SCIENCE CHINA
Information Sciences

February 2024, Vol. 67, Iss. 2, 120101:1–120101:15
https://doi.org/10.1007/s11432-023-3864-6

• RESEARCH PAPER •
Special Topic: Evolutionary Learning and Optimization

An evolutionary multiobjective method based on dominance and decomposition for feature selection in classification

Jing LIANG[1,2,3], Yuyang ZHANG[1,2], Ke CHEN[1,2*], Boyang QU[4], Kunjie YU[1,2], Caitong YUE[1,2] & Ponnuthurai Nagaratnam SUGANTHAN[5]

[1]*School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou* 450001*, China;*
[2]*State Key Laboratory of Intelligent Agricultural Power Equipment, Luoyang* 471000*, China;*
[3]*School of Electrical Engineering and Automation, Henan Institute of Technology, Xinxiang* 453003*, China;*
[4]*School of Electronics and Information, Zhongyuan University of Technology, Zhengzhou* 450007*, China;*
[5]*KINDI Center for Computing Research, College of Engineering, Qatar University, Doha* 999043*, Qatar*

Received 17 April 2023/Revised 30 June 2023/Accepted 14 August 2023/Published online 26 January 2024

Abstract Feature selection in classification can be considered a multiobjective problem with the objectives of increasing classification accuracy and decreasing the size of the selected feature subset. Dominance-based and decomposition-based multiobjective evolutionary algorithms (MOEAs) have been extensively used to address the feature selection problem due to their strong global search capability. However, most of them face the problem of not effectively balancing convergence and diversity during the evolutionary process. In addressing the aforementioned issue, this study proposes a unified evolutionary framework that combines two search forms of dominance and decomposition. The advantages of the two search methods assist one another in escaping the local optimum and inclining toward a balance of convergence and diversity. Specifically, an improved environmental selection strategy based on the distributions of individuals in the objective space is presented to avoid duplicate feature subsets. Furthermore, a novel knowledge transfer mechanism that considers evolutionary characteristics is developed, allowing for the effective implementation of positive knowledge transfer between dominance-based and decomposition-based feature selection methods. The experimental results demonstrate that the proposed algorithm can evolve feature subsets with good convergence and diversity in a shorter time compared with 9 state-of-the-art feature selection methods on 20 classification problems.

Keywords evolutionary algorithms, feature selection, multiobjective optimization, knowledge transfer, classification

1 Introduction

Classification involves categorizing unknown samples into different groups based on their features [1]. With the development of information technology in today's world, the dataset is to be classified as a higher dimensionality [2–4]. However, irrelevant and redundant features present in such data do not contribute to class prediction and may result in poor classification performance due to the "curse of dimensionality" [5, 6]. One popular method of dimensionality reduction is feature selection, which can boost classification accuracy and enhance computational efficiency by selecting subsets of original features that are relevant [5].

Feature selection is often viewed as a multiobjective problem with two potentially conflicting goals: minimizing the classification error rate and minimizing the size of feature subsets [7–9]. The methods of feature selection are generally classified into three categories: filter [10], wrapper [11–14], and embedded methods [15]. Filter methods rank all features based on a few criteria, such as relevance, distance, or information gain. Then, features with lower ranks are filtered out, and some valid features are selected. However, this method does not consider the interaction between features. Thus, it consumes the least

* Corresponding author (email: chenkezixf@zzu.edu.cn)

computational cost but is the least effective. The wrapper method directly evaluates the classification performance of feature subsets based on a machine-learning classifier, which performs better than the filter method, even though the wrapper method usually consumes many computing resources. The computational cost of the embedded method is higher than the filter method and lower than the wrapper method because the embedded method performs algorithms and model training to obtain feature weights to select. After weighing computational performance and efficiency, the wrapper method can be viewed as an effective feature selection method used in this report.

For contemporary researchers, it is popular to use multiobjective evolutionary algorithms (MOEAs) based on a population search mechanism for feature selection. Dominance-based and decomposition-based feature selection methods are extensively used due to their outstanding performance. Dominance-based feature selection methods assign all individuals to fronts with different ranks based on the dominant relationship between two objectives. When applied to the feature selection problem [16–18], the individuals on the Pareto front have the highest classification accuracy or the smallest feature subset size. The individuals on other fronts who are more than those on the Pareto front are usually considered suboptimal solutions. Therefore, they can provide more choices of feature combinations for evolution. Additionally, to maintain diversity during the evolution, NSGA-II [19] uses the crowding distance metric to select dissimilar parents to generate offspring and environmental selection of offspring. However, the dominance-based MOEAs are slow to converge due to the high randomness of the search and the need to weigh the two objectives and calculate the front surface to which all individuals belong.

Decomposition-based feature selection methods decompose the original problem into multiple subproblems [20–22]. The solution obtained by convergence on each subproblem forms the final Pareto front. The best-known decomposition-based multiobjective evolutionary algorithm framework is MOEA/D [23]. These subproblems are usually represented as weight vectors reflecting different preferences for the two objectives, and the convergence of all subproblems is governed based on an ideal point. This expedites the algorithm to converge to the promising region. The search experience of adjacent subproblems is utilized to generate offspring and update the subpopulation in the current subproblem during the evolutionary process. However, this operation may result in an elitist solution repeatedly appearing on multiple subproblems. This can lead to a decrease in diversity, waste of computational resources, and a lack of exploration in promising regions of the decision space, thus falling into local optima. Therefore, the research on evolutionary multiobjective feature selection methods remains a challenge.

Some current studies also combine dominance and decomposition methods. Chen et al. [24] proposed to define a specific D-dominance (DrEA) in each subpopulation after decomposition. The decomposition-based crowding metric was then designed to retain the solutions with good distributions. Li et al. [25] proposed a paradigm (MOEA/DD) combining dominance-based and decomposition-based approaches to solve optimization problems with many objectives. The objective space is divided into different subregions by designing to generate widely distributed weight vectors in a high-dimensional objective space. In the population update phase, a hierarchical selection in the form of a combination of Pareto dominance, local density estimation, and scaling functions is proposed. These two studies found that combining dominance and decomposition approaches is an effective approach to balancing diversity and convergence. The success lies in considering the preferences of both objectives and population distribution in the evolutionary process. However, the above two algorithms are only limited to combining dominance and decomposition in the environmental selection strategy to guarantee the diversity of the objective space without fully exploring the promising regions of the decision space, and their combination strategies do not consider the evolutionary state of the population. Therefore, it is crucial to design a framework that considers the evolutionary state of individuals and populations while combining dominance and decomposition methods.

Based on the above, we proposed a multiobjective feature selection method (DDFS) combining dominance and decomposition. Considering the diversity of solutions of NSGA-II in the evolutionary process and the fast convergence capability of MOEA/D, they are applied to two subpopulations to coevolve by knowledge transfer [26–32]. The proposed method aims to overcome the shortcomings of the two original MOEAs and use different search methods to balance diversity and convergence. The final achievement is to find feature subsets with higher classification accuracy and to select fewer features. The main contributions of this paper are as follows:

(1) A new framework for multiobjective feature selection based on dominance and decomposition is proposed for selecting diverse feature subsets with high classification accuracy. It aims to improve the dominance-based and decomposition-based algorithms according to problem characteristics while allowing

the subpopulations of the two search methods to capture each other's search knowledge and evolve toward the global optimum. This is the first study to construct dominance-based and decomposition-based subpopulations to address the multiobjective feature selection problem.

(2) A new environmental selection strategy in the decomposition-based subpopulation is proposed to avoid duplicate solutions. Specifically, the value of the angle between each individual and the origin of the objective space is used as the reference to assign the weight vector, allowing the adjustment of solutions to more suitable weight vectors. During the environmental selection, only one parent and one child are present on each weight vector to prevent repeated substitutions between neighbors. The diversity of solutions in the population is enhanced.

(3) A knowledge transfer mechanism is proposed to increase the diversity of decomposition-based subpopulations while accelerating the convergence speed of dominance-based subpopulations. This mechanism uses the elitist solutions of the decomposed subpopulation to accelerate the convergence of subpopulations based on dominance. Solutions on non-Pareto fronts are used to improve the diversity of decomposition-based subpopulations.

2 Related work

2.1 Dominance-based feature selection methods

In the multiobjective feature selection problem, reducing the classification error rate and reducing the number of selected features are often contradictory objectives. They must be weighed simultaneously during the optimization process. The objective of dominance-based MOEAs is to evaluate the dominance relationship by comparing the goodness of these two objectives between two possible solutions. For a minimization task, as shown in (1), $\boldsymbol{y}$ dominates $\boldsymbol{x}$ if neither of the two objectives of $\boldsymbol{x}$ is better than the two objectives of $\boldsymbol{y}$.

$$\forall i : f_i(\boldsymbol{y}) \leqslant f_i(\boldsymbol{x}) \wedge \exists j : f_j(\boldsymbol{y}) < f_j(\boldsymbol{x}). \tag{1}$$

The set of multiple solutions not dominated by any other solution is known as the Pareto optimal solution set. This can offer the user multiple choices. When applied to feature selection tasks, dominance-based MOEAs can identify feature subsets with the lowest classification accuracy or the fewest selected features. The classical dominance-based multiobjective algorithms are the NSGA-II [19] and SPEA-2 [33]. They both use information about crowding and density to maintain the diversity of the population. However, they often fall into local optima rapidly due to the large search space of feature selection.

Improving the search strategy is a popular research direction in dominance-based approaches. Huang et al. [17] proposed a feature selection method based on NSGA-II to predict customer loss in telecommunications services. The original NSGA-II is modified to search for global nondominated feature subsets based on selected local optimal solutions of different sizes. Cheng et al. [34] designed a population steering matrix to reflect the importance of features and the importance of individuals in the population. This matrix reduces irrelevant features and guides the population to search for promising regions. However, the steering matrix needs to be updated frequently, along with the calculation of the importance of each individual. Therefore, this method consumes a large amount of computational cost. Cheng et al. [35] proposed a feature selection method based on refining the granularity where a granularity contains multiple features. As features are removed from the granularity during evolution, some feature combinations can eventually be found where the features are relevant. However, this method faces the drawback of losing important features as the granularity is eliminated.

Some improved feature selection algorithms for the environmental selection phase have also achieved good results. Xu et al. [16] proposed a dominance-based multiobjective algorithm to remove duplicate feature subsets in the decision space. High-quality solutions are selected by analyzing the distances of solutions in the objective space. Wang et al. [18] proposed a feature selection method by modifying the dominance relationship into a multiobjective evolutionary algorithm of grid dominance. It aims to preserve diverse solutions for the next generation and enhance the convergence pressure using the grid dominance method. However, these improved methods do not always identify nondominated feature subsets on the edge of the Pareto front. Wang et al. [36] proposed a niche-based multiobjective feature selection approach. First, an improved mutation operator combines the niche and global information to help the offspring jump out of local optima. Second, a newly developed environment selection mechanism

allows storing the same feature subsets by relaxing the Pareto dominance relationship. However, niche-based multiobjective feature selection algorithms usually involve many sensitive parameters, such as radius size and number of niches. The difficulty in determining their optimal values has a substantial impact on the performance of the algorithm.

In summary, dominance-based approaches usually have good population diversity. However, these methods still face the limitation that they only focus on the center of the Pareto front while ignoring the search on the edge of the Pareto front.

2.2 Decomposition-based feature selection methods

Another way to solve the multiobjective problem is to decompose the original problem into multiple subproblems. Reference points or weight vectors generally represent these subproblems. When applied to the feature selection problem, the reference points represent the proportion of features selected. Different weight vectors represent different preferences for the two conflicting goals and are used to evaluate individual fitness. As shown in (2),

$$\text{fitness} = \alpha \times f_1 + (1 - \alpha) \times f_2, \tag{2}$$

where f_1 and f_2 represent the classification error rate and the feature selection ratio, the different values of α make the population converge to the Pareto front from multiple directions in the objective space.

Demir et al. [20] found that decomposition-based feature selection algorithms usually find only low-quality feature subsets in the central region of the Pareto front. Moreover, on the same dataset, the feature subsets found by MOEA/D usually have a lower classification accuracy than those found by the dominance-based algorithm. Nguyen et al. [22] proposed a decomposition-based framework based on reference points to address feature selection. This framework uses static reference points to decompose the search direction. Dynamic reference points are used to enhance the search in objective conflict regions. However, the decomposition-based approach generates many similar or duplicate solutions in the neighborhood-based reproduction and environmental selection, thus substantially reducing the diversity of solutions in the population. Jiao et al. [32] proposed a decomposition-based multiform framework to solve the multiobjective feature selection problem. Two simplified single-objective auxiliary tasks were constructed to explore new regions in the objective space, which provided useful information for the multiobjective main task. Furthermore, the distribution of solutions in the objective space filters similar or duplicate feature subsets to improve population diversity. However, knowledge transfer in this method is only performed in the environmental selection phase, failing to fully explore the feature space. Liang et al. [37] proposed a multiform framework in which many single-objective tasks assist a multiobjective task. First, the similarity of the weight vectors between the elite solutions of the main task and auxiliary tasks is considered to improve the probability of positive knowledge transfer. Second, a diversity enhancement mechanism is designed to improve the searchability of the algorithm in promising regions. The experimental results showed that the proposed algorithm performs better than other advanced feature selection algorithms, especially for high-dimensional problems.

In summary, decomposition-based methods usually converge quickly to promising regions because the complex multiobjective feature selection problem is decomposed into multiple single-objective subproblems. However, these methods generally require frequent calculations of distances and updates of reference points or weight vectors. Therefore, there is often a high computational cost in the decomposition-based method.

2.3 Motivation

The analysis in Subsections 2.1 and 2.2 shows that the dominance-based search method faces the problem of slow convergence, whereas the decomposition-based method tends to fall into local optima. However, diverse solutions can usually be found by dominance-based algorithms. In addition, the decomposition-based algorithms can converge quickly to some promising regions. Inspired by complementary strengths, we proposed combining these two search methods into a novel evolutionary framework that can achieve the following advantages. First, this novel evolutionary framework can escape local optima and find diverse feature subsets with better classification accuracy. Second, dominance-based and decomposition-based subpopulations can exploit their strengths to search for and share knowledge. Therefore, this report explores how to develop the exchange of experiences and complementary advantages between these two search approaches.

Algorithm 1 Framework of the proposed DDFS algorithm

Input: N: population size.
Output: $\mathcal{P}_N$ and $\mathcal{P}_M$: decision variables of dominance-based and decomposition-based subpopulation.
1: $\mathcal{P}_N \leftarrow$ Initialize $N/2$ solutions for the dominance-based subpopulation;
2: $\mathcal{P}_M \leftarrow$ Initialize $N/2$ solutions for the decomposition-based subpopulation;
3: **while** the termination criterion is not met **do**
4: Select $N/2$ parents for the dominance-based subpopulation based on the tournament method;
5: Select N parents for the decomposition-based subpopulation based on neighborhoods;
6: $\mathcal{O}_N \leftarrow$ Offspring generation for $\mathcal{P}_N$ using the GA operator;
7: $\mathcal{O}_M \leftarrow$ Offspring generation for $\mathcal{P}_M$ using the GA-half operator;
8: Evaluate the objective values of $\mathcal{O}_N$ and $\mathcal{O}_M$, respectively;
9: $\mathcal{P}_N \leftarrow$ Environmental selection from $\mathcal{P}_N \cup \mathcal{O}_N$;
10: Update the ideal point Z^*;
11: $\mathcal{P}_M \leftarrow$ Environmental selection from $\mathcal{P}_M \cup \mathcal{O}_M$ based on Algorithm 2;
12: Perform nondominated sorting on $\mathcal{P}_N \cup \mathcal{P}_M$;
13: Assign individuals for $\mathcal{P}_N$ and $\mathcal{P}_M$ respectively based on Algorithm 3;
14: **end while**
15: **return** Outputs.

3 Proposed method

3.1 Framework of the proposed DDFS algorithm

The framework of the proposed DDFS method is described in Algorithm 1. First, each of the two search approaches is assigned half of the initial population to better balance diversity and convergence (lines 1 and 2). The encoding takes the form of a binary to express whether each feature is selected. Second, parents are selected for each individual in the two subpopulations respectively (lines 4 and 5). Offspring is generated by the genetic algorithm (GA) operator and evaluated for both subpopulations (lines 6–8). Third, nondominated ranking and crowding distance are used to select individuals for the next generation in the dominance-based subpopulation (line 9). This ensures the diversity of solutions in the dominance-based subpopulation. For the decomposition-based subpopulation, an improved environmental selection strategy is used to assign each individual to an appropriate weight vector while avoiding duplicate solutions (line 11). Finally, solutions at different ranks in the merged population are divided into two subpopulations with different search methods to achieve information exchange (lines 12 and 13).

3.2 Improved environmental selection strategy

In the original MOEA/D, each solution gets assigned a unique weight vector. The neighborhood including similar subproblems is divided for each solution by determining the distance between the weight vectors. During the environmental selection phase, the original MOEA/D allows one excellent child to replace all parents in their neighborhood who are not as good as them. As shown in Figure 1, the parents P_2, P_3, and P_4 belong to the same neighborhood. C_2, C_3, and C_4 are duplicated offspring on the decision and objective spaces. When the child C_2 is compared with these three parents, C_2, C_3, and C_4 replace P_2, P_3, and P_4, respectively. This causes the solutions on W_2, W_3, and W_4 to be duplicates of C_2 in the population of next generation. Obviously, this would make some offspring heavily duplicate in the new generation population. These duplicate offspring not only waste evaluation time and computational resources, but also greatly reduce the population diversity in the next generation. Another problem is which weight vector the elite offspring should be placed on to compare with their corresponding parents. Finding more suitable weight vectors for the offspring to be placed can facilitate convergence and avoid losing valuable feature subsets. Therefore, we propose an improved environmental selection strategy to rearrange the offspring and avoid duplicate offspring.

The specific steps are shown in Algorithm 2. First, the angle values of the objective vector are calculated for all parent and offspring individuals (lines 1 and 4). They are sequentially arranged on the weight vectors in the order from smallest to largest (lines 2–4). In this way, the selection of the parent and child on the more suitable weight vector is more effective in helping the algorithm to converge. Then, the Tchebycheff method is used to calculate the values of the aggregation functions of the parents and children (lines 5–8). As shown in (3),

$$\begin{aligned} \min \quad & g(\boldsymbol{x} \mid \lambda, Z^*) = \max\{\lambda_i(f_i(\boldsymbol{x}) - Z^*)\} \\ \text{s.t.} \quad & \boldsymbol{x} \in \Omega, \end{aligned} \tag{3}$$

where g is the aggregation function's value, $\boldsymbol{x}$ is the decision vector, λ is the weight vector, and Z^* is the ideal point. After the previous arrangement, one parent and one child exist on each weight vector. Finally,

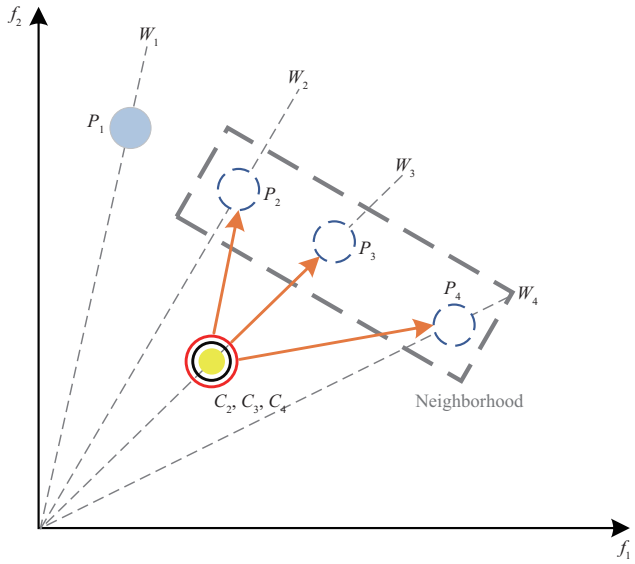

Figure 1 (Color online) Example for environmental selection of the original MOEA/D. C_2, C_3, and C_4 represent three duplicated offspring located on W_2, W_3, and W_4, respectively.

Algorithm 2 Improved environmental selection strategy

Input: $\mathcal{P}_M$ and $\mathcal{O}_M$: the parents and offspring of decomposition-based subpopulation.
$\mathcal{E}_{P_M}$ and $\mathcal{E}_{O_M}$: objective values of $\mathcal{P}_M$ and $\mathcal{O}_M$.
f_1 and f_2: selected feature ratio and classification error rate of each individual.
$N/2$: population size of $\mathcal{P}_M$.
Output: $\mathcal{P}_M$ Mobj: decision variables and objective values of decomposition-based subpopulation.
Calculate the angle value based on the objective vector of each individual in $\mathcal{P}_M$ according to $\arctan(f_2/f_1)$;
Rearrange the order of the parents according to the angle values from smallest to largest;
Arrange individuals with smaller angle values to weight vectors with larger preferences for the objective f_1;
Calculate the angle value of each individual in $\mathcal{O}_M$ and correspond to the weight vector in the same way as in $\mathcal{P}_M$;
Calculate the ideal point Z^* according to the objective values of all individuals;
RoadP $= \max(\mathcal{E}_{P_M}) - Z^*$;
RoadO $= \max(\mathcal{E}_{O_M}) - Z^*$;
Calculate the values of the aggregation function of the parent and child on each weight vector according to the Tchebycheff method;
for $i = 1 : N/2$ **do**
    Select an individual with the smaller aggregation function value from the parent and child;
end for
return Outputs.

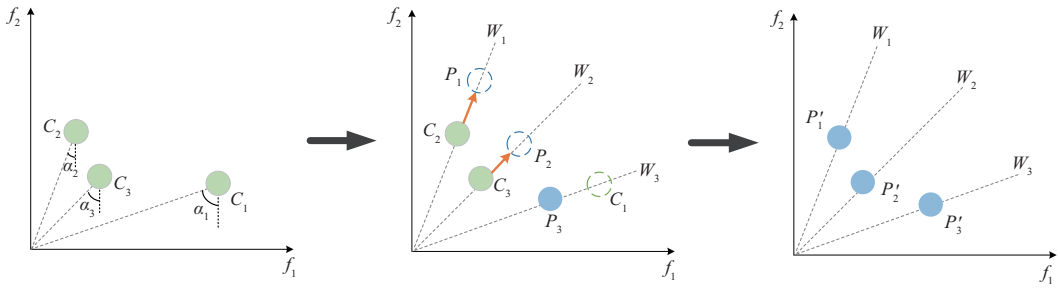

Figure 2 (Color online) Presentation of the improved environmental selection strategy.

only the one with the smaller g is chosen for the next generation. After the improved environmental selection strategy, there are not a large number of duplicate solutions in the population. The proposed algorithm is better to select the solution that is closer to the ideal point on each weight vector. As a result, the convergence speed and exploration capability of the algorithm are further enhanced.

Figure 2 illustrates an example of an improved environmental selection strategy. First, three offspring are rearranged as C_2, C_3, and C_1 according to $\alpha_2 < \alpha_3 < \alpha_1$ in the objective space. Then, C_2 and C_3 replace P_1 and P_2 on the weight vectors W_1 and W_2, respectively, according to the Tchebycheff method. P_3 is retained and not replaced by C_1. Finally, the three retained solutions P'_1, P'_2, and P'_3 are the parents of the next generation. Improved environmental selection strategy increases the diversity of the

Algorithm 3 Knowledge transfer mechanism

Input: $\mathcal{P}_N$ and $\mathcal{P}_M$: the individuals of dominance-based and decomposition-based subpopulations.
$\mathcal{E}_{P_N}$ and $\mathcal{E}_{P_M}$: objective values of $\mathcal{P}_N$ and $\mathcal{P}_M$.
Output: $\mathcal{P}_N$, $\mathcal{P}_M$, $\mathcal{E}_{P_N}$, and $\mathcal{E}_{P_M}$;
1: $\mathcal{P}_A \leftarrow [\mathcal{P}_N, \mathcal{P}_M]$;
2: $\mathcal{E}_{P_A} \leftarrow [\mathcal{E}_{P_N}, \mathcal{E}_{P_M}]$;
3: Perform a nondominant sorting operation on $\mathcal{E}_{P_A}$;
4: Index $\leftarrow$ Get the indexes of individuals on non-Pareto fronts;
5: $\mathcal{P}_M = \mathcal{P}_A(\text{Index}(1 : N/2))$;
6: $\mathcal{E}_{P_M} = \mathcal{E}_{P_A}(\text{Index}(1 : N/2))$;
7: $\mathcal{P}_N = \mathcal{P}_A \setminus \mathcal{P}_M$;
8: $\mathcal{E}_{P_N} = \mathcal{E}_{P_A} \setminus \mathcal{E}_{P_M}$;
9: **return** Outputs.

population to jump out of local optima.

3.3 Knowledge transfer mechanism

The knowledge transfer mechanism is the core part of the proposed framework, which achieves the sharing of experiences between the two search methods. To ensure the efficiency of knowledge transfer, two subpopulations are designed for dominance-based and decomposition-based, respectively, each of which owns half of all individuals.

Algorithm 3 demonstrates the specific process of knowledge transfer. First, two subpopulations are merged into one whole (lines 1 and 2). A nondominated sorting operation is performed on the whole population to obtain the front orders of all individuals (line 3). What to transfer: The dominance-based subpopulation has better diversity. And the decomposition-based subpopulation has the ability to converge quickly with ideal point guidance. Therefore, there are currently better convergent solutions on the obtained Pareto front, which are assigned to the dominance-based subpopulation. How to transfer: Second, starting from the second front, $N/2$ individuals are sequentially selected for assignment to the decomposition-based subpopulation (lines 4–6). This is because the diversity of feature subsets from other dominated fronts tends to be better than those on the Pareto front. And suboptimal feature subsets also have the potential to evolve into new optimal feature subsets. Finally, the remaining $N/2$ solutions including those on the Pareto front are assigned to the dominance-based subpopulation (lines 7 and 8). When to transfer: Both dominance-based and decomposition-based subpopulations are rearranged to maintain the balance of convergence and diversity in each generation.

According to the proposed knowledge transfer mechanism, the knowledge in the elitist solutions is transferred to the dominance-based subpopulation. In this way, the dominance-based subpopulation is allowed to generate more diverse suboptimal solutions. In addition, the previous suboptimal solutions are transferred to the decomposition-based subpopulation. Due to the neighborhood-based local search operation, suboptimal solutions may evolve into new nondominated solutions to help the proposed algorithm jump out of local optima.

3.4 Computational complexity

The computational cost of the proposed method DDFS mainly comes from the nondominated sorting operation and the crowding distance metric. Specifically, first, the computational complexity of the parent selection and reproduction operators is $O(N/2)$ in each subpopulation, respectively. N is the population size. Second, the improved environmental selection strategy has a lower computational complexity of $O(N/2)$ than the original one. Third, the computational complexity of the nondominated ranking in the dominance-based subpopulation and the merged population is $O(M \cdot (N/2)^2)$ and $O(M \cdot N^2)$, respectively. Finally, the computational complexity of the crowding distance metric in the dominance-based subpopulation is $O(M \cdot (N_F)^2)$, where N_F is the number of nondominated solutions and M is the number of objectives. Thus, the overall complexity of the proposed method in each generation is $2 \cdot 2 \cdot O(N/2) + O(M \cdot (N/2)^2) + O(M \cdot N^2) + O(M \cdot (N_F)^2) = O(M \cdot N^2)$. In addition, the computation time of DDFS is shorter than NSGA-II and MOEA/D because it is able to discover feature subsets with fewer features. This is verified in Subsection 5.4.

Table 1 Basic information of the 20 datasets

No.	Dataset	#Features	#Classes	Smallest class (%)	Largest class (%)	#Instances	Instances/features
1	Wine	13	3	26.9	39.9	178	13.69
2	Australian	14	2	32.2	67.8	690	49.29
3	Zoo	16	7	4.0	40.6	101	6.31
4	Climate	18	2	8.5	91.5	540	30.00
5	Hepatitis	19	2	20.7	79.3	155	8.16
6	SPECTF Heart	22	2	20.6	79.4	267	12.14
7	Flags	28	7	2.0	31.0	194	6.93
8	Ionosphere	34	2	35.9	64.1	351	10.32
9	Sonar	60	2	46.6	53.4	208	3.47
10	Hill Valley	100	2	49.3	50.7	606	6.06
11	Musk-1	166	2	43.5	56.5	476	2.85
12	Arrhythmia	279	13	0.4	54.0	452	1.62
13	LSVT	310	2	30.0	70.0	126	0.41
14	Madelon	500	2	50.0	50.0	2600	5.20
15	Isolet5	617	26	3.7	3.8	1559	2.53
16	SRBCT	2308	4	13.2	34.9	83	0.04
17	DLBCL	5469	2	24.7	75.3	77	0.01
18	Brain1	5920	5	4.4	66.6	90	0.02
19	Carcinom	9182	11	3.4	15.5	174	0.02
20	Leukemia2	11225	3	27.7	38.8	72	0.01

4 Experimental setups

4.1 Classification datasets

The 20 datasets used in this experiment are all derived from the real world and downloaded from https://archive.ics.uci.edu/ml/datasets.php. First, the number of features ranges from 13 (i.e., Wine) to 11225 (i.e., Leukemia2) as shown in Table 1. Second, 9 datasets are multiclassification problems with categories greater than 2. Moreover, they come from different domains such as diseases (Hepatitis, Arrhythmia, Brain1, and Leukemia2), physics (Sonar and Ionosphere), and speech recognition (Isolet5). In addition, in-depth analyses are performed for all datasets by calculating their class imbalance ratios and instance-to-feature ratios. It can be found that most datasets have unbalanced categories, which can cause a huge challenge for classification. Finally, instance-to-feature ratios in these datasets vary from 0.01 to 49.29, which provides a challenge to the classification time of the algorithm. Therefore, these datasets are sufficient to test the effectiveness of all compared algorithms in terms of classification accuracy and time consumed.

4.2 Comparison algorithms

To observe the effectiveness of our proposed algorithm, nine advanced algorithms are selected for comparison experiments. They are NSGA-II [19], MOEA/D [23], SPEA-2 [33], DrEA [24], MOEA/DD [25], PMMOEA [38], DAEA [16], SM-MOEA [34], and MFFS [32]. Among them, NSGA-II is the classical algorithm for updating populations based on dominance relationships. MOEA/D is a classical algorithm based on decomposition. Each individual represents a subproblem. Multiple neighborhoods of similar subproblems are used to coevolve to the Pareto front. SPEA-2 is used to calculate the fitness value of each individual by their dominance relationship. Nondominated solutions are present in the archive to guide the evolution. It is worth noting that PMMOEA, DAEA, SM-MOEA, and MFFS have been newly proposed for feature selection in the last three years. PMMOEA is specifically designed to solve sparse optimization by limiting the dimensionality of the offspring and maintaining sparsity. SM-MOEA uses the feature association matrix to evolve better feature subsets. These algorithms are selected as comparison algorithms to assure the fairness of the experiments since they have solid theoretical foundations and exhibit good performance in feature selection. Bold numbers in the tables of experimental results in Section 5 indicate that the algorithm corresponding to them produced the best results.

Table 2 Mean IGD performance on test data[a)]

Dataset	NSGA-II	MOEA/D	SPEA-2	DrEA	MOEA/DD
Wine	7.26E−02±3.29E−04(+)	7.78E−02±1.26E−02(+)	4.61E−02±1.49E−03(+)	3.50E−02±1.31E−02(+)	1.76E−02±1.32E−02(=)
Australian	4.42E−02±0.00e+00(+)	8.02E−02±0.00e+00(+)	**3.14E−02±7.31E−18(−)**	1.05E−01±4.05E−02(+)	4.97E−02±1.40E−02(+)
Zoo	7.23E−02±1.89E−02(+)	8.75E−02±1.50E−02(+)	4.52E−02±3.18E−03(+)	7.16E−02±1.69E−02(+)	**1.68E−02±1.22E−02(−)**
Climate	4.38E−02±1.33E−02(+)	1.05E−01±3.83E−02(+)	3.10E−02±9.44E−03(=)	8.00E−02±7.61E−03(+)	7.15E−02±5.38E−03(+)
Hepatitis	2.32E−01±3.46E−02(+)	1.80E−01±4.46E−02(+)	1.16E−01±3.89E−02(+)	4.85E−02±3.20E−02(=)	**1.55E−02±1.65E−02(−)**
SPECTF Heart	8.63E−02±1.33E−02(+)	9.83E−02±1.72E−02(+)	9.42E−02±7.74E−03(+)	7.29E−02±1.91E−02(+)	4.43E−02±1.17E−02(−)
Flags	5.72E−02±6.23E−03(+)	9.84E−02±1.46E−17(+)	6.21E−02±5.80E−03(+)	7.15E−02±3.93E−02(+)	5.69E−02±1.12E−02(+)
Ionosphere	2.98E−02±4.34E−03(+)	3.21E−02±7.08E−03(+)	1.79E−02±4.30E−03(+)	5.53E−02±1.94E−02(+)	1.22E−02±3.45E−03(+)
Sonar	9.43E−02±1.28E−02(+)	7.99E−02±1.00E−02(+)	3.42E−01±2.57E−03(+)	5.93E−02±1.75E−02(=)	6.34E−02±8.72E−03(+)
Hill Valley	3.11E−02±1.13E−02(+)	5.58E−02±1.99E−02(+)	3.03E−02±1.13E−02(+)	3.28E−02±5.32E−03(+)	2.29E−02±6.21E−03(+)
Musk-1	5.19E−02±1.20E−03(+)	8.31E−02±3.06E−02(+)	7.75E−02±1.38E−02(=)	8.53E−02±1.63E−02(+)	3.91E−02±1.27E−02(−)
Arrhythmia	8.94E−02±1.10E−02(+)	1.08E−01±8.87E−03(+)	3.17E−02±5.64E−03(+)	6.71E−02±2.31E−02(+)	1.19E−01±3.46E−02(+)
LSVT	1.86E−01±1.06E−02(+)	2.01E−01±8.27E−03(+)	1.12E−01±1.06E−02(+)	9.60E−02±3.25E−03(+)	7.35E−02±6.37E−03(+)
Madelon	1.48E−01±7.69E−03(+)	1.68E−01±7.92E−03(+)	7.13E−02±7.77E−03(+)	4.86E−02±2.46E−02(+)	8.40E−02±1.50E−02(+)
Isolet5	1.83E−01±6.27E−03(+)	2.02E−01±1.01E−02(+)	1.28E−01±8.43E−03(+)	4.14E−02±6.43E−03(+)	7.04E−02±1.37E−02(+)
SRBCT	1.88E−01±1.17E−02(+)	2.02E−01±8.35E−03(+)	1.64E−01±1.09E−02(+)	6.14E−02±1.64E−02(+)	2.52E−01±1.93E−02(+)
DLBCL	1.91E−01±1.54E−03(+)	2.00E−01±2.50E−03(+)	5.44E−02±1.30E−03(+)	4.71E−02±3.01E−02(+)	3.33E−01±1.64E−02(+)
Brain1	2.18E−01±8.43E−03(+)	2.11E−01±2.93E−17(+)	3.34E−01±8.05E−03(+)	1.37E−01±2.75E−02(+)	3.50E−01±9.91E−03(+)
Carcinom	2.49E−01±3.38E−03(+)	2.44E−01±0.00e+00(+)	4.00E−01±4.55E−03(+)	8.96E−02±9.27E−03(+)	4.00E−01±7.69E−03(+)
Leukemia2	2.28E−01±2.81E−02(+)	1.98E−01±0.00e+00(+)	3.59E−01±3.07E−02(+)	9.69E−02±6.46E−02(+)	3.97E−01±2.41E−02(+)
+/=/−	20/0/0	20/0/0	17/2/1	18/2/0	15/1/4
Friedman's rank	7.1	8.8	6.225	5.875	5.575

Dataset	PMMOEA	DAEA	SM-MOEA	MFFS	DDFS
Wine	2.37E−02±3.41E−02(+)	3.76E−02±2.03E−02(+)	3.94E−02±1.52E−02(+)	3.89E−02±1.65E−02(+)	**1.02E−02±1.31E−4**
Australian	4.43E−02±0.00e+00(+)	6.82E−02±1.49E−02(+)	6.13E−02±7.10E−03(+)	4.08E−02±1.46E−17(+)	3.49E−02±3.03E−03
Zoo	4.88E−02±6.49E−03(+)	7.75E−02±8.57E−03(+)	9.36E−02±1.50E−02(+)	5.89E−02±1.19E−02(+)	1.89E−02±1.19E−02
Climate	3.39E−02±2.24E−03(+)	5.78E−02±8.75E−03(+)	5.00E−02±1.61E−02(+)	8.09E−02±2.01E−04(+)	**2.76E−02±8.57E−03**
Hepatitis	2.34E−02±2.26E−03(=)	1.24E−01±3.30E−02(+)	1.67E−01±3.60E−03(+)	4.66E−02±4.67E−02(=)	4.79E−02±4.21E−02
SPECTF Heart	**4.21E−02±9.99E−04(−)**	7.43E−02±7.85E−03(+)	1.06E−01±1.95E−02(+)	7.60E−02±2.93E−03(+)	6.34E−02±1.08E−02
Flags	5.02E−02±3.96E−03(+)	9.18E−02±9.19E−03(+)	8.95E−02±6.00E−03(+)	6.18E−02±1.89E−02(+)	**4.31E−02±1.45E−02**
Ionosphere	1.44E−02±9.77E−03(+)	4.58E−02±1.14E−02(+)	3.51E−02±7.39E−03(+)	3.15E−02±8.42E−03(+)	**1.02E−02±3.68E−03**
Sonar	**5.58E−02±2.40E−03(−)**	7.87E−02±5.77E−03(+)	7.31E−02±1.55E−02(+)	6.15E−02±1.01E−02(+)	5.93E−02±1.33E−02
Hill Valley	2.31E−02±2.30E−03(+)	3.71E−02±1.69E−02(+)	3.15E−02±2.45E−02(+)	2.19E−02±1.20E−03(+)	**1.12E−02±7.05E−03**
Musk-1	7.77E−02±1.00E−02(+)	**3.37E−02±1.42E−02(−)**	9.37E−02±1.27E−02(+)	5.06E−02±7.20E−03(+)	7.05E−02±7.27E−03
Arrhythmia	2.09E−02±8.17E−03(=)	1.02E−01±1.02E−02(+)	4.28E−02±1.36E−02(+)	5.09E−02±1.43E−02(+)	**1.89E−02±1.43E−02**
LSVT	5.58E−02±1.99E−02(+)	3.75E−02±4.86E−02(+)	9.04E−02±3.38E−03(+)	4.01E−02±2.88E−02(+)	**1.58E−02±8.67E−03**
Madelon	2.49E−02±1.92E−03(=)	1.20E−01±9.24E−03(+)	5.03E−02±1.16E−03(+)	2.55E−02±1.26E−03(=)	**2.08E−02±1.54E−03**
Isolet5	1.83E−02±2.65E−03(=)	9.85E−02±7.19E−03(+)	4.34E−02±6.66E−03(+)	2.19E−02±1.67E−02(+)	**1.80E−02±3.80E−03**
SRBCT	5.63E−02±1.29E−02(+)	7.71E−02±1.65E−02(+)	9.39E−02±4.67E−02(+)	8.01E−02±1.37E−02(+)	**2.84E−02±3.05E−02**
DLBCL	1.58E−02±7.40E−03(−)	7.44E−02±1.27E−02(+)	3.66E−02±1.62E−02(=)	**1.53E−02±4.64E−03(−)**	3.40E−02±2.27E−02
Brain1	5.74E−02±2.51E−03(+)	8.75E−02±1.46E−17(+)	8.17E−02±4.47E−02(+)	2.94E−01±4.94E−03(+)	**1.73E−02±7.86E−03**
Carcinom	7.76E−02±2.34E−03(+)	**4.98E−02±7.27E−03(=)**	7.43E−02±2.37E−02(+)	7.17E−02±9.27E−03(+)	5.08E−02±1.47E−02
Leukemia2	6.84E−02±3.67E−03(+)	1.39E−01±4.02E−02(+)	1.13E−01±7.15E−02(+)	8.56E−02±3.14E−02(+)	**3.84E−02±2.38E−02**
+/=/−	13/4/3	18/1/1	19/1/0	17/2/1	–
Friedman's rank	2.9	6.05	6.4	4.3	**1.775**

a) The bold numbers indicate that the corresponding algorithm achieves the best performance on corresponding dataset.

4.3 Performance metrics

In the experimental analysis phase, we computed hypervolume (HV) [39] and inverted generation distance (IGD) [40] to measure the diversity and convergence of all compared algorithms. In calculating the HV metric, the classification error rate and the ratio of selected features to the total number of features are first normalized to the scale $[0, 1]$. Then, the reference point of the HV metric is set to $[1, 1]$ [16]. The larger the HV value calculated in this way, the better the performance of the algorithm. When calculating the IGD metrics, the Pareto front on each dataset is unknown. Because feature selection is a discrete problem and the search space is huge, it is almost impossible to find the true Pareto front. Therefore, to try to ensure the fairness of the experiments, we first combine the final populations found by all compared algorithms in 30 experiments into one joint population. Then the first front obtained after nondominant sorting of the joint population is used as the reference front for calculating the IGD metric. In general, the smaller the IGD value, the better the performance of the algorithm.

In addition, the Wilcoxon test with a significance level of 0.05 is used to test whether the performance of DDFS is significantly different from that of other competing algorithms. And, "+", "−", and "=" represent that DDFS is significantly better, worse, or not significantly different from the compared algorithms, respectively. For the computation time of all algorithms, the Friedman test yields the relative ranking of all algorithms.

4.4 Parameter settings

In the comparison experiments, first, all the algorithms are from the source code provided by the authors [41]. Second, each algorithm is run 30 times on each dataset, where the random seed is set to

Table 3 Mean HV performance on test data[a)]

Dataset	NSGA-II	MOEA/D	SPEA-2	DrEA	MOEA/DD
Wine	8.24E−01±5.40E−03(+)	8.09E−01±1.34E−02(+)	8.31E−01±4.41E−03(+)	8.68E−01±1.00E−02(+)	8.79E−01±7.08E−03(=)
Australian	6.22E−01±1.13E−16(+)	5.95E−01±3.60E−02(+)	6.22E−01±1.38E−03(+)	5.93E−01±9.28E−03(+)	6.36E−01±1.14E−02(+)
Zoo	7.98E−01±1.94E−02(+)	7.53E−01±1.47E−02(+)	5.50E−01±3.07E−02(+)	8.06E−01±1.61E−02(+)	**8.55E−01±2.24E−03(−)**
Climate	8.99E−01±7.92E−03(+)	8.69E−01±1.53E−02(+)	9.04E−01±4.49E−03(+)	9.01E−01±4.42E−03(+)	9.12E−01±4.00E−03(=)
Hepatitis	7.39E−01±2.17E−02(+)	7.40E−01±2.87E−02(+)	7.62E−01±1.29E−02(+)	8.30E−01±5.49E−03(=)	**8.45E−01±1.39E−02(−)**
SPECTF Heart	7.02E−01±9.06E−03(+)	7.05E−01±9.47E−03(+)	7.04E−01±4.88E−03(+)	7.01E−01±2.58E−02(+)	7.67E−01±9.49E−03(−)
Flags	6.15E−01±1.36E−02(+)	5.65E−01±0.00e+00(+)	6.21E−01±1.50E−02(+)	6.07E−01±3.50E−02(+)	6.53E−01±1.07E−02(+)
Ionosphere	9.13E−01±9.66E−03(+)	9.11E−01±1.24E−02(+)	9.23E−01±3.06E−03(+)	9.12E−01±1.41E−02(+)	9.32E−01±6.07E−03(=)
Sonar	8.15E−01±2.09E−02(+)	8.41E−01±1.02E−02(=)	5.50E−01±3.07E−02(+)	8.59E−01±2.51E−02(+)	8.61E−01±1.65E−02(=)
Hill Valley	5.48E−01±1.13E−02(+)	5.10E−01±2.25E−02(+)	5.75E−01±1.54E−02(+)	5.72E−01±1.22E−02(+)	5.84E−01±1.38E−02(+)
Musk-1	8.28E−01±1.93E−02(−)	7.83E−01±2.92E−02(+)	8.16E−01±2.11E−02(−)	7.82E−01±3.02E−02(+)	8.35E−01±2.12E−02(−)
Arrhythmia	6.06E−01±1.08E−02(+)	5.70E−01±1.87E−02(+)	6.60E−01±9.81E−03(+)	6.28E−01±2.53E−02(+)	5.97E−01±3.04E−02(+)
LSVT	8.73E−01±4.65E−02(+)	8.21E−01±6.54E−02(+)	8.96E−01±3.11E−02(+)	9.03E−01±2.50E−02(+)	9.18E−01±2.11E−02(+)
Madelon	6.07E−01±1.44E−02(+)	5.82E−01±2.32E−02(+)	8.85E−01±1.01E−02(+)	7.69E−01±5.83E−02(+)	7.07E−01±2.07E−02(+)
Isolet5	6.90E−01±1.03E−02(+)	6.62E−01±1.03E−02(+)	7.83E−01±1.05E−02(+)	7.73E−01±1.20E−02(+)	8.24E−01±1.32E−02(+)
SRBCT	8.32E−01±6.53E−03(+)	8.12E−01±2.01E−02(+)	8.79E−01±1.06E−02(+)	9.66E−01±2.10E−02(=)	8.01E−01±9.75E−03(+)
DLBCL	7.91E−01±1.36E−02(+)	7.62E−01±1.80E−02(+)	6.60E−01±1.46E−02(+)	9.63E−01±3.74E−02(=)	6.31E−01±2.26E−02(+)
Brain1	6.04E−01±1.24E−02(+)	6.11E−01±4.52E−16(+)	5.02E−01±1.20E−02(+)	8.26E−01±2.68E−02(+)	5.34E−01±2.14E−02(+)
Carcinom	6.81E−01±1.12E−02(+)	6.84E−01±1.13E−16(+)	5.25E−01±1.53E−02(+)	8.43E−01±1.65E−02(+)	5.48E−01±1.12E−02(+)
Leukemia2	6.78E−01±6.36E−02(+)	7.39E−01±3.39E−16(+)	5.67E−01±6.27E−02(+)	8.93E−01±8.23E−02(+)	4.44E−01±4.29E−02(+)
+/=/−	20/0/0	19/1/0	20/0/0	17/3/0	11/4/5
Friedman's rank	7.425	8.65	6.825	5.5	4.75

Dataset	PMMOEA	DAEA	SM-MOEA	MFFS	DDFS
Wine	**8.81E−01±5.00E−03(=)**	8.52E−01±6.33E−03(+)	8.44E−01±3.55E−03(+)	8.45E−01±5.63E−05(+)	8.79E−01±2.34E−16
Australian	6.37E−01±2.81E−04(+)	6.18E−01±8.20E−03(+)	6.13E−01±1.32E−02(+)	6.24E−01±3.55E−03(+)	**6.40E−01±7.15E−03**
Zoo	8.14E−01±1.09E−02(+)	7.69E−01±8.36E−03(+)	7.37E−01±1.65E−02(+)	7.97E−01±9.82E−03(+)	8.44E−01±4.23E−03
Climate	**9.12E−01±2.74E−03(=)**	8.86E−01±4.52E−03(+)	8.95E−01±4.61E−03(+)	8.72E−01±9.31E−05(+)	9.10E−01±4.23E−03
Hepatitis	8.40E−01±1.62E−02(−)	7.86E−01±1.53E−02(+)	7.82E−01±9.53E−03(+)	8.29E−01±2.26E−16(=)	8.27E−01±3.97E−03
SPECTF Heart	**7.70E−01±1.74E−02(−)**	7.28E−01±9.45E−03(+)	6.74E−01±1.29E−02(+)	7.10E−01±8.96E−03(+)	7.37E−01±7.42E−03
Flags	6.44E−01±9.44E−03(+)	5.69E−01±1.83E−02(+)	5.89E−01±1.65E−02(+)	6.48E−01±3.62E−03(+)	**6.61E−01±3.87E−02**
Ionosphere	9.32E−01±6.83E−03(=)	9.02E−01±1.10E−02(+)	9.08E−01±7.24E−03(+)	9.19E−01±1.05E−04(+)	**9.32E−01±5.90E−03**
Sonar	**8.81E−01±2.10E−02(−)**	8.54E−01±2.63E−02(=)	8.38E−01±1.91E−02(+)	8.45E−01±1.86E−02(+)	8.63E−01±2.31E−02
Hill Valley	5.83E−01±8.98E−03(+)	5.33E−01±1.29E−02(+)	5.54E−01±8.31E−03(+)	5.91E−01±1.67E−02(+)	**5.97E−01±1.56E−02**
Musk-1	8.17E−01±2.43E−02(−)	**8.37E−01±2.07E−02(−)**	7.67E−01±1.91E−02(+)	8.26E−01±2.49E−02(−)	7.97E−01±1.35E−02
Arrhythmia	6.80E−01±1.47E−02(=)	6.29E−01±7.87E−03(+)	6.35E−01±1.83E−02(+)	6.22E−01±7.35E−03(+)	**6.82E−01±2.51E−02**
LSVT	9.22E−01±3.61E−02(+)	9.53E−01±3.02E−02(+)	9.04E−01±2.53E−02(+)	9.32E−01±4.01E−02(+)	**9.84E−01±1.77E−02**
Madelon	8.97E−01±4.43E−03(=)	5.93E−01±8.29E−03(+)	7.55E−01±8.06E−02(+)	8.97E−01±2.89E−03(=)	**8.98E−01±3.03E−03**
Isolet5	8.29E−01±1.23E−02(=)	7.32E−01±4.86E−03(+)	7.54E−01±1.21E−02(+)	8.18E−01±1.47E−02(+)	**8.31E−01±1.15E−02**
SRBCT	9.42E−01±4.49E−02(+)	**9.72E−01±5.71E−03(−)**	8.50E−01±6.03E−02(+)	9.33E−01±3.69E−02(+)	9.65E−01±5.95E−02
DLBCL	9.89E−01±2.88E−02(−)	9.81E−01±3.43E−03(−)	9.53E−01±3.86E−02(+)	**9.91E−01±2.25E−02(−)**	9.65E−01±5.21E−02
Brain1	8.63E−01±4.05E−02(+)	7.49E−01±5.84E−03(+)	7.52E−01±4.29E−02(+)	8.27E−01±3.80E−02(+)	**8.70E−01±2.69E−02**
Carcinom	8.62E−01±4.02E−02(+)	8.00E−01±1.24E−02(+)	7.97E−01±4.78E−02(+)	8.19E−01±3.67E−02(+)	**8.81E−01±2.31E−02**
Leukemia2	9.33E−01±4.50E−02(+)	8.51E−01±5.48E−02(+)	8.58E−01±7.13E−02(+)	9.18E−01±9.55E−02(+)	**9.44E−01±4.22E−02**
+/=/−	9/6/5	16/1/3	20/0/0	16/2/2	–
Friedman's rank	2.45	5.8	7.1	4.325	**2.175**

a) The bold numbers indicate that the corresponding algorithm achieves the best performance on corresponding dataset.

reproduce each performance. Third, the customizable parameters of each algorithm are consistent with the descriptions in their papers. This is to ensure the fairness of the experiments. Notably, the population size of each comparison algorithm is set to 200 to ensure diversity and efficiency. The maximal iteration is set to 100.

For the division of each dataset, 80% of the original dataset is taken as the training set, while the remaining 20% is the test set [16,42]. Considering the unbalanced proportion of categories in the dataset, each category is divided according to the above proportion. It is worth noting that this division is constant for each dataset for 30 experiments, but the division is different for different datasets. This is to ensure that the results are statistically significant for 30 times. The KNN classifier with 5-fold cross-validation is employed to avoid inaccurate evaluation of feature subsets due to bias in the training data. And the value of k is set to 5 to balance accuracy and efficiency [43].

5 Results and discussions

5.1 IGD and HV performances

Tables 2 and 3 present the IGD and HV metrics obtained by the Pareto front evaluated on the test data. Notably, feature subsets found in the training data may suffer from overfitting. The test data may dominate some nondominated feature subsets. Therefore, showing the computed metrics on the test set demonstrates the generalization performance of algorithms in feature selection is important.

Table 2 shows that the top-ranked DDFS outperforms NSGA-II and SM-MOEA on all datasets in terms of IGD. Although DDFS does not achieve the best IGD value on eight datasets, it was considerably better

Table 4 Mean HV performance on test data in validation experiments of improved environmental selection strategy[a)]

Dataset	DDFS-O	DDFS-B	DDFS
Wine	8.79E−01±2.34E−16(=)	**8.81E−01±3.01E−03(−)**	8.79E−01±2.34E−16
Australian	**6.43E−01±1.05E−02(−)**	6.36E−01±8.41E−03(+)	6.41E−01±5.04E−03
Zoo	8.38E−01±1.59E−02(+)	8.24E−01±8.09E−03(+)	**8.44E−01±4.23E−03**
Climate	9.09E−01±5.03E−03(=)	9.04E−01±3.96E−03(+)	**9.10E−01±4.23E−03**
Hepatitis	**8.28E−01±1.42E−02(−)**	8.19E−01±1.22E−02(+)	8.27E−01±3.25E−03(+)
SPECTF Heart	7.31E−01±1.01E−02(+)	7.32E−01±1.71E−02(+)	**7.37E−01±7.42E−03**
Flags	6.52E−01±2.89E−02(+)	6.38E−01±1.42E−02(+)	**6.61E−01±3.87E−02**
Ionosphere	9.27E−01±8.37E−03(+)	9.19E−01±1.23E−02(+)	**9.32E−01±5.90E−03**
Sonar	8.68E−01±1.42E−02(+)	8.62E−01±2.92E−02(+)	**8.70E−01±2.17E−02**
Hill Valley	5.80E−01±1.62E−02(+)	5.01E−01±6.22E−02(+)	**5.97E−01±1.56E−02**
Musk-1	7.80E−01±1.53E−02(+)	7.44E−01±3.07E−02(+)	**7.97E−01±1.35E−02**
Arrhythmia	6.74E−01±1.87E−02(+)	**6.85E−01±1.69E−02(−)**	6.82E−01±2.51E−02
LSVT	9.42E−01±2.64E−02(+)	9.22E−01±3.51E−02(+)	**9.84E−01±1.77E−02**
Madelon	8.94E−01±6.63E−03(=)	8.50E−01±8.00E−02(+)	**8.98E−01±3.03E−03**
Isolet5	8.00E−01±1.45E−02(+)	7.98E−01±2.32E−02(+)	**8.31E−01±1.15E−02**
SRBCT	9.49E−01±4.84E−02(+)	9.69E−01±3.85E−02(=)	**9.69E−01±3.80E−02**
DLBCL	9.09E−01±7.23E−02(+)	**9.96E−01±3.74E−03(−)**	9.65E−01±5.21E−02
Brain1	8.57E−01±4.33E−02(+)	8.42E−01±4.97E−02(+)	**8.70E−01±2.69E−02**
Carcinom	8.59E−01±2.66E−02(+)	8.61E−01±1.37E−02(+)	**8.81E−01±2.31E−02**
Leukemia2	8.91E−01±8.95E−02(+)	8.64E−01±6.19E−02(+)	**9.52E−01±5.09E−02**
+/=/−	15/3/2	16/1/3	–
Friedman's rank	2.1875	2.4062	**1.4062**

a) The bold numbers indicate that the corresponding algorithm achieves the best performance on corresponding dataset.

than the other nine compared algorithms on most datasets. It indicates that the Pareto front found by DDFS has the best convergence even on high-dimensional datasets with small samples (e.g., Brain1 and Leukemia2). This is because the proposed knowledge transfer mechanism uses the potential solutions of the dominance-based search approach to help DDFS escape local optima and find feature subsets with higher classification accuracy. Table 3 shows that the proposed DDFS beats the other nine advanced algorithms on most datasets in terms of HV. This proves that the proposed DDFS algorithm better balances convergence and diversity. Specifically, first, DDFS significantly outperforms NSGA-II, MOEA/D, SPEA-2, DrEA, and SM-MOEA on almost all datasets. Second, it can be found that DDFS does not perform as well as MOEA/DD on five low-dimensional and medium-dimensional datasets due to a two-layer weight vector generation method in MOEA/DD that can divide diverse subregions. Moreover, they can enhance population diversity by providing diverse search directions and density estimation. However, our proposed DDFS performs significantly better than MOEA/DD on high-dimensional datasets. The proposed improved environmental selection strategy can substantially filter redundant features and find feature subsets with a few features. Finally, although DDFS does not perform as well as DAEA and MFFS on four datasets, it still significantly outperforms them on 16 datasets. These results proved that our proposed algorithm could find feature subsets with high classification accuracy and few features even on class-imbalanced datasets (e.g., Flags, Arrhythmia, and Brain1).

5.2 Major component contribution analysis

Two validation experiments were organized to observe the effectiveness of the two main components proposed in this paper. Four variants of the algorithm were run 30 times independently on all data sets. Friedman tests were performed on their HV metrics.

Table 4 presents that the effectiveness of the improved environmental selection strategy was verified by comparing it with two other environmental selection strategies. Specifically, DDFS-O is the replacement of the proposed improved environmental selection strategy with a neighborhood-based environmental selection operator [22], and DDFS-B is the replacement of the proposed improved environmental selection strategy with a best-fit-based environmental selection operator [44]. Notably, the DDFS-O and DDFS-B are much worse than DDFS on most datasets because the proposed improved environmental selection strategy significantly improves the ability of the algorithm to explore the search space and thus find a better subset of features. Table 5 depicts the results of experiments to validate the effectiveness of the

Table 5 Mean HV performance on test data in validation experiments of knowledge transfer mechanism[a)]

Dataset	DDFS-I	DDFS-E	DDFS
Wine	8.79E−01±2.34E−16(=)	8.79E−01±2.34E−16(=)	**8.79E−01±2.34E−16**
Australian	6.40E−01±1.93E−03(=)	6.40E−01±7.20E−03(=)	**6.40E−01±7.15E−03**
Zoo	8.41E−01±8.98E−04(+)	8.26E−01±9.68E−03(+)	**8.44E−01±4.23E−03**
Climate	9.03E−01±1.79E−03(+)	9.04E−01±2.84E−03(+)	**9.10E−01±4.23E−03**
Hepatitis	**8.36E−01±1.09E−02(−)**	8.27E−01±3.83E−03(=)	8.27E−01±3.25E−03
SPECTF Heart	**7.64E−01±2.92E−03(−)**	7.53E−01±7.59E−03(−)	7.37E−01±7.42E−03
Flags	6.43E−01±2.97E−02(+)	6.35E−01±1.60E−02(+)	**6.61E−01±3.87E−02**
Ionosphere	9.20E−01±6.63E−03(+)	9.20E−01±9.25E−03(+)	**9.32E−01±5.90E−03**
Sonar	8.57E−01±1.22E−02(+)	**8.76E−01±2.47E−02(−)**	8.63E−01±2.31E−02
Hill Valley	5.96E−01±6.22E−02(=)	5.89E−01±1.59E−02(+)	**5.97E−01±1.56E−02**
Musk-1	**8.11E−01±1.70E−02(−)**	7.96E−01±2.14E−02(=)	7.97E−01±1.35E−02
Arrhythmia	**6.85E−01±1.67E−02(−)**	6.74E−01±2.27E−02(+)	6.82E−01±2.51E−02
LSVT	9.75E−01±2.18E−02(+)	9.52E−01±3.27E−02(+)	**9.84E−01±1.77E−02**
Madelon	**8.99E−01±3.21E−03(=)**	8.92E−01±6.13E−03(=)	8.98E−01±3.03E−03
Isolet5	8.08E−01±9.23E−03(+)	7.93E−01±1.03E−02(+)	**8.31E−01±1.15E−02**
SRBCT	**9.79E−01±9.37E−03(−)**	9.51E−01±7.14E−02(+)	9.65E−01±5.95E−02
DLBCL	9.67E−01±3.88E−02(=)	**9.78E−01±4.13E−02(−)**	9.65E−01±5.21E−02
Brain1	8.62E−01±3.67E−02(+)	**8.72E−01±2.67E−02(=)**	8.70E−01±2.69E−02
Carcinom	8.54E−01±2.04E−02(+)	8.61E−01±1.20E−02(+)	**8.81E−01±2.31E−02**
Leukemia2	9.38E−01±4.89E−02(+)	**9.83E−01±2.72E−02(−)**	9.52E−01±5.09E−02
+/=/−	10/5/5	10/6/4	–
Friedman's rank	2.125	2.1562	**1.7188**

a) The bold numbers indicate that the corresponding algorithm achieves the best performance on corresponding dataset.

knowledge transfer mechanism. Here, DDFS-I denotes replacing the proposed knowledge transfer mechanism with an implicit knowledge transfer [30], and DDFS-E denotes replacing the proposed knowledge transfer mechanism with an explicit knowledge transfer [27–29]. The results of the Friedman test for HV metrics proved that our proposed knowledge transfer mechanism is effective. It can be seen that DDFS significantly outperforms DDFS-I and DDFS-E on ten datasets, especially on the low-dimensional datasets. Therefore, we conclude that our two proposed new strategies have better results on the feature selection problem.

5.3 Nondominated front distribution analysis

Figure 3 illustrates the distribution of the nondominated feature subsets found by each algorithm on the test data of nine representative datasets. The Pareto front with the median HV value among the 30 experiments is plotted, reflecting the overall effect of the algorithm. First, Figure 3 shows that the front found by DDFS has the best classification accuracy compared with other algorithms. Even on high-dimensional datasets (e.g., Leukemia2), feature subsets with a classification error rate of 0 can be found because the proposed knowledge transfer mechanism can achieve better convergence while maintaining diversity. Second, our algorithm solves the high-dimensional problem better compared to other algorithms. Specifically, DDFS can filter out redundant features and select only a few key features, thus achieving higher classification accuracy. Third, the Pareto front found by DDFS has the most diverse solutions compared to other algorithms. The reason is that the improved environmental selection strategy retains more potential solutions. The diversity of populations is effectively increased, thus evolving new nondominated solutions. Finally, DDFS finds the combination of features with the highest classification accuracy on most datasets, although the number of selected features is large. These solutions should be considered when the user wants high classification accuracy and does not care about the classification cost. Although the classification accuracy of the subset with the fewest features chosen is not high, it remains the most cost-effective solution when accuracy requirements can be met. Moreover, DDFS provides other diverse options if the user needs to trade off classification accuracy and cost.

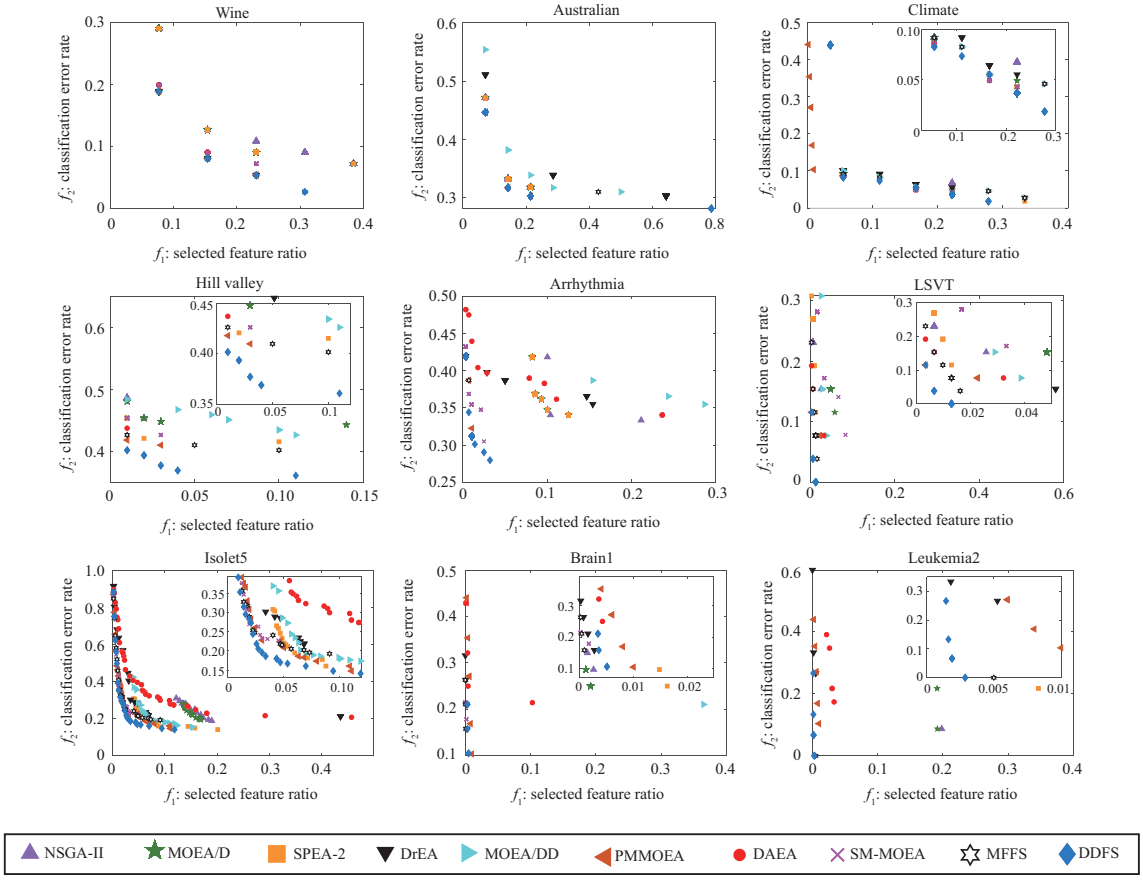

Figure 3 (Color online) Distributions of the median Pareto fronts obtained on 9 representative datasets by each algorithm in the objective space.

5.4 Computing time

To examine the performance of the proposed DDFS in terms of computation time, we compared the average computation time of all algorithms over 30 experiments using 20 datasets. In the experiments, the Friedman test was used to calculate the ranking of each algorithm, as depicted in Table 6. Generally, the number of samples and features in the dataset significantly affects the computation time of the feature selection algorithm. This is because the KNN classifier computes the distance over many samples to predict the labels, leading to an increase in time cost. Furthermore, the high dimensionality of the feature vector also leads to longer computation time. Specifically, Australia and Isolet5 take longer to compute than other datasets with small sample sizes, and Carcinom and Leukemia2 take longer to compute than other datasets with low feature dimensions. Overall, DDFS takes the shortest computation time using 13 datasets. Furthermore, DDFS is the best performer based on the Friedman test because the decomposition-based subpopulation can identify Pareto feature subsets with a few critical features shared by the proposed knowledge transfer mechanism to guide population evolution. This reduces the evaluation time of feature subsets in the training process.

6 Conclusion

This report aimed to develop an evolutionary multiobjective feature selection method that combined dominance-based and decomposition-based approaches. Inspired by the complementary advantages, the knowledge obtained by different search approaches was shared to select smaller feature subsets with higher classification accuracy.

First, the environmental selection strategy in the decomposition-based subpopulation based on neighborhood replacement was improved. The angle value of each individual in the objective space for the coordinate axis was calculated using its objective value, and the angle values were used to rearrange

Table 6 Average time consumed by 10 algorithms on 20 datasets (s)[a)]

Dataset	NSGA-II	MOEA/D	SPEA-2	DrEA	MOEA/DD	PMMOEA	DAEA	SM-MOEA	MFFS	DDFS
Wine	63.82	31.83	28.61	52.72	61.06	71.60	76.51	23.97	73.59	**23.45**
Australian	52.42	28.65	30.96	325.60	100.84	310.64	75.01	**26.42**	294.67	105.66
Zoo	56.26	26.37	27.70	30.63	52.54	117.92	64.72	23.72	109.63	**22.01**
Climate	54.89	29.10	30.07	212.48	125.83	199.88	85.41	**26.04**	193.43	75.42
Hepatitis	54.43	26.44	28.06	46.18	63.89	47.27	45.15	24.29	39.39	**20.50**
SPECTF Heart	53.33	36.42	28.73	70.04	87.30	125.42	43.79	**24.69**	84.33	30.21
Flags	52.69	27.70	28.29	52.11	81.46	92.39	49.68	25.49	77.60	**25.22**
Ionosphere	59.59	34.62	30.32	155.07	92.74	112.83	54.83	**24.97**	110.61	40.94
Sonar	49.61	42.53	29.17	92.88	67.78	76.94	55.06	24.79	73.30	**23.78**
Hill Valley	379.02	129.24	251.47	400.30	308.85	229.77	174.19	126.72	362.36	**99.60**
Musk-1	482.81	329.63	276.18	297.94	337.15	370.28	455.76	203.46	228.18	**198.41**
Arrhythmia	59.17	36.87	37.11	292.97	153.63	229.37	69.74	**29.78**	244.74	63.24
LSVT	79.35	67.42	55.66	51.86	83.60	86.65	63.72	99.23	**50.35**	65.69
Madelon	4238.03	3487.35	1977.46	2317.04	2306.96	2865.00	3215.75	2052.79	4160.67	**1806.51**
Isolet5	201.43	205.36	**200.83**	1049.99	642.72	919.81	453.14	788.01	1583.27	465.02
SRBCT	102.79	59.61	38.89	153.86	133.53	136.78	84.14	75.63	113.65	**30.68**
DLBCL	123.14	130.58	62.64	608.22	396.63	786.83	246.38	167.95	345.86	**60.15**
Brain1	233.45	162.80	81.64	707.81	553.32	527.73	220.43	244.12	730.15	**68.87**
Carcinom	411.78	252.16	186.46	3035.21	1718.34	2697.78	556.53	496.66	2390.59	**179.44**
Leukemia2	426.74	246.28	151.22	1095.36	913.72	1255.46	523.89	616.08	2135.37	**102.79**
Friedman's rank	5.95	3.60	2.80	7.85	7.20	8.50	5.80	3.25	7.65	**2.40**

a) The bold numbers indicate that the corresponding algorithm achieves the best performance on corresponding dataset.

each individual to a more appropriate weight vector, avoiding many duplicate solutions and enhancing the diversity of the proposed algorithm. Furthermore, a mechanism for transferring useful knowledge between subpopulations of the two search methods was developed. Considering their respective evolutionary characteristics, the advanced search experience assisted each other in escaping local optima. The experimental results fully demonstrate the effectiveness of our proposed algorithm.

In the future, we will select a limited number of relevant features during the initialization phase. It is an interesting research direction to reduce the computational time during the evolutionary process. Furthermore, the proposed algorithm has prospective applications in cancer classification, image recognition, and text extraction.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61876169, 61922072, 62206255, 62176238, 62106230), National Key R&D Program of China (Grant No. 2022YFD2001200), Natural Science Foundation of Henan Province (Grant No. 222300420088), Program for Science & Technology Innovation Teams in Universities of Henan Province (Grant No. 23IRTSTHN010), Program for Science & Technology Innovation Talents in Universities of Henan Province (Grant No. 23HASTIT023), and China Postdoctoral Science Foundation (Grant Nos. 2022M712878, 2022TQ0298, 2021T140616, 2021M692920).

References

1 Xue B, Zhang M J, Browne W N, et al. A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Computat, 2015, 20: 606–626
2 Tsamardinos I, Borboudakis G, Katsogridakis P, et al. A greedy feature selection algorithm for Big Data of high dimensionality. Mach Learn, 2019, 108: 149–202
3 Khaire U M, Dhanalakshmi R. Stability of feature selection algorithm: a review. J King Saud University-Comput Inf Sci, 2022, 34: 1060–1073
4 Li J, Liu H. Challenges of feature selection for big data analytics. IEEE Intell Syst, 2017, 32: 9–15
5 Gui J, Sun Z N, Ji S Q, et al. Feature selection based on structured sparsity: a comprehensive study. IEEE Trans Neural Netw Learn Syst, 2016, 28: 1490–1507
6 Yan C K, Ma J J, Luo H M, et al. A novel feature selection method for high-dimensional biomedical data based on an improved binary clonal flower pollination algorithm. Hum Hered, 2019, 84: 34–46
7 Chen K, Xue B, Zhang M J, et al. Evolutionary multitasking for feature selection in high-dimensional classification via particle swarm optimization. IEEE Trans Evol Computat, 2021, 26: 446–460
8 Chen K, Xue B, Zhang M, et al. An evolutionary multitasking-based feature selection method for high-dimensional classification. IEEE Trans Cybern, 2020, 52: 7172–7186
9 Chen K, Xue B, Zhang M, et al. Correlation-guided updating strategy for feature selection in classification with surrogate-assisted particle swarm optimization. IEEE Trans Evol Computat, 2021, 26: 1015–1029
10 Hancer E, Xue B, Zhang M. Differential evolution for filter feature selection based on information theory and feature ranking. Knowledge-Based Syst, 2018, 140: 103–119
11 Wang P, Xue B, Liang J, et al. Multiobjective differential evolution for feature selection in classification. IEEE Trans Cybern, 2023, 53: 4579–4593

12 Zhang Y, Wang Y H, Gong D W, et al. Clustering-guided particle swarm feature selection algorithm for high-dimensional imbalanced data with missing values. IEEE Trans Evol Computat, 2021, 26: 616–630
13 Song X F, Zhang Y, Gong D W, et al. Surrogate sample-assisted particle swarm optimization for feature selection on high-dimensional data. IEEE Trans Evol Computat, 2022, 27: 595–609
14 Liu S L, Wang H D, Peng W, et al. A surrogate-assisted evolutionary feature selection algorithm with parallel random grouping for high-dimensional classification. IEEE Trans Evol Computat, 2022, 26: 1087–1101
15 Maldonado S, López J. Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification. Appl Soft Computing, 2018, 67: 94–105
16 Xu H, Xue B, Zhang M J. A duplication analysis-based evolutionary algorithm for biobjective feature selection. IEEE Trans Evol Computat, 2020, 25: 205–218
17 Huang B, Buckley B, Kechadi T M. Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. Expert Syst Appl, 2010, 37: 3638–3646
18 Wang P, Xue B, Zhang M J, et al. A grid-dominance based multiobjective algorithm for feature selection in classification. In: Proceedings of IEEE Congress on Evolutionary Computation, Krakow, 2021. 2053–2060
19 Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Computat, 2002, 6: 182–197
20 Demir K, Nguyen B H, Xue B, et al. A decomposition based multiobjective evolutionary algorithm with ReliefF based local search and solution repair mechanism for feature selection. In: Proceedings of IEEE Congress on Evolutionary Computation, London, 2020. 1–8
21 Nguyen B H, Xue B, Ishibuchi H, et al. Multiple reference points MOEA/D for feature selection. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, Berlin, 2017. 157–158
22 Nguyen B H, Xue B, Andreae P, et al. Multiple reference points-based decomposition for multiobjective feature selection in classification: static and dynamic mechanisms. IEEE Trans Evol Computat, 2019, 24: 170–184
23 Zhang Q F, Li H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans Evol Computat, 2007, 11: 712–731
24 Chen L, Liu H L, Tan K C, et al. Evolutionary many-objective algorithm using decomposition-based dominance relationship. IEEE Trans Cybern, 2018, 49: 4129–4139
25 Li K, Deb K, Zhang Q F, et al. An evolutionary many-objective optimization algorithm based on dominance and decomposition. IEEE Trans Evol Computat, 2014, 19: 694–716
26 Zhang L J, Xie Y L, Chen J J, et al. A study on multiform multi-objective evolutionary optimization. Memetic Comp, 2021, 13: 307–318
27 Qiao K J, Yu K J, Qu B Y, et al. An evolutionary multitasking optimization framework for constrained multiobjective optimization problems. IEEE Trans Evol Computat, 2022, 26: 263–277
28 Liang J, Qiao K J, Yu K J, et al. Utilizing the relationship between unconstrained and constrained Pareto fronts for constrained multiobjective optimization. IEEE Trans Cybern, 2023, 53: 3873–3886
29 Qiao K J, Yu K J, Qu B Y, et al. Dynamic auxiliary task-based evolutionary multitasking for constrained multiobjective optimization. IEEE Trans Evol Computat, 2023, 27: 642–656
30 Gupta A, Ong Y S, Feng L. Multifactorial evolution: toward evolutionary multitasking. IEEE Trans Evol Computat, 2015, 20: 343–357
31 Qiao K J, Liang J, Yu K J, et al. Evolutionary constrained multiobjective optimization: scalable high-dimensional constraint benchmarks and algorithm. IEEE Trans Evol Computat, 2023. doi: 10.1109/TEVC.2023.3281666
32 Jiao R W, Xue B, Zhang M J. Benefiting from single-objective feature selection to multiobjective feature selection: a multiform approach. IEEE Trans Cybern, 2023, 53: 7773–7786
33 Zitzler E, Laumanns M, Thiele L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. TIK-report, 2001, 103 doi: 10.3929/ethz-a-004284029
34 Cheng F, Chu F X, Xu Y, et al. A steering-matrix-based multiobjective evolutionary algorithm for high-dimensional feature selection. IEEE Trans Cybern, 2021, 52: 9695–9708
35 Cheng F, Cui J J, Wang Q J, et al. A variable granularity search-based multiobjective feature selection algorithm for high-dimensional data classification. IEEE Trans Evol Computat, 2022, 27: 266–280
36 Wang P, Xue B, Liang J, et al. Differential evolution-based feature selection: a niching-based multiobjective approach. IEEE Trans Evol Computat, 2022, 27: 296–310
37 Liang J, Zhang Y Y, Qu B Y, et al. A multiform optimization framework for multi-objective feature selection in classification. IEEE Trans Evol Computat, 2023. doi: 10.1109/TEVC.2023.3284867
38 Tian Y, Lu C, Zhang X Y, et al. A pattern mining-based evolutionary algorithm for large-scale sparse multiobjective optimization problems. IEEE Trans Cybern, 2020, 52: 6784–6797
39 While L, Hingston P, Barone L, et al. A faster algorithm for calculating hypervolume. IEEE Trans Evol Computat, 2006, 10: 29–38
40 Zitzler E, Thiele L, Laumanns M, et al. Performance assessment of multiobjective optimizers: an analysis and review. IEEE Trans Evol Computat, 2003, 7: 117–132
41 Tian Y, Cheng R, Zhang X, et al. PlatEMO: a MATLAB platform for evolutionary multi-objective optimization. IEEE Comput Intell Mag, 2017, 12: 73–87
42 Xue B, Zhang M J, Browne W N. Particle swarm optimisation for feature selection in classification: novel initialisation and updating mechanisms. Appl Soft Computing, 2014, 18: 261–276
43 Hu P, Pan J S, Chu S C. Improved binary grey wolf optimizer and its application for feature selection. Knowledge-Based Syst, 2020, 195: 105746
44 Hong R, Xing L N, Zhang G T. Ensemble of selection operators for decomposition-based multi-objective evolutionary optimization. Swarm Evolary Computation, 2022, 75: 101198