Shahid Bahonar
University of Kerman

Journal of Mahani Mathematical Research

Print ISSN: 2251-7952 Online ISSN: 2645-4505

JMMR 2012

ANT-EHFS: ANT COLONY OPTIMIZATION EQUIPPED WITH AN ENSEMBLE OF HEURISTICS THROUGH FUZZY LOGIC FOR FEATURE SELECTION

N.Z. Joodaki, M.B. Dowlatshahi ✉, and M. Joodaki

Article type: Research Article

(Received: 18 February 2022, Received in revised form: 29 April 2022)

(Accepted: 11 July 2022, Published Online: 07 August 2022)

Abstract. One of the crucial stages in machine learning in high- dimensional datasets is feature selection. Unrelated features weaknesses the efficiency of the model. However, merging several feature selection strategies is routine to solve this problem, the way to integrate feature selection methods is problematic. This paper presents a new ensemble of heuristics through fuzzy Type-I based on Ant Colony Optimization (ACO) for ensemble feature selection named Ant-EHFS. At first, three feature selection methods are run. Then, the Euclidean Distance between each pair of features is computed as a heuristic (an M×M matrix is constructed), that M is the total of features. After that, a Type-I fuzzy is used individually to address various uncertainty of feature selections and estimate trustworthiness for each feature, as another heuristic. A complete weighted graph based on combining the two heuristics is then built.Finally, ACO is applied to the complete graph for finding features that have the highest relevance together in the features space, which in each ant considers the reliability rate and Euclidean Distance of the destination node together for moving between nodes of the graph. Five and eight robust and well-known ensemble feature selection methods and primary feature selection methods, respectively, have been compared with Ant-EHFS on six high-dimensional datasets to show the proposed method's performance. The results have shown that the proposed method outperforms five ensemble feature selection methods and eight primary feature selections in Accuracy, Precision, Recall, and F1-score metrics.

Keywords: : Ant colony optimization, High-dimensional data, Feature selection, Ensemble feature selection, an ensemble of algorithms, Type-I fuzzy, population-based optimization algorithms.
2020 MSC: 94D05.

1. Introduction

Using high-dimensional data is widespread in different machine learning algorithms, and caused problems like insufficient memory and the time-consuming learning process. Many fields, such as data mining, image processing, and

---

✉ dowlatshahi.mb@lu.ac.ir, ORCID: 0000-0002-4862-0626

DOI: 10.22103/jmmr.2022.19039.1207
Publisher: Shahid Bahonar University of Kerman
How to cite: N.Z. Joodaki, M.B. Dowlatshahi, M. Joodaki, *Ant-EHFS: Ant colony optimization equipped with an ensemble of heuristics through fuzzy logic for feature selection*, J. Mahani Math. Res. 2023; 12(2): 29-56.

bioinformatics, need to find the best features among their high-dimensional data [23].

Dimensionality reduction strategies are used mainly to deal with these problems, and the strategies can be divided into two categories of feature selection and feature extraction. The features shift into a new space in feature extraction methods, but feature selection methods will select the best features based on their importance. In other words, in feature selection methods, select a subset of relevant features and ignore irrelevant features [24]. There are many feature selection methods that each method has its metrics. In other words, for decreasing irrelevant features that negatively affect the performance of models, feature selection strategies can be classified into three categories filter, wrapper, and embedded [29].

Filter methods select features without using a learning algorithm to examine chosen features and their effect on performance during the selection [34]. As mentioned previously, there are many feature selection methods, and each of them considers metrics like information, distance, consistency, similarity, and statistical measures [34]. Here some feature selection methods in the filter, wrapper, and embedded strategies are mentioned briefly.

A strategy via Mutual Information through filter methods for discovering the most useful group of features has been offered by [51]. In this method, three critical points, comprising maximizing relevance, minimizing monotony, and maximizing Complementary Information, have been evaluated for uncovering the most suitable group. Chi-square [58] is based on chi-square statistics and distribution and examines the relevance between features. relieFF [36] aims to identify conditional dependencies between features. Correlation [60], finds a set of independent variables that have positively correlated to the dependent variable but not correlated to other independent variables. Method named RFS [38] detects the cover between features and target variables based on R-value. Maximal Information Coefficient [9] discovers the similarity between features based on the coefficient. There is further information about feature selection methods in [37, 19] .

However, for choosing more relevant features, wrapper methods use a learning model for assessing the selected features by a classification model during the feature selection process [8]. The wrapper methods more positive effect on performance than filter methods. Three critical disadvantages of wrapper methods are the risk of overfitting, computational complexity, and a time-consuming process. The wrapper methods are classified into sequential selections, and heuristic search approaches [48]. The sequential selection methods try to reduce inappropriate features until it gets the best fulfillment. The heuristic search approaches by seeking between features attempt to obtain the most valuable collection of the appropriate features [15, 12].

Regarding the mentioned shortages of filter selection methods and wrapper methods, using embedded methods is expected. The embedded methods employ filter methods and wrapper methods together; in other words, the embedded methods decrease the risk of overfitting and computational complexity and are not a time-consuming process. Recently, many embedded feature selection methods have been proposed, like min, max, mean and medium, plurality vote, Borda voting, and weighted Borda voting; they first use a group of filter methods and then use a learning model to assess selected features [18]]. A set of methods relevant to ensemble feature selection is mentioned in the following. For the classification problem in [13], an aggregate of embedded methods and filter methods was applied to the miRNA dataset. To find a relevant group of features [61] has used a distance-based filter method and then has employed a weighted bootstrapping search method.

To doing combinatorial optimization problems, population-based metaheuristic algorithms have been considering [62]. Evolutionary algorithms such as genetic algorithms [3, 55], gravitational search [14, 16, 17, 62], particle swarm optimization [1, 2], and swarm intelligence-based algorithms like ant colonies [18, 56, 53, 52] are in the population-based metaheuristic algorithms group. For example, the Ant Colony Optimization (ACO) for feature selection in the multi-label dataset was used in [47]; they have adjusted the ACO to gain the most assuring features with the lowest redundancy and highest relevancy with class labels in the multi-label datasets.

To obtain the optimal solution for a complicated optimization problem, it is better to utilize various search strategies. The same search strategy may make Population-based Optimization Algorithms (POA) trapping into one local optimum. Consequently, an ensemble of various strategies could make POA incredibly efficient for complicated optimization problems [59]. Therefore, the ensemble strategy is one of the most suitable ways for the robustness and efficiency of POA [59].

In other words, an ensemble POA with influential and outstanding ensemble components is more likely capable of dealing with different sorts of complicated optimization problems. Furthermore, the ensemble of various strategies with fitting adaptation mechanisms could cause search strategies of different abilities to support each other, thus significantly increasing the performance of a POA [41]. Using a single heuristic in the ACO cannot always significantly deal with POA [59]. Concerning success in [47, 28] for selecting the most relevant features, we were inspired to use the ACO based on an ensemble of heuristic. Further, in the case of using a combination of heuristics to reach a better result, it should note that some feature selection methods suggest an incorrect rank for some features that are different from the majority of ranks and produce noise in ranks which will have a bad effect on heuristics. Concerning that, techniques based on fuzzy logic are more suitable for integrating different feature selection methods and avoiding noise [20, 32, 39, 42, 33]. In other words, fuzzy logic presents extra flexibility for the uncertainty of feature selections when each feature's ranks for different feature selections are uncertain. Accordingly, we have applied fuzzy Type-I to decrease the impact of incorrect ranks and calculate a reliability rate for each feature.

While heuristics have been investigated as successful methods for nearly three decades to help ACO achieve reasonable solutions to different optimization problems, no attempt has been reported to incorporate various heuristics in ACO to create an effective ensemble of heuristics. This paper recommends a novel ensemble feature selection using swarm intelligence Ant colony optimization with an efficient ensemble of heuristics through fuzzy logic. The ACO imitates real ants' behavior; when ants find food, they build some paths between food and their nest, and after that, they choose a path as the best path between food and their nest based on a group agreement. We have adjusted the real ants' behavior to introduce a novel ensemble feature selection in the proposed method. The main contributions of the proposed method are as follows:

- For the first time, applying the ACO algorithm for ensemble feature selection,
- Defining an efficient ensemble of heuristics in ACO through fuzzy logic for finding the highest relevant features,

This paper's design is classified as follows: In Section 2, we studied some corresponding ensemble feature selection methods. Section 3 is provided the fundamental concepts. The offered method is defined in Section 4, and the final outcomes and discussion are proposed in Section 5. Ultimately, in Section 6, a conclusion is given.

2. Related works

As stated above, tackling massive data and high dimensional datasets is vital in various fields. In the same vein, reducing the dimensions of datasets by finding the more optimal subset of features is crucial [32, 50, 31, 4, 45, 44]. Recently, combining several feature selection methods has been regular. In this section, some of the new and base methods are proposed.

Generally, heterogeneous and homogeneous techniques are two primary groups of the ensemble feature selection method. In heterogeneous strategies, various sorts of feature selection methods are applied, while in the homogeneous, feature selection methods have the same type [14]. For example, Ensemble-based Filter Feature Selection (EFFS) [25] is a heterogeneous strategy. Four filter feature ranking approaches involving information gain, gain ratio, chi-squared, and relieFF are utilized to achieve the most proper feature subset. The method of [1] has employed some voting techniques to develop an ensemble feature selection. Arora et al. have engaged various Nature-Inspired methods to build an ensemble feature selection [35]. Wang et al. have occupied the Genetic Algorithm in their design to determine the best rank for features [54]. EFSPF [30] is a novel algorithm that uses the PageRank algorithm and fuzzy inference as an ensemble feature selection method. In other words, three based feature selection methods are first used, and then the fuzzy Type-I is applied to umpire between the ranks of feature selection methods. Hashemi et al. [27] measure the correlation between the target and other features. Next, a weighted graph is built based on determining the Euclidean Distance between features. Finally, the weighted PageRank algorithm is employed to rank the features. MFS-MCDM [26] assesses the importance of features in terms of their connection with the target. In other words, it is a novel strategy for decreasing the dimensionality of multi-label datasets. A new mixed feature selection technique called M-DFIFS through the feature's Dynamic Feature Importance (DFI) index was offered [57]. This method, Dynamic Feature Importance, is described through feature monotony and feature priority. Another method is called PEFS [24]. Firstly, for uncovering the non-dominate in a matrix made via different feature selection methods, the bi-objective optimization issue is used. Then, the crowding space in the bio-objective area is utilized to keep the non-dominated features. The second phase persists so long as whole non-dominated features are seen.

3. Fundamental concepts

3.1. **Ant colony optimization.** Ant colony optimization is a metaheuristic swarm intelligence algorithm; it is used for solving optimization problems and is utilized in a wide variety of field [11]. Ant Colony Optimization (ACO) imitates real ants' behavior; when ants find food, they build paths between food and their nest. In each path, they leave some chemical material as the pheromone trail. After a while, the pheromone trail starts to evaporate, and if ants use the paths again, they will reinforce the pheromone trail. So, the paths which have a solid pheromone trail will be chosen by other ants with a high probability. The different problems are adjusted based on real ants' behavior to be solved by ACO. To imitate ants' behavior to solve problems, we need to add some artificial behaviors to their genuine behaviors [10]. Regarding [47], the adjusted ACO for feature selection is the following.

Imagine we have a fully connected graph $G = (F, E)$ which F shows a set of features $F = f_1, f_2, \ldots . f_n$ as the nodes, and E denotes the relationships between features, respectively. At first, m ants with some initial pheromone value start to travel on m features of graph G randomly. In other words, each ant with a stochastic greedy rule travels on graph G. Then, a global updating rule will update the pheromone values of features while ants finish their travels. As mentioned previously, ants try to select a feature with high probability; so, the heuristic function for calculating probability moving from feature i to feature j for ant k is the following:

$$P_{ij}{}^{k}(t) = \frac{[\tau_i][\eta(f_i, f_j)]^{\beta}}{\Sigma_{u \in Ni}[\eta(f_i, f_j)]^{\beta}}, \forall j \in N_i^k \tag{1}$$

In Eq.(1) , the ${P_{ij}^{k}}(t)$ is the probability for ant k to move from feature i to feature j at time t, and τ_i denotes the pheromone value of feature i,N_i^k shows the neighbors of ant k,$\eta(f_i, f_j)$ is the heuristic information between feature i and feature j. The probability will calculate for all neighbors of node i. To specify the importance of heuristic information and pheromone value, use the β arameter in the range [0, 1]. It is obvious if $\beta = 0$, the importance of heuristic information and pheromone value will be equal.

Every time each ant makes decide to leave feature i to the next feature. For choosing the next feature, a random variable R (in the range [0, 1]) will be generated; for choosing one node to move, the roulette wheel selection is used [21]. As mentioned previously, a global updating rule will update the pheromone values of features while ants finish their travels. Eq.(2) is used as a global updating rule. ρ is a pheromone debasement scale parameter,$\Delta\tau_i$ shows the pheromone supplement value for feature i, which during the updating pheromone values will be added.

$$\tau_i(t+1) = (1-\rho)\tau_i(t) + \Delta\tau_i. \tag{2}$$

3.2. **Fuzzy inference structure.** The methods based on fuzzy are more appropriate for combining different feature selection techniques and tackling noise [20, 32, 39, 42]. On the other hand, fuzzy logic gives added flexibility for solve the uncertainty of various strategies.

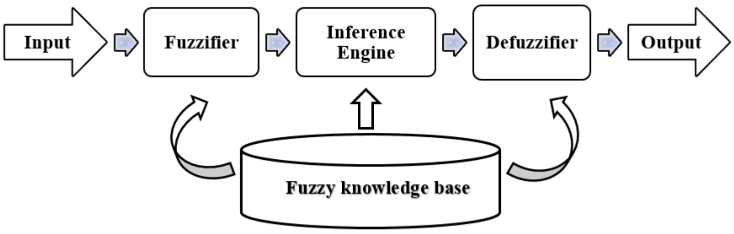

FIGURE 1. Structure of the fuzzy inference [30] .

The general design of the fuzzy inference system is illustrated in Fig.1. A group of if-then rules constructs the knowledge base of the system. The fuzzifier component uses the membership function in the knowledge base to shift numeric input into fuzzy linguistic variables.

The inference motor utilizes the if-then rules to shift fuzzy data into fuzzy outcomes. Also, the defuzzification component employs the membership functions to shift the fuzzy inference motor's outcome into numerical yields [30]. In Section 4, rules of the fuzzy system and membership functions for building a voter are given.

4. Proposed Method

As mentioned previously, using a single heuristic in the ACO cannot always significantly deal with the POA. So, in this paper, an ensemble of heuristics for feature selection has been proposed. In addition, in the case of applying a combination of heuristics to reach a better result, it should note that some feature selection methods offer a mistaken rank as a noise that will negatively influence heuristics. The fuzzy logic suggests additional flexibility to tackle uncertainty when the ranks of methods are uncertain.

In the proposed method, three effective-based methods are run. Then, fuzzy Type-I is involved in calculating importance (a reliability rate) as a heuristic for each feature based on views of feature selection methods. Then, a $1 \times M$ matrix is created (M demonstrates the number of features). Then the Euclidean Distance of estimated ranks between each pair of features as another heuristic is computed, and an $M \times M$ matrix is built. After that, a weighted graph based on a combination of the two heuristics will be constructed; Finally, the ACO is applied to the complete graph, and each ant considers the reliability rate and Euclidean Distance of the destination node together for moving between nodes of the graph.

4.1. **Heuristic 1: Applying Fuzzy Type-I schema.** For gaining more valuable results, involving several feature selection methods can be more beneficial. However, some methods may not correctly rank some features and cause noise. As stated earlier, fuzzy logic proposes additional flexibility for reducing noise. An efficient heuristic based on fuzzy logic that pursues the majority rule (denotes a rank far away from the other ranks is likely far from the valid rank [43]) has been created to discover a trustworthiness rate per feature.

First, three powerful methods comprising Maximal Information Coefficient [9], t-test [7], and RFS [38], have been operated separately to rank. Then, the contrast between anticipated ranks in the scope (0,1) is calculated. After that, the fuzzy voter for adjudicating is set as follows.

It can be seen in Fig.2 that the three inputs $r1$, $r2$, and $r3$ are supposed to characterize the foreseen rank of feature $F1$ by three methods containing Maximal Information Coefficient, t-test, and RFS, respectively. The Type-I fuzzy voter scheme gives a trustworthiness rate as a heuristic for the feature $F1$ by following definitions; the first phase is to calculate the length between the anticipated rank for $F1$ by these three methods, i.e., $D12 = |r1 - r2|$, $D13 = |r1 - r3|$ and $D23 = |r2 - r3|$. Three membership procedures called Small, Medium, and Large (notice Fig. 3) are utilized to fuzzify the contrasts. After that, for estimating agreement, the fuzzy rules (some of the fuzzy rules operated in this study are presented in Table 1) and the membership procedures comprising Vlow, Low, Medium, High, and Vhigh are involved (notice Fig.4). For instance, one rule has been shown in the following:

"If $D12$ is Small and $D13$ is Large, and $D23$ is Large, then output is Low" For instance, in this rule, if the contrast between Maximal Information Coefficient and t-test, the contrast between Maximal Information Coefficient and RFS, and the difference between t-test and RFS are Small, Large, and Large membership procedures, respectively, then the outcome is Low membership procedure.

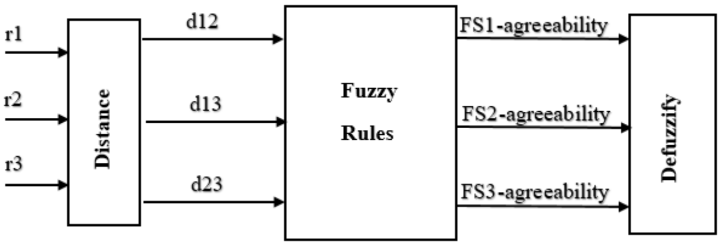

FIGURE 2. The schema of the voter [40].

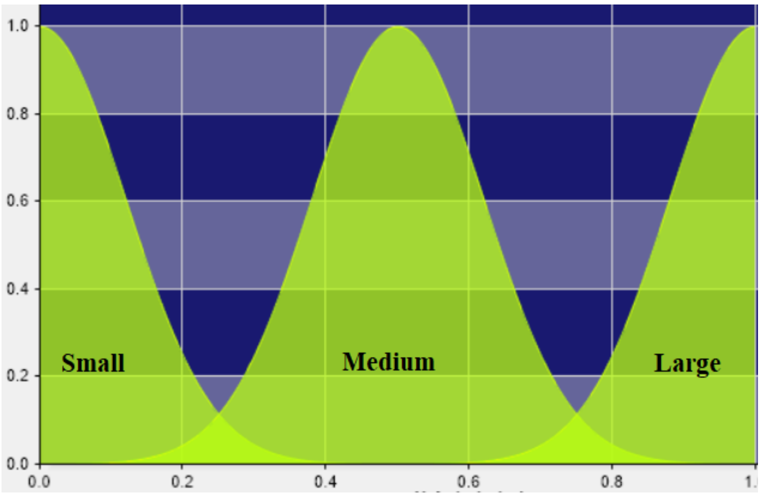

FIGURE 3. The membership procedures for fuzzifier [30].

Finally, the centroid Cj of N (count of fuzzy clusters) fuzzy clusters (Bj) is gained from the rules. Furthermore, the membership value, (μ_{B_j}) of the inputs of the intakes of each fuzzy cluster is estimated. Eventually, for computing a trustworthiness rate as a heuristic for the $F1$ feature Eq. (3) is operated:

$$R_{Fi} = \frac{\Sigma_i^N \mu_{B_j} c_j}{\Sigma_i^N \mu_{B_j}}. \tag{3}$$

Table 1. Some of linguistic rules [30].

D12 membership value	D13 membership value	D23 membership value	Output of rule
Small	Small	Small	Vhigh
Small	Medium	Small	Vhigh
Small	Large	Small	High
Large	Large	Large	Vlow
Small	Large	Large	Low
Small	Medium	Medium	Medium
Medium	Large	Large	Vlow
Medium	Medium	Medium	Medium
Medium	Large	Medium	Medium

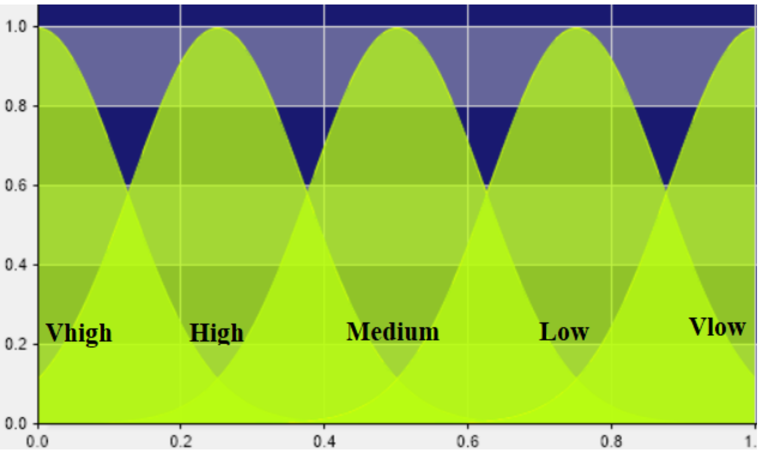

Figure 4. The membership procedures for agreement [40] .

Table 2. Parameters for fuzzification [30].

Small		**Medium**		**Large**	
σ	m	σ	m	σ	m
0.12	0	0.12	0.5	0.12	1

Our voter system is made by Small, Medium, and Large membership procedures to shift spans into fuzzy linguistic variables. The Vhigh, High, Medium, Low, and Vlow membership procedures are utilized to compute agreement. The Gaussian operations are picked, and also estimated ranks were standardized in the scope (0,1), the parameters for fixing the Gaussian function are adjusted in Tables 2 and 3.

4.2. **Heuristic 2: Applying Euclidean distance for building a complete weighted graph.** After applying three feature selection methods, their

Table 3. Parameters for agreement [30].

Vhigh		**High**		**Medium**		**Low**		**Vlow**	
σ	m	σ	m	σ	m	σ	m	σ	m
0.12	0	012	0.25	0.12	0.5	0.12	0.75	0.12	1

outcomes are combined into a matrix named Matrix of Ranks in Fig. 5, in this Figure, rows and columns are ranks and methods, respectively. Then, the Rf_{ij} represents the rank of feature i which is calculated with method j. Finally, the Euclidean distance between features is calculated based on a heuristic's three feature selection methods. For instance, for features 3 and 4, Euclidean distance is computed by Eq. (4):

$$ED_{3,4} = \sqrt{(Rf_{31} - Rf_{41})^2 + (Rf_{32} - Rf_{42})^2 + (Rf_{33} - Rt_{43})^2}. \tag{4}$$

$$Marix\ of\ Ranks = \begin{bmatrix} Rf_{11} & \cdots & Rf_{1n} \\ \vdots & \ddots & \vdots \\ Rf_{m1} & \cdots & Rf_{mn} \end{bmatrix}$$

Figure 5. Matrix of Ranks.

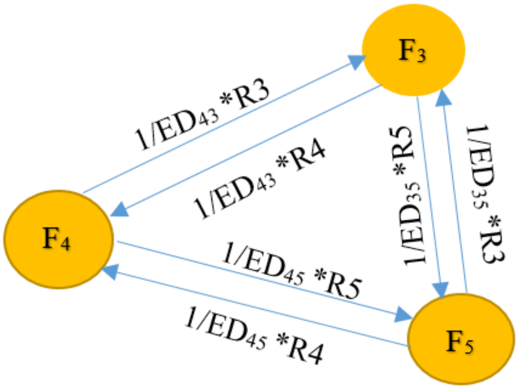

Figure 6. An instance of an entire graph.

For developing an entire graph, the Multiplication between the destination node's trustworthiness and the inverse of the Euclidean distance is set the importance of the edge. For example, in Fig. 6, an entire graph between three features is presented.

R demonstrates the trustworthiness of the goal node, and $ED_j i$ denotes the Euclidean length between nodes j and i. Ultimately, the ACO is applied to the complete graph, and each ant considers the trustworthiness rate and Euclidean Distance of the goal node together for transferring between nodes of the graph.

4.3. **Applying Ant Colony Optimization.** For applying ACO, we need to initialize some parameters in Eq. (5), which are written following:

$$P_{ij}{}^{k}(t) = \frac{[\tau_i][\eta(f_i, f_j)]^{\beta}}{\Sigma_{u \in Ni}[\eta(f_i, f_j)]^{\beta}}, \forall j \in N_i^k. \tag{5}$$

In Eq.1 the first step, the pheromone matrixτ for all nodes is initialized with a constant value as τ_0 In this paper, each element in matrix τ_0 for all link between nodes i and j in the complete graph is initialized with Eq. (6).

$$\tau_{ij} = \frac{1}{mean(alldistancesinmatrixED)}. \tag{6}$$

As mentioned previously, using a single heuristic in the ACO cannot always significantly deal with problems, and the ensemble of various strategies could cause search strategies with different abilities to support each other, thus significantly increase the performance. Consequently, here, the heuristic information, which is an ensemble of heuristics for each pair of nodes i and j, $\tau(f_i, f_j)$, is considered the Multiplication between inverse the calculated distance between node i and node j by Eq. (4), $\frac{1}{ED_{ij}}$ and the reliability of the destination node (R by Eq. (3)). Here, we consider the equal importance between heuristic information and pheromone, so the β parameter will be 1.

4.3.1. *Global pheromone updating rule.* In this paper, the numbers of ants and iterations are forty and forty, respectively. At the end of each iteration, the pheromone matrix τ will be updated; in other words, the pheromone value for each node by Eq. (7), which the $ant_k.Cost$ is the sum of weights' edges of travel of ant k which passed node i, and the parameter ρ is considered 0.05. The flowchart of the Ant-EHFS is shown in Algorithm 1. It should be noted, that the fuzzy and ACO implementations are on MATLAB. The code and datasets are on this path, https://github.com/mehdijoodaki/Ant-EHFS .

$$\tau_i(t+1) = (1-\rho)\tau_i(t) + \Delta_i = (1-\rho)\tau_i(t) + \frac{1}{ant_k, Cost}. \tag{7}$$

5. Results

The proposed method has employed six datasets to compare with the most delinquent ensemble methods. Five latest methods, PEFSS, Borda voting, Weighted Borda voting, M-DFIF, and plurality vote for the ensemble feature section, have been picked to describe our method's potency. Furthermore, eight based methods comprising Chi-square, relieFF, Fisher-score, t-test, Maximum Information Coefficient, RFS, Correlation and, Mutual Information (MI) have been contrasted with our method. In Table 4, some required details regarding are brought, including names of datasets, total of samples, total of instances, and classes.

Algorithm 1: Ant colony optimization equipped with an ensemble of heuristics through fuzzy logic for feature selection

```
Input: Ranked $M$ features of dataset $D$ by $N$
Output: Rank $M$ features base on their importance
1. Compute the rank of features by three based methods (Matrix of Ranks);
2. Compute Euclidean distance features as a heuristic (ED Matrix);
3. Apply the fuzzy Type-I to estimate a trustworthiness rate as heuristic for
   each feature
4. Make a graph based on a combination of the two heuristics;
5. Initializing the pheromone trails;
6. Repeat
7.  For each ant, Do
8.    Choosing the next feature according to the calculated
      probability of neighbors (take the reliability rate and Euclidean
      Distance of the destination node together);
9.   Update the pheromone:
10.     Evaporation
11.     Reinforcement
12. Until stopping criteria;
13. Output: final pheromone vector;
14. Choose $M$ features according to the pheromone trail;
```

Our analysis utilized these six datasets to verify our method's performance in metrics including, Accuracy, Recall, F1-score and Precision. Accuracy is the ratio of correctly organized and F1-score figured with Recall and Precision. That Recall is the sum of genuine positive instances divided by the whole total of genuine positive instances. Precision is the count of genuine positive instances divided by whole instances as positive. Consequently, the mentioned metrics are described by below Equations.

$$Recall = \frac{TP}{TP + FN}. \tag{8}$$

Table 4. Details about six datasets.

Dataset	**Class**	**Instance**	**Feature**
Jaffe [40]	10	213	676
ORL [49]	40	400	1024
Yale [5]	15	165	1024
WarpPIE10P [46]	10	210	2421
Lung[6]	5	203	3312
GLIOMA[18]	4	50	4434

$$presision = \frac{TP}{TP + FP}. \tag{9}$$

$$F1 - score = 2 * \frac{(Precision * Recall)}{Precision + Recall}. \tag{10}$$

The KNN algorithm [22] is applied to estimate the classification performance. Individually dataset is split randomly into 80% and 20% for the train and test datasets, respectively. Then, the 100 high features are selected by operating the ANT-EHFS for each dataset. For per dataset, the KNN algorithm is executed in 21 periods. In other words, firstly, one feature is picked, and then in each period, five features are added until all one hundred are chosen.

Figs. 7-18 illustrate the proposed method's accuracy compared to other algorithms. Also, Tables 1A-12A in Appendix A and 5-10 indicate ANT-EHFS comparisons in metrics, including the F1-score, Precision, and Recall, respectively. The Mean in Tables 5-10 shows the average scores of the 21 periods, and also, the winning sign indicates the number of periods that each method has a more score than the others. For instance, in Table 6, the proposed method wins 16 times from 21 runs, so the winning metric is equal to 16/21, and the mean of the F1 score is 0.72 percent for it. Regarding these results, the proposed method's strength is remarkably better than other methods based on the noted metrics.

The Wilcoxon trial is a nonparametric form engaged for hypothesis investigations; it counts the p-value of paired inputs to investigate their variety. If the p-value is smaller than or equivalent to a typical significance level (=0.05), then the null hypothesis (the null hypothesis indicates that there is no significant difference between the actions of the two methods) is denied. The plus sign (+) and minus (-) both decide whether a strategy is better than others, while the equal sign (=) points out that the actions of the methods are similar. Tables 11-13 show the comparison outcomes of the Wilcoxon test of our method and its rival's methods based on Average Recall, Average precision, and Average F1 score in six datasets. Tables 11-13 indicate that the ANT-EHFS rejects the null hypothesis and is better than others.

5.1. **Discussion.**

According to Tables 5-13, Figs. 7-18 and 1A-12A in Appendix A ANT-EHFS has acquired more excellent Accuracy, Precision, Recall, F1-score, and better Wilcoxon trial outcome than other methods. This fantastic accomplishment is due to operating a fitting ensemble of heuristics. The voter based on fuzzy logic is a trustworthy and adaptable method that reckons uncertainty between views of various methods and lessens noise. Thus, ANT-EHFS has attempted to apply ACO and reduce noise results on the ensemble of heuristics; it is apparent that it has had the most profitable performance over its competitors because it has efficiently conducted noise.

In other words, Views of features selection methods around features are not usually exact; as a result, some features selection methods assume a mistaken rank and generate noise in ranks. Therefore, when we are going to employ an ensemble of heuristics, the noise will negatively affect the model results. ANT-EHFS has incredible results than its ensemble competitions and based feature selection methods (Table 5-13 and Fig 7-12). In marked difference to current robust ensemble methods, including Weighted Board Count, Plurality Voting, PEFS, M-DFIFS, and ANT-EHFS in almost all datasets (see Fig 7-12), has earned the most increased average accurateness. For example, in Table 13, the ANT-EHFS's average F1-score and M-DFIFS, PEFS, Plurality Voting, Weighted Board Count, and Board Count are 0.83, 0.70, 0.73, 0.68, 0.72, and 0.67 percent, respectively. The significant causes for more acceptable results are that in ANT-EHFS, three feature selection methods with unlike views are employed as a heuristic. Secondly, fuzzy Type-I adds a weight (a trustworthiness rate) as another heuristic (to manage various feature selections' uncertainty). Ultimately, merging the noted two heuristics, which generates a reliable graph, has brought about these marvelous results. In other words, fuzzy Type-I has declined the influence of faulty ranks in the ensemble of heuristics.

It is noticeable from Tables 5-13, and Fig 13-18 that ANT-EHFS, because of lowering noise and making a trustworthy graph (with an ensemble of heuristics), excellent performance has been achieved by ACO than based methods comprising Mutual Information, t-test, Fisher-score, relieFF, Chi-square, Maximum Information Coefficient, RFS and Correlation, for example, average F1-score for ANT-EHFS and cited based methods are 0.83, 0.57, 0.50, 0.73, 0.64,0.64, 0.72, 0.65 and 0.59 percent, respectively.

6. **Conclusion**

High-dimensional data have numerous features that most features are outside or duplicative. Likewise, these features are expensive in computation and memory; they also hurt the learning model's performance. Consequently, one of the crucial tasks in this area is uncovering the most relevant features. One strategy is to discover the highest relevant features by utilizing multiple feature selection techniques and ensemble strategies in this area. In this study, we have presented the ACO with an ensemble of heuristics as a novel method. Here, the Euclidean distance between the ranks as heuristic has been used. Then, fuzzy Type-I has been employed for figuring a trustworthiness rate for features in terms of various ranks of methods as another heuristic. After that, an entire graph based on a mix of the two heuristics has been made; ultimately, the ACO is executed on the graph, and each ant considers the trustworthiness rate and Euclidean Distance of the destination node together for transferring between them. In addition, the complexity of the proposed method depends on the time of the finding a global optimum by ACO, so, for future work reducing this time by increasing the number of feature would be interesting [29].

Table 5. F1-score for the Jaffe dataset.

Number of features	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting	M-DFIFS	MI	PEFS	EFSPF
1	0.50	0.77	0.83	0.80	0.70	0.76	0.86	0.70	0.90	0.92	0.91	0.90	0.68	**0.94**	0.5
5	0.95	0.86	0.96	0.94	0.80	**0.98**	0.87	0.91	0.96	0.95	0.97	0.97	0.96	0.97	0.81
10	0.98	0.92	0.93	0.95	0.85	**0.99**	0.96	0.93	0.98	0.97	0.98	**0.99**	0.96	0.94	0.82
15	**1.00**	0.91	0.90	0.90	0.78	0.95	0.91	0.88	0.91	0.93	0.95	0.94	0.86	0.95	0.97
20	**1.00**	0.97	0.92	0.94	0.91	0.94	0.96	0.92	0.92	0.95	0.95	0.98	0.89	0.96	**1.00**
25	**1.00**	0.94	0.93	0.90	0.80	0.97	0.89	0.90	0.98	0.99	1.00	0.97	0.85	0.98	**1.00**
30	**1.00**	0.99	0.98	0.95	0.84	0.99	0.94	0.94	0.99	0.99	0.99	0.99	0.94	0.98	**1.00**
35	**1.00**	0.91	0.92	0.92	0.77	0.94	0.96	0.94	0.97	0.92	0.93	0.99	0.90	0.95	**1.00**
40	**1.00**	0.94	0.93	0.94	0.82	0.97	0.98	0.94	0.96	0.95	0.94	0.99	0.92	0.96	**1.00**
45	**1.00**	0.89	0.96	0.94	0.77	0.96	0.95	0.94	0.92	0.95	0.94	0.86	0.91	0.94	**1.00**
50	**1.00**	0.91	0.94	0.95	0.88	0.98	0.99	0.96	0.98	0.97	0.97	0.93	0.93	0.95	**1.00**
55	**1.00**	0.86	0.87	0.98	0.71	0.96	0.91	0.81	0.87	0.94	0.92	0.97	0.95	0.98	**1.00**
60	**1.00**	0.93	0.88	0.97	0.81	0.97	0.95	0.90	0.89	0.95	0.92	0.96	0.96	0.93	**1.00**
65	**1.00**	0.94	0.91	0.93	0.75	0.91	0.92	0.95	0.93	0.93	0.92	0.98	0.91	0.92	**1.00**
70	**1.00**	0.95	0.92	0.93	0.85	0.94	0.95	0.95	0.95	0.94	0.94	0.98	0.97	0.96	**1.00**
75	**1.00**	0.94	0.92	0.96	0.79	0.97	0.93	0.95	0.95	0.99	0.93	**1.00**	0.97	0.98	**1.00**
80	**1.00**	**1.00**	0.92	0.98	0.90	0.99	0.97	0.93	0.95	0.98	0.93	**1.00**	0.97	0.94	**1.00**
85	**1.00**	0.88	0.89	0.89	0.78	0.96	0.88	0.86	0.90	0.95	0.90	**1.00**	0.93	0.96	**1.00**
90	**1.00**	0.93	0.87	0.96	0.82	0.97	0.91	0.87	0.89	0.92	0.95	**1.00**	0.93	0.97	**1.00**
95	**1.00**	0.89	0.90	0.95	0.78	0.96	0.92	0.88	0.93	0.94	0.93	0.96	0.94	0.93	**1.00**
100	**1.00**	0.93	0.92	0.95	0.84	0.96	0.89	0.88	0.92	0.95	0.95	0.97	0.97	0.97	**1.00**
Mean	**0.97**	0.92	0.91	0.93	0.81	0.95	0.93	0.90	0.94	0.95	0.94	0.97	0.92	0.96	0.96
Win	**18/21**	1/21	0	0	0	2/21	0	0	0	0	0	5/21	0	1/21	17/21

Table 6. F1-score for the ORL dataset.

Number of features	**ANT-EHFS**	**Correlation**	**relieFF**	**Fisher Score**	**t-test**	**MIC**	**RFS**	**Chi-Squared**	**Borda Count**	**Weighted Borda Count**	**Plurality Voting**	**M-DFIFS**	**MI**	[HTML]FFFFFF**PEFS**	**EFSPF**
1	0.08	0.21	0.21	0.35	0.15	0.49	0.41	0.38	0.45	0.38	0.40	0.42	0.20	0.73	**0.80**
5	0.33	0.41	0.56	0.63	0.27	0.56	0.62	0.49	0.66	0.61	0.50	0.63	0.29	**0.74**	0.36
10	0.49	0.59	0.62	0.69	0.48	0.66	0.70	0.59	0.70	0.65	0.59	0.67	0.37	**0.73**	0.54
15	0.59	0.27	0.50	0.62	0.34	0.64	0.67	0.51	0.62	0.56	0.47	0.62	0.34	**0.75**	0.6
20	0.69	0.47	0.57	0.72	0.36	**0.75**	0.72	0.57	0.69	0.61	0.50	0.66	0.37	0.72	0.66
25	**0.75**	0.35	0.52	0.69	0.30	0.68	0.59	0.54	0.59	0.61	0.55	0.65	0.31	0.74	0.68
30	**0.79**	0.51	0.62	0.65	0.42	0.65	0.65	0.57	0.72	0.68	0.55	0.74	0.37	0.72	0.75
35	**0.79**	0.38	0.56	0.63	0.22	0.62	0.69	0.42	0.62	0.55	0.46	0.58	0.30	0.68	0.78
40	0.79	0.51	0.59	0.65	0.41	0.70	0.69	0.54	0.67	0.57	0.60	0.64	0.39	0.70	**0.82**
45	0.78	0.31	0.49	0.64	0.38	0.61	0.68	0.49	0.68	0.57	0.56	0.64	0.37	0.74	**0.81**
50	0.77	0.57	0.65	0.67	0.46	0.72	0.74	0.59	0.66	0.64	0.51	0.70	0.44	0.67	**0.82**
55	0.77	0.31	0.58	0.58	0.33	0.58	0.68	0.48	0.59	0.56	0.51	0.68	0.36	0.71	**0.80**
60	**0.82**	0.48	0.64	0.69	0.40	0.60	0.73	0.52	0.66	0.66	0.54	0.68	0.38	0.80	**0.82**
65	**0.82**	0.38	0.58	0.69	0.26	0.60	0.65	0.48	0.64	0.61	0.52	0.66	0.47	0.73	**0.82**
70	0.81	0.53	0.57	0.73	0.39	0.61	0.67	0.55	0.60	0.66	0.54	0.63	0.57	0.77	**0.85**
75	**0.85**	0.24	0.49	0.64	0.35	0.63	0.66	0.37	0.57	0.51	0.45	0.64	0.49	0.75	0.84
80	**0.84**	0.43	0.55	0.72	0.37	0.66	0.69	0.49	0.66	0.61	0.52	0.61	0.57	0.72	**0.84**
85	**0.84**	0.28	0.46	0.56	0.28	0.54	0.64	0.57	0.65	0.60	0.45	0.65	0.30	0.74	**0.84**
90	**0.86**	0.48	0.52	0.62	0.43	0.63	0.69	0.52	0.68	0.64	0.50	0.66	0.32	0.67	0.84
95	**0.86**	0.42	0.58	0.65	0.24	0.62	0.62	0.43	0.63	0.52	0.61	0.68	0.52	0.72	**0.86**
100	**0.86**	0.56	0.66	0.73	0.36	0.63	0.71	0.60	0.69	0.66	0.62	0.69	0.57	0.70	085
Mean	0.72	0.41	0.55	0.65	0.34	0.63	0.66	0.50	0.64	0.59	0.52	0.64	0.40	0.72	**0.76**
Win	**11/21**	0	0	0	0	1/21	0	0	0	0	0	0	0	3/21	**11/21**

F1-score for the Yale dataset.

Number of features	**ANT-EHFS**	**Correlation**	**relieFF**	**Fisher Score**	**t-test**	**MIC**	**RFS**	**Chi-Squared**	**Borda Count**	**Weighted Borda Count**	**Plurality Voting**	**M-DFIFS**	**MI**	**PEFS**	**EFSPF**
1	0.09	0.26	0.30	**0.53**	0.17	0.40	0.19	0.38	0.19	0.31	0.43	0.26	0.27	0.47	0.09
5	0.29	0.24	0.37	0.53	0.24	0.42	0.33	0.49	0.32	0.41	0.44	0.40	0.30	**0.54**	0.31
10	0.30	0.37	0.40	0.52	0.35	0.44	0.39	0.41	0.47	0.40	0.51	**0.58**	0.35	0.50	0.57
15	0.46	0.34	0.42	0.53	0.29	0.48	0.38	0.45	0.48	0.47	0.51	0.54	0.39	**0.59**	0.50
20	0.40	0.29	0.42	0.55	0.25	0.44	0.43	0.43	0.36	0.50	0.46	0.53	0.45	**0.66**	0.55
25	0.41	0.45	0.46	**0.68**	0.37	0.36	0.31	0.50	0.42	0.53	0.51	0.48	0.28	0.58	0.60
30	0.53	0.52	0.51	0.54	0.32	0.56	0.37	0.58	0.40	0.57	0.56	0.47	0.28	0.56	**0.60**
35	0.56	0.22	0.43	0.43	0.27	0.43	0.42	0.53	0.31	0.50	0.43	0.52	0.32	**0.59**	0.58
40	0.47	0.28	0.39	0.43	0.32	0.48	0.39	0.47	0.40	0.50	0.47	0.54	0.46	0.45	**0.65**
45	0.51	0.41	0.47	0.48	0.27	0.48	0.43	0.42	0.53	0.45	0.42	0.60	0.29	0.54	**0.62**
50	0.62	0.38	0.53	0.53	0.38	0.57	0.44	0.45	0.53	0.45	0.53	0.57	0.32	0.50	**0.64**
55	0.58	0.36	0.36	0.62	0.36	0.62	0.33	0.50	0.42	0.54	0.62	0.54	0.44	0.43	**0.68**
60	0.60	0.42	0.49	0.56	0.36	0.56	0.37	0.55	0.52	0.55	**0.64**	0.52	0.50	0.45	0.53
65	**0.58**	0.33	0.43	0.57	0.34	0.55	0.22	0.46	0.44	0.52	0.52	0.45	0.41	0.48	0.52
70	0.52	0.37	0.39	**0.60**	0.43	0.56	0.35	0.47	0.46	0.56	0.49	0.39	0.43	0.50	**0.60**
75	0.52	0.30	0.37	0.44	0.35	0.47	0.30	0.42	0.44	0.50	0.55	0.42	0.26	0.43	**0.65**
80	0.57	0.40	0.44	0.54	0.42	0.44	0.47	0.50	0.42	0.55	0.47	0.35	0.30	0.50	**0.64**
85	0.59	0.42	0.49	0.61	0.27	**0.64**	0.46	0.50	0.48	0.41	0.60	0.50	0.33	0.42	0.61
90	0.60	0.47	0.50	**0.66**	0.41	0.62	0.46	0.54	0.51	0.54	0.57	0.61	0.39	0.48	0.62
95	0.58	0.35	0.42	0.46	0.29	0.58	0.38	0.50	0.40	0.48	0.45	0.46	0.33	0.61	**0.62**
100	0.58	0.38	0.46	0.47	0.41	0.61	0.43	0.40	0.46	0.49	0.50	0.46	0.35	0.49	**0.62**
Mean	0.49	0.36	0.43	0.54	0.33	0.51	0.37	0.47	0.43	0.49	0.51	0.49	0.35	0.51	**0.56**
Win	1/21	0	0	4/21	0	1/21	0	0	0	0	1/21	1/21	0	4/21	**10/21**

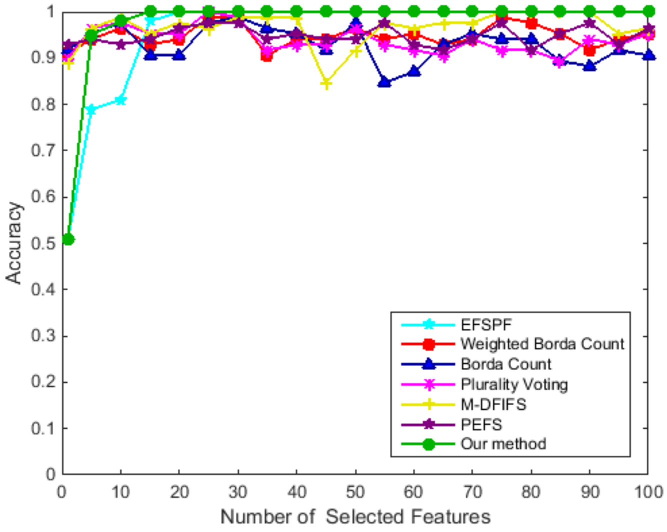

Figure 7. Jaffe

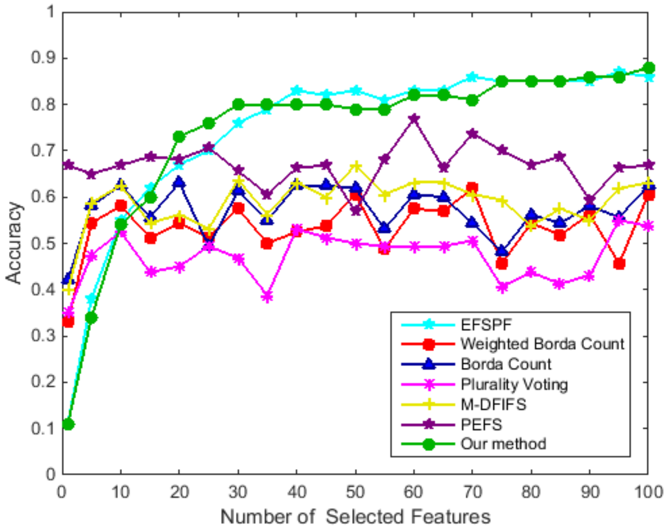

Figure 8. ORL

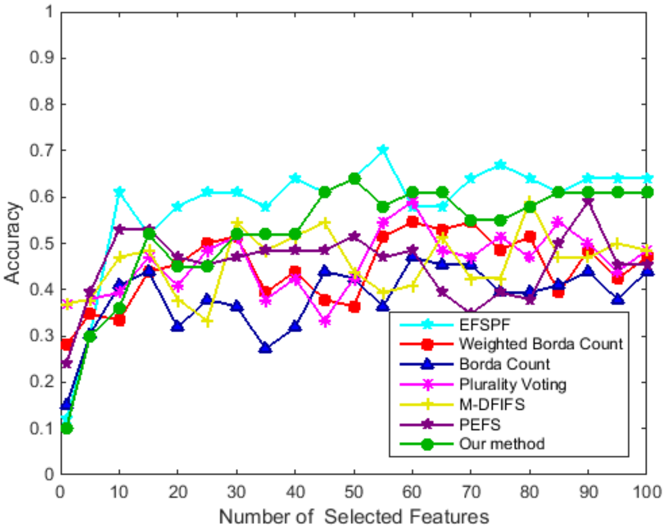

FIGURE 9. Yale

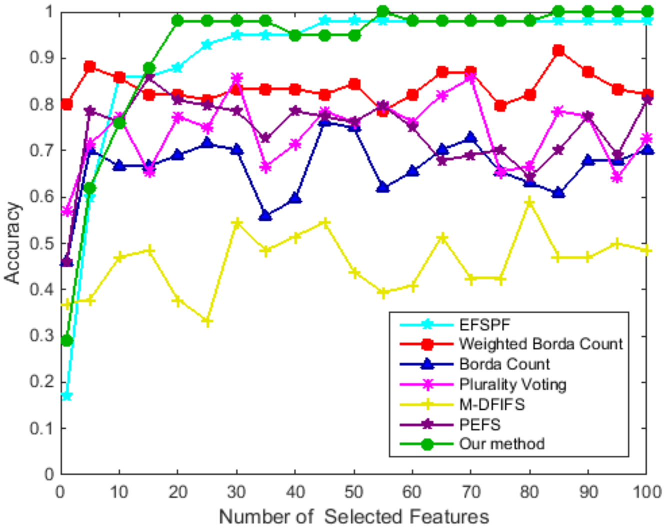

FIGURE 10. WarpPIE10P

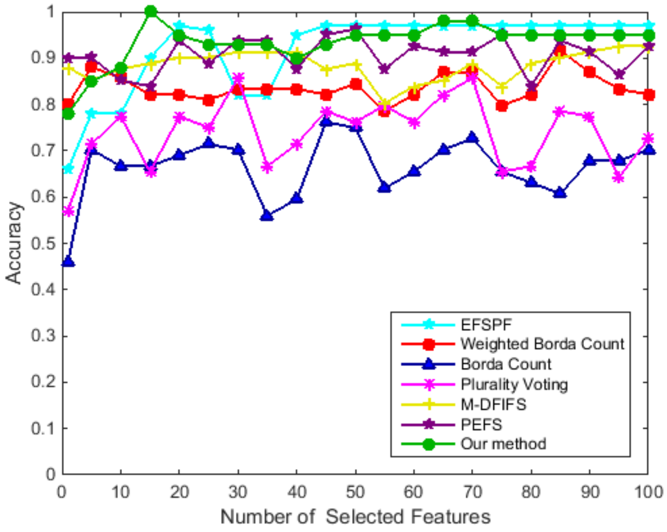

Figure 11. Lung

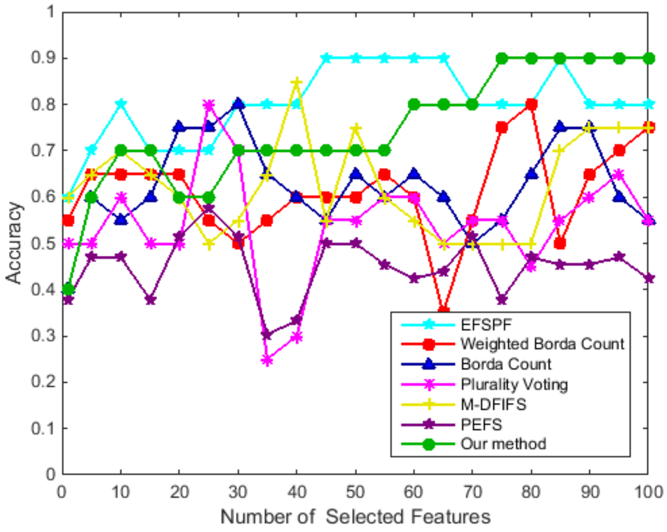

Figure 12. GLIOMA

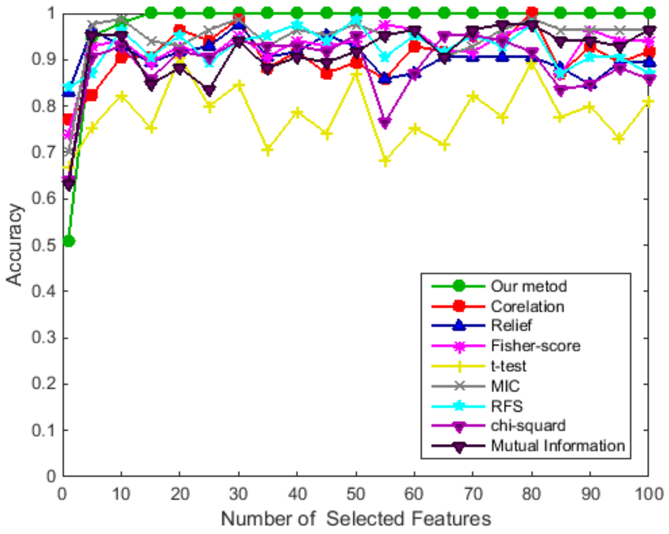

Figure 13. Jaffe

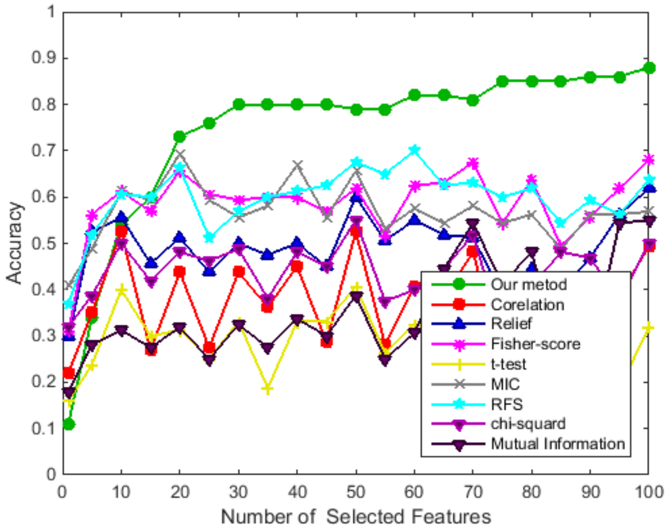

Figure 14. ORL

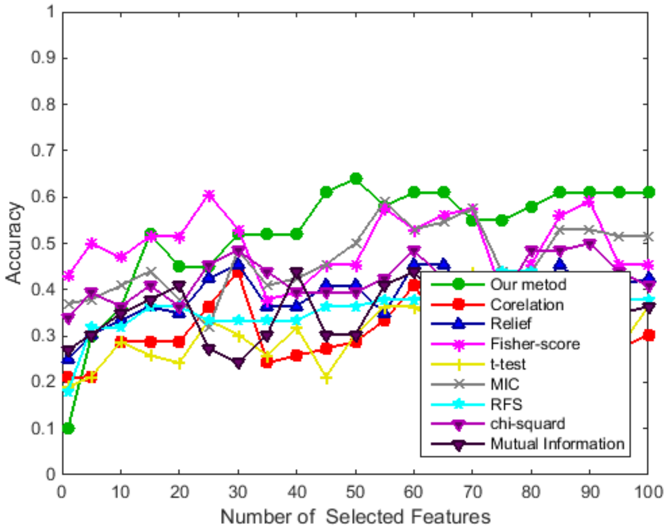

Figure 15. Yale

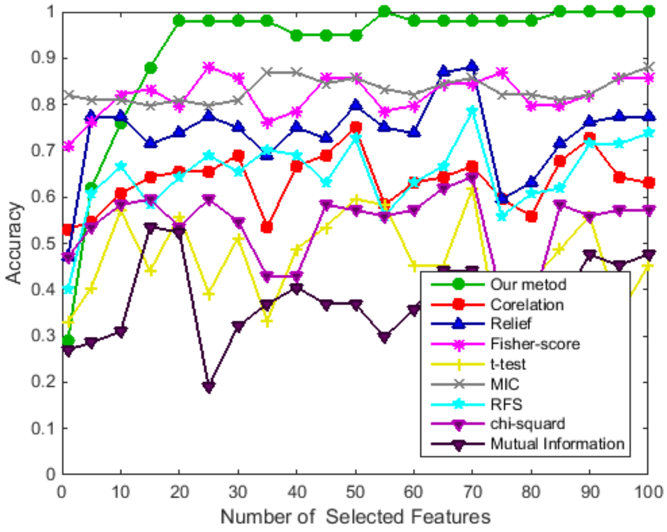

Figure 16. WarpPIE10P

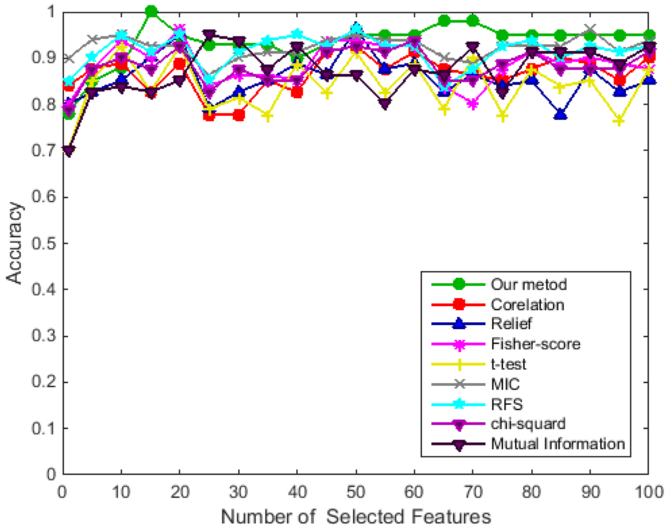

FIGURE 17. Lung

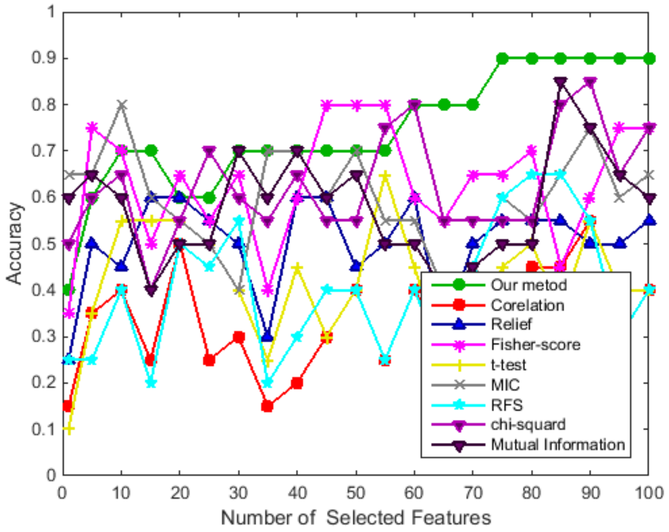

FIGURE 18. GLIOMA

Table 7. F1-score for the WarpPIE10P dataset.

Number of features	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting	M-DFIFS	MI	PEFS	EFSPF
1	0.29	0.54	0.52	0.76	0.38	0.82	0.46	0.52	0.54	0.83	0.63	0.52	0.34	**0.84**	0.16
5	0.62	0.60	0.80	0.80	0.42	0.80	0.67	0.62	0.74	**0.90**	0.74	0.80	0.30	0.81	0.58
10	0.78	0.68	0.82	0.86	0.60	0.83	0.72	0.66	0.69	0.88	0.81	0.77	0.35	**0.89**	0.85
15	**0.87**	0.65	0.79	0.85	0.51	0.81	0.67	0.66	0.71	0.85	0.71	0.86	0.56	0.84	0.85
20	**0.98**	0.70	0.82	0.84	0.58	0.83	0.73	0.59	0.74	0.86	0.82	0.82	0.55	0.92	0.87
25	**0.98**	0.66	0.80	0.90	0.45	0.87	0.74	0.64	0.74	0.82	0.77	0.80	0.20	0.87	0.93
30	**0.98**	0.71	0.76	0.88	0.54	0.84	0.72	0.61	0.72	0.85	0.87	0.83	0.36	0.88	0.95
35	**0.98**	0.58	0.72	0.77	0.35	0.87	0.73	0.48	0.63	0.84	0.71	0.74	0.40	0.92	0.95
40	**0.95**	0.71	0.78	0.79	0.51	0.88	0.72	0.49	0.65	0.85	0.75	0.83	0.44	0.89	**0.95**
45	0.95	0.71	0.78	0.90	0.57	0.89	0.67	0.63	0.82	0.85	0.82	0.78	0.43	0.87	**0.98**
50	0.95	0.77	0.82	0.90	0.64	0.89	0.76	0.63	0.80	0.88	0.80	0.79	0.43	0.86	**0.98**
55	**1.00**	0.66	0.78	0.82	0.60	0.82	0.58	0.58	0.62	0.82	0.82	0.83	0.32	0.85	0.98
60	**0.98**	0.68	0.77	0.82	0.53	0.86	0.66	0.63	0.67	0.88	0.79	0.80	0.37	0.89	**0.98**
65	**0.98**	0.69	0.88	0.87	0.47	0.83	0.70	0.64	0.74	0.90	0.84	0.72	0.49	0.90	**0.98**
70	**0.98**	0.72	0.90	0.87	0.64	0.88	0.80	0.68	0.75	0.90	0.87	0.74	0.49	0.90	**0.98**
75	**0.98**	0.63	0.68	0.88	0.28	0.84	0.63	0.49	0.72	0.82	0.70	0.73	0.34	0.87	**0.98**
80	**0.98**	0.60	0.72	0.83	0.45	0.84	0.67	0.48	0.68	0.85	0.73	0.67	0.33	0.79	**0.98**
85	**1.00**	0.73	0.73	0.83	0.50	0.81	0.68	0.62	0.64	0.92	0.81	0.74	0.41	0.89	0.98
90	**1.00**	0.80	0.80	0.84	0.60	0.84	0.74	0.59	0.70	0.88	0.80	0.80	0.53	0.79	0.98
95	**1.00**	0.67	0.83	0.87	0.39	0.89	0.77	0.62	0.74	0.87	0.72	0.74	0.49	0.82	0.98
100	**1.00**	0.70	0.82	0.86	0.52	0.91	0.77	0.62	0.74	0.85	0.77	0.84	0.51	0.90	0.98
Mean	**0.92**	0.67	0.78	0.84	0.5	0.85	0.69	0.59	0.70	0.86	0.77	0.77	0.41	0.87	0.90
Win	**16/21**	0	0	0	0	0	0	0	0	1/21	0	0	0	2/21	8/21

Table 8. F1-score for the Lung dataset.

Number of features	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting	M-DFIFS	MI	PEFS	EFSPF
1	0.68	0.69	0.69	0.60	0.28	0.70	0.73	0.53	0.61	0.70	0.80	**0.82**	0.50	0.71	0.81
5	0.83	0.81	0.79	0.88	0.63	0.91	0.94	0.86	0.69	0.91	**0.95**	0.80	0.82	0.71	0.81
10	0.84	0.86	0.79	0.94	0.75	0.90	0.96	0.88	0.82	0.90	**0.98**	0.83	0.80	0.82	0.86
15	**1.00**	0.75	0.79	0.84	0.56	0.85	0.84	0.80	0.69	0.84	0.80	0.69	0.65	0.61	0.91
20	0.94	0.80	0.88	0.91	0.72	0.83	0.88	0.72	0.71	0.79	0.82	0.70	0.66	0.76	**0.97**
25	0.90	0.59	0.63	0.68	0.59	0.70	0.71	0.67	0.65	0.66	0.67	0.71	0.90	0.86	**0.95**
30	**0.90**	0.59	0.67	0.71	0.62	0.75	0.76	0.72	0.68	0.67	0.72	0.72	0.88	0.89	**0.90**
35	**0.92**	0.71	0.67	0.75	0.50	0.70	0.88	0.76	0.68	0.64	0.73	0.73	0.79	0.75	0.90
40	0.89	0.71	0.80	0.74	0.72	0.71	0.92	0.62	0.81	0.81	0.90	0.73	0.86	0.82	**0.94**
45	0.92	0.89	0.83	0.88	0.58	0.89	0.92	0.89	0.82	0.87	0.91	0.68	0.68	0.89	**0.94**
50	0.94	0.91	0.95	0.87	0.69	0.94	0.96	0.92	**0.97**	0.87	0.96	0.70	0.68	0.77	0.94
55	0.94	0.86	0.68	0.89	0.59	0.84	0.90	0.85	0.61	0.83	**0.96**	0.64	0.65	0.77	0.94
60	0.94	0.90	0.68	0.75	0.70	0.90	0.91	0.93	0.68	0.91	**0.96**	0.68	0.83	0.90	0.94
65	**0.98**	0.87	0.73	0.83	0.57	0.88	0.79	0.79	0.82	0.70	0.81	0.66	0.67	0.73	0.94
70	**0.98**	0.83	0.77	0.58	0.71	0.87	0.86	0.79	0.87	0.80	0.83	0.70	0.74	0.86	0.94
75	**0.94**	0.77	0.76	0.81	0.54	0.87	0.88	0.70	0.71	0.81	0.87	0.63	0.62	0.89	**0.94**
80	**0.94**	0.81	0.78	0.86	0.67	0.84	0.88	0.86	0.71	0.84	0.87	0.69	0.85	0.78	**0.94**
85	**0.94**	0.89	0.53	0.78	0.74	0.87	0.90	0.68	0.67	0.90	0.90	0.71	0.81	0.89	**0.94**
90	**0.94**	0.89	0.87	0.86	0.66	0.96	0.92	0.68	0.75	0.90	0.85	0.73	0.86	0.88	**0.94**
95	**0.94**	0.65	0.60	0.86	0.46	0.83	0.70	0.68	0.73	0.80	0.82	0.75	0.89	0.83	**0.94**
100	**0.94**	0.84	0.65	0.79	0.67	0.89	0.86	0.86	0.75	0.87	0.84	0.75	0.89	0.81	**0.94**
Mean	**0.92**	0.79	0.74	0.80	0.62	0.84	0.86	0.77	0.73	0.81	0.85	0.72	0.76	0.80	**0.92**
Win	**10/21**	0	0	0	0	1/21	0	0	1/21	0	4/21	1/21	0	0	**10/21**

Table 9. F1-score for the GLIOMA dataset.

Number of features	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting	M-DFIFS	MI	[HTML]FFFFFF**PEFS**	[HTML]FFFFFF**EFSPF**
1	0.40	0.16	0.21	0.33	0.15	0.48	0.22	0.52	0.30	0.21	0.58	0.63	0.49	0.47	**0.81**
5	0.50	0.48	0.47	0.73	0.36	0.56	0.25	0.62	0.72	0.80	0.58	0.67	0.67	0.50	**0.81**
10	0.69	0.49	0.51	0.81	0.52	0.79	0.48	0.65	0.68	0.76	0.65	0.68	0.63	0.56	**0.86**
15	0.74	0.46	0.64	0.79	0.51	0.48	0.25	0.40	0.67	0.78	0.44	0.77	0.55	0.47	**0.91**
20	0.63	0.67	0.74	0.67	0.60	0.44	0.67	0.64	0.82	0.82	0.41	0.67	0.62	0.55	**0.97**
25	0.58	0.26	0.63	0.67	0.44	0.64	0.38	0.69	0.86	0.60	0.85	0.61	0.43	0.64	**0.95**
30	0.67	0.25	0.46	0.49	0.43	0.49	0.47	0.62	0.88	0.60	0.76	0.64	0.80	0.53	**0.90**
35	0.67	0.18	0.46	0.65	0.33	0.59	0.24	0.53	0.65	0.69	0.39	0.58	0.62	0.37	**0.90**
40	0.67	0.23	0.65	0.60	0.53	0.54	0.27	0.60	0.62	0.71	0.34	0.87	0.62	0.36	**0.94**
45	0.67	0.41	0.67	0.40	0.24	0.64	0.38	0.46	0.60	0.69	0.43	0.42	0.64	0.48	**0.94**
50	0.67	0.33	0.54	0.51	0.37	0.76	0.53	0.46	0.68	0.69	0.43	0.57	0.68	0.49	**0.94**
55	0.67	0.32	0.55	0.73	0.60	0.53	0.28	0.76	0.70	0.70	0.62	0.52	0.46	0.45	**0.94**
60	0.80	0.48	0.66	0.75	0.53	0.53	0.44	0.81	0.75	0.64	0.53	0.53	0.52	0.38	**0.94**
65	0.80	0.29	0.25	0.32	0.34	0.32	0.21	0.49	0.63	0.40	0.43	0.43	0.37	0.50	**0.94**
70	0.80	0.35	0.42	0.39	0.40	0.37	0.45	0.47	0.51	0.58	0.48	0.43	0.41	0.54	**0.94**
75	0.90	0.38	0.58	0.67	0.45	0.50	0.61	0.63	0.61	0.78	0.57	0.54	0.55	0.47	**0.94**
80	0.90	0.47	0.60	0.65	0.56	0.61	0.66	0.64	0.71	0.82	0.48	0.59	0.54	0.51	**0.94**
85	0.90	0.54	0.43	0.84	0.21	0.72	0.67	0.85	0.73	0.64	0.66	0.73	0.91	0.53	**0.94**
90	0.90	0.58	0.61	0.84	0.65	0.81	0.41	0.89	0.80	0.73	0.53	0.82	0.82	0.48	**0.94**
95	0.90	0.31	0.42	0.79	0.32	0.58	0.22	0.62	0.63	0.70	0.59	0.72	0.63	0.55	**0.94**
100	0.90	0.30	0.50	0.77	0.38	0.64	0.34	0.72	0.47	0.76	0.47	0.74	0.47	0.50	**0.94**
Mean	0.73	0.38	0.52	0.64	0.42	0.57	0.40	0.62	0.67	0.67	0.53	0.63	0.59	0.49	**0.92**
Win	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**21/21**

Table 10. Wilcoxon trial in term of Average Precision.

Datasets	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting		M-DFIFS	MI	PEFS	EFSPF
Jaffe	0.97	0.92	0.91	0.94	0.81	0.95	0.93	0.91	0.93	0.94	0.94	0.97	0.92		0.95	0.99
ORL	0.77	0.42	0.56	0.65	0.36	0.63	0.67	0.53	0.65	0.61	0.52	0.65	0.40		0.72	0.77
Yale	0.52	0.38	0.45	0.55	0.32	0.52	0.37	0.52	0.44	0.49	0.53	0.50	0.36		0.52	0.60
WarpPIE0	0.93	0.70	0.79	0.85	0.51	0.85	0.71	0.62	0.72	0.87	0.78	0.78	0.43		0.87	0.90
Lung	0.91	0.86	0.82	0.87	0.66	0.86	0.91	0.84	0.79	0.90	0.91	0.73	0.81		0.85	0.92
GLIOMA	0.79	0.40	0.54	0.64	0.44	0.56	0.40	0.62	0.68	0.72	0.54	0.61	0.58		0.51	0.85
Wilcoxon		+	+	+	+	+	+	+	+	+	=	+	+		+	+

References

[1] Laith Mohammad Abualigah and Ahamad Tajudin Khader. "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering".

Table 11. Wilcoxon trial in term of Average Recall.

Datasets	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting			M-DFIFS	MI	PEFS	EFSPF
Jaffe	**0.97**	0.91	0.92	0.93	0.81	0.95	0.93	0.90	0.94	0.95	0.94	**0.97**	0.92			0.95	0.96
ORL	**0.74**	0.41	0.55	0.64	0.33	0.63	0.65	0.49	0.63	0.58	0.52	0.64	0.39			0.71	0.74
Yale	0.52	0.35	0.42	**0.53**	0.35	0.51	0.38	0.44	0.42	0.49	0.49	0.48	0.35			0.51	0.61
WarpPIE0	**0.92**	0.66	0.66	0.84	0.50	0.85	0.68	0.57	0.69	0.86	0.77	0.77	0.41			0.86	0.90
Lung	**0.93**	0.73	0.68	0.74	0.58	0.80	0.81	0.72	0.69	0.74	0.81	0.71	0.73			0.78	0.92
GLIOMA	**0.90**	0.37	0.52	0.64	0.42	0.59	0.41	0.63	0.66	0.67	0.54	0.64	0.60			0.47	0.80
Wilcoxon		+	+	+	+	+	+	+	+	+	+	+	+			+	+

Table 12. Wilcoxon trial in term of Average F1-score.

Datasets	ANT-EHFS	Correlation	relieFF	Fisher Score	t-test	MIC	RFS	Chi-Squared	Borda Count	Weighted Borda Count	Plurality Voting			M-DFIFS	MI	PEFS	EFSPF
Jaffe	**0.97**	0.92	0.91	0.93	0.81	0.95	0.93	0.90	0.94	0.95	0.94	**0.97**	0.92			0.96	0.96
ORL	**0.72**	0.41	0.55	0.65	0.34	0.63	0.66	0.50	0.64	0.59	0.52	0.64	0.40			**0.72**	0.76
Yale	0.49	0.36	0.43	**0.54**	0.33	0.51	0.37	0.47	0.43	0.49	0.51	0.49	0.35			0.51	0.56
WarpPIE0	**0.92**	0.67	0.68	0.84	0.5	0.85	0.69	0.59	0.70	0.86	0.77	0.77	0.41			0.87	0.90
Lung	**0.92**	0.79	0.74	0.80	0.61	0.83	0.86	0.77	0.73	0.81	0.85	0.72	0.76			0.80	0.79
GLIOMA	**0.73**	0.38	0.52	0.64	0.42	0.57	0.40	0.62	0.67	0.67	0.53	0.63	0.59			0.49	0.92
Wilcoxon		+	+	+	+	+	+	+	+	+	+	=	+			=	=

In: *The Journal of Supercomputing* 73.11 (2017), pp. 4773–4795. ISSN: 0920-8542.

[2] Laith Mohammad Abualigah, Ahamad Tajudin Khader, and Essam Said Hanandeh. “A new feature selection method to improve the document clustering using particle swarm optimization algorithm”. In: *Journal of Computational Science* 25 (2018), pp. 456–466. ISSN: 1877-7503.

[3] Laith Mohammad Qasim Abualigah and Essam S Hanandeh. “Applying genetic algorithms to information retrieval using vector space model”. In: *International Journal of Computer Science, Engineering and Applications* 5.1 (2015), p. 19. ISSN: 2231-0088.

[4] Amir Hossein Atashkar, Nasser Ghadiri, and Mehdi Joodaki. “Linked data partitioning for RDF processing on Apache Spark”. In: *2017 3th International Conference on Web Research (ICWR)*. IEEE, pp. 73–77. ISBN: 1538604205.

[5] Peter N. Belhumeur, Joao P Hespanha, and David J. Kriegman. “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”. In: *IEEE Transactions on pattern analysis and machine intelligence* 19.7 (1997), pp. 711–720.

[6] Arindam Bhattacharjee et al. “Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses”. In: *Proceedings of the National Academy of Sciences* 98.24 (2001), pp. 13790–13795.

[7] C Alan Boneau. “The effects of violations of assumptions underlying the t test”. In: *Psychological bulletin* 57.1 (1960), p. 49. ISSN: 1939-1455.

[8] Paul S Bradley and Olvi L Mangasarian. “Feature selection via concave minimization and support vector machines.” In: *ICML*. Vol. 98. Citeseer. 1998, pp. 82–90.

[9] Elias F Combarro et al. “Introducing a family of linear measures for feature selection in text categorization”. In: *IEEE transactions on Knowledge and Data Engineering* 17.9 (2005), pp. 1223–1232.

[10] Marco Dorigo, Mauro Birattari, and Thomas Stutzle. "Ant colony optimization". In: *IEEE computational intelligence magazine* 1.4 (2006), pp. 28–39. ISSN: 1556-603X.

[11] Marco Dorigo and Gianni Di Caro. "Ant colony optimization: a new meta-heuristic". In: *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*. Vol. 2. IEEE. 1999, pp. 1470–1477.

[12] Mohammad Bagher Dowlatshahi, Vali Derhami, and Hossein Nezamabadi-pour. "Ensemble of filter-based rankers to guide an epsilon-greedy swarm optimizer for high-dimensional feature subset selection". In: *Information* 8.4 (2017), p. 152.

[13] Mohammad Bagher Dowlatshahi, Vali Derhami, and Hossein Nezamabadi-pour. "A novel three-stage filter-wrapper framework for miRNA subset selection in cancer classification". In: *Informatics*. Vol. 5. 1. Multidisciplinary Digital Publishing Institute. 2018, p. 13.

[14] Mohammad Bagher Dowlatshahi and Hossein Nezamabadi-Pour. "GGSA: a grouping gravitational search algorithm for data clustering". In: *Engineering Applications of Artificial Intelligence* 36 (2014), pp. 114–121. ISSN: 0952-1976.

[15] Mohammad Bagher Dowlatshahi, Hossein Nezamabadi-Pour, and Mashaallah Mashinchi. "A discrete gravitational search algorithm for solving combinatorial optimization problems". In: *Information Sciences* 258 (2014), pp. 94–107. ISSN: 0020-0255.

[16] Mohammad Bagher Dowlatshahi, Hossein Nezamabadi-Pour, and Mashaallah Mashinchi. "A discrete gravitational search algorithm for solving combinatorial optimization problems". In: *Information Sciences* 258 (2014), pp. 94–107.

[17] Mohammad Bagher Dowlatshahi and Mehdi Rezaeian. "Training spiking neurons with gravitational search algorithm for data classification". In: *2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. IEEE, pp. 53–58. ISBN: 1467387371.

[18] Peter Drotár, Matej Gazda, and Liberios Vokorokos. "Ensemble feature selection using election methods and ranker clustering". In: *Information Sciences* 480 (2019), pp. 365–380.

[19] Richard O Duda, Peter E Hart, et al. *Pattern classification*. John Wiley & Sons, 2006.

[20] Mahdi Eftekhari et al. *How Fuzzy Concepts Contribute to Machine Learning*. 2022.

[21] David E Goldberg. "Genetic algorithms in search". In: *Optimization, and MachineLearning* (1989).

[22] Gongde Guo et al. "KNN model-based approach in classification". In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2003, pp. 986–996.

[23] Isabelle Guyon and André Elisseeff. “An introduction to variable and feature selection”. In: *Journal of machine learning research* 3.Mar (2003), pp. 1157–1182.

[24] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. “A Pareto-based ensemble of feature selection algorithms”. In: *Expert Systems with Applications* (2021), p. 115130. ISSN: 0957-4174. DOI: `https://doi.org/10.1016/j.eswa.2021.115130`. URL: `https://www.sciencedirect.com/science/article/pii/S0957417421005716`.

[25] Amin Hashemi and Mohammad Bagher Dowlatshahi. “MLCR: a fast multi-label feature selection method based on K-means and L2-norm”. In: *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*. IEEE, pp. 1–7. ISBN: 1728159377.

[26] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. “MFS-MCDM: Multi-label feature selection using multi-criteria decision making”. In: *Knowledge-Based Systems* 206 (2020), p. 106365. ISSN: 0950-7051.

[27] Amin Hashemi, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. “MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality”. In: *Expert Systems with Applications* 142 (2020), p. 113024. ISSN: 0957-4174.

[28] Amin Hashemi et al. “Ant Colony Optimization equipped with an ensemble of heuristics through Multi-Criteria Decision Making: A case study in ensemble feature selection”. In: *Applied Soft Computing* (2022), p. 109046.

[29] Nazrul Hoque, Dhruba K Bhattacharyya, and Jugal K Kalita. “MIFS-ND: A mutual information-based feature selection method”. In: *Expert Systems with Applications* 41.14 (2014), pp. 6371–6385. ISSN: 0957-4174.

[30] Mehdi Joodaki, Mohammad Bagher Dowlatshahi, and Nazanin Zahra Joodaki. “An ensemble feature selection algorithm based on PageRank centrality and fuzzy logic”. In: *Knowledge-Based Systems* (2021), p. 107538. ISSN: 0950-7051.

[31] Mehdi Joodaki, Nasser Ghadiri, and Amir Hossein Atashkar. “Protein complex detection from PPI networks on Apache Spark”. In: *2017 9th International Conference on Information and Knowledge Technology (IKT)*. IEEE, pp. 111–115. ISBN: 1538625474.

[32] Mehdi Joodaki et al. “A scalable random walk with restart on heterogeneous networks with Apache Spark for ranking disease-related genes through type-II fuzzy data fusion”. In: *Journal of Biomedical Informatics* (2021), p. 103688. ISSN: 1532-0464.

[33] Nazanin Zahra Joodaki, Mohammad Bagher Dowlatshahi, and Mehdi Joodaki. “A novel ensemble feature selection method through Type I fuzzy”. In: *2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*. IEEE. 2022, pp. 1–6.

[34] Alan Jović, Karla Brkić, and Nikola Bogunović. “A review of feature selection methods with applications”. In: *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*. Ieee. 2015, pp. 1200–1205.

[35] Hamidreza Rashidy Kanan and Karim Faez. “An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system”. In: *Applied Mathematics and Computation* 205.2 (2008), pp. 716–725. ISSN: 0096-3003.

[36] Igor Kononenko. “Estimating attributes: analysis and extensions of RELIEF”. In: *European conference on machine learning*. Springer, pp. 171–182.

[37] Cosmin Lazar et al. “A survey on filter techniques for feature selection in gene expression microarray analysis”. In: *IEEE/ACM transactions on computational biology and bioinformatics* 9.4 (2012), pp. 1106–1119.

[38] Jimin Lee, Nomin Batnyam, and Sejong Oh. “RFS: Efficient feature selection method based on R-value”. In: *Computers in biology and medicine* 43.2 (2013), pp. 91–99.

[39] Ondrej Linda and Milos Manic. “Interval type-2 fuzzy voter design for fault tolerant systems”. In: *Information Sciences* 181.14 (2011), pp. 2933–2950.

[40] Michael J Lyons, Julien Budynek, and Shigeru Akamatsu. “Automatic classification of single facial images”. In: *IEEE transactions on pattern analysis and machine intelligence* 21.12 (1999), pp. 1357–1362.

[41] Rammohan Mallipeddi et al. “Differential evolution algorithm with ensemble of parameters and mutation strategies”. In: *Applied soft computing* 11.2 (2011), pp. 1679–1696. ISSN: 1568-4946.

[42] Adele Mehranfar et al. “A Type-2 fuzzy data fusion approach for building reliable weighted protein interaction networks with application in protein complex detection”. In: *Computers in biology and medicine* 88 (2017), pp. 18–31. ISSN: 0010-4825.

[43] Majid Mohammadi and Jafar Rezaei. “Ensemble ranking: Aggregation of rankings produced by different multi-criteria decision-making methods”. In: *Omega* 96 (2020), p. 102254. ISSN: 0305-0483.

[44] Seyed Vahid Moravvej et al. “A method Based on an Attention Mechanism to Measure the Similarity of two Sentences”. In: *2021 7th International Conference on Web Research (ICWR)*. 2021, pp. 238–242. DOI: `10.1109/ICWR51868.2021.9443135`.

[45] SV Moravvej et al. “Efficient GAN-based method for extractive summarization”. In: *Journal of Electrical and Computer Engineering Innovations (JECEI)* (2021).

[46] Songyot Nakariyakul. “High-dimensional hybrid feature selection using interaction information-guided search”. In: *Knowledge-Based Systems* 145 (2018), pp. 59–66.

[47] Mohsen Paniri, Mohammad Bagher Dowlatshahi, and Hossein Nezamabadi-pour. “MLACO: A multi-label feature selection algorithm based on ant colony optimization”. In: *Knowledge-Based Systems* 192 (2020), p. 105285. ISSN: 0950-7051.

[48] Marjan Kuchaki Rafsanjani and Mohammad Bagher Dowlatshahi. “Using gravitational search algorithm for finding near-optimal base station location in two-tiered WSNs”. In: *International Journal of Machine Learning and Computing* 2.4 (2012), p. 377.

[49] Ferdinando S Samaria and Andy C Harter. “Parameterisation of a stochastic model for human face identification”. In: *Proceedings of 1994 IEEE workshop on applications of computer vision.* IEEE. 1994, pp. 138–142.

[50] Samaneh Samiei, Mehdi Joodaki, and Nasser Ghadiri. “A Scalable Pattern Mining Method Using Apache Spark Platform”. In: *2021 7th International Conference on Web Research (ICWR).* IEEE, pp. 114–118. ISBN: 1665404264.

[51] Sadia Sharmin et al. “Simultaneous feature selection and discretization based on mutual information”. In: *Pattern Recognition* 91 (2019), pp. 162–174. ISSN: 0031-3203.

[52] Sina Tabakhi and Parham Moradi. “Relevance–redundancy feature selection based on ant colony optimization”. In: *Pattern recognition* 48.9 (2015), pp. 2798–2811. ISSN: 0031-3203.

[53] Sina Tabakhi, Parham Moradi, and Fardin Akhlaghian. “An unsupervised feature selection algorithm based on ant colony optimization”. In: *Engineering Applications of Artificial Intelligence* 32 (2014), pp. 112–123. ISSN: 0952-1976.

[54] Sina Tabakhi, Parham Moradi, and Fardin Akhlaghian. “An unsupervised feature selection algorithm based on ant colony optimization”. In: *Engineering Applications of Artificial Intelligence* 32 (2014), pp. 112–123. ISSN: 0952-1976.

[55] El-Ghazali Talbi. *Metaheuristics: from design to implementation.* Vol. 74. John Wiley Sons, 2009. ISBN: 0470496908.

[56] Youchuan Wan et al. “A feature selection method based on modified binary coded ant colony optimization algorithm”. In: *Applied Soft Computing* 49 (2016), pp. 248–258. ISSN: 1568-4946.

[57] Guangfen Wei et al. “A novel hybrid feature selection method based on dynamic feature importance”. In: *Applied Soft Computing* 93 (2020), p. 106337. ISSN: 1568-4946.

[58] Ian H Witten and Eibe Frank. “Data mining: practical machine learning tools and techniques with Java implementations”. In: *Acm Sigmod Record* 31.1 (2002), pp. 76–77.

[59] Guohua Wu, Rammohan Mallipeddi, and Ponnuthurai Nagaratnam Suganthan. "Ensemble strategies for population-based optimization algorithms–A survey". In: *Swarm and evolutionary computation* 44 (2019), pp. 695–711. ISSN: 2210-6502.

[60] Xiao Yu, Qing Li, and Jin Liu. "Scalable and parallel sequential pattern mining using spark". In: *World Wide Web* 22.1 (2019), pp. 295–324. ISSN: 1573-1413.

[61] Jixiong Zhang, Yanmei Xiong, and Shungeng Min. "A new hybrid filter/wrapper algorithm for feature selection in classification". In: *Analytica chimica acta* 1080 (2019), pp. 43–54.

[62] Wenhao Zhao, Li Yan, and Yunsheng Zhang. "Geometric-constrained multi-view image matching method based on semi-global optimization". In: *Geo-spatial information science* 21.2 (2018), pp. 115–126. ISSN: 1009-5020.

Nazanin Zahra Joodak
Orcid number: 0000-0002-3954-6751
Department of Computer Engineering
Faculty of Engineering
Lorestan University
Khorramabad, Iran
Email address: `joodaki.nz@fe.lu.ac.ir`

Mohammad Bagher Dowlatshahi
Orcid number: 0000-0002-4862-0626
Department of Computer Engineering
Faculty of Engineering
Lorestan University
Khorramabad, Iran
Email address: `dowlatshahi.mb@lu.ac.ir`

Mehdi Joodaki
Orcid number: 0000-0002-2184-7118
Department of Computer Engineering
Isfahan University of Technology
Isfahan, Iran
Email address: `mehdi.joodaki@ec.iut.ac.ir`