Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 12, Number 1, Spring 2023, pp. 20-28
Received: 06/04/2020, Revised: 03/30/2021, Accepted: 04/24/2021

Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach

Amin Hashemi[1], Mohammad Bagher Dowlatshahi[2*] and Hossein Nazamabadi-pour[3]
1- Department of Computer Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran.
2*- Department of Computer Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran.
3- Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran
[1]hashemi.am@fe.lu.ac.ir, [2*]Dowlatshahi.mb@lu.ac.ir, and [3]Nezami@uk.ac.ir

Corresponding author's address: Mohammad Bagher Dowlatshahi, Department of Computer Engineering, Faculty of Engineering, Lorestan University, Khorramabad, Iran

Abstract- Ensemble feature selection methods are used to improve the robustness of feature selection algorithms. These approaches are a combination of several feature selection methods to achieve the final ranking of features. The reason for using such approaches is derived from the fact that the variety of different methods is more effective than only one method. Each feature selection algorithm may find feature subsets that can be considered local optima in the feature subsets space. Ensemble feature selection is a solution to address this problem. In this paper, we have proposed a bi-objective feature selection algorithm based on Pareto-based ranking. The maximum relevancy and minimum redundancy are considered as our two objectives. Both of the objectives are obtained by the ensemble of three feature selection methods. The final evaluation of features is according to a bi-objective optimization process and the crowding distance of features in this space for ranking the features. The proposed method results are compared with recent ensemble feature selection algorithms and simple feature selection algorithms. The results show that our classification accuracy method is superior to other similar methods and performs in a short time.

Keywords- Ensemble feature selection, Pareto-based Ranking, Bi-objective optimization, Crowding distance.

I. INTRODUCTION

Using high-dimensional data is widespread in different machine learning algorithms, and they have caused some problems like inadequate memory and time-consuming in the learning process. Many fields, such as data mining, image processing, bioinformatics, need to find the best features among their high-dimensional data. Feature selection [12], [13] is an effective way to deal with these problems. Feature selection as a dimensionality reduction technique selects essential features and removes the irrelevant and redundant ones from the dataset. This procedure can reduce the dimensionality of data, the computational cost, time complexity, storage complexity, and also increase the classification accuracy [7], [17], [33], [35]. We can categorize feature selection methods from two perspectives: label prospective and search strategy [30]. Feature Selection has three modes base on labels: supervised, semi-supervised, and unsupervised [57]. A supervised feature selection method selects features based on label information. Supervised methods are applied to single-label (SL) and multi-label (ML) data [3], [21], [24]–[26], [39]–[41]. In ML data, unlike SL, each sample contains more than one label, and the correlation between features and labels is considered as feature relevancy [42], [57]. Unsupervised methods refer to the methods that evaluate the features based on training samples without class labels [57]. Semi-supervised methods work with a type of data that includes some labeled samples [49].

Based on the search strategy, feature selection techniques can be divided into filter, embedded, and wrapper methods. In the filter-based methods, each feature will be assessed by static experiment and essential specification data; hence it is fast for high-dimensional data [22], [54]. On the other hand, wrapper methods evaluate the features in the learning procedure. These methods assess the feature subsets to

determine the best values according to their predictive power [33]. Finally, embedded methods seek to address the weaknesses of the two previous techniques and use the strengths of both

The ensemble of feature selection algorithms is a new technique that has recently been introduced. In this approach, the combination of several feature selection methods is used to obtain the final feature subset. The rationale for this approach is based on the old proverb "Two heads are better than one" and on the fact that the variety of different methods makes the feature selection method work better than only one method. Each feature selection algorithm may find feature subsets that can be considered local optima in the feature subsets space. Ensemble feature selection can be helpful to address this problem [4].

An optimization process [10], [11], [14]–[16], [36], [37], [44], [45] is finding the best solution out of all the possible solutions. Optimization is widely used in real-world issues. Many optimization problems consist of multiple objectives that need to be optimized simultaneously. These problems are called multi-objective optimization problems and occur in many engineering and medical tasks [31]. The conflict between objectives leads to a set of trade-off solutions called the Pareto optimal set instead of an optimal one. The solutions in the Pareto optimal set are called non-dominated solutions. A non-dominated solution is a solution that is not worse than any other solution and is better than at least one solution, which means that there is no solution to dominate it. On the other hand, Pareto-based methods have shown excellent performance with two or three objectives.

An ensemble of feature selection algorithms called PEFS [23] is recently proposed based on a Pareto-based approach. This method used a two objectives process to rank the features based on four feature selection algorithms. The average and minimum rank of each feature based on feature selection methods are considered as the objectives. An issue that is not addressed in this method is considering the dependence on the class label and the redundancy of the features independently. This article has proposed a feature selection algorithm based on the maximum relevancy with the class label and minimum redundancy to improve the performance of the PEFS method. In the **m**inimum **R**edundancy **M**aximum **R**elevance **E**nsemble **F**eature **S**election (mRMR-EFS) method, we model the ensemble feature selection problem to a Pareto-based optimization problem with two objectives. These two objectives are the minimum redundancy and maximum relevancy, and for each of them, an ensemble of three feature selection methods is considered. The ensemble of three relevancy-based and three redundancy-based methods are regarded as our two conflicting objectives. Thus, the combination of ensemble feature ranking methods and the concept of Pareto dominance give higher ranking to the features that are not dominated by other features. We first combined the results obtained by three different relevancy-based and redundancy-based feature selection methods using two aggregation methods in the proposed method. Finally, we evaluated these two aggregation results using a bi-objective optimization. The non-dominant features are ranked in the bi-objective space according to their crowding distance. These features are removed from the space so that other features are ranked based on this strategy. This approach can balance the relevancy with class label and redundancy of features and give higher ranking to the most relevant features with the class label and the least redundancy. To show the superiority of the proposed method over other methods, six real data sets, including five biological data and one image data with different dimensions, have been used. In the obtained experiments, the proposed method in most datasets has a significant improvement in terms of classification accuracy and algorithm runtime compared to other methods.

The structure of this paper is organized as follows: Section II deals with reviewing related methods. Section III will describe the fundamental concepts used in the proposed method, and section IV represents the proposed method in detail. Section 5V includes the experimental results, and Section VI represents the conclusion and expression of future works.

II. RELATED WORKS

Ensemble-based Filter Feature Selection (EFFS) [52] is a heterogeneous approach for wearable sensor-based human activity recognition. In this method, the aggregation of four filter feature ranking methods includes information gain, gain ratio, chi-squared, and ReliefF, are considered to achieve the final feature set. The aggregation method used in EFFS is a weighted mean, and the weight for each technique is obtained from multiple experiments.

In [48], two different homogeneous and heterogeneous approaches are presented. They used several aggregators for combining the results of various rankers. In [6], an ensemble feature selection method is presented for high dimensional data. In this method, the combination of different rankers is done based on the reliability assessment-based aggregation (RAA) technique. Das et al. [9] proposed a new ensemble FS method based on a bi-objective genetic algorithm. This algorithm tries to find the best subset of dynamic mating features and considers rough set theory and information theory as their two objective functions. Ansari et al. [1] proposed an ensemble FS method for sentiment classification. They used the concept of hesitant fuzzy sets to combine the results of different filter FS methods. This approach selects top-k ranked features based on the relevancy score. In [20], a technique based on Maximum Relevancy and Minimum Redundancy (MRMR) is presented for ensemble FS using Hesitant Fuzzy Sets (HFSs). This algorithm is a filter-based method, and the results of different rankers are combined using the concept of information energy of HFSs. In [18], several ensemble FS methods are presented based on some basic techniques like max, min,… and election methods such as Borda-count, weighted Borda-count, and plurality voting. They also performed several clustering-based methods using the mean-shift algorithm. In these methods, eight filter-based feature selection methods are applied ReliefF, Maximum Information Coefficient (MIC), Robust Feature Selection (RFS), Gini-index, Correlation coefficient, Anova-based FS, t-test FS, and Fisher-score. All these methods are based on a rank aggregation procedure.

EFS-MI [28] is an ensemble feature selection algorithm that used the aggregation ranks assigned to features by filter

feature ranking methods. This method used gain ratio, information gain, chi-squared, ReliefF, and symmetric uncertainty as to the based feature ranking methods. The ensemble process of this method is based on feature-class and feature-feature mutual information. Recently other ensemble feature selection procedures are proposed for high dimensional datasets with a low number of instances. In this article, some serial and parallel techniques have been offered [53]. Another approach is presented to deal with ensemble feature selection algorithms' sensitivity and minimize the training error. The NSGA-III is used to find the optimal feature subsets [38].

III. FUNDAMENTAL CONCEPT

A. Ensemble Feature Selection

Ensemble feature selection is the process of combining multiple models instead of one model. The rationale for this approach is based on the old proverb "Two heads are better than one" and on the fact that the variety of different methods makes the feature selection method work better than only one method. Various studies have shown that these methods usually achieve better accuracy than individual methods because of the diversity of these approaches and control of variance. Ensembles have recently been used in machine learning techniques such as supervised classification, regression, and optimization. In the area of feature selection, these methods have also recently been widely used. Ensemble FS methods can be categorized into two main groups: The first group is called homogeneous ensembles, which exploit data diversity. The data is partitioned into multiple partitions in these methods, and a feature selection method is implemented on each partition. Finally, the results on each partition are aggregated to achieve the final feature subset (Fig. 1(a)). The second group is called heterogeneous ensembles, which exploit the diversity of functions. In these methods, multiple feature selection methods are executed on the data, and the results of these FS methods aggregate to find the best subset of features (Fig. 2(b)) [5], [4].

B. Pareto-based solutions

In multi-objective optimization problems (MOPs), the conflict between objectives leads to a set of trade-off solutions called the Pareto optimal set instead of an optimal one. This paper used minimization optimization for the Pareto optimization concepts without loss of generality [32], [51]. A MOP is formulated as follows:

$$\begin{cases} ma\,x\,F(x) = \left(f_1(x), f_2(x), \dots\dots, f_n(x)\right), \\ s.c.\,x \in S \end{cases} \quad (1)$$

where $n\ (n \geq 2)$ denotes the number of objectives, $x = (x_1, \dots, x_k)$ is the decision variables vector, and S represents the set of all candidate solutions. Finally, $F(x) = (f_1(x), f_2(x), \dots\dots, f_n(x)$ is the objectives vector that should be optimized. Since we consider the MOP a maximization problem, the vector F can be defined as a benefit function that estimates the quality of each solution.

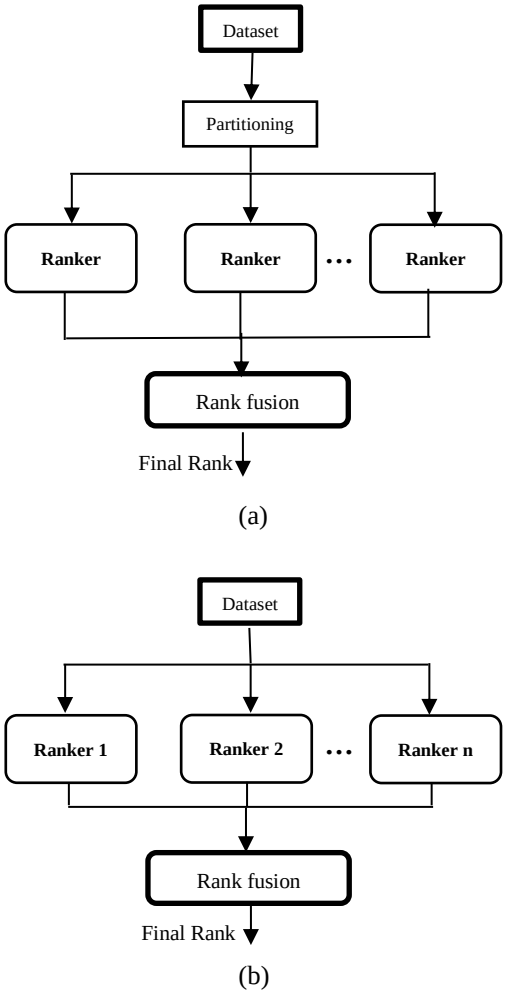

Fig. 1. Block diagram of (a) Homogeneous and (b) Heterogeneous ensemble FS.

In multi-objective optimization problems, the following concepts for minimization problems are widely discussed:

1) Pareto Dominance: An objective vector $u = (u_1, \dots\dots, u_n)$ is said to dominate $v = (v_1, \dots\dots, v_n)$ (denoted by $u \prec v$) if and only if no component of v is more significant than the corresponding component of u and at least one component of u is strictly significant, that is:

$$\forall i \in \{1, \dots, n\}: \ u_i \leq v_i \wedge \exists i \in \{1, \dots, n\}: \ u_i < v_i\,. \quad (2)$$

Eq. 2 is shown that if we say u dominates v, then all the components of u should be greater than v.

2) Pareto Optimality: A solution $x^* \in S$ is Pareto optimal if, for every $x \in S$, $F(x)$ does not dominate $F(x^*)$, i.e., $F(x) \nprec F(x^*)$.

3) Pareto Optimal Set: For a given $MOP(F, S)$, the Pareto optimal set is defined as $\mathcal{P}^* = \{x \in S \mid \nexists x' \in S, F(x') \prec F(x)\}$.

4) Pareto Optimal Front: The projection of the Pareto optimal set on the objective space is considered as Pareto optimal front.

C. Crowding Distance

The crowding distance of a solution represents the density of other solutions surrounding it. Fig. 2 shows the calculation of the crowding distance of solution S in a bi-objective space. This distance is based on estimating the largest cubed enclosing S so that any other solution is not included. The crowding distance of points a and b are set to infinite. Then, the crowding distance is obtained by normalizing the calculated distance with the difference between the maximum and minimum distance on the same Pareto front [46], [55].

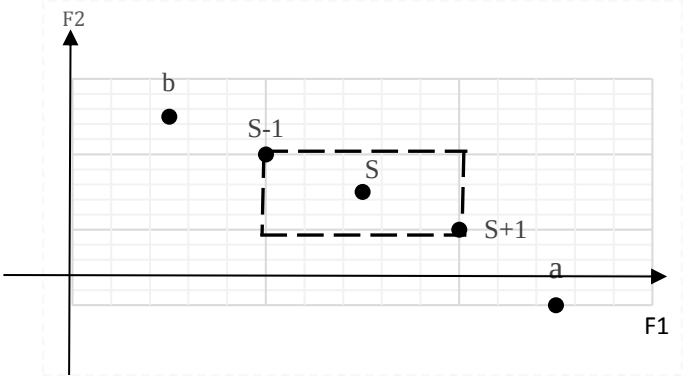

Fig. 2. Crowding distance of point S in a bi-objective space

IV. PROPOSED METHOD

In this section, we discuss our proposed method in detail. This method used a filter-based heterogeneous procedure. In this method, we used the combination of ensemble FS methods and a Pareto-based optimization procedure. Algorithm 1 shows the step-by-step process of the proposed method.

A. Motivation

Ensemble methods can be used to improve the robustness of FS algorithms. Each FS algorithm may find feature subsets that can be considered local optima in the feature subsets space. Ensemble feature selection can help address this problem by combining the outputs of multiple FS algorithms and improving accuracy.

In the real world, many problems consist of multiple objectives that need to be optimized simultaneously. The conflict between goals leads to a set of trade-off solutions called the Pareto optimal set instead of an optimal one. The solutions in the Pareto optimal set are called non-dominated solutions. A non-dominated solution is a solution that is not worse than any other solution and is better than at least one solution, which means that there is no solution to dominate it. Based on experiments, the Pareto-based methods have shown excellent performance with two or three objectives. The main reason is that with the increase in the number of the objective, almost every solution becomes non-dominated that loses the performance of Pareto-based algorithms.

Recently a new ensemble of feature selection algorithms called PEFS is proposed based on the concept of Pareto dominance. This method used a rank fusion strategy based on multiple feature selection algorithms. In this method, the relevancy between features and class labels and the redundancy of features are not modeled independently. Thus to improve the performance of this method, we have proposed a minimum redundancy maximum relevancy (mRMR) strategy. For this purpose, we used three relevancy-based and three redundancy-based feature selection methods. Therefore we have a bi-objective optimization problem with conflicting objectives.

Algorithm 1: Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach (MRMR-EFS)

Input: $N \times M$ Feature data matrix $\boldsymbol{X}$, $N \times 1$ label vector $\boldsymbol{Y}$, feature index **F**

Output: Feature ranking vector w

1. $\boldsymbol{w} = \emptyset$;
2. R1=Calculate the relevancy matrix (ranks) based on three relevancy-based FS methods;
3. R2=Calculate the redundancy matrix (ranks) based on three redundancy-based FS methods ;
4. F1=Calculate the minimum value for each feature based on the R1 matrix ;
5. F2=Calculate the average value for each feature based on the R2 matrix;
6. **While** ($\mathbf{F} \sim = \mathbf{0}$)
7. Perform non-dominated sorting on *F1* and *F2* vectors;
8. S = non-dominated feature subset;
9. Pareto-num{features $\in \boldsymbol{S}$} = $\boldsymbol{k}$;
10. $\boldsymbol{k} = \boldsymbol{k} + \mathbf{1}$;
11. F = F - S;
12. **end while**

% Sort the features with the same Pareto-num according to their crowding distance in objective space

13. $\boldsymbol{d}$ = Calculate the crowding distance of each solution (feature);
14. R = ParetoNum + 1/(1+$\boldsymbol{d}$);
15. $\boldsymbol{w}$ = sort the features based on their values in R in ascending order;

B. MRMR-EFS: The proposed minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach

Now we describe the steps of algorithm 1 as our proposed method. We define an empty vector (w) as a feature ranking vector to add its features in the first step.

In step 2, we have applied three FS methods (Fisher-Score [19], MIC [19], and Information Gain (IG)) to a dataset to achieve the feature relevance rankings (*R1*). The structure of the *R1* matrix is presented as follows:

$$\mathrm{R1} = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ \vdots & \vdots & \vdots \\ r_{1M} & r_{2M} & r_{3M} \end{bmatrix} \quad (1)$$

where the first column of the R1 matrix represents the rank of each feature assigned by IG, the second one is Fisher-score, and the last column refers to the MIC method.

To obtain our first objective vector, the minimum value of each row of the *R1* matrix is representing the relevance value as follows:

$$F1 = [mean(R1(1,:), mean(R1(2,:), \dots, mean(R1(M,:)] \quad (2)$$

In step 3, we have applied three FS methods (LLCFS [56], CFS [34], and Cosine distance) to a dataset to achieve the feature redundancy rankings (*R2*). To obtain our first objective vector, the minimum value of each row of the R1 matrix is representing the relevance value as follows:
The structure of the obtained *R2* matrix is presented as follows:

$$R2 = \begin{bmatrix} d_{11} & d_{21} & d_{31} \\ d_{12} & d_{22} & d_{32} \\ \vdots & \vdots & \vdots \\ d_{1M} & d_{2M} & d_{3M} \end{bmatrix} \quad (3)$$

where the first column of the R2 matrix represents the rank of each feature assigned by LLCFS, the second one is CFS, and the last column refers to the Cosine distance method.

To obtain our second objective vector, the average value of each row of the *R2* matrix is representing the rankings according to the redundancy-based method as follows:

$$F2 = [Mean(R2(1,:)), Mean(R2(2,:)), \dots, Mean(R2(M,:))] \quad (4)$$

In steps 6 to 12 of Algorithm (1), a non-dominated sorting with two following objectives is performed, and the algorithm assigns a Pareto number to each feature. In step 13, we set up our secondary measure to sort the same Pareto number features. To do this, the crowding distance of each feature is calculated in the bi-objective space and stored in vector *d*. In the last step of Algorithm (1), first, we normalize the crowding distance of features to a value in the interval [0, 1] and then set a score to each feature based on the following equation:

$$R = \text{ParetoNum} + 1 / (1+d) \quad (5)$$

Now we can sort the features based on their value in R in ascending order and store the results in a *w* vector that the user can select a desired number of features.

V. Experimental Results

To measure the performance of our proposed method, we have compared it with six simple FS methods which: Fisher-Score [19], MIC [19], LLCFS [56], ReliefF [47], Gini-index [19], and CFS [34]. We have also compared our method to the five ensemble FS method based on rank aggregation. These methods are E-Borda [18] which uses the Borda-count method, E-WBorda [18] uses weighted Borda-count, E-Plu [18] uses plurality voting, and PEFS [23], and EFFS [52] use the mean and minimum rank which assigned to each feature by feature selection methods.

To measure the performance of the proposed mRMR-EFS and competing methods, we used the accuracy [5] metric.

A. Datasets

We used six real-world datasets obtained to measure the performance of the proposed method with comparing methods. Table 1 contains the properties of the following datasets.

TABLE 1
Description of the Used Datasets

Dataset	Instances	Features	Domain	Reference
DLBCL	77	5470	Biology	[50]
Shipp	77	7130	Biology	[50]
Pomeroy	60	7129	Biology	[43]
Semeion	1593	257	Image	[2]
Lung	203	3313	Biology	[30]
NCI60	64	6831	Biology	[29]

B. Results

For all comparing methods, the value of each parameter is set based on the recommendations by that corresponding paper. K-nearest neighbors (K-NN) classifier is used to compare the classification performance of the comparing algorithms and considered the number of neighbors equal to 5. For each test, randomly, 60% of the samples are chosen as training data and the remaining 40% considered as test data. The reported results are averaged results achieved by 20 separate runs for each method; for testing each method, as the user determines the number of selected features, we change the size of features subset from 10 to 100, which results in 100 different runs on each dataset. In our method, the number of features is determined by the user. All the computations have been done on windows 8.1-64bit machine with Intel® Core (TM) i5-M460 and 4GB Ram, using MATLAB® 9.4.0.813654 (R2018a).

1) Comparison between the proposed method and based FS methods

Figs. 3 to 8 show the classification performance for accuracy criterion comparing based FS methods. In these figures, the horizontal axis indicates the number of selected features, and the vertical axis represents comparison criteria. To evaluate the proposed method, we used 10 different intervals for comparison. First, we have compared the different methods with the top 10 features in the ranking system, and then each time, we have added 10 features to the number of features selected by the user. For each number of features, 20 different runs have been performed and the results obtained are the average of these different runs. As a result, for each dataset, each feature selection method is performed 200 times separately.

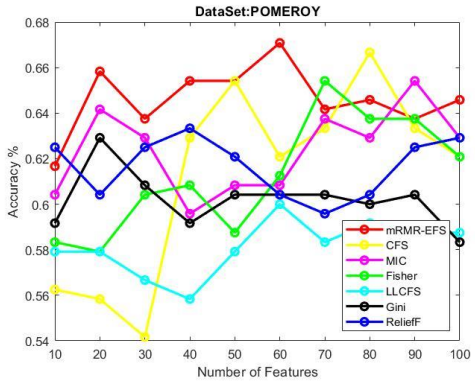

Fig. 3. Accuracy for Pomeroy dataset

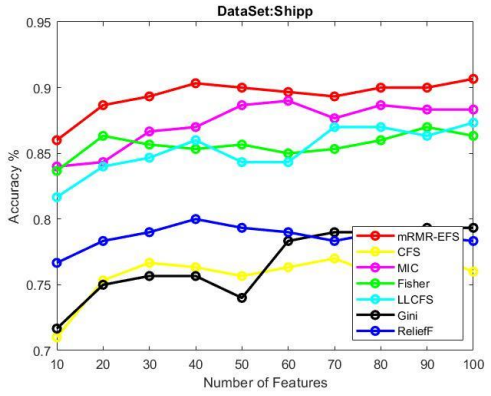

Fig. 4. Accuracy for Shipp dataset

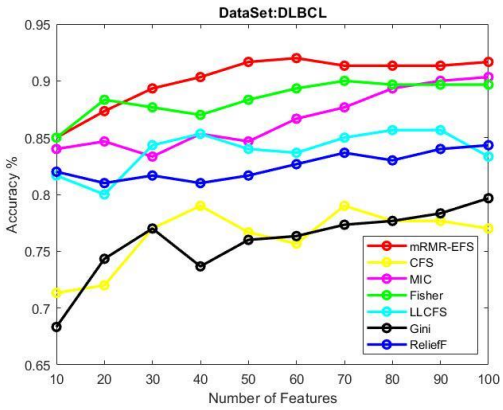

Fig. 5. Accuracy for DLBCL dataset

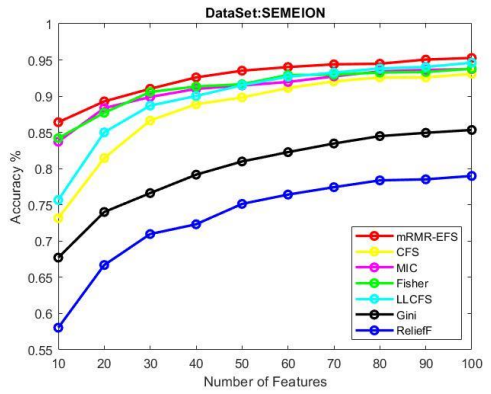

Fig. 6. Accuracy for Semeion dataset

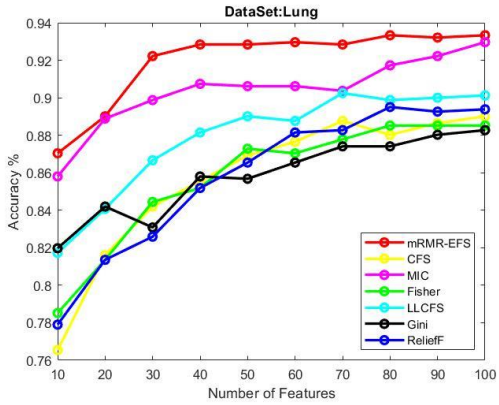

Fig. 7. Accuracy for Lung dataset

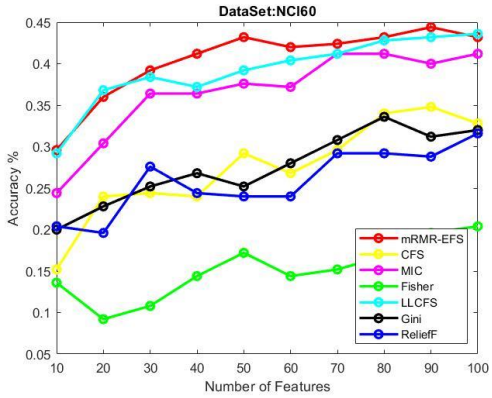

Fig. 8. Accuracy for NCI60 dataset

The obtained results of mRMR-EFS and all comparing methods are also compared statistically. For this purpose, The Friedman test [27] is applied to the obtained results. The desired significance level for the post-hoc test is set to 0.05. If the result obtained by the Friedman test is less than the significance level, we perform a test for pairwise comparison of variables according to Conover [8]. The results of the Friedman test on mRMR-EFS against the other methods on 6 datasets are shown in the last row of Tables 2. The values show the obtained p-values of each method comparing the mRMR-EFS method by the Friedman test. Also, the (+) sign indicates that our method statistically wins the comparing method, and respectively signs (=) and (-) show the tie and loss. The number of overall win/tie/loss of our method against others is shown in Table 2.

TABLE 2
THE OBTAINED P-VALUES BY THE FRIEDMAN TEST IN TERMS OF ACCURACY AGAINST BASED FS METHODS.

Dataset	mRMR - EFS VS CFS	mRMR- EFS VS LLCFS	mRMR- EFS VS Fisher-Score	mRMR- EFS VS MIC	mRMR- EFS VS Gini-index	mRMR- EFS VS Relief
Shipp	0.0005(+)	0.0005(+)	0.0018(+)	0.0004(+)	0.0004(+)	0.0005(+)
DLBCL	0.0005(+)	0.0005(+)	0.008(+)	0.0005(+)	0.0005(+)	0.0005(+)
Pomeroy	0.0087(+)	0.1659(=)	0.0087(+)	0.0153(+)	0.0016 (+)	0.0018(+)
Lung	0.0005(+)	0.0005(+)	0.0005(+)	0.0016(+)	0.0005(+)	0.0005(+)
NCI60	0.0005(+)	0.0005(+)	0.0005(+)	0.0015(+)	0.0005(+)	0.0005(+)
Semeion	0.0005(+)	0.0153(+)	0.0056(+)	0.0018(+)	0.0005(+)	0.0005(+)

TABLE 3
ACCURACY OF ENSEMBLE METHODS FOR SHIPP DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	**0.8600**	0.7133	0.7167	0.8233	0.8467	0.8000
20	**0.8867**	0.7500	0.7300	0.8733	0.8433	0.8133
30	**0.8933**	0.7500	0.7667	0.8533	0.8667	0.8400
40	**0.9033**	0.7667	0.7833	0.8567	0.8667	0.8300
50	**0.9000**	0.7600	0.7967	0.8633	0.8767	0.8200
60	**0.8967**	0.7833	0.7933	0.8600	0.8800	0.8300
70	**0.8932**	0.7767	0.7867	0.8767	0.8833	0.8433
80	**0.9000**	0.7867	0.7900	0.8883	0.8900	0.8467
90	**0.9000**	0.7700	0.7633	0.8700	0.8800	0.8533
100	**0.9067**	0.7733	0.7533	0.8800	0.8933	0.8533
Friedman Test		**0.0005 (+)**	**0.0005 (+)**	**0.0018 (+)**	**0.0005 (+)**	**0.0005 (+)**

TABLE 4
ACCURACY OF ENSEMBLE METHODS FOR DLBCL DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	**0.8500**	0.7433	0.8133	0.8433	0.8400	0.8100
20	**0.8733**	0.7833	0.8167	0.8700	0.8667	0.8300
30	0.8933	0.7733	0.8367	**0.8967**	0.8833	0.8367
40	0.9033	0.7633	0.8200	**0.9200**	0.9000	0.8200
50	0.9167	0.7533	0.8267	**0.9333**	0.9000	0.8300
60	0.9200	0.7667	0.8267	**0.9300**	0.9067	0.8367
70	0.9133	0.7900	0.8067	**0.9267**	0.9100	0.8433
80	0.9133	0.7733	0.8100	**0.9067**	0.9000	0.8467
90	0.9133	0.7833	0.8300	**0.9133**	0.8967	0.8400
100	0.9167	0.7967	0.8433	**0.9200**	0.9167	0.8427
Friedman Test		**0.0005 (+)**	**0.0004 (+)**	**0.2938(=)**	**0.0051 (+)**	**0.0005(+)**

TABLE 5
ACCURACY OF ENSEMBLE METHODS FOR SEMEION DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	**0.8641**	0.6958	0.7391	0.8378	0.8532	0.8107
20	**0.8928**	0.7783	0.8702	0.8721	0.8898	0.8920
30	0.9102	0.8251	0.9003	0.8954	**0.9124**	0.9086
40	**0.9259**	0.8440	0.9100	0.9096	0.9217	0.9166
50	**0.9352**	0.8724	0.9174	0.9220	0.9301	0.9309
60	0.9402	0.8793	0.9223	0.9331	**0.9424**	0.9367
70	0.9440	0.8895	0.9254	0.9425	**0.9476**	0.9396
80	0.9447	0.8981	0.9287	0.9440	**0.9501**	0.9388
90	**0.9505**	0.9064	0.9279	0.9469	**0.9505**	0.9429
100	**0.9527**	0.9097	0.8256	0.9493	0.9526	0.9443
Friedman Test		**0.0005 (+)**	**0.0018 (+)**	**0.0029(+)**	**1 (+)**	**0.0153(+)**

TABLE 6
ACCURACY OF ENSEMBLE METHODS FOR NCI60 DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	0.2960	0.2160	0.1840	0.2600	0.3000	**0.3280**
20	0.3600	0.2760	0.2000	0.3280	0.3280	**0.4360**
30	0.3920	0.2640	0.2120	0.3080	0.3560	**0.4080**
40	**0.4120**	0.2760	0.2160	0.3640	0.3480	0.4040
50	0.4320	0.2680	0.1960	0.3520	0.3640	**0.4400**
60	0.4200	0.3240	0.2320	0.3480	0.3840	**0.4200**
70	0.4240	0.3240	0.2720	0.3640	0.3720	**0.4400**
80	0.4320	0.3120	0.2760	0.3880	0.3960	**0.4440**
90	0.4440	0.3040	0.2560	0.3680	0.4000	**0.4640**
100	0.4320	0.3280	0.2520	0.3480	0.4120	**0.4640**
Friedman Test		**0.0005 (+)**	**0.0005 (+)**	**0.0018(+)**	**0.0056 (+)**	**0.0358(-)**

TABLE 7
ACCURACY OF ENSEMBLE METHODS FOR POMEROY DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	0.6167	**0.6333**	0.6250	0.6083	0.6125	0.6167
20	**0.6583**	0.5875	0.6375	0.5750	0.5833	0.6333
30	**0.6375**	0.6000	0.5958	0.5917	0.5833	0.5958
40	**0.6542**	0.6083	0.6042	0.6000	0.5625	0.6000
50	**0.6542**	0.6167	0.6042	0.6292	0.5542	0.5875
60	**0.6708**	0.5917	0.6125	0.6292	0.5917	0.6042
70	**0.6417**	0.5750	0.6167	0.6333	0.5875	0.5917
80	**0.6458**	0.6000	0.6208	0.6298	0.6000	0.6042
90	**0.6375**	0.5958	0.6042	0.6292	0.6042	0.6167
100	**0.6458**	0.5917	0.6167	0.6250	0.6083	0.6083
Friedman Test		**0.0018 (+)**	**0.0056 (+)**	**0.0005 (+)**	**0.0005 (+)**	**0.0029(+)**

TABLE 8
ACCURACY OF ENSEMBLE METHODS FOR LUNG DATASET.

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
10	**0.8704**	0.8148	0.8407	0.8185	0.8519	0.8111
20	**0.8901**	0.8420	0.8605	0.8704	0.8802	0.8642
30	**0.9222**	0.8667	0.8580	0.8778	0.9025	0.8926
40	**0.9284**	0.8753	0.8790	0.8815	0.9111	0.8951
50	**0.9284**	0.8926	0.9778	0.8815	0.9160	0.8914
60	**0.9296**	0.8914	0.8926	0.8988	0.9222	0.9062
70	**0.9284**	0.8963	0.9086	0.9000	0.9247	0.9123
80	**0.9233**	0.8988	0.9160	0.9062	0.9222	0.9099
90	**0.9321**	0.9074	0.9160	0.9074	0.9309	0.9198
100	**0.9333**	0.9074	0.9247	0.9074	0.9321	0.9099
Friedman Test		**0.0005 (+)**	**0.0005 (+)**	**0.0005 (+)**	**0.0029 (+)**	**0.0005(+)**

2) Comparison between the proposed method and ensemble FS methods

Tables 3 to 8 show the classification performance for accuracy criterion comparing ensemble FS methods. The obtained results of MRMR-EFS and all comparing methods are also compared statistically. For this purpose, The Friedman test [27] is applied to the obtained results. The desired significance level for the post-hoc test is set to 0.05. If the result obtained by the Friedman test is less than the significance level, we perform a test for pairwise comparison of variables according to Conover [8].

The Friedman test results on mRMR-EFS against the other methods on 6 datasets are shown in the last row of Tables 3-8. In the last row of tables, the p-values of each method comparing the MRMR-EFS method obtained by the Friedman test are recorded. Also, symbol (+) in the last row of each table indicates that our method statistically wins the comparing method, and respectively signs (=) and (-) show the tie and loss. Finally, we showed the number of overall win/tie/loss of our method against base and ensemble methods in Table 9, and Table 10 represents the run-time of ensemble algorithms.

C. Discussion

In this method, we treated the problem of ensemble feature selection as a bi-objective optimization approach. We considered the FS methods as our decision matrix and tried to construct a ranking system for features by the concept of Pareto dominance. We tried to combine redundancy-based and relevancy-based FS methods to achieve the most relevant and less redundant feature subset. To do this, the ensemble of three redundancy-based methods and three relevancy-based methods are considered as our two conflicting goals. So our optimization process is a combination of redundancy and relevancy methods. We used two objectives for this optimization. These two objectives are the average and minimum values assigned by FS methods to features. We used the combination of two ensemble feature selection method to improve the performance of the proposed method compared to the competitive methods. In this paper, we tried to improve another Pareto-based method called PEFS proposed for ensemble feature selection. This method just considers the optimization based on some feature selection methods and has not modeled the relevancy-based and redundancy-based methods separately.

We classified the results of the proposed method compared to other methods into two groups. The first group is the results of comparing the proposed method against based FS methods. The results of this category are presented in Figs 3 to 8. The second group is the results of comparing the proposed method against ensemble FS methods. These results are also presented in Tables 3 to 8. The overall win/tie/loss in Table 9 and the averaged run-time of algorithms are recorded in Table 10. The results show that our method is superior to other methods in all evaluation criteria. According to these results, the MRMR-EFS method is swift, and according to the values in Table 10, the proposed method is so much faster than all ensemble methods. If we consider d as the number of features and L as the number of simple feature selection methods used in the ensemble process. In our method, we used non-dominated

sorting, mean, and min functions whose computational complexity is $O(d^2L)$, $O(dL)$, and $O(dL)$, respectively. Thus we can say that the computational complexity of our method is $O(d^2L + 2dL)$. Since the value of the L parameter is constant in our method and it is equal to 3, then we can conclude that the overall computational complexity is $O(d^2)$.

TABLE 9
THE WIN/TIE/LOSS RESULTS OF MRMR-EFS AGAINST THE ENSEMBLE FS METHODS BASED ON FRIEDMAN TEST

MRMR-EFS against	Accuracy (win/tie/loss)
E-Borda	6/0/0
E-Wborda	6/0/0
E-Plu	5/1/0
PEFS	5/1/0
EFFS	5/0/1
CFS	6/0/0
Fisher-Score	6/0/0
LLCFS	5/1/0
MIC	6/0/0
ReliefF	6/0/0
Gini-index	6/0/0
Total	62/3/1

TABLE 10
THE AVERAGE RUNTIME OF MRMR-EFS AGAINST THE ENSEMBLE FS METHODS

M/fea	mRMR-EFS	E-Borda	E-Wborda	E-Plu	PEFS	EFFS
NCI60	5.84	1220	1110	1114	**4.55**	1101
DLBCL	3.64	670	791	688	**2.45**	671
Semeion	5.28	670	791	688	**3.27**	671
Shipp	5.51	2008	2013	2010	**3.67**	2007
Lung	8.30	3059	3063	3057	**6.87**	3050
Pomeroy	4.68	3763	3770	3760	**3.20**	3757

VI. CONCLUSION AND FUTURE WORKS

In this work, we proposed a new feature selection algorithm for ensemble learning called, the mRMR-EFS which maps the features selection process to a Pareto-based procedure. This method is based on a filter-based strategy and used a heterogeneous approach for ensemble learning. In this method, after we obtained the scores of features based on multiple FS methods and construct objective vectors, we deliver this data as our objectives to a non-dominated sorting method. The results of different datasets show the optimality and efficiency of the proposed method. The disadvantage of this method is that it considers the impact of all feature selection algorithms equally in the ensemble process. We think that for achieving better accuracy, a weighting strategy can be useful. We intend to improve our work using a weighting strategy for feature selection algorithms and use other approaches for ensemble FS, especially graph-based approaches. We also try to extend our work to other types of feature selection, like online feature selection.

REFERENCES

[1] G. Ansari, T. Ahmad, and M. N. Doja, "Ensemble of feature ranking methods using hesitant fuzzy sets for sentiment classification," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 5, pp. 599–608, 2019.

[2] M. Bache, K. & Lichman, "Repository, UCI Machine Learning," *Irvine, CA: University of California*, 2013. .

[3] H. Bayati, M. B. Dowlatshahi, and M. Paniri, "MLPSO: A Filter Multi-label Feature Selection Based on Particle Swarm Optimization," in *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, 2020, pp. 1–6.

[4] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Inf. Fusion*, vol. 52, pp. 1–12, 2019.

[5] V. Bolón-Canedo and A. Alonso-Betanzos, "Evaluation of ensembles for feature selection," in *Intelligent Systems Reference Library*, vol. 147, 2018, pp. 97–113.

[6] A. Ben Brahim and M. Limam, "Ensemble feature selection for high dimensional data: a new method and a comparative study," *Adv. Data Anal. Classif.*, pp. 1–16, 2017.

[7] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[8] C. W. Coakley and W. J. Conover, "Practical Nonparametric Statistics," *J. Am. Stat. Assoc.*, vol. 95, no. 449, p. 332, 2000.

[9] A. K. Das, S. Das, and A. Ghosh, "Ensemble feature selection using bi-objective genetic algorithm," *Knowledge-Based Syst.*, vol. 123, pp. 116–127, 2017.

[10] M. B. Dowlatshahi, V. Derhami, and H. Nezamabadi-Pour, "Fuzzy particle swarm optimization with nearest-better neighborhood for multimodal optimization," *Iran. J. Fuzzy Syst.*, vol. 17, no. 4, pp. 7–24, 2020.

[11] M. B. Dowlatshahi and V. Derhami, "Winner Determination in Combinatorial Auctions using Hybrid Ant Colony Optimization and Multi-Neighborhood Local Search," *J. AI Data Min.*, vol. 5, no. 2, pp. 169–181, 2017.

[12] M. B. Dowlatshahi, V. Derhami, and H. Nezamabadi-pour, "A novel three-stage filter-wrapper framework for miRNA subset selection in cancer classification," *Informatics*, vol. 5, no. 1, 2018.

[13] M. B. Dowlatshahi, V. Derhami, and H. Nezamabadi-Pour, "Ensemble of filter-based rankers to guide an epsilon-greedy swarm optimizer for high-dimensional feature subset selection," *Inf.*, vol. 8, no. 4, 2017.

[14] M. B. Dowlatshahi, M. Kuchaki Rafsanjani, and B. B. Gupta, "An energy aware grouping memetic algorithm to schedule the sensing activity in WSNs-based IoT for smart cities," *Appl. Soft Comput.*, vol. 108, p. 107473, Sep. 2021.

[15] M. B. Dowlatshahi and H. Nezamabadi-Pour, "GGSA: A Grouping Gravitational Search Algorithm for data clustering," *Eng. Appl. Artif. Intell.*, vol. 36, pp. 114–121, 2014.

[16] M. B. Dowlatshahi, H. Nezamabadi-Pour, and M. Mashinchi, "A discrete gravitational search algorithm for solving combinatorial optimization problems," *Inf. Sci. (Ny).*, vol. 258, pp. 94–107, 2014.

[17] M. B. Dowlatshahi and M. Rezaeian, "Training spiking neurons with gravitational search algorithm for data classification," in *1st Conference on Swarm Intelligence and Evolutionary Computation, CSIEC 2016 - Proceedings*, 2016, pp. 53–58.

[18] P. Drotár, M. Gazda, and L. Vokorokos, "Ensemble feature selection using election methods and ranker clustering," *Inf. Sci. (Ny).*, vol. 480, pp. 365–380, 2019.

[19] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification," *New York John Wiley, Sect.*, vol. 10, p. l, 2001.

[20] M. K. Ebrahimpour and M. Eftekhari, "Ensemble of feature selection methods: A hesitant fuzzy sets approach," *Appl. Soft Comput. J.*, vol. 50, pp. 300–312, 2017.

[21] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality," *Expert Syst. Appl.*, vol. 142, 2020.

[22] A. Hashemi, M. Bagher Dowlatshahi, and H. Nezamabadi-pour, "VMFS: A VIKOR-based multi-target feature selection," *Expert Syst. Appl.*, p. 115224, May 2021.

[23] A. Hashemi, M. Bagher Dowlatshahi, and H. Nezamabadi-pour, "A pareto-based ensemble of feature selection algorithms," *Expert Syst. Appl.*, vol. 180, p. 115130, Oct. 2021.

[24] A. Hashemi and M. B. 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)*, 2020, pp. 1–7.

[25] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MFS-MCDM: Multi-label feature selection using multi-criteria decision making," *Knowledge-Based Syst.*, p. 106365, Aug. 2020.

[26] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "A bipartite matching-based feature selection for multi-label learning," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 2, pp. 459–475, Feb. 2021.

[27] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," *Math. Intell.*, 2017.

[28] N. Hoque, M. Singh, and D. K. Bhattacharyya, "EFS-MI: an ensemble feature selection method for classification," *Complex Intell. Syst.*, 2018.

[29] J. Z. Huang, "An Introduction to Statistical Learning: With Applications in R By Gareth James, Trevor Hastie, Robert Tibshirani, Daniela Witten," *J. Agric. Biol. Environ. Stat.*, vol. 19, no. 4, pp. 556–557, Dec. 2014.

[30] J. Li *et al.*, "Feature Selection: A Data Perspective," *ACM Comput. Surv.*, vol. 50, 2017.

[31] M. Li, S. Yang, and X. Liu, "Bi-goal evolution for many-objective optimization problems," *Artif. Intell.*, vol. 228, pp. 45–65, 2015.

[32] C. Von Lücken, B. Barán, and C. Brizuela, "A survey on multi-objective evolutionary algorithms for many-objective problems," *Comput. Optim. Appl.*, vol. 58, no. 3, pp. 707–756, 2014.

[33] J. Miao and L. Niu, "A Survey on Feature Selection," in *Procedia Computer Science*, 2016, vol. 91, pp. 919–926.

[34] K. Michalak and H. Kwasnicka, "Correlation based feature selection method," *Int. J. Bio-Inspired Comput.*, vol. 2, no. 5, pp. 319–332, 2010.

[35] W. C. Mlambo, N, "A survey and comparative study of filter and wrapper feature selection techniques," *Int. J. Eng. Sci.*, vol. 5, no. 8, pp. 57–67, 2016.

[36] E. Momeni, M. B. Dowlatshahi, F. Omidinasab, H. Maizir, and D. J. Armaghani, "Gaussian Process Regression Technique to Estimate the Pile Bearing Capacity," *Arab. J. Sci. Eng.*, Jun. 2020.

[37] E. Momeni, A. Yarivand, M. Bagher Dowlatshahi, and D. Jahed Armaghani, "An Efficient Optimal Neural Network Based on Gravitational Search Algorithm in Predicting the Deformation of Geogrid-Reinforced Soil Structures," *Transp. Geotech.*, p. 100446, 2020.

[38] W. W. Y. Ng, Y. Tuo, J. Zhang, and S. Kwong, "Training error and sensitivity-based ensemble feature selection," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 10, pp. 2313–2326, Oct. 2020.

[39] H. Noormohammadi and M. B. Dowlatshahi, "Feature Selection in Multi-label Classification based on Binary Quantum Gravitational Search Algorithm," in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, 2021, pp. 1–6.

[40] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MLACO: A multi-label feature selection algorithm based on ant colony optimization," *Knowledge-Based Syst.*, vol. 192, 2020.

[41] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-pour, "Ant-TD: Ant colony optimization plus temporal difference reinforcement learning for multi-label feature selection," *Swarm Evol. Comput.*, vol. 64, p. 100892, Jul. 2021.

[42] R. B. Pereira, A. Plastino, B. Zadrozny, and L. H. C. Merschmann, "Categorizing feature selection methods for multi-label classification," *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 57–78, Jan. 2018.

[43] S. L. Pomeroy *et al.*, "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, no. 6870, pp. 436–442, 2002.

[44] M. K. Rafsanjani and M. B. Dowlatshahi, "Using Gravitational Search Algorithm for Finding Near-optimal Base Station Location in Two-Tiered WSNs," *Int. J. Mach. Learn. Comput.*, pp. 377–380, 2012.

[45] M. K. Rafsanjani, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "Gravitational search algorithm to solve the K-of-N lifetime problem in two-tiered WSNs," *Iran. J. Math. Sci. Informatics*, vol. 10, no. 1, pp. 81–93, 2015.

[46] C. R. Raquel and P. C. Naval, "An effective use of crowding distance in multiobjective particle swarm optimization," in *GECCO 2005 - Genetic and Evolutionary Computation Conference*, 2005, pp. 257–264.

[47] O. Reyes, C. Morell, and S. Ventura, "Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context," *Neurocomputing*, vol. 161, 2015.

[48] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: Homogeneous and heterogeneous approaches," *Knowledge-Based Syst.*, vol. 118, pp. 124–139, 2017.

[49] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A Survey on semi-supervised feature selection methods," *Pattern Recognit.*, vol. 64, pp. 141–158, 2017.

[50] M. A. Shipp *et al.*, "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning," *Nat. Med.*, vol. 8, no. 1, pp. 68–74, 2002.

[51] E. G. Talbi, *Metaheuristics: From Design to Implementation*. 2009.

[52] Y. Tian, J. Zhang, J. Wang, Y. Geng, and X. Wang, "Robust human activity recognition using single accelerometer via wavelet energy spectrum features and ensemble feature selection," *Syst. Sci. Control Eng.*, 2020.

[53] C. F. Tsai and Y. T. Sung, "Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches," *Knowledge-Based Syst.*, 2020.

[54] B. Venkatesh and J. Anuradha, "A review of Feature Selection and its methods," *Cybern. Inf. Technol.*, vol. 19, no. 1, pp. 3–26, 2019.

[55] W. Xu, A. Chong, O. T. Karaguzel, and K. P. Lam, "Improving evolutionary algorithm performance for integer type multi-objective building system design optimization," *Energy Build.*, vol. 127, pp. 714–729, 2016.

[56] H. Zeng and Y. M. Cheung, "Feature selection and kernel learning for local learning-based clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1532–1547, 2011.

[57] R. Zhang, F. Nie, X. Li, and X. Wei, "Feature selection with multi-view data: A survey," *Inf. Fusion*, vol. 50, pp. 158–167, 2019.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۲، شماره ۱، بهار ۱۴۰۲، صفحه ۲۰-۲۸
دریافت: ۱۳۹۹/۰۳/۱۵، بازنگری: ۱۴۰۰/۰۱/۱۰، پذیرش: ۱۴۰۰/۰۲/۰۴

انتخاب ویژگی شورایی مبتنی بر حداقل افزونگی، حداکثر همبستگی: یک رویکرد دو هدفه بر اساس مفهوم غلبه پارتو

امین هاشمی[۱]، محمد باقر دولتشاهی*[۲]، حسین نظام آبادی پور[۳]
۱- دانشکده فنی و مهندسی، دانشگاه لرستان، خرم آباد، ایران.
*۲- دانشکده فنی و مهندسی، دانشگاه لرستان، خرم آباد، ایران.
۳- دانشکده مهندسی برق ، دانشگاه شهید باهنر، کرمان، ایران.
[1]hashemi.am@fe.lu.ac.ir, [2*]dowlatshahi.mb@lu.ac.ir, and [3]nezam@uk.ac.ir

* نشانی نویسنده مسئول: محمد باقر دولتشاهی، خرم آباد، دانشگاه لرستان، دانشکده مهندسی فنی و مهندسی، گروه مهندسی کامپیوتر.

چکیده- برای بهبود الگوریتم‌های انتخاب ویژگی، روش‌های شورایی مورد استفاده قرار می‌گیرند. در این رویکردها نتایج چندین روش انتخاب ویژگی با هم ترکیب می‌شوند تا مجموعه ویژگی نهایی حاصل شود. انتخاب ویژگی شورایی بر اساس این حقیقت است که تنوع روش‌های انتخاب ویژگی بهتر از تنها یک روش عمل می‌کند. هر الگوریتم انتخاب ویژگی ممکن است یک اپتیموم محلی را در فضای ویژگی‌ها در نظر بگیرد. در نتیجه روش‌های انتخاب ویژگی شورایی برای حل این مشکلات مورد استفاده قرار می‌گیرند. در این مقاله ما یک الگوریتم انتخاب ویژگی شورایی بر اساس رتبه‌دهی مبتنی بر مفهوم غلبه پارتو برای بهبود دقت دسته‌بندی روش‌های انتخاب ویژگی شورایی حاضر و روش‌های پایه انتخاب ویژگی ارائه داده‌ایم. این روش با استفاده از یک فرآیند بهینه‌سازی دوهدفه و مفهوم فاصله ازدحام، ویژگی‌ها در این فضا و در نظر گرفتن میزان همبستگی با برچسب کلاس و نیز افزونگی هر ویژگی به رتبه‌دهی آنها می پردازد. ما این روش را با روش‌های انتخاب ویژگی شورایی جدید و الگوریتم‌های پایه انتخاب ویژگی مقایسه کرده‌ایم. نتایج نشان‌دهنده برتری روش در معیار دقت دسته‌بندی است و همچنین در زمان کوتاه‌تری نسبت به سایر روش‌ها اجرا می‌شود.

واژه‌های کلیدی: انتخاب ویژگی شورایی ، فاصله ازدحامی ، بهینه‌سازی دو هدفه، رتبه‌دهی مبتنی بر مفهوم غلبه پارتو