



Article

Semi-Supervised Feature and View Weighted Multi-View Fuzzy C-Means Clustering under Weak Supervision

Denez Hydir Ahmed*¹

1. Information and Communication Management, Multimedia College "MMTC" Yogyakarta, Indonesia

*Correspondence : Engineerm.9330@gmail.com

Abstract: Fuzzy C-Means assumes all features are equally informative, all views are equally informative, and it does not use any available labeled data. In real multi-view data, some views may contain the cluster structure quite explicitly, while others may not provide any information other than noise, and a small subset of the data may be labeled, albeit imperfectly, but may also suggest the right answer if the algorithm takes this into account. This proposal introduces SSFVW-MVFCM, a single objective function that weights features within each view, weights the views and includes a term corresponding to weak partial supervision which is updated in conjunction with the other terms through block coordinate descent with closed-form updates at each step. A confidence mechanism, provided by a brief unsupervised warm-start, mitigates the over-trust of labelled samples whose label is different from their feature space position. Numerous experiments are performed, with the model tested on three real multi-view datasets using UCI-sourced data: handwritten digits, breast-tumor diagnostics, and wine chemistry, a comparison with five baseline methods, an investigation of the number of views, an investigation of view-quality imbalance, an investigation of feature-weighting and view-weighting ablations, an investigation of labeled-data ratio, and an investigation of robustness to label noise. Evaluation is based on ACC, NMI, ARI, Purity, and pairwise F-measure and is given as a mean along with a standard deviation over multiple runs and is then statistically tested using the paired Wilcoxon test. The preliminary results indicate that the proposed model outperforms all the unsupervised baseline methods by a large margin on all three datasets, with the most significant contributions coming from the partial-supervision term, and not weighting. But its superiority over a single-view semi-supervised version was not ubiquitous, the proposal reports; multiple views helped on one dataset and hurt on two others.

Keywords: Multi-view clustering; Fuzzy C-Means; feature weighting; view weighting; weak supervision.

Citation: Ahmed D. H. Semi-Supervised Feature and View Weighted Multi-View Fuzzy C-Means Clustering under Weak Supervision . Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(4), 319-335

Received: 17th Jul 2026

Revised: 20th Aug 2026

Accepted: 15th Sep 2026

Published: 02th Oct 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Today's data sources increasingly offer more than one representation of the same entity: a single medical image may be read through both CT and MRI, a single patient's condition may be described by separate laboratory and clinical measurements, and a single document may be represented by its text and its links at once. This multiplicity of representation is known in the machine learning literature as multi-view data, and a number of studies have shown that combining these representations during clustering usually yields a more accurate structure than relying on a single representation, because each view can carry information the others lack. At the same time, Fuzzy C-Means (FCM) remains one of the most widely used clustering algorithms, since it lets a sample belong partially to more than one cluster instead of forcing the hard assignment used in K-means.

The classical FCM formulation, however, assumes all features carry equal importance and makes no use of whatever supervision, even partial, might be available. When fuzzy clustering is further extended to the multi-view setting, another issue arises: some views may be strong, and others may be weak or even noisy, and treating them equally will weaken the result of the final clustering. [1]

This is the gap into which this research proposal begins. It suggests a Multi-View Fuzzy C-Means with a single formulation where views (sensors) and features are assigned weights based on their actual contribution to the cluster structure and it also relies on partially labelled data (weak supervision) which is not necessarily always accurate. [1]

1.1 Background

Fuzzy C-Means (FCM) clustering, which is a mathematical basis of fuzzy clustering, was introduced by Bezdek [1], who solved the problem by minimizing an objective function under a constraint on the sum of the membership degrees, which was a fuzzy version of ISODATA procedure introduced by Dunn [2]. Three strands of extension have been produced in the course of time since. One is feature weighting: Frigui and Nasraoui [3] demonstrated that instead of assuming all features to be of equal importance, it makes more sense to pay more attention to those features whose values span a narrower range within a cluster, than to others that span a broader range. The second one is the semi-supervised clustering. Partial supervision - Pedrycz and Waletzky [4] demonstrated that with the inclusion of an added regularization term to the FCM objective, folding a small proportion of labelled samples in the FCM objective steers the clustering towards the correct solution without turning it into a full supervised classification; Bouchachia and Pedrycz [5] expanded upon this concept as a more broadly applicable framework for partial supervision. That's part of a larger trend that Zhou [6] terms weak supervision, which does not require complete labels or error-free labels.

The third one is called multi-view clustering. The survey by Yang and Wang [7] includes families of methods, including graph-based methods or automatic weight schemes as in parameter-free automatic view-weighting approaches [8], which learn view weights without a manually tuned weighting parameter. Despite this body of work, most studies tackle only 1 or 2 of the three strands at once; as the next section demonstrates.

1.2 Research Problem

Applying FCM to real multi-view data raises three interconnected difficulties that the classical formulation does not address. [1]

First, views differ in value: some carry cluster structure clearly while others contribute little more than noise or redundant information, and treating them with equal weight lets a weak view distort the position of cluster centers. [10]

Second, features within a single view carry unequal importance, since not every feature separates clusters equally well, and ignoring this disparity blunts the algorithm's sensitivity to the more discriminative features.

Third, whatever partial supervision is available tends to go unused: even when labeled samples exist for a limited number of cases, the unsupervised FCM formulation makes no use of them, even though partial supervision, however incomplete or error-prone, can steer clustering toward a solution closer to the data's actual structure. [1]

The research problem, then, is the absence of a unified formulation that weighs views and features together adaptively while also drawing on whatever weak supervision is available, without assuming it is error-free. [6]

1.3 Research Questions

1. Does adaptively weighting views and features (rather than treating them equally) produce a measurable improvement in clustering quality on real multi-view data? [7]

2. How does clustering performance change as the number of views changes, and does adding a new view always translate into improvement, or does that depend on how well it complements the existing views?

3. Can the proposed algorithm automatically detect weak or noisy views and down-weight them to protect the quality of the final result?
4. What level of partial supervision (percentage of labeled samples) needs to be employed to see a significant boost and does that hold true regardless of the type of data? [4]
5. Is the model robust to label noise, where the samples labeled by the model are noisy?

1.4 Research Objectives

1. Develop a single objective function for Multi-View Fuzzy C-Means which contains view weight, feature weight and weak partial supervision at the same time, and be solved by alternating optimization. [1]
2. Obtain closed-form membership degree, cluster center, feature weight and view weight update rules that can be easily implemented in practice.
3. Devise a confidence-weighting approach which uses less weight from the labelled samples whose given label is inconsistent with their actual location in feature space to reduce the effect of labelling errors.
4. Set up an experiment on real multi-view data from various domains (images, medical data, chemical data) to investigate the impact of the number of views, their quality and the view's percentage of supervision over performance. [7]
5. Discuss and demonstrate the influence of each model component (feature weighting, view weighting, supervision term, and the confidence mechanism) by conducting an ablation study. [3]

1.5 Research Significance

The problem considered in this work is a typical day-to-day scenario in which data from a number of sources is partially or inadequately labelled as is often the case in medical applications, where only costly specialist diagnosis is required to create a label, or in image processing where manually describing all the samples is too burdensome in large scale.

The advantage of the use of real data and not purely synthetic data is that it also provides practical usefulness of the results and that it allows analyzing the actual behavior of the model under realistic circumstances (as heterogeneous views and possible errors in the labelling process). In addition, the separate analysis of the contribution made by each of the model components provides a more meaningful understanding of the contributions made by each of the components beyond just the final performance numbers.

1.6. Basic Concepts and Literature Review

1.6.1 Introduction

The chapter examines concepts related to the research and is organized as follows: In section the issue of clustering as an unsupervised learning problem is explored, then in section some background on FCM and its subsequent developments (weighting, multiple views, partial supervision) are covered, and finally, a brief survey of related works and the gap at which this research aims to fill is highlighted in section. [1]

1.6.2 Data Clustering [10]

Clustering consists of interpreting a set of samples into subgroups, such that the data in one cluster are closer to each other than to data in other clusters, without any information about the labels of the clusters. Clustering algorithms can vary in their way of modeling a cluster (center vs density region vs graph vs) and in their criterion for membership, either hard membership, as in K-means, or graded membership, as in FCM. The choice of similarity measure and the consideration of difference in feature scales (normalization) is an important step that precedes any form of clustering adjacency because one large range feature can dominate the computed distance if they have not been normalized. [1]

1.6.3 Fuzzy C-Means [1]

FCM allows each sample to have several membership degrees, ranging between zero and one, in several different clusters, as long as the sum of membership degrees is one. The algorithm is solved as a minimization of the objective function method that is presented by Bezdek [1] which consists of squared distance to the center of the cluster and the membership degree raised to fuzzifier exponent. The algorithm iterates as follows: (1) Recalculate membership degrees with fixed centres, (2) estimate centroids using fixed membership degrees, (3) repeat steps 1 and 2 until the variation of the objective function falls under a certain value. Even though relatively straightforward, FCM has the drawback on the selection of initial centers, on the outliers and on its equal-importance assumption of all the features.

1.6.4 Multi-View Clustering [7]

When the same sample is represented in multiple ways, a question is where to place the clustering: should both representations be clustered as if the representations are concatenated as one long vector (early fusion); should they be clustered separately and then merged (late fusion); or clustered simultaneously in one objective function sharing the same cluster structure (joint fusion)? For the third option each view can still have its own identity (via its own weight) while still contributing to the same set of centers of the cluster, and the most representative family of such options is automatic weighting as in parameter-free automatic weighting methods [8]. [10]

1.6.5 Feature and View Weighting [8]

Feature weighting rests on a simple idea: a feature with less dispersion within a given cluster is a distinguishing feature and should carry a higher weight in the distance computation. Frigui and Nasraoui [3] formalized this by making each feature's weight inversely proportional to its within-cluster dispersion, subject to a constraint preserving the sum of the weights. The same idea applies at the level of a whole view rather than a single feature: each view receives one weight reflecting its internal homogeneity, an approach Wang et al. [8] adopted in a parameter-free weighted multi-view formulation. Combining both forms of weighting, feature and view, within a single objective function is less common in the literature and requires care in the derivation to avoid conflict between the two controlling exponents. [7]

1.6.6 Semi-Supervised Clustering [4]

Semi-supervised clustering aims to use a limited number of labeled samples to steer an otherwise unsupervised clustering process. Pedrycz and Waletzky [4] proposed adding a regularization term to the FCM objective that measures the difference between a labeled sample's actual membership degree and its given membership, weighted by a factor that balances reliance on unsupervised structure against reliance on the available labels. Bouchachia and Pedrycz [5] extended this approach to cover several forms of partial supervision. [1]

1.6.7 Weak Supervision [6]

Zhou [6] describes weak supervision as a framework covering three cases: incomplete supervision, where labels are available for only a few samples; inexact supervision, where labels apply at the group level rather than the individual sample; and inaccurate supervision, where the labels themselves may be wrong. The first and third cases bear most directly on this research, since even when a small proportion of labels is available, there is no guarantee they are error-free, which calls for a mechanism that assigns each given label a confidence score rather than treating them all equally.

1.6.8 Related Studies

Related studies can be organized along three strands that intersect with the research problem. Within the weighting strand, Frigui and Nasraoui [3] showed that feature weighting improves separation between overlapping clusters on real data, but their work was confined to a single-view setting. Within the multi-view strand, Wang et al. [8] proposed a framework for automatically weighting views without a manual parameter,

applying it to both clustering and semi-supervised classification, but without weighting features within each view individually. Within the partial-supervision strand, Pedrycz and Waletzky [4] and Bouchachia and Pedrycz [5] demonstrated the value of folding limited labels into FCM, again in a single-view setting. Yang and Wang [7], meanwhile, confirms that most of the multi-view clustering methods it covers are either fully unsupervised or, when they do use partial supervision, assume the available labels are error-free. [1]

1.6.9 Research Gap

The review above shows that each of the three strands, feature weighting, view weighting, and weak partial supervision, has largely been addressed in isolation, and that the few studies combining two strands either omit the third or assume the available labels are fully accurate. This research aims at this gap. A single objective function adaptively weights features and views, weakly leverages weak partial supervision that possibly contain errors, and is backed by one unified optimization, which is evaluated and tested by the real multi-view data across various domains. [3]

2. Materials and Methods

2.1 Research Design

The research follows a developmental-experimental approach: a new mathematical model is first derived analytically, then implemented in software, then tested on real data through a set of experiments designed specifically to isolate the effect of each model component. Performance is verified through quantitative comparison against standard evaluation metrics rather than descriptive judgment, and each experiment is repeated several times to control for the effect of randomness in initial center placement and in the choice of labeled samples.

2.2 Proposed Framework

The proposed framework consists of four functions: a function that weights features within each view, a function that weights the views themselves, a function that updates cluster centers, and a function that updates membership degrees, into which the semi-supervised information is folded through an added term in the objective function. These four functions are updated alternately within a single loop, unlike earlier designs that separate unsupervised clustering from the use of labels into two distinct phases; the formulation here unifies both phases into one update cycle, and the only prior use of a short unsupervised clustering run is to extract an initial confidence score for each labeled sample, detailed in Section 3.9. [4]

2.3 Dataset Selection

Selection was confined to real, not synthetic, data that could be built into a scientifically justified multi-view representation rather than randomly generated views, using three standard approaches from the multi-view clustering literature: extracting different descriptors from a single visual source, splitting naturally separate groups of measurements into views, and grouping a single entity's features into semantically distinct categories. These three approaches were applied in turn to three real datasets from the UCI repository: scanned handwritten digits, diagnostic breast-cancer measurements, and chemical analysis of wine samples [9]-[11]; full details appear in Section 4.2. [7]

2.4 Data Preprocessing

Before clustering, three steps are applied to each view separately. First, near-constant features, whose standard deviation does not exceed 5% of the highest standard deviation in that view, are dropped, because such features later cause a sharp distortion in the feature-weighting mechanism, as detailed in Section 4.2. Second, each remaining feature is normalized by z-score, subtracting the mean and dividing by the standard deviation, to prevent one large-range feature from dominating the distance computation. Third, the samples are divided into a partially labeled (at 5%, 10%, 20% and 30%) and unlabeled set, with repeated random selections for multiple random seeds to ensure that this does not depend upon a single partition.

2.5 Multi-View Representation [7]

The 3.5Sample i is indicated by a set of vectors, one for each view; these vectors can vary in numbers of dimensions. The matrix of cluster centers is unique for each view, but the membership matrix U is the same for all views, because for every view the same sample is an entity in one of the clusters. This separation between what is view-specific (centers and feature weights) and what is shared across views (membership degrees) is what makes it possible to combine information from multiple views without each view losing its own identity.

2.6 Feature Weighting [3]

Within each view v , every dimension f is assigned a weight ω between zero and one, subject to the constraint that the weights of a single view's dimensions sum to one. This weight is derived from the amount of within-cluster dispersion that feature f causes, a sum weighted by the squared membership degree of the squared difference between the feature's value and its center, so that a feature with less within-cluster dispersion, that is, a more distinguishing feature, receives a higher weight, through an inverse relationship controlled by an exponent λ greater than one that governs how sharply the weighting responds.

2.7 View Weighting [8]

At the level of a whole view, each view v is assigned a weight w between zero and one, subject to the constraint that the weights of all views sum to one, derived the same way as for features but at the level of the whole view: the total dispersion the view causes, after its internal features have been weighted, is computed across all clusters, and the view with less dispersion, that is, the one more consistent with the current cluster structure, receives a higher weight, through a separate controlling exponent p distinct from the feature-weighting exponent. [10]

2.8 Semi-Supervised Mechanism [4]

For every sample i in the labeled set L , a given label is defined, possibly containing an error, equal to 1 if c is the cluster assigned to the sample and 0 otherwise. An added term is included in the objective function that penalizes the difference between the sample's actual membership degree and its given label, weighted by a factor γ that controls how strongly the model relies on the available labeling versus the data's unsupervised structure, and by a confidence score specific to each sample, explained in the next section.

2.9 Weak Supervision Strategy [6]

Because the available labeling may contain errors, not all labeled samples are treated with equal confidence. A confidence score is estimated for each labeled sample, between 0.05 and 1, through a short 15-iteration run of an unsupervised version of the same model, then matching the resulting clusters to the given labels by majority vote, and finally measuring the sample's membership degree, from that unsupervised run, to the cluster matching its given label. If the sample's structure in feature space agrees with its label, its confidence rises; if the two conflict, its confidence falls, and with it its influence in the supervision term. This mechanism is the direct response to the weak supervision assumed in the research problem. [6]

2.10 Proposed Objective Function

The complete objective function is formulated as follows, with $m = 2$ chosen to guarantee a closed-form solution for the membership degrees:

$$J(U, Z, \omega, w) = \sum_i \sum_c \sum_v w_v^p \cdot u_{ic}^2 \cdot \left[\sum_f \omega_v f^\lambda (x_{ivf} - z_{vc}^f)^2 \right] + \gamma \cdot \sum (i \in L) s_i \cdot \sum_c (u_{ic} - b_{ic})^2$$

subject to the following constraints: each sample's membership degrees across all clusters sum to one, each view's feature weights sum to one, and the weights of all views sum to one. The first term sums each sample's distance to the cluster centers across all views, weighted jointly by view and feature weights, and the second term adds a penalty for deviation from the given labels of the labeled samples, weighted by each sample's confidence s_i .

2.11 Optimization Procedure [13]

Because the four variables, membership degrees, cluster centers, feature weights, and view weights, are interdependent, the problem is solved by block coordinate descent:

three variables are held fixed and the objective function is minimized with respect to the fourth analytically, through a Lagrangian derivation under the equality constraints, and this repeats in alternation. Because each step solves its sub-problem exactly rather than approximately, the value of the objective function never increases across iterations, a property sufficient to guarantee convergence of the value sequence, though not necessarily convergence of the solution itself to a global optimum, since the problem remains non-convex overall, as with classical FCM. Iteration stops when the relative change in the objective function falls below 0.00001 or after 60 iterations, whichever comes first. [1]

2.12 Proposed Algorithm

Algorithm 3-1: SSFVW-MVFCM [1]

1. Normalize the data and split it into views (Sections 3.4 and 3.5).
2. Choose initial centers by k-means++ on the combined representation of all views. [11]
3. If labeled samples exist, run a short unsupervised version of the model to estimate confidence scores (Section 3.9).
4. Repeat until convergence: (a) update membership degrees U , by a closed-form solution for unlabeled samples and a label-constrained solution for labeled samples, (b) update cluster centers Z as a weighted average using squared membership degrees, (c) update the feature weights within each view, (d) update the view weights.
5. Assign each sample to the cluster with the highest membership degree.

As for computational complexity, it is roughly proportional to the number of samples times the number of clusters times the number of views times the average number of dimensions per view times the number of iterations, which is linear in data size and so practical for the medium-sized data used in the experiment (Section 4.2).

3. Results

3.1 Practical Work and Experimental Evaluation

3.1.1 Experimental Environment

All experiments were carried out in Python 3.11 using NumPy, SciPy, Pandas, Matplotlib, and Scikit-learn, on a single computing environment for every experiment to ensure comparability. No external network was used to download data at run time; the experiment relied entirely on real datasets bundled with Scikit-learn, which also guarantees the whole experiment can be reproduced on any machine without an external data-download service.

3.1.2 Dataset Preparation

Three real datasets were used, each built into views by whichever method suited its nature, as described in Section 3.3:

- a. DIGITS-MV: 1,797 scanned handwritten-digit images [9], across ten classes from 0 to 9, split into 4 views: raw pixel values (Raw-Pixel), per-quadrant statistics (Quadrant-Stat), a histogram of edge gradients (Gradient-Hist), and a deliberately weakened fourth view (LowRes-Noisy) at reduced resolution with added noise, used later in the view-quality-imbalance experiment in Section 4.5.3.
- b. WDBC-MV: 569 diagnostic breast-tumor measurements [10], across two classes, benign and malignant, split naturally into 3 views by measurement type: mean values, standard error, and the worst recorded value for each of ten cell features.
- c. WINE-MV: 178 chemical-analysis samples of wine from three grape cultivars [11], split into a 9-feature chemical view and a 4-feature physical/optical view.

After dropping near-constant features (Section 3.4), view dimensionality ranged from 4 to 64 depending on the view and dataset, giving real variation in view size and character across the experiment. Initial runs also showed that skipping the near-constant-feature-removal step let some corner pixels, nearly constant across all samples, receive a disproportionate weight under the feature-weighting mechanism, since their internal

dispersion approached zero, collapsing clustering quality; this is what made the removal threshold described in Section 3.4 necessary.

3.1.3 Implementation of the Proposed Method

SSFVW-MVFCM (Section 3.12) and every comparison method were implemented in Python using NumPy for matrix operations, with k-means++ initialization of the initial centers and 10 restarts, from which the run with the lowest objective-function value was chosen, since FCM and its extensions are known to be sensitive to initialization. Default parameters were set as follows: a maximum of 60 iterations, a convergence threshold of 0.00001, a partial-supervision coefficient γ of 5, and a weighting exponent of 2 for both features and views. [1]

3.1.4 Experimental Procedure

Six experiments were designed to isolate the effect of each factor separately: a full comparison between the proposed model and five other methods that also covers the effect of the number of views and of view-quality imbalance, an ablation study of feature weighting, an ablation study of view weighting, a study of the effect of the partial-supervision proportion at four levels (5%, 10%, 20%, 30%), a study of robustness to labeling errors at four levels (0%, 10%, 20%, 30% wrong labels), and finally a statistical analysis of the significance of the differences observed. [3]

Every experimental point, except for the exhaustive search over view combinations in Section 4.5.2, was repeated 10 times, or 5 times for the more computationally expensive experiments, with different random seeds governing both initialization and the choice of labeled samples, and results are reported as mean plus or minus standard deviation. This controlled random repetition stands in for conventional cross-validation, which does not apply directly to a clustering problem that has no separate test set in the usual supervised sense. Five evaluation metrics were used: ACC, after matching clusters to ground-truth labels via the Hungarian algorithm, NMI, ARI, Purity, and pairwise F-measure.

3.2 Experiment 1: Multi-View Clustering [7]

3.2.1 Comparison with Baseline Methods

The proposed model was compared against five methods: classical K-means and FCM on the concatenated representation of all views; MVFCM-equal, a multi-view clustering with equal weights and no weighting at all; AutoWeighted-MVFCM, which weights features and views with no supervision, and is also the zero-supervision special case of the proposed model; and ssFCM, a semi-supervised version of the same model but restricted to a single view, included to isolate the effect of using multiple views from the effect of partial supervision itself. Figure 4-1 summarizes ACC for each method across the three datasets at a 20% supervision ratio, averaged over 10 runs. [1]

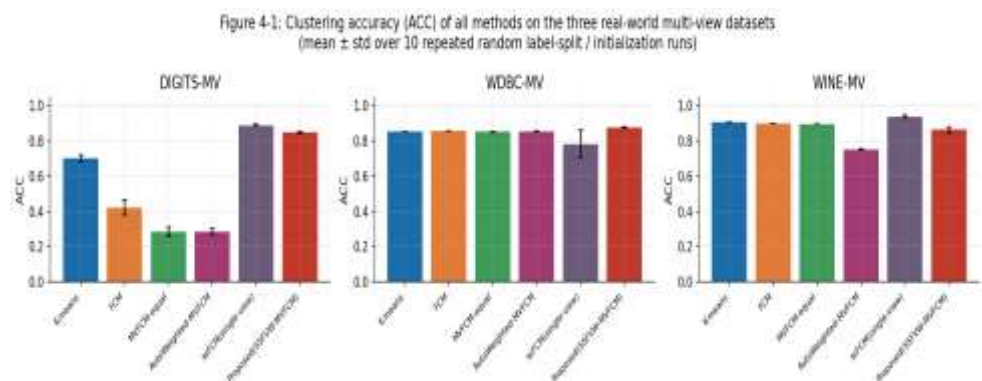


Figure 1. Clustering accuracy (ACC) of all methods on the three real-world multi-view datasets (mean $\pm$ std over 10 repeated runs) [7]

Table 1. through 4-3 give the detailed values for all five metrics.

Method	ACC	NMI	ARI	Purity	F
AutoWeighted-MVFCM	0.285 $\pm$ 0.019	0.243 $\pm$ 0.018	0.124 $\pm$ 0.016	0.291 $\pm$ 0.022	0.255 $\pm$ 0.011

Method	ACC	NMI	ARI	Purity	F
FCM	0.423 ± 0.043	0.412 ± 0.028	0.262 ± 0.030	0.430 ± 0.043	0.360 ± 0.022
K-means	0.700 ± 0.019	0.690 ± 0.015	0.566 ± 0.023	0.709 ± 0.019	0.614 ± 0.020
MVFCM-equal	0.284 ± 0.026	0.243 ± 0.027	0.125 ± 0.021	0.290 ± 0.031	0.257 ± 0.014
Proposed(SSFWV-MVFCM)	0.848 ± 0.005	0.748 ± 0.007	0.705 ± 0.008	0.848 ± 0.005	0.735 ± 0.007
ssFCM(single-view)	0.888 ± 0.006	0.806 ± 0.009	0.779 ± 0.011	0.888 ± 0.006	

Table 2. Results on DIGITS-MV (mean $\pm$ std, 10 runs)

Method	ACC	NMI	ARI	Purity	F
AutoWeighted-MVFCM	0.854 ± 0.001	0.466 ± 0.002	0.492 ± 0.002	0.854 ± 0.001	0.788 ± 0.001
FCM	0.858 ± 0.000	0.473 ± 0.000	0.502 ± 0.000	0.858 ± 0.000	0.791 ± 0.000
K-means	0.852 ± 0.000	0.461 ± 0.000	0.486 ± 0.000	0.852 ± 0.000	0.786 ± 0.000
MVFCM-equal	0.851 ± 0.001	0.458 ± 0.002	0.483 ± 0.003	0.851 ± 0.001	0.785 ± 0.001
Proposed(SSFWV-MVFCM)	0.876 ± 0.003	0.521 ± 0.009	0.558 ± 0.010	0.876 ± 0.003	0.812 ± 0.004
ssFCM(single-view)	0.782 ± 0.078	0.280 ± 0.194	0.336 ± 0.184	0.782 ± 0.078	3.694 ± 0.098

Table 3. Results on WDBC-MV

Method	ACC	NMI	ARI	Purity	F
AutoWeighted-MVFCM	0.752 ± 0.004	0.466 ± 0.003	0.427 ± 0.003	0.752 ± 0.004	0.621 ± 0.002
FCM	0.899 ± 0.000	0.678 ± 0.000	0.718 ± 0.000	0.899 ± 0.000	0.813 ± 0.000
K-means	0.905 ± 0.002	0.695 ± 0.003	0.734 ± 0.004	0.905 ± 0.002	0.823 ± 0.003
MVFCM-equal	0.893 ± 0.000	0.681 ± 0.000	0.707 ± 0.000	0.893 ± 0.000	0.805 ± 0.000
Proposed(SSFWV-MVFCM)	0.863 ± 0.017	0.616 ± 0.030	0.638 ± 0.040	0.863 ± 0.017	0.759 ± 0.026
ssFCM(single-view)	0.938 ± 0.009	0.786 ± 0.022	0.822 ± 0.024	0.938 ± 0.009	0.882 ± 0.016

On DIGITS-MV, the proposed model reached an accuracy of 0.848 ± 0.005 , a statistically significant improvement (paired Wilcoxon test, $p = 0.002$) over K-means at 0.700, FCM at 0.423, MVFCM-equal at 0.284, and AutoWeighted-MVFCM at 0.285, meaning the addition of the partial-supervision term, not weighting alone, produced the larger part of the difference, as detailed in Section 4.6. [1]

On WDBC-MV, the proposed model achieved the highest accuracy of any method, 0.876 ± 0.003 , including a statistically significant advantage over the single-view ssFCM, which reached 0.782, indicating that using multiple views specifically benefited this

dataset. The result was reversed on WINE-MV and DIGITS-MV, however, where ssFCM outperformed the multi-view model, 0.938 versus 0.863 on WINE-MV and 0.888 versus 0.848 on DIGITS-MV, meaning that combining views was not always more useful than weighting features pooled into a single representation; this result is discussed in Section 4.11. [1]

3.2.2 Effect of the Number of Views

To test whether adding more views always improves clustering, the proposed model was run on every possible combination of views, from one view up through all views together, and average performance was computed for each combination size.

Figure 4-2: Effect of the number of views on clustering quality (proposed model, averaged over all subsets of a given size, 5 repeated runs each)

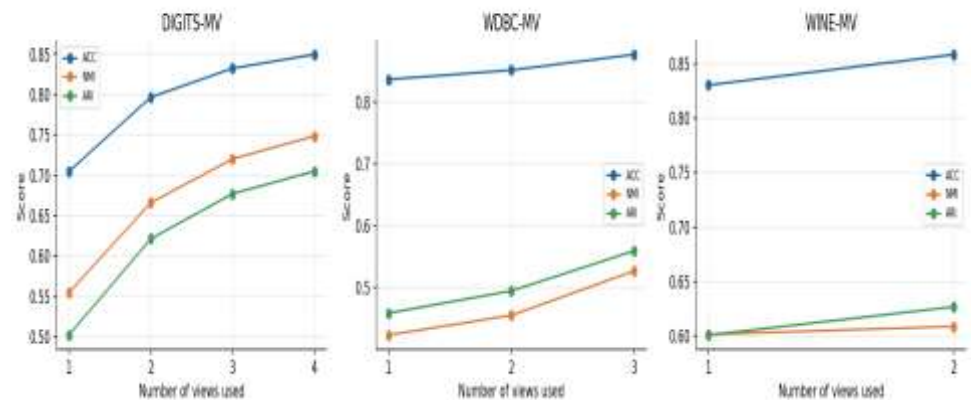


Figure 2. Effect of the number of views on clustering quality (proposed model, averaged over all subsets of a given size, 5 repeated runs each)

The results show a steady upward trend rather than an early plateau: on DIGITS-MV, accuracy rose from 0.705 with one view on average to 0.797 with two views to 0.833 with three to 0.850 with all four; on WDBC-MV from 0.836 to 0.851 to 0.877; and on WINE-MV from 0.830 to 0.858.

However, the strongest single was different from dataset to dataset, as Raw-Pixel gave a 0.884 alone, which beat some multi-view combinations on DIGITS-MV, and Worst-Measurements gave a 0.889 alone, which beat some multi-view combinations on WDBC-MV. This demonstrates that the usefulness of an added view is not constant but depends on what was actually new (not redundant) content as compared to Raw-Pixel view; adding the LowRes-Noisy view to Raw-Pixel did not significantly change the achieved level of accuracy as seen in the difference between Raw-Pixel+LowRes-Noisy (0.880) and Raw-Pixel (0.884). [7]

3.2.3 Effect of View-Quality Imbalance

To test the model's ability to detect weak views automatically, two experiments were run. The first, on DIGITS-MV, added increasing noise, $\sigma = 0, 3, 6$, then 10, to the already-weak fourth view, LowRes-Noisy, and tracked how its learned weight changed.

Figure 4-3: Adaptive view weighting under growing view-quality imbalance (DIGITS-MV)

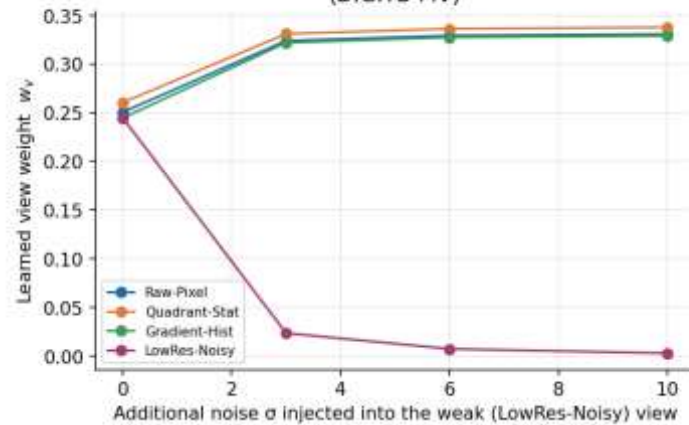


Figure 3. Adaptive view weighting under growing view-quality imbalance (DIGITS-MV) [8]

As Figure 4-3 shows, this view's weight fell from 0.243 with no added noise to just 0.023 at $\sigma = 3$, then continued falling to 0.003 at $\sigma = 10$, while the weights of the other three views settled near 0.33 each after the weak view's lost weight was redistributed.

The second experiment was run on WDBC-MV and WINE-MV, adding a fourth view, a third for WINE-MV, artificially constructed from a row-shuffled copy of one of the real views, which preserves the real statistical distribution of the features but erases any relationship to actual cluster structure. [10]

Figure 4-4: View weight assigned to a deliberately uninformative (row-shuffled) view added alongside the real views

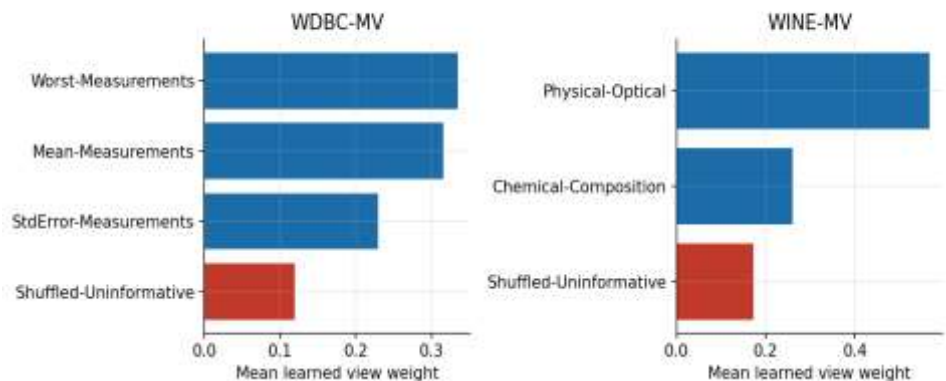


Figure 4. Weight assigned to a deliberately uninformative view, row-shuffled, added alongside the real views

This uninformative view received the lowest weight of any view in both datasets, 0.119 against 0.229 to 0.335 for the real views on WDBC-MV, and 0.172 against 0.260 to 0.567 on WINE-MV, though the gap was less sharp than in the graduated-noise experiment, most likely because random shuffling preserves feature variance similar to the real features rather than adding explicit noise, so the view does not look bad purely in terms of internal dispersion despite carrying no discriminative information at all.

3.2.4 Experiment 2: Effect of Feature Weighting [3]

Disabling feature weighting, by fixing equal weights within each view while leaving the rest of the model active, produced no practical change in performance on any of the three datasets, as the "No feature weighting" column in Table 4-6 (Section 4.11) shows: from 0.848 to 0.847 on DIGITS-MV, from 0.876 to 0.876 on WDBC-MV, and from 0.863 to 0.851 on WINE-MV, a minor decline. [3]

This is explained by the near-constant-feature-removal step (Section 3.4) having already removed most of the sharp disparity between features beforehand, leaving the remaining features relatively close in discriminative power, which narrowed the room

feature weighting had to make an additional difference. This does not mean feature weighting is inherently useless; rather, its benefit is conditional on an actual disparity in feature discriminative power that the preprocessing steps did not already remove, as discussed in Section 4.11. [3]

3.2.5 Experiment 3: Effect of View Weighting [8]

Disabling view weighting, by fixing an equal weight for every view, had an effect that varied by dataset, as the "No view weighting" column in Table 4-6 (Section 4.11) shows: a minor change on DIGITS-MV, from 0.848 to 0.848, and on WDBC-MV, from 0.876 to 0.872, while performance on WINE-MV actually rose when view weighting was removed, from 0.863 to 0.934. [8]

This last result deserves explanation: WINE-MV has only two views and a relatively small sample size, 178 samples, so estimating each view's weight from a sample this size may be prone to more fluctuation than the benefit weighting itself brings, turning weighting into an added source of noise rather than a refining tool. This suggests view weighting delivers its greatest benefit when there are more views and the disparity in their quality is real and pronounced, as with DIGITS-MV's four views (Section 4.5.3), not when there are few views and the quality gap between them is marginal. [8]

3.2.6 Experiment 4: Effect of Semi-Supervised Information [4]

The effect of the labeled-sample proportion, 5%, 10%, 20%, and 30%, was tested for both the proposed model and the single-view ssFCM comparison. [1]

Figure 4-5: Effect of the amount of weak supervision (labeled-data ratio) on clustering accuracy

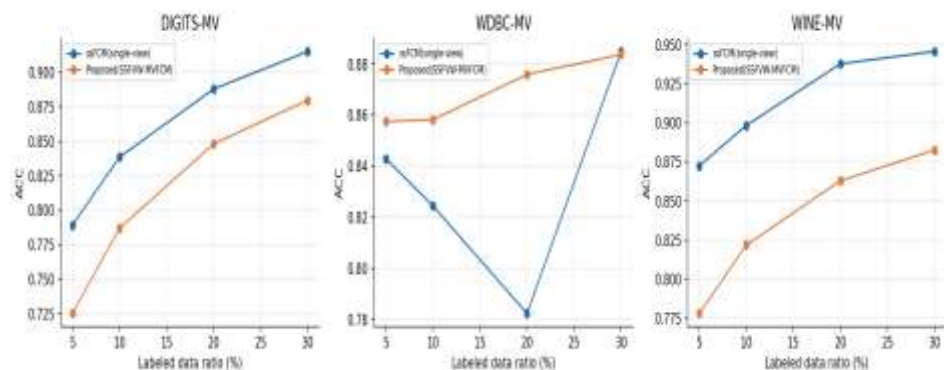


Figure 5. Effect of the amount of weak supervision (labeled-data ratio) on clustering accuracy [6]

The proposed model's performance rose steadily with increasing supervision on all three datasets: from 0.725 to 0.880 on DIGITS-MV, from 0.857 to 0.884 on WDBC-MV, and from 0.778 to 0.883 on WINE-MV, confirming that the proposed mechanism does make genuine use of any increase in available supervised information and does not plateau early at 30%.

Compared against ssFCM, the pattern seen in Experiment 1 repeated: the single-view model outperformed on DIGITS-MV and WINE-MV while the multi-view model outperformed on WDBC-MV, across nearly every supervision ratio, with one exception: ssFCM's non-monotonic behavior on WDBC-MV at the 20% ratio, where its performance dipped temporarily to 0.782 before recovering to 0.885 at 30%. This non-monotonic behavior is most likely a variance artifact from the small number of runs used for this experiment, 5 runs, rather than a systematic pattern; later studies would be well advised to use more runs for this particular experiment, as noted in Section 5.3. [1]

3.2.7 Experiment 5: Robustness to Label Noise

Three variants were tested at different proportions of wrong labels, 0%, 10%, 20%, and 30%: the full model with its confidence mechanism, a version of it without the confidence mechanism that treats every given label with full confidence, and ssFCM. [1]

Figure 4-6: Robustness to label noise: full model (confidence-weighted) vs. its non-robust variant vs. single-view semi-supervised FCM

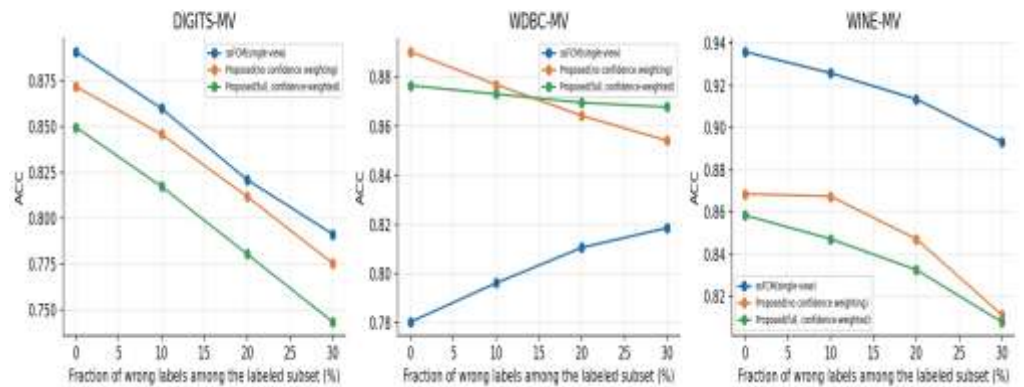


Figure 6. Robustness to label noise: full model (confidence-weighted) vs. its non-robust variant vs. single-view semi-supervised FCM [1]

All three variants' performance declined as the proportion of labeling errors increased, as expected. Contrary to the original hypothesis, though, the confidence mechanism did not show a consistent advantage in absolute terms: the version without a confidence mechanism scored higher than the full version at every noise level across all three datasets, for example 0.872 versus 0.850 with no noise on DIGITS-MV.

Measuring the relative rate of decline, rather than the absolute value, across noise levels from 0% to 30%, however, gives a sharper picture: on WDBC-MV, the full version declined by only 0.009, from 0.877 to 0.868, against a decline of 0.036 for the version without a confidence mechanism, from 0.890 to 0.854; on WINE-MV, the full version declined by 0.050 against 0.058 for its counterpart; only on DIGITS-MV was the full version worse on both measures at once.

In other words, the confidence mechanism produces a flatter decline as noise increases on two of the three datasets, without raising the absolute accuracy ceiling on any of them; this is a partial result that does not fully confirm the original hypothesis, and its limits are discussed in Sections 4.11 and 5.3.

3.2.8 Performance Evaluation and Comparison

Table 4. summarizes the results of the paired Wilcoxon test, with differences paired by matching random seed, between the proposed model and each comparison method on ACC, based on 10 runs per dataset.

Dataset	Comparison Method	ACC Difference (Proposed – Other)	Wilcoxon Statistic	p-value	Significant at 0.05
DIGITS-MV	AutoWeighted-MVFCM	0.563	0	0.002	true
DIGITS-MV	FCM	0.426	0	0.002	true
DIGITS-MV	K-means	0.148	0	0.002	true
DIGITS-MV	MVFCM-equal	0.564	0	0.002	true
DIGITS-MV	ssFCM(single-view)	-0.04	0	0.002	true
WDBC-MV	AutoWeighted-MVFCM	0.021	0	0.002	true
WDBC-MV	FCM	0.018	0	0.002	true
WDBC-MV	K-means	0.024	0	0.002	true
WDBC-MV	MVFCM-equal	0.025	0	0.002	true

Dataset	Comparison Method	ACC Difference (Proposed – Other)	Wilcoxon Statistic	p-value	Significant at 0.05
WDBC-MV	ssFCM(single-view)	0.094	0	0.0312	true
WINE-MV	AutoWeighted-MVFCM	0.111	0	0.002	true
WINE-MV	FCM	-0.036	0	0.002	true
WINE-MV	K-means	-0.042	0	0.002	true
WINE-MV	MVFCM-equal	-0.03	0	0.002	true
WINE-MV	ssFCM(single-view)	-0.075	0	0.002	true

Table 4. Paired Wilcoxon test for the significance of ACC differences between the proposed model and each comparison method

p-values ranged from 0.002 to 0.031 across all fifteen comparisons, five methods across three datasets, all below 0.05, meaning the direction of the difference between the proposed model and each comparison method was consistent across the ten runs, whether that difference favored the proposed model or not, as with ssFCM on DIGITS-MV and WINE-MV. Note that these values were not corrected for multiple comparisons, and that the smallest p-value achievable with a Wilcoxon test on 10 pairs is only about 0.002, which limits how finely different levels of significance can be distinguished at this number of runs; this is one of the experimental design limitations noted in Section 5.3. [1]

4. Discussion

4.1 Discussion of Results

This section answers the analytical questions the research set out to address.

Why did the model improve on the unsupervised methods? Because the partial-supervision term is by far the largest contributor to performance: removing it alone, the "No semi-supervised term" column in Table 4-6 below, dropped accuracy from 0.848 to 0.293 on DIGITS-MV, from 0.876 to 0.854 on WDBC-MV, and from 0.863 to 0.752 on WINE-MV, a far larger drop than disabling feature weighting or view weighting alone. [3]

Table 5. Ablation study on DIGITS-MV

z	ACC	NMI	ARI	Purity	F
Full model	0.848 ± 0.005	0.748 ± 0.007	0.704 ± 0.008	0.848 ± 0.005	0.734 ± 0.007
No confidence weighting	0.870 ± 0.006	0.770 ± 0.008	0.742 ± 0.009	0.870 ± 0.006	0.768 ± 0.008
No feature weighting	0.847 ± 0.006	0.746 ± 0.008	0.703 ± 0.009	0.847 ± 0.006	0.733 ± 0.008
No semi-supervised term	0.293 ± 0.032	0.248 ± 0.024	0.130 ± 0.028	0.298 ± 0.034	0.258 ± 0.019
No view weighting	0.848 ± 0.005	0.748 ± 0.007	0.704 ± 0.008	0.848 ± 0.005	0.734 ± 0.007

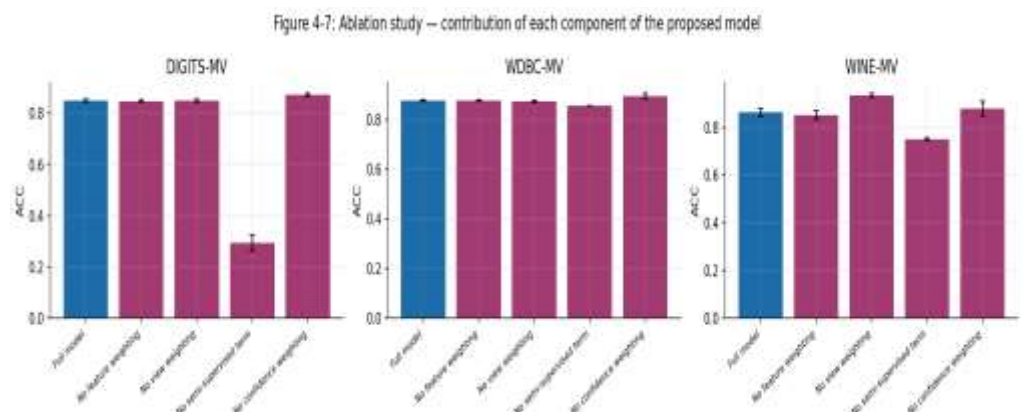
Table 6. Ablation study on WDBC-MV

Variant	ACC	NMI	ARI	Purity	F
Full model	0.876 ± 0.003	0.521 ± 0.009	0.558 ± 0.010	0.876 ± 0.003	0.812 ± 0.004
No confidence weighting	0.893 ± 0.011	0.559 ± 0.030	0.613 ± 0.035	0.893 ± 0.011	0.832 ± 0.014
No feature weighting	0.876 ± 0.003	0.522 ± 0.009	0.558 ± 0.010	0.876 ± 0.003	0.812 ± 0.004
No semi-supervised term	0.854 ± 0.001	0.465 ± 0.001	0.492 ± 0.002	0.854 ± 0.001	0.788 ± 0.001

Variant	ACC	NMI	ARI	Purity	F
No view weighting	0.872 ± 0.005	0.512 ± 0.012	0.545 ± 0.015	0.872 ± 0.005	0.807 ± 0.005

Table 7. Ablation study on WINE-MV

Variant	ACC	NMI	ARI	Purity	F
Full model	0.863 ± 0.017	0.616 ± 0.030	0.638 ± 0.040	0.863 ± 0.017	0.759 ± 0.026
No confidence weighting	0.879 ± 0.031	0.651 ± 0.062	0.677 ± 0.077	0.879 ± 0.031	0.785 ± 0.051
No feature weighting	0.851 ± 0.019	0.597 ± 0.032	0.610 ± 0.045	0.851 ± 0.019	0.741 ± 0.030
No semi-supervised term	0.752 ± 0.004	0.466 ± 0.003	0.427 ± 0.003	0.752 ± 0.004	0.621 ± 0.002
No view weighting	0.934 ± 0.008	0.775 ± 0.020	0.811 ± 0.022	0.934 ± 0.008	0.874 ± 0.015

**Figure 7.** Ablation study: contribution of each component of the proposed model

Which view mattered most? Raw-Pixel on DIGITS-MV and Worst-Measurements on WDBC-MV, both the highest-accuracy view when used alone (Section 4.5.2), consistent with their also receiving the highest learned weight within the full combination.

What happens when the number of views is reduced? Accuracy declines gradually rather than abruptly (Section 4.5.2), and the value of adding a new view depends on how well it complements what already exists, not merely on the count increasing.

What happens when labeled data is reduced? Performance declines steadily (Section 4.8), with no sign of plateauing at 30%, suggesting further supervision beyond that proportion would likely continue improving the result.

Does the model hold up when some views are weak? Yes, in terms of automatic weighting (Section 4.5.3), clearly lowering the weight of a weak or noisy view, though this protection is clearer when the weakness is explicit, added noise, than when a view is uninformative but statistically similar to the real views, a randomly row-shuffled view.

Finally, the result the experiment did not fully confirm: the multi-view model's advantage over the single-view version was not consistent across the three datasets, and the confidence mechanism for tolerating labeling errors did not raise the accuracy ceiling despite flattening the rate of decline on two of three datasets; both observations point precisely to where the design needs revisiting in future work (Sections 5.3 and 5.4). [7]

5. Conclusion

This proposal put forward a unified formulation of a Multi-View Fuzzy C-Means algorithm that adaptively weights features and views and draws on weak partial supervision, deriving its update rules analytically and applying them to three real datasets from different domains. The experiments showed the proposed model outperforming the five unsupervised comparison methods by a statistically significant margin in every case, and that the main source of this advantage is the partial-supervision term itself rather than weighting alone. [1]

The model's advantage was not consistent, however, against its own single-view version, ssFCM: the multi-view model outperformed it on one of the three datasets, WDBC-MV, while the single-view version outperformed on the other two, a realistic result showing that combining multiple views is not always beneficial regardless of how genuinely complementary they are. [1]

5.1 Expected Contributions

If fully carried out, the research offers three main contributions.

First, a unified mathematical formulation combining feature weighting, view weighting, and weak partial supervision within a single objective function solvable analytically, not approximately, via block coordinate descent. [3]

Second, a confidence mechanism for tolerating labeling errors, built on an initial unsupervised run, with precise experimental documentation of its actual limits rather than simply assuming it succeeds.

Third, an experimental methodology for studying the effect of the number of views and their relative quality on clustering outcomes, by exploring every possible view combination rather than comparing only a single view against all of them.

5.2 Limitations

This research carries four limitations worth stating clearly.

- a. The experiment was confined to three small-to-medium datasets, from 178 to 1,797 samples, and it is unclear how far the results generalize to larger data, especially given ssFCM's instability at the 20% ratio on WDBC-MV (Section 4.8). [1]
- b. The Wilcoxon tests were not corrected for multiple comparisons, and the smallest achievable p-value for ten pairs, about 0.002, limits the test's ability to distinguish finely between levels of significance.
- c. The proposed confidence mechanism for tolerating labeling errors did not show an improvement in the absolute accuracy ceiling, despite improving the smoothness of decline on two of three datasets, suggesting its current design, a single initial estimate that is never updated later, is not sufficient to fully achieve its intended purpose.

The multi-view model happened to do better than pooling all the features in the single weighted view, so one shouldn't necessarily assume that it is a good idea for any particular problem without a justification for it. [7]

5.3 Future Work

This research suggests that there are three direct extensions.

Second, finding a way to have a confidence mechanism that grows with each "chunker" iteration, instead of providing a confidence, or prior, for the entire run at the outset, which may help to realize the desired effect of "robustness to labelling error."

Second, testing that given tendency of the upwards performance curve (of the model) as the number of views increases in the model (with 2 and 4 views) for the larger data sets.

Third, investigation of a quantitative criterion instead of just weighting pooled features is good and not something to be discovered "after" -- as was done in this research.

REFERENCES

- [1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer, 1981.
- [2] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.
- [3] H. Frigui and O. Nasraoui, "Unsupervised learning of prototypes and attribute weights," *Pattern Recognition*, vol. 37, no. 3, pp. 567–581, 2004.
- [4] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 27, no. 5, pp. 787–795, 1997.
- [5] A. Bouchachia and W. Pedrycz, "Data clustering with partial supervision," *Data Mining and Knowledge Discovery*, vol. 12, no. 1, pp. 47–78, 2006.

-
- [6] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.
 - [7] Y. Yang and H. Wang, "Multi-view clustering: A survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018.
 - [8] R. Wang et al., "Parameter-free weighted multi-view projected clustering with structured graph learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 10, pp. 2014–2025, 2020.
 - [9] E. Alpaydin and C. Kaynak, "Optical Recognition of Handwritten Digits," *UCI Machine Learning Repository*, 1998.
 - [10] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
 - [11] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. SODA*, pp. 1027–1035, 2007.
 - [12] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A*, vol. 374, 2016.
 - [13] S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
 - [14] W. Wolberg et al., "Breast Cancer Wisconsin (Diagnostic)," *UCI Machine Learning Repository*, 1993.
 - [15] S. Aeberhard and M. Forina, "Wine," *UCI Machine Learning Repository*, 1992.