Iranian Journal of Fuzzy Systems
Volume 20, Number 7, (2023), pp. 55-70
Original paper

Unsupervised feature selection: A fuzzy multi-criteria decision-making approach

M. B. Dowlatshahi[1] and A. Hashemi[2]

[1,2]*Department of Computer Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran*
dowlatshahi.mb@lu.ac.ir, Hashemi.am@fe.lu.ac.ir

Abstract

Feature selection (FS) has shown remarkable performance in decreasing the dimensionality of high-dimensional datasets by selecting a good subset of features. Labeling high-dimensional data can be expensive and time-consuming as labeled samples are not always available. Therefore, providing effective unsupervised FS methods is essential in machine learning. This article provides a fuzzy multi-criteria decision-making method for unsupervised FS in which an ensemble of unsupervised FS rankers is utilized to evaluate the features. These methods are aggregated based on a fuzzy TOPSIS method. This is the first time a fuzzy multi-criteria decision-making approach has been used for an FS problem. Multiple comparisons are made to show the optimality and effectiveness of the proposed strategy against multiple competing FS methods. Our approach regarding two classification metrics, F-score and accuracy, appears superior to comparable strategies. Also, it is performing so swiftly.

Keywords: Unsupervised feature selection, fuzzy TOPSIS, fuzzy multi-criteria decision-making, high-dimensional data, ensemble of feature selection methods.

1 Introduction

Endless amounts of information are produced daily from different sources, including the world wide web and social networks [2, 30]. Generally speaking, redundant and irrelevant features exist in datasets. The redundant features do not provide new information and are nearly a blend of other features. The irrelevant features also do not influence accomplishing the result. Learning algorithms suffer from a decrease in learning accuracy and get time-consuming due to the presence of these features in data. FS is an effective dimensionality reduction technique to reduce these effects in the learning process by eliminating irrelevant and redundant features and keeping the most practical ones. This strategy can diminish data's time complexity, dimensionality, and computational cost and increment the classification performance [11, 18, 38]. Supervision and search strategy can be categorized as two kinds of FS strategies [12]. FS methods are known as three main classes considering the supervision point of view: supervised [23], semi-supervised [10], and unsupervised [37] are three classes of FS algorithms according to the supervision perspective viewpoint.

In supervised FS methods, features are evaluated on their ability to predict class labels. Single-label (SL) [23] and multi-label (ML) data [33] can be assessed by supervised FS methods. If each training sample has more than one class label, it is called ML data. Otherwise, it is called SL data [22]. Due to the data's lack of class label information, the relationship between features is the measure of assessing features by unsupervised methods [37]. Datasets processed by semi-supervised methods contain a small fraction of labeled samples among unlabeled samples [10, 30].

Filter, wrapper, and embedded are three kinds of FS methods based on the search strategy. In filter methods, the features are evaluated before the learning process by statistical criteria, including ReliefF, Symmetric Uncertainty (SU), Fisher Score, and Mutual Information (MI) [19, 34]. A learning algorithm is the main criterion for evaluating possible feature subsets in wrapper strategies [14]. Embedded strategies use the advantages of filter and wrapper methods by embedding a learning algorithm in the FS process [1, 19].

Corresponding Author: M. B. Dowlatshahi
Received: March 2023; Accepted: Invited paper.
https://doi.org/10.22111/IJFS.2023.7630

Labeling data samples is time-consuming, laborious, and costly in real-world applications. Also, these methods are not biased and decrease the possibility of overfitting since supervised methods may not be able to deal with new data classes. Thus, unsupervised FS methods are essential since labeled samples are not always available. Effective methods for unsupervised FS should be provided based on the cases mentioned. Thus, we aim to provide an FS algorithm using an ensemble of filter strategies in this article [6, 25, 33].

The reason for using the ensemble strategy in this article is that the robustness of FS methods can be improved by using this approach. Ensemble techniques have improved FS methods in the last decade. On the other hand, in unsupervised FS, the features are evaluated solely according to their relationships since class information is not available. Therefore, using one criterion alone may lead to a local optimum, and using several unsupervised evaluation criteria can lead to a better result. Because each feature is evaluated based on multiple criteria, and as a result, the selected set can be considered as the opinions of several experts [21, 25, 26, 27]. Based on these reasons, an ensemble of FS methods is provided in this paper. A fuzzy multi-criteria decision-making modeling is used to do this aggregation process.

Real-world decision-making problems usually cannot be concluded by evaluating only one criterion to create an optimal decision because these problems are complex and inconsistent. In fact, considering only one indicator of the problem in the decision-making process can lead to a local optimum. A more attractive approach is to simultaneously pay attention to all indicators related to the problem. Multi-criteria decision-making specializes in developing and implementing decision-making tools and methods to solve complex decision-making problems. In general, multi-criteria decision-making seeks to evaluate multiple alternatives based on the opinion of a group of decision-makers or several different criteria [13, 29]. FS is an NP-hard problem that researchers have sought to provide effective methods for in past decades. Recently, multi-criteria decision-making methods have been applied to different FS problems [20, 24, 25, 26]. This modeling has achieved significant results in this field. However, this strategy has not been used for unsupervised FS problems.

This paper provides a method using multi-criteria decision-making modeling for unsupervised FS. A fuzzy multi-criteria decision-making approach is applied to combine multiple FS methods. Decision-making problems usually face goals and limitations that cannot be determined precisely in many real-world problems. After that, Professor Lotfi Askarzadeh and Bellman [5] published the article, and for the first time, they defined fuzzy sets in multi-criteria decision-making where the decision makers make judgments based on uncertain and incomplete information and knowledge. In this article, we have transformed the scores that the FS methods assign to each feature into fuzzy numbers because, in real-world problems, decisions are usually made based on this point of view.

In the proposed method, three unsupervised feature ranking methods are considered as our decision-making criteria. Then, we transformed the ranks of features into scores. These scores are transformed into fuzzy triangular numbers, and then the fuzzy TOPSIS method is utilized to achieve the final feature rankings.

The main contributions of the proposed method are listed below:

1. A decision-making council can perform better than just one decision-making indicator. We use multiple objectives for decision-making in unsupervised streaming FS for better performance.

2. The unsupervised FS problem is modeled as a fuzzy multi-criteria decision-making problem for the first time.

3. Fuzzy TOPSIS method and fuzzy triangular numbers are utilized to solve the ensemble of unsupervised FS in a fuzzy system.

4. An ensemble of filter rankers is used in the proposed method. Thus, it is so swift.

Different real-world datasets and seven competitive unsupervised FS algorithms have been employed to evaluate the proposed method's performance. Classification accuracy and F-score are the primary metrics in estimating the performance of algorithms. The reported results show the effectiveness of the proposed method. All the algorithms are also compared based on their running time, and the proposed method reported short run-time in this comparison. The remainder of the article is structured as follows: The literature review is presented in Section 2. Section 3 describes fuzzy triangular numbers, fuzzy multi-criteria decision-making, and the fuzzy TOPSIS method. A detailed explanation of the proposed algorithm is provided in Section 4, and Section 5 contains the experimental results. Finally, the conclusions of this article are presented in Section 6.

2 Related works

Some unsupervised FS methods are reviewed in this section. Lin et al. [32] proposed an embedded-based unsupervised FS technique using reliable local structure-preserving and orthogonal basis clustering. This method is called OCLSP and has utilized two procedures to discover the cluster separation process by orthogonal basis clustering and preserving the local structure of data using graph regularization. Han et al. [17] proposed an autoencoder-based FS using a two-layered autoencoder network. Row-sparse regularization of the hidden layer weights is the main metric in evaluating features.

Dependence Guided Unsupervised FS (DGUFS) [16] is an unsupervised FS method based on an embedded strategy using a clustering approach. In addition to selecting features, DGUFS is conducted to divide the data in a typical manner. Also, it increases the interdependence between the desired features, cluster labels, and original data. The gaps between the selected features and the original data are filled by the learned cluster labels. As a result, two dependency-driven terms are proposed for the model.

Zhu et al. proposed the FRFSO method [43] based on non-convex regularized self-representation. The representation of features is performed based on a linear combination of features in this embedded-based unsupervised FS strategy. Iterative reweighted least square and augmented Lagrangian methods are used to solve the non-convex regularized self-representation model and ensure convergence in this method.

SCAFS is proposed by Huang et al. [28] based on utilizing many random subspaces to enhance the performance in unsupervised FS tasks using an embedded search strategy. First, some primary feature partitions with similarly sized random subspaces are generated by a balanced subspace randomization scheme. Several k-nearest-neighbor graphs are then constructed on these subspaces. Then, the Laplacian scores of the features in each subspace are in terms of the locality to preserve power. The extracted feature score vectors for different subspaces in different base feature partitions are combined into a full score vector for all features. It considers structural information in different subspaces and dramatically enhances unsupervised FS methods.

Xie et al. [40] provide three unsupervised FS methods using a filter strategy. The main rules in these three methods are cosine similarity and standard deviation. These methods are SCAFS (Standard deviation and Anti-Cosine similarity-based FS), SCRFS (Standard deviation and Reciprocal Cosine similarity-based Feature Selection), and SCEFS (Standard deviation and Exponent Cosine similarity-based Feature Selection). In order to quickly recognize features that have both independence and high distinctness from the original features, they put all features into a two-dimensional space with distinctness as the x coordinate and independence as the y coordinate. The feature's contribution to classification is the rectangle bounded by the feature coordinate lines and axes . It is referred to as feature value, where they can be sorted based on these values.

Beiranvand et al. [4] provide an unsupervised FS using a bipartite graph matching model and the principal component analysis (PCA) concept. This method uses PCA to generate orthogonal features. Then the orthogonal and original features are sides of a bipartite graph where the similarity between these two kinds of features is considered as the weights of graph edges between them. Finally, the Hungarian algorithm is performed to achieve the maximum match.

3 Preliminaries

3.1 Fuzzy triangular numbers

A fuzzy set $\bar{a}$ in a universal set X is determined by a membership function $\mu\bar{a}\left(x\right)$ in which it maps each $x \in$ X to a real-value number in the range $[0,1]$. The value of $\mu\bar{a}\left(x\right)$ is called the membership degree of x in $\bar{a}$. The closer this value is to one, the higher the degree of membership [3].

A fuzzy triangular number is shown by $\bar{a} = (a_1, a_2, a_3)$ in which it shows the membership function $\mu\bar{a}\left(x\right)$ on fuzzy triangular number $\bar{a}$ [3, 35]. A fuzzy triangular system is illustrated in Figure 1.

Based on Figure 1, $\mu\bar{a}\left(a_2\right) = 1$ and $\mu\bar{a}\left(a_1\right) = \mu\bar{a}\left(a_3\right) = 0$. The fuzzy numerical rule can be defined as follows [35]:

$$\mu\bar{a}\left(x\right) = \begin{cases} 0 & x \leq a_1 \\ \dfrac{x-a_1}{a_2-a_1} & a_1 \geq x \geq a_2 \\ \dfrac{x-a_2}{a_3-a_2} & a_2 \geq x \geq a_3 \\ 0 & x \geq a_3 \end{cases} \tag{1}$$

If we assume $\bar{a} = (a_1, a_2, a_3)$ and $\bar{b} = (b_1, b_2, b_3)$ as two triangular numbers, the distance between them is obtained through the following relationship [35]:

$$d\left(\bar{a}, \bar{b}\right) = \sqrt{\frac{1}{3}\left[\left(a_1-b_1\right)^2 + \left(a_2-b_2\right)^2 + \left(a_3-b_3\right)^2\right]}. \tag{2}$$

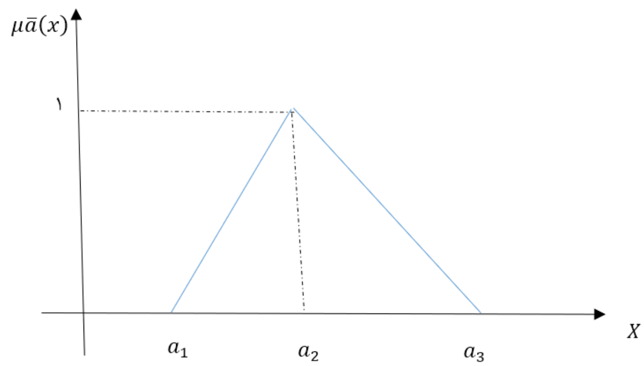

Figure 1: Fuzzy triangular system

Also, the product of these two numbers is defined as follows [35]:

$$\bar{a}.\bar{b} = (a_1, a_2, a_3) \,.\, (b_1, b_2, b_3) = (a_1.b_1, a_2.b_2, a_3.b_3) \,. \tag{3}$$

The transformation scales transform linguistic modifications into fuzzy numbers, usually used in multi-criteria decision-making problems from a 1 to 9 scoring system for alternatives and criteria. These intervals provide a uniform representation from 1 to 9 for the fuzzy triangular numbers representing the five linguistic ratings. Consider a multi-criteria decision-making scoring system where one verbally scores alternatives and criteria. Table 1 shows the five types of linguistic scoring and their corresponding fuzzy numbers [35].

Table 1: Fuzzy scoring for linguistic variables

Alternative evaluation	Weights of criteria	Fuzzy number
Very weak	Very low	$(\mathbf{1}, \mathbf{1}, \mathbf{3})$
Weak	Low	$(\mathbf{1}, \mathbf{3}, \mathbf{5})$
Fair	Medium	$(\mathbf{3}, \mathbf{5}, \mathbf{7})$
Good	High	$(\mathbf{5}, \mathbf{7}, \mathbf{9})$
Very good	Very high	$(\mathbf{7}, \mathbf{9}, \mathbf{9})$

3.2 Fuzzy multi-criteria decision-making and fuzzy TOPSIS

Multi-criteria decision-making is one of the important branches of decision-making issues. This approach is used in problems where several options are measured based on different indicators. Each of these indicators has a different effect on the final result. For example, consider a person looking to buy a house and offer him various options. To buy a house, he considers various indicators like the size of the house, the location, the number of rooms, and the price of the house. Finally, based on these indicators, the desired option is reached, which is the best option by establishing a tradeoff between the indicators [20, 25].

In fuzzy multi-criteria decision-making, the alternatives are assessed according to linguistic values [29]. Fuzzy TOPSIS is a well-known method in this category in which the best alternative determines by the closest distance to the fuzzy positive ideal solution (FPIS) and the most significant distance to the fuzzy negative ideal solution (FNIS). FNIS and FPIS are the worst and best solutions, respectively. Now we describe the steps of the fuzzy TOPSIS method [35, 36].

In an MCDM problem, a weight vector and a decision matrix should be initialized first. Therefore, the ratings of d alternatives construct the decision matrix (X) according to criteria h, as follows:

$$W = \ [w_1, w_2, \ldots, w_h] \,. \tag{4}$$

Sometimes, different criteria have different degrees of impact on the decision-making process. Therefore, the weights of the criteria should be defined as a vector:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1h} \\ x_{21} & x_{22} & \cdots & x_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \dots & x_{dh} \end{bmatrix}. \tag{5}$$

where in the fuzzy TOPSIS method, each value in the decision matrix (X) and weight vector (W) is represented by fuzzy triangular numbers, as in Table 1. For example, $x_{11} = \left(x^1{}_{11}, x^2{}_{11}, x^3{}_{11}\right)$ and $w_1 = \left(w^1{}_1, w^2{}_1, w^3{}_1\right)$.

Considering that in an MCDM process, the values of each criterion may be in different ranges, this issue can have a negative impact on decision-making. Thus, at the beginning of the process, the decision matrix should be normalized so that the values of all indicators be on the same level and the normalized decision matrix $\bar{R} = [\bar{r}_{ij}]$ obtains. The following equations are used to normalize the decision matrix.

$$\bar{r}_{ij} = \left(\frac{x^1_{ij}}{c^*_{ij}}, \frac{x^2_{ij}}{c^*_{ij}}, \frac{x^3_{ij}}{c^*_{ij}}\right) \quad \text{where} \quad c^*_{ij} = \max(x^1_{ij}, x^2_{ij}, x^3_{ij}). \tag{6}$$

Equation 6 is used for beneficial criteria, and Equation 7 for loss criteria.

$$\bar{r}_{ij} = \left(\frac{x^1_{ij}}{c^*_{ij}}, \frac{x^2_{ij}}{c^*_{ij}}, \frac{x^3_{ij}}{c^*_{ij}}\right) \quad \text{where} \quad c^*_{ij} = \min(x^1_{ij}, x^2_{ij}, x^3_{ij}). \tag{7}$$

Equation 8 shows the structure of the normalized decision matrix.

$$\bar{R} = \begin{bmatrix} \bar{r}_{11} & \bar{r}_{12} & \cdots & \bar{r}_{1h} \\ \bar{r}_{21} & \bar{r}_{22} & \cdots & \bar{r}_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{r}_{d1} & \bar{r}_{d2} & \dots & \bar{r}_{dh} \end{bmatrix}. \tag{8}$$

As we mentioned earlier, different criteria have different degrees of impact on selecting the alternatives in MCDMs. To influence these weight values, they are multiplied in the normalization matrix. The following equation is used to achieve this goal.

$$\bar{V} = [\bar{v}_{ij}] \quad \text{where} \quad \bar{v}_{ij} = \bar{r}_{ij} \times \; w_j = \; \left(\bar{r}^1_{ij}, \bar{r}^2_{ij}, \bar{r}^3_{ij}\right) \times \left(w^1{}_j, w^2{}_j, w^3{}_j\right) = \left(\bar{r}^1_{ij} \times w^1{}_j, \bar{r}^2_{ij} \times w^2{}_j, \bar{r}^3_{ij} \times w^3{}_j\right). \tag{9}$$

Equation 10 shows the structure of the weighted decision matrix.

$$\bar{V} = \begin{bmatrix} \bar{v}_{11} & \bar{v}_{12} & \cdots & \bar{v}_{1h} \\ \bar{v}_{21} & \bar{v}_{22} & \cdots & \bar{r}_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{v}_{d1} & \bar{v}_{d2} & \dots & \bar{r}_{dh} \end{bmatrix}. \tag{10}$$

After achieving the weighted decision matrix, the FPIS and FNIS are calculated using the following relations:

$$A^+ = \left(a^+_1, a^+_2, \dots\dots\dots, a^+_d\right) \text{ where } a^+_i = \left(\max\left\{\bar{v}_{ij}\right\}, \max\left\{\bar{v}_{ij}\right\}, \max\left\{\bar{v}_{ij}\right\}\right) \quad i = 1, 2, \dots, d \quad j = 1, 2, \dots, h, \tag{11}$$

$$A^- = \left(a^-_1, a^-_2, \dots\dots\dots, a^-_d\right) \text{ where } a^-_i = \left(\min\left\{\bar{v}_{ij}\right\}, \min\left\{\bar{v}_{ij}\right\}, \min\left\{\bar{v}_{ij}\right\}\right) \quad i = 1, 2, \dots, d \quad j = 1, 2, \dots, h. \tag{12}$$

Then the distance of every alternative from FPIS and FNIS is computed using Equation 2 as follows:

$$A^+_i = \sum_{j=1}^{h} d\left(\bar{v}_{ij}, A^+_j\right), \;\; A^-_i = \sum_{j=1}^{h} d\left(\bar{v}_{ij}, A^-_j\right), \tag{13}$$

Then, the closeness coefficient (CC) is computed using the following equation:

$$\bar{C}_i = \frac{A_i^-}{A_i^- + A_i^+}\ ,\ \text{where } i = 1, 2, \ldots.., d. \tag{14}$$

We can rank the alternatives based on the value of CC in descending order. Algorithm 1 shows the steps of the fuzzy TOPSIS method in Figure 2.

Algorithm 1: Fuzzy TOPSIS

Input: Weight vector: $\boldsymbol{W} = [\boldsymbol{w_1}.\boldsymbol{w_2} \ldots.\boldsymbol{w_h}]$, $\boldsymbol{d} \times \boldsymbol{h}$ Decision matrix X

Output: ranks of alternatives

1. Normalize the decision matrix using Eq.6 and Eq.7

2. Obtain the weighted decision matrix using Eq.9

3. Compute FNIS and FPIS using Eq.11 and Eq.12

4. Calculate the distance of each alternative from FNIS and FPIS using Eq.13

5. Compute CC for each alternative using Eq.14

6. Sort the alternatives using CC values in descending order

Figure 2: Pseudo code of fuzzy TOPSIS

4 Proposed method

The proposed algorithm is described in detail in this section. The proposed algorithm is implemented based on an ensemble of filter strategies for unsupervised FS called UFS-FMCDM. A fusion is conducted using a fuzzy multi-criteria decision-making approach using the fuzzy TOPSIS method. The procedure of the proposed method is presented in algorithm 2 in Figure 3.

Proposing effective FS methods for unsupervised data is a critical challenge in machine learning because the data samples are not always labeled due to the cost of labeling and the time-consuming nature of this process. This paper uses a council of three unsupervised FS methods to evaluate features better. An unsupervised data M, including n data samples and d features, is presented as the following matrix:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & m_{22} & \cdots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \ldots & m_{nd} \end{bmatrix}. \tag{15}$$

Three unsupervised FS methods, including CFS (Correlation-based FS), Laplacian score, and LLCFS [41] are used for the fusion process. In CFS, first, the linear relationship between the features is captured using the Pearson correlation coefficient:

$$\text{correlation } (X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}, \tag{16}$$

where $Cov(X, Y)$ refers to the covariance between variable X and Y. Also σ_X and σ_Y are the standard deviation of X and Y, respectively. In the CFS method, the minimum correlation of each feature is calculated, and then the features are ranked based on these values in ascending order.

In LLCFS, a learning-based clustering is used for the FS procedure. Also, the Laplacian score [42] is computed using a neighborhood graph. These two methods can obtain the ranking of features. Thus, we have three rankings of

Algorithm 2: Proposed UFS-FMCDM

Input: M (n× $\boldsymbol{d}$ data), Weights (W)

Output: Features ranking

1. R1= Computer feature ranking 1 using the CFS method

2. R2= Computer feature ranking 2 using the Laplacian score method

3. R3= Computer feature ranking 3 using the LLCFS method

4. Construct the rank matrix (R)

5. Convert the rank matrix to scores by subtracting all rank values from the number of features plus one (d+1) and obtaining matrix $\bar{\boldsymbol{M}}$

6. Normalize the matrix $\bar{\boldsymbol{M}}$ by min-max method and achieve matrix $\boldsymbol{M}^*$

7. Convert the values of the matrix $\boldsymbol{M}^*$ and vector W into fuzzy triangular numbers and obtain matrix $\boldsymbol{X}$

8. Perform fuzzy TOPSIS method in algorithm 1

9. Final feature ranking

Figure 3: Pseudo code of UFS-FMCDM

features based on the three mentioned methods, and the decision matrix is formed as follows:

$$R = \begin{bmatrix} R1\,(1) & R2\,(1) & R3\,(1) \\ R1\,(2) & R2\,(2) & R3\,(2) \\ \vdots & \vdots & \vdots \\ R1\,(d) & R2\,(d) & R3\,(d) \end{bmatrix}. \tag{17}$$

where $R1\,(2)$ indicates the ranking that the first FS method assigned to feature 2. These steps are the first four steps of the proposed method.

Now, to consider the problem as maximization, we subtract all the ranks from the number of features plus one:

$$\bar{M} = (d+1) - R. \tag{18}$$

It means that the feature with rank 2 gets a score equal to $d-2$. This is done in step 5. To model the problem into a fuzzy system, the scores in $\bar{M}$ are normalized by the min-max method in the range $[0,1]$ and then these values and the weights are converted to fuzzy triangular numbers based on the rules in Table 2. This is done in Steps 7 and 8. The weight of our criteria in this problem is set as follows: CFS (0.4), LLCFS (0.3), and Laplacian-score (0.3).

Table 2: The rules of converting scores to triangular numbers

Scores or Weights	Fuzzy number
Less or than 0.2	$(\mathbf{1},\mathbf{1},\mathbf{3})$
More than 0.2 and less or than 0.4	$(\mathbf{1},\mathbf{3},\mathbf{5})$
More than 0.4 and less or than 0.6	$(\mathbf{3},\mathbf{5},\mathbf{7})$
More than 0.6 and less or than 0.8	$(\mathbf{5},\mathbf{7},\mathbf{9})$
More than 0.8 and less or than 1	$(\mathbf{7},\mathbf{9},\mathbf{9})$

Now that the decision matrix and the weight vector are constructed, the fuzzy TOPSIS method in algorithm 1 is applied to rank the features. Let's use a simple example to understand the proposed method better. Let's assume we have 5 features in our dataset and want to use the proposed method to rank them. The ranks of features based on three FS methods are obtained as follows:

$$R = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 4 & 5 & 4 \\ 5 & 4 & 5 \\ 2 & 3 & 3 \end{bmatrix}.$$

Now the rank matrix is converted to scores in Step 5.

$$\bar{M} = 6 - \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 4 & 5 & 4 \\ 5 & 4 & 5 \\ 2 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 4 & 5 \\ 3 & 5 & 4 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \\ 4 & 3 & 3 \end{bmatrix}.$$

The scores are then normalized by the min-max method:

$$M^* = \begin{bmatrix} 1 & 0.75 & 1 \\ 0.5 & 1 & 0.75 \\ 0.25 & 0 & 0.25 \\ 0 & 0.25 & 0 \\ 0.75 & 0.5 & 0.5 \end{bmatrix}.$$

Now the values are converted to fuzzy triangular numbers based on Table 2:

$$X = \begin{bmatrix} (7,9,9) & (5,7,9) & (7,9,9) \\ (3,5,7) & (7,9,9) & (5,7,9) \\ (1,3,5) & (1,1,3) & (1,3,5) \\ (1,1,3) & (1,3,5) & (1,1,3) \\ (5,7,9) & (3,5,7) & (3,5,7) \end{bmatrix}.$$

$$W = [(1,3,5)\,,(1,3,5)\,,(1,3,5)].$$

In the first step of the fuzzy TOPSIS method, the fuzzy triangular numbers are normalized:

$$\bar{R} = \begin{bmatrix} (0.777,1,1) & (0.555,0.777,1) & (0.777,1,1) \\ (0.429,0.6,1) & (0.777,1,1) & (0.555,0.777,1) \\ (0.2,0.6,1) & (0.333,0.333,1) & (0.2,0.6,1) \\ (0.333,0.333,1) & (0.2,0.6,1) & (0.333,0.333,1) \\ (0.555,0.777,1) & (0.429,0.6,1) & (0.429,0.6,1) \end{bmatrix}.$$

$$W = [(0.2,0.6,1)\,,(0.2,0.6,1)\,,(0.2,0.6,1)]\;.$$

Since all our criteria are beneficial in this problem, just Equation 5 is used to normalize the values. For example, for calculating r_{11} the following procedure is conducted:

$$c^*_{11} = \max{(3,5,7)} = 7.$$

$$r_{11} = \left(\tfrac{3}{7},\tfrac{5}{7},\tfrac{7}{7}\right) = (0.429,0.6,1).$$

Then the weighted decision matrix is computed:

$$\bar{V} = \begin{bmatrix} (0.155,0.6,1) & (0.111,0.466,1) & (0.155,0.6,1) \\ (0.085,0.36,1) & (0.155,0.6,1) & (0.111,0.466,1) \\ (0.4,0.36,1) & (0.066,0.199,1) & (0.4,0.36,1) \\ (0.066,0.199,1) & (0.4,0.36,1) & (0.066,0.199,1) \\ (0.111,0.466,1) & (0.085,0.36,1) & (0.085,0.36,1) \end{bmatrix}.$$

For example, for calculating $\bar{v}_{11}$ the following procedure is conducted:

$$\bar{v}_{11} = (0.777,1,1)\,.\,(0.2,0.6,1) = (0.777 \times 0.2, 1 \times 0.6, 1 \times 1) = (0.155,0.6,1).$$

Now the FPIS and FNIS should be obtained. FPIS and FNIS represent the maximum and minimum values of each column.

$$A^{+} = \left[a_1^{+}\left(1,1,1\right), a_2^{+}\left(1,1,1\right), a_3^{+}\left(1,1,1\right)\right].$$

$$A^{-} = \left[a_1^{-}\left(0.066, 0.066, 0.066\right), a_2^{-}\left(0.066, 0.066, 0.066\right), a_3^{-}\left(0.066, 0.066, 0.066\right)\right].$$

For example, for calculating a_1^{+} , the maximum value among all the values in column 1 is 1 thus

$$a_1^{+} = a_1^{+}\left(1,1,1\right).$$

In the next step, the distance of each feature is computed based on 5 against the FPIS and FNIS.

$$S1 = \begin{bmatrix} 0.5397 & 0.5987 & 0.5397 \\ 0.6446 & 0.5397 & 0.5987 \\ 0.5065 & 0.7104 & 0.5065 \\ 0.7104 & 0.5065 & 0.7104 \\ 0.5987 & 0.6446 & 0.6446 \end{bmatrix}.$$

$$S2 = \begin{bmatrix} 0.6232 & 0.5872 & 0.6232 \\ 0.5654 & 0.6232 & 0.5872 \\ 0.5973 & 0.5447 & 0.5973 \\ 0.5447 & 0.5973 & 0.5447 \\ 0.5872 & 0.5654 & 0.5654 \end{bmatrix}.$$

$$S^{+} = \begin{bmatrix} 1.6781 \\ 1.7830 \\ 1.8156 \\ 1.8615 \\ 18615 \end{bmatrix}, \qquad S^{-} = \begin{bmatrix} 1.8336 \\ 1.7758 \\ 1.7292 \\ 1.7074 \\ 1.7074 \end{bmatrix}.$$

For example, S_1^{+} is computed as follows. The distance of three triangular numbers of feature 1 is computed against the FPIS:

$$S_{11}^{+} = \sqrt{\tfrac{1}{3}\left[\left(0.155-1\right)^2 + \left(0.6-1\right)^2 + \left(1-1\right)^2\right]} = \sqrt{\tfrac{1}{3}\left[0.7740+0.16\right]} = \sqrt{0.2913} = 0.5397.$$

$$S_{12}^{+} = 0.5987, \qquad S_{13}^{+} = 0.5397.$$

$$S_1^{+} = 0.5397 + 0.5987 + 0.5987 = 1.6781.$$

In the final step, CC value is computed as follows for all features.

$$CC = \begin{bmatrix} 0.5251 \\ 0.4990 \\ 0.4878 \\ 0.4874 \\ 0.0001 \end{bmatrix}.$$

For feature 1, CC is computed as follows:

$$CC = \frac{1.8336}{1.6781 + 1.8336} = 0.5251.$$

Based on CC, the features are ordered as follows:

$$\text{Ranks} = \begin{bmatrix} F1 \\ F2 \\ F3 \\ F4 \\ F5 \end{bmatrix}.$$

5 Experimental studies

Experimental settings and datasets are described in this section. Then, the Experimental results are presented to demonstrate the comparison between the proposed method and competing methods.

5.1 Datasets

Ten real-world datasets are utilized to evaluate the proposed method's performance against competitive methods. Table 3 provides details of the datasets. These datasets include Glioma, Yale, ORL, WarpAR10P, WarpPIE10P [1] , Sorlie [2] , Khan [31], Chiaretti [8], Christensen [9], and West [39].

Table 3: The main characteristics of the datasets.

Dataset	Number of instances	Number of features	Number of classes	Domain
Glioma	50	4435	4	Biological Data
Sorlie	85	457	5	Microarray Data
WarpAR10P	130	2400	10	Face Image Data
WarpPIE10P	210	2420	10	Face Image Data
Yale	165	1024	15	Face Image Data
Khan	63	2309	4	Microarray Data
Chiaretti	217	1414	3	Microarray Data
Christensen	214	1414	2	Image Data
West	49	7129	2	Microarray Data
ORL	400	1025	40	Face Image Data

5.2 Results

Seven unsupervised FS methods are utilized to compare with the proposed method, including AEFS [17], OCLSP [32], DGUFS [16], FRFSO [43], SCAFS[28], SCAFS [40], SCRFS [40], and SCEFS [40]. These methods are reviewed in Sections 2.

The values of the parameters of all competing algorithms were fixed considering the recommendations of the following articles for all the simulations. The k-nearest neighbor (KNN) is used to achieve the classification performance of all algorithms based on accuracy and F-score metrics. The number of nearest neighbors in the KNN classifier is set to 10. Ten various subsets of features are chosen in the range $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ to assess the performance of the FS algorithms. All the simulations are performed on a windows server 2013-64 bit machine with 64 GB Ram and Intel (R) Xeon (R) Gold 6254 CPU with 16 3.10 GHz processors using MATLAB R2018a.

The experimental results are validated based on hold-out validation. 70 percent of all training instances are randomly selected as the training set in each separate run, and the rest are set as the test set. The reported results of each algorithm are the average of 30 separate runs. The user in the proposed method determines the number of selected features. Figure 4 and Figure 5 show the reported results based on ten datasets. Two well-known metrics, Accuracy and F-measure [7] are the main metrics in evaluating the classification performance of methods.

A statistical comparison is performed to evaluate the significance of the results according a nonparametric Friedman test [15]. The significance level of this statistical test was considered to be 0.05. Table 4 and Table 5 show the obtained p-values of pair-wise statistical comparisons, and the overall win/tie/loss of the proposed method against the competing methods are presented in Table 6. In tables Table 4 and Table 5, the (+) sign shows the superiority of the proposed method over the competing methods statistically. Also, (=) and (−) signs indicate the equal or worst performance of the proposed method comparing that method. Also, Table 7 reports the run-time of algorithms.

5.3 Discussions

In this paper, we have utilized fuzzy multi-criteria decision-making modeling to evaluate the features in unsupervised datasets. The proposed method outperforms competing methods considering classification metrics based on the obtained results and statistical tests. The proposed method is filter-based, and there are three filter-based methods in competitive methods, including SCAFS, SCEFS, and SCRFS. The other four competitive methods are embedded. Embedded methods are expected to provide higher accuracy than filter methods due to the learning algorithm used in the FS

[1] https://jundongl.github.io/scikit-feature/datasets.html
[2] https://search.r-project.org/CRAN

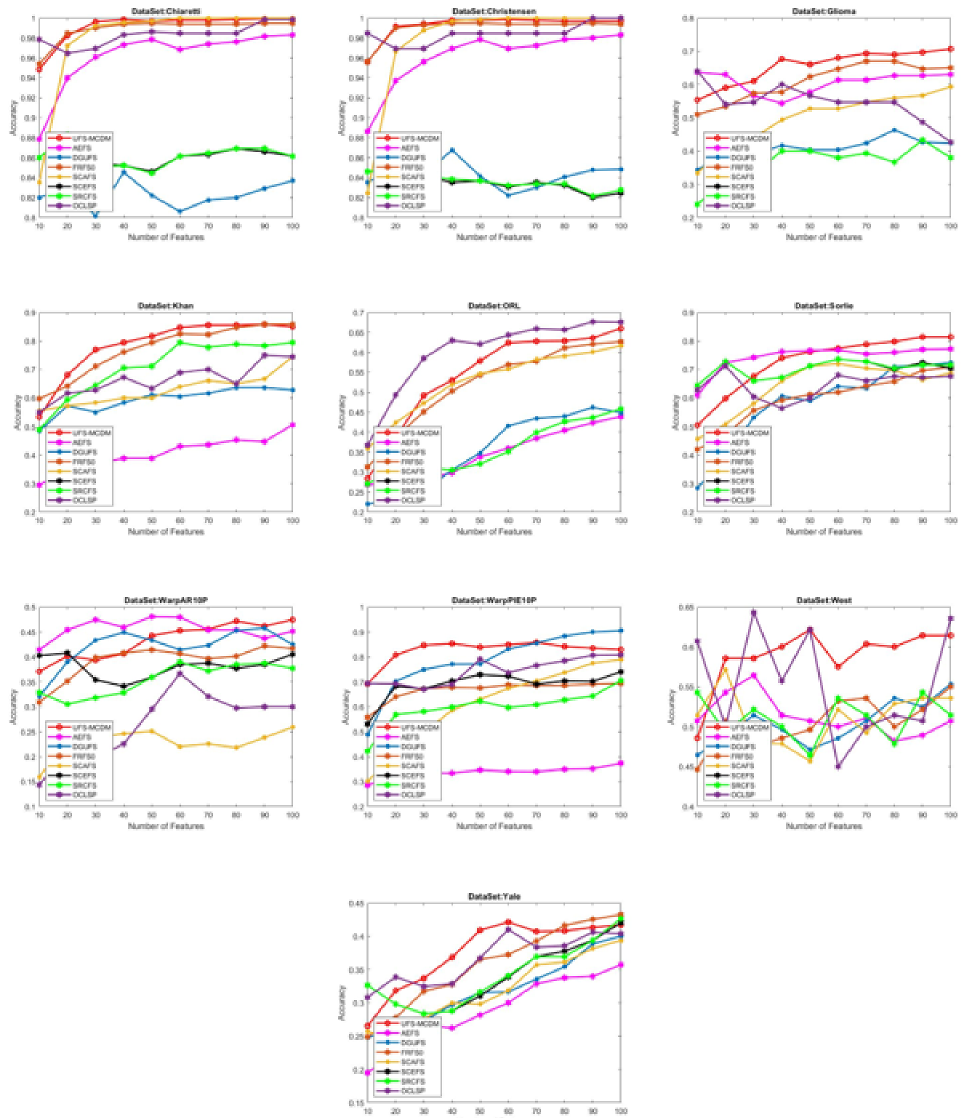

Figure 4: The results based on classification accuracy

Table 4: The obtained p-values by the Friedman test for the accuracy metric

UFS -MCDM against	AEFS	DGUFS	FRFSO	SCAFS	SCEFS	SCRFS	OCLSP
Glioma	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0153(+)
Sorlie	0.0005(+)	0.0018(+)	0.0377(+)	0.0018(+)	0.0005(+)	0.0005(+)	0.0833(=)
WarpAR10P	0.0005(-)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)
WarpPIE10P	0.0051(+)	0.0005(+)	0.0004(+)	0.0005(+)	0.5959(=)	0.2111(=)	0.0005(+)
Yale	0.0005(+)	0.0056(+)	0.0056(+)	0.0133(+)	0.0018(+)	0.0018(+)	0.4884(=)
Khan	0.0377(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0051(+)
Chiaretti	0.0003(+)	0.0003(+)	0.0009(+)	1.0000(=)	0.0003(+)	0.0003(+)	0.0011(+)
Christensen	0.0003(+)	0.0003(+)	0.0010(+)	0.0143(+)	0.0003(+)	0.0003(+)	0.0038(+)
West	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	1.0000(=)
ORL	0.0005(+)	0.0005(+)	0.0133(+)	0.0008(+)	0.0005(+)	0.0005(+)	0.4795(=)

Figure 5: The results based on F-measure

Table 5: The obtained p-values by the Friedman test for the F-measure metric

UFS -MCDM against	AEFS	DGUFS	FRFSO	SCAFS	SCEFS	SCRFS	OCLSP
Glioma	0.0005(+)	0.0005(+)	0.0005(-)	0.0005(+)	0.0005(+)	0.0005(+)	0.0056(+)
Sorlie	0.0005(+)	0.0018(+)	0.0153(+)	0.0018(+)	0.0005(+)	0.0005(+)	0.0833(=)
WarpAR10P	0.0005(-)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)
WarpPIE10P	0.0377(+)	0.0005(+)	0.0056(+)	0.0005(+)	0.7290(=)	0.7290(=)	0.0005(+)
Yale	0.0018(+)	0.0153(+)	0.1659(=)	0.0377(+)	0.0153(+)	0.0153(+)	0.2987(=)
Khan	0.0377(+)	0.0005(+)	0.1659(=)	0.0005(+)	0.0005(+)	0.0005(+)	0.0377(+)
Chiaretti	0.0004(+)	0.0004(+)	0.0106(+)	1.0000(=)	0.0004(+)	0.0004(+)	0.1441(=)
Christensen	0.0003(+)	0.0003(+)	0.0010(+)	0.0143(+)	0.0003(+)	0.0003(+)	0.0130(+)
West	0.0018(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)	0.0005(+)
ORL	0.1659(=)	0.0005(+)	0.0377(+)	0.0833(=)	1.0000(=)	1.0000(=)	0.0377(-)

Table 6: The overall win/tie/loss based on the Friedman test

UFS -MCDM against	AEFS	DGUFS	FRFSO	SCAFS	SCEFS	SCRFS	OCLSP
Glioma	9/0/1	10/0/0	10/0/0	9/1/0	9/1/0	9/1/0	6/4/0
Sorlie	9/0/1	10/0/0	7/2/1	8/2/0	8/2/0	8/2/0	6/3/1
ORL	18/0/2	20/0/0	17/2/1	17/3/0	17/3/0	17/3/0	12/7/1

Table 7: The run-time of algorithms (seconds)

Dataset	UFS-MCDM	AEFS	DGUFS	FRFSO	SCAFS	SCEFS	SCRFS	OCLSP
Glioma	0.44	0.11	2.40	0.21	1.51	1.46	1.49	2364.70
Sorlie	0.12	0.22	0.25	0.85	0.02	0.02	0.02	6.50
WarpAR10P	0.47	1.78	2.93	35.85	0.30	0.30	0.31	206.36
WarpPIE10P	0.82	2.05	4.41	38.52	0.34	0.34	0.35	183.11
Yale	0.28	0.90	0.80	6.06	0.07	0.07	0.12	38.12
Khan	0.25	0.11	2.40	0.21	0.34	0.34	0.34	241.95
Chiaretti	0.53	1.41	0.43	21.74	0.15	0.15	0.16	27.60
Christensen	0.53	1.41	0.43	21.74	0.15	0.14	0.14	29.19
West	0.76	5.36	2.67	398.54	3.12	3.14	3.14	8847.40
ORL	0.94	0.89	4.66	6.18	0.08	0.08	0.09	21.24

process. Because the filter methods only evaluate the features without evaluation in the learning environment. However, it can be seen that the proposed method outperforms these methods.

This is the first time the unsupervised FS problem is modeled as a multi-criteria decision-making problem. In unsupervised FS, due to the lack of class labels, it is challenging to make decisions about features, and different methods usually have variance in the results and provide inferior results in some datasets. Therefore, combining these methods and considering them as experts can reduce the variance and stabilize the results. On the other hand, each FS method examines the features from one aspect, and considering each method alone can lead to local optimality. Therefore, combining these methods can lead to a better result, as seen in the results obtained in the article.

In this article, a fuzzy approach is used in multi-criteria decision-making. The concept of fuzzy triangular numbers is used to express the judgment of our experts. The reason for using this fuzzy approach is based on the fact that in an unsupervised FS problem, due to the absence of a class label, we only consider an estimate of how different a feature is from other features, which is not an accurate criterion for evaluating features in this space. Because there may be different features in the dataset that are not compelling. Therefore, we decided to put these values in different intervals and check the values of each interval vaguely and non-deterministically in the fuzzy environment to test our hypothesis.

Also, the execution time of all methods is shown in Table 7. The proposed method is not the fastest based on these values, especially compared to filter-based methods. Therefore, in each dataset, three FS algorithms are also implemented in addition to the proposed algorithm. However, the proposed method is swift and has been implemented in all datasets in less than one second.

6 Conclusions

A new FS algorithm is proposed in this article called UFS-MCDM for unsupervised learning. UFS-MCDM is an ensemble of filter-based FS algorithms that combines three methods using a fuzzy multi-criteria decision-making approach. At first, unlabeled data samples are delivered to the three FS methods that can determine the feature's importance as our experts. Then the ratings are delivered to the fuzzy TOPSIS method to achieve the final ranking of features based on the judgment of the FS methods. The efficiency and optimality of the proposed method are visible based on the obtained results. For future works, we plan to investigate the proposed algorithm in other FS problems, including semi-supervised and multi-label classification. On the other hand, this algorithm can be used in various machine learning applications, and we are looking to test this algorithm in different applications.

Acknowledgement

The authors wish to express their appreciation for several excellent suggestions for improvements in this paper made by the referees.

References

[1] N. AlNuaimi, M. M. Masud, M. A. Serhani, N. Zaki, *Streaming feature selection algorithms for big data: A survey*, Applied Computing and Informatics, **18** (2022), 113-135.

[2] H. Bayati, M. B. Dowlatshahi, A. Hashemi, *MSSL: A memetic-based sparse subspace learning algorithm for multi-label classification*, International Journal of Machine Learning and Cybernetics, **13**(11) (2022), 3607-3624.

[3] B. Bede, L. Stefanini, *Generalized differentiability of fuzzy-valued functions*, Fuzzy Sets and Systems, **230** (2013), 119-141.

[4] F. Beiranvand, V. Mehrdad, M. B. Dowlatshahi, *Unsupervised feature selection for image classification: A bipartite matching-based principal component analysis approach*, Knowledge-Based Systems, **250** (2022), 109085. DOI: 10.1016/j.knosys.2022.109085.

[5] R. E. Bellman, L. A. Zadeh, *Decision-making in a fuzzy environment*, Management Science, **17** (1970), 141-164.

[6] V. Bolón-Canedo, A. Alonso-Betanzos, *Ensembles for feature selection: A review and future trends*, Information Fusion, **52** (2019), 1-12.

[7] G. Chandrashekar, F. Sahin, *A survey on feature selection methods*, Computers and Electrical Engineering, **40** (2014), 16-28.

[8] S. Chiaretti, X. Li, R. Gentleman, A. Vitale, M. Vignetti, F. Mandelli, J. Ritz, R. Foa, *Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival*, Blood, **103**(7) (2004), 2771-2778.

[9] B. C. Christensen, E. A. Houseman, C. J. Marsit, S. Zheng, M. R. Wrensch, J. L. Wiemels, H. H. Nelson, M. R. Karagas, J. F. Padbury, R. Bueno, D. J. Sugarbaker, R. F. Yeh, J. K. Wiencke, K. T. Kelsey, *Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CPG island context*, PLOS Genetics, **5** (2009). DOI: 10.1371/journal.pgen.1000602.

[10] A. Dalvand, M. B. Dowlatshahi, A. Hashemi, *SGFS: A semi-supervised graph-based feature selection algorithm based on the Page-Rank algorithm*, 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 2022.

[11] X. Deng, Y. Li, J. Weng, J. Zhang, *Feature selection for text classification: A review*, Multimedia Tools and Applications, **79** (2019), 3797-3816.

[12] P. Dhal, C. Azad, *A comprehensive survey on feature selection in the various fields of machine learning*, Applied Intelligence, **52** (2022), 4543-4581.

[13] L. Fei, Y. Deng, *Multi-criteria decision making in Pythagorean fuzzy environment*, Applied Intelligence, **50** (2020), 537-561.

[14] V. Feofanov, E. Devijver, M. R. Amini, *Wrapper feature selection with partially labeled data*, Applied Intelligence, **52** (2022), 12316-12329.

[15] M. Friedman, *A comparison of alternative tests of significance for the problem of m rankings*, The Annals of Mathematical Statistics, **11** (1940), 86-92.

[16] J. Guo, W. Zhu, *Dependence guided unsupervised feature selection*, 32nd AAAI Conference on Artificial Intelligence, **2018**, 2018.

[17] K. Han, Y. Wang, C. Zhang, C. Li, C. Xu, *Autoencoder inspired unsupervised feature selection*, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2018.

[18] E. Hancer, B. Xue, M. Zhang, *A survey on feature selection approaches for clustering*, Artificial Intelligence Review, **53** (2020), 4519-4545.

[19] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *MGFS: A multi-label graph-based feature selection algorithm via Page-Rank centrality*, Expert Systems with Applications, **142** (2020), 113024. DOI:10.1016/j.eswa.2019.113024.

[20] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *MFS-MCDM: Multi-label feature selection using multi-criteria decision making*, Knowledge-Based Systems, **206** (2020), 106365. DOI:10.1016/j.knosys.2020.106365.

[21] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach*, Journal of Soft Computing and Information Technology, **50** (2021).

[22] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *A bipartite matching-based feature selection for multi-label learning*, International Journal of Machine Learning and Cybernetics, **12** (2021), 459-475.

[23] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *A pareto-based ensemble of feature selection algorithms*, Expert Systems with Applications, **180** (2021), 115130. DOI:10.1016/j.eswa.2021.115130.

[24] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *VMFS: A VIKOR-based multi-target feature selection*, Expert Systems with Applications, **182** (2021), 115224. DOI:10.1016/j.eswa.2021.115224.

[25] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, *Ensemble of feature selection algorithms: A multi-criteria decision-making approach*, International Journal of Machine Learning and Cybernetics, **13** (2022), 49-69.

[26] A. Hashemi, M. Joodaki, N. Z. Joodaki, M. B. Dowlatshahi, *Ant colony optimization equipped with an ensemble of heuristics through multi-criteria decision making: A case study in ensemble feature selection*, Applied Soft Computing, **124** (2022), 109046. DOI:10.1016/j.asoc.2022.109046.

[27] A. Hashemi, M. R. Pajoohan, M. B. Dowlatshahi, *Online streaming feature selection based on Sugeno fuzzy integral*, 2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), IEEE, Bam, Iran, 2022.

[28] D. Huang, X. Cai, C. D. Wang, *Unsupervised feature selection with multi-subspace randomization and collaboration*, Knowledge-Based Systems, **182** (2019), 104856. DOI:10.1016/j.knosys.2019.07.027.

[29] C. Kahraman, S. C. Onar, B. Oztaysi, *Fuzzy multicriteria decision-making: A literature review*, International Journal of Computational Intelligence Systems, **8** (2015), 637-666.

[30] F. Karimi, M. B. Dowlatshahi, A. Hashemi, *SemiACO: A semi-supervised feature selection based on ant colony optimization*, Expert Systems with Applications, **214** (2023), 119130. DOI:10.1016/j.eswa.2022.119130.

[31] J. Khan, J. S. Wei, M. Ringnér, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, P. S. Meltzer, *Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks*, Nature Medicine, **7** (2001), 1-7.

[32] X. Lin, J. Guan, B. Chen, Y. Zeng, *Unsupervised feature selection via orthogonal basis clustering and local structure preserving*, IEEE Transactions on Neural Networks and Learning Systems, **33**(11) (2021), 6881-6892.

[33] M. Miri, M. B. Dowlatshahi, A. Hashemi, M. K. Rafsanjani, B. B. Gupta, W. Alhalabi, *Ensemble feature selection for multilabel text classification: An intelligent order statistics approach*, International Journal of Intelligent Systems, **37**(12) (2022), 11319-11341.

[34] W. C. Mlambo, *A survey and comparative study of filter and wrapper feature selection techniques*, The International Journal Of Engineering And Science (IJES), **5** (2016), 57-67.

[35] S. Nădăban, S. Dzitac, I. Dzitac, *Fuzzy TOPSIS: A general view*, Procedia Computer Science, **91** (2016), 823-831.

[36] M. M. Salih, B. B. Zaidan, A. A. Zaidan, M. A. Ahmed, *Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017*, Computers and Operations Research, **104** (2009), 207-227.

[37] S. Solorio-Fernández, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, *A review of unsupervised feature selection methods*, Artificial Intelligence Review, **53** (2020), 907-948.

[38] B. Venkatesh, J. Anuradha, *A review of feature selection and its methods*, Cybernetics and Information Technologies, **19** (2019), 3-26.

[39] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. Olson, J. R. Marks, J. R. Nevins, *Predicting the clinical status of human breast cancer by using gene expression profiles*, Proceedings of the National Academy of Sciences USA, **98**(20) (2001), 11462-11467.

[40] J. Xie, M. Wang, S. Xu, Z. Huang, P. W. Grant, *The unsupervised feature selection algorithms based on standard deviation and cosine similarity for genomic data analysis*, Frontiers in Genetics, **12** (2021). DOI:10.3389/fgene.2021.684100.

[41] H. Zeng, Y. M. Cheung, *Feature selection and kernel learning for local learning-based clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33** (2011), 1532-1547.

[42] L. Zhu, L. Miao, D. Zhang, *Iterative Laplacian score for feature selection*, Communications in Computer and Information Science, (2012), 80-87. DOI:10.1007/978-3-642-33506-8_11.

[43] P. Zhu, W. Zhu, W. Wang, W. Zuo, Q. Hu, *Non-convex regularized self-representation for unsupervised feature selection*, Image and Vision Computing, **60** (2017), 22-29.