

Article

# A New Validity Index Based on Fuzzy Energy and Fuzzy Entropy Measures in Fuzzy Clustering Problems

Ferdinando Di Martino <sup>1,2,\*</sup> and Salvatore Sessa <sup>1,2</sup>

<sup>1</sup> Università degli Studi di Napoli Federico II, Dipartimento di Architettura, Via Toledo 402, 80134 Napoli, Italy

<sup>2</sup> Università degli Studi di Napoli Federico II, Centro Interdipartimentale di Ricerca "Alberto Calza Bini", Via Toledo 402, 80134 Napoli, Italy; sessa@unina.it

\* Correspondence: fdimarti@unina.it; Tel.: 390812538908; Fax: 39081238905; Mobile: 393334529362

**Abstract.** Two well-known drawbacks in fuzzy clustering are the requirement of assign in advance the number of clusters and random initialization of cluster centers.; the quality of the final fuzzy clusters depends heavily on the initial choice of the number of clusters and the initialization of the clusters, then it is necessary to apply a validity index to measure the compactness and the separability of the final clusters and run the clustering algorithm several times. We propose a new fuzzy C-means algorithm in which a validity index based on the concepts of maximum fuzzy energy and minimum fuzzy entropy is applied to initialize the cluster centers and to find the optimal number of clusters and initial cluster centers in order to obtain a good clustering quality, without increasing time consumption. We test our algorithm on UCI machine learning classification datasets comparing the results with the ones obtained by using well-known validity indices and variations of FCM using optimization algorithms in the initialization phase. The comparison results show that our algorithm represents an optimal trade-off between the quality of clustering and the time consumption.

**Keywords:** FCM; validity index; fuzzy energy; fuzzy entropy

---

## 1. Introduction

A validity index is a measure applied in fuzzy clustering to evaluate the compactness of clusters and the separability among clusters.

Numerous validity indices have been applied to measure the compactness and separateness of clusters detected by applying the fuzzy C-means algorithm [1,2] (for short, FCM).

The two well-known main drawbacks of the FCM are the random setting of the initial clusters and the requirement of assign in advance the number of clusters. The initial selection of the cluster centers can affect the performances of the algorithm in terms of efficiency and number of iterations need to obtain the convergence. Moreover, the quality of the final fuzzy clusters depends on the choice of the number of clusters, then it is necessary to use a validity index to evaluate what is the optimal number of clusters.

A simple technique applied to solve these problems is to execute several times the clustering algorithm varying the initial centers of the clusters and the number of clusters and to choice the optimal clustering using a validity index to measure the quality of the final clustering. However, this technique can be computationally expensive as the clustering algorithm has to be run many times.

In [3,4] a technique is proposed which is based on the subtractive clustering algorithm to initialize the clusters, but this method needs to set the maximum peak and the maximum radius parameters.

In [5] a technique, called *Fuzzy Silhouette*, is proposed: this method generalizes the Average Silhouette Width Criterion [6] applied for evaluate the quality of crisp clustering. The authors of [5] show that the proposed validity measure, unlike to other well-known validity measures as Fuzzy Hypervolume and Average Partition Density [7] and Xie-Beni [8] index, can be used as an objective function of an evolutionary algorithm to find automatically the number of clusters: however, this approach requires running FCM many times for each cluster number selection.

In [9] a new optimization method based on the density of the grid cells is proposed to find the optimal initial cluster centers and number of clusters: this approach can reduce run times in high dimensional clustering.

The K-means algorithm is used in [10] to initialize the centers of the clusters; then, the Partition Coefficient [1,11] and Partition Entropy [12] validity measures are calculated to find the optimal number of clusters. The drawback of this method is the high time consuming and it can be unsuitable for managing massive datasets.

Some authors propose hybrid FCM variations in which metaheuristic approaches are applied to optimize the initialization of the cluster centers. In [13] a kernel FCM algorithm is proposed in which an evolutive method is applied in order to find the initial cluster centers. A Genetic Algorithm (GA) is proposed in [14] to find the optimal initial FCM cluster centers in image segmentation problems. A Particle Swarm Optimization (PSO) algorithm is proposed in [15] to find the optimal FCM initial cluster centers for sentiment clustering. Three hybrid FCM algorithms, based on Differential Evolution, GA, and PSO methods, are proposed in [16] to optimize the cluster centers initialization. These algorithms, while guaranteeing a higher quality of results, require too long execution times; they too are unsuitable for handling high dimensional data.

In this paper we propose a FCM variation in which a new validity index based on the De Luca & Termini Fuzzy Entropy and Fuzzy Energy concepts [17,18] is used to optimize the initialization of the clusters and to find the optimal number of clusters. Our aim is to reach a trade-off between the time consumption and the quality of the clustering algorithm.

Recently a weighted FCM variation based on the De Luca & Termini Fuzzy Entropy is proposed in order to optimize the initialization of the cluster centers in [19]: to initialize the cluster centers the authors execute initially a weighted FCM algorithm, in which the weight assigned to a data point is given by a fuzziness measure obtained by calculating the mean fuzzy entropy of the data point and then the initial cluster centers are found when the mean fuzzy entropy of the clustering converges as well.

The algorithm proposed in [19] is less time-consuming than hybrid algorithms using meta-heuristic approaches, but like the algorithm proposed in [10], it applies an iterative method of pre-processing to initialize cluster centers. Furthermore, it does not detect the optimal number of clusters that must be set in advance.

In the proposed algorithm the validity measure of the quality of clustering based on the fuzzy energy and fuzzy entropy is calculated both in the pre-processing phase to find the optimal initial cluster centers and to determine the optimal number of clusters.. We set the number of clusters and randomly assign cluster centers several times, by choosing as initial cluster centers those for which the clustering validity index is greatest; finally, the FCM algorithm runs. We repeat this process by increasing the number of clusters up to a maximum number. After obtaining the final clusters for each setting of the number of clusters, we choose that one with the largest validity index.

In Section 2, we give a brief review on the Fuzzy Energy and Fuzzy Entropy measures of a fuzzy set and of the FCM algorithm. In Section 3, we introduce the proposed FCM algorithm based on the fuzzy energy and entropy-based validity index. In Section 4, we present several experimental results to demonstrate the features of the proposed index by applying to FCM. In Section 5, we present our conclusions.

## 2. Preliminaries

### 2.1. Fuzzy Energy and Entropy measures

Let  $X$  be a universe of discourse and  $F(X) = \{A: X \rightarrow [0, 1]\}$  be the set of all fuzzy sets defined on  $X$ . Moreover let  $A \in F(X)$ ,  $B \in F(Y)$  be two fuzzy sets defined on the sets  $X$  and  $Y$ , respectively, and let  $R \subseteq F[X \times Y]$  be a fuzzy relation on  $X \times Y$ .

In [17,18] are defined two categories of fuzziness measures of fuzzy sets: fuzzy energy and fuzzy entropy. If  $X = \{x_1, \dots, x_m\}$  is a discrete set with cardinality  $m$ , the energy measure of fuzziness of the fuzzy set  $A \in F(X)$  is given by:

$$E(A) = \sum_{i=1}^m e(A(x_i)) \quad (1)$$

where  $e: [0,1] \rightarrow [0, 1]$  is a continuous function called *fuzzy energy function*. The following restriction are required for the function  $e$ :

- 1)  $e(0) = 0$
- 2)  $e(1) = 1$
- 3)  $e$  is monotonically increasing.

The simplest fuzzy energy function is given by the identity  $e(u) = u$  with  $u \in [0,1]$ . A more general formula for  $e(u)$  is:

$$e(u) = u^p \quad (2)$$

where  $p > 0$  is a positive number.

The minimal value of the fuzzy energy measure is 0 and the maximal value is given by  $E(A) = \text{Card}(X) = m$ , where  $\text{Card}(X)$  is the cardinality of the set  $X$ .

The energy measures of a fuzzy set  $A$  can be seen as a measure of information contained in this fuzzy set. If  $E(A) = 0$  then  $A$  coincides with the empty set; if  $E(A) = m$ , then  $A$  coincides with the set  $X$ .

The entropy measure of fuzziness of the fuzzy set  $A$  is given by:

$$H(A) = \sum_{i=1}^m h(A(x_i)) \quad (3)$$

where  $h: [0,1] \rightarrow [0,1]$  is a continuous function called *fuzzy entropy function*. The following restriction are required for the function  $h$ :

- 4)  $h(1) = 0$
- 5)  $h(u) = h(1-u)$
- 6)  $h$  is monotonically increasing in  $[0, \frac{1}{2}]$
- 7)  $h$  is monotonically decreasing in  $[\frac{1}{2}, 1]$

The simplest fuzzy entropy function is given by:

$$h(u) = \begin{cases} 2u & \text{if } u < \frac{1}{2} \\ 2(1-u) & \text{if } u \geq \frac{1}{2} \end{cases} \quad (4)$$

This fuzzy entropy function has a minimal value 0 when  $u$  is 0 or 1 and a maximum value 1 when  $u = \frac{1}{2}$ .

De Luca and Termini in [21] propose the following fuzzy entropy function:

$$h(u) = \begin{cases} 0 & \text{if } u = 0 \\ -u \cdot \log_2(u) - (1-u) \cdot \log_2(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{if } u = 1 \end{cases} \quad (5)$$

This fuzzy entropy function has a maximum value 1 when  $u = \frac{1}{2}$ ; it is called Shannon's function.

The entropy measures of a fuzzy set  $A$  can be seen as a measure of the fuzziness contained in this fuzzy set. If  $H(A) = 0$  then for each element  $x_i, i = 1, \dots, m, A(x_i) = 0$  or  $A(x_i) = 1$  and  $A$  coincides with a subset of the set  $X$ ; if  $H(A) = m$ , then for each element  $x_i, i = 1, \dots, m, A(x_i) = \frac{1}{2}$  and the fuzziness of  $A$  is maximum

A problem is to find the fuzzy set from a family of fuzzy sets of  $F(X)$  with the highest information content and the lowest fuzziness.

## 2.2. Fuzzy C-means algorithm

Let  $\mathbf{X} = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$  be a set of  $N$  data points in the  $n$ -dimensional space  $\mathbb{R}^n$  where  $x_j = (x_{j1}, \dots, x_{jn})$  and  $\mathbf{V} = \{v_1, \dots, v_C\} \subset \mathbb{R}^n$  be the set of centers of the  $C$  clusters. Let  $\mathbf{U}$  be the  $C \times N$  partition matrix where  $u_{ij}$  is the membership degree of the  $j$ th data point  $x_j$  to the  $i$ th cluster  $v_i$ .

The FCM algorithm [1,2] is based on the minimization of the following objective function:

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^p d_{ij}^2 = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^p \|x_j - v_i\|^2 \quad (6)$$

where  $d_{ij} = \|x_j - v_i\|$  is the Euclidean distance between the center  $v_i$  of the  $i$ th cluster and the  $j$ th object  $x_j$ ,  $p \in [1, +\infty)$  is the fuzzifier parameter (a constant which affects the membership values and defines the degree of fuzziness of the partition). For  $m = 1$ , FCM become an hard C-means clustering; the more  $m$  tends towards  $+\infty$  the more the fuzziness level of the clusters grows.

By considering the following constraints:

$$\sum_{i=1}^C u_{ij} = 1 \quad \forall j \in \{1, \dots, N\} \quad (7)$$

$$0 < \sum_{j=1}^N u_{ij} < N \quad \forall i \in \{1, \dots, C\} \quad (8)$$

and applying the Lagrange multipliers, we obtain the following solutions for (1) :

$$v_i = \frac{\sum_{j=1}^N u_{ij}^p x_j}{\sum_{j=1}^N u_{ij}^p} \quad i \in \{1, \dots, C\} \quad (9)$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m-1}}} \quad i \in \{1, \dots, C\}, \quad j \in \{1, \dots, N\} \quad (10)$$

An iterative process is proposed in [2] as follows: initially the membership degrees are assigned randomly; in each iteration the cluster centers are calculated by (4), then the membership degree components are calculated by (5). The iterative process stops at the  $t$ th iteration when

$$|\mathbf{U}^{(t)} - \mathbf{U}^{(t-1)}| < \varepsilon \quad i = 1, \dots, C; \quad j = 1, \dots, N \quad (11)$$

where  $\varepsilon > 0$  is a parameter assigned a priori to stop the iteration process and

$$|\mathbf{U}^{(t)} - \mathbf{U}^{(t-1)}| = \max_{\substack{i=1, \dots, C \\ j=1, \dots, N}} \{|u_{ij}^{(t)} - u_{ij}^{(t-1)}|\} \quad i = 1, \dots, C; \quad j = 1, \dots, N \quad (12)$$

The pseudocodes of the FCM algorithm is shown below.

Algorithm: FCM

**Input:** Dataset  $X = \{x_1, \dots, x_N\}$

**Output:** Cluster centers  $V = \{v_1, \dots, v_C\}$ ; Partition matrix  $U$

**Arguments:** number of clusters  $C$ ; fuzzifier  $p$ ; stop iteration threshold  $\varepsilon$

1. Set  $p, \varepsilon, C$  to the values of the arguments
2. Initialize randomly the partition matrix  $U$
3. **Repeat**
4.       Calculate  $v_i \quad i = 1, \dots, C$  by using (4)
5.       Calculate  $u_{ij} \quad i = 1, \dots, C \quad j = 1, \dots, N$  by using (5)
6. **Until**  $|U^{(t)} - U^{(t-1)}| > \varepsilon \quad |U^{(t)} - U^{(t-1)}| > \varepsilon$
7. **Return**  $V, U$

### 3. The proposed FCM algorithm based on a Fuzzy Energy and Entropy validity index

Let  $X = \{x_1, \dots, x_N\}$  the set of data point with cardinality  $N$ . We consider the fuzzy set  $A_i \in F(X)$  where  $A_i(x_j) = u_{ij}$  is the membership degree of the  $j$ th data point to the  $i$ th cluster.

We propose a new validity index based on the fuzzy energy and fuzzy energy measures to evaluate the compactness of clusters and the separability among clusters.

By using (1) and (3), respectively, we can evaluate the fuzzy energy and the fuzzy entropy of the  $i$ th cluster, measuring the fuzzy entropy and the fuzzy energy of the fuzzy set  $A_i$ , given by

$$E(A_i) = \frac{1}{N} \sum_{j=1}^N e(A(x_j)) = \frac{1}{N} \sum_{j=1}^N e(u_{ij}) \quad i = 1, 2, \dots, C \quad (13)$$

$$H(A_i) = \frac{1}{N} \sum_{j=1}^N h(A(x_j)) = \frac{1}{N} \sum_{j=1}^N h(u_{ij}) \quad i = 1, 2, \dots, C \quad (14)$$

where the fuzzy energy and entropy are normalized dividing them by the cardinality  $N$  of the dataset.

Fuzzy energy (13) measures the quantity of information contained in the  $i$ th cluster and fuzzy entropy (14) measures the fuzziness of the  $i$ th cluster, namely the quality of the information therein contained.

For example, a cluster with low fuzzy entropy has low fuzziness, so it is compact; however, if it also has a low fuzzy energy, then the information which it contains is low. Hence, even if compact, a very small number of data points will belong to this cluster and this could be due to the presence of noise or outliers in the data. Moreover, a cluster with a high value of fuzzy entropy has high fuzziness and low compactness.

We set the function (2) as fuzzy energy function, where  $p$  is given by the value of the fuzzifier parameter. The fuzzy entropy function  $h(u)$  is given by the Shannon function (5).

We measure the energy and the entropy of the clustering given by the averages of the energy and entropy of the  $C$  clusters.:

$$E = \frac{1}{C} \sum_{i=1}^C E(A_i) \quad (15)$$

and

$$H = \frac{1}{C} \sum_{i=1}^C H(A_i) \quad (16)$$

Respectively. The proposed validity index, called Partition Energy-Entropy (for short, PEH) is given by the difference between the energy and the entropy of the clustering.

$$PEH = E - H \quad (17)$$

This index varies in the range (-1,1); the optimal clustering is the one that maximizes PEH; the greater the value of PEH, the more the clusters are compact and well separated from each other.

We propose a new algorithm, called PEHFCM, in which the PEH index is used to initialize the cluster centers and to find the optimal number of clusters.

In addition to the fuzzifier and iteration error threshold parameters, further arguments of the algorithm are the maximum number of clusters  $C_{max}$  and the number of random selections of initial  $C$  clusters  $S_{max}$ . The PEHFCM algorithm is composed of FOR loop in which the number of clusters is initially set to 2 and then cyclically iterated until the  $C_{max}$  value. In each cycle,  $S_{max}$  sets of cluster centers are initially selected, for each of which the PEH index is calculated. The optimal set of initial cluster centers is the one for which the PEH indicator is maximum. Subsequently, a variation of the FCM algorithm is performed, called FCMV, which, unlike FCM, uses the set of initial cluster centers  $V^0$  as a further argument instead of setting it randomly. Finally, the PEH index of the final clustering is calculated.

The PEHFCM algorithm returns the optimal number of  $C^*$  clusters and the respective sets of cluster centers  $V^*$  and partition matrix  $U^*$  corresponding to the highest PEH validity index.

Below we show the algorithm PEHFCM and the algorithm FCMV called by PEHFC.-

---

Algorithm: *PEHFCM*

---

**Input:** Dataset  $X = \{x_1, \dots, x_N\}$

**Output:** Cluster centers  $V = \{v_1, \dots, v_C\}$ ; Partition matrix  $U$ , optimal number of clusters  $C^*$

**Arguments:** max num of clusters  $C_{max}$ ; max num of random selections of the initial cluster centers  $S_{max}$ , fuzzifier  $p$ ; stop iter threshold  $\varepsilon$

1. Set  $p, \varepsilon, C_{max}$  to the values of the arguments
  2.  $C^* := 1$
  3.  $PEH^* := -1$
  4. **For**  $c = 2$  **to**  $C_{max}$
  5.     **For**  $k = 1$  **to**  $S_{max}$
  6.         Set randomly the partition matrix  $U$
  7.         Calculate the value of the cluster centers  $v_i$  by (9)  $i = 1, \dots, c$
  8.         Calculate  $E$  by (15)
  9.         Calculate  $H$  by (16)
  10.          $PEH := E - H$
  11.         **If**  $PEH > PEH^*$  **Then**
  12.              $V^0 := V_k$
  13.         **End if**
  14.      $PEH^* := -1$
  15.     Set randomly
  16.     Call FCMV( $X, V^0, p, \varepsilon, C$ )
  17.     Calculate  $E$  by (15)
  18.     Calculate  $H$  by (16)
  19.      $PEH := E - H$
  20.     **If**  $PEH > PEH^*$  **Then**
  21.          $PEH^* := PEH$
  22.          $C^* := c$
  23.          $V^* := V$
  24.          $U^* := U$
  25.     **End if**
  26. **Next**  $c$
-

---

27. **Return**  $\mathbf{U}^*, \mathbf{V}^*, C^*$

---

Algorithm: FCMV

---

**Input:** Dataset  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  Initial cluster centers  $\mathbf{V}^0 = \{\mathbf{v}_1^0, \dots, \mathbf{v}_C^0\}$

**Output:** Cluster centers  $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_C\}$ ; Partition matrix  $\mathbf{U}$

**Arguments:** Initial cluster centers  $\mathbf{V}^0 = \{\mathbf{v}_1^0, \dots, \mathbf{v}_C^0\}$ ; number of clusters  $C$ ; fuzzifier  $p$ ; stop iteration threshold  $\varepsilon$

1. Set  $p, \varepsilon, C$  to the values of the arguments
  2.  $\mathbf{v}_i := \mathbf{v}_i^0 \quad i = 1, \dots, C$
  3. Calculate  $u_{ij} \quad i = 1, \dots, C \quad j = 1, \dots, N$  by using (5)
  4. **Repeat**
  5.     Calculate  $\mathbf{v}_i := \quad i = 1, \dots, C$  by using (4)
  6.     Calculate  $u_{ij} \quad i = 1, \dots, C \quad j = 1, \dots, N$  by using (5)
  7. **Until**  $|U^{(t)} - U^{(t-1)}| > \varepsilon \quad |\mathbf{U}^{(t)} - \mathbf{U}^{(t-1)}| > \varepsilon$
  8. **Return**  $\mathbf{V}, \mathbf{U}$
- 

To measure the performances of the proposed algorithm we compare the results with the ones obtained by applying FCM and applying our method by using other well-known validity indices: Partition Coefficient (PC) [1,11], Partition Entropy (PE) [12], Fukuyama and Sugeno (FS) [20] and Xie-Beni (XB) [8], described below.

The PC validity index is given by the formula

$$PC(C) = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 \quad \frac{1}{C} \leq PC(C) \leq 1 \quad (18)$$

It measures the crispness of the clusters. The value  $C^*$  is obtained when PC is maximum.

The PE validity index is given by:

$$PE(C) = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^N u_{ij} \log_2 u_{ij} \quad 0 \leq PE(C) \leq \log_2 C \quad (19)$$

It measures the mean fuzziness of the clusters; the optimal number of clusters  $C^*$  is obtained when PE is minimum.

The FS validity index is given by:

$$FS(C) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|^2 - \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \|\mathbf{x}_j - \bar{\mathbf{v}}\|^2 \quad (20)$$

where  $\bar{\mathbf{v}}$  is the average of the cluster centers. The first term in (20) measures the compactness of the clusters, the other one the separability among the same clusters. The optimal number of clusters  $C^*$  is obtained when FS is maximum.

The XB validity index is given by the formula:

$$XB(C) = \frac{\sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|^2}{N \cdot \min_{\substack{i=1, \dots, C \\ j=1, \dots, N}} \left( \|\mathbf{x}_j - \mathbf{v}_i\|^2 \right)} \quad (21)$$

The numerator measures the compactness of the clusters, the denominator indicates the separability between clusters. The optimal number of clusters  $C^*$  is obtained when XB assumes the minimum value.

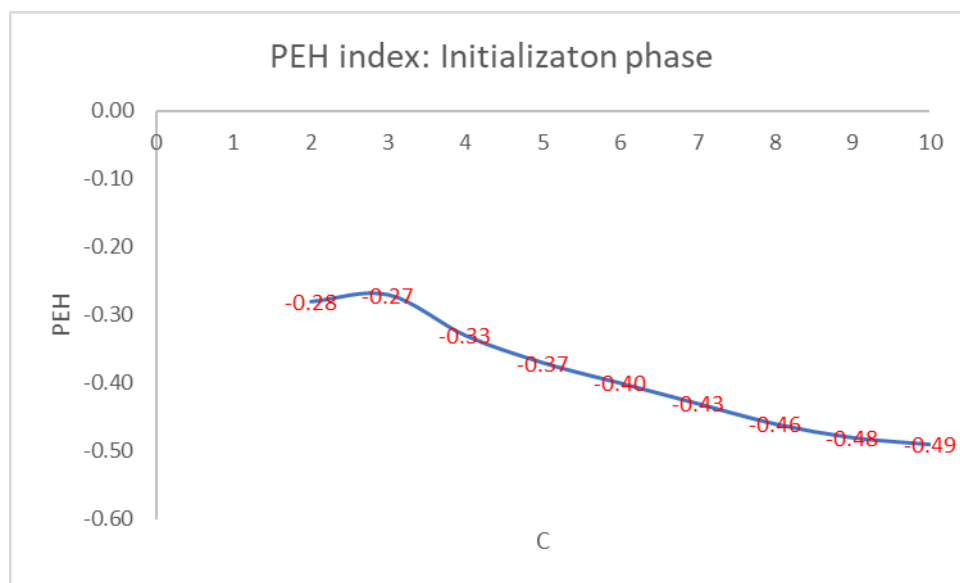
We complete our comparisons by comparing our method with hybrid metaheuristic algorithms.

The comparison tests are performed on well-known UC Irvine (UCI) machine learning classification datasets (<http://archive.ics.uci.edu/ml/datasets.html>)). We measure the quality of the results in terms of accuracy, precision, recall and F1-score [21,22].

#### 4. Results

We show the results obtained on a set of over 40 classification UCI machine learning datasets. In all experiments we used an Intel core i5 3.2 GHz processor,  $m = 2$ ,  $\varepsilon = 0.01$  and  $S_{max} = 100$ .

For brevity, we only show in detailed the results obtained on the well-known Iris flower dataset. This dataset contains 150 data points with 4 features given by the length and the width of the sepals and petals measured in centimeters; 50 data points are classified as belonging to the type of IRIS flower Iris Setosa, 50 data points to the type Iris Versicolor and 50 data points to the type Iris Virginica. Only the class Iris Setosa is linearly separable from the other two, which are not linearly separable. We set the max number of clusters  $C_{max}$  to 10. In Fig. 1 we show the values of the PEH index of the best initial cluster centers obtained for each setting of the number of clusters.

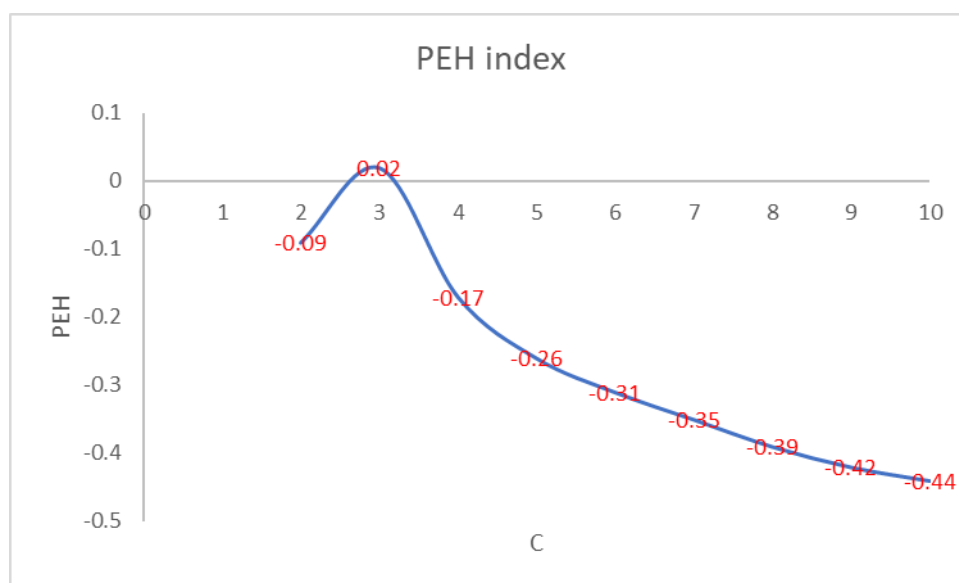


**Figure 1.** PEH index calculated in the initialization phase by varying the number of clusters.

As can be seen from Fig. 1, among the maximum values of the PEH index obtained by varying the number of clusters, the highest is obtained by setting  $C = 3$ .

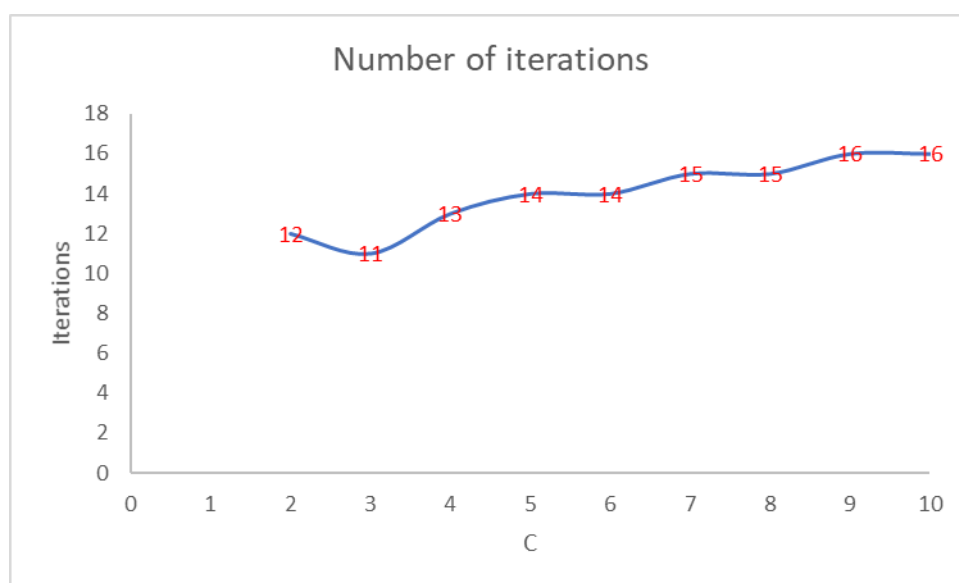
Fig.2 show the PEH index of the final clustering measured for each number of clusters. The highest value (0.02) is obtained for  $C = 3$ , then the optimal number of clusters is 3.





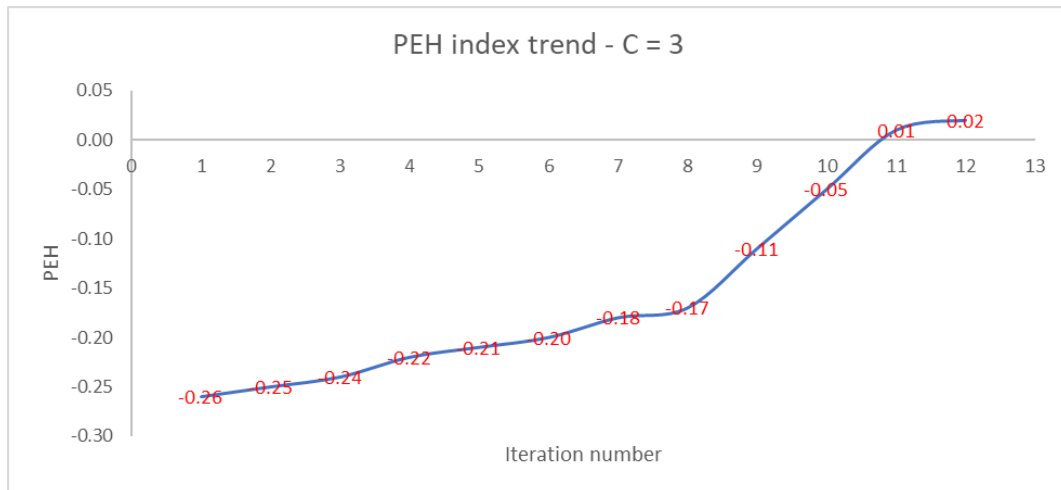
**Figure 2.** PEH index of the final clustering by varying the number of clusters.

Fig.2 shows that the number of iterations increases as the PEH value of the initial clustering decreases. Fig.3 shows the trend of the number of iterations necessary to reach the convergence by varying the number of clusters. The least number of iterations 12 is obtained for  $C = 3$ .



**Figure 3.** Number of iterations of the algorithm by varying the number of clusters.

Like the PEH index of the final clustering, the number of iterations increases as the PEH value of the initial clustering decreases. In Fig. 4 we show the trend of the PEH in any iteration for  $C = 3$ .



**Figure 4.** Trend of PEH with respect to the iteration number for  $C = 3$ .

The PEH index increase slightly, then increases rapidly after the 8th iteration and reaches a plateau at the 12th iteration. To compare the performances of the PEH index with the ones of the PC, PE, FS and XB validity indices.

Table 1 shows the optimal number of clusters found using the validity index, the number of iterations necessary to the convergence and the running time.

**Table 1.** Dataset Iris: validity indices comparison of number of iterations and running time.

Method	Number of clusters	Iterations	Running Time (s)
FCM + PC	2	15	0.158
FCM + PE	2	15	0.157
FCM + FS	3	13	0.138
FCM + XB	3	13	0.135
PEHFCM	3	11	<b>0.103</b>

The best results are obtained by executing PEHFCM with respect to FCM+PC and FCM+PA (resp., FCM + FS and FCM + XB) when the optimal number of clusters obtained is 2 (resp., 3). In both cases the least number of iterations and the shortest execution time are achieved using PEHFCM. In addition, we compare the results obtained by executing PEHFCM with the ones obtained via the Entropy-based weighted FCM algorithm (EwFCM) [19] and the metaheuristic PSOFCM proposed in [15].

Table 2 shows the running time, the accuracy, precision, recall and F1-Score obtained by executing FCM+FS, FCM+XB, PEHFCM, EwFCM and PSOFCM.

**Table 2.** Dataset Iris: Comparisons of running time and classification performances.

Index	FCM+FS	FCM+XB	PEHFCM	EwFCM	PSOFCM
<b>Running time (s)</b>	0.138	0.135	0.103	0.112	0.134
<b>Accuracy</b>	94.22%	94.67%	96.00%	96.44%	96.44%
<b>Precision</b>	91.33%	92.00%	94.00%	94.67%	94.67%
<b>Recall</b>	91.34%	92.00%	94.01%	94.66%	94.67%
<b>F1 Score</b>	91.33%	92.00%	94.00%	94.66%	94.67%

The results in Table2 show that the best classification performances are given by EwFCM and PSOFCM. PEHFCM has the shortest running time but classification performances comparable with EwFCM and PSOFCM.

These results are confirmed testing other UCI machine datasets. Now we present the results obtained on the wine dataset. This dataset is given by 178 data points having 13 features; each data

point represents an Italian wine derived from a specific crop and their features provide information on its chemical composition. The dataset is partitioned in three classes, corresponding to three crops.

In Table 3 we show the results obtained by considering the five validity indices.

**Table 3.** Dataset Wine: validity indices comparison of number of iterations and running time.

Method	Number of clusters	Iterations	Running Time (s)
FCM + PC	2	17	0.166
FCM + PE	2	17	0.168
FCM + FS	3	16	0.152
FCM + XB	3	16	0.151
PEHFCM	3	12	0.116

Even in this case, PEHFCM provides the best number of iterations and running time.

Table 4 shows the running time and the classification performances of all the compared algorithms.

**Table 4.** Dataset Wine: Comparisons of running time and classification performances.

Index	FCM+FS	FCM+XB	PEHFCM	EwFCM	PSOFCM
<b>Running time (s)</b>	0.152	0.151	0.116	0.137	0.145
<b>Accuracy</b>	90.16%	90.34%	92.91%	93.50%	93.81%
<b>Precision</b>	85.89%	86.03%	88.67%	90.04%	90.61%
<b>Recall</b>	86.09%	86.18%	88.61%	90.48%	90.84%
<b>F1 Score</b>	85.99%	86.10%	88.64%	90.26%	90.72%

Also here the results obtained on the dataset Wine show that PEHFCM provides the shortest execution time and classification performances comparable to those obtained by using EwFCM and PSOFCM. We summarize the results obtained on all the classification UCI machine learning datasets used in our tests, calculating:

- the *mean percent of gain (or loss) of running time*. If  $T_c$  is a running time calculated by running a FCM-based method and  $T_{CPEH}$  is the one calculated with PEHFCM, this index is given by the average of the percentage of  $(T_c - T_{CPEH})/T_{CPEH}$ . This value is equal to 0 for PEHFCM,
- the *mean percentage gain (or loss) of a classification index*. If  $I_c$  is a classification index value obtained by running a FCM-based method and  $I_{CPEH}$  is the one obtained with PEHFCM, this index is given by the average of the percentage of  $(I_c - I_{CPEH})/I_{CPEH}$ . This value is equal to 0 for PEHFCM.

If the value of a summarized index is positive, then executing the algorithm we obtain a gain in term of running time or of the classification index analyzed; conversely, we get a loss if that value is negative. In Table 5 we show these results.

**Table 5.** Average percentage of gain of running time and classification indices.

Index	FCM+FS	FCM+XB	PEHFCM	EwFCM	PSOFCM
<b>Percentage gain of running time</b>	-39.16%	-39.01%	0.00%	-28.06%	-31.57%
<b>Percentage gain of accuracy</b>	-3.14%	-3.11%	0.00%	+0.91%	+0.93%
<b>Percentage gain of precision</b>	-3.28%	-3.25%	0.00%	+1.74%	+1.77%
<b>Percentage gain of recall</b>	-3.22%	-3.26%	0.00%	+1.71%	+1.74%
<b>Percentage gain of F1 Score</b>	-3.25%	-32.5%	0.00%	+1.72%	+1.76%

The results in Table 5 shows that PEHFCM provides the best running time; the running times measured executing the other FCM-based algorithms is more 28% than longer than the one obtained by running PEHFCM. The gain of accuracy, precision, recall and F1-score obtained executing EwFCM and PSOFM is less than 2%.

## 5. Conclusions

We propose a variation of FCM in which a validity index based on the fuzzy energy and fuzzy entropy of the clustering, in order to find an optimal initialization of the cluster centers and the optimal number of clusters.

The proposed method represents a trade-off between the running time and the clustering performances: it aims to overcome the problems of initializing cluster centers and setting the number of clusters a priori, without, at the same time, requiring long execution times due to the pre-processing phase necessary to optimize the initialization of cluster centers.

The results of experimental tests applied on well-known UCI machine learning classification datasets show that the PEHFCM algorithm provides shorter running times than the EwFCM and PSOFM algorithms, which use an optimization method based on fuzzy entropy and a metaheuristic method PSO-based to determine the initial cluster centers, respectively. Furthermore, PEHFCM provides classifier performance comparable to EwFCM and PSO FCM. In the future we intend to test PEHFCM on high-dimensional datasets, in which it is essential to guarantee high performances both in terms of quality of results and execution times.

**Author Contributions:** Conceptualization, F.D.M. and S.S.; methodology, F.D.M. and S.S.; software, F.D.M. and S.S.; validation, F.D.M. and S.S.; formal analysis, F.D.M. and S.S.; investigation, F.D.M. and S.S.; resources, F.D.M. and S.S.; data curation, F.D.M. and S.S.; writing—original draft preparation, F.D.M. and S.S.; writing—review and editing, F.D.M. and S.S.; visualization, F.D.M. and S.S.; supervision, F.D.M. and S.S..

**Funding:** This research received no external funding.

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

## References

1. Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms. *Kluwer Academic Publishers*: Norwell, MA, USA, **1981**.
2. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203.
3. Chen, J.; Qin, Z.; Jia, J. A Weighted Mean Subtractive Clustering Algorithm. *Inf. Technol. J.* **2008**, *7*, 356–360.
4. Yang, Q.; Zhang, D.; Tian, F. A initialization method for Fuzzy C-means algorithm using Subtractive Clustering. In Proceedings of the Third International Conference on Intelligent Networks and Intelligent Systems, Shenyang, China, 1–3 November **2010**, Volume 10, pp. 393–396.
5. Campello, R.J.G.B.; Hruschka E.R. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems* **2006**, *157*, 2858–2875.
6. Everitt, B.S.; Landau, S.; Leese, M., Stahl, D. *Cluster Analysis*, fifth edition, Arnold, Paris, **2011**, 343 pp., ISBN: 978-0-470-74991-3.
7. Gath, I.; Geva, A. B. Unsupervised optimal fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11* (7), 773–781.
8. Xie X. L.; Beni I. G. A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1991**, *13*, 841–847.
9. Zou, K.-Q.; Wang, Z.-P.; Pei, S.-J.; Hu, M. A New Initialization Method for Fuzzy c-Means Algorithm Based on Density. In *Advances in Multimedia, Software Engineering and Computing Vol.2*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, **2009**, Volume 54, pp. 547–553.
10. Hayet, D.; Hamza, C. ; Mostefa, B.; Nadjette, D. H. Initialization Methods for K-means and Fuzzy Cmeans Clustering Algorithm, International Conference on Automatic, Télécommunication, and Signals, Annaba, Algeria, November **2015**, 6 pp., doi: 10.13140/RG.2.1.3971.4968.
11. Trauert, E. On the meaning of Dunn's partition coefficient for fuzzy clusters, *Fuzzy Sets and Systems*, **1988**, *25*, 217–242.
12. Bezdek, J.C. Cluster validity with fuzzy sets. *Journal of Cybernetics*, **1973**, *3*, 58–73.

13. Ding, Y.; Fu, X. Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm. *Neurocomputing* **2016**, *188*, 233–238.
14. Li, D.; Han, Z.; Zhao, J. A Novel Level Set Method with Improved Fuzzy C-Means Based on Genetic Algorithm for Image Segmentation. In Proceedings of the 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM); Chengdu, China, 10–11 August 2017, pp. 151–157.
15. Siringoringo, R.; Jamaluddin, J. Initializing the Fuzzy C-Means Cluster Center with Particle Swarm Optimization for Sentiment Clustering. In *Proceedings 1st International Conference of SNIKOM 2018*; IOP Publishing: Bristol, UK, **2019**, Volume 1361, p. 012002.
16. Franco, D.G.D.B.; Steiner, M.T.A. Clustering of solar energy facilities using a hybrid fuzzy c-means algorithm initialized by metaheuristics. *J. Clean. Prod.* **2018**, *191*, 445–457.
17. De Luca, A.; Termini, S. *Entropy and energy measures of fuzzy sets*. In *Advances in Fuzzy Set Theory and Applications*; Gupta, M.M., Ragade, R.K., Yager, R.R., Eds.; North-Holland: Amsterdam, The Netherlands, **1979**; pp. 321–338.
18. De Luca, A.; Termini, S. A definition of non-probabilistic entropy in the setting of fuzzy sets theory. *Inf. Control*, **1972**, *20*, 301–312.
19. Cardone, B.; Di Martino, F. A Novel Fuzzy Entropy-Based Method to Improve the Performance of the Fuzzy C-Means Algorithm. *Electronics*, **2020**, *9*(4), 554, 11 pp.
20. Fukuyama Y.; Sugeno M. A new method of choosing the number of clusters for the fuzzy C-means method, *Proceedings of Fifth Fuzzy Systems Symposium*, Kobe, **1989**, pp. 247–250.
21. Ming, T.K. *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I.; Springer: Berlin/Heidelberg, Germany, **2011**; ISBN 978-0-387-30164-8.
22. Tharwat, A. Classification assessment methods. *Appl. Comput. Informatics* **2018**, *13*.
23. Dunn J. C.. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, **1973**, *3* (3): 32–57.