

Article

Weighted Multiview K-Means Clustering with L2 Regularization

Ishtiaq Hussain ^{1,*}, Yessica Nataliani ², Mehboob Ali ³, Atif Hussain ⁴, Hana M. Mujlid ⁵, Faris A. Almaliki ⁵
and Nouf M. Rahimi ⁶¹ Department of Applied Mathematics, Chung Yuan Christian University, Chung-Li, Taoyuan 32023, Taiwan² Department of Information Systems, Satya Wacana Christian University, Salatiga 50711, Indonesia; yessica.nataliani@uksw.edu³ Department of Statistics, Government College Gilgit, Gilgit Baltistan 15100, Pakistan; mehboob.ali@gcgilgit.edu.pk⁴ School of Artificial Intelligence, Xidian University, Xi'an 710126, China; 24176215842@stu.xidian.edu.cn⁵ Department of Computer Engineering, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; hmujlid@tu.edu.sa (H.M.M.); m.faris@tu.edu.sa (F.A.A.)⁶ Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23218, Saudi Arabia; nmraheeme@uj.edu.sa

* Correspondence: ishtiaqhucane@gmail.com

Abstract: In the era of big data, cloud, internet of things, virtual communities, and interconnected networks, the prominence of multiview data is undeniable. This type of data encapsulates diverse feature components across varying perspectives, each offering unique insights into the same underlying samples. Despite being sourced from diverse settings and domains, these data serve the common purpose of describing the same samples, establishing a significant interrelation among them. Thus, there arises a necessity for the development of multiview clustering methodologies capable of leveraging the wealth of information available across multiple views. This study introduces two novel weighted multiview k-means algorithms, W-MV-KM and weighted multiview k-means using L2 regularization, W-MV-KM-L2, designed specifically for clustering multiview data. These algorithms incorporate feature weights and view weights within the k-means (KM) framework. Our approach emphasizes a weighted multiview learning strategy, which assigns varying degrees of importance to individual views. We evaluate the clustering performance of our algorithms on seven diverse benchmark datasets spanning dermatology, textual, image, and digit domains. Through extensive experimentation and comparisons with existing methods, we showcase the superior effectiveness and utility of our newly introduced W-MV-KM-L2 algorithm.

Keywords: k-means; clustering; multiview; weighted multiview k-means (W-MV-KM)

Citation: Hussain, I.; Nataliani, Y.; Ali, M.; Hussain, A.; Mujlid, H.M.; Almaliki, F.A.; Rahimi, N.M.

Weighted Multiview K-Means Clustering with L2 Regularization.

Symmetry **2024**, *16*, 1646. <https://doi.org/10.3390/sym16121646>

Academic Editor: Hongkun Xu

Received: 28 October 2024

Revised: 5 December 2024

Accepted: 9 December 2024

Published: 12 December 2024



Copyright: © 2024 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 basic methods to uncover the underline structure within data. Clustering divides a dataset into clusters, where data points contained within each cluster display strong similarity, while those belonging to separate clusters demonstrate significant dissimilarity [1,2]. In the era of the internet of things (IoT) and big data, a substantial volume of data is generated across various areas of interest. The collected data originate from various sources or multiple views with diverse representations, including documents in different languages, mobile internet usage, social networks, multimedia data analysis, information and image retrieval, healthcare applications, and scientific research. In multivariate analysis, clustering methods can generally be categorized into two main approaches: probability model-based clustering and nonparametric clustering. This study focuses on the latter, emphasizing nonparametric clustering techniques. Several widely used nonparametric clustering methods are used in the literature, including the k-means [3,4], fuzzy c-means [5–7], and possibilistic c-means [8,9]. However, while these clustering techniques may work well for low-dimensional data, they are not well suited

for high-dimensional datasets. Multiview learning is gaining ground as a promising path within the field of pattern recognition and machine learning. Unlike conventional single-view clustering, multiview clustering leverages valuable feature knowledge from diverse perspectives to enhance clustering performance; multiview data encompass more shared, complementary, and redundant information. Numerous methodologies for multiview clustering methods have been proposed in order to increase the efficacy.

In the multiview clustering literature, Bickel and Scheffer [10] were the first to introduce multiview clustering as a method for handling multiview data. Since then, numerous researchers in the multiview clustering literature have contributed to advancements across various domains, such as in the paradigm of unsupervised multiview clustering (Yang and Hussain [11] and Hussain et al. [12]). In the field of multiview graph learning, various researchers have made significant contributions [13–17]. For incomplete multiview clustering, algorithms were proposed by [18–22]. In the area of tensor-based multiview clustering, notable algorithms have been proposed by Liu and Song [23], Li et al. [24], Chen et al. [25] and Benjamin and Yang [26]. Regarding multiview feature learning, several schemes have been developed. For example, Xu et al. [27] proposed feature learning for contrastive clustering, Zhang et al. [28] introduced sparse feature selection for images, Tang et al. [29] introduced consensus learning for unsupervised feature selection, and Xu et al. [30] presented a weighted multiview scheme for feature selection. In weighted multiview clustering, several researchers have made valuable contributions. Jiang et al. [31] introduced collaborative fuzzy clustering, Wang and Chen [32] proposed minimax optimization for fuzzy clustering, Yang and Sinaga [33] developed a feature reduction algorithm for k-means, and Yang and Sinaga [34] also proposed a collaborative scheme for FCM clustering. Some recent weighted multiview clustering techniques have been proposed. Benjamin and Yang [35] presented an algorithm for PCM using L2 regularization, Khan et al. [36] introduced a weighted concept-based scheme for incomplete multiview clustering, Zhou et al. [37] proposed evidential c-means clustering in a multiview scenario, Liu et al. [38] introduced an adaptive scheme with feature preference, Ouadfel and Abd Elaziz [39] proposed a multi-objective gradient optimizer approach-based scheme, Houfar et al. [40] introduced a binary weighted scheme in a multiview scenario, and Liu et al. [41] proposed a weighted scheme in evidential clustering in a multiview paradigm.

The rest of this study is structured as follows: Section 2 introduces a literature review. In Section 3, we introduce the weighted multiview k-means algorithms designed for clustering multiview datasets. Experimental results and comparisons with existing methods are discussed in Section 4. Discussions are outlined in Section 5. Lastly, Section 6 concludes this study and offers some future recommendations, particularly suggesting the use of a point symmetry-based distance instead of Euclidean distance to better capture cluster symmetry behaviors [42,43].

2. Literature Review

In this section, we review some multiview clustering algorithms that we compared with our proposed algorithms. We use two hard clustering algorithms built on multiview k-means, and two soft clustering algorithms based on multiview fuzzy c-means are discussed.

Xu et al. [30] introduced a multiview clustering algorithm based on k-means, termed the multiview weighted algorithm with feature selection (WMCFS). This algorithm incorporates two approaches for data points and feature selection. The objective function of WMCFS, as described by Xu et al. [30], is as follows:

$$J_{WMCFS}(U, A, W) = \sum_{h=1}^s (v_h)^\alpha \sum_{k=1}^c \sum_{i=1}^n z_{ik} \sum_{j=1}^{d_h} \left(w_j^h (x_{ij}^h - a_{kj}^h) \right)^2 + \tau \sum_{h=1}^s \sum_{j=1}^{d_h} \left(w_j^h \right)^2 \quad (1)$$

The parameter α is utilized to regulate the sparsity of view weights, while τ is employed to regulate the sparsity of feature weights. In WMCFS, the balancing parameter τ is utilized to control the feature weights within each view.

Yang and Sinaga [33] introduced the feature reduction multiview k-means (FRMVKM) clustering algorithm. The objective function of FRMVKM [33] is defined as:

$$J_{FRMVKM}(U, A, V, W) = \sum_{h=1}^s (v_h)^\alpha \sum_{k=1}^c \sum_{i=1}^n z_{ik} \sum_{j=1}^{d_h} w_j^h \delta_j^h (x_{ij}^h - a_{kj}^h)^2 + \frac{n}{d_h} \sum_{h=1}^s \sum_{j=1}^{d_h} w_j^h \ln w_j^h \delta_j^h \quad (2)$$

In this formulation, δ_j^h represents the parameter used to control the feature weights within the k -th cluster of each view, while $\alpha > 0$ denotes the view weight.

Jiang et al. [31] presented another extension to the multiview clustering algorithm, termed WV-Co-FCM. In this approach, they accounted for varying weights for each view, incorporating parameters to control the distribution of these view weights. The objective function of WV-Co-FCM is expressed as:

$$J_{WV-Co-FCM}(U, A, V) = \sum_{h=1}^s v_h \left[\sum_{i=1}^n \sum_{k=1}^c (z_{ik}^h)^m (a_{ik}^h)^2 + \Delta_h \right] + \lambda \sum_{h=1}^s v_h \ln v_h \quad (3)$$

where v_h is the h -th view weight with $\sum_{h=1}^s v_h = 1$, $\lambda > 0$ used to regulate view weights, $\Delta_h =$

$$\sum_{i=1}^n \alpha_{ik}^h \sum_{k=1}^c z_{ik}^h \left(1 - (z_{ik}^h)^{m-1} \right) - \sum_{i=1}^n \tau_{ik}^h \sum_{k=1}^c z_{ik}^h \left(1 - (z_{ik}^h)^{m-1} \right), \text{ and } d_{ik}^h = \sqrt{\sum_{j=1}^{d_h} (x_{ij}^h - a_{kj}^h)^2}.$$

WV-Co-FCM is designed to address multiview data, where weights play a crucial role in the final step. Specifically, without the influence of weights, the membership of the cluster center and each object are updated independently.

Wang and Chen [32] presented the Minmax-FCM clustering algorithm, which operates without a collaboration step and considers different weights for each view. Minmax-FCM is constructed based on the single-view FCM, where it evaluates the least distance between the membership matrix (U^*) and cluster centers (A^h) in each view. The extreme value within views (v_h) influences the smallest separation between (U^*) and (A^h). Consequently, the consistent clustering results in Minmax-FCM are achieved through Minmax optimization, aiming to minimize dissimilarities across diverse views. The formulation of Minmax-FCM [35] is as follows:

$$\min_{U^*, \{A^h\}_{h=1}^s} \max_{\{v_h\}_{h=1}^s} \sum_{h=1}^s (v_h)^\alpha Q^h \quad (4)$$

In the Minmax-FCM scenario, $Q^h = \sum_{i=1}^n \sum_{k=1}^c (z_{ik}^*)^m \|x_i^h - a_k^h\|^2$ and α serve as user-defined parameters to regulate the distribution of views.

While previous works, such as WMCFS, FRMVKM, WVCofCM, and Minmax-FCM, have made significant contributions to multiview clustering by focusing on feature selection, relevance filtering and different weights for views, there remains a gap in the effective integration of feature and view weights within the k-means framework. Specifically, existing algorithms focus either exclusively on feature selection (e.g., WMCFS) or weight adjustment (e.g., FRMVKM and WVCofCM) but fail to comprehensively combine feature and view weights within a unified clustering strategy. Although the Minmax FCM takes view weights into account, it does not include a collaboration step, which further limits its ability to effectively capture the interdependencies between features and views. In response to this gap, we propose two novel algorithms, W-MV-KM and W-MV-KN-L2-Norm, both of which are based on the k-means algorithm and integrate feature weights and view weights in a synergistic manner. These algorithms provide an innovative approach by extending the k-means framework with weighted strategies to better capture the complex relationships in multiview datasets, improve clustering performance, and ensure more accurate feature selection. This contribution addresses the limitations of previous methods and presents a more comprehensive solution for multiview clustering.

3. Weighted Multiview K-Means Clustering

In this section, we introduce two weighted multiview k-means clustering algorithms tailored for multiview data. We additionally explore the clustering performance of weighted multiview k-means in comparison to existing multiview clustering algorithms. Consider a multiview dataset denoted as X , comprising h views and d_h features, where $X = \{x_1, x_2, \dots, x_n\}$ in the Euclidean space \mathbb{R}^{d_h} where $x_i = \{x_i^h\}_{h=1}^s$. Let us consider the view vector weight $v_h \in [0, 1]$, $V = [v_h]_{1 \times s}$ with $\sum_{h=1}^s v_h = 1$. let $Z = [z_{ik}]_{n \times c}$, where $z_{ik} = 1$ assuming the data point x_i belongs to the k -th cluster, $z_{ik} = 0$ otherwise, i.e., $z_{ik} \in [0, 1]$, let $W = [w_j]_{1 \times d_h}$ where $w_j = \{w_j^h\}_{h=1}^s$, let $A = \{a_1, a_2, \dots, a_c\}$ be the c clusters. Therefore, we introduce our first weighted multiview k-means (W-MV-KM) algorithm, which assigns distinct weights to both views and features. Our objective is to devise a weighted scheme that aims to discern the significance of views as well as features within each view. The objective function for the proposed W-MV-FCM is

$$J(V, Z, A, W) = \sum_{h=1}^s (v_h)^\alpha \sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2 \tag{5}$$

Theorem 1. Assume $z_{ik} \in [0, 1]$, let $W = [w_j]_{1 \times d_h}$, $\sum_{j=1}^{d_h} w_j^h = 1$ and v_h is the view weight for the h -th view $\sum_{h=1}^s v_h = 1$. The constraints for minimizing the objective function of W-MV-KM $J(V, Z, A, W)$ are as follows:

$$a_{kj}^h = \frac{\sum_{i=1}^n z_{ik} x_{ij}^h}{\sum_{i=1}^n z_{ik}} \tag{6}$$

$$z_{ik} = \begin{cases} 1 & \text{if } \sum_{h=1}^s (v_h)^\alpha \sum_{j=1}^{d_h} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2 = 1 \leq k \leq c \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$v_h = \left[\sum_{h'=1}^s \left(\frac{\sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2}{\sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_{h'}} (w_j^{h'})^\tau (x_{ij}^{h'} - a_{kj}^{h'})^2} \right)^{1/\alpha-1} \right]^{-1} \tag{8}$$

$$w_j^h = \left[\sum_{j'=1}^{d_h} \left(\frac{(v_h)^\alpha \sum_{i=1}^n \sum_{k=1}^c z_{ik} (x_{ij}^h - a_{kj}^h)^2}{(v_h)^\alpha \sum_{i=1}^n \sum_{k=1}^c z_{ik} (x_{ij'}^h - a_{kj'}^h)^2} \right)^{1/\tau-1} \right]^{-1} \tag{9}$$

Proof. For the optimization of the proposed W-MV-KM, the Lagrange multiplier technique is used. $J(V, Z, A, W) = \sum_{h=1}^s (v_h)^\alpha \sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2 + \lambda_1 \left(\sum_{h=1}^s v_h - 1 \right) + \lambda_2 \left(\sum_{j=1}^{d_h} w_j^h - 1 \right)$ regarding the cluster center, membership, view weight, and feature weight, the constraints for minimizing the W-MV-KM are as follows. Initially, we hold V , W , and Z fixed and solve the updated equation for the cluster center. By taking the partial deriva-

tive of the $J(V, Z, A, W)$ with respect to a_{kj}^h and making them equal to zero, we have $\frac{\partial J}{\partial a_{kj}^h} = (v_h)^\alpha \sum_{i=1}^n z_{ik} (w_j^h)^\tau (-2) (x_{ij}^h - a_{kj}^h) = 0$; solving for a_{kj}^h , we obtain the updated Equation (6). Subsequently, with V, A , and W held constant, we solve the updated equation for the membership. By taking the partial derivative of the Lagrangian concerning z_{ik} and equating it to 0, we obtain Equation (7). Assuming Z, A , and W are fixed, we proceed with the partial derivative of the Lagrangian $J(V, Z, A, W)$ with respect to v_h and equate the result to 0. $\frac{\partial J}{\partial v_h} = \alpha v_h^{\alpha-1} \sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2 + \lambda_1 = 0$. We obtain the updated Equation (8) for the view weight. Given fixed Z, A , and V , we compute the partial derivative of the Lagrangian $J(V, Z, A, W)$ with respect to v_h and set it equal to zero, such that $\frac{\partial J}{\partial w_j^h} = (v_h)^\alpha \sum_{i=1}^n \sum_{k=1}^c z_{ik} \tau (w_j^h)^{\tau-1} (x_{ij}^h - a_{kj}^h)^2 + \lambda_2 = 0$. Upon solving the equation, we can drive the updating Equation (9) for the feature weight w_j^h . Therefore, the proposed W-MV-KM algorithm can be summarized as Algorithm 1. \square

Algorithm 1. The W-MV-KM Algorithm.

Input : Dataset $X = \{x_1, x_2, \dots, x_n\}$ with $x_i = \{x_i^h\}_{h=1}^s$ and $x_i^h = \{x_{ij}^h\}_{j=1}^{d_h}$,
 “c” number of cluster and $\varepsilon > 0$

Output : a_{kj}^h, z_{ik}, w_j^h and v_h .

Initialization : Randomly generate $Z^{(t)}$. Initialize the feature weight

$W^{h(0)} = [w_j^h]_{1 \times d_h}, 1 \leq j \leq d_h$ with $\sum_{j=1}^{d_h} w_j = 1$ by $w_j^h = 1/d_h$, initialize view weight

$V^{(0)} = [v_h]_{1 \times s}$ (user may define $v_h = 1/s \forall h$), and set $t = 1$.

Step 1: Compute the cluster centers $A^{h(t)}$ using $Z^{(t-1)}$ by Equation (6).

Step 2: Update the membership matrix $Z^{(t)}$ using $A^{h(t)}, V^{(t-1)}$ and $W^{h(t-1)}$ by Equation (7).

Step 3: Update the view weight $V^{(t)}$ by (8) using $Z^{(t)}, A^{h(t)}$ and $W^{h(t-1)}$.

Step 4: Update the feature weight $W^{h(t)}$ using (9) $Z^{(t)}, A^{h(t)}$ and $V^{(t)}$.

Step 5: If $\| \|W^{h(t)}\| - \|W^{h(t-1)}\| \| < \varepsilon$, then stop;

Else go back to Step 1.

In W-MV-KM, α and τ represent two exponent parameters utilized for regulating the distribution of feature weights and view weights. When both α and τ are set to zero, the W-MV-KM objective function simplifies to the k-means objective function.

We aim to discover an optimal combination of α and τ values to minimize the intra-cluster variance of the h -th view and j -th features. The parameter α regulates the distribution of view weights. Let $F_1(h) = \sum_{k=1}^c \sum_{i=1}^n \sum_{j=1}^{d_h} z_{ik} (w_j^h)^\tau (x_{ij}^h - a_{kj}^h)^2$ then Equation (8) will can be

expressed as $v_h = 1 / \left(\sum_{h'=1}^s (F_1(h) / F_1(h'))^{\frac{1}{\alpha-1}} \right)$, where $F_1(h)$ is the intra-cluster variance.

When values of $\alpha \leq 1$, Equation (8) suggests that only one view is chosen for clustering, which is not suitable for handling multiview datasets. As the values of α increase, the distribution of view weights becomes more uniform, indicating that more views will contribute to the clustering process. Conversely, the parameter τ , which governs the distribution of

feature weights, can be examined as follows: Let us use $F_2(h) = \sum_{i=1}^s v_h^\alpha \sum_{k=1}^c \sum_{i=1}^n z_{ik} (x_{ij}^h - a_{kj}^h)^2$ to represent the intra-cluster variance of the j -th feature. With this, Equation (9) can be expressed as $w_j^h = 1 / \left(\sum_{j=1}^{d_{h'}} (F_2(j) / F_2(j'))^{\frac{1}{\tau-1}} \right)$. If $\tau = 1$, this indicates the selection of only one feature. Then, we need to select values of τ that stabilize the algorithm and ultimately yield favorable results for W-MV-KM.

The clustering outcomes are influenced by the parameters α and τ in the W-MV-KM algorithm, posing challenges in both the control and estimation of results. Hence, to manage and stabilize these effects, we set the values of both α and τ to 2. This is how we apply regularization to the objective function for both the feature and view weights. We subsequently introduce another multiview learning approach for W-MV-KM, integrating L2 regularization, which we refer to as W-MV-KM-L2. In W-MV-KM-L2, L2 regularization is applied to both the features and view weights. The objective function of the W-MV-KM-L2 is formulated in the following:

$$J(V, Z, A, W) = \sum_{h=1}^s (v_h)^2 \sum_{i=1}^n \sum_{k=1}^c \sum_{j=1}^{d_h} z_{ik} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 + \alpha \sum_{h=1}^s \|v_h\|_2^2 + \tau \sum_{j=1}^{d_h} \|w_j^h\|_2^2 \quad (10)$$

$$\sum_{h=1}^s v_h = 1, \sum_{j=1}^{d_h} w_j^h = 1, \sum_{j=1}^{d_h} \|w_j^h\|_2^2 = \sum_{j=1}^{d_h} (w_j^h)^2, \sum_{h=1}^s \|v_j^h\|_2^2 = \sum_{h=1}^s (v_h)^2.$$

In Algorithm 2, we incorporate the exponential parameters α and τ with the regularization framework. This integration serves to control the distribution of both features and views. In W-MV-KM, the parameters α and τ are determined by the user, while in W-MV-KM-L2, we calculate the parameters within the regularization terms by means of the equations $\alpha = c/d_h$ and $\tau = \sum_{i=1}^n \|x_i^h - \bar{x}_i^h\|^2 / n$. Likewise, we derive the updating equations for the W-MV-KM-L2 objective function through the utilization of the Lagrange multiplier. Consequently, we derive Theorem 2, as presented below.

Theorem 2. Suppose $\sum_{h=1}^s v_h = 1, v_h \in [0, 1]$ and $\sum_{j=1}^{d_h} w_j^h = 1, w_j \in [0, 1]$. The necessary conditions for minimizing the objective function $J_{W-MV-KM-L2}(V, Z, A, W)$ are

$$a_{kj}^h = \frac{\sum_{i=1}^n z_{ik} x_{ij}^h}{\sum_{i=1}^n z_{ik}} \quad (11)$$

$$z_{ik} = \begin{cases} 1 & \text{if } \sum_{h=1}^s (v_h)^2 \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 = 1 \leq k \leq \min_{h=1}^s c \sum_{h=1}^s (v_h)^2 \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

$$v_h = \sum_{h'=1}^s \left[\frac{\sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 + \alpha}{\sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_{h'}} (w_j^{h'})^2 (x_{ij}^{h'} - a_{kj}^{h'})^2 + \alpha} \right]^{-1} \quad (13)$$

$$w_j^h = \sum_{j'=1}^{d_h} \left[\frac{(v_h)^2 \sum_{i=1}^n \sum_{k=1}^c z_{ik} (x_{ij}^h - a_{kj}^h)^2 + \tau}{(v_h)^2 \sum_{i=1}^n \sum_{k=1}^c z_{ik} (x_{ij'}^h - a_{kj'}^h)^2 + \tau} \right]^{-1} \quad (14)$$

Proof. To demonstrate the optimization problem described in Equation (10), we employ the Lagrange multipliers. Consequently, we construct the Lagrangian function as follows. We first set W , V and Z to solve the updated equation for cluster centers a_{kj}^h . We achieve this by computing partial derivatives with respect to a_{kj}^h and subsequently setting them equal to zero, yielding the following equation $\frac{\partial J}{\partial (a_{kj}^h)} = (v_h)^2 \sum_{i=1}^n z_{ik} (w_j^h)^2 (x_{ij}^h - a_{kj}^h) (-2) = 0$.

Upon solving for a_{kj}^h , we obtained the updated Equation (11) $a_{kj}^h = \frac{\sum_{i=1}^n z_{ik} x_{ij}^h}{\sum_{i=1}^n z_{ik}}$. Subsequently, by fixing A , W , and V , we proceed to solve the updated equation for the membership degree z_{ik} . After computing the partial derivative of $J(V, Z, A, W)$ with respect to z_{ik} and setting it equal to zero, we obtain $\frac{\partial J}{\partial z_{ik}} = \sum_{h=1}^s (v_h)^2 \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 = 0$. From this, we obtain the updated Equation (12). Consequently, with fixed Z , A , and W , we compute the partial derivative of $J(Z, A, W, V)$ with respect to v_h and equating it to zero, we have $\frac{\partial J}{\partial v_h} = 2(v_h) \sum_{i=1}^n \sum_{k=1}^c z_{ik} \sum_{j=1}^{d_h} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2 + 2\alpha(v_h) + \lambda_1 = 0$. Further solving this equation leads to the derivation of the updated Equation (13). We compute the partial derivative of $J(V, Z, A, W)$ with respect to w_j^h and set it to zero. The results obtained are as follows $\frac{\partial J}{\partial w_j^h} = (v_h)^2 \sum_{i=1}^n \sum_{k=1}^c z_{ik} (2) (w_j^h) (x_{ij}^h - a_{kj}^h)^2 + 2\tau(w_j^h) + \lambda_3 = 0$; therefore, the updated Equation (14) is obtained. \square

Algorithm 2. The W-MV-KM-L2 Algorithm

Input : Dataset $X = \{x_1, x_2, \dots, x_n\}$ with $x_i = \{x_i^h\}_{h=1}^s$ and $x_i^h = \{x_{ij}^h\}_{j=1}^{d_h}$, “ c ” number of cluster and $\varepsilon > 0$.

Output: a_{kj}^h , z_{ik} , w_j^h and v_h .

Initialization : Randomly generate $Z^{(t)}$. Initialize the feature weight $W^{h(0)} = [w_j^h]_{1 \times d_h}$, $1 \leq j \leq d_h$ with $\sum_{j=1}^{d_h} w_j = 1$ by $w_j^h = 1/d_h$, initialize view weight $V^{(0)} = [v_h]_{1 \times s}$ (user may define $v_h = 1/s \forall h$), and set $t = 1$.

Step 1: Calculate $\alpha = c/d_h$ and $\tau = \sum_{i=1}^n (x_i^h - \bar{x}_i^h)^2 / n$ where τ is the covariance of the data points in the h^{th} view.

Step 2: Compute the cluster centers $A^{h(t)}$ using $Z^{(t-1)}$ by Equation (11).

Step 3: Update the membership matrix $Z^{(t)}$ using $A^{h(t)}$, $W^{h(t-1)}$ and $V^{(t-1)}$ by Equation (12).

Step 4: Update the view weight $V^{(t)}$ using (13) $Z^{(t)}$, $W^{h(t-1)}$ and $A^{h(t)}$.

Step 5: Update the feature weight $W^{h(t)}$ using Equation (14) $Z^{(t)}$, $V^{(t)}$ and $A^{h(t)}$.

Step 6: If $\| \|W^{h(t)}\| - \|W^{h(t-1)}\| \| < \varepsilon$, then stop.

Otherwise, go back to Step 2.

The W-MV-KM-L2 incorporates parameters α and τ , which play a vital role in determining the classification of view weights and feature weights. To regulate the behavior of α and τ , we select appropriate values of parameters in a manner that ensures the stability of the algorithm, thus enabling the computation of final estimates. The parameters α and τ in the W-MV-KM-L2 objective function are computed through $\alpha = c/d_h$ and

$\tau = \sum_{i=1}^n (x_i^h - \overline{x_i^h})^2 / n$, where c signifies the total number of clusters and d_h signifies the total number of features in each view. If the value of d_h is larger than the value of α , it becomes very small and consequently does not significantly contribute to the determination of view weights. The role of c becomes crucial in regulating the values of α .

Let $G_1(h) = \sum_{k=1}^c \sum_{i=1}^n \sum_{j=1}^{d_h} z_{ik} (w_j^h)^2 (x_{ij}^h - a_{kj}^h)^2$ be the intra-cluster variances in the h -th view.

Then, it becomes $v_h = \sum_{h'=1}^s [G_1(h) + \alpha_h / G_1(h') + \alpha_{h'}]$. Smaller values of α_h will result in smaller weights, whereas large values of $\alpha_{h'}$ will lead to larger view weights v_h . This demonstrates that α_h supervises the weights in each view. Thus, view weights are utilized to update the feature weights w_j^h in each view. Meanwhile, the other parameter

τ is estimated using the sample variance for each view, such that $\tau = \sum_{i=1}^n (x_i^h - \overline{x_i^h})^2 / n$

and it is utilized to regulate the distribution in each view. From Equation (14), we regard $G_2(j) = \sum_{h=1}^s (v_h)^2 \sum_{k=1}^c \sum_{i=1}^n z_{ik} (x_{ij}^h - a_{kj}^h)^2$ as the intra-cluster variance of the features in view.

Thus, Equation (14) becomes $w_j^h = \sum_{j'=1}^{d_h} \left[\frac{G_2(j) + \tau_j}{G_2(j') + \tau_{j'}} \right]$. If the values of τ_j are larger, then the feature weights also become larger, whereas if the values of τ_j are smaller, then the feature weights also become smaller. Hence, τ_j is employed to diminish the feature weights that are irrelevant. Consequently, features with larger weights are prioritized over those with smaller weights.

Computational Complexity: We conduct an analysis of computational complexity for both the proposed W-MV-KM algorithm and the W-MV-KM-L2 algorithm. These algorithms can be segmented into four parts: (1) compute the cluster center a_k^h with $O(nsc)$; (2) update the membership degree z_{ik} with $O(nsc2d)$; (3) update the view weight v_h with $O(nscd_h^2)$; (4) update the feature weights v_h with $O(nscd_h^2)$. The total computational complexity for the W-MV-KM and W-MV-KM-L2 algorithm is $O(nsc^2d_h + nsc^2d_h)$, where n represents the number of data points, s denotes the number of views, c signifies the number of clusters, and d_h stands for the dimensionality of data points.

4. Experimental Comparisons and Results

In this section, two synthetic and seven real-life datasets are used to demonstrate the performance of the proposed W-MV-KM-L2 algorithm. We compare the W-MV-KM-L2 algorithm with five existing algorithms, W-MV-KM, WMCFS [30], WV-Co-FCM [31], Minmax-FCM [32], and FRMVKM [33], in this section. For the experiments, we use the same initializations for the cluster center, feature weight, and view weights. For measuring the clustering performance, we use the following evaluation measures: accuracy rate (AR), Jaccard index (JI) [44], Fowlkes–Mallows index (FMI) [45], Rand Index (RI) [46], and normalized mutual information (NMI) [47]. Accuracy rate (AR) is the term used to describe the proportion of correct predictions or classifications made by a model or system, expressed as a percentage of the total number of predictions or classifications attempted. The Jaccard index (JI) [44] shows how similar two sets are by comparing the number of items they have in common to the overall count of unique items in both sets. The FMI [45] is a metric utilized to assess the similarity between two clustering results. It calculates the geometric mean of the pairwise similarity between points in the same clusters across two different clustering results. The RI [46] measures how similar two sets of clusters are. It counts pairs of data points that are either in the same cluster in both sets or different clusters in both sets, and the NMI [47] is like a friendly compass that helps us navigate through clusters of data. It measures how well two different clustering results match up, considering the size and arrangement of the clusters. The greater the magnitudes of AR, JI, FMI, RI, and NMI, the more exceptional the clustering performance becomes.

The experiments are executed in Matlab 2020a with identical initializations. They are repeated 100 times using distinct random initializations of the WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM, and W-MV-KM-L2 algorithms. For Minmax-FCM and WV-Co-FCM, the fuzziness index m is 2 for the methods.

Example 1. We generate a dataset with two distinct views using a Gaussian mixture model: $\sum_{h=1}^2 \frac{1}{2} N(\mu_i^h, \Sigma_i^h)$, where h represents the view index. These two views are then merged to form the samples (x_i^1, x_i^2, c_k) . The dataset consists of two clusters across both views, with 1500 data points sampled for each view. For the first view, the cluster means are set to $\mu_1^{(1)} = (1.5 \ 1.5)$ and $\mu_2^{(1)} = (7 \ 7)$. In the second view, the cluster means are $\mu_1^{(2)} = (2.5 \ -2.5)$ and $\mu_2^{(2)} = (5 \ 5)$. The covariance matrices $\Sigma_1^{(i)} = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}$, $i = 1, \dots, 4$, are applied to both views. Each view consists of two main features: $x_1^{(1)}$ & $x_2^{(1)}$ for view 1, and $x_1^{(2)}$ & $x_2^{(2)}$ for view 2. Additionally, we introduce one noise feature per view. For view 1, the noise feature is $x_3^{(1)}$, sampled from a uniform distribution in the range $[0, 5]$, and for view 2, the noise feature is $x_3^{(2)}$, sampled from a uniform distribution in the range $[0, 10]$, named Artificial 1 dataset. A 2D and 3D graph illustrating the dataset is shown in Figure 1a–d for both view 1 and view 2, respectively.

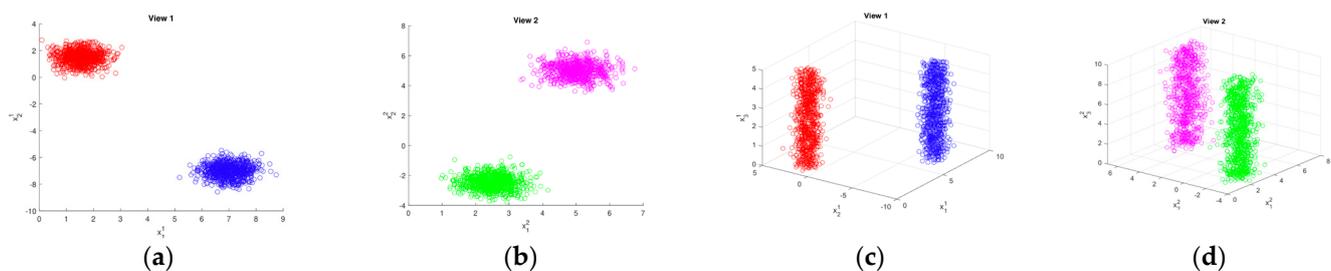


Figure 1. A 2D and 3D graphical representation of 2-cluster dataset with 2 views, each containing 1500 data points. (a) 2D visualization of view 1; (b) 2D visualization of view 2; (c) 3D visualization of view 1; and (d) 3D visualization of view 2.

Figure 2 shows the clustering visualization of the two algorithms W-MV-KM and W-MV-KM-L2 for both view 1 and view 2 in 3D representation. Table 1 shows the clustering performance of the proposed algorithms with the existing algorithms, where the W-MV-KM-L2 algorithm performs well, as compared to other algorithms in all evaluation measures. In this example, we try different values for two parameters α and τ and report the values of the algorithms in Table 2.

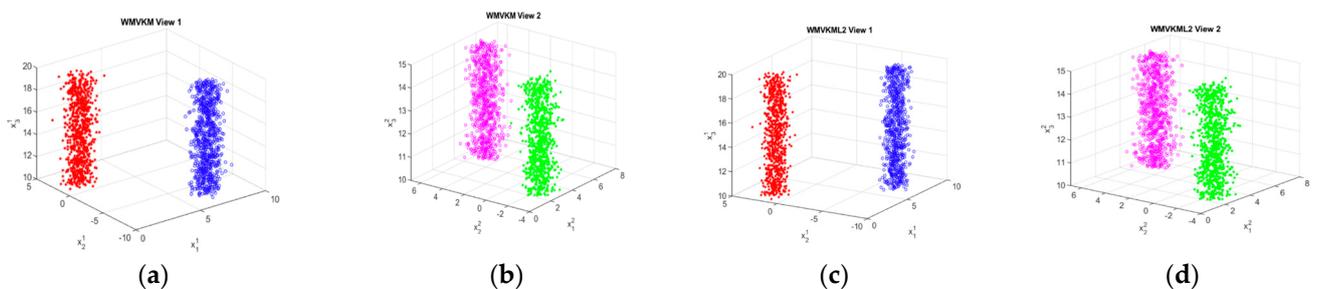


Figure 2. Clustering performance of WMVKM and WMVKML2 in 3D. (a) WMVKM visualization of view 1; (b) WMVKM visualization of view 2; (c) WMVKML2 visualization of view 1; and (d) WMVKML2 visualization of view 2.

Table 1. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for Artificial 1 multiview data based on average (STD).

Methods	AR	Jl	FMI	RI	NMI
WMCFS	0.844 (0.212)	0.764 (0.310)	0.826 (0.230)	0.826 (0.230)	0.647 (0.464)
FRMVKM	0.960 (0.116)	0.926 (0.187)	0.949 (0.132)	0.949 (0.132)	0.889 (0.280)
WV-Co-FCM	0.500 (0.000)	0.500 (0.000)	0.707 (0.000)	0.500 (0.000)	0.706 (0.000)
Minmax-FCM	0.500 (0.000)	0.500 (0.000)	0.707 (0.000)	0.500 (0.000)	0.706 (0.000)
W-MV-KM	0.718 (0.311)	0.686 (0.333)	0.765 (0.764)	0.768 (0.250)	0.530 (0.499)
W-MV-KM-L2	0.960 (0.137)	0.947 (0.180)	0.960 (0.135)	0.960 (0.135)	0.920 (0.270)

Table 2. Clustering sensitivity analysis of the parameters α and τ .

Algorithms	α	τ
W-MV-KM	1, 2, 3...	1, 2, 3...
W-MV-KM-L2	$\alpha = c/d_h$	$\tau = \frac{1}{n} \sum_{i=1}^n (x_i^h - \bar{x}_i^h)^2$

In W-MV-KM, we try different values for the two parameters, as mentioned in Table 2. If we choose for both parameters $\alpha = \tau = 1$, then the algorithm is unable to give results; if we choose $\alpha = 2$, then the algorithm performs, but the results are very poor as compared to other existing algorithms. Thus, we propose W-MV-KM-L2, where the values of α and τ are determined by the following formulas, as discussed in Section 3. In the W-MV-KM algorithm, if we keep values of $\alpha < 2$, then the algorithm is unable to perform. So, for this, it should be $\alpha \geq 2$. If we increase the parameter values $\alpha = \tau \geq 2$, even then, the clustering performance is not improved. In both algorithms, W-MV-KM and W-MV-KM-L2, the view weight should be ≥ 2 because if it is < 2 , then only one view is selected, which is against the basic principal of multiview clustering.

Example 2. In this example, we use the synthetic data used in [35], called Syn1. To implement the dataset in the proposed and existing algorithms, we keep different initializations for existing algorithms: we choose $\alpha = 2$ and $\tau = 0.001$ for WMCFS; for FRMVKM, we assume $\alpha = 2$; and for WV-Co-FCM, $\tau = 1.3$. These serve as the parameters. With the given initializations, we apply these algorithms to the dataset, and their clustering performance is summarized in Table 3. The results indicate that W-MV-KM-L2 obtained the highest values across all evaluation metrics.

Table 3. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for Syn1 multiview data based on average (STD).

Methods	AR	Jl	FMI	RI	NMI
WMCFS	0.861 (0.218)	0.807 (0.303)	0.855 (0.227)	0.855 (0.227)	0.710 (0.453)
FRMVKM	0.924 (0.154)	0.870 (0.249)	0.906 (0.180)	0.906 (0.180)	0.805 (0.371)
WV-Co-FCM	0.943 (0.132)	0.902 (0.184)	0.938 (0.125)	0.929 (0.147)	0.842 (0.291)
Minmax-FCM	0.500 (0.000)	0.499 (0.000)	0.707 (0.000)	0.499 (0.000)	0.000 (0.000)
W-MV-KM	0.856 (0.291)	0.868 (0.265)	0.901 (0.198)	0.901 (0.198)	0.802 (0.297)
W-MV-KM-L2	0.972 (0.112)	0.960 (0.158)	0.970 (0.118)	0.970 (0.118)	0.940 (0.237)

Example 3. In this example, we use the synthetic dataset used in [48] to implement the W-MV-KM-L2, W-MV-KM, Minmax-FCM, WV-Co-FCM, FRMVKM, and WMCFS in the Syn500 dataset. For our initializations, we choose $\alpha = 4$ and $\tau = 0.01$ for WMCFS; for FRMVKM, we

assume $\alpha = 2$; and for WV-Co-FCM, $\tau = 4$. For Minmax-FCM and WV-Co-FCM, the fuzziness index is $m = 2$. With the given initializations, we apply these algorithms to the dataset, and their clustering performance is summarized in Table 4. The results indicate that W-MV-KM-L2 obtained the highest values across all evaluation metrics.

Table 4. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for Syn500 multiview data based on average (STD).

Methods	AR	Jl	FMI	RI	NMI
WMCFS	0.671 (0.138)	0.429 (0.116)	0.593 (0.101)	0.592 (0.102)	0.152 (0.178)
FRMVKM	0.873 (0.015)	0.644 (0.029)	0.784 (0.023)	0.778 (0.021)	0.532 (0.062)
WV-Co-FCM	0.740 (0.000)	0.446 (0.000)	0.617 (0.000)	0.614 (0.000)	0.179 (0.000)
Minmax-FCM	0.500 (0.000)	0.499 (0.000)	0.706 (0.000)	0.499 (0.000)	0.000 (0.000)
W-MV-KM	0.740 (0.152)	0.499 (0.118)	0.658 (0.106)	0.656 (0.104)	0.265 (0.189)
W-MV-KM-L2	0.880 (0.027)	0.657 (0.017)	0.793 (0.013)	0.784 (0.013)	0.523 (0.027)

We first normalized all the real datasets. We tested the performance of our proposed algorithms with seven real datasets. Seven real-world multiview datasets serve as the benchmark to assess the efficacy of the W-MV-KM-L2 algorithm. These are the Minist4 dataset [49], Handwritten4 (HW) [50], Caltech2 dataset [51], UCI Derm dataset [52], HumanEva 3D Motion [53–55] dataset, UCI 3views dataset [50], and Microsoft Research Cambridge Volume 1 (MSRC-V1) dataset [56], respectively. The characteristics of these seven real datasets are displayed in Table 5 in terms of cluster number c , data type s , the data number n , the feature dimension d_h , and the view numbers.

Table 5. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for Syn500 multiview data based on average (STD).

Methods	C	n	s	d_1	d_2	d_3	d_4	d_5
Caltech2	3	2033	2	300	256	-	-	-
HumanEva 3D Motion	5	5000	2	48	48	-	-	-
Derm	6	366	2	11	12	-	-	-
Minist4	4	4000	3	30	9	30	-	-
UCI 3views	10	2000	3	240	76	6	-	-
HW1256	10	2000	4	76	216	47	6	-
MSRC-V1	7	210	5	24	576	512	256	254

Example 4. In this example, we use seven real datasets for comparing W-MV-KM-L2 with all the algorithms, such as WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM and W-MV-KM. Various combinations of α and τ will result in different distributions for the view and feature weight vectors, respectively. This is the case for WMCFS and WV-Co-FCM, where α and τ are user defined. For FRMVKM, we use $\alpha = 2$ for all datasets. In W-MV-KM-L2, the regularization parameters α and τ are used to control the sparsity of the view weights and feature weights, respectively. The parameters α and τ in W-MV-KM-L2 are simply calculated by $\alpha = c/d_h$ and $\tau = \sum_{i=1}^n (x_i^h - \bar{x}_i^h)^2 / n$. We compare the results of the proposed algorithms with existing algorithms. To achieve this, we utilize real datasets and apply the algorithms using the assumed values of τ in Table 6.

For each given value of α , we ran 100 simulations for each algorithm using different seed generations across all clustering methods and recorded their clustering performance

metrics. This process was repeated for all real datasets listed in Table 6. We present only the overall average mean and standard deviation of the performance measures for each algorithm and compare them with the results of W-MV-KM-L2. The outcomes, summarized in Table 7, show that W-MV-KM-L2 consistently achieves higher results for AR, RI, NMI, JI, and FMI across all real datasets.

Table 6. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for Syn500 multiview data based on average (STD).

	Minist4	Handwritten4	Caltech2	Derm	Human Eva 3D	UCI 3views	MSRC-V1
WMCFS	0.1	0.001	0.001	0.001	0.01	0.01	0.01
WV-Co-FCM	3	5	3	2	4	2	3

Table 7. Clustering performances of WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 for seven real multiview data based on average (STD).

Dataset	Methods	AR	JI	FMI	RI	NMI
Minist4	WMCFS	0.661 (0.166)	0.445 (0.157)	0.600 (0.145)	0.795 (0.075)	0.483 (0.175)
	FRMVKM	0.444 (0.163)	0.267 (0.103)	0.411 (0.123)	0.698 (0.062)	0.229 (0.170)
	WV-Co-FCM	0.250 (0.000)	0.250 (0.000)	0.500 (0.000)	0.250 (0.000)	0.000 (0.000)
	Minmax-FCM	0.250 (0.000)	0.250 (0.000)	0.500 (0.000)	0.250 (0.000)	0.000 (0.000)
	W-MV-KM	0.209 (0.016)	0.171 (0.001)	0.293 (0.001)	0.609 (0.002)	0.043 (0.000)
	W-MV-KM-L2	0.789 (0.126)	0.568 (0.097)	0.720 (0.083)	0.853 (0.051)	0.637 (0.085)
HW1256	WMCFS	0.470 (0.055)	0.284 (0.026)	0.444 (0.032)	0.877 (0.010)	0.522 (0.028)
	FRMVKM	0.445 (0.078)	0.260 (0.056)	0.412 (0.072)	0.869 (0.056)	0.459 (0.072)
	WV-Co-FCM	0.100 (0.000)	0.100 (0.000)	0.000 (0.000)	0.100 (0.000)	0.316 (0.000)
	Minmax-FCM	0.200 (0.000)	0.142 (0.000)	0.340 (0.000)	0.529 (0.000)	0.264 (0.000)
	W-MV-KM	0.322 (0.185)	0.186 (0.092)	0.338 (0.110)	0.676 (0.325)	0.300 (0.226)
	W-MV-KM-L2	0.470 (0.055)	0.284 (0.026)	0.444 (0.032)	0.877 (0.010)	0.522 (0.028)
Caltech2	WMCFS	0.748 (0.150)	0.558 (0.141)	0.707 (0.114)	0.792 (0.082)	0.543 (0.144)
	FRMVKM	0.404 (0.076)	0.343 (0.046)	0.532 (0.054)	0.513 (0.058)	0.144 (0.120)
	WV-Co-FCM	0.394 (0.000)	0.354 (0.000)	0.595 (0.000)	0.354 (0.000)	0.000 (0.000)
	Minmax-FCM	0.390 (0.006)	0.351 (0.006)	0.587 (0.015)	0.362 (0.019)	0.013 (0.012)
	W-MV-KM	0.437 (0.068)	0.337 (0.018)	0.517 (0.019)	0.546 (0.051)	0.132 (0.050)
	W-MV-KM-L2	0.758 (0.161)	0.591 (0.160)	0.731 (0.126)	0.808 (0.092)	0.600 (0.161)
Derm	WMCFS	0.673 (0.093)	0.536 (0.099)	0.695 (0.083)	0.868 (0.040)	0.734 (0.067)
	FRMVKM	0.677 (0.085)	0.476 (0.070)	0.643 (0.062)	0.850 (0.034)	0.680 (0.054)
	WV-Co-FCM	0.475 (0.038)	0.290 (0.023)	0.488 (0.023)	0.636 (0.040)	0.322 (0.038)
	Minmax-FCM	0.306 (0.000)	0.199 (0.000)	0.446 (0.000)	0.199 (0.000)	0.000 (0.000)
	W-MV-KM	0.633 (0.135)	0.501 (0.128)	0.668 (0.105)	0.815 (0.170)	0.669 (0.192)
	W-MV-KM-L2	0.734 (0.117)	0.629 (0.131)	0.769 (0.099)	0.881 (0.131)	0.797 (0.168)
HumanEva 3D motion	WMCFS	0.448 (0.057)	0.304 (0.027)	0.480 (0.031)	0.722 (0.038)	0.412 (0.038)
	FRMVKM	0.200 (0.000)	0.200 (0.000)	0.447 (0.000)	0.304 (0.000)	0.000 (0.000)
	WV-Co-FCM	0.200 (0.000)	0.200 (0.000)	0.447 (0.000)	0.200 (0.000)	0.000 (0.000)
	Minmax-FCM	0.339 (0.000)	0.227 (0.001)	0.473 (0.001)	0.621 (0.007)	0.324 (0.005)
	W-MV-KM	0.200 (0.000)	0.200 (0.000)	0.447 (0.000)	0.200 (0.000)	0.000 (0.000)
	W-MV-KM-L2	0.691 (0.114)	0.533 (0.093)	0.693 (0.077)	0.865 (0.044)	0.691 (0.072)

Table 7. Cont.

Dataset	Methods	AR	JI	FMI	RI	NMI
UCI 3views	WMCFS	0.471 (0.067)	0.285 (0.033)	0.446 (0.035)	0.87 (0.079)	0.526 (0.060)
	FRMVKM	0.316 (0.044)	0.202 (0.021)	0.351 (0.027)	0.813 (0.029)	0.375 (0.037)
	WV-Co-FCM	0.100 (0.000)	0.100 (0.000)	0.316 (0.000)	0.100 (0.000)	0.000 (0.000)
	Minmax-FCM	0.200 (0.000)	0.142 (0.000)	0.340 (0.000)	0.529 (0.000)	0.264 (0.000)
	W-MV-KM	0.228 (0.043)	0.154 (0.016)	0.277 (0.012)	0.789 (0.203)	0.336 (0.099)
	W-MV-KM-L2	0.654 (0.070)	0.450 (0.051)	0.620 (0.048)	0.920 (0.011)	0.697 (0.032)
MSRC-VI	WMCFS	0.543 (0.080)	0.314 (0.049)	0.477 (0.055)	0.844 (0.020)	0.513 (0.054)
	FRMVKM	0.297 (0.047)	0.137 (0.017)	0.240 (0.240)	0.782 (0.007)	0.225 (0.033)
	WV-Co-FCM	0.143 (0.000)	0.139 (0.000)	0.373 (0.000)	0.139 (0.000)	0.000 (0.000)
	Minmax-FCM	0.286 (0.005)	0.181 (0.000)	0.371 (0.002)	0.559 (0.007)	0.234 (0.002)
	W-MV-KM	0.358 (0.036)	0.163 (0.010)	0.281 (0.015)	0.786 (0.009)	0.298 (0.024)
	W-MV-KM-L2	0.590 (0.078)	0.343 (0.044)	0.511 (0.049)	0.856 (0.016)	0.547 (0.045)

Example 5. In this example, we compare the number of iterations required for convergence, along with the running time, for the WMCFS, FRMVKM, WV-Co-FCM, Minmax-FCM, W-MV-KM, and W-MV-KM-L2 clustering algorithms. The number of iterations needed for convergence is reported in Table 8. The initial parameters are set as follows: $m = 2$ is used as the fuzzifier value for Minimax-FCM and WV-Co-FCM. The convergence (error) tolerance for all the algorithms is set to be 1×10^{-3} . The running time per second achieved by all the six algorithms is reported in Table 9, and the lowest running time for any dataset achieved by any algorithm is highlighted in bold face.

Table 8. Number of iterations to converge the algorithms.

Datasets	WMCFS	FRMVKM	WV-Co-FCM	Minmax-FCM	WMVKM	W-MV-KM-L2
Minist4	8	5	≥ 20	≥ 50	24	19
Handwritten4	3	6	≥ 20	≥ 50	16	29
Caltech2	3	4	≥ 20	≥ 50	7	3
Derm	3	5	≥ 20	≥ 50	6	11
HumanEva 3D Motion	19	2	≥ 20	≥ 50	3	5
UCI3views	3	4	≥ 20	≥ 50	26	17
MSRC-V1	8	5	≥ 20	≥ 50	24	19

Table 9. The running time (RT) achieved through the proposed and other existing algorithms: WMCFS, FR-MV-KM, WV-Co-FCM, Minmax-FCM, W-MV-KM and W-MV-KM-L2 (the best is marked in boldface).

Datasets	WMCFS	FRMVKM	WV-Co-FCM	Minmax-FCM	WMVKM	W-MV-KM-L2
Minist4	0.5790	0.4200	204.3660	8.7610	1.9390	0.9830
Handwritten4	0.7980	0.7420	424.6910	12.7080	4.9650	6.0090
Caltech2	0.6210	0.5410	30.0290	7.7870	1.6920	0.4800
Derm	0.2100	0.2600	15.4410	0.8510	0.4840	0.3130
HumanEva 3D Motion	2.5780	0.3790	172.9800	15.7600	0.9250	0.6910
UCI3views	0.7430	0.6440	296.3110	13.1700	8.3240	3.9570
MSRC-V1	0.3280	0.3280	47.3710	2.6450	0.9830	0.7260

Example 6. In this example, we use the multiview cluster validity indices proposed by Yang and Hussain [11], the multiview Dunn (MV-Dunn) index, and multiview generalized Dunn (MV-G-Dunn) index. We implement the multiview validity indices only based on k -means such as WMCFS, FRMVKM, and W-MV-KM-L2. We consider the WMCFS+MV-Dunn, WMCFS+MV-G-Dunn, FRMVKM+MV-Dunn, FRMVKM+MV-G-Dunn, W-MV-KM-L2+MV-Dunn, and W-MV-KM-L2+MV-G-Dunn regularization with 40 different initializations. The estimated number of clusters, along with their percentages, determined using the MV-Dunn and MV-G-Dunn indices, are presented in Table 10 for WMCFS, Table 11 for FRMVKM, and Table 12 for WMVKML2.

Table 10. Results of WMCFS with MV-Dunn and MV-G-Dunn.

Datasets\Indices	MV-Dunn	MV-G-Dunn
Artificial 1	2 (65%)	2 (70%)
Syn1	2 (50%)	2 (72.5%)
Syn500	2 (42.5%)	2 (17.5%)
Caltech2	3 (10%)	3 (0%)
HumanEva 3D motion	5 (2.5%)	5 (0%)
Derm	6 (10%)	6 (0%)
Minist4	4 (15%)	4 (0%)
UCI3views	10 (2.5%)	10 (0%)
HW1256	10 (2.5%)	10 (0%)
MSRC-IV	7 (5%)	7 (0%)

Table 11. Results of FRMVKM with MV-Dunn and MV-G-Dunn.

Datasets\Indices	MV-Dunn	MV-G-Dunn
Artificial 1	2 (100%)	2 (2.5%)
Syn1	2 (100%)	2 (0%)
Syn500	2 (100%)	2 (0%)
Caltech2	3 (0%)	3 (0%)
HumanEva 3D motion	5 (0%)	5 (0%)
Derm	6 (0%)	6 (0%)
Minist4	4 (0%)	4 (0%)
UCI3views	10 (0%)	10 (0%)
HW1256	10 (0%)	10 (0%)
MSRC-IV	7 (0%)	7 (0%)

Table 12. Results of WMVKML2 with MV-Dunn and MV-G-Dunn.

Datasets\Indices	MV-Dunn	MV-G-Dunn
Artificial 1	2 (100%)	2 (2.5%)
Syn1	2 (100%)	2 (0%)
Syn500	2 (100%)	2 (0%)
Caltech2	3 (2.5%)	3 (0%)
HumanEva 3D motion	5 (0%)	5 (0%)
Derm	6 (2.5%)	6 (0%)

Table 12. Cont.

Datasets\Indices	MV-Dunn	MV-G-Dunn
Minist4	4 (0%)	4 (0%)
UCI3views	10 (0%)	10 (0%)
HW1256	10 (0%)	10 (0%)
MSRC-IV	7 (0%)	7 (0%)

5. Discussion

Defining the structure of multiview data through L2 regularization for k-means and incorporating view weights and feature weights into its objective function presents a significant challenge. In this manuscript, we introduce two innovative clustering algorithms for multiview data, namely W-MV-KM and W-MV-KM-L2, building on the foundation of k-means. In W-MV-KM, users are tasked with assigning weights to both features and views. The parameters α and τ are utilized to regulate the distribution of weights and features in W-MV-KM, respectively. As the values of α and τ are user-determined for these weights, identifying an optimal combination of view and feature weights can present a challenge. To tackle this concern, we suggest an alternative method called W-MV-KM-L2, which integrates L2 regularization into the W-MV-KM framework. The parameters α and τ utilized for the regularization of W-MV-KM-L2 are subsequently estimated. This approach enhances the significance of results for both view and feature weights, ultimately leading to an improvement in clustering performance.

6. Conclusions

In the empirical findings for real-world datasets for both the proposed algorithm and existing algorithms, W-MV-KM-L2 outperformed in clustering performance. In our future work, we aim to explore making the W-MV-KM-L2 algorithm independent of initialization requirements. Additionally, we plan to enhance the W-MV-KM-L2 algorithm by incorporating the ability to determine an unsupervised optimal number of clusters. Since point symmetry-based distance [42] is more effective at representing cluster symmetry behavior [43] compared to Euclidean distance, we plan to extend the W-MV-KM-L2 algorithm in future work by incorporating point symmetry-based distance to enhance its ability to present cluster symmetry.

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

Funding: This research was funded by Taif University, Saudi Arabia, under project number (TU-DSPP2024-139).

Data Availability Statement: All data generated or analyzed during this study are included in this article.

Acknowledgments: The authors wish to acknowledge the anonymous reviewers and editors for their valuable feedback and insights.

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

References

- Jain, A.K.; Dubes, R.C. *Algorithms for Clustering Data*; Prentice-Hall, Inc.: Hoboken, NJ, USA, 1988.
- Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley: New York, NY, USA, 1990.
- MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.

4. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2020**, *31*, 651–666. [[CrossRef](#)]
5. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Kluwer: Norwell, MA, USA, 1981.
6. Ruspini, E.H.; Bezdek, J.C.; Keller, J.M. Fuzzy clustering: A historical perspective. *IEEE Comput. Intell. Mag.* **2019**, *14*, 45–55. [[CrossRef](#)]
7. Yang, M.S.; Nataliani, Y. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. *Pattern Recognit.* **2017**, *71*, 45–59. [[CrossRef](#)]
8. Krishnapuram, R.; Keller, J.M. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1993**, *1*, 98–110. [[CrossRef](#)]
9. Yang, M.S.; Benjamin, J.B. Sparse possibilistic c-means clustering with Lasso. *Pattern Recognit.* **2023**, *138*, 109348. [[CrossRef](#)]
10. Bickel, S.; Scheffer, T. Multi-view clustering. In Proceedings of the 4th IEEE International Conference on Data Mining, Brighton, UK, 1–4 November 2004; pp. 19–26.
11. Yang, M.S.; Hussain, I. Unsupervised multi-view K-means clustering algorithm. *IEEE Access* **2023**, *11*, 13574–13593. [[CrossRef](#)]
12. Hussain, I.; Sinaga, K.P.; Yang, M.S. Unsupervised multiview fuzzy c-means clustering algorithm. *Electronics* **2023**, *12*, 4467. [[CrossRef](#)]
13. Shi, S.; Nie, F.; Wang, R.; Li, X. Multi-view clustering via nonnegative and orthogonal graph reconstruction. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 201–214. [[CrossRef](#)]
14. Yu, W.; Xing, L.; Nie, F.; Li, X. Multi-View Fuzzy Clustering Based on Anchor Graph. *IEEE Trans. Fuzzy Syst.* **2023**, *32*, 755–766. [[CrossRef](#)]
15. Wang, R.; Wang, P.; Wu, D.; Sun, Z.; Nie, F.; Li, X. Multi-view and multi-order structured graph learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 14437–14448. [[CrossRef](#)]
16. Yang, B.; Zhang, X.; Li, Z.; Nie, F.; Wang, F. Efficient multi-view K-means clustering with multiple anchor graphs. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 6887–6900. [[CrossRef](#)]
17. Zhao, M.; Yang, W.; Nie, F. Auto-weighted orthogonal and nonnegative graph reconstruction for multi-view clustering. *Inf. Sci.* **2023**, *632*, 324–339. [[CrossRef](#)]
18. Yang, M.; Li, Y.; Hu, P.; Bai, J.; Lv, J.; Peng, X. Robust multi-view clustering with incomplete information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1055–1069. [[CrossRef](#)]
19. Lin, Y.; Gou, Y.; Liu, Z.; Li, B.; Lv, J.; Peng, X. Completer: Incomplete multi-view clustering via contrastive prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 11174–11183.
20. Cai, Y.; Che, H.; Pan, B.; Leung, M.F.; Liu, C.; Wen, S. Projected cross-view learning for unbalanced incomplete multi-view clustering. *Inf. Fusion* **2024**, *105*, 102245. [[CrossRef](#)]
21. Chao, G.; Jiang, Y.; Chu, D. Incomplete Contrastive Multi-View Clustering with High-Confidence Guiding. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 11221–11229.
22. Zhan, J.; Cai, M. A cost-minimized two-stage three-way dynamic consensus mechanism for social network-large scale group decision-making: Utilizing K-nearest neighbors for incomplete fuzzy preference relations. *Expert Syst. Appl.* **2024**, *263*, 125705. [[CrossRef](#)]
23. Liu, Z.; Song, P. Deep low-rank tensor embedding for, multi-view subspace clustering. *Expert Syst. Appl.* **2024**, *237*, 121518. [[CrossRef](#)]
24. Li, J.; Gao, Q.; Wang, Q.; Yang, M.; Xia, W. Orthogonal non-negative tensor factorization based multi-view clustering. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 18186–18202.
25. Chen, M.S.; Li, Z.Y.; Lin, J.Q.; Wang, C.D.; Huang, D. Sparse Graph Tensor Learning for Multi-View Spectral Clustering. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *8*, 3534–3543. [[CrossRef](#)]
26. Benjamin, J.B.; Yang, M.S. Tensor-Based Possibilistic C-Means Clustering. *IEEE Trans. Fuzzy Syst.* **2024**, *32*, 5939–5950. [[CrossRef](#)]
27. Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; He, L. Multi-level feature learning for contrastive multi-view clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 16051–16060.
28. Zhang, Y.; Wu, J.; Cai, Z.; Philip, S.Y. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Trans. Multimed.* **2020**, *22*, 2844–2857. [[CrossRef](#)]
29. Tang, C.; Zheng, X.; Liu, X.; Zhang, W.; Zhang, J.; Xiong, J.; Wang, L. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Trans. Knowl. Data Eng.* **2021**, *34*, 4705–4716. [[CrossRef](#)]
30. Xu, Y.M.; Wang, C.D.; Lai, J.H. Weighted multi-view clustering with feature selection. *Pattern Recognit.* **2016**, *53*, 25–35. [[CrossRef](#)]
31. 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](#)] [[PubMed](#)]
32. 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](#)]
33. Yang, M.S.; Sinaga, K.P. A feature-reduction multi-view k-means clustering algorithm. *IEEE Access* **2019**, *7*, 114472–114486. [[CrossRef](#)]
34. Yang, M.S.; Sinaga, K.P. Collaborative feature-weighted multi-view fuzzy c-means clustering. *Pattern Recognit.* **2021**, *119*, 108064. [[CrossRef](#)]
35. Benjamin, J.B.M.; Yang, M.S. Weighted multiview possibilistic c-means clustering with L2 regularization. *IEEE Trans. Fuzzy Syst.* **2021**, *30*, 1357–1370. [[CrossRef](#)]

36. Khan, G.A.; Khan, J.; Anwar, T.; Ashraf, Z.; Javed, M.H.; Diallo, B. Weighted Concept Factorization based Incomplete Multi-view Clustering. *IEEE Trans. Artif. Intell.* **2024**, *5*, 5699–5708. [[CrossRef](#)]
37. Zhou, K.; Zhu, Y.; Guo, M.; Jiang, M. MvWECM: Multi-view Weighted Evidential C-Means clustering. *Pattern Recognit.* **2024**, *159*, 111108. [[CrossRef](#)]
38. Liu, Z.; Huang, H.; Letchmunan, S.; Deveci, M. Adaptive weighted multi-view evidential clustering with feature preference. *Knowl.-Based Syst.* **2024**, *294*, 111770. [[CrossRef](#)]
39. Ouadfel, S.; Abd Elaziz, M. A multi-objective gradient optimizer approach-based weighted multi-view clustering. *Eng. Appl. Artif. Intell.* **2021**, *106*, 104480. [[CrossRef](#)]
40. Houfar, K.; Samai, D.; Dornaika, F.; Benlamoudi, A.; Bensid, K.; Taleb-Ahmed, A. Automatically weighted binary multi-view clustering via deep initialization (AW-BMVC). *Pattern Recognit.* **2023**, *137*, 109281. [[CrossRef](#)]
41. Liu, Z.; Huang, H.; Letchmunan, S. weighted multi-view evidential clustering. In *International Conference on Artificial Neural Networks*; Springer Nature: Cham, Switzerland, 2023; pp. 265–277.
42. Bandyopadhyay, S.; Saha, S. A point symmetry-based clustering technique for automatic evolution of clusters. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 1441–1457. [[CrossRef](#)]
43. Fu, J.; Hsiao, C.A. Decoding intelligence via symmetry and asymmetry. *Sci. Rep.* **2024**, *14*, 12525. [[CrossRef](#)] [[PubMed](#)]
44. 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.
45. Fowlkes, E.B.; Mallows, C.L. A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* **1983**, *78*, 553–569. [[CrossRef](#)]
46. Rand, W.M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.* **1971**, *66*, 846–850. [[CrossRef](#)]
47. Cover, T.M. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 1999.
48. 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](#)]
49. Fei-Fei, L.; Fergus, R.; Perona, P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004; p. 178.
50. Dua, D.; Graff, C. UCI Machine Learning Repository. 2017, Volume 7, p. 62. Available online: <http://archive.ics.uci.edu/ml> (accessed on 8 December 2024).
51. Xu, P.; Deng, Z.; Choi, K.S.; Cao, L.; Wang, S. Multi-view information-theoretic co-clustering for co-occurrence data. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 379–386.
52. Tao, H.; Hou, C.; Qian, Y.; Zhu, J.; Yi, D. Latent complete row space recovery for multi-view subspace clustering. *IEEE Trans. Image Process.* **2020**, *29*, 8083–8096. [[CrossRef](#)]
53. Ma, Z.; Nie, F.; Yang, Y.; Uijlings, J.R.; Sebe, N.; Hauptmann, A.G. Discriminating joint feature analysis for multimedia data understanding. *IEEE Trans. Multimed.* **2012**, *14*, 1662–1672. [[CrossRef](#)]
54. Ning, H.; Xu, W.; Gong, Y.; Huang, T. Discriminative learning of visual words for 3D human pose estimation. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
55. Yang, Y.; Xu, D.; Nie, F.; Yan, S.; Zhuang, Y. Image clustering using local discriminant models and global integration. *IEEE Trans. Image Process.* **2010**, *19*, 2761–2773. [[CrossRef](#)] [[PubMed](#)]
56. Winn, J.; Jojic, N. Locus: Learning object classes with unsupervised segmentation. In Proceedings of the Tenth IEEE International Conference on Computer Vision, ICCV'05, Washington, DC, USA, 17–20 October 2005; Volume 1, pp. 756–763.

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.