Scientia Iranica E (2021) 28(4), 2277–2293

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

A new validity index for fuzzy-possibilistic c-means clustering

M.H. Fazel Zarandi[a,*], S. Sotudian[a], and O. Castillo[b]

a. *Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran.*
b. *Tijuana Institutes of Technology, Tijuana, Mexico.*

Received 16 January 2018; received in revised form 6 September 2019; accepted 1 March 2021

KEYWORDS
Fuzzy-possibilistic clustering;
Cluster validity index;
Exponential separation;
Medical pattern recognition;
Microarray gene expression.

Abstract. In some complicated datasets, due to the existence of noisy data points and outliers, cluster validity indices can yield conflicting results in terms of determining the optimal number of clusters. This paper presents a new validity index for fuzzy-possibilistic C-means clustering called Fuzzy-Possibilistic (FP) index, which works well in the presence of clusters that vary in shape and density. Moreover, like most of the clustering algorithms, Fuzzy-Possibilistic C-Means (FPCM) is susceptible to some initial parameters. In this regard, in addition to the number of clusters, FPCM requires a priori selection of the degree of fuzziness (m) and the degree of typicality (η). Therefore, an efficient procedure was presented for determining optimal values of m and η. The proposed approach is evaluated using several synthetic and real-world datasets. Final computational results demonstrate the capabilities and reliability of the proposed approach compared with several well-known fuzzy validity indices in the literature. Furthermore, to clarify the ability of the proposed method in real applications, the proposed method is implemented in microarray gene expression data clustering and medical image segmentation.

1. Introduction

Clustering is an unsupervised pattern classification method that determines the intrinsic grouping in a set of unlabeled data. There are a large number of algorithms for clustering based on crisp [1], probabilistic [2], fuzzy [3], and possibilistic methods [4]. Hard-clustering methods limit each point of a dataset to exactly one cluster. However, since Zadeh introduced the notion of fuzzy sets which produced the idea of allowing for membership functions to exist in all clusters [5], fuzzy clustering has been extensively applied to various fields of science such as engineering and medical sciences [6–8].

In clustering algorithms, there are no predefined classes; as a result, we need to determine the optimal or near-optimal number of clusters before clustering. In this regard, compactness and separation are two measures of clustering assessment and selection of an optimal clustering scheme [9]. The closeness of cluster elements represents compactness, while isolation between clusters indicates separation.

So far, a considerable number of validity indices have been developed to evaluate the clustering quality (see Section 2). In these approaches, to find the optimal or near-optimal number of clusters, clustering algorithms should be executed several times for each

*. *Corresponding author. Tel.: +98 21 64545378*
E-mail addresses: zarandi@aut.ac.ir (M.H. Fazel Zarandi); shahab7290@aut.ac.ir (S. Sotudian); ocastillo@tectijuana.mx (O. Castillo)

doi: 10.24200/sci.2021.50287.1614

cluster number and its outputs be implemented into the cluster validity index in order to find the optimal or near-optimal number of clusters. Thus, to achieve an optimal prototype using a validity index, two conditions are unavoidable:

1. An algorithm that can find the best initial parameters of the clustering algorithm;
2. A validity function for assessing the worthiness of cluster schemes for various clusters.

Once these two necessities are met, the strategy of finding an optimal number of clusters becomes straightforward, i.e., determining the best initial parameters and using the validity function to choose the best number of clusters.

All clustering algorithms are susceptible to some initial parameters. For example, Fuzzy C-Means (FCM) may give various clustering results with different degrees of fuzziness. Therefore, even though the number of clusters is given, these algorithms may yield different results for the optimal number of clusters. In the current study, we use Fuzzy-Possibilistic C-Means (FPCM) clustering instead of FCM and its fuzzy counterparts and discuss the reason for this shortly. Therefore, to satisfy the first condition, a new algorithm is proposed for determining the best initial parameters of FPCM clustering including the degree of fuzziness (m) and typicality (η). Firstly, the algorithm reconstructs the original dataset from the outputs of the FPCM algorithm for different values of m, η and the number of clusters. Then, the differences between the predicted dataset and the original dataset are determined using Root Mean Squared Error (RMSE). Finally, the best values of η and m are obtained by minimizing the Cumulative Root Mean Square Error (CRMSE) for every pair of (m, η).

In the second condition, a novel validity index is proposed for FPCM clustering called FP index. The major difficulty in measuring the compactness of a validity index is the significant variability in the density, shape, and number of patterns in each cluster. To solve this problem, we assess the dispersion of the data for each cluster and consider the shape and density of clusters using the properties of the Fuzzy-Possibilistic (FP) covariance matrix as a measure of compactness. Also, the essential characteristic of a validity index is its capability to handle noise and outliers. Since FCM and cluster validity indices designed on its basis are quite susceptible to noise, we use FPCM instead of FCM and its fuzzy counterparts. Moreover, an FP exponential-type separation is used in the separation part of the proposed FP index because an exponential operation is extremely effective in dealing with Shannon entropy [10].

The proposed framework is one of the very first FP approaches in the literature. In the forthcoming sections, upon using artificial and well-known datasets, capabilities of the proposed approach will be tested and then, it will be implemented for clustering several real microarray datasets and medical images.

The remainder of this paper is organized as follows. The next section reviews several cluster validity indices and also discusses their advantages and disadvantages. A new cluster validity index is then proposed for FP clustering in Section 3. A method for determining the parameters of the proposed index is presented in Section 4. Section 5 gives the comparisons of experimental results for a variety of datasets and the proposed method will be implemented in microarray gene expression data clustering and medical image segmentation. Finally, conclusions are presented in Section 6.

2. Background

2.1. Fuzzy-possibilistic C-Means (FPCM) clustering

FCM clustering and its variations are the most renowned methods in the literature. FCM was first proposed by Dunn [11] and then, generalized by Bezdek [3]. A disadvantage of the FCM clustering algorithm is that it is susceptible to noise. To attenuate this effect, Krishnapuram and Keller eliminated the membership constraint in FCM and proposed the Possibilistic C-Means (PCM) algorithm [4]. The superiority of PCM is that it is extremely robust in the presence of outliers. However, PCM has several defects, i.e., it considerably relies on good initialization and has an undesirable propensity to generate coincident clusters [12].

To address these shortcomings, Pal and Bezdek defined FPCM clustering that merges the attributes of both FCM and PCM. FPCM overcomes the noise susceptibility of FCM and also resolves the coincident clusters problem of PCM. They believed that typicalities and memberships were indispensable for defining the accuracy feature of data substructure in the clustering problem. In this regard, they defined the objective function of FPCM as follows [13]:

$$\min_{(U,T,V,X)} \left\{ J_{FPCM}(U,T,V,X) = \sum_{i=1}^{c}\sum_{j=1}^{N}(t_{ij}^{\eta}+u_{ij}^{m})D^2(x_j,v_i) \right\}, \quad (1)$$

with the following constraints:

$$\begin{cases} \sum_{i=1}^{c} u_{ij} = 1 & \forall j \in (1,2,...,N) \\ \sum_{j=1}^{N} t_{ij}^{\eta} = 1 & \forall i \in (1,2,...,c) \end{cases} \quad (2)$$

where $X = \{x_1, x_2, ..., x_N\} \subseteq \mathbb{R}^d$ is the dataset in the d-dimensional vector space, u_{ij} is the degree of belonging of the jth data to the ith cluster, $V = \{v_1, v_2, ..., v_c\}$ is the prototype of clusters, $D(x_i, v_i)$ is the distance between the jth data and the ith cluster center, m is the degree of fuzziness, t_{ij} is the typicality, and U and T are fuzzy and possibilistic partition matrices, respectively. η is a suitable positive number, c is the number of clusters, and N is the number of data. This objective function can be solved through an iterative procedure in which the degrees of membership, typicality, and cluster centers are updated via [13]:

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{2/(m-1)} \right)^{-1},$$

$$1 \le i \le c, \qquad 1 \le j \le N \tag{3}$$

$$t_{ij} = \left(\sum_{k=1}^{N} \left(\frac{D(x_j, v_i)}{D(x_j, v_k)} \right)^{2/(\eta-1)} \right)^{-1},$$

$$1 \le i \le c, \qquad 1 \le j \le N, \tag{4}$$

$$v_i = \frac{\sum_{k=1}^{N} (t_{ik}^{\eta} + u_{ik}^{m}) x_k}{\sum_{k=1}^{N} (t_{ik}^{\eta} + u_{ik}^{m})}, \qquad 1 \le i \le c. \tag{5}$$

2.2. Validity indices for fuzzy clustering

In this subsection, some methods are employed for quantitative assessment of the clustering results, known as cluster validity methods. According to the work of Wang and Zhang [14], these methods can be grouped into three main categories:

1. Indices comprising only the membership values;
2. Indices comprising the membership values and dataset;
3. Other approaches.

The earliest validity indices for fuzzy clustering, the partition coefficient V_{PC}, and the partition entropy V_{PE} were introduced by Bezdek [15]. These indices are examples of the indices comprising only the membership values. Their essential drawback is the lack of any connection to the geometrical structure [14]. Some researchers have considered fuzzy memberships and the data structure to resolve this disadvantage. In the current paper, we compare the performance of the proposed validity index with those of fifteen popular cluster validation indices in the literature. Table 1 lists these cluster validity indices. In this table, x_i is the jth data point, c is the number of clusters, v_i's are cluster centers, u_{ij} is the degree of belonging of the jth data to the ith cluster, and N is the total number of patterns in a given dataset. The last three indices in this table are based on the general type-2 fuzzy logic. Higher-order fuzzy clustering algorithms are very well suited to dealing with high levels of uncertainties present in a majority of real-world applications. However, the immense computational complexity associated with such clustering algorithms has been a great obstacle to the practical applications [16].

Now, we focus our attention on a well-known index from the second category, which is the Partition Coefficient And Exponential Separation (PCAES) index proposed by Wu and Yang [21]. V_{PCAES} only utilizes membership values to validate the compactness measure and does not consider the structure of data, i.e., the relative distance between objects and cluster centers [9]. For this reason, it performed poorly in compactness measure. In order to solve this problem, we use the FP covariance matrix and membership values in the proposed compactness measure. In this way, we involve characteristics like density, shape, and patterns in the proposed index.

Moreover, V_{PCAES} takes advantage of the exponential function to validate the separation measure and, also, it involves the distance between the mean of cluster centers and cluster centers. The stimulus behind taking the exponential function is that an exponential operation is extremely effective in coping with Shannon entropy [27,28]; Wu and Yang asserted that an exponential-type distance would yield a robust property. Nevertheless, the experimental results demonstrate that this index produces inappropriate results when the cluster centers are close to each other [9]. Figure 1 illustrates an example of limited scope in which V_{PCAES} loses its capability to indicate the appropriate number of clusters. Intuitively, we know that there are 7 fuzzy clusters in this dataset. In Section 4, it is demonstrated that V_{PCAES} will detect four clusters. This problem occurs because V_{PCAES} calculates the separation between clusters using only centroid distances. To solve these problems in the proposed index, membership values and centroid

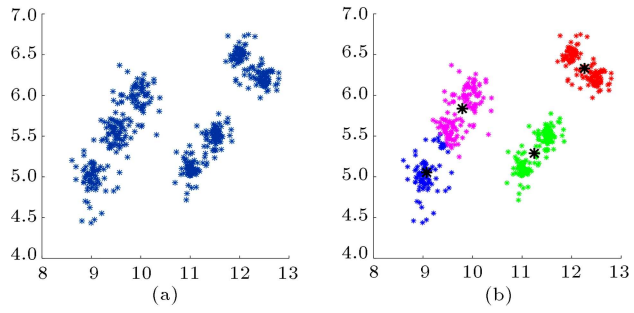


Figure 1. (a) A dataset that consists of seven clusters. (b) The result of Partition Coefficient And Exponential Separation (PCAES) validity index.

Table 1. Fifteen well-known validity indices for fuzzy clustering.

Name/authors	Function	Ref.
Partition Coefficient	$\max_{2\le c\le C_{\max}} \mathrm{V_{PE}}(U,V,X) = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^2$	[15]
Partition Entropy	$\min_{2\le c\le C_{\max}} \mathrm{V_{PE}}(U,V,X) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}\log u_{ij}$	[15]
Fukuyama and Sugeno	$\max_{2\le c\le C_{\max}} \mathrm{V_{FS}}(U,V,X) = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m \parallel x_j - v_i \parallel^2 - \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m \parallel v_i - \bar{v} \parallel^2$, $\bar{v} = \frac{\sum v_i}{c}$	[17]
Xie and Beni	$\min_{2\le c\le C_{\max}} \mathrm{V_{XB}}(U,V,X) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m \left\|x_j - v_i\right\|^2}{N.\min_{i,j}\|v_i-\bar{v}\|^2}$	[18]
Kwon	$\min_{2\le c\le C_{\max}} \mathrm{V_K}(U,V,X) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^2\|x_j-v_i\|^2 + \frac{1}{c}\sum_{i=1}^{c}\|v_i-\bar{v}\|^2}{\min_{i\ne k}\|v_i-v_k\|^2}$, $\bar{v} = \frac{\sum_{j=1}^{N} x_j}{N}$	[19]
Gath and Geva	$\min_{2\le c\le C_{\max}} \mathrm{V_{FHV}}(U,V,X) = \sum_{i=1}^{c}\left[\det(F_i)\right]^{\frac{1}{2}}$, $F_i = \frac{\sum_{j=1}^{N}\left(u_{ij}^m\right)(x_i-v_i)\left(x_j-v_i\right)^T}{\sum_{j=1}^{N}\left(u_{ij}^m\right)}$	[20]
Wu and Yang	$\max_{2\le c\le C_{\max}} \mathrm{V_{PCAES}}(U,V,X) = \sum_{i=1}^{c}\sum_{j=1}^{N}\frac{{u_{ij}}^2}{u_M} - \sum_{i=1}^{c}\exp\left(-\min_{i\ne k}\left\{\frac{\|v_i-v_k\|^2}{B_T}\right\}\right)$, $u_M = \min_{1\le i\le c}\left(\sum_{j=1}^{N} u_{ij}^m\right)$, $B_T = \sum_{s=1}^{c}\frac{\|v_s-\bar{v}\|^2}{c}$, $\bar{v} = \sum_{j=1}^{N}\frac{x_j}{N}$	[21]
Zhang et al.	$\min_{2\le c\le C_{\max}} \mathrm{V_W}(U,V) = \frac{Var^N(U,V)}{Sep^N(c,V)}$, $Var^N(U,V) = Var(U,V)/\max_c(Var(U,V))$, $Sep^N(U,V) = Sep(c,V)/\max_c(sep(c,V))$, $Sep(c,V) = 1-\max_{i\ne j}\left(\max_{x_k\in X}\min(u_{ik},u_{jk})\right)$, $Var(U,V) = \left(\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}d^2(x_j,v_i)/n(i)\right)\times\left(\frac{c+1}{c-1}\right)^{1/2}$	[22]
Rezaee	$\min_{2\le c\le C_{\max}} \mathrm{V_{SC}}(c,U) = Sep^N(c,U) + Comp^N(c,U)$, $Comp^N(c,U) = Comp(c,U)/\max_c(Comp(c,U))$, $Sep^N(c,U) = Sep(c,U)/\max_c(Sep(c,U))$, $Comp(c,U) = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^2 \parallel x_j - v_i \parallel^2$, $Sep(c,U) = \frac{2}{c(c-1)}\sum_{p\ne q}^{c}\left[\sum_{j=1}^{N}\left(\min\left(u_{F_p}(x_j), u_{F_q}(x_j)\right)\right)\times h(x_j)\right]$, $h(x_j) = -\sum_{i=1}^{c} u_{F_p}(x_j)\log_a u_{F_q}(x_j)$	[23]
Zhang et al.	$\max_{2\le c\le C_{\max}} \mathrm{V_{WGLI}} = (2MMD + Q_B)/3$, $MMD = \frac{1}{n}\sum_{j=1}^{N}\max_{1\le i\le c} u_{ij}$, $Q_B = \sum_i(e_{ij}-a_ia_j)$, $j = \max_k(e_{ik})$, $a_i = \sum_i e_{ij} = \frac{1}{2M}\sum_{i\in V_l}\sum_{j\in V} A(i,j)$. A is the adjacency matrix and M is the number of edges in a bipartite network.	[24]
Fazel Zarandi et al.	$\max_{2\le c\le C_{\max}} \mathrm{V_{ECAS}}(c) = \frac{EC_{comp}(c)}{\max_c\left(EC_{comp}(c)\right)} - \frac{ES_{sep}(c)}{\max_c\left(ES_{sep}(c)\right)}$, $EC_{comp}(c) = \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m\exp\left(-\left(\frac{\|x_j-v_i\|^2}{\beta_{comp}} + \frac{1}{c+1}\right)\right)$, $\beta_{comp} = \frac{\sum_{k=1}^{N}\|x_k-\bar{v}\|^2}{n(i)}$, $\bar{v} = \sum_{j=1}^{N}\frac{x_j}{N}$, $n(i)$ is the number of data in cluster i, $ES_{sep}(c) = \sum_{i=1}^{c}\exp\left(-\min_{i\ne j}\left\{\frac{(c-1)\|v_i-v_j\|^2}{\beta_{sep}}\right\}\right)$ $\beta_{comp} = \frac{\sum_{b=1}^{c}\|v_b-\bar{v}\|^2}{c}$	[9]

Table 1. Fifteen well-known validity indices for fuzzy clustering (continued).

Name/authors	Function	Ref.
Fazel Zarandi et al.	$\max_{2\le c\le C_{\max}} \mathrm{V_{FNT}}(U,V,X) = \frac{2}{c(c-1)} \sum_{p\neq q}^{c} S_{rel}(A_p, A_q)$ $S_{rel}(A_p, A_q)$ is the relative similarity between two fuzzy sets A_p and A_q.	[25]
Askari et al.	$\min_{2\le c\le C_{\max}} \mathrm{V_{GPF1}} = \sum_{i=1}^{c} R_i^r \left(\sqrt{\prod_{q=1}^{r}\lambda_{qi}}\right)^{-1} \Big/ \sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m$ λ_{qi} is qth eigenvalue of fuzzy covariance norm matrix	[26]
Askari et al.		[26]
Askari et al.	$\min_{2\le c\le C_{\max}} \mathrm{V_{GPF3}} = \left(c\sum_{i=1}^{c} R_i^r \left(\sqrt{\prod_{q=1}^{r}\lambda_{qi}}\right)^{-1}\right) \Big/ \left(\left(\sum_{i=1}^{c}\sum_{j=1}^{N} u_{ij}^m\right)\left(\sum_{k=1}^{c}\sum_{i=1}^{c}\sum_{j=1}^{N} \lvert u_{kj}-u_{ij}\rvert^m\right)\right)$	[26]

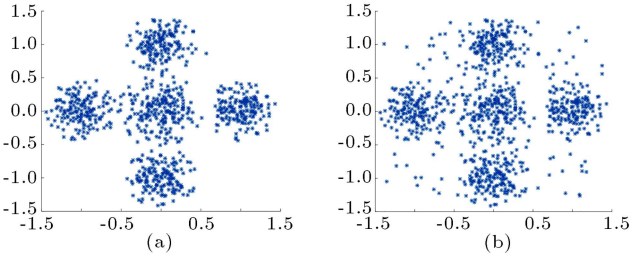


Figure 2. (a) A dataset that consists of five clusters. (b) Previous dataset + 100 noisy points.

distances are used to improve the separation measure.

What is more, a substantial feature of validity index is its capability to handle noise and outliers. Because of the noise sensitivity of FCM and the structure of compactness measure in $\mathrm{V_{PCAES}}$, it is very susceptible to noise. To demonstrate the noise sensitivity of PCAES validity index, we considered a 5-cluster dataset and the optimum number of clusters obtained using $\mathrm{V_{PCAES}}$ was 5. Additionally, we added 100 noisy points to the previous dataset and due to the noise sensitivity of $\mathrm{V_{PCAES}}$, only four well-separated clusters could be detected in this noisy dataset. These datasets are depicted in Figure 2. To solve this problem in the proposed index, we use FPCM clustering instead of FCM or PCM clustering. FPCM clustering overcomes the noise susceptibility of FCM and resolves the coincident clusters problem of PCM. In the next section, we will propose a new validity index for FPCM clustering in order to overcome these shortcomings.

3. The proposed validity index

In the previous section, the most widely used validity indices found in the literature were reviewed and the disadvantages of some of these methods were explained. Moreover, some solutions were suggested to address these issues. Now, a new validity index is proposed for FPCM clustering which considers differences in cluster density, shape, and orientation and works well in the presence of noise. We will demonstrate that this validity index can effectively address these issues.

Definition: Let $X = \{x_1, x_2, \ldots, x_n | x_i \in \Re^p\}$ be an FP c-partition of the dataset with c cluster centers v_i, such that $V = \{v_1, v_2, \ldots, v_c\}$ and u_{ij} as the fuzzy membership of data point x_j belonging to the jth cluster and t_{ij} as typicality of data point x_j belonging to ith cluster.

The FP validity index has the following form:

$$V_{FP}(U,T,V,X) = Comp\,(c,U,T,V,X) + Sep\,(c,U,T,V), \quad (6)$$

where $Comp\,(c,U,T,V,X)$ is the compactness of the FP c-partition which is defined as follows:

$$Comp\,(c,U,T,V,X) = \sum_{i=1}^{c} \frac{1}{trace(F_i)} \sum_{j=1}^{N} \left(t_{ij}^{\eta} + u_{ij}^{m}\right) \parallel x_j - v_i \parallel^2, \quad (7)$$

where m is degree of fuzziness, η is degree of typicality, and F_i is the FP covariance matrix of the ith cluster which is defined as follows:

$$F_i = \frac{\sum_{j=1}^{N} \left(t_{ij}^{\eta} + u_{ij}^{m}\right)\left(x_j - v_i\right)\left(x_j - v_i\right)^T}{\sum_{j=1}^{N}\left(t_{ij}^{\eta} + u_{ij}^{m}\right)}. \quad (8)$$

In the compactness part, if intra-cluster dispersion increases, then clusters become less compact. Thus, the sum of FP variations of clusters is an appropriate indication of the compactness of clusters. It is worth mentioning that Eq. (7) combines the advantages of fuzzy and possibilistic modeling with the power of the covariance matrix as a measure of compactness.

A significant obstacle to measuring the compactness is the considerable variation in the density, shape, and number of patterns of each cluster. To resolve this problem, we can evaluate the variation of data for

each cluster using the attributes of the FP covariance matrix. In general, when compactness of a cluster is greater than that of another cluster, the trace of that cluster covariance matrix will be less than the other. Owing to this inverse correlation between the trace of the cluster covariance matrix and compactness, we have put this term in the denominator to show this inverse correlation. Moreover, we use the trace of a matrix instead of its determinant because the computational complexity of computing the determinant is much greater than the complexity of trace.

$Sep\ (c,U,T,V)$ is the FP exponential-type separation of clusters which is defined as follows:

$$Sep\,(c,\,U,\,T,\,V) = \sum_{i=1}^{c}\sum_{j=1}^{N}\left(t_{ij}^{\eta}+u_{ij}^{m}\right) \exp\left(-\min_{i\neq j}\left(\left(\frac{\parallel v_i - v_j \parallel}{\parallel v_i - \bar{v} \parallel}\right)^{m}\right)\right), \quad (9)$$

where $\bar{v} = \frac{\sum v_i}{c}$. The FP exponential-type separation is similar to the exponential function of the separation measure in $\mathrm{V_{PCAES}}$. According to the research conducted by Wu and Yang, an exponential-type distance is more robust based on the influence function analysis [10]. Furthermore, we have combined the fuzziness and possibility in each row of U and T with the exponential-type separation. In Section 2, $\mathrm{V_{PCAES}}$ was found to have inappropriate results when the cluster centers were close to each other. The experimental results indicate that FP index can correctly determine the number of clusters for this type of dataset (see Section 4).

$Sep\,(c,\,U,\,T,\,V)$ and $Comp\,(c,U,T,V,X)$ have different scales; as a result, they need to be normalized before calculating $\mathrm{V_{FP}}$. First, we explain each of them with respect to $c = 2, 3, ..., c_{\max}$ as follows:

$$Sep\,(c,U,T,V) = \{Sep\,(2,U,T,V)\,, Sep\,(3,U,T,V)\,, ..., Sep\,(c_{\max},U,T,V)\}, \quad (10)$$

$$Comp\ (c,U,T,V,X) = \{Comp\ (2,U,T,V,X)\,, Comp\ (3,U,T,V,X)\,, ..., Comp\ (c_{\max},U,T,V,X)\}. \quad (11)$$

For each measure, the maximum values are computed as follows:

$$Sep_{\max} = \max_{c}\left(Sep\,(c,U,T,V)\right), \quad (12)$$

$$Comp_{\max} = \max_{c}\left(Comp\,(c,U,T,V,X)\right). \quad (13)$$

Then, the normalized separation and compactness can be computed as:

$$Sep^{N}\,(c,U,T,V) = \frac{Sep(c,U,T,V)}{Sep_{\max}}, \quad (14)$$

$$Comp^{N}(c,U,T,V,X) = \frac{Comp\,(c,U,T,V,X)}{Comp_{\max}}. \quad (15)$$

Consequently, the proposed FP cluster validity index $\mathrm{V_{FP}}$ can be redefined as follows:

$$\mathrm{V_{FP}}\,(U,T,V,X) = Comp^{N}(c,U,T,V,X) + Sep^{N}\,(c,U,T,V)\,. \quad (16)$$

In the proposed validity index, the large values for the compactness measure over and separation measure over c are indicative of a compact partition and well-separated clusters, respectively. Therefore, the optimum value of c is obtained by maximizing $\mathrm{V_{FP}}\,(U,T,V,X)$ over $c = 2, 3, ..., c_{\max}$.

In addition, the time and space complexity of validity indices depends on the underlying clustering algorithms. The time complexity of FCM and FPCM clustering algorithms is $O\left(tkNn^2\right)$ where t, k, n, and N are the numbers of iterations, clusters, features, and objects, respectively. Moreover, the space complexity of these two algorithms is $O\left(Nn + kN + n^2\right)$. Of note, since solving most of the common optimization formulations of clustering is NP-hard (in particular, solving the popular FCM and FPCM clustering problems), solving validity indices is also NP-hard.

4. A procedure for determining the parameters of the proposed method

In addition to the number of clusters, FPCM and its different extensions require a priori selection of the degree of fuzziness (m) and degree of typicality (η). During the past few decades, different ranges and values for the optimum degree of fuzziness have been proposed. Here, studies that have proposed a range or a method for determining the optimal degree of fuzziness are briefly reviewed. Then, an efficient procedure for determining optimal values for m and η is presented. Bezdek was one of the first scientists who introduced a heuristic procedure for finding an optimum value for m [29]. McBratney and Moore [30] observed that the objective function value J_m decreased monotonically upon augmenting m. Furthermore, they demonstrated that the greatest change in J_m occurred around $m = 2$. Choe and Jordan [31] proposed an algorithm to find the optimum using the concept of fuzzy decision theory. Yu et al. [32] defined two theoretical rules to select the weighting exponent in the FCM. Through their approach, they revealed the relationship between the stability of the fixed points of the FCM and the dataset

itself. Okeke and Karnieli [33] presented a procedure using the output of the fuzzy clustering. Their method predicts the original data using the idea of linear mixture modeling. The formula for reconstructing the original dataset has the following form:

$$\tilde{X} = \tilde{V}\tilde{U}, \tag{17}$$

where $\tilde{X}$ is the vector of predicted dataset, $\tilde{V}$ the vector of the FCM output centers, and $\tilde{U}$ the matrix of membership functions. Next, the differences between the predicted and original datasets are specified through the following formula [33]:

$$\sigma = \| X - \tilde{X} \| \qquad \sigma > 0, \quad \forall m. \tag{18}$$

Finally, the degree of fuzziness which corresponds to the minimum value of σ is the optimum value [33]. Since the values of η and m play an important role in the FP index, an algorithm is proposed to tackle this problem. In the proposed algorithm, the FPCM clustering is run for different values of m, η, and c. Then, the original dataset is reconstructed from the outputs of FPCM algorithm using the following formulas:

$$\tilde{X}^U = \tilde{V}\tilde{U}, \tag{19}$$

$$\tilde{X}^T = \tilde{V}\tilde{T}^N, \tag{20}$$

where $\tilde{X}^U$ is the vector of the predicted dataset based on the membership functions matrix, $\tilde{U}$ the matrix of membership functions for the FPCM algorithm, $\tilde{V}$ the vector of the FPCM centers, $\tilde{X}^T$ the vector of the predicted dataset based on the normalized typicality matrix, and $\tilde{T}^N$ the normalized typicality matrix defined below:

$$\tilde{T}^N(c, N) = \frac{\tilde{T}(c, N)}{\sum_c \tilde{T}(c, N)}. \tag{21}$$

Then, the difference between the predicted and original datasets is determined by the Root Mean Squared Error (RMSE) as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \tilde{x}_i)^2}, \tag{22}$$

where N is the total number of samples, and x_i and $\tilde{x}_i$ are the actual and predicted data points, respectively. In this respect, $RMSE_{Total}$ can be defined as follows:

$$RMSE_{Total} = RMSE_T + RMSE_U, \tag{23}$$

where $RMSE_T$ and $RMSE_U$ are the root mean squared errors computed by $\tilde{X}^T$ and $\tilde{X}^U$, respectively. Then, Cumulative Root Mean Square Error (CRMSE) for every pair of (m, η) is defined as follows:

$$CRMSE(m, \eta) = \sum_{c=2}^{c_{\max}} RMSE_{Total}(m, \eta, c), \tag{24}$$

where $c_{\max}$ is the maximum number of clusters. Finally, optimal values for m and η can be found by minimizing CRMSE over η and m. The steps of the proposed algorithm can be seen in Algorithm 1.

Algorithm 1 runs the FPCM clustering and computes V_{FP} with respect to $c = 2, 3, \ldots, c_{\max}$. There

- Define the initial parameters:
  - Set $c = 2$ and determine the maximum number of the clusters ($c_{\max}$)
  - Set $m = 1.1$ and determine the maximum value for the degree of fuzziness ($m_{\max}$)
  - Set $\eta = 1.1$ and determine the maximum value for the degree of typicality ($\eta_{\max}$)

For $c = 2$ to $c_{\max}$

For $m = 1.1$ to $m_{\max}$

For $\eta = 1.1$ to $\eta_{\max}$

Compute **the** fuzzy prototypes $\tilde{V}$, membership functions ($\tilde{U}$), and typicality ($\tilde{T}$) using the FPCM algorithm.
Reconstruct the original dataset using Equations 19 and 20.

Compute $RMSE_{Total}$, the difference between the original and predicted datasets, through Equation 23.

End for

End for

End for

Compute $CRMSE$ for every pair of (m, η) using Equation 24.

Find m^* and η^* such that $CRMSE(m^*, \eta^*) = \min_{\eta} \min_{m} (CRMSE)$.

Compute the FP index (V_{FP}) with $m = m^*$ and $\eta = \eta^*$ for $c = 2, 3, \ldots, c_{\max}$

Determine the optimum value of c by maximizing V_{FP} over $c = 2, 3, \ldots, c_{\max}$.

Algorithm 1. The proposed algorithm for determining the suitable values of c, m, and η.

Table 2. Cumulative Root Mean Square Error (CRMSE) values for different η and m.

	m											
η	**1.2**	**1.6**	**2**	**2.2**	**2.6**	**3**	**3.4**	**3.8**	**4.2**	**4.4**	**4.6**	**5**
1.2	6.225	6.344	6.285	6.319	6.302	6.265	6.357	6.316	6.228	6.329	6.278	6.367
1.6	6.333	6.328	6.383	6.408	6.343	6.305	6.333	6.287	6.238	6.321	6.278	6.390
2	6.350	6.343	6.281	6.340	6.307	6.314	6.310	6.351	6.336	6.357	6.363	6.294
2.2	6.316	6.190	6.338	6.252	6.381	6.405	6.597	6.342	6.322	6.316	6.281	6.421
2.6	6.313	6.307	6.346	6.440	6.341	6.331	6.239	6.244	6.254	6.338	6.330	6.313
3	6.312	6.292	6.305	6.311	6.355	6.265	6.282	6.313	6.318	6.360	6.221	6.327
3.4	6.249	6.278	6.325	6.342	6.305	6.303	6.244	6.301	6.289	6.297	6.382	6.309
3.8	6.278	6.293	6.542	6.252	6.337	6.350	6.305	6.318	6.289	6.210	6.290	6.321
4.2	6.363	6.365	6.306	6.315	6.406	6.323	6.354	6.330	6.327	6.365	6.273	6.269
4.4	6.308	6.302	6.658	6.289	6.339	6.338	6.328	6.288	6.306	6.649	6.339	6.288
4.6	6.265	6.328	6.367	6.337	6.275	6.313	6.338	6.456	6.349	6.335	6.316	6.281
5	6.287	6.263	6.313	6.298	6.322	6.292	6.354	6.347	6.358	6.292	6.277	6.421

is no universal agreement on what value to use for $c_{\max}$. The value of $c_{\max}$ can be determined in accordance with the user's knowledge about the dataset; however, since this is not always feasible, a lot of researchers use $c_{\max} = \sqrt{N}$ instead [34]. Furthermore, the variation in FP index values for all experimental datasets demonstrates that the maximum value of V_{FP} exists between 2 and $\sqrt{N}$ (see Section 4). In order to exhibit the behavior of Algorithm 1, the dataset shown in Figure 2(b) is used as the input data. Let $c_{\max} = 8$, $m_{\max} = 5$, and $\eta_{\max} = 5$ be the initial values for Algorithm 1 (the theoretical rules proposed by Yu et al. are employed to define $m_{\max}$ and $\eta_{\max}$). Table 2 shows the cumulative root mean square error of this dataset. The elements of this table are the degree of typicality (η) and degree of fuzziness (m) considered as the input variables and CRMSE as results. For instance, for $m = 1.2$, and $\eta = 2.2$, CRMSE is 6.316. According to Table 2, the suitable values of m and η can be found by $CRMSE(m^*, \eta^*) = \min_{\eta} \min_{m} (CRMSE)$. Therefore, the suitable values of m and η are 1.6 and 2.2, respectively. Finally, the optimal number of clusters obtained through Algorithm 1 is five with $m = 1.6$ and $\eta = 2.2$. Figure 3 presents the variation in the proposed index values with the number of clusters for this dataset.

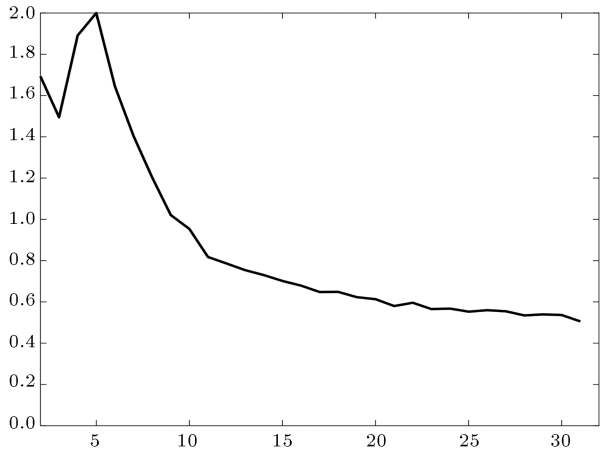


Figure 3. The variation in the proposed index values with the number of clusters.

5. Experimental results

In this section, to ascertain the effectiveness of FP index, we conducted comparisons between FP index and some well-known indices in the literature, as reviewed in Section 2. In the next subsections, FP index will be evaluated using several synthetic and real-world datasets. Moreover, in order to clarify the ability of the proposed method in real applications, the proposed method is implemented in microarray gene expression data clustering and medical image segmentation. In the computational experiments, all the indices are computed using the same input in order to achieve comparable results. In this regard, the clustering algorithm is run and then, the resulting U matrix, the prototypes of clusters, and the other inputs needed for the indices are used for all the indices.

5.1. Artificial and real-world datasets

Eight artificial and five well-known datasets were considered for experiments. These eight artificial datasets are called Dataset_2_12, Dataset_2_5, Dataset_2_6, Dataset_2_7, Dataset_2_8, Dataset_2_10, Dataset_3_3, and Dataset_3_4. The names imply the number of clusters that actually exists in the data and its dimensions. For instance, in

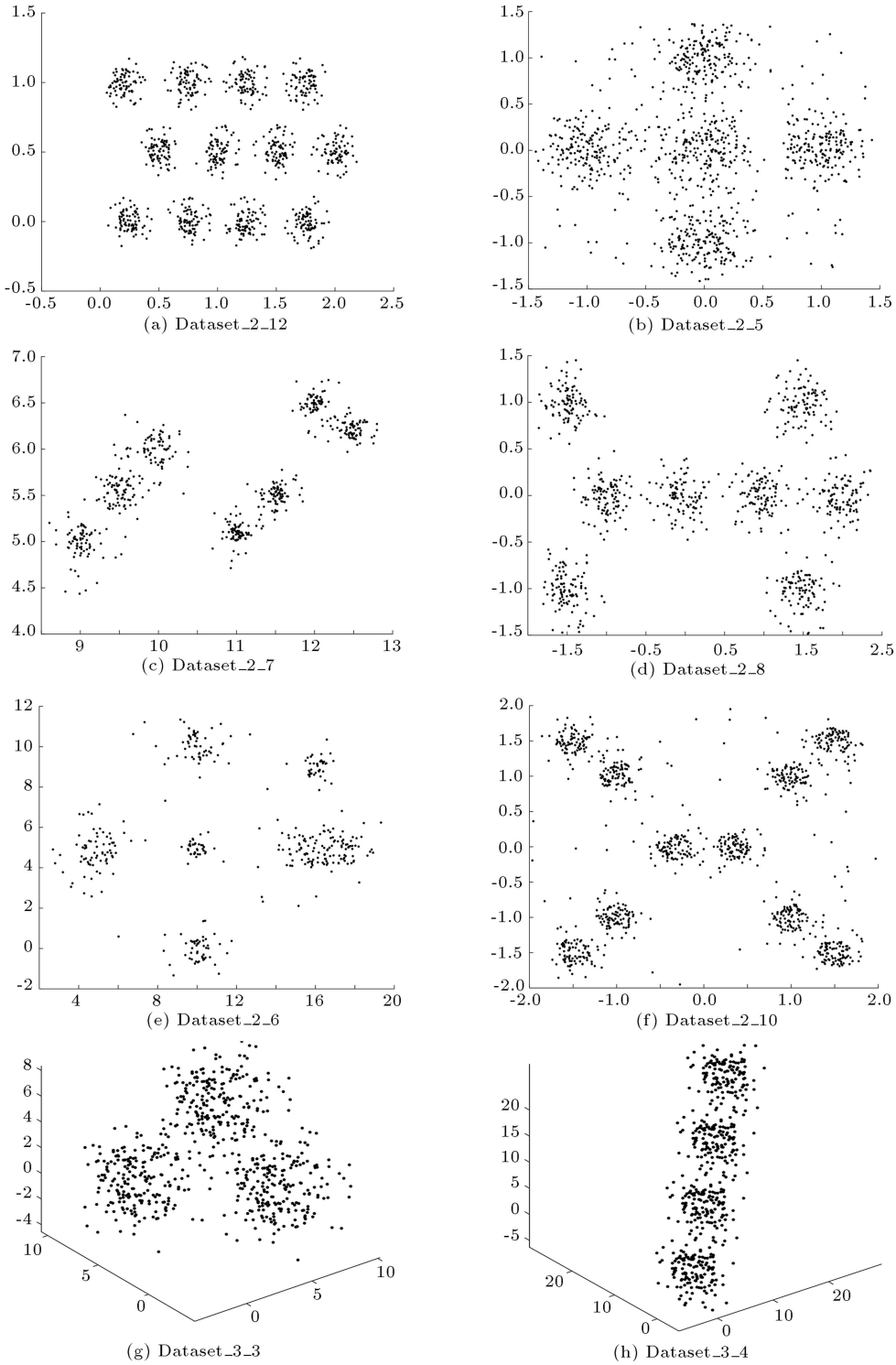


Figure 4. The eight artificial datasets.

Dataset_2_5, there are five clusters and two dimensions of the data. As observed, the artificial datasets include two- and three-dimensional data where the number of clusters varies from three to twelve. These datasets are demonstrated in Figure 4. In addition, six well-known datasets including Bupa Liver Disorder, Wine, Iris, Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC), and Mammographic mass were used in this study, all of which were real-life datasets freely accessible at [35]. The real-world datasets are characterized by four to thirty dimensions and the number of clusters varies from two to three.

In this section, experiment results are presented

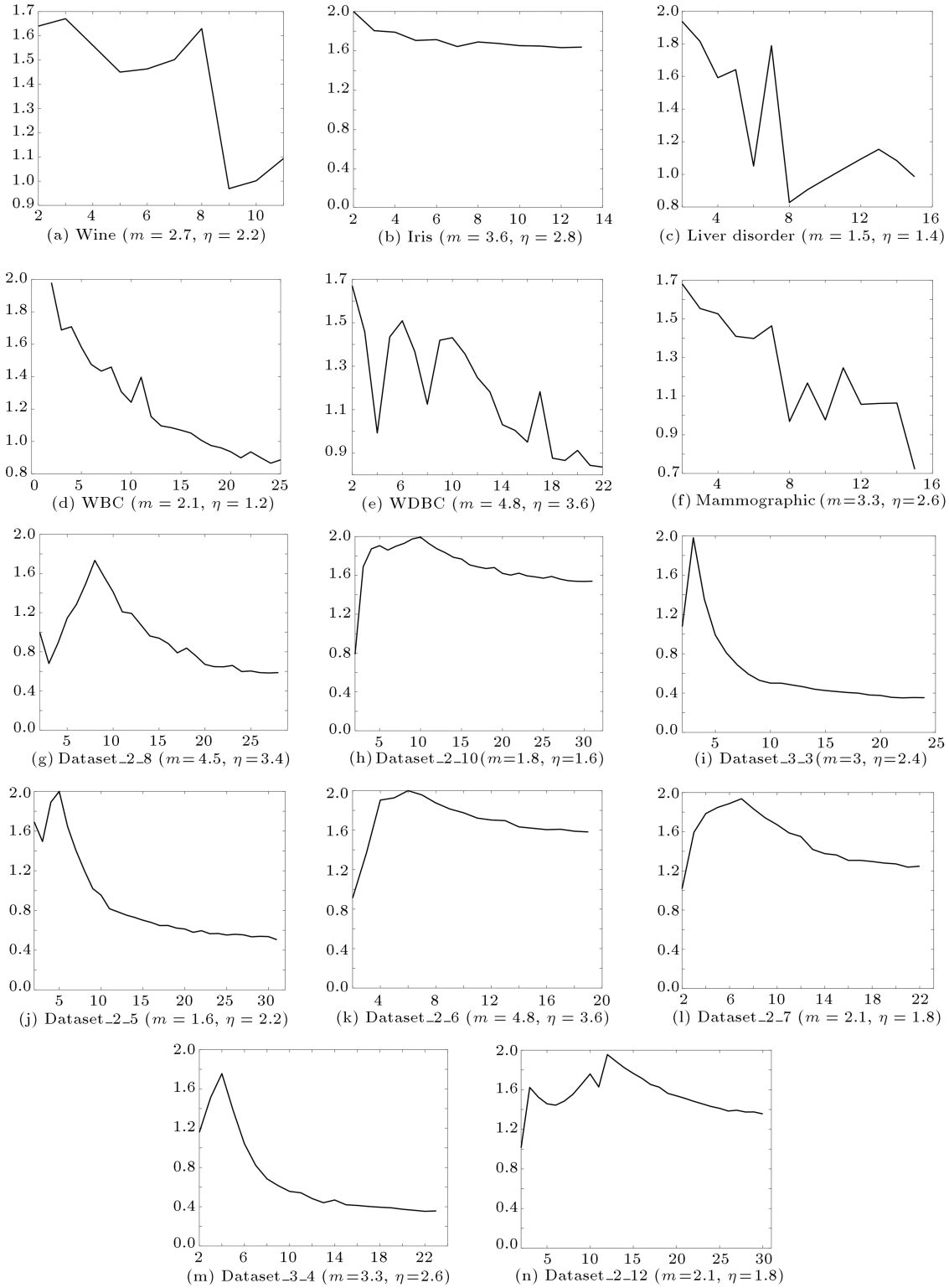


Figure 5. The variation in the proposed index values with the number of clusters for all datasets.

to compare the V_{FP} index with fifteen other indices including V_{PC}, V_{PE}, V_{FS}, V_{XB}, V_K, V_{FHV}, V_{PCAES}, V_W, V_{SC}, V_{WGLI}, V_{ECAS}, V_{FNT}, V_{GPF1}, V_{GPF2}, and V_{GPF3}. In the proposed index, the optimum value of c is obtained by maximizing $V_{FP}(U, T, V, X)$ over $c = 2, 3, \ldots, c_{max}$. Figure 5 shows the variation of V_{FP} with c for all of datasets. The maximum value of the index corresponds to the optimum number of clusters. These values for each dataset are listed in Figure 5. For example, the proposed V_{FP} index reaches

Table 3. The optimal number of clusters obtained by each cluster validity index.

Dataset	C^*	PC	PE	FS	XB	K	FHV	PCAES	W	SC	WGLI	ECAS	FNT	GPF1	GPF2	GPF3	FP
Dataset_2_12	12	**10**	**13**	7	12	12	12	12	12	12	12	12	12	12	12	12	12
Dataset_2_5	5	**4**	**2**	5	**4**	**4**	5	**4**	**4**	**4**	**4**	**4**	**4**	4	5	5	5
Dataset_2_6	6	**7**	**7**	6	6	6	**7**	6	**7**	6	6	6	**7**	**7**	6	6	6
Dataset_2_7	7	**2**	**2**	7	**6**	7	7	**4**	7	**6**	7	**6**	**6**	7	7	7	7
Dataset_2_8	8	**2**	**2**	**7**	**7**	8	8	8	8	8	8	8	8	8	8	8	8
Dataset_2_10	10	10	10	**5**	**8**	**5**	10	**5**	10	10	**8**	10	10	10	10	10	10
Dataset_3_3	3	**2**	**2**	3	**2**	**2**	3	**2**	3	3	3	**2**	**2**	3	3	3	3
Dataset_3_4	4	**2**	**2**	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Liver Disorder	2	2	2	**4**	2	2	**18**	2	2	2	2	2	2	2	2	2	2
Wine	3	**2**	**2**	**13**	3	3	3	3	3	3	3	3	**2**	3	3	3	3
Iris	2,3	2	2	**5**	2	2	3	2	2	2	2	2	2	2	2	2	2
WBC	2	2	2	**12**	2	2	2	2	2	2	2	2	2	2	2	2	2
WDBC	2	2	2	**12**	2	2	2	2	2	2	2	2	2	2	2	2	2
Mammographic mass	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

the maximum ($V_{FP} = 2$) at $c^* = 2$ for the Iris dataset, which properly reveals the underlying cluster number.

Furthermore, Table 3 summarizes the results obtained when the 15 different validity indices were applied to the above-mentioned datasets. The column c^* in Table 3 gives the actual number of clusters for each dataset, and other columns show the optimal cluster numbers obtained from each index. In this table, the highlighted entries correspond to the incorrect result of the indices.

According to this table, the validity indices V_{FP}, V_{GPF2}, and V_{GPF3} can correctly recognize the correct number of clusters for all of the datasets. Of note, the general type-2 fuzzy clustering algorithms outperformed the type-1 fuzzy clustering algorithms in many computational experiments [36]. This happens mainly because a general type-2 fuzzy set offers a way to model higher levels of uncertainty resulting from additional degrees of freedom provided by its third dimension [36]. However, the general type-2 fuzzy is computationally much more complex than type-1 fuzzy, particularly the defuzzifier process which is a very costly operation [36]. As a result, the immense computational complexity associated with general type-2 fuzzy clustering algorithms becomes a great obstacle to practical applications. In this respect, although V_{GPF2} and V_{GPF3} perform with high accuracy, the proposed validity index in this study achieved the same result with much less computational complexity.

In addition, $V_{FP}, V_{FHV}, V_W, V_{GPF1}, V_{GPF2}$, and V_{GPF3} correctly recognize the number of clusters for datasets where the cluster centers are close to each other (Dataset_2_10 and Dataset_2_7). According to the results, for noisy datasets (i.e., Dataset_2_6 and Dataset_2_5), the validity indices containing only the membership values are very susceptible to noises; however, some of the validity indices comprising the membership values and the dataset (i.e., V_{FP} and V_{FHV}) are robust to noise.

5.2. Analysis of gene expression data

Studies on microarray gene expression have been the main focus of researchers over the last few years. The main objective of these studies was to find the biologically considerable knowledge hidden under a large volume of gene expression data. In particular, recognizing gene groups that exhibit similar expression patterns (co-expressed genes) allows identifying the set of genes involved in the same biological process; therefore, we can characterize unfamiliar biological facts. Clustering algorithms have exhibited an excellent capability to find the underlying patterns in microarray gene expression profiles [37].

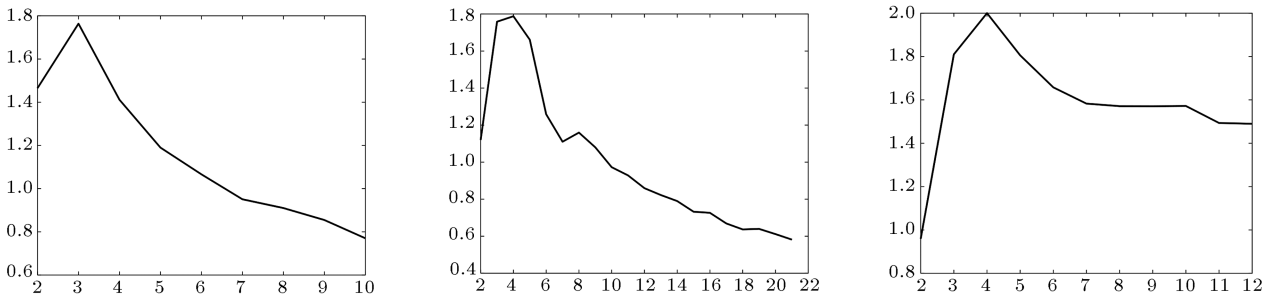
With consideration of a set of genes, a clustering algorithm divides the genes into a number of distinct clusters based on certain similarity measures [38]. Each cluster corresponds to a specific macroscopic phenotype, such as clinical syndromes or cancer types [39]. In fact, a clustering algorithm should identify a set of clusters such that genes within a cluster possess high similarity as compared to those in different clusters; this is not possible without knowing the optimal number of clusters.

In this subsection, three microarray gene expression datasets namely, yeast sporulation, Rat CNS, and Arabidopsis thaliana, are tested and the capability of the FP index is analyzed from different perspectives. These datasets are adopted from [40–42]. For more information on the features of these datasets, refer to [1].

After implementing the proposed validity index, the optimum number of clusters for Arabidopsis thaliana and yeast sporulation datasets and Rat CNS dataset is 4 and 3 clusters, respectively. The variation

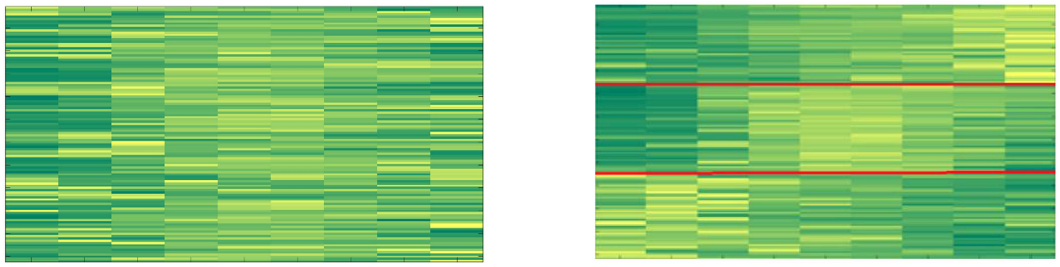
Table 4. The optimal number of clusters obtained by each cluster validity index.

Dataset	c^*	PC	PE	FS	XB	K	FHV	PCAES	W	SC	WGLI	ECAS	FNT	GPF1	GPF2	GPF3	FP
Rat CNS	3	**2**	**2**	**7**	**2**	**2**	**8**	**2**	**2**	3	3	**2**	**2**	3	3	3	3
Yeast	4	**2**	**2**	**3**	**2**	**2**	**7**	**2**	4	**3**	**2**	4	4	4	4	4	4
Arabidopsis	4	**2**	**2**	**7**	4	**3**	**7**	**2**	4	**2**	4	4	**3**	4	4	4	4



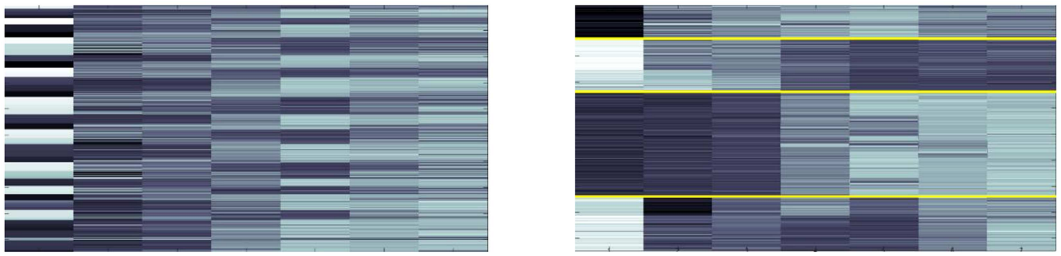
(a) Rat CNS (m=2.2, η=2.4) (b) Yeast (m = 3.1, η = 2.3) (c) Arabidopsis (m = 2.6, η = 3.2)

Figure 6. The variation in FP index values with the number of clusters for all microarray gene expression datasets.



(a) A random ordering of Rat CNS dataset (b) The proposed method

Figure 7. Eisen plot for Rat CNS dataset.

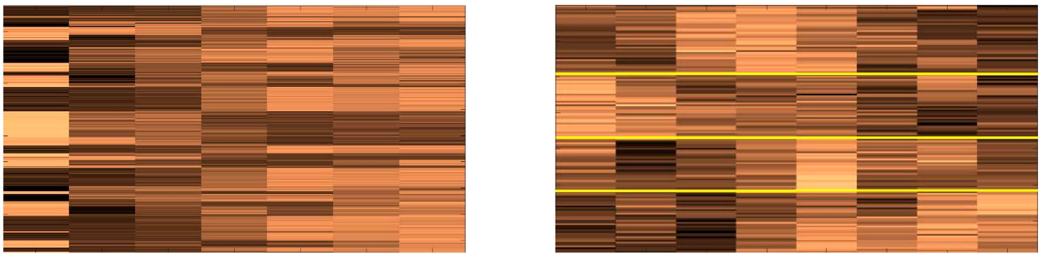


(a) A random ordering of yeast sporulation dataset (b) The proposed method

Figure 8. Eisen plot for yeast sporulation dataset.

of V_{FP} with the number of clusters for these datasets is depicted in Figure 6. In fact, based on [40–42], the optimum number of clusters for Arabidopsis thaliana and yeast sporulation datasets is 4 clusters and for Rat CNS dataset is 3 clusters. Therefore, the proposed validity index detected the correct number of clusters for all of these datasets. Table 4 summarizes the results obtained when these fifteen different validity indices are applied to these microarray gene expression datasets. As observed earlier, the FP, GPF1, GPF2, and GPF3 indices can correctly determine the number of clusters for all of these datasets. However, given the high complexity of GPF1, GPF2, and GPF3 indices compared to the proposed index, we can conclude that FP index yields the best result for the gene expression datasets.

Moreover, to ensure a better understanding of the microarray gene expression context, we used Eisen plot as a visual tool. An Eisen plot is a two-dimensional false color plot that visualizes the expression levels of many genes across several samples. Every row in the Eisen plot demonstrates a gene expression profile across the sample [43]. We have also generated a random sequence of genes to make a simpler distinction between the effects of FP index and a random sequence of genes for each dataset. These plots are depicted in Figures 7–9. Lines in each Eisen plot are the boundaries of clusters. Here, given that the number of clusters has



(a) A random ordering of Arabidopsis Thaliana dataset

(b) The proposed method

Figure 9. Eisen plot for Arabidopsis thaliana dataset.

(a) Data 1 (b) Data 2 (c) Data 3

Figure 10. (a) CT image revealing a large cystic lesion. (b) CT image demonstrating a large liver hepatic angiomyolipoma. (c) Grey scale ultrasound showing two focal non-specific hypoechoic liver lesions.

been determined properly, the genes within a cluster possess high similarity as compared to the genes in other clusters. Moreover, the genes in different clusters are properly separated.

It can be seen that FP index performs really well in determining the suitable number of clusters of the gene expression datasets. However, it should be noted that because of the complicated nature of the gene expression datasets, it is difficult to find a single partitioning that can be claimed to be the optimal partition. From the figures, it is apparent that the expression profiles of the genes of a cluster are similar to each other and they usually have similar color patterns. Moreover, these figures also demonstrate how cluster profiles for various groups of genes differ from each other, while the profiles within a group are similar.

5.3. Medical image segmentation

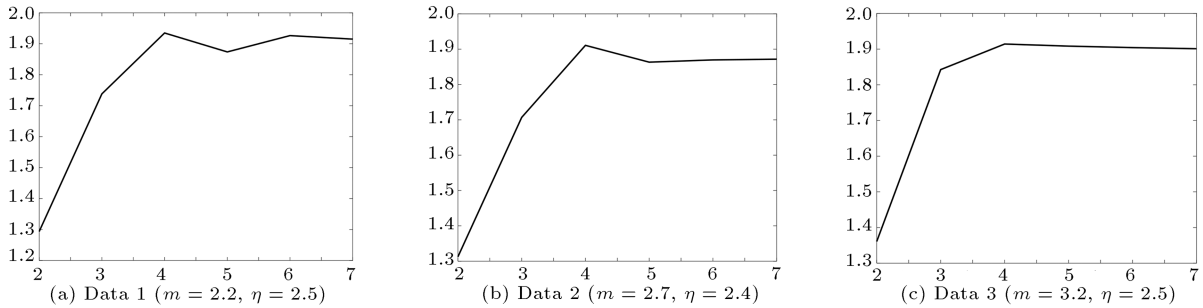
In this subsection, V_{FP} is experimented on several liver images to indicate the applicability of this approach to medical image segmentation. In general, segmentation is the process of dividing an image into regions with similar properties such as color, brightness, texture, and contrast [1]. The existence of noise and low contrast in medical images are critical barriers that may hinder achieving a good segmentation system. Thus, in order to check the applicability of our method in this area, three medical images of the liver are used. For this purpose, two CT images and an ultrasound image of the liver have been adopted from [44]. These medical data are shown in Figure 10.

The proposed index is performed for the datasets corresponding to these images, and the number of clusters obtained is four for all of the datasets. The variation in FP index values with the number of clusters for these datasets is given in Figure 11. Moreover, the results of segmentation by FPCM clustering when the number of clusters is four are shown in Figures 12–14.

Table 5 summarizes the results obtained when the fifteen different validity indices were applied to the datasets corresponding to these images. As you can see, the segmentation results show that FP, ECAS, GPF1, GPF2, and GPF3 indices can successfully recognize the optimal number of clusters for all of these datasets. Consequently, FP index can effectively segment the

Table 5. The optimal number of clusters obtained by each cluster validity index.

Dataset	PC	PE	FS	XB	K	FHV	PCAES	W	SC	WGLI	ECAS	FNT	GPF1	GPF2	GPF3	FP
Data1	**2**	**2**	**5**	**10**	4	**17**	**2**	4	4	**2**	4	4	4	4	4	4
Data2	**2**	**2**	**5**	**5**	**2**	4	4	**8**	4	4	4	4	4	4	4	4
Data3	**2**	**2**	**11**	**9**	**2**	**2**	**2**	4	**5**	4	4	**5**	4	4	4	4



(a) Data 1 ($m = 2.2$, $\eta = 2.5$) (b) Data 2 ($m = 2.7$, $\eta = 2.4$) (c) Data 3 ($m = 3.2$, $\eta = 2.5$)

Figure 11. The variation in Fuzzy-Possibilistic (FP) index values with the number of clusters for the medical image datasets.

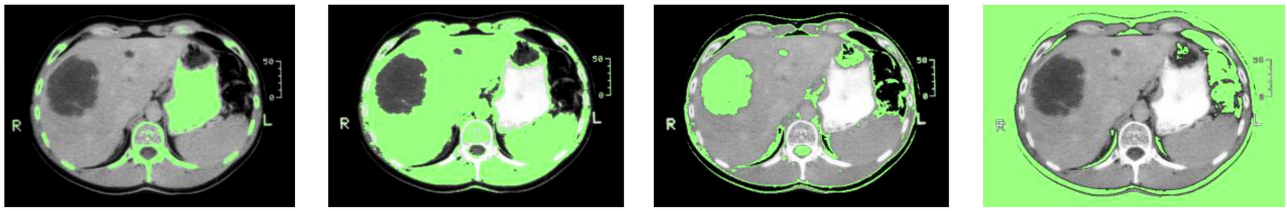


Figure 12. The segmented Data 1.

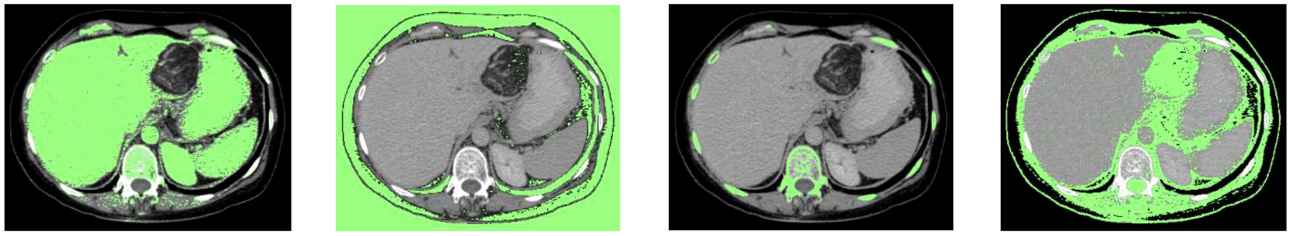


Figure 13. The segmented Data 2.

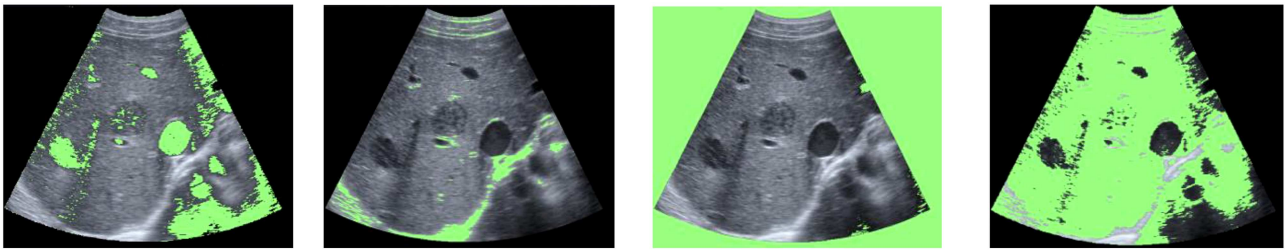


Figure 14. The segmented Data 3.

cysts and lesions from the CT images and the ultrasound image, despite the gray level resemblance of adjoining organs and the different gray levels of hepatic cysts and lesions in the images.

In general, the sensitivity of the ultrasound images was significantly less than that of CT images in detecting lesions. Ultrasound images are not as detailed as those from CT or MRI scans. Their applications are also limited in some parts of the body because the sound waves cannot go through the air (such as in the lungs) or bone. The main objective of this experiment was to show the capability of the proposed method to detect the lesions in ultrasound images despite the gray level resemblance of adjoining organs and different gray levels of hepatic cysts and lesions in these images. As observed in Figure 14, the proposed method successfully detects two lesions. However, due to the low quality of this kind of images, the final result is not as good as CT images.

6. Conclusions

The present study managed to investigate several fuzzy validity indices and discuss their advantages and disadvantages. According to the observation, in case the cluster centers were close to each other, some of these indices produced incorrect results. Moreover, most of these indices were susceptible to noise. To

overcome these shortcomings, a new fuzzy-possibilistic validity index called FP index was proposed in this study.

Furthermore, FPCM, like other clustering algorithms, was susceptible to some initial parameters. In this regard, in addition to the number of clusters, FPCM required a priori selection of the degree of fuzziness (m) and degree of typicality (η). Therefore, an efficient procedure for determining the optimal values of m and η was required.

In order to demonstrate the efficiency of FP index, the proposed index was assessed using eight artificial and five well-known datasets. The results of the experiments demonstrated the effectiveness and flexibility of the FP validity index in terms of sensitivity to cluster overlapping and difference in the cluster shape and density, in comparison with several other well-known approaches in the literature. Moreover, the applications of the proposed approach in real microarray gene expression datasets and medical image segmentation were discussed. In both applications, the proposed method, which was robust in the presence of noise, exhibited an excellent performance in determining the proper number of clusters.

References

1. Duda, R.O. and Hart, P.E., *Pattern Classification and Scene Analysis*, **3**, New York: Wiley (1973).
2. Gallegos, M.T. and Ritter, G. "Probabilistic clustering via Pareto solutions and significance tests", *Advances in Data Analysis and Classification*, **12**(2), pp. 179–202 (2018).
3. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer Science and Business Media (2013).
4. Krishnapuram, R. and Keller, J.M. "A possibilistic approach to clustering", *IEEE Transactions on Fuzzy Systems*, **1**(2), pp. 98–110 (1993).
5. Mendel, J.M. "Type-2 fuzzy sets", In *Uncertain Rule-Based Fuzzy Systems*, pp. 259–306, Springer, Cham (2017).
6. Sotudian, S., Zarandi, M.F., and Turksen, I.B. "From type-I to type-II fuzzy system modeling for diagnosis of hepatitis", *World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng.*, **10**(7), pp. 1238–1246 (2016).
7. Haldar, N.A.H., Khan, F.A., Ali, A., et al. "Arrhythmia classification using mahalanobis distance based improved fuzzy c-means clustering for mobile health monitoring systems", *Neurocomputing*, **220**, pp. 221–235 (2017).
8. Zarandi, M.H., Fazel, A., Seifi, H., Esmaeeli, et al. "A type-2 fuzzy hybrid expert system for commercial burglary", In *North American Fuzzy Information Processing Society Annual Conference*, pp. 41–51 (2017).
9. Fazel Zarandi, M.H., Faraji, M.R., and Karbasian, M. "An exponential cluster validity index for fuzzy clustering with crisp and fuzzy data", *Sci. Iran. Trans. E Ind. Eng.*, **17**, pp. 95–110 (2010).
10. Wu, K.L. and Yang, M.S. "Alternative c-means clustering algorithms", *Pattern Recognition*, **35**, pp. 2267–2278 (2002).
11. Dunn, J.C., *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*, *Journal of Cybernetics*, pp. 32–57 (1973).
12. Krishnapuram, R. and Keller, J.M. "The possibilistic c-means algorithm: insights and recommendations', *IEEE Transactions on Fuzzy Systems*, **4**(3), pp. 385–393 (1996).
13. Pal, N.R., Pal, K., and Bezdek, J.C. "A mixed c-means clustering model. In Fuzzy Systems", *Proceedings of 6th International Fuzzy Systems IEEE*, **1**, pp. 11–21 (1997).
14. Wang, W. and Zhang, Y. "On fuzzy cluster validity indices", *Fuzzy Sets and Systems*, **158**(19), pp. 2095–2117 (2007).
15. Bezdek, J.C., Keller, J., Krisnapuram, R., et al., *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer Science and Business Media (2006).
16. Wijayasekara, D., Linda, O., and Manic, M. "Shadowed type-2 fuzzy logic systems", In *T2FUZZ*, pp. 15–22 (2013).
17. Fukuyama, Y. and Sugeno, M. "A new method of choosing the number of clusters for the fuzzy c-means method", In *Proc. 5th Fuzzy Syst. Symp.*, **247**, pp. 247–250 (1989).
18. Xie, X.L. and Beni, G. "A validity measure for fuzzy clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(8), pp. 841–847(1991).
19. Kwon, S.H. "Cluster validity index for fuzzy clustering", *Electron Lett.*, **34**(22), pp. 2176–2178 (1998).
20. Gath, I. and Geva, A.B. "Unsupervised optimal fuzzy clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7), pp. 773–780 (1989).
21. Wu, K.L. and Yang, M.S. "A cluster validity index for fuzzy clustering", *Pattern Recognition Letters*, **26**(9), pp. 1275–1291 (2005).
22. Zhang, Y., Wang, W., Zhang, X., et al. "A cluster validity index for fuzzy clustering", *Information Sciences*, **178**(4), pp. 1205–1218 (2008).
23. Rezaee, B. "A cluster validity index for fuzzy clustering", *Fuzzy Sets and Systems*, **161**(23), pp. 3014–3025 (2010).
24. Zhang, D., Ji, M., Yang, J., et al. "A novel cluster validity index for fuzzy clustering based on bipartite modularity", *Fuzzy Sets and Systems*, **253**, pp. 122–137 (2014).
25. Zarandi, M.H.F., Neshat, E., and Türkşen, I.B. "Retracted article: A new cluster validity index for

fuzzy clustering based on similarity measure", In *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, Springer, Berlin, Heidelberg, pp. 127–135 (2007).

26. Askari, S., Montazerin, N., and Zarandi, M.F. "Generalized possibilistic fuzzy c-means with novel cluster validity indices for clustering noisy data", *Applied Soft Computing*, **53**, pp. 262–283 (2017).
27. Pal, N.R. and Pal, S.K. "Entropy: A new definition and its applications", *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(5), pp. 1260–1270 (1991).
28. Pal, N.R. and Pal, S.K. "Some properties of the exponential entropy", *Information Sciences*, **66**(1-2), pp. 119–137 (1992).
29. Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Algorithms*, Plenum Press, New York (1981).
30. McBratney, A.B. and Moore, A.W. "Application of fuzzy sets to climatic classification", *Agricultural and Forest Meteorology*, **35**(1-4), pp. 165–185 (1985).
31. Choe, H. and Jordan, J.B. "On the optimal choice of parameters in a fuzzy c-means algorithm", *IEEE International Conference on Fuzzy Systems*, pp. 349–354 (1992).
32. Yu, J., Cheng, Q., and Huang, H. "Analysis of the weighting exponent in the FCM", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **34**(1), pp. 634–639 (2004).
33. Okeke, F. and Karnieli, A. "Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm", *Ecological Informatics*, **1**(1), pp. 117–124 (2006).
34. Bezdek, J.C., *Pattern Recognition in Handbook of Fuzzy Computation*, IOP Publishing Ltd., Boston, MA (1998).
35. UCI Machine Learning Repository, Retrieved October 21 (2018). http://www.ics.uci.edu/mlearn/ databases.html.
36. Torshizi, A.D., Zarandi, M.F., and Türksen, I.B. "Computing centroid of general type-2 fuzzy set using constrained switching algorithm", *Scientia Iranica, Transactions E, Industrial Engineering*, **22**(6), p. 2664 (2015).
37. Jothi, R., Mohanty, S.K., and Ojha, A. "DK-means: a deterministic K-means clustering algorithm for gene expression analysis", *Pattern Analysis and Applications*, pp. 1–19 (2017).
38. Hosseini, B. and Kiani, K. "FWCMR: A scalable and robust fuzzy weighted clustering based on MapReduce with application to microarray gene expression", *Expert Systems with Applications*, **91**, pp. 198–210 (2018).
39. Daxin, J., Tang, C., and Zhang, A. "Cluster analysis for gene expression data: a survey", *IEEE Transactions on Knowledge and Data Engineering*, **16**(11), pp. 1370–1386 (2004).
40. The Transcriptional Program of Sporulation in Budding Yeast. (n.d.), Retrieved October 21, (2018). http://www.sciencemag.org/content/ 282/5389/699.long
41. *Validating Clustering for Gene Expression Data. (n.d.)*, Retrieved October **21** (2018). http://faculty.washington.edu/ kayee/cluster
42. Biological Data Analysis using Clustering (n.d.), Retrieved October **21** (2018). http://homes.esat.kuleuven.be/thijs Work/Clustering.html
43. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. "Cluster analysis and display of genome-wide expression patterns", *Proceedings of the National Academy of Sciences*, **95**(25), pp. 14863–14868 (1998).
44. Open-edit radiology resource, Retrieved October 21 (2018). http:// Radiopaedia.org

Biographies

Mohammad Hossein Fazel Zarandi is a Professor at the Department of Industrial Engineering at Amirkabir University of Technology, Tehran, Iran and a member of the Knowledge-Information Systems Laboratory of the University of Toronto, Canada. His main research interests focus on intelligent information systems, soft computing, computational intelligence, fuzzy sets and systems, multi-agent systems, networks, metaheuristics, and optimization. He has authored several books, scientific papers, and technical reports in the above areas, most of which are accessible on the web. He has also taught several courses in fuzzy systems engineering, decision support systems, management information systems, artificial intelligence and expert systems, systems analysis and design, scheduling, neural networks, simulations, and production planning and control in several universities in Iran and North America.

Shahabeddin Sotudian is currently a PhD Student at the Division of Systems Engineering at Boston University. He completed his MSc and BSc in Industrial and Systems Engineering at Amirkabir University of Technology. His research interests lie in machine learning, numerical optimization, and expert systems. He has published several papers in international journals and conferences around the world.

Oscar Castillo holds the Doctor in Science degree (Doctor Habilitatus) in Computer Science from the Polish Academy of Sciences (with the Dissertation "Soft Computing and Fractal Theory for Intelligent Manufacturing"). He is a Professor of Computer Science in the Graduate Division, Tijuana Institute of Technology, Tijuana, Mexico. In addition, he is serving as the Research Director of Computer Science

and Head of the Research Group on Hybrid Fuzzy Intelligent Systems. Currently, he is President of HAFSA (Hispanic American Fuzzy Systems Association) and Past President of IFSA (International Fuzzy Systems Association). Prof. Castillo is also Chair of the Mexican Chapter of the Computational Intelligence Society (IEEE). He also belongs to the Technical Committee on Fuzzy Systems of IEEE and to the Task Force on "Extensions to Type-1 Fuzzy Systems". He is also a member of NAFIPS, IFSA and IEEE. He belongs to the Mexican Research System (SNI Level 3). His research interests are Type-2 fuzzy logic, fuzzy control, neuro-fuzzy and genetic-fuzzy hybrid approaches. He has published over 300 journal papers, 10 authored books, 40 edited books, 200 papers in conference proceedings, and more than 300 chapters in edited books.