

Gene Expression Data Analysis Using Feature Weighted Robust Fuzzy c -Means Clustering

Vikas Singh¹, *Student Member, IEEE*, and Nishchal K. Verma², *Senior Member, IEEE*

Abstract—Clustering of gene expression data has been proven to be very useful in various applications, i.e., identifying the natural structure inherent in gene expression, understanding gene functions, mining relevant information from noisy data, and understanding gene regulation. In all these applications, genes, i.e., features, play a crucial role in characterizing them into different groups. These features may be relevant, irrelevant, or redundant, but they have different contributions during the clustering process. This paper presents a novel approach by considering the effect of features during the clustering process. In the proposed method, the fuzzy c -means the objective function is modified using a weighted Euclidean distance between the features with a monotonically decreasing function. The monotonically decreasing function helps control the features' contribution during the clustering process to partition the data into more relevant clusters. The proposed approach is validated, and performance is presented in various clustering performance measures on the different standard datasets. These clustering performance measures have also been compared with multiple state-of-the-art methods.

Index Terms— K -medoids, K -means, fuzzy c -means, alternate optimization, gene expression, features, cluster validation.

I. INTRODUCTION

CLUSTERING is an unsupervised learning method used to organize the unlabelled data sample into groups with respect to each data sample's underlying properties based on some similarity measures [1]–[4]. The data samples are unlabelled in the real application and have a different distribution. It is challenging to categorize them into meaningful clusters. Various clustering algorithms have been developed in the literature by considering these challenges. Hierarchical clustering (HC) organizes data into a hierarchical structure according to the proximity matrix, and a binary tree generally depicts results, or dendrogram [5]. The most successful application of HC is in the area of neuroimaging and bioinformatics [6]. In model-based clustering, data points in different clusters were generated by different probability distributions [7], and the number of clusters of a given dataset is identified by estimating the parameters using maximum likelihood (ML) or expectation-maximization (EM) [8]. The major flaw with these algorithms is that they have a slow convergence rate and are sensitive to the initial parameters [9].

Manuscript received 7 February 2020; revised 5 January 2021, 17 December 2021, and 16 February 2022; accepted 1 March 2022. Date of publication 8 March 2022; date of current version 30 December 2022. (Corresponding author: Vikas Singh.)

The authors are with the Department of Electrical Engineering, IIT, Kanpur 208016, India (e-mail: vikkyk@iitk.ac.in; nishchal@iitk.ac.in).

Digital Object Identifier 10.1109/TNB.2022.3157396

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

Partition-based clustering is well known for its simplicity and efficiency in large-scale clustering datasets, in which the dataset is divided into several subsets. K -medoids [10] and K -means (KM) [11] are one of the efficient partition-based algorithms that solve well-known clustering problem. The K -medoids are the most appropriate data point within a cluster, which have two main advantages: first, it has no limitations on the type of attributes; second, the medoids are chosen by predominant data points within a cluster and, therefore, it is less sensitive to the outliers. However, KM is the centroid-based method that works only for the numerical attributes and can be affected by a single outlier [12]. Ruspini [13], and Bezdek [14] have presented the fuzzy versions of KM algorithms, where each data point can belong to more than one cluster with a distinct membership degree. The fuzzy version of algorithms are widely used in real applications, where data generated from the system are uncertain, imprecise, and vague, and difficult to handle with KM and K -medoids because providing a hard partition of the data, whereas the fuzzy clustering-based method provides a soft partition of the data [15]–[19]. The major problem with these algorithms remains the same, i.e., all features are treated equally important during the clustering process and easily affected by the outliers, and difficult to find a meaningful cluster in the dataset.

In the literature, various feature weight clustering algorithms have been studied. Huang *et al.* [20] have presented a weighted KM (WKM) algorithm in which variable weights are multiplied with the dissimilarity measure. These variable weights are updated iteratively, and unimportant variables are eliminated by choosing variables with smaller weights. Jing *et al.* [21] modified the WKM by adding the entropy term in the objective function. The entropy term helps minimize the within-cluster dispersion and maximize the negative entropy on the current data partition in each iteration. In [22], authors have presented sparse KM by putting the L_1 constraint on the feature weights. The L_1 constraint helps to make some feature weight become zero. To improve the learning performance of the FCM, Wang *et al.* [23] have presented the weighted FCM (WFCM), in which variable weights are multiplied with the dissimilarity measure similar to the KM. The variable weights in WFCM are learned through the gradient descent technique by following the approach presented in [24]. In [25], the authors have presented two different feature-weighted FCM algorithms. The first approach has added the L_2 constraint on the feature weight; however, in the second approach, they have added the discriminant exponent on the feature weights. In [26], a new approach has been presented, which automatically computes individual feature weight and simultaneously

reduces the dataset's irrelevant features. Although these variants of feature-weighted clustering perform well, they are unable to find the optimal number of clusters with consistent performance in the data.

Clustering becomes a crucial tool for gene expression data analysis, which helps to unveil new cancer subtypes or to identify groups of genes that respond similarly to a specific experimental condition [27]. Several clustering algorithms have been studied for the gene expression dataset in the literature. In [28], Sherlock has analyzed the various clustering approaches for large-scale gene expression data and discussed the advantages and drawbacks of the algorithms. Smet, *et al.* [29] have presented an adaptive quality-based clustering of gene expression profiles to determine the actual number of clusters. In [30], authors have applied the fuzzy c -means clustering method to investigate trends in patient conditions, which doctors can use for disease diagnosis. Kannan, *et al.* [31] have proposed an effective fuzzy c -means by incorporating the membership function of fuzzy c -means for cancer subtypes in the gene expression cancer database. Herwig, *et al.* [32] have presented a sequential k -means algorithm with additional refinements that can handle high-throughput data in the order of hundreds of thousands of data items measured on hundreds of variables. In [33], a kernel-based clustering method for gene selection with gene expression data has been discussed in which weights are iteratively learned at the time of optimizing the clustering objective function. Handhayani and Hiryanto [34] have presented intelligent kernel k -means for clustering gene expression data without prior knowledge of clusters.

The variable weighting of features in gene expression data clustering helps identify the meaningful clusters with improved performance. It also overcomes the problem of an equal weighting of features in KM, K -medoids, and FCM based algorithms. The cluster-dependent features have two main advantages: First, they guide the clustering process to partition the data into more meaningful clusters. Second, they can be used in the subsequent steps of a learning system to improve their learning behavior. One more advantage of feature-dependent clustering is that the weighted feature simulates more dimensions to identify meaningful clusters when data is sparse and contains several outliers.

The rest of the paper is organized as follows: Section II presents the robust FCM with feature discrimination. Experimentation and comparisons with state-of-the-art methods are discussed in Section III. Finally, Section IV concludes the complete paper.

The major contributions are briefly summarized as follows:

- (a) Features of data play a crucial role in characterizing them into different groups. The effect of these features is incorporated in cluster identification by simultaneously weighting them in an unsupervised manner. As shown in (1), the objective function of FCM is modified by the monotonically decreasing function, which controls the weights of the features during the clustering process.
- (b) The weighted feature helps to eliminate the problem of trivial solutions during the clustering process since each feature in the clusters is multiplied by different weights based on the dispersion of the feature in the cluster.

- (c) The weighted feature helps to simulate more dimensions for identifying meaningful clusters when data is sparse and contains more outliers.

II. PROPOSED METHODOLOGY: ROBUST FCM WITH FEATURE DISCRIMINATION

The features in the data might be relevant or irrelevant, but they have a different contribution to the clustering process. Considering the effect of different features' contributions during the clustering process provides the best clustering performance with the optimal number of clusters in data. In this section, the fuzzy c -means objective function is modified using weighted Euclidean distance between the features with a monotonically decreasing function, which controls the contribution of the features during the clustering process. The monotonically decreasing function also overcome the trivial solutions during the clustering process since the feature weight is always finite. The proposed approach described below are as follows:

Let us consider a data matrix $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{p \times n}$, p and n are the number of features and number of samples respectively. Here, $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]' \in R^p$ is the i^{th} sample in the data matrix. To group the data matrix \mathbf{x} into c clusters, the following objective function can be minimized

$$F(U, V, W) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \sum_{k=1}^p w_{ik} d_{ijk}^2 + \sum_{i=1}^c \eta_i \sum_{k=1}^p (w_{ik} \log w_{ik} - w_{ik}) \quad (1)$$

subject to

$$\sum_{i=1}^c \mu_{ij} = 1, \quad 0 \leq \mu_{ij} \leq 1, \quad 1 \leq j \leq n$$

$$\text{and } \sum_{k=1}^p w_{ik} = 1, \quad 0 \leq w_{ik} \leq 1, \quad 1 \leq i \leq c \quad (2)$$

where, $d_{ijk} = (x_{jk} - v_{ik})^2$, is the dissimilarity measure between j^{th} sample and i^{th} cluster, x_{ik} is the value of k^{th} feature of j^{th} sample, v_{ik} is the value of k^{th} feature of i^{th} cluster, $U = [\mu_{ij}]$ is an $c \times n$, fuzzy partition matrix, μ_{ij} is the membership degree value of the i^{th} cluster of j^{th} sample, m is any real number greater than 1, $V = [v_{ik}]$ is an $c \times p$, matrix containing the cluster centers, $W = [w_{ik}]$ is an $c \times p$, weight matrix, w_{ik} is the weight value of k^{th} feature to i^{th} cluster, η_i is the input parameter used to control the feature weight.

The objective function, as proposed in (1), is a constrained nonlinear optimization problem whose solutions are unknown. The main objective is to minimize F with respect to U , V , and W using alternating optimization methods. In the alternating optimization, we first fix $U = \hat{U}$ and $W = \hat{W}$ and minimize F with respect to V . Then we fix $U = \hat{U}$ and $V = \hat{V}$ and minimize F with respect to W . Afterward, fix $V = \hat{V}$ and $W = \hat{W}$ and minimize F with respect to U .

TABLE I
PERFORMANCE MEASURE IN TERMS OF AR, RI, AND NMI (IN %)

Datasets		KM [11]	EWKM [21]	AFKM [35]	FCM [15]	SCAD1 [25]	SCAD2 [25]	FRFCM [26]	Proposed Approach
ALLML [36]	AR	65.28	70.28	72.22	66.67	68.06	66.67	85.33	86.11
	RI	55.91	57.86	59.31	54.93	55.91	54.93	65.46	75.74
	NMI	13.94	07.90	11.20	13.94	10.04	20.92	39.32	53.06
Ovarian cancer [37]	AR	55.73	55.73	55.73	64.81	71.30	57.41	77.50	79.63
	RI	50.46	50.46	50.46	54.18	58.88	50.87	65.00	67.41
	NMI	00.71	00.71	07.00	07.44	18.76	02.59	32.40	35.10
Colon cancer [36]	AR	53.23	56.20	53.23	54.80	54.84	54.80	59.70	61.29
	RI	49.39	50.30	49.39	49.70	49.66	49.70	51.50	51.77
	NMI	04.59	01.60	01.27	01.30	01.77	01.30	02.70	01.81
Seeds [38]	AR	89.43	58.76	89.52	89.52	89.52	89.52	86.20	89.52
	RI	87.38	54.38	87.44	87.44	87.44	87.44	84.20	87.53
	NMI	69.70	55.63	69.49	69.49	69.49	69.49	67.50	70.31
Iris [38]	AR	88.67	89.78	90.67	82.67	88.67	95.60	94.70	96.00
	RI	87.37	88.39	88.93	82.78	87.37	94.60	93.40	94.95
	NMI	74.19	77.80	72.44	65.22	74.19	85.60	83.30	86.42
Thyroid [38]	AR	80.93	49.30	70.70	85.58	85.58	73.95	86.50	87.44
	RI	74.53	53.00	64.73	79.32	79.33	66.93	79.40	80.39
	NMI	41.12	32.10	35.69	44.00	43.80	43.03	49.90	53.02
Glioma [36]	AR	48.00	64.00	62.00	56.00	54.00	32.00	62.00	68.00
	RI	70.12	74.37	73.96	70.94	72.65	27.02	74.66	76.00
	NMI	44.45	51.86	49.41	46.74	49.61	07.15	49.52	52.81
Lung cancer [36]	AR	49.75	62.44	51.72	55.56	59.61	59.80	71.88	73.89
	RI	61.09	64.11	58.98	51.48	56.67	59.16	64.32	66.29
	NMI	45.75	49.16	42.50	27.33	19.99	29.99	37.41	38.71
Lymphoma cancer [38]	AR	54.03	56.46	56.25	56.25	56.25	57.29	57.25	57.38
	RI	75.81	78.75	78.32	69.87	69.87	70.96	78.83	79.06
	NMI	62.12	62.61	61.81	45.72	45.72	46.83	62.61	69.98

First for the fixed U and W the objective $F(U, V, W)$ is minimized with respect to V as

$$\frac{\partial F(U, V, W)}{\partial v_{ik}} = -2 \sum_{j=1}^n \mu_{ij}^m w_{ik} (x_{jk} - v_{ik}) = 0 \quad (3)$$

From (3) the cluster center can be obtained using

$$v_{ik} = \frac{w_{ik} \sum_{j=1}^n \mu_{ij}^m x_{jk}}{w_{ik} \sum_{j=1}^n \mu_{ij}^m} \quad (4)$$

There are two cases arises depending on the value of w_{ik} .

Case 1: If $w_{ik} = 0$, the k^{th} feature is completely irrelevant relative to the i^{th} cluster. Hence, for any value of v_{ik} , the values of this feature will not contribute to the overall weighted distance computation. Therefore, in this case, any random value can be chosen for v_{ik} .

Case 2: If $w_{ik} \neq 0$, the k^{th} feature has some relevance to the i^{th} cluster, then the (4) is written as

$$v_{ik} = \frac{\sum_{j=1}^n \mu_{ij}^m x_{jk}}{\sum_{j=1}^n \mu_{ij}^m} \quad (5)$$

Then for given $V = \hat{V}$, the constraint optimization problem in (1) is changed into unconstrained minimization problem using Lagrangian multiplier technique as follows:

$$\begin{aligned} \hat{F}(U, W, \alpha, \delta) &= \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \sum_{k=1}^p w_{ik} d_{ijk}^2 \\ &+ \sum_{i=1}^c \eta_i \left(\sum_{k=1}^p (w_{ik} \log w_{ik} - w_{ik}) \right) \\ &- \sum_{j=1}^n \alpha_j \left(\sum_{i=1}^c \mu_{ij} - 1 \right) - \sum_{i=1}^c \delta_i \left(\sum_{k=1}^p w_{ik} - 1 \right) \end{aligned} \quad (6)$$

where, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ and $\delta = [\delta_1, \delta_2, \dots, \delta_k]$ are the vectors containing the Lagrangian multipliers. If the variables $(\hat{U}, \hat{W}, \hat{\alpha}, \hat{\delta})$ are the optimal values of $\hat{F}(U, W, \alpha, \delta)$, then the gradient with respect to these variable are vanishes.

$$\frac{\partial \hat{F}(U, W, \alpha, \delta)}{\partial \hat{\mu}_{ij}} = m \mu_{ij}^{m-1} \sum_{k=1}^p w_{ik} d_{ijk}^2 - \alpha_j = 0 \quad 1 \leq j \leq n, \quad 1 \leq i \leq c \quad (7)$$

$$\frac{\partial \hat{F}(U, W, \alpha, \delta)}{\partial \hat{w}_{ik}} = \sum_{j=1}^n \mu_{ij}^m d_{ijk}^2 + \eta_i \log w_{ik} - \delta_i = 0 \quad 1 \leq i \leq c, \quad 1 \leq k \leq p \quad (8)$$

$$\frac{\partial \hat{F}(U, W, \alpha, \delta)}{\partial \hat{\alpha}_j} = \sum_{i=1}^c \mu_{ij} - 1 = 0 \quad (9)$$

and,

$$\frac{\partial \hat{F}(U, W, \alpha, \delta)}{\partial \hat{\delta}_i} = \sum_{k=1}^p w_{ik} - 1 = 0 \quad (10)$$

the (7) and (8) can be simplified as

$$\mu_{ij} = \left(\frac{\alpha_j}{m \sum_{k=1}^p w_{ik} d_{ijk}^2} \right)^{\frac{1}{m-1}} \quad (11)$$

$$w_{ik} = \exp \left(\frac{-\sum_{j=1}^n \mu_{ij} d_{ijk}^2}{\eta_i} \right) \exp \left(\frac{\delta_i}{\eta_i} \right) \quad (12)$$

By substituting (11) in (9) we have

$$\sum_{i=1}^c \mu_{ij} = \sum_{i=1}^c \frac{\alpha_j}{m \sum_{k=1}^p w_{ik} d_{ijk}^2} = 1 \quad (13)$$

Similarly, (12) in (10)

$$\begin{aligned} \sum_{k=1}^p w_{ik} &= \sum_{k=1}^p \exp\left(\frac{-\sum_{j=1}^n \mu_{ij} d_{ijk}^2}{\eta_i}\right) \exp\left(\frac{\delta_i}{\eta_i}\right) = 1 \\ &= \exp\left(\frac{\delta_i}{\eta_i}\right) \sum_{k=1}^p \exp\left(\frac{-\sum_{j=1}^n \mu_{ij} d_{ijk}^2}{\eta_i}\right) = 1 \\ \exp\left(\frac{\delta_i}{\eta_i}\right) &= \frac{1}{\sum_{k=1}^p \exp\left(\frac{-\sum_{j=1}^n \mu_{ij} d_{ijk}^2}{\eta_i}\right)} \end{aligned} \quad (14)$$

By simplifying (13) and (11) we obtain the fuzzy partition, i.e., j^{th} sample to the i^{th} cluster as

$$\mu_{ij} = \frac{1}{\sum_{r=1}^c \left(\tilde{d}_{ij}^2 / \tilde{d}_{rj}^2\right)^{\frac{1}{m-1}}} \quad (15)$$

Similarly, from (14) and (12) we obtain the feature weight of k^{th} to i^{th} cluster as

$$w_{ik} = \frac{\exp\left(\frac{-D_{ik}}{\eta_i}\right)}{\sum_{q=1}^p \exp\left(\frac{-D_{iq}}{\eta_i}\right)} \quad (16)$$

where, $\tilde{d}_{ij}^2 = \sum_{k=1}^p w_{ik} d_{ijk}^2$ is the weighted Euclidean distance, $\sum_{r=1}^c \tilde{d}_{rj}^2$ is the sum of weighted Euclidean distance in all cluster, $D_{ik} = \sum_{j=1}^n \mu_{ij}^m d_{ijk}^2$ is the measure of dispersion of the i^{th} cluster with k^{th} feature, and $\sum_{q=1}^p \exp\left(\frac{-D_{iq}}{\eta_i}\right)$ is the measure of total dispersion of i^{th} cluster with all features in the data.

The choice of parameter η_i in (16) is the essential to the performance of the proposed approach since they reflect the importance of the second relative to the first term in (1). The control parameter η_i is used to control the feature weights, since, η_i is positive, and makes the weight w_{ik} as given in (16) is inversely proportional to total dispersion of i^{th} cluster to all features in data. The small of this term will make large w_{ik} , i.e., k^{th} features in i^{th} cluster are more important. If η_i in (16) is too large, the second term will dominate, and all features in cluster i will be relevant and assigned equal weights of $\frac{1}{p}$. The parameters η_i in (16) is computed by (17) in iteration, t , as follows

$$\eta_i = K \frac{\sum_{j=1}^n (\mu_{ij}^m)^{t-1} \sum_{k=1}^p (w_{ik})^{t-1} (d_{ijk}^2)^{t-1}}{\sum_{k=1}^p (w_{ik}^{t-1} - w_{ik}^{t-1} \log w_{ik}^{t-1})} \quad (17)$$

The parameter K is a constant, and the superscript $(t-1)$ denote their values in iteration $(t-1)$. The proposed approach is briefly summarized in Algorithm 1.

III. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the performance of robust FCM with feature discrimination on binary and multi-class gene expression datasets. The proposed approach is also investigated on other biological datasets, as shown in Table II. The proposed approach and state-of-the-art methods are validated on the same platform, and detailed experimentation are explained in the following subsections:

Algorithm 1 Robust FCM With Simultaneous Feature Discrimination

- 1: Give the number of cluster c , initial iteration number $t = 0$, randomly initialize cluster center $V^{(0)}$, randomly initialize feature weight $W^{(0)}$, and initialize η^0 randomly
- 2: **Repeat**
- 3: Update the feature weight matrix W using (16)
- 4: Update the partition matrix U using (15)
- 5: Update the cluster center matrix V using (5)
- 6: Update the parameter η_i using (17)
- 7: $t = t + 1$
- 8: **Until** The objective function obtains its local minimum value or centers stabilize

TABLE II
DATASET DESCRIPTIONS

Datasets	# of samples	# of features	# of classes
ALLAML [36]	72	7129	2
Ovarian cancer [39]	216	4000	2
Colon cancer [40]	62	2000	2
Seeds [38]	210	7	3
Iris [38]	150	4	3
Thyroid [38]	215	5	3
Glioma [36]	50	4434	4
Lung cancer [36]	203	3312	5
Lymphoma cancer [36]	96	4026	9

A. Performance Measure and Cluster Validation

The performance of the proposed approach is analysed in terms of various clustering performance measures such as accuracy rate (AR), rand index (RI) [41], Fuzzy rand index [42] and normalized mutual information (NMI) [43]. These performance measures are mathematically written as follows:

$$AR = \frac{1}{n} \sum_{j=1}^c n(c_j) \quad (18)$$

$$RI = \frac{f_1 + f_2}{f_1 + f_2 + f_3 + f_4} \quad (19)$$

$$NMI = \frac{I(X : Y)}{[H(X) + H(Y)]/2} \quad (20)$$

where, $n(c_j)$ is the number of data points correctly retrieved in cluster j , n is the total number of data points. The large value of AR represents better clustering performance. The RI index as defined in (19) is used to measure the similarity between the two clustering partitions. Let c is the number of true clusters, and c^* is the number of clusters obtained through the clustering algorithm. For a pair of points (x_i, x_j) , f_1 is the number of pairs of points that are the same in clusters c and c^* , f_2 is the number of pairs of points that are same in cluster c and different in cluster c^* , f_3 is the number of pairs of points that are different in cluster c , and same in cluster c^* , and f_4 is the number of pairs of points that are different in clusters c , and c^* . However, NMI is used to measure the information of a sample to contribute to making the correct classification decision, and its values always lie in between 0 and 1. As shown in Table I the performance of the proposed approach is better in the term of AR, RI, and NMI except for the colon and lung cancer dataset where NMI is small in comparison to state-of-the-art methods. Based

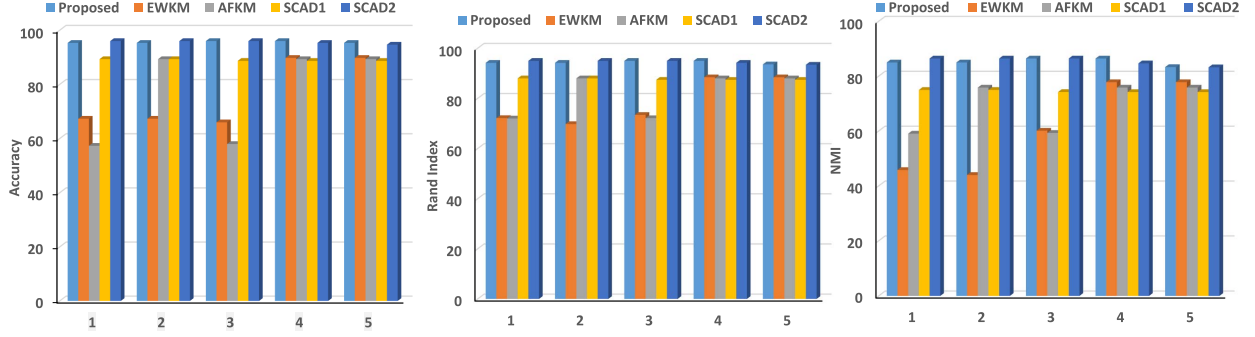


Fig. 1. Comparisons of AR, RI and NMI with EWKM, AFKM, SCAD1 and SCAD2 on the IRIS dataset under different parameter values.

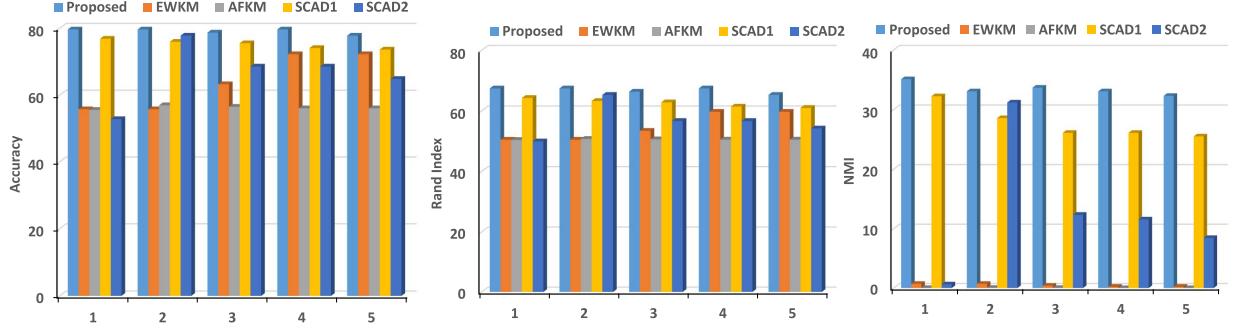


Fig. 2. Comparisons of AR, RI and NMI with EWKM, AFKM, SCAD1 and SCAD2 on the Ovarian cancer dataset under different parameter values.

on these clustering performance measures, we can assure that the proposed clustering process with feature discrimination provides the better partition of dataset either of datasets that are high-dimensional or sparse.

The above clustering performance measures are supervised, i.e., the number of cluster c is known. However, it is generally unknown in clustering. Therefore, to show the effectiveness we have used different cluster validity indices, i.e., Partition Coefficient (PC) [44], Xie and Beni's (XB) Index [45], Classification Entropy (CE) [46], and Dunn's Index (DI) [47]. These validity indices are mathematically written as follows:

$$PC = \frac{1}{n} \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^2 \quad (21)$$

$$CE = -\frac{1}{n} \sum_{j=1}^c \sum_{i=1}^n \mu_{ij} \log \mu_{ij} \quad (22)$$

$$XB = \frac{\sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^2 \|x_j - v_i\|^2}{n \min_{ij} \|x_j - v_i\|^2} \quad (23)$$

$$DI = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{i \in c} \{ \max_{x, y \in c} d(x, y) \}} \right\} \right\} \quad (24)$$

The value of PC always lies in between $[1/c, 1]$ if PC is closer to 1 represents the best partition, whereas closer to $1/c$, the partition becomes fuzzier. The CE is used to measure the fuzziness of the cluster partition, smaller the values of CE represents the optimal cluster. In XB, the numerator defines the compactness of the fuzzy partition, and denominator denotes the strength between clusters, smaller the values of XB move towards the optimal cluster. However, DI is used to measure the compactness in the well-separated clusters. The large value of DI represents better clustering.

The cluster validity index, i.e., PC, CE, XB, and DI, are computed and compared with state-of-the-art methods, as given in Table III. The cluster validity measure shows that the proposed approach can provide the optimal clustering result with improved clustering performance in terms of AR, RI, FRI and NMI.

B. Effect of Parameter Variation on Clustering Performance

In this subsection, we have investigated the effect of parameter variation on the performance of the proposed approach, along with state-of-the-art. The proposed approach is validated on small, high dimension and sparse datasets and compared with seven different state-of-the-art methods, i.e., K -means (KM) [11], entropy weighted K -means (EWKM) [21], Agglomerative fuzzy K -means (AFKM) [35], fuzzy c -means (FCM) [15], simultaneous clustering and attribute discrimination (SCAD1) [25], and SCAD2 [25] and a feature-reduction FCM (FRFCM) [26]. The KM algorithm gives equal importance to all features because it cannot discriminate between relevant and irrelevant features. This problem is eliminated by the EWKM, where different feature weight is assigned during the clustering. But the weight values are dependent on the initial cluster center, and if it changes, the algorithm will lead to different performances and be unable to converge. In the AFKM, the entropy help to find the optimal cluster, but the problem remains the same, i.e., all the features have equal importance during the clustering. FCM does the soft partition of the data. Still, it treats all features equally during the clustering process. In the SCAD, each feature has a different feature weight; however, the data partition is not consistent. In the proposed approach, the monotonically decreasing function controls the weights of the features

TABLE III
PERFORMANCE MEASURE IN TERMS OF PC, CE, XB AND DI

Datasets	True cluster	Validity indices	AFKM [35]	FCM [15]	SCAD1 [25]	SCAD2 [25]	FRFCM [26]	Proposed Approach
			Optimal cluster (values)	Optimal cluster (values)	Optimal cluster (values)	Optimal cluster (values)	Optimal cluster (values)	Optimal cluster (values)
ALLAML [36]	2	PC	2 (0.789)	2 (0.500)	2 (0.500)	2 (0.500)	2 (0.500)	2 (1.000)
		CE	2 (0.336)	2 (0.693)	2 (0.693)	2 (0.693)	2 (0.693)	2 (0.000)
		XB	2 (1.049)	2 (0.709)	2 (0.709)	2 (0.709)	2 (0.709)	2 (1.329)
		DI	2 (0.423)	2 (0.379)	2 (0.379)	2 (0.346)	2 (0.379)	2 (0.464)
Ovarian cancer [37]	2	PC	2 (0.915)	2 (0.779)	2 (0.773)	2 (0.500)	2 (0.873)	2 (0.978)
		CE	2 (0.143)	2 (0.365)	2 (0.374)	2 (0.693)	2 (0.073)	2 (0.038)
		XB	2 (2.583)	2 (2.072)	2 (2.087)	2 (2.755)	2 (2.643)	2 (3.195)
		DI	2 (0.118)	2 (0.118)	2 (0.118)	2 (0.093)	2 (0.118)	2 (0.112)
Colon cancer [40]	2	PC	2 (0.500)	2 (0.568)	2 (0.568)	2 (0.504)	2 (0.868)	2 (0.959)
		CE	2 (0.693)	2 (0.622)	2 (0.622)	2 (0.689)	2 (0.068)	2 (0.066)
		XB	2 (0.982)	2 (0.997)	2 (0.997)	2 (0.988)	2 (1.025)	2 (1.510)
		DI	2 (0.311)	2 (0.326)	2 (0.326)	2 (0.187)	2 (0.312)	2 (0.395)
Seeds [38]	3	PC	3 (0.571)	3 (0.726)	3 (0.974)	3 (0.686)	3 (0.938)	3 (0.976)
		CE	3 (0.677)	3 (0.499)	3 (0.044)	3 (0.564)	3 (0.032)	3 (0.042)
		XB	3 (3.902)	3 (3.880)	3 (6.942)	3 (3.725)	3 (6.652)	3 (7.865)
		DI	3 (0.056)	2 (0.085)	3 (0.086)	3 (0.047)	3 (0.078)	3 (0.089)
Iris [38]	3	PC	3 (0.798)	3 (0.783)	3 (0.783)	3 (0.832)	3 (0.883)	3 (0.991)
		CE	3 (0.299)	3 (0.395)	3 (0.395)	3 (0.311)	3 (0.251)	3 (0.017)
		XB	3 (3.614)	3 (4.231)	3 (4.230)	3 (4.248)	3 (4.330)	3 (4.679)
		DI	3 (0.030)	3 (0.105)	3 (0.105)	3 (0.091)	3 (0.084)	3 (0.064)
Thyroid [39]	3	PC	3 (0.797)	3 (0.649)	3 (0.754)	3 (0.648)	3 (0.934)	3 (0.978)
		CE	3 (0.322)	3 (0.618)	3 (0.394)	3 (0.613)	3 (0.021)	3 (0.038)
		XB	3 (4.114)	3 (2.892)	3 (5.442)	3 (3.934)	3 (7.832)	3 (8.088)
		DI	3 (0.238)	3 (0.024)	3 (0.020)	3 (0.018)	3 (0.022)	3 (0.017)
Glioma [36]	4	PC	4 (0.465)	4 (0.252)	4 (0.263)	4 (0.528)	4 (0.552)	4 (0.951)
		CE	4 (0.926)	4 (1.382)	4 (1.360)	4 (0.659)	4 (0.965)	4 (0.106)
		XB	4 (0.607)	4 (0.406)	4 (0.405)	4 (0.793)	4 (0.459)	4 (1.195)
		DI	4 (0.434)	4 (0.474)	4 (0.397)	4 (0.422)	4 (0.432)	4 (0.533)
Lung cancer [36]	5	PC	5 (0.308)	5 (0.200)	5 (0.200)	5 (0.200)	5 (0.569)	5 (0.779)
		CE	5 (1.309)	5 (1.609)	5 (1.609)	5 (1.609)	5 (1.609)	5 (0.424)
		XB	5 (0.419)	5 (0.292)	5 (0.292)	5 (0.292)	5 (0.785)	5 (0.992)
		DI	5 (0.431)	5 (0.342)	5 (0.364)	5 (0.386)	5 (0.356)	5 (0.462)
Lymphoma [36]	9	PC	9 (0.991)	9 (0.111)	9 (0.111)	9 (0.111)	9 (0.111)	9 (0.991)
		CE	9 (0.002)	9 (2.197)	9 (2.197)	9 (2.197)	9 (2.197)	9 (0.023)
		XB	9 (1.153)	9 (0.137)	9 (0.137)	9 (0.137)	9 (0.137)	9 (1.153)
		DI	9 (0.704)	9 (0.468)	9 (0.647)	9 (0.468)	9 (0.647)	9 (0.647)

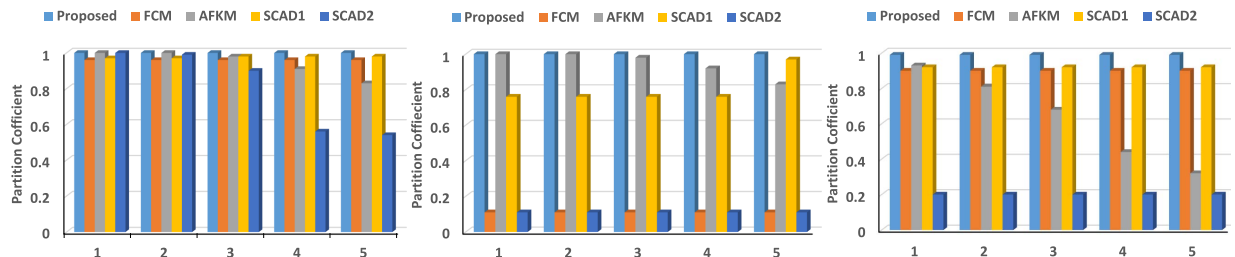


Fig. 3. Comparison of partition coefficient with FCM AFKM, SCAD1 and SCAD2 on ALLAML, lymphoma and lung dataset under different parameter.

TABLE IV
COMPUTATIONAL COMPLEXITY OF THE METHODS

Methodology	Computational Complexity
KM [11]	$O(nkm)$
EWKM [21]	$O(nkm^2)$
AFKM [35]	$O(nk^2m)$
FCM [15]	$O(nk^2m)$
SCAD1 [25]	$O(nk^2m + nkm^2)$
SCAD2 [25]	$O(nk^2m + nkm^2)$
FRFCM [26]	$O(nk^2m + nkm^2)$
Proposed Approach	$O(nk^2m + nkm^2)$

during the clustering and helps find the optimal number of the partition. The weight and fuzzy partition control parameters of the EWKM and AFKM are heuristics; for SCAD, weight control parameters are updated automatically in each iteration, but the fuzzy partition is not consistent. However, in FRFCM, the weight control parameters discard the unimportant features but are unable to preserve the high-dimension datasets' actual cluster. In the proposed approach, the weight control parameter is updated automatically in each iteration, and an optimal partition is obtained due to the relevant feature weight

assignment during the clustering process. The computational complexity of the proposed approach is also computed and compared with state-of-the-art methods, as shown in Table IV. In the proposed approach, the hyperparameter $m \in (1, 3]$ and variable $K \in [1, 10]$ are varied to find the best partition of the data. As shown in figure 1 and 2, for the IRIS and ovarian cancer dataset, AR, RI, and NMI values are plotted by varying the values of K for the fixed hyperparameter. For the IRIS and ovarian cancer dataset, as shown in figure 1 and 2, the AR, RI, and NMI are plotted for five different values K at $m = 2$. In the case of SCAD2, the weight exponent (q) is varied from $q \in [2, 10]$, however in the case of SCAD1, the variable $K \in [10, 100]$ and $m = 2$ for both the cases. Whereas, for the EWKM and AFKM, the variable is varied from $K \in [10, 100]$ and for the FRFCM, the variables are varied from $K \in [1, 10]$ with $m = 2$. Figure 1 and 2 showed that the performance of the proposed approach is better and stable in comparison to state-of-the-art methods. In figure 3 partition coefficient are also plotted for ALLAML, lymphoma, and lung dataset under different parameter values,

which shows that the proposed approach even able to provide the optimal cluster in the dataset.

IV. CONCLUSION

This paper has presented a novel clustering algorithm for clustering high-dimensional data with improved performance. In this approach, the objective function of FCM is changed by adding the monotonically decreasing function, which controls the weights of the features during the clustering process and helps in identifying the better clustering structure of the data. The major advantage of the proposed approach is that the clustering performance is consistent and insensitive to the initial cluster center due to different feature weights assigned to each cluster in the clustering process. The clustering performance is compared with various state-of-the-art methods, which shows that the proposed approach is a new clustering algorithm to partition the data with improved performance. In the future, the correlation between the features can be considered to minimize the effect of redundant features, and the proposed approach is extended for the mixed attributes.

REFERENCES

- [1] X. Z. Kong *et al.*, "Robust and efficient biomolecular clustering of tumor based on p -norm singular value decomposition," *IEEE Trans. Nanobiosci.*, vol. 16, no. 5, pp. 341–348, 2017.
- [2] R. Babuška, "Fuzzy clustering algorithms with applications to rule extraction," in *Fuzzy System Medical*. Heidelberg, Germany: Physica, 2000, pp. 139–173.
- [3] Y. Ling, X. Pan, G. Li, and X. Hu, "Clinical documents clustering based on medication/symptom names using multi-view nonnegative matrix factorization," *IEEE Trans. Nanobiosci.*, vol. 14, no. 5, pp. 500–504, Jul. 2015.
- [4] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.
- [5] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.
- [6] I. Cattinelli, G. Valentini, E. Palesu, and N. A. Borghese, "A novel approach to the problem of non-uniqueness of the solution in hierarchical clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1166–1173, Jul. 2013.
- [7] H. Zeng and Y.-M. Cheung, "Learning a mixture model for clustering with the completed likelihood minimum message length criterion," *Pattern Recognit.*, vol. 47, no. 5, pp. 2011–2030, 2014.
- [8] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
- [9] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *J. Amer. Stat. Assoc.*, vol. 97, no. 458, pp. 611–631, Jun. 2002.
- [10] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K -medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.
- [11] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Nov. 1967, vol. 1, no. 233, pp. 281–297.
- [12] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping Multidimensional Data*. Berlin, Germany: Springer, 2006, pp. 25–71.
- [13] E. H. Ruspini, "A new approach to clustering," *Inf. Control*, vol. 15, no. 1, pp. 22–32, 1969.
- [14] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.
- [15] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
- [16] P. D'Urso, L. De Giovanni, and R. Massari, "Garch-based robust clustering of time series," *Fuzzy Sets Syst.*, vol. 305, pp. 1–28, Oct. 2016.
- [17] B. Lafuente-Rego, P. D'Urso, and J. Vilar, "Robust fuzzy clustering based on quantile autocovariances," *Statistical*, vol. 61, pp. 2393–2448, Dec. 2018.
- [18] P. D'Urso, L. De Giovanni, and R. Massari, "Robust fuzzy clustering of multivariate time trajectories," *Int. J. Approx. Reasoning*, vol. 99, pp. 12–38, Aug. 2018.
- [19] P. D'Urso, L. De Giovanni, and R. Massari, "Trimmed fuzzy clustering of financial time series based on dynamic time warping," *Ann. Oper. Res.*, vol. 299, pp. 1379–1395, Apr. 2019.
- [20] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in K -means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657–668, May 2005.
- [21] L. Jing, M. K. Ng, and J. Z. Huang, "An entropy weighting K -means algorithm for subspace clustering of high-dimensional sparse data," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1026–1041, Aug. 2007.
- [22] D. M. Witten and R. Tibshirani, "A framework for feature selection in clustering," *J. Amer. Stat. Assoc.*, vol. 105, no. 490, pp. 713–726, 2010.
- [23] X. Wang, Y. Wang, and L. Wang, "Improving fuzzy C-means clustering based on feature-weight learning," *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1123–1132, Jul. 2004.
- [24] D. S. Yeung and X. Z. Wang, "Improving performance of similarity-based clustering by feature weight learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 556–561, Apr. 2002.
- [25] H. Frigui and O. Nasraoui, "Unsupervised learning of prototypes and attribute weights," *Pattern Recognit.*, vol. 37, no. 3, pp. 567–581, 2004.
- [26] M.-S. Yang and Y. Nataliani, "A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 817–835, Apr. 2018.
- [27] P. A. Jaskowiak, R. J. Campello, and I. G. Costa, "On the selection of appropriate distances for gene expression data clustering," *BMC Bioinf.*, vol. 15, p. S2, Jan. 2014.
- [28] G. Sherlock, "Analysis of large-scale gene expression data," *Current Opinion Immunol.*, vol. 12, no. 2, pp. 201–205, 2000.
- [29] F. De Smet, J. Mathys, K. Marchal, G. Thijs, B. De Moor, and Y. Moreau, "Adaptive quality-based clustering of gene expression profiles," *Bioinformatics*, vol. 18, no. 5, pp. 735–746, May 2002.
- [30] V. Kunwar, A. S. Sabitha, T. Choudhury, and A. Aggarwal, "Chronic kidney disease using fuzzy C-means clustering analysis," *Int. J. Bus. Anal.*, vol. 6, no. 3, pp. 43–64, Jul. 2019.
- [31] S. R. Kannan, R. Devi, S. Ramathilagam, and T. P. Hong, "Effective fuzzy possibilistic C-means: An analyzing cancer medical database," *Soft Comput.*, vol. 21, no. 11, pp. 2835–2845, Jun. 2017.
- [32] R. Herwig, A. J. Poustka, C. Müller, C. Bull, H. Lehrach, and J. O'Brien, "Large-scale clustering of cDNA-fingerprinting data," *Genome Res.*, vol. 9, no. 11, pp. 1093–1105, Nov. 1999.
- [33] H. Chen, Y. Zhang, and I. Gutman, "A kernel-based clustering method for gene selection with gene expression data," *J. Biomed. Informat.*, vol. 62, pp. 12–20, Aug. 2016.
- [34] T. Handhayani and L. Hiryanto, "Intelligent kernel K -means for clustering gene expression," *Proc. Comput. Sci.*, vol. 59, pp. 171–177, Feb. 2015.
- [35] M. J. Li, M. K. Ng, Y.-M. Cheung, and J. Z. Huang, "Agglomerative fuzzy K -means clustering algorithm with selection of number of clusters," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 11, pp. 1519–1534, Nov. 2008.
- [36] Accessed: Dec. 21, 2019. [Online]. Available: <http://featureselection.asu.edu/datasets.php>
- [37] Accessed: Dec. 21, 2019. [Online]. Available: <http://cilab.ujn.edu.cn/datasets.htm>
- [38] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [39] Accessed: Dec. 21, 2019. [Online]. Available: <https://in.mathworks.com/help/stats/sample-data-sets.html>
- [40] U. Alon *et al.*, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proc. Nat. Acad. Sci. USA*, vol. 96, no. 12, pp. 6745–6750, 1999.
- [41] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Statist. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
- [42] R. J. G. B. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, pp. 833–841, May 2007.
- [43] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 507–514.
- [44] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *J. Math. Biol.*, vol. 1, no. 1, pp. 57–71, May 1974.
- [45] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.
- [46] J. C. Bezdek, "Cluster validity with fuzzy sets," Taylor & Francis Group, 1973, pp. 58–73, vol. 3, no. 3.
- [47] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, 2008.