \alpha_l \ln \alpha_l < 0$, we obtain $-e^{-1}\eta_3 < \eta_3 \alpha_k \left(\ln \alpha_k - \sum_{l=1}^{c} \alpha_l \ln \alpha_l \right) < \eta_3(-\alpha_k E)$. Third, we apply $\sum_{k=1}^{c} \alpha_k = 1$ and $\alpha_k < 1/2$, then we can obtain $\left(\ln \alpha_k - \sum_{l=1}^{c} \alpha_l \ln \alpha_l \right) < 0$. Fourth, because $\alpha_k > 0$ then $-e^{-1}\eta_3 > -\max\{\alpha_k | \alpha_k < 1/2, k = 1, 2, \ldots, c\}$. So we obtain $\eta_3 < \max\{\alpha_k e | \alpha_k < 1/2, k = 1, 2, \ldots, c\} < e/2$. Note that η_3 is a solution to the matric α_k, then justifying an appropriate input value of this η_3 is essential. The estimation of η_3's can easily be formulated, if the dispersion between $\alpha_k^{(t+1)}$ and $\alpha_k^{(t)}$ is not too small, and to enhance the competition, it must be large. If the dispersion between $\alpha_k^{(t+1)}$ and $\alpha_k^{(t)}$ is large, then to maintain stability η_3 will be small. Thus, we introduce an updating equation for η_3 as

$$\eta_3 = \frac{\sum_{k=1}^{c} \exp\left\{ -rn \left| \alpha_k^{t+1} - \alpha_k^{t} \right| \right\}}{c}, \tag{21}$$

In addition, we consider the restriction $\max_{1 \leq k \leq c} \alpha_k^{(t+1)} \leq 1$. But, $\max_{1 \leq k \leq c} \alpha_k^{(t+1)} \leq \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) + \frac{\eta_3}{\eta_1} \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(\ln \max_{1 \leq k \leq c} \alpha_k^{(t)} - \sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right) \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) + \eta_3 \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(-\sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right) \Big], \left[\max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) + \frac{\eta_3}{\eta_1} \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(\ln \max_{1 \leq k \leq c} \alpha_k^{(t)} - \sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right) \right. < \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) + \eta_3 \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(-\sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right) \Big]$. Thus, if $\max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) + \eta_3 \max_{1 \leq k \leq c} \alpha_k^{(t)} \left(-\sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right) \leq 1$, then the restriction will be held. It employs that

$$\eta_3 \leq \frac{\left(1 - \max_{1 \leq k \leq c} \left(\frac{1}{n} \sum_{i=1}^{n} \mu_{ik} \right) \right)}{\left(-\max_{1 \leq k \leq c} \alpha_k^{(t)} \sum_{\updownarrow=1}^{c} \alpha_{\updownarrow}^{(t)} \ln \alpha_{\updownarrow}^{(t)} \right)} \tag{22}$$

Thus, by combining Equations (21) and (22), we obtain

$$\eta_3 = \min\left(\frac{\sum_{k=1}^{c}\exp\left(-rn\left|\alpha_k^{(t+1)} - \alpha_k^{(t)}\right|\right)}{c}, \frac{1 - \max\limits_{1\le k\le c}\left(\frac{1}{n}\sum_{i=1}^{n}\mu_{ik}\right)}{\left(-\max\limits_{1\le k\le c}\alpha_k^{(t)}\sum_{\updownarrow=1}^{c}\alpha_{\updownarrow}^{(t)}\ln\alpha_{\updownarrow}^{(t)}\right)}\right) \tag{23}$$

We use Figure 5 to demonstrate the parameter sensitivity of η_1, η_2, η_3, and β for Artificial Datasets 1 and 2.

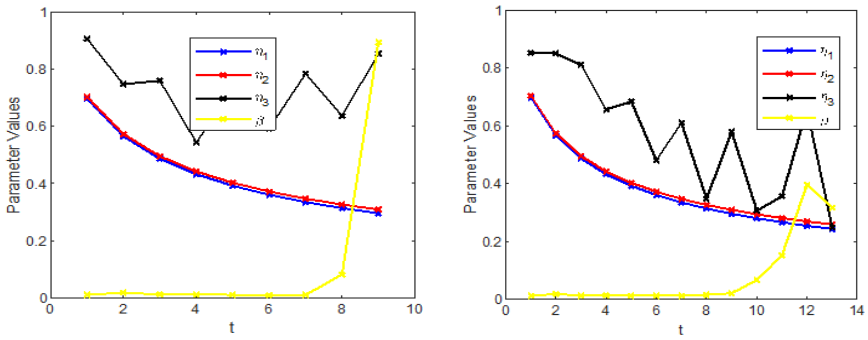


Figure 5. Parameter sensitivity of η_1, η_2, η_3, and β for (**a**) Artificial 1; (**b**) Artificial 2.

The formulation of multiview learning is significant as the growing of data is massive. Each data point can be represented by different views and hence various restrictions including the parameter estimations are required. In the U-MV-FCM, we propose several parameter estimators to explore patterns in the multiview data. There are many possible combinations of the parameters to process the data, but not all combinations can perform well. Thus, the multiview data can be patterned by using hundreds of possible combinations of the parameters before choosing the one that reveals meaningful insights behind the data. Our proposed U-MV-FCM aims to find the optimal number of clusters with high-quality cluster centroids in an efficient manner. One of the important parameters to determine the success of the U-MV-FCM is r. In relation to η_3 estimator in (23), we also consider r as a nonzero element. We then formulate r as $r = \min\limits_{1\le h\le s}\left\{\min\left\{1, 2/t * d_h^{\lfloor t*d_h/2-1\rfloor}\right\}\right\}$, where $\lfloor a\rfloor$ denotes the largest integer that is no more than a. Thus, the proposed U-MV-FCM clustering algorithm is presented in Algorithm 1:

We next demonstrate the parameters and cluster behaviors for the proposed U-MV-FCM clustering algorithm. Parameter selection is one of the important step in clustering data, as the high-quality structured data will be obtained if the choice of the combination parameters is provided. The relation and interaction between η_1, η_2, η_3, and β can be parameterized and lead the proposed U-MV-FCM to give the optimal number of clusters c. As displayed in Figure 6a–d, the proposed U-MV-FCM reduced the clusters number for both views that are the first and second views of Artificial 1 from originally c = 600 to c = 327 after t = 2 and continuously reduced to c^* = 2 after t = 9 with AR = 1.00. For Artificial 2 as shown in Figure 7a–f, the proposed U-MV-FCM processed the data from the three views from originally c = 500 to 295 after the second iteration and converged after fourteen iterations with c^* = 5 and AR = 0.9980.

Algorithm 1. The U-MV-FCM clustering algorithm

Input: Dataset $X = \{x_1, \ldots, x_n\}$ with $x_i = \left\{x_i^h\right\}_{h=1}^{s}$ and $x_i^h = \left\{x_{ij}^h\right\}_{j=1}^{d_h}$.

Output: a_{kj}^h, μ_{ik},v_h and α_k.

Initialization: Give initial $c^{(0)} = n$, $\alpha_k^{(0)} = 1/n$, $a_{k,h}^{(0)} = x_i$, initial $\eta_3^{(0)} = 1$, initialize view weight $V^{(t)} = [v_h]_{1\times s}$ (user may define $v_h = 1/s \forall h$), and set iteration counter $t = 0$, and fit $\varepsilon > 0$.

Step 1: Calculate $\eta_1^{(t)}, \eta_2^{(t)}$ by Equations (17) and (18).

Step 2: Calculate $\beta^{(t)}$ by Equation (14).

Step 3: Compute the membership $U^{(t)}$ using $A^{h^{(t-1)}}$, $V^{h^{(t-1)}}$, $\alpha^{(t-1)}$, $c^{(t-1)}$, $\eta_1^{(t)}, \eta_2^{(t)}$, and $\beta^{(t)}$ by Equation (10).

Step 4: Update $\alpha^{(t)}$ with$\eta_1^{(t)}, \eta_3^{(t)}$, $U^{(t)}$ and $\alpha^{(t-1)}$ by Equation (13).

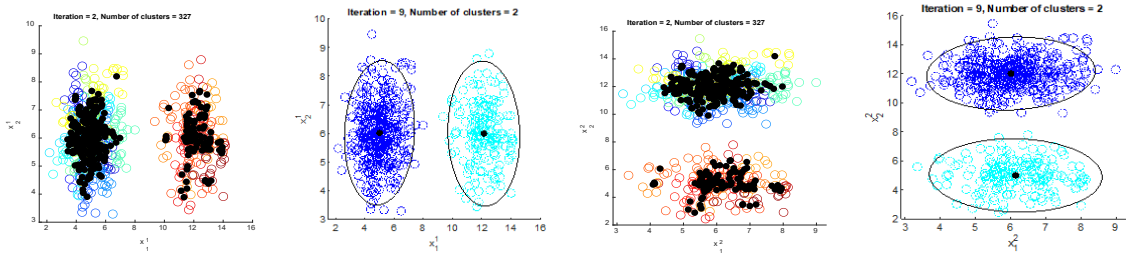
Step 5: Compute $\eta_3^{(t)}$ with $\alpha^{(t)}$ and $\alpha^{(t-1)}$ by Equation (23).

Step 6: Update $c^{(t-1)}$ to $c^{(t)}$ by discarding those clusters with $\alpha^{(t)} \leq 1/n$ and adjust $\alpha^{(t)}$ and $U^{(t)}$ by Equations (19) and (20). IF $t \geq 60$ and $c^{(t-60)} - c^{(t)} = 0$, THEN let $\eta_3^{(t)} = 0$.

Step 7: Update the view weight $V^{h^{(t)}}$ using $c^{(t)}$, $A^{h^{(t-1)}}$, $\beta^{(t)}$, and $U^{(t)}$ by Equation (11).

Step 8: Update $A^{h^{(t)}}$ with $c^{(t)}$ and $U^{(t)}$ by Equation (12).

Step 9: IF $\left|\left\|A^{h^{(t)}}\right\| - \left\|A^{h^{(t-1)}}\right\|\right| < \varepsilon$, then stop;
ELSE let $t = t + 1$ and return to Step 1.



(**a**) (**b**) (**c**) (**d**)

Figure 6. (**a**) Processes of U-MV-FCM for first view after the second iteration; (**b**) Convergence results for first view; (**c**) Process of U-MV-FCM for second view after the second iteration; (**d**) Convergence results for second view.

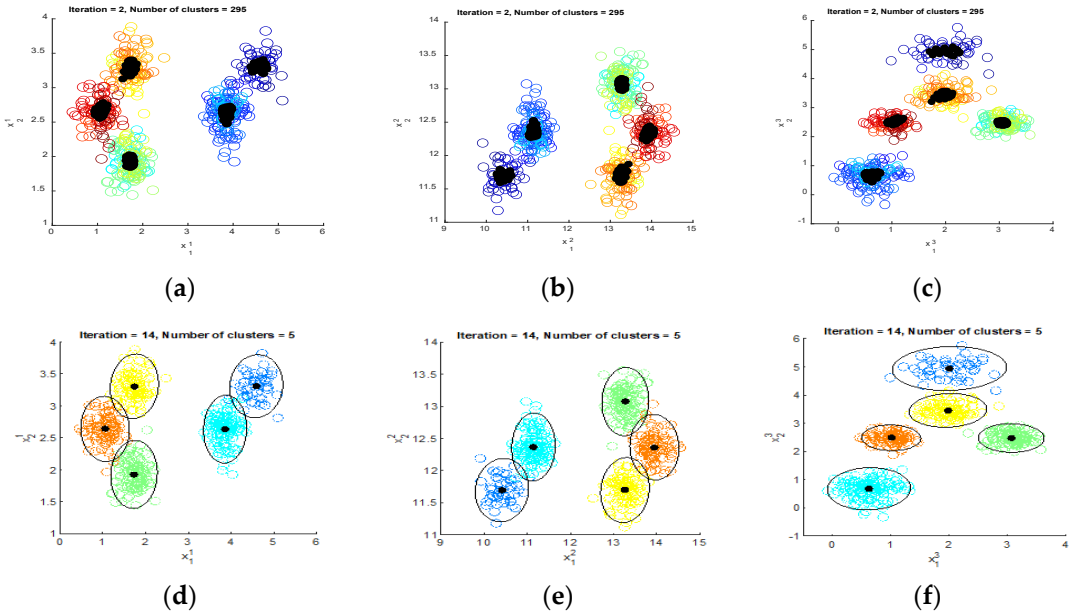


Figure 7. Processes of the U-MV-FCM after second iteration for: (**a**) first view; (**b**) second view; (**c**) third view; Convergence results for: (**d**) first view; (**e**) second view; (**f**) third view.

4. Experimental Results and Comparisons

In this section, we show the comparison results of the proposed U-MV-FCM with the the Co-FKM, MinMax-FCM, and WV-Co-FCM algorithms with consideration of biological, image, text, and webpage news datasets. The six benchmark datasets used are Wikipedia Articles [44,45], Prokaryotic [46], WebKB [47,48], 3-Sources [49], Reuters [50] and Extended YaleB [51]. These three artificial datasets such as Artificial 1, Artificial 2, and Syn500 [52] are also involved in experiments to evaluate our proposed U-MV-FCM algorithm. Table 2 represents the brief descriptions of these six benchmark dataset numbers in terms of data types, view s, cluster c, data points n, and feature dimension d_h.

Table 2. Characteristics of the MV datasets: name, number of clusters c, number of views v, the feature dimension d_h, number of data points n and data types.

Dataset	Prokaryotic	Wikipedia Articles	3-Sources	Reuters	WebKB	Extended YaleB
c	4	10	6	6	3	10
n	552	693	169	1200	203	640
s	3	2	3	5	3	3
d_1	393	128	3560	2000	1703	1024
d_2	3	10	3631	2000	230	256
d_3	-	-	-	2000	-	-
d_4	-	-	-	2000	-	-
d_5	-	-	-	2000	-	-
Type	Biological	Text	Text	Text	Web Page	Face

4.1. Performance Evaluation of MV-FCM Clustering Algorithms

In this subsection, the proposed U-MV-FCM and these existing algorithms such as the Co-FKM, MinMax-FCM, and WV-Co-FCM are validated by using the three artificial datasets and six benchmark multiview datasets. In order to evaluate the performance of the aforementioned MV-FCM clustering algorithms, we utilize the performance evaluation measures including AR (accuracy rate), RI (rand index) [53], FMI (Fowlkes and Mallows

index) [54], NMI (normalized mutual information) [55], and JI (Jaccard index) [56]. In general, these performance evaluation measures are the most commonly used indices in the literature. The values of these external validation measures range from 0 to 1. The higher the rate, the better the performance. For a fair comparison, simulation for the Co-FKM, MinMax-FCM, and WV-Co-FCM would be designed over 51 different random initializations. For quantitative evaluation, we only report the average of 51 different random initializations and then summarize the average of fuzziness m ranging from 1.1–2.0. It should be noted that our proposed U-MV-FCM aims at finding the optimal number of clusters with a high-quality cluster centroids in an efficient manner. Furthermore, our proposed U-MV-FCM is designed to automatically assign the viewpoint data as initial cluster centers. While in the existing algorithms, users need to specify a certain number as initialization for clusters. Therefore, in these experimental setups of Co-FKM, WV-Co-FCM, and MinMax-FCM, we assign the accurate clusters number on each multiview data as their initial number of clusters.

Findings and Discussion: The results regarding the performance of the proposed U-MV-FCM with the existing MV-FCM clustering algorithms on the nine multiview datasets have been demonstrated. Table 3 shows the clusters number obtained by the proposed U-MV-FCM as well as the information regarding the actual number of clusters. The optimal number of clusters c^* obtained by the proposed U-MV-FCM is displayed inside the parentheses. It can be observed that the proposed U-MV-FCM obtains an optimal number of clusters for Artificial Data 1 ($c^* = 2$), for Artificial Data 2 ($c^* = 5$), Syn500 ($c^* = 2$), and the real-world datasets Prokaryotic ($c^* = 4$), Wikipedia Articles ($c^* = 10$), WebKB ($c^* = 3$), Extended YaleB ($c^* = 10$), 3-Source ($c^* = 3$), Reuters ($c^* = 2$), respectively.

Table 3. Number of clusters obtained by U-MV-FCM.

Datasets	Cluster No.	Obtained Cluster No.
Artificial Data 1	2	**2**
Artificial Data 2	5	**5**
Syn500	2	**2**
Prokaryotic	4	**4**
Wikipedia Articles	10	**10**
WebKB	3	**3**
Extended YaleB	10	**10**
3-sources	6	**3**
Reuters	6	**2**

We established the results of the U-MV-FCM in terms of AV-AR, AV-FMI, AV-RI, AV-NMI, and AV-JI in Tables 4–10. It should be noted that the best result is marked in boldface, and the second one is marked in underlining. It can be witnessed that the proposed U-MV-FCM performs superior in the case of the evaluation measure as well as the comparison to other state-of-the-art algorithms on Artificial Data 1, Artificial Data 2, Syn500, Prokaryotic, Wikipedia Articles, and WebKB datasets. In the case of 3-Sources and Reuters datasets, the proposed U-MV-FCM failed to recognize the optimal number of clusters c^* as six. It is apparent that the MV-FCM unveils a higher AR, FMI, RI, and JI in most of the real MV datasets.

Table 4. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for Artificial Data 1.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 2)	**1**	**1**	**1**	**1**	**1**
Co-FKM (with c = 2)	0.785	0.730	0.798	0.383	0.676
MinMax-FCM (with c = 2)	0.700	0.579	0.761	0.000	0.579
U-MV-FCM (without c)	**1**	**1**	**1**	**1**	**1**

Table 5. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for Artificial Data 2.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 5)	0.579	0.744	0.659	0.669	0.452
Co-FKM (with c = 5)	0.531	0.702	0.568	0.505	0.376
MinMax-FCM (with c = 5)	0.274	0.216	0.465	0.000	0.216
U-MV-FCM (without c)	**0.998**	**0.998**	**0.996**	**0.993**	**0.992**

Table 6. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for syn500 Multiview Data.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 2)	0.759	0.638	0.640	0.223	0.474
Co-FKM (with c = 2)	0.546	0.545	0.610	0.102	0.432
MinMax-FCM (with c = 2)	0.500	0.499	0.706	0.000	0.499
U-MV-FCM (without c)	**0.874**	**0.779**	**0.786**	**0.560**	**0.647**

Table 7. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for Prokaryotic Multiview Data.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 4)	0.484	0.572	0.487	0.232	0.320
Co-FKM (with c = 4)	0.455	0.552	0.469	0.114	0.304
MinMax-FCM (with c = 4)	0.313	0.596	0.394	0.214	0.240
U-MV-FCM (without c)	**0.773**	**0.778**	**0.686**	**0.502**	**0.515**

Table 8. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for Wikipedia Articles Multiview Data.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 10)	0.124	0.715	0.172	0.056	0.087
Co-FKM (with c = 10)	0.507	0.848	0.439	0.493	0.278
MinMax-FCM (with c = 10)	0.144	0.664	0.188	0.049	0.090
U-MV-FCM (without c)	**0.592**	**0.887**	**0.478**	**0.550**	**0.314**

Table 9. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for WebKB Multiview Data.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 3)	0.583	0.644	0.539	0.234	0.355
Co-FKM (with c = 3)	0.577	0.693	0.574	0.350	0.403
MinMax-FCM (with c = 3)	0.527	0.393	0.627	0.000	0.393
U-MV-FCM (without c)	**0.699**	**0.727**	**0.666**	**0.403**	**0.498**

Table 10. Clustering Performances of WV-Co-FCM, MinMax-FCM, Co-FKM, and U-MV-FCM for Extended YaleB Multview Data.

Algorithms\Measures	**Evaluation Methods (Average)**				
	AV-AR	**AV-RI**	**AV-FMI**	**AV-NMI**	**AV-JI**
WV-Co-FCM (with c = 10)	0.100	0.099	**0.314**	0.000	**0.099**
Co-FKM (with c = 10)	**0.171**	**0.830**	0.155	**0.153**	0.084
MinMax-FCM (with c = 10)	0.100	0.099	**0.314**	0.000	**0.099**
U-MV-FCM (without c)	0.136	0.825	0.146	0.128	0.079

It can be clearly seen from Table 4 that the evaluation measures for the proposed U-MV-FCM performs better compared to the other algorithms except the WV-Co-FCM for Artificial Data 1.

Table 5 shows that our proposed U-MV-FCM performed the best compared to the other algorithms, and the WV-Co-FCM is ranked second on the base of the performance evaluation measures. It is important to note that, in our proposed U-MV-FCM, the optimal c is calculated automatically, while in the other algorithms the performance measures are calculated with a given number of clusters as initialization.

Table 6 shows that our proposed U-MV-FCM performed the best compared to the other algorithms. The WV-Co-FCM is ranked second on the base of the performance evaluation measures for AV-AR, AV-RI, AV-NMI, and the MinMax-FCM for AV-FMI and AV-JI, respectively. It is worth mentioning that our proposed U-MV-FCM calculates the optimal c in an automatic way, while in the other algorithms the performance evaluation measures are calculated with a given number of clusters.

From Table 7, it is evident that our proposed U-MV-FCM performed the best compared to the other algorithms. The WV-Co-FCM is ranked second in respect to the performance evaluation measures for AV-AR, AV-FMI, AV-NMI, AV-JI, and the MinMax-FCM for AV-RI, respectively. We mention that our proposed U-MV-FCM calculates the optimal c in an automatic way, but in the other algorithms, the performance evaluation measures are calculated with a given number of clusters.

It can be seen from Table 8 that our proposed U-MV-FCM performed the best compared to the other algorithms, and the Co-FKM is ranked second in the performance evaluation measures. It is worth saying that our proposed U-MV-FCM calculates the optimal c in an automatic way. But, in the other algorithms, the performance evaluation measures are calculated with a given number of clusters.

Table 9 shows that our proposed U-MV-FCM performed the best compared to the other algorithms, while the Co-FKM is ranked second in respect to the performance evaluation measures for AV-RI, AV-NMI, AV-JI, WV-Co-FCM for AV-AR and the MinMax-FCM for AV-FMI, respectively. It is worth mentioning that our proposed U-MV-FCM calculates the optimal c in an automatic way, while in the other algorithms the performance evaluation measures are calculated with a given number of clusters in initialization.

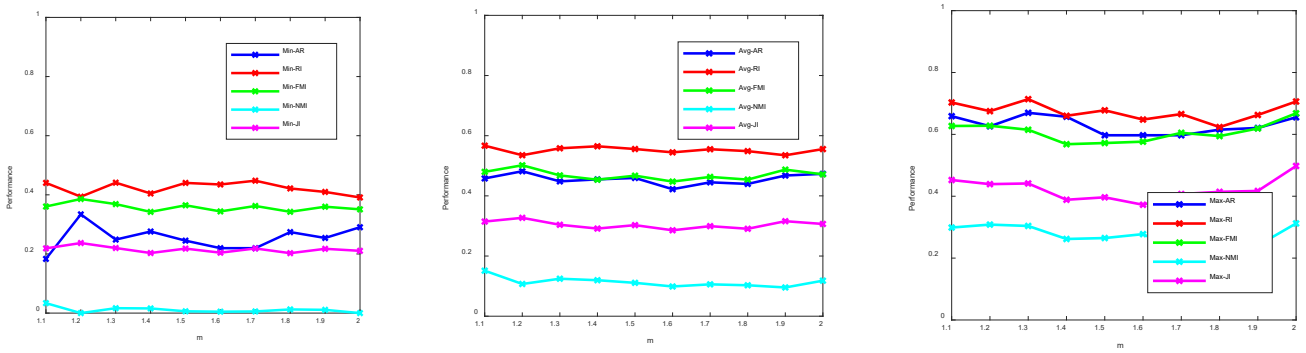
Co-FKM performs the best for the Extended YaleB dataset in terms of AV-AR, AV-RI, and AV-NMI, while for AV-FMI and AV-JI, the WV-Co-FCM and MinMax-FCM performs the best, and our proposed algorithm ranked second.

The running time (RTs) for all algorithms on the nine datasets are reported in Table 11. The RTs on Artificial Data 1, Artificial Data 2, Syn500, Prokaryotic, Wikipedia Articles, 3-Sources, WebKB, Reuters, and Extended YaleB for the MinMax-FCM are less compared to our proposed U-MV-FCM. However, it is noteworthy that the MinMax-FCM and the other compared algorithms depend on the fuzziness parameter m. Hence, if we calculate and report the total running time of the MinMax-FCM under 51 different random initializations for ten different m's, then the introduced U-MV-FCM is faster and more efficient.

Table 11. The Running Time (RT) obtained by the proposed U-MV-FCM, Co-FKM, MinMax-FCM, and WV-Co-FCM (The best is marked in boldface, and the second is marked in underline).

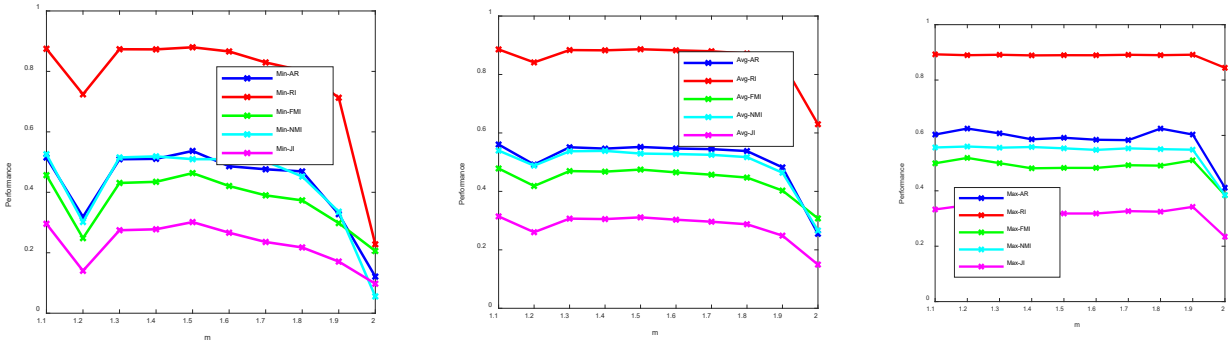
Algorithms\Measures	**The Performances of RT (In Seconds)**			
	Co-FKM	**MinMax-FCM**	**WV-Co-FCM**	**U-MV-FCM**
Artificial 1	7.141	**1.082**	10.73	<u>2.346</u>
Artificial 2	51.83	**1.434**	52.16	<u>2.361</u>
Syn500	3.633	**0.356**	4.811	<u>0.960</u>
Prokaryotic	37.55	**2.127**	23.57	<u>4.528</u>
Wikipedia Articles	90.72	**1.320**	4.990	<u>4.594</u>
3-Sources	563.2	**9.108**	18.54	<u>11.49</u>
WebKB	3.408	**1.314**	14.77	<u>2.476</u>
Reuters	1767.8	**75.75**	478.71	<u>668.1</u>
Extended YaleB	422.7	**14.99**	238.1	<u>83.08</u>

For visual comparison, the performance results for the Co-FKM, WV-Co-FCM, and MinMax-FCM algorithms in terms of minimum values, average values, and maximum values on Prokaryotic, Wikipedia Articles, WebKB, and Extended YaleB datasets are displayed in Figures 8–16.



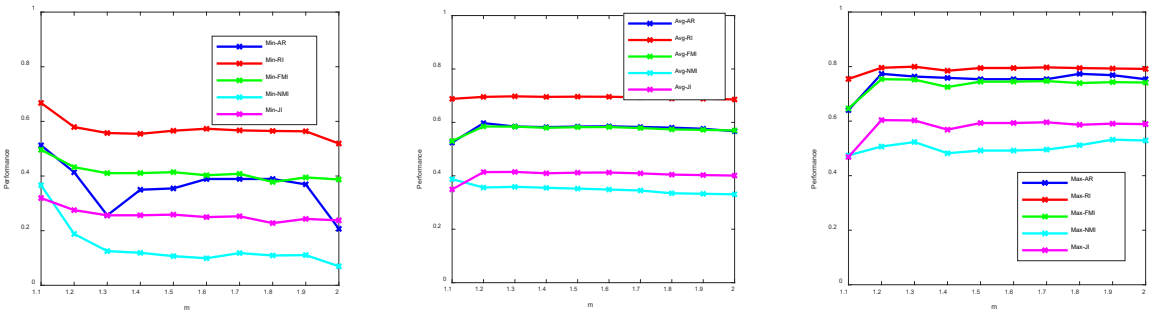
(**a**) (**b**) (**c**)

Figure 8. The performance results of Co-FKM for Prokaryotic data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



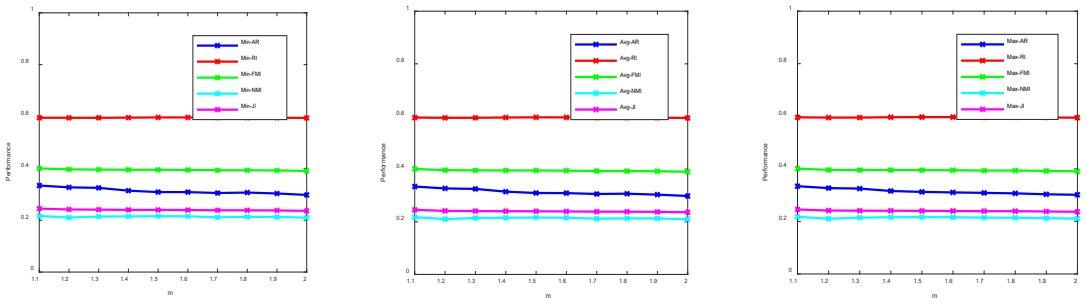
(**a**) (**b**) (**c**)

Figure 9. The performance results of Co-FKM for Wikipedia Articles data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



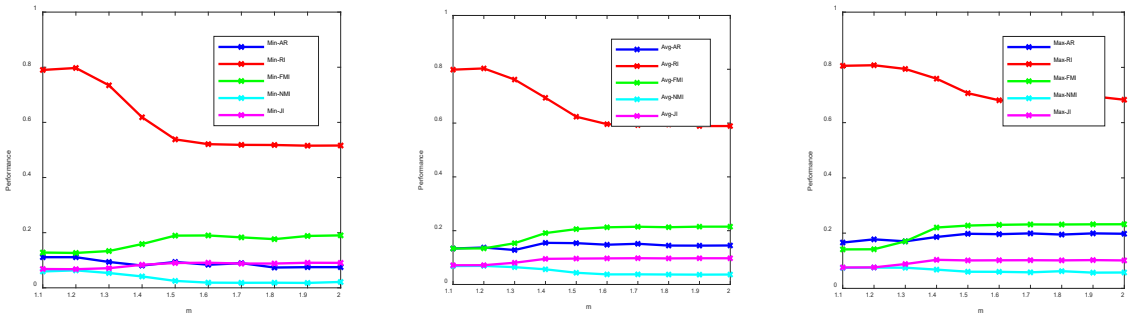
(**a**) (**b**) (**c**)

Figure 10. The performance results of Co-FKM for WebKB data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



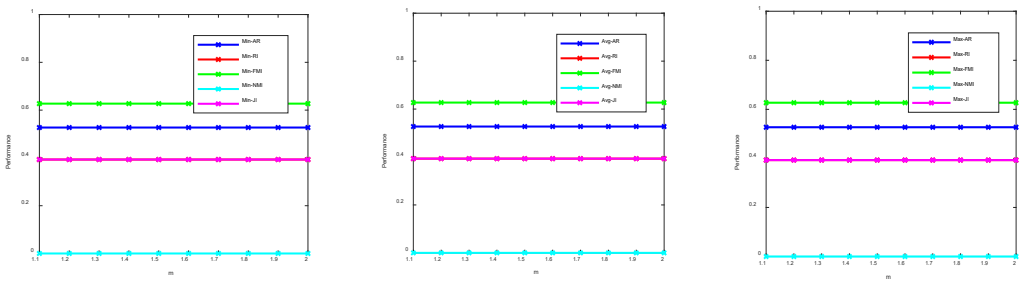
(**a**) (**b**) (**c**)

Figure 11. The performance results of MinMax-FCM for Prokaryotic data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



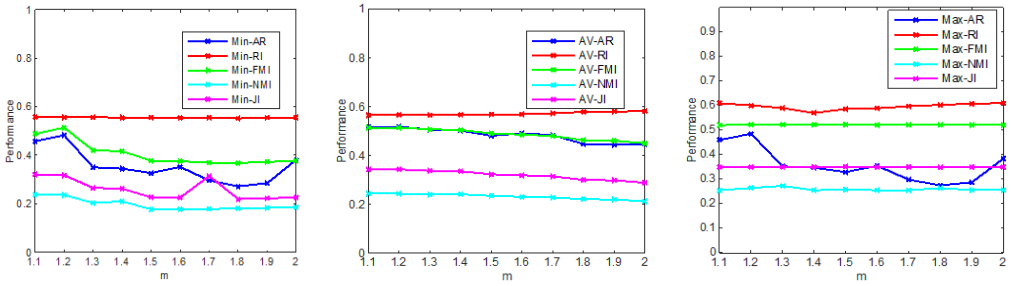
(**a**) (**b**) (**c**)

Figure 12. The performance results of MinMax-FCM for Wikipedia Articles data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



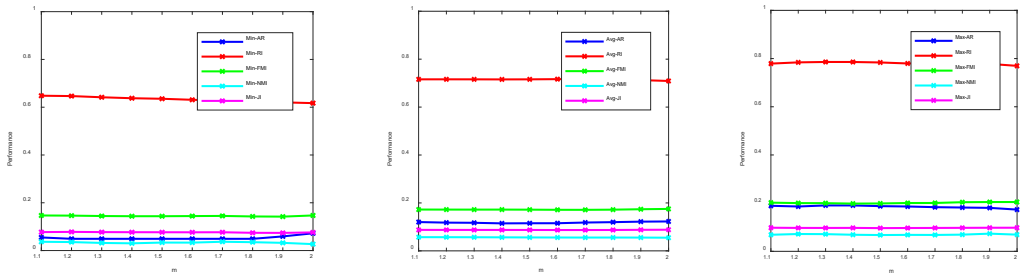
(**a**) (**b**) (**c**)

Figure 13. The performance results of MinMax-FCM for WebKB data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



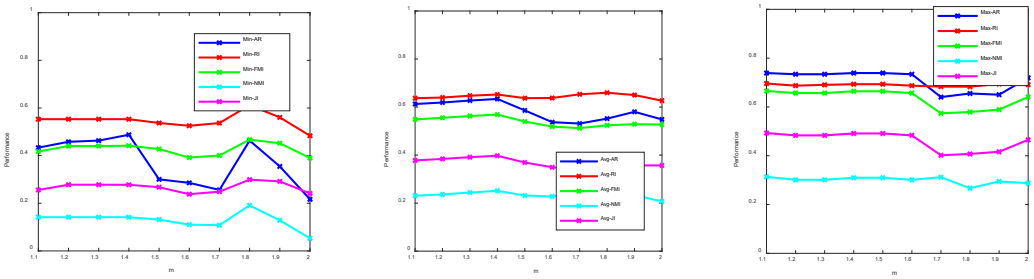
(**a**) (**b**) (**c**)

Figure 14. The performance results of WV-Co-FCM for Prokaryotic data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



(**a**) (**b**) (**c**)

Figure 15. The performance results of WV-Co-FCM for Wikipedia Articles data (**a**) minimum values; (**b**) average values; (**c**) maximum values.



(**a**) (**b**) (**c**)

Figure 16. The performance results of WV-Co-FCM for WebKB data (**a**) minimum values; (**b**) average values; (**c**) maximum values.

Figure 8 shows the minimum, average, and maximum values for the Co-FKM on Prokaryotic data; the Min-AR, Min-RI, Min-FMI, and Min-JI in Figure 8a are varied over the entire m, while the Min-NMI is stable. The Avg-AR, Avg-RI, Avg-FMI, Avg-NMI, and Avg-JI in Figure 8b show fluctuating patterns. Figure 8c shows a higher variability.

Figure 9 shows the performance results for the Co-FKM on the Wikipedia Articles dataset. The Min-AR, Min-RI, Min-FMI, Min-NMI, and Min-JI in Figure 9a show a downward pattern. The Avg-AR, Avg-RI, Avg-FMI, Avg-NMI, and Avg-JI in Figure 9b show a downward trend when m = 1.1, but stable when m = 1.3, and a decreasing movement when m = 1.9. The Max-AR, Max-FMI, and Max-JI in Figure 9c is stable when m =1.1, but a descending trend at m = 1.8. The Max-RI and Max-NMI have a downward trend when m = 1.9.

Figure 10 portrays the performance results for the Co-FKM on WebKB data. The Min-AR, Min-RI, and Min-NMI in Figure 10a show an irregular pattern, initially decreases and reaches stability when m = 1.9. The Min-FMI and Min-JI show a downward trend when m = 1.1, but

stable when m = 1.3. The Avg-AR, Avg-NMI, Avg-FMI, and Avg-JI in Figure 10b is stable when m = 1.2. The Avg-RI in Figure 10c reaches stability over the entire m.

Figures 11–13 display the performance results for the MinMax-FCM on Prokaryotic, Wikipedia Articles, and WebKB data, respectively, while Figures 14–16 show the performance for WV-Co-FCM on Prokaryotic, Wikipedia Articles and WebKB data, respectively.

4.2. The Effect of Parameters for MV-FCM Clustering Algorithms

It is commonly known that (single-view) FCM is basically affected by the fuzziness parameter m. The m values are generated to control the distribution of intra-clusters in the data. Since m values are varied, it leads to different structures. In this sense, m is vital to trial the data for producing a recommended result. At the same time, MV-FCM, as a high considerably clustering algorithm, considers more parameters to maintain good performance. Since the MV-FCM clustering algorithms are developed to handle multiview data, it is reasonable if the objective function became more complex. The MV-FCM involves additional variables such as view weights, feature-view weights, an agreement between memberships, etc. These additional variables are driven by some parameters to handle their distribution, and mostly are generated by users. Due to these complexities, it is highly recommended to measure these MV-FCM performances by employing different values of the parameters. The sequences of m with 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0 are implemented in this experiment. The parameter setting of Co-FKM, MinMax-FCM, and WV-Co-FCM is varied and control different tasks. Since these parameters are correlated to each other, parameter tunings are required to obtain a precious pattern. In this sense, an experiment to evaluate the consistency of m with fixed combination parameters on three MV-FCM algorithms is considerable. Table 12 reports the parameter setting for these three MV-FCM clustering algorithms of Co-FKM, MinMax-FCM, and WV-Co-FCM. We mention that our proposed U-MV-FCM clustering algorithm is free of parameters.

Table 12. Parameter setting for Co-FKM, MinMax-FCM, and WV-Co-FCM.

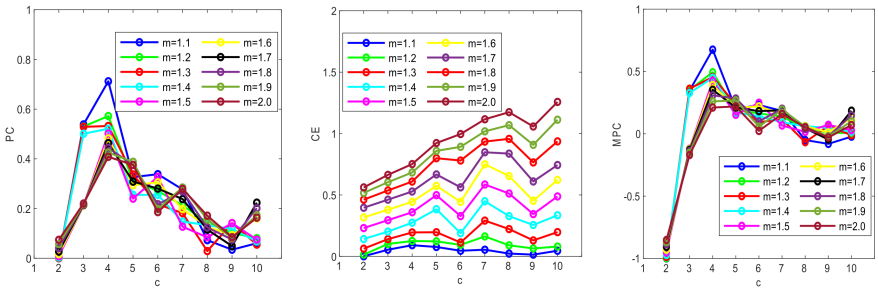
Algorithms	Datasets	m	β	η
Co- FKM	Prokaryotic	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0	3	-
	Wikipedia Articles			
	WebKB			
	Extended YaleB			
MinMax- FCM	Prokaryotic	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0	$\frac{s-1}{s}$	-
	Wikipedia Articles			
	WebKB			
	Extended YaleB			
WV-Co-FCM	Prokaryotic	1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0	4	$\frac{s-1}{s}$
	Wikipedia Articles			
	WebKB			
	Extended YaleB			

4.3. Ensembling Membership-Based Indices to MV-FCM Clustering Algorithms

For further experiments, we consider ensembling MV-FCM clustering algorithms with some membership-based internal indices. These MV-FCM clustering algorithms of Co-FKM, MinMax-FCM, and WV-Co-FCM will be employed with these membership-based CVIs of Partition Coefficient (PC) [32], Partition Entropy (PE) [33], and Modified Partition Coefficient (MPC) [36]. For MV-FCM cases, the cluster centers are initially generated and updated separately for each view. In this sense, different clusters number will lead to different output. Since cluster-center-based CVIs were designed only for (single view) FCM

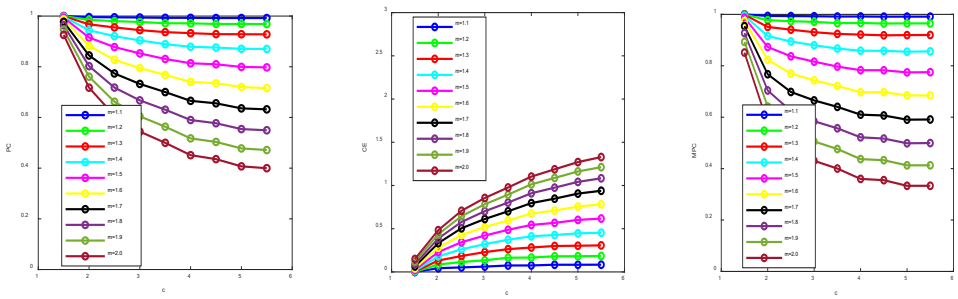
clustering algorithm, it is not recommended to perform them for the MV-FCM clustering algorithms. However, membership-based CVIs are applicable to the MV-FCM clustering algorithms. The MV-FCM clustering algorithms usually generate their memberships in two ways. One is generated by treating the importance of each data-view membership equally. Another one is to generate the membership for each view (local-membership), and then to estimate a global solution by collaborating local-memberships based on their agreement/disagreement degree. In this sense, membership-based CVIs are well considered for this experiment. These membership-based CVIs are PC, PE, and MPC. In order to verify the effectiveness of Co-FKM, MinMax-FCM, and WV-Co-FCM algorithms with the CVIs of PC, PE, and MPC for finding an optimal number of clusters, we ensemble the MV-FCM clustering algorithms with PC, PE, and MPC, which are termed as Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC, respectively.

Findings and Discussion: These Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC are implemented on three artificial and five real multiview datasets, called Artificial 1, Artificial 2, Syn500, Prokaryotic, Wikipedia Articles, 3-Source, WebKB, and Reuters, respectively. Figures 17–19 visualize the distribution of these Co-FKM+PC, Co-FKM+PE, and Co-FKM+MPC on the three artificial multiview datasets, with m ranging from 1.1 to 2.0. As it can be seen, the weights for Co-FKM+PC, Co-FKM+PE, and Co-FKM+MPC follow a similar distribution. In this sense, each m mostly produces the same optimal number of clusters. The same behavior is also presented for CVIs- ensembled MinMax-FCM and WV-Co-FCM. As shown in Figures 20–23, in cases of Artificial 2 and Syn500 MV datasets, the pattern distributions of these MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC give the same optimal number of clusters. As noted, Figures 17–23 are generated based on a random initialization.



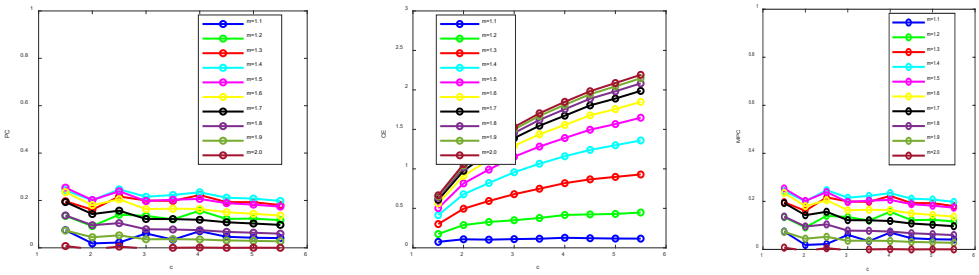
(**a**) (**b**) (**c**)

Figure 17. Cluster validity indices results on Artificial 1 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) Co-FKM+PC; (**b**) Co-FKM+PE; (**c**) Co-FKM+MPC.



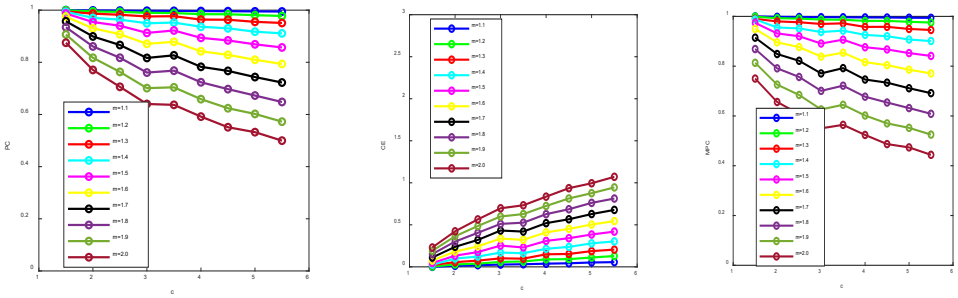
(**a**) (**b**) (**c**)

Figure 18. Cluster validity indices results on Artificial 2 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) Co-FKM+PC; (**b**) Co-FKM+PE; (**c**) Co-FKM+MPC.



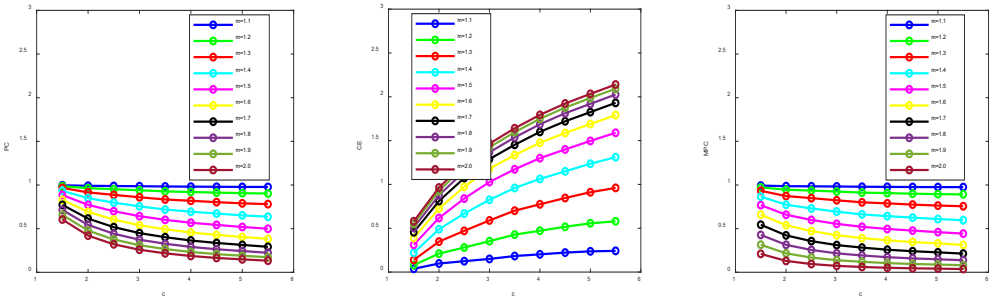
(**a**) (**b**) (**c**)

Figure 19. Cluster validity indices results on Syn500 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) Co-FKM+PC; (**b**) Co-FKM+PE; (**c**) Co-FKM+MPC.



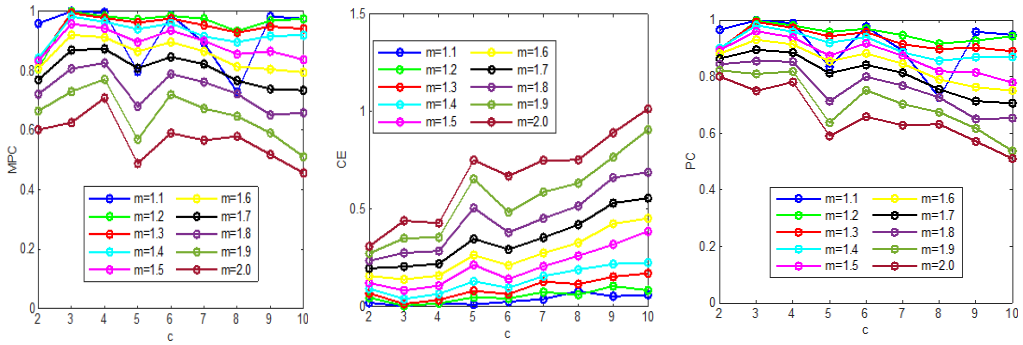
(**a**) (**b**) (**c**)

Figure 20. Cluster validity indices results on Artificial 2 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) MinMax-FCM+PC; (**b**) MinMax-FCM+PE; (**c**) MinMax-FCM+MPC.



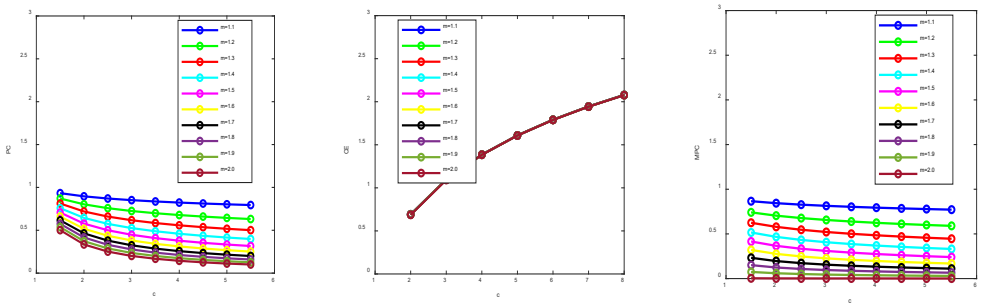
(**a**) (**b**) (**c**)

Figure 21. Cluster validity indices results on Syn500 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) MinMax-FCM+PC; (**b**) MinMax-FCM+PE; (**c**) MinMax-FCM+MPC.



(**a**) (**b**) (**c**)

Figure 22. Cluster validity indices results on Artificial 2 data with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) WV-Co-FCM+PC; (**b**) WV-Co-FCM+PE; (**c**) WV-Co-FCM+MPC.



(**a**) (**b**) (**c**)

Figure 23. Cluster validity indices results on Syn500 with m ranging from 1.1 to 2.0 and $c_{\min} = 2$ to $c_{\max} = 10$ for (**a**) WV-Co- (**a**) WV-Co- FCM+PC; (**b**) WV-Co-FCM+PE; (**c**) WV-Co-FCM+MPC.

We generate 51 different random initializations for Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC with m = 2 on the eight multiview datasets. The results based on 51 different random initializations on the eight datasets are reported in Table 13. In terms of the three artificial and five real multiview datasets, Table 13 illustrates that WV-Co-FCM+PC, MinMax-FCM+PE, and MinMax-FCM+MPC are the best investigators, while the second investigators are Co-FKM+PC, Co-FKM+PE, and WV-Co-FCM+MPC. Furthermore, we report the running time (RTs) for these CVIs- ensembled Co-FKM, MinMax-FCM, and WV-Co-FCM on the eight multiview datasets in Table 14. It can be seen that MinMax-FCM+PC, MinMax-FCM+PE, and MinMax+MPC perform faster as compared to the others.

Table 13. Optimal cluster number obtained by Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC with m = 2 and t_max=10 on eight MV datasets (Percentages of correct c in parantheses).

Algorithms	Artificial 1	Artificial 2	Syn500	Prokaryotic	Wikipedia Articles	3-Sources	Reuters	WebKB
Co-FKM+PC	**2** (13.7%)	**5 (15.7%)**	**2 (100%)**	**4** (0%)	**10** (11.7%)	**6** (1.9%)	**6** (3.9%)	**3** (0%)
MinMax-FCM+PC	**2 (100%)**	2 (0%)	**2 (100%)**	**4** (0%)	**10** (0%)	**6** (0%)	**6** (0%)	**3** (0%)
WV-Co-FCM+PC	**2** (98.0%)	2 (0%)	**2 (100%)**	**4** (7.84%)	**10 (13.7%)**	**6** (0%)	**6** (0%)	**3** (0%)
Co-FKM+PE	**2** (66.7%)	**5** (3.9%)	**2 (100%)**	**4** (0%)	**10** (0%)	**6** (0%)	**6** (0%)	**3** (0%)
MinMax-FCM+PE	**2 (100%)**	**5** (0%)	**2 (100%)**	**4** (0%)	**10** (0%)	**6** (0%)	**6** (0%)	**3** (0%)
WV-Co-FCM+PE	**2** (98.0%)	**5** (0%)	**2 (100%)**	**4** (0%)	**10** (0%)	**6** (0%)	**6** (0%)	**3** (0%)
Co-FKM+MPC	**2** (7.8%)	**5** (13.7%)	**2** (54.9%)	**4 (29.4%)**	**10** (11.8%)	**6** (3.9%)	**6 (11.7%)**	**3** (17.6%)
MinMax-FCM+ MPC	**2 (100%)**	**5** (3.9%)	**2** (72.6%)	**4**(13.7%)	**10** (2.0%)	**6 (9.8%)**	**6** (7.8%)	**3** (23.5%)
WV-Co-FCM+MPC	**2** (98.0%)	**5** (0%)	**2** (66.7%)	**4** (7.8%)	**10 (13.7%)**	**6** (0%)	**6** (0%)	**3 (25.4%)**

Table 14. Running time for Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC with m = 2 and t_max = 10 on eight MV datasets.

Algorithms	Artificial 1	Artificial 2	Syn500	Prokaryotic	Wikipedia Articles	3-Sources	Reuters	WebKB
Co-FKM+PC	42.53	72.89	34.9	113.7	63.5	751.6	251.1	88.3
MinMax-FCM+PC	**0.98**	**1.06**	0.97	**2.8**	**1.5**	**29.3**	**78.3**	**2.5**
WV-Co-FCM+PC	37.86	51.43	16.2	85.3	68.1	277.6	704.7	26.4
Co-FKM+PE	37.26	70.89	31.8	113.4	55.1	582.2	202.1	84.5
MinMax-FCM+PE	1.13	1.15	**0.88**	3.9	1.8	33.9	67.16	3.2
WV-Co-FCM+PE	32.86	53.36	17.1	82.4	37.8	285.7	749.6	26.7
Co-FKM+MPC	42.53	72.89	34.9	113.7	63.5	751.6	251.1	88.3
MinMax-FCM+MPC	**0.98**	**1.06**		**2.8**	**1.5**	**29.3**	**78.3**	**2.5**
WV-Co-FCM+MPC	37.86	51.43	16.2	85.3	68.1	277.6	704.7	26.4

5. Conclusions

In this paper, we proposed a novel multiview fuzzy c-means (MV-FCM) clustering algorithm to find the optimal number of clusters with high-quality cluster centroids in an efficient manner, called U-MV-FCM. The proposed U-MV-FCM can automatically produce an optimal number of clusters and simultaneously improve the accuracy rate without a parameter setting. The view-points on the multiview data are assigned as the initial cluster centers, and the proposed U-MV-FCM can directly reduce the number of clusters and automatically produce an optimal number of clusters. To assess the performance of the proposed U-MV-FCM algorithm, a comparative analysis with the existing MV-FCM algorithms, such as the Co-FKM, MinMax-FCM, and WV-Co-FCM, was made. Furthermore, experimental results based on the CVIs-ensembled MV-FCM clustering algorithms are generated, namely Co-FKM+PC, Co-FKM+PE, Co-FKM+MPC, MinMax-FCM+PC, MinMax-FCM+PE, MinMax-FCM+MPC, WV-Co-FCM+PC, WV-Co-FCM+PE, and WV-Co-FCM+MPC. The effectiveness of experimental results is conducted by using three artificial and six benchmark multiview datasets, including Artificial Data 1, Artificial Data 2, Syn500, Prokaryotic, Wikipedia Articles, 3-Sources, WebKB, Reuters and Extended YaleB. For the Artificial 1, Artificial 2, Syn500, Prokaryotic, Wikipedia Articles, WebKb datasets, our proposed method outperforms these existing algorithms on the base of five validity indices, namely AV-AR, AV-FMI, AV-RI, AV-NMI and AV-JI, except Extended YaleB dataset in which our proposed method shows the second better performance.

Based on the experimental results, it is evident that the U-MV-FCM outperformed the existing algorithms on most datasets, as indicated by the higher accuracy rates and more accurate cluster identification with automatically finding an optimal number of clusters.

Thus, it can also be concluded that the learning approach for multiview data in the U-MV-FCM is suitable to obtain more a accurate number of clusters and has better clustering accuracy than others, such as these CVIs-ensembled MV-FCM clustering algorithms. In addition, the U-MV-FCM is more stable from the aspect of the variance of multiple results as our parameter settings are more robust and so it improves the performance of fuzzy clustering algorithms in multiview scenarios. As in the U-MV-FCM, we use the idea of global membership; thus, to study further we suggest the scheme of collaboration in membership, and we can extend it to feature extraction for the U-MV-FCM. On the other hand, we mention that if the desired selected clusters are not applicable to MV data, then the proposed U-MV-FCM failed to evaluate the accurate number of clusters c^*. This will be our further research to improve the proposed U-MV-FCM clustering algorithm. The work presented in this paper, introducing the U-MV-FCM clustering algorithm, holds significant promise and potential impact in the field of multiview clustering. The conclusions drawn from this study highlight several key aspects that underline the significance of this work. The U-MV-FCM algorithm outperforms existing multiview clustering algorithms in terms of accuracy and the identification of an optimal number of clusters across various datasets. On the other hand, our research emphasizes the stability of the U-MV-FCM in which the algorithm is less prone to variations in results, due to the robust parameter settings. The research has real-world implications, as the improved multiview clustering offered by the U-MV-FCM can benefit various domains, such as data analysis, information retrieval, bioinformatics, social network analysis, and more. It can lead to more accurate insights and decision-making in these application areas. In general, the U-MV-FCM may fail to evaluate the accurate number of clusters in case the selected clusters are not applicable to multiview data, and so future research may explore methods to address this limitation. This might involve the development of adaptive clustering techniques that can adjust the number of clusters. On the other hand, the U-MV-FCM does not consider feature weights. Thus, an advanced U-MV-FCM with feature weights and feature reduction will be our further research topic.

Author Contributions: Conceptualization, I.H. and M.-S.Y.; methodology, I.H., K.P.S. and M.-S.Y.; software, I.H. and K.P.S.; validation, I.H. and K.P.S.; formal analysis, I.H., K.P.S. and M.-S.Y.; investigation, I.H., K.P.S. and M.-S.Y.; data curation, I.H. and K.P.S.; writing—original draft preparation, I.H.; writing—review and editing, M.-S.Y.; visualization, I.H., K.P.S. and M.-S.Y.; supervision, M.-S.Y.; funding acquisition, M.-S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Ministry of Science and technology (MOST) of Taiwan under Grant MOST 111-2118-M-033-001.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jain, A.K.; Dubes, R.C. *Algorithms for Clustering Data*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1988.
2. Chang-Chien, S.J.; Hung, W.L.; Yang, M.S. On mean shift-based clustering for circular data. *Soft Comput.* **2012**, *16*, 1043–1060. [CrossRef]
3. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
4. Banfield, J.D.; Raftery, A.E. Model-based Gaussian and non-Gaussian clustering. *Biometrics* **1993**, *1*, 803–821. [CrossRef]
5. Zhong, S.; Ghosh, J. A unified framework for model-based clustering. *J. Mach. Learn. Res.* **2003**, *4*, 1001–1037.
6. Yu, J.; Chaomurilige, C.; Yang, M.S. On convergence and parameter selection of the EM and DA-EM algorithms for Gaussian mixtures. *Pattern Recognit.* **2018**, *77*, 188–203. [CrossRef]
7. Chamroukhi, F.; Nguyen, H.D. Model-based clustering and classification of functional data. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1298. [CrossRef]
8. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Davis, CA, USA, 21 June–18 July 1967; Volume 1, No. 14. pp. 281–297.

9. Inbarani, H.H.; Azar, A.T. Leukemia image segmentation using a hybrid histogram-based soft covering rough k-means clustering algorithm. *Electronics* **2020**, *9*, 188. [CrossRef]
10. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
11. Ruspini, E.H. A new approach to clustering. *Inf. Control* **1969**, *15*, 22–32. [CrossRef]
12. Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybern.* **1975**, *3*, 32–57. [CrossRef]
13. Chaomurilige, C.; Yu, J.; Yang, M.S. Analysis of parameter selection for Gustafson–Kessel fuzzy clustering using Jacobian matrix. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 2329–2342. [CrossRef]
14. Chaomurilige, C.; Yu, J.; Yang, M.S. Deterministic annealing Gustafson-Kessel fuzzy clustering algorithm. *Inf. Sci.* **2017**, *417*, 435–453. [CrossRef]
15. Cardone, B.; Di Martino, F. A novel fuzzy entropy-based method to improve the performance of the fuzzy C-means algorithm. *Electronics* **2020**, *9*, 554. [CrossRef]
16. Kumar, D.; Agrawal, R.K.; Kumar, P. Bias-corrected intuitionistic fuzzy c-means with spatial neighborhood information approach for human brain MRI image segmentation. *IEEE Trans. Fuzzy Syst.* **2020**, *30*, 687–700. [CrossRef]
17. Wang, E.; Lee, H.; Do, K.; Lee, M.; Chung, S. Recommendation of Music Based on DASS-21 (Depression, Anxiety, Stress Scales) Using Fuzzy Clustering. *Electronics* **2022**, *12*, 168. [CrossRef]
18. Dhillon, I.S. Co-clustering documents and words using bipartite spectral graph partitioning. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 26–29 August 2001; pp. 269–274.
19. Bickel, S.; Scheffer, T. Multi-view clustering. In Proceedings of the 4th IEEE International Conference on Data Mining ICDM, Brighton, UK, 1–4 November 2004; Volume 4, pp. 19–26.
20. Cleuziou, G.; Exbrayat, M.; Martin, L.; Sublemontier, J.H. CoFKM: A centralized method for multiple-view clustering. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 752–757.
21. Jiang, Y.; Chung, F.L.; Wang, S.; Deng, Z.; Wang, J.; Qian, P. Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans. Cybern.* **2014**, *45*, 688–701. [CrossRef]
22. Wang, Y.; Chen, L. Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources. *Expert Syst. Appl.* **2017**, *72*, 457–466. [CrossRef]
23. Zeng, S.; Wang, X.; Cui, H.; Zheng, C.; Feng, D. A unified collaborative multikernel fuzzy clustering for multiview data. *IEEE Trans. Fuzzy Syst.* **2017**, *26*, 1671–1687. [CrossRef]
24. Benjamin, J.B.; Yang, M.S. Weighted multiview possibilistic c-means clustering with L2 regularization. *IEEE Trans. Fuzzy Syst.* **2022**, *30*, 1357–1370. [CrossRef]
25. Huang, S.; Kang, Z.; Xu, Z. Auto-weighted multi-view clustering via deep matrix decomposition. *Pattern Recognit.* **2020**, *97*, 107015. [CrossRef]
26. Chen, Y.; Wang, S.; Zheng, F.; Cen, Y. Graph-regularized least squares regression for multi-view subspace clustering. *Knowl.-Based Syst.* **2020**, *194*, 105482. [CrossRef]
27. Tan, J.; Shi, Y.; Yang, Z.; Wen, C.; Lin, L. Unsupervised multi-view clustering by squeezing hybrid knowledge from cross view and each view. *IEEE Trans. Multimed.* **2020**, *23*, 2943–2956. [CrossRef]
28. Yang, M.S.; Sinaga, K.P. Collaborative feature-weighted multi-view fuzzy c-means clustering. *Pattern Recognit.* **2021**, *119*, 108064. [CrossRef]
29. Yang, M.S.; Hussain, I. Unsupervised multi-view K-means clustering algorithm. *IEEE Access* **2023**, *11*, 13574–13593. [CrossRef]
30. Papakostas, C.; Troussas, C.; Krouska, A.; Sgouropoulou, C. Personalization of the learning path within an augmented reality spatial ability training application based on fuzzy weights. *Sensors* **2022**, *22*, 7059. [CrossRef] [PubMed]
31. Papakostas, C.; Troussas, C.; Krouska, A.; Sgouropoulou, C. PARSAT: Fuzzy logic for adaptive spatial ability training in an augmented reality system. *Comput. Sci. Inf. Syst.* **2023**, *20*, 1389–1417. [CrossRef]
32. Lengyel, A.; Botta-Dukát, Z. Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecol. Evol.* **2019**, *9*, 13231–13243. [CrossRef] [PubMed]
33. Yang, S.C.H.; Lengyel, M.; Wolpert, D.M. Active sensing in the categorization of visual patterns. *eLife* **2016**, *5*, e12215. [CrossRef] [PubMed]
34. Xu, S.; Qiao, X.; Zhu, L.; Zhang, Y.; Xue, C.; Li, L. Reviews on determining the number of clusters. *Appl. Math. Inf. Sci.* **2016**, *10*, 1493–1512. [CrossRef]
35. Yang, M.S.; Nataliani, Y. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. *Pattern Recognit.* **2017**, *71*, 45–59. [CrossRef]
36. Pedrycz, W. Collaborative fuzzy clustering. *Pattern Recognit. Lett.* **2002**, *23*, 1675–1686. [CrossRef]
37. Zhu, l.; Chung, F.L.; Wang, S. Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2009**, *39*, 578–591.
38. Bezdek, J.C. Numerical taxonomy with fuzzy sets. *J. Math. Biol.* **1974**, *1*, 57–71. [CrossRef]
39. Bezdek, J.C. Cluster validity with fuzzy sets. *J. Cybern.* **1973**, *3*, 58–73. [CrossRef]
40. Gath, I.; Geva, A.B. Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 773–780. [CrossRef]

41. Xie, X.L.; Beni, G.A. validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 841–847. [CrossRef]
42. Wu, K.L.; Yang, M.S.; Hsieh, J.N. Robust cluster validity indexes. *Pattern Recognit.* **2009**, *42*, 2541–2550. [CrossRef]
43. Roubens, M. Pattern classification problems and fuzzy sets. *Fuzzy Sets Syst.* **1978**, *1*, 239–253. [CrossRef]
44. Pereira, J.C.; Coviello, E.; Doyle, G.; Rasiwasia, N.; Lanckriet, G.R.; Levy, R.; Vasconcelos, N. On the role of correlation and abstraction in cross-modal multimedia retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 521–535. [CrossRef]
45. Wang, H.; Yang, Y.; Li, T. Multi-view clustering via concept factorization with local manifold regularization. In Proceedings of the 2016 IEEE 16th international conference on data mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 1245–1250.
46. Brbić, M.; Piškorec, M.; Vidulin, V.; Kriško, A.; Šmuc, T.; Supek, F. The landscape of microbial phenotypic traits and associated genes. *Nucleic Acids Res.* **2016**, *44*, gkw964. [CrossRef]
47. Lu, Q.; Getoor, L. Link-based Classification. In Proceedings of the 20th International Conference on Machine Learning (ICML), Washington, DC, USA, 21–24 August 2003; pp. 496–503.
48. Wang, H.; Yang, Y.; Liu, B. GMC: Graph-based multi-view clustering. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1116–1129. [CrossRef]
49. Greene, D.; Cunningham, P. A matrix factorization approach for integrating multiple data views. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Bled, Slovenia, 6–10 September 2009; Springer: Berlin/Heidelberg, Germany; pp. 423–438.
50. Lewis, D.D.; Yang, Y.; Russell-Rose, T.; Li, F. Rcv1: A new benchmark collection for text categorization research. *J. Mach. Learn. Res.* **2004**, *5*, 361–397.
51. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]
52. Rong, W.; Zhuo, E.; Peng, H.; Chen, J.; Wang, H.; Han, C.; Cai, H. Learning a consensus affinity matrix for multi-view clustering via subspaces merging on Grassmann manifold. *Inf. Sci.* **2021**, *547*, 68–87. [CrossRef]
53. Rand, W.M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **1971**, *66*, 846–850. [CrossRef]
54. Fowlkes, E.B.; Mallows, C.L. A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* **1983**, *78*, 553–569. [CrossRef]
55. Cover, T.M. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
56. Jaccard, P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaudoise Sci. Nat.* **1901**, *37*, 241–272.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.