

Article

# A Hybrid Initialization and Effective Reproduction-Based Evolutionary Algorithm for Tackling Bi-Objective Large-Scale Feature Selection in Classification

Hang Xu <sup>1,\*</sup> , Chaohui Huang <sup>1</sup>, Hui Wen <sup>2</sup> , Tao Yan <sup>1</sup> , Yuanmo Lin <sup>1</sup> and Ying Xie <sup>1</sup>

<sup>1</sup> School of Mechanical, Electrical & Information Engineering, Putian University, Putian 351100, China; jxlfhch@163.com (C.H.); yantaoshu@aliyun.com (T.Y.); linyuanmo@ptu.edu.cn (Y.L.); xieyingahu@126.com (Y.X.)

<sup>2</sup> New Engineering Industry College, Putian University, Putian 351100, China; wen\_hui81@163.com

\* Correspondence: xuhang@ptu.edu.cn

**Abstract:** Evolutionary algorithms have been widely used for tackling multi-objective optimization problems, while feature selection in classification can also be seen as a discrete bi-objective optimization problem that pursues minimizing both the classification error and the number of selected features. However, traditional multi-objective evolutionary algorithms (MOEAs) can encounter setbacks when the dimensionality of features explodes to a large scale, i.e., the curse of dimensionality. Thus, in this paper, we focus on designing an adaptive MOEA framework for solving bi-objective feature selection, especially on large-scale datasets, by adopting hybrid initialization and effective reproduction (called HIER). The former attempts to improve the starting state of evolution by composing a hybrid initial population, while the latter tries to generate more effective offspring by modifying the whole reproduction process. Moreover, the statistical experiment results suggest that HIER generally performs the best on most of the 20 test datasets, compared with six state-of-the-art MOEAs, in terms of multiple metrics covering both optimization and classification performances. Then, the component contribution of HIER is also studied, suggesting that each of its essential components has a positive effect. Finally, the computational time complexity of HIER is also analyzed, suggesting that HIER is not time-consuming at all and shows promising computational efficiency.

**Keywords:** bi-objective optimization; evolutionary algorithm; effective reproduction; hybrid initialization; large-scale feature selection

**MSC:** 68W50



**Citation:** Xu, H.; Huang, C.; Wen, H.; Yan, T.; Lin, Y.; Xie, Y. A Hybrid Initialization and Effective Reproduction-Based Evolutionary Algorithm for Tackling Bi-Objective Large-Scale Feature Selection in Classification. *Mathematics* **2024**, *12*, 554. <https://doi.org/10.3390/math12040554>

Academic Editor: Pedro A. Castillo Valdivieso

Received: 28 December 2023

Revised: 9 February 2024

Accepted: 10 February 2024

Published: 12 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Evolutionary algorithms [1] have been used as a common tool to solve optimization problems in the past decades. When the number of optimizing objectives is more than one, the problems become multi-objective optimization problems (MOPs) [2] and the algorithms become multi-objective evolutionary algorithms (MOEAs) [3]. Compared with other meta-heuristics [4], MOEAs have the superior feature of population-based global search mode and no need for domain knowledge, which is suitable to find a set of nondominated solutions in every generation of evolution. There have been numerous MOEAs proposed and designed all over the world, and they can be roughly categorized into the following frameworks: dominance-based MOEAs such as the classic nondominated sorting genetic algorithm (NSGA) [5], the improved NSGA with fast nondominated sorting (NSGA-II) [6], the improved NSGA for many objectives (NSGA-III) [7,8], a new dominance-relation-based algorithm [9], and a strengthened dominance-relation-based one [10]; decomposition-based MOEAs such as the classic MOEA based on decomposition (MOEA/D) [11], an improved MOEA/D with differential evolution (MOEA/D-

DE) [12], an improved MOEA/D with stable mating (MOEA/D-STM) [13], and a hierarchical decomposition-based MOEA (MOEA/HD) [14]; indicator-based MOEAs such as the classic hypervolume-based MOEA [15], the Minkowski distance-based MOEA [16], the polar-metric-based MOEA [17], and an indicator-based MOEA for many objectives [18]; surrogate-based MOEAs such as an offline data-driven one [19], an ensemble surrogate-based one [20], and one based on decomposition and multi-classifiers [21]; cooperative coevolutionary MOEAs such as the competitive-cooperative coevolutionary paradigm for dynamic multi-objective optimization [22], and so on [23,24]; multi-tasking-based MOEAs such as the multi-objective multi-factorial MOEA [25], and so on [26,27].

Moreover, owing to their flexible architecture and versatile capability, MOEAs have also been applied into many real-world complex optimization problems [28–32] including discrete optimization problems such as network community detection problems [33], neural network search problems [34], task offload problems [35,36], and feature selection problems [37–39]. In particular, feature selection has been widely used as a data preprocessing and dimensionality reduction technique for tackling large-scale classification datasets by selecting only a subset of useful features [40]. When pursuing the minimization of both the classification error and the number of selected features, feature selection becomes an MOP, which is quite suitable for MOEAs to deal with [41].

However, the “curse of dimensionality” still arises from the exponential increase in the total features in big data, challenging the search capability of traditional MOEA frameworks in addressing high-dimensional feature selection [42] with large-scale and sparse decision space [43]. This is fairly common in the age of big data when the features are large-scale but the samples are relatively inadequate. As the dimensionality grows, not only will the search space inevitably become expanded to being large scale and sparse for populations to search [44], but also the relationships between different features will become even more complicated to handle.

There have been various existing MOEAs attempting to solve discrete bi-objective feature selection problems, but many of them are either inefficient in addressing high-dimensional datasets or in need of too many complicated parameter settings and techniques. For example, Xue et al. [45] proposed an improved initialization method for feature selection, inspired by the classic forward and backward search ideas, but this method was only designed for single-objective particle swarm optimization. In light of this, Xu et al. [46] proposed a segmented initialization method to be integrated into existing MOEA frameworks, but its key parameter setting was fixed and inflexible for different datasets. Furthermore, Nguyen et al. [37] designed an improved decomposition-based MOEA with static or dynamic reference points for solving feature selection, but the highest dimensionality tested was only 649, far from large scale. Subsequently, Xu et al. [47] also proposed a nondominated sorting-based MOEA with solution duplication analyses before environmental selection, but still not focusing on high-dimensional datasets. Recently, the decision variable pre-analysis idea [48,49] has become popular for addressing large-scale optimization [50] in sparse decision space. For instance, Tian et al. [51] designed a large-scale MOEA framework for searching in sparse decision space, but its single feature performance pre-analysis process consumes a large number of objective function evaluations, which can in return obstruct normal evolution if the computational resource is inadequate.

To overcome the above-mentioned drawbacks caused by high-dimensional feature selection, in this paper, we propose a hybrid initialization method and an effective reproduction-based MOEA framework, with the aim of boosting the search ability of MOEAs with a promising start of a hybrid initial population, as well as improving both the diversity and convergence performance via reproducing high-quality offspring. Moreover, regarding the aforementioned shortcomings in existing MOEAs, the framework of HIER should be as simple and robust as possible, with adaptive parameter settings dynamically adjusted according to the tested dataset and the population state. To this end, in the hybrid initialization process, a set of adaptively generated extra initial populations is used to vastly explore

the objective space and to exploit forward searching areas early, with relatively smaller numbers of selected features. Furthermore, in the effective reproduction process, a totally random mating method is adopted for fairness of each parent solution, and an effective crossover operation is conducted for generating more valid offspring, and a self-adaptively set dynamic mutation scale is utilized for increasing variations within each offspring. In a word, the hybrid initialization method boosts the search ability of MOEAs, and the effective reproduction method helps to balance the population diversity and convergence, thereby combining both of them will maximize the efforts of better facing the challenges from the “curse of dimensionality” in addressing high-dimensional feature selection.

The major contributions of this paper are summarized as follows:

- First of all, a hybrid initialization method, abbreviated as **HI**, is proposed in order to boost the search ability of MOEAs in addressing high-dimensional bi-objective feature selection, by composing a promising hybrid initial population which vastly explores the objective space and adaptively exploits its forward areas.
- As a supplement, an effective reproduction method, abbreviated as **ER**, is also proposed to balance the diversity and convergence factors, and to further increase the offspring quality for better variations, via adopting an effective crossover operation and a dynamic mutation scale.
- Comprehensive experiments are conducted in this work regarding the general performance and component contributions versus different state-of-the-art MOEAs, in terms of multiple metrics, tested on a series of 20 datasets. The empirical results and analyses confirm the search advantages of HIER as well as its high efficiency.

The remainder of this paper is organized as follows. First, the related works are introduced in Section 2. Then, the proposed HIER algorithm is detailed in Section 3. After that, the experiment setups are provided in Section 4, and the empirical results are studied in Section 5. Finally, the conclusions and future work are given in Section 6.

## 2. Related Works

### 2.1. Bi-Objective Optimization Problem

Normally, feature selection [52] can be defined as a multi-objective optimization problem that can be shown as follows:

$$\begin{aligned} & \text{minimize } \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_M(\mathbf{x}))^T \\ & \text{subject to } \mathbf{x} = (x_1, x_2, \dots, x_D), x_i = \{0, 1\} \end{aligned} \quad (1)$$

where  $M$  is the dimension of objective space, i.e., objectives to be optimized, and  $D$  is the dimension of decision space, i.e., the total number of features to be selected. In this paper,  $M$  is set to two, and  $\mathbf{F}(\mathbf{x})$  is the objective vector of  $\mathbf{x}$ , while  $f_i(\mathbf{x})$  denotes the corresponding value in the direction of the  $i$ th objective (so-called  $f_i$  direction). Moreover,  $\mathbf{x} = (x_1, x_2, \dots, x_D)$  is the decision vector of a solution where 1 means selecting a feature and 0 means not selecting it. Here, the first objective function  $f_1(\mathbf{x})$  can be defined as follows:

$$f_1(\mathbf{x}) = \sum_{i=1}^D x_i / D \quad (2)$$

where the function value discretely ranges from 0 to 1, i.e.,  $\in \{0, 1/D, 2/D, \dots, 1\}$ , denoting the ratio of currently selected features. In addition, the second objective function  $f_2(\mathbf{x})$  denotes the classification error rate related to the classification results of the previously selected features in  $\mathbf{x}$ , the value of which also discretely ranges from 0 to 1, limited by the size of samples. If the classification accuracy obtained from a specific classifier, e.g., KNN, is denoted as  $\theta$ , then the second objective function  $f_2(\mathbf{x})$  can be formalized as follows:

$$f_2(\mathbf{x}) = 1 - \theta \quad (3)$$

where, for example, if the classification accuracy obtained is 89%, then the  $f_2$  objective value will become 11% or 0.11.

## 2.2. Evolutionary Feature Selection

Being widely used for decades, evolutionary feature selection [53] can be roughly classified into two categories: wrapper- or filter-based approaches [54,55]. Normally, wrapper-based approaches [56,57] utilize a classification model, like SVM or KNN [58], as a black box to evaluate the classification accuracy. By contrast, filter-based approaches [59,60] are mostly independent of any classifier, ignoring the classification results of currently selected features during evolution. Therefore, a wrapper-based approach is normally more accurate than a filter-based one but generally consumes more computational time. In this paper, we focus on studying wrapper-based approaches for bi-objective evolutionary feature selection, seeking to improve the optimization and classification performance.

In fact, during recent years, many evolutionary algorithms have been developed for solving feature selection. For example, Chen et al. [42,61] proposed two algorithms based on multi-tasking for high-dimensional classification, and Xue et al. [62] also proposed efficient initialization and updating mechanisms for particle swarm optimization. However, these three approaches are only designed for a single objective and may be infeasible for multiple objectives. As for MOEAs, Xu et al. [47] proposed a duplication-analysis-based algorithm, which is specially designed to handle bi-objective feature selection optimization. Moreover, Tian et al. [63] proposed an MOEA for sparse decision space, which can also be used to solve large-scale feature selection. Thus, these two MOEAs have been adopted in the experiment as algorithms to compare our proposed one against. Nevertheless, there are still many other MOEAs that have been reported in recent years that can be used for tackling multi-objective feature selection, such as the variable granularity search-based algorithm in [64], the surrogate-assisted and filter-based algorithm in [65], and the steering-matrix-based algorithm in [66].

## 3. Proposed Algorithm

In this section, we first introduce the general framework of the proposed HIER algorithm, and then further explain its essential components, i.e., the hybrid initialization and effective reproduction processes. Finally, we give more discussions of how the proposed approaches take effect, with two simple examples shown in figures. It is also worth noting that in all the pseudocode of this paper, the variable *Rand* is used as a randomly generated probability in order to make a yes or no decision.

### 3.1. General Framework

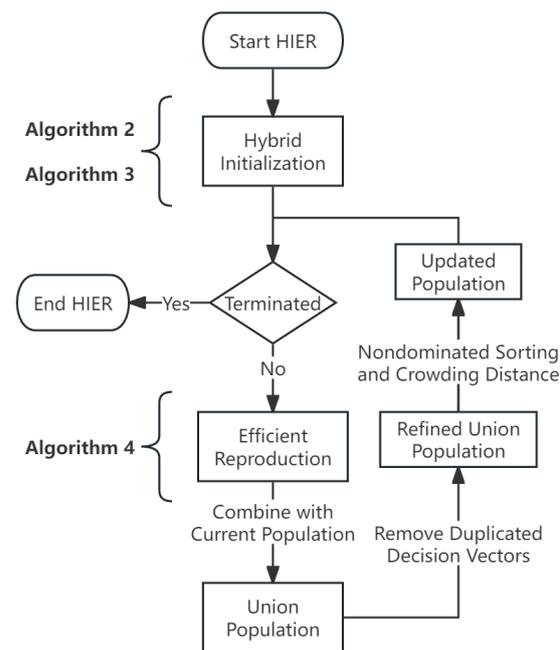
The general framework of the proposed HIER algorithm is given by the pseudocode Algorithm 1. The population size  $N$  and the feature dimension  $D$  are input, while the final optimized population *Pop* is output after terminating the evolutionary feature selection. In Algorithm 1, the initialization and reproduction processes are, respectively, conducted by Algorithms 2 and 4, which will both be explained later in the following sections. The environmental selection process, which truncates the combination of the previous population and the new offspring into the current population, is almost the same as that in the most well-known traditional dominance-based framework NSGA-II [6], except that HIER additionally removes all the duplicated solutions in the decision space beforehand. Moreover, for a better explanation of the methodology, a flow chart for the general framework of HIER is illustrated in Figure 1, where each evolutionary step in Algorithm 1 is presented, along with the invoked algorithms.

**Algorithm 1** *GeneralFramework*( $N, D$ )

---

**Input:** population size  $N$ , feature dimension  $D$ ;  
**Output:** final population  $Pop$ ;  
1:  $Pop = HI(N, D)$ ; // initialization by Algorithm 2  
2: **while** termination criterion not reached **do**  
3:    $Osp = ER(Pop)$ ; // reproduction by Algorithm 4  
4:    $Pop \leftarrow$  remove all the duplicated decision vectors  
   from  $Pop \cup Osp$ ;  
5:    $Pop \leftarrow$  select  $N$  best solutions from  $Pop$  by  
   nondominated sorting and crowding distances;  
6: **end while**

---



**Figure 1.** Flowchart for the general framework of HIER.

### 3.2. Hybrid Initialization

HIEA adopts an adaptive and hybrid initialization mechanism specially designed for large-scale feature selection, given by the pseudocode Algorithm 2 accompanied by Algorithm 3. In Algorithm 2, we first generate an initial population in a traditional random sampling way as the starting point. Then, the number of extra initial populations to be adaptively generated is roughly estimated, according to the ratio between the feature dimension and the population size. Normally, a larger value of the above ratio means a greater number of extra initial populations is going to be generated. All the extra initial populations are obtained by Algorithm 3, with a so-called distribution probability parameter input. This parameter actually controls the probable distribution area of a new population generated in the objective space. In other words, the value of the input distribution probability acts as a symmetric axis where all the objective vectors are distributed around in the  $f_1$  direction. Generally speaking, in Algorithm 3, each variable value of the decision vector is randomly decided to obtain the value zero or one, where one means selecting that feature and zero means not selecting it. Back to Algorithm 2, all the hybrid populations, including the starting initial population and all the extra initial populations, are at last truncated by the same nondominated sorting and crowding distance methods used in the previously introduced general framework to reserve only the best  $N$  unique solutions as the final initial population we want. It should also be noted that the exponential base in line 5 of Algorithm 2 is set to 0.5 for the following two reasons. First, 0.5 is the starting point of the adaptively set population initialization, which is right in the middle of the objective space,

appropriately balancing both diversity and convergence factors. Second, the proposed hybrid initialization method (HI) is based on the idea of binary search, which constantly reduces its previous exploring area by half, i.e., 0.5 as the base.

---

**Algorithm 2**  $HI(N, D)$ 


---

**Input:** population size  $N$ , feature dimension  $D$ ;

**Output:** initial population  $Pop$ ;

- 1:  $Pop \leftarrow$  generate a traditional initial population by randomly sampling  $N$  decision vectors from the  $D$ -dimensional decision space (i.e., feature space);
  - 2:  $K = \lfloor \log_2(D/N) \rfloor$ ; // get the adaptive number of extra initial populations
  - 3: **if**  $K > 0$  **then**
  - 4:     **for**  $i = 1, 2, \dots, K$  **do**
  - 5:          $Pop = Pop \cup NewPop(N, D, 0.5^{i+1})$ ;
  - 6:     **end for**
  - 7:      $Pop \leftarrow$  use nondominated sorting and crowding distances to select  $N$  best unique decision vectors from  $Pop$ ;
  - 8: **end if**
- 

---

**Algorithm 3**  $NewPop(N, D, P)$ 


---

**Input:** population size  $N$ , feature dimension  $D$ ; distribution probability  $P$ ;

**Output:** new population  $Pop$ ;

- 1:  $Pop = Zeros(N, D)$ ; // 0 matrix of order  $N$  by  $D$
  - 2: **for**  $i = 1, 2, \dots, N$  **do**
  - 3:     **for**  $j = 1, 2, \dots, D$  **do**
  - 4:         **if**  $Rand < P$  **then**
  - 5:              $Pop(i, j) = 1$ ; // select this feature
  - 6:         **end if**
  - 7:     **end for**
  - 8: **end for**
- 

### 3.3. Effective Reproduction

Accompanied by the previously introduced hybrid initialization, an effective reproduction process is specially designed for HIER to cope with large-scale feature dimensionality. The pseudocode Algorithm 4 shows how this reproduction process works, including the mating, crossover, and mutation procedures. In Algorithm 4, all the parents are randomly selected from the current population, with every solution holding equal opportunities for mating, and no preference for elite solutions. In the for loop of Algorithm 4, the efficiency and validity are both ensured by only performing crossover within the different decision variables between the pairwise parents. Moreover, for mutation, we first obtain the number of selected features in the parent solution (i.e., value 1 decision variables), and then use it to estimate the scale of the genes to be randomly mutated. Normally, the mutation scale is positively related to the number of already selected features, but the chances for larger mutation scales are set to be much smaller than the traditional one-gene mutation scale. This delicate principle is conducted by lines 10 to 14 in Algorithm 4, while  $Rand$  means generating a random probability or a vector of probabilities (i.e.,  $Rand(1, D)$ ). For example, if  $t = 10,000$  in line 8 of Algorithm 4, which apparently is a large-scale case, then  $r$  in line 9 could be any integer within 100, like 10 for instance. Then, if  $r = 10$ , then the dynamically set mutation scale  $s$  in line 11 will probably be set to a random integer value around 10, meaning the later mutation operation in line 15 will swap probably 10 features among all the currently selected features. This actually grants those solutions that have selected a high number of features an opportunity to mutate more genes rather than just one. And more subtly, this probability  $1/r$  is inversely proportional to the mutation scale  $s$ . In this way, the mutation scale can be dynamically set according to the current evolutionary state of the

solutions, while a solution with high-dimensional selected features will have a delicately controlled chance to perform a much bolder mutation operation within the previously set mutation scale.

**Algorithm 4**  $ER(Pop)$

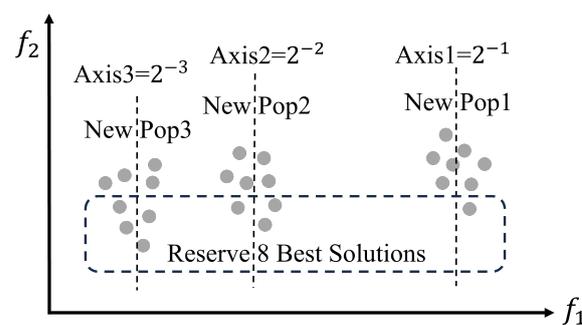
```

Input: current population  $Pop$ ;
Output: offspring set  $Osp$ ;
1:  $Pars \leftarrow$  randomly select  $N$  pairs of solutions from  $Pop$  as parents; // mating
2: for  $i = 1, 2, \dots, N$  do
3:    $par_1, par_2 \leftarrow$  get the pairwise parents in  $Pars(i)$ ;
4:    $dv \leftarrow$  find different decision variable indexes
     within  $par_1$  and  $par_2$ ;
5:    $k \leftarrow$  get a random integer between  $Size(dv)$ ;
6:    $j \leftarrow$  randomly select  $k$  indexes from  $dv$ ;
7:    $par_1(j) = par_2(j)$ ; // crossover
8:    $t \leftarrow$  get the number of value 1 variables in  $par_1$ ;
9:    $r \leftarrow$  get a random integer within  $\lceil \sqrt{t} \rceil$ ;
10:  if  $Rand < 1/r$  then
11:     $s = Rand(1, D) < r/D$ ;
12:  else
13:     $s = Rand(1, D) < 1/D$ ;
14:  end if
15:   $par_1(s) = \neg par_1(s)$ ; // mutation
16:   $Osp(i) = par_1$ ; // get the new offspring
17: end for

```

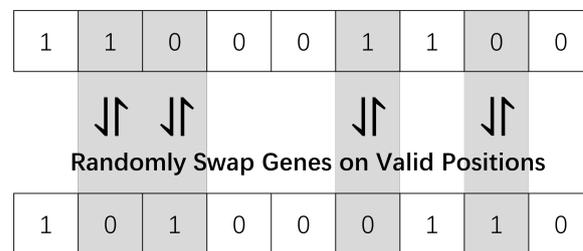
3.4. More Discussions

Figure 2 gives a simple example of how to compose a hybrid initial population with three other newly generated initial populations adaptively distributed in the objective space. According to Algorithm 2, the newly generated initial populations start from the middle of the objective space, i.e., from *Axis1* in Figure 2. Then, using 0.5 as the exponential base, the next generated new initial population is distributed around *Axis2*, i.e.,  $0.25$  or  $2^{-2}$ . The same goes for the third newly generated initial population, distributed around *Axis3*, i.e.,  $2^{-3}$ , in the objective space. Finally, by truncating the three newly generated initial populations together into a hybrid one with the preset population size (reserving eight solutions in this example), the final hybrid initial population then not only explores more diverse genes by those adaptively distributed new initial populations across the objective space, but also exploits more elite genes through the earlier search in the forward areas of the objective space. It should also be noted that the distribution axis of each newly generated initial population in Figure 2 is controlled by the input  $P$  parameter in Algorithm 3, where a smaller  $P$  means selecting less features and a larger one means selecting more, in each newly generated solution within a population.



**Figure 2.** An example of composing a hybrid initial population composed of three new populations: NewPop1 with Axis1, NewPop2 with Axis2, and NewPop3 with Axis3.

Figure 3 gives a simple example of how to conduct the effective crossover operation in reproduction. The upper and lower parts of Figure 3, respectively, show two decision vectors for parent solutions, with 1 meaning selecting that position of the feature and 0 meaning not selecting it. It can be seen from the directions and positions of the interactive arrows located in the middle of Figure 3 that only variables in the gray areas can be swapped, i.e., from 1 to 0, or from 0 to 1. These gray areas are in fact valid areas where the upper and lower variables in two decision vectors have distinct values. Thus, crossover operations within the above-introduced valid areas can be effective, as they make no invalid operations, such as swapping 0 to 0, or 1 to 1. Combining the above-introduced effective crossover with the mating and mutation operations in Algorithm 4, the reproduction process of HIER is then efficiently improved in the following three aspects: first, the mating is based on fair and random selection from the current population to bring in more diversity; second, the crossover only focuses on swapping different variables between two parent solutions in order to make valid offspring; third, the mutation adaptively adjusts the scale of variation to skip potential local optima and to help find more varied offspring.



**Figure 3.** An example of how to perform effective crossover in reproduction.

In HIER, although both hybrid initialization and effective reproduction have considered the balance between exploration and exploitation, yet the former still focuses more on the convergence boosting and the latter focuses more on the diversity maintaining. Compared with other algorithms, there are two major advantages of HIER, which also implies its uniqueness, as explained in the following. First, the adaptively distributed initial populations vastly explore the objective space and the finally composed hybrid initial population that starts the evolution with a relatively smaller number of selected features within each solution helps to exploits the forward area of objective space. The adaptive generation mechanism of new initial populations also takes advantage of the binary search idea which reduces its previous search range to the half and thereby increases the exploration efficiency. Second, the effective crossover operation helps to make valid offspring, while the dynamically set mutation scale can adaptively control the balance between exploration and exploitation, thereby more delicately balancing the diversity and convergence during evolution. Therefore, via combining the hybrid initialization and effective reproduction, HIER makes the most complementary contribution to improving both the exploration and exploitation factors in addressing high-dimensional feature selection.

## 4. Experiment Setups

### 4.1. Datasets for Test Problems

In this work, a total of 20 open-source classification datasets [67] were used to test the optimization performances of MOEAs in tackling the bi-objective feature selection problem. Details of those datasets are shown in Table 1, where the number of features, samples, and classes are shown in different columns. It can be seen that the number of total features in each dataset varies from 100 to 10,509, which covers a wide range of feature dimensions but concentrates on the high-dimensional ones. Moreover, the number of samples ranges from 50 to 606, and that of the classes from 2 to 15, which also indicates the generality and comprehensiveness of the test problems.

**Table 1.** Attributes for each classification dataset used as test problems.

No.	Datasets	Features	Samples	Classes
1	HillValley	100	606	2
2	MUSK1	166	476	2
3	Arrhythmia	278	452	13
4	Yale	1024	165	15
5	Colon	2000	62	2
6	SRBCT	2308	83	4
7	AR10P	2400	130	10
8	PIE10P	2420	210	10
9	Leukemia1	5327	72	3
10	Tumor9	5726	60	9
11	TOX171	5748	171	4
12	Brain1	5920	90	5
13	Leukemia2	7070	72	2
14	ALLAML	7129	72	2
15	Carcinom	9182	174	11
16	Nci9	9712	60	9
17	Arcene	10,000	200	2
18	Orlraws10P	10,304	100	10
19	Brain2	10,367	50	4
20	Prostate	10,509	102	2

#### 4.2. Algorithms for Comparison Analyses

In this paper, six state-of-the-art MOEAs (i.e., NSGA-II [6], MOEA/D [11], HypE [15], MOEA/HD [14], SparseEA [63], and DAEA [47]) are tested to compare their performance results with the proposed HIER algorithm. The above MOEAs are selected as comparison algorithms versus HIER for the following reasons. First, NSGA-II, MOEA/D, and HypE are among the most classic and well-known MOEAs, which are based on dominance, decomposition, and indicator, respectively. These three algorithms will test the advantages of HIER over traditional MOEAs in addressing feature selection problems. Second, MOEA/HD is a recently published MOEA, based on the combination of dominance and decomposition, which is specifically designed for solving complex MOPs. This algorithm is used to compare the performance of HIER in tackling the discrete complicated optimization environment of feature selection. Third, DAEA and SparseEA are both recently published MOEAs based on dominance, which are specifically designed for tackling large-scale feature selection problems. These two algorithms will challenge HIER in efficiently searching the sparse and large-scale decision space on high-dimensional datasets.

#### 4.3. Metrics for Performance Results

In this study, multiple performance indicators are used to measure the general performances of each algorithm in terms of both optimization and classification. To be more specific, the hypervolume (HV) [68] metric is used as the main indicator to measure the MOEAs' optimization performance, with its reference point set to (1, 1). As a supplement, the minimum classification error (MCE) and number of selected features (NSFs) metrics [46,47] are used to measure the best classification performance obtained on the final test data, respectively, reflecting the  $f_2$  and  $f_1$  objective values of a solution. For example, if solution  $x^*$  within a population obtains the best classification accuracy on the final test data (this could be any specific tested dataset), then  $f_2(x^*)$  denotes the current MCE value for that population and  $f_1(x^*)$  denotes the current NSF value. The formalized definitions of  $f_1$  and  $f_2$  are, respectively, shown in Equations (2) and (3) in Section 2.1. Generally speaking, greater HV values mean better performance, while smaller MCE and NSF values are preferred. Finally, the Wilcoxon's test with a significance level of 5% is adopted to identify the differences in pairwise comparisons.

#### 4.4. Settings for Computational Environments

In this work, all the comparison algorithms use the same traditional initialization method for fairness, while the reproduction methods and other parameter settings are inherited from the studies in which they were presented. In addition, all the algorithm codes are programmed and run on an open-source MATLAB platform [69]. During evolution, each classification dataset is randomly divided into the training and test subsets with a proportion of about 70/30, following the stratified split process [47]. Moreover, the KNN ( $K = 5$ ) model is adopted for classification, with 10-fold cross-validation on the training data to avoid feature selection bias [70]. Finally, each experiment is independently run 20 times with a series of randomly preset starting seeds, while the population size is set to 100 and the termination criterion (the number of objective function evaluations) is set to 10,000 (about 100 generations) for each algorithm.

### 5. Experiment Studies

#### 5.1. General Performance Studies

The general performance of each algorithm is shown in Tables 2–5. More specifically, Table 3 gives the multi-objective optimization performances based on the final nondominated solutions obtained by each algorithm on the test data, which are measured by the widely used HV performance indicator. First of all, Table 2 gives the overall Friedman’s test on all seven algorithms, showing the mean performance ranks among them: HIER still ranks the first in performance for all three metrics on both the training and test data. Furthermore, in Table 3, the proposed HIER performs the best on almost every dataset in terms of the HV metric, and only loses on MUSK1 to SparseEA and DAEA, which is a relatively low-dimensional dataset. By contrast, Table 4 combined with Table 5 gives the classification performance of each algorithm based on the classification accuracy (i.e., the MCE metric) and efficiency (i.e., the NSF metric). In detail, Table 4 shows the minimum classification error while Table 5 shows the related number of selected features. In Table 4, HIER performs the best on every dataset in terms of the MCE metric, showing its outstanding superiority in classification accuracy. In Table 5, HIER performs the best on almost every dataset in terms of the NSF metric, and only loses on HillValley and MUSK1, which are relatively low-dimensional datasets, generally showing excellent classification efficiency. Therefore, based on the above studies, it is suggested that HIER generally performs the best in most of the test instances, compared with the other six MOEAs, in terms of all three metrics.

**Table 2.** Mean ranks calculated by Friedman’s test on both training and test data.

Metric	Data	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
HV	Train	1.1975	4.5800	3.4475	4.7475	5.5650	6.4850	1.9775
	Test	1.2138	4.5250	3.6162	5.0713	5.3213	6.0563	2.1963
MCE	Train	1.2125	4.2800	5.0888	3.9987	4.7988	5.8812	2.7400
	Test	1.8475	4.3800	4.6850	4.7263	4.6387	4.1350	3.5875
NSF	Train	1.3575	4.5287	2.8575	4.6937	5.4763	6.7300	2.3563
	Test	1.3700	4.4150	2.9987	4.9212	5.4825	6.5850	2.2275

**Table 3.** Mean HV performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
HillValley	6.2892e-01	5.8491e-01	† 6.2589e-01	6.0937e-01	† 6.2585e-01	† 6.2597e-01	† 6.2599e-01
	±8.46e-03	±2.51e-02	±1.04e-02	± 1.81e-02	±8.54e-03	±8.77e-03	±8.50e-03
MUSK1	8.8189e-01	8.2291e-01	8.6667e-01	8.2231e-01	8.4699e-01	8.9529e-01	9.0313e-01
	±1.84e-02	±2.49e-02	±1.64e-02	±2.14e-02	±3.17e-02	±1.42e-02	±1.55e-02

Table 3. Cont.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
Arrhythmia	6.9949e-01 ±1.49e-02	6.3085e-01 ±2.66e-02	6.6549e-01 ±1.58e-02	6.1568e-01 ±3.67e-02	4.8263e-01 ±1.69e-02	5.3277e-01 ±5.28e-02	† 6.9537e-01 ±1.60e-02
Yale	7.2988e-01 ±3.46e-02	4.8777e-01 ±1.34e-02	5.1271e-01 ±2.62e-02	4.7391e-01 ±2.55e-02	4.9196e-01 ±3.39e-02	4.8107e-01 ±1.64e-02	6.0505e-01 ±2.48e-02
Colon	8.8458e-01 ±5.47e-02	5.5002e-01 ±2.65e-02	6.0785e-01 ±4.78e-02	5.4386e-01 ±2.77e-02	5.2395e-01 ±3.30e-02	4.9869e-01 ±2.01e-02	6.6802e-01 ±3.95e-02
SRBCT	8.8158e-01 ±7.70e-02	2.8407e-01 ±2.05e-03	3.1776e-01 ±2.03e-03	2.8506e-01 ±2.29e-03	2.5538e-01 ±1.81e-03	2.4792e-01 ±1.69e-03	3.0281e-01 ±1.96e-02
AR10P	7.9190e-01 ±3.81e-02	3.6309e-01 ±2.01e-02	3.7087e-01 ±2.26e-02	3.4460e-01 ±1.96e-02	3.4495e-01 ±1.75e-02	3.2227e-01 ±1.11e-02	4.3142e-01 ±2.05e-02
PIE10P	9.5463e-01 ±2.49e-02	6.0231e-01 ±1.06e-02	6.4575e-01 ±1.13e-02	5.8837e-01 ±1.23e-02	5.8831e-01 ±1.22e-02	5.4344e-01 ±6.44e-03	6.9820e-01 ±1.57e-02
Leukemia1	9.4797e-01 ±3.25e-02	5.2904e-01 ±1.81e-02	5.4315e-01 ±3.10e-02	5.1931e-01 ±1.15e-02	5.1321e-01 ±2.40e-02	4.8596e-01 ±1.64e-02	6.0516e-01 ±2.04e-02
Tumor9	5.0677e-01 ±5.87e-02	2.7901e-01 ±2.80e-02	2.8321e-01 ±2.54e-02	2.6172e-01 ±3.52e-02	2.6731e-01 ±2.35e-02	2.6728e-01 ±1.96e-02	3.0436e-01 ±2.38e-02
TOX171	8.3099e-01 ±3.79e-02	4.8294e-01 ±8.58e-03	4.8764e-01 ±1.97e-02	4.7065e-01 ±1.61e-02	4.7759e-01 ±1.65e-02	4.5753e-01 ±1.19e-02	5.4227e-01 ±1.92e-02
Brain1	7.8591e-01 ±3.82e-02	4.7180e-01 ±3.11e-03	4.9062e-01 ±1.09e-02	4.7043e-01 ±3.61e-03	4.5347e-01 ±1.00e-02	4.3124e-01 ±1.78e-03	5.1273e-01 ±8.53e-03
Leukemia2	9.4408e-01 ±5.57e-02	5.3600e-01 ±8.95e-03	5.4496e-01 ±1.94e-02	5.3301e-01 ±1.69e-02	5.1258e-01 ±1.79e-02	4.9718e-01 ±9.84e-03	6.0216e-01 ±1.67e-02
ALLAML	9.5646e-01 ±4.54e-02	5.2052e-01 ±1.52e-02	5.3575e-01 ±1.34e-02	5.1175e-01 ±1.64e-02	5.0598e-01 ±1.44e-02	4.8530e-01 ±1.52e-02	5.8265e-01 ±1.83e-02
Carcinom	8.8720e-01 ±2.73e-02	5.1803e-01 ±1.09e-02	5.2327e-01 ±1.55e-02	5.0847e-01 ±1.18e-02	5.0915e-01 ±1.10e-02	4.8714e-01 ±8.18e-03	5.8095e-01 ±1.18e-02
Nci9	5.0449e-01 ±7.48e-02	2.4060e-01 ±2.54e-02	2.6158e-01 ±2.94e-02	2.3947e-01 ±2.57e-02	2.3696e-01 ±2.21e-02	2.2538e-01 ±2.00e-02	2.7073e-01 ±2.75e-02
Arcene	8.6704e-01 ±2.45e-02	3.6248e-01 ±1.10e-03	3.7242e-01 ±1.85e-03	3.6265e-01 ±2.03e-03	3.4447e-01 ±1.24e-03	3.3745e-01 ±1.29e-03	3.8590e-01 ±2.75e-03
Orlraws10P	9.6479e-01 ±2.83e-02	5.3898e-01 ±7.53e-03	5.4470e-01 ±9.58e-03	5.3648e-01 ±5.90e-03	5.2969e-01 ±8.00e-03	5.0571e-01 ±3.88e-03	5.9507e-01 ±8.65e-03
Brain2	7.2102e-01 ±7.46e-02	3.9029e-01 ±2.15e-02	3.8245e-01 ±2.43e-02	3.7751e-01 ±2.55e-02	3.7819e-01 ±2.12e-02	3.6871e-01 ±1.66e-02	4.3346e-01 ±2.82e-02
Prostate	9.4399e-01 ±4.03e-02	4.6288e-01 ±1.29e-02	4.5987e-01 ±1.51e-02	4.5114e-01 ±1.06e-02	4.5588e-01 ±1.16e-02	4.4194e-01 ±8.71e-03	5.2044e-01 ±1.53e-02

Table 4. Mean MCE performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
HillValley	4.0055e-01 ±9.75e-03	4.2225e-01 ±1.67e-02	† 4.0440e-01 ±1.17e-02	4.1346e-01 ±1.44e-02	† 4.0220e-01 ±8.84e-03	† 4.0412e-01 ±1.00e-02	† 4.0412e-01 ±9.68e-03
MUSK1	8.5315e-02 ±1.74e-02	† 9.5105e-02 ±2.18e-02	1.1538e-01 ±2.04e-02	9.8601e-02 ±1.80e-02	† 9.7203e-02 ±2.33e-02	1.0210e-01 ±1.64e-02	† 9.1608e-02 ±1.84e-02
Arrhythmia	3.2554e-01 ±1.68e-02	3.7914e-01 ±3.29e-02	3.5072e-01 ±1.91e-02	3.8921e-01 ±4.51e-02	4.6367e-01 ±1.71e-02	4.3885e-01 ±4.18e-02	† 3.3129e-01 ±1.80e-02

Table 4. Cont.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
Yale	2.9333e-01 ±3.86e-02	3.4778e-01 ±2.53e-02	3.6222e-01 ±3.82e-02	3.5667e-01 ±3.85e-02	3.4000e-01 ±5.20e-02	3.1333e-01 ±2.78e-02	† 2.9889e-01 ±3.01e-02
Colon	1.2632e-01 ±6.01e-02	2.1316e-01 ±4.35e-02	2.0789e-01 ±6.50e-02	2.1053e-01 ±4.52e-02	2.3421e-01 ±5.26e-02	1.9737e-01 ±3.77e-02	1.6842e-01 ±5.01e-02
SRBCT	1.2800e-01 ±8.47e-02	6.4000e-01 ±1.14e-16	6.4000e-01 ±1.14e-16	6.4000e-01 ±1.14e-16	6.4000e-01 ±1.14e-16	6.4000e-01 ±1.14e-16	6.4000e-01 ±1.14e-16
AR10P	2.2750e-01 ±4.21e-02	4.9000e-01 ±3.38e-02	5.2000e-01 ±3.77e-02	5.1375e-01 ±3.09e-02	5.1500e-01 ±3.08e-02	5.0750e-01 ±2.00e-02	4.7000e-01 ±3.10e-02
PIE10P	4.8333e-02 ±2.75e-02	9.6667e-02 ±1.28e-02	1.0333e-01 ±2.27e-02	1.0167e-01 ±1.61e-02	1.0167e-01 ±1.42e-02	1.0167e-01 ±1.07e-02	8.5833e-02 ±1.82e-02
Leukemia1	5.6818e-02 ±3.57e-02	1.6136e-01 ±3.12e-02	1.8182e-01 ±4.17e-02	1.6818e-01 ±2.14e-02	1.7045e-01 ±4.14e-02	1.5909e-01 ±3.13e-02	1.4773e-01 ±2.90e-02
Tumor9	5.4167e-01 ±6.47e-02	5.9722e-01 ±5.06e-02	6.0556e-01 ±4.38e-02	6.2222e-01 ±6.40e-02	6.1111e-01 ±4.42e-02	5.8056e-01 ±3.81e-02	6.1111e-01 ±4.03e-02
TOX171	1.8396e-01 ±4.19e-02	2.2358e-01 ±1.53e-02	2.4057e-01 ±3.56e-02	2.3302e-01 ±3.02e-02	2.2736e-01 ±2.84e-02	2.1321e-01 ±2.38e-02	2.1792e-01 ±2.84e-02
Brain1	2.3519e-01 ±4.21e-02	2.5926e-01 ±0.00e+00	2.5926e-01 ±0.00e+00	2.5926e-01 ±0.00e+00	2.5926e-01 ±0.00e+00	2.5926e-01 ±0.00e+00	2.5926e-01 ±0.00e+00
Leukemia2	6.1364e-02 ±6.13e-02	1.3182e-01 ±1.40e-02	1.4545e-01 ±2.80e-02	1.3182e-01 ±2.91e-02	1.4545e-01 ±2.80e-02	1.2727e-01 ±1.87e-02	1.2045e-01 ±2.67e-02
ALLAML	4.7727e-02 ±5.00e-02	1.5682e-01 ±2.75e-02	1.5909e-01 ±2.33e-02	1.6591e-01 ±3.05e-02	1.6364e-01 ±2.28e-02	1.5000e-01 ±2.99e-02	1.5000e-01 ±2.60e-02
Carcinom	1.2308e-01 ±3.02e-02	1.4231e-01 ±1.91e-02	1.5000e-01 ±2.30e-02	1.4904e-01 ±2.15e-02	1.5000e-01 ±2.03e-02	† 1.3846e-01 ±1.60e-02	† 1.3269e-01 ±1.64e-02
Nci9	5.4474e-01 ±8.24e-02	6.5789e-01 ±4.68e-02	6.3421e-01 ±5.26e-02	6.6053e-01 ±4.67e-02	6.4211e-01 ±4.39e-02	6.5526e-01 ±4.00e-02	6.3158e-01 ±4.83e-02
Arcene	1.4583e-01 ±2.70e-02	4.3333e-01 ±1.14e-16	4.3333e-01 ±1.14e-16	4.3333e-01 ±1.14e-16	4.3333e-01 ±1.14e-16	4.3333e-01 ±1.14e-16	4.3333e-01 ±1.14e-16
Orlraws10P	3.8333e-02 ±3.11e-02	1.0500e-01 ±1.22e-02	1.1167e-01 ±1.63e-02	1.0333e-01 ±1.03e-02	1.0500e-01 ±1.22e-02	1.0167e-01 ±7.45e-03	1.0333e-01 ±1.03e-02
Brain2	3.0667e-01 ±8.21e-02	3.7667e-01 ±3.91e-02	3.9667e-01 ±4.03e-02	3.9333e-01 ±4.79e-02	3.9000e-01 ±3.91e-02	3.7000e-01 ±3.40e-02	3.7333e-01 ±4.54e-02
Prostate	6.1290e-02 ±4.43e-02	2.4194e-01 ±2.45e-02	2.6129e-01 ±2.94e-02	2.5645e-01 ±1.95e-02	2.4355e-01 ±1.95e-02	2.2742e-01 ±1.65e-02	2.2581e-01 ±2.56e-02

Table 5. Mean NSF performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
HillValley	4.3500e+00 ±3.01e+00	1.1200e+01 ±6.46e+00	† 3.2500e+00 ±1.92e+00	† 6.6000e+00 ±4.91e+00	† 3.8500e+00 ±3.45e+00	† 4.0500e+00 ±2.91e+00	† 3.5500e+00 ±3.10e+00
MUSK1	2.8650e+01 ±1.12e+01	† 3.3950e+01 ±1.13e+01	2.1150e+01 ±1.08e+01	† 2.9950e+01 ±9.45e+00	† 2.8450e+01 ±1.21e+01	† 2.6800e+01 ±1.50e+01	† 2.7800e+01 ±8.97e+00
Arrhythmia	9.4500e+00 ±2.86e+00	1.6550e+01 ±6.20e+00	1.2150e+01 ±3.87e+00	1.8850e+01 ±1.14e+01	5.0900e+01 ±3.31e+00	3.6950e+01 ±7.40e+00	† 1.0050e+01 ±5.69e+00
Yale	1.9000e+01 ±8.35e+00	3.3400e+02 ±2.99e+01	2.6825e+02 ±1.46e+01	3.3800e+02 ±1.10e+01	3.2915e+02 ±1.68e+01	3.8475e+02 ±2.03e+01	2.3105e+02 ±4.03e+01

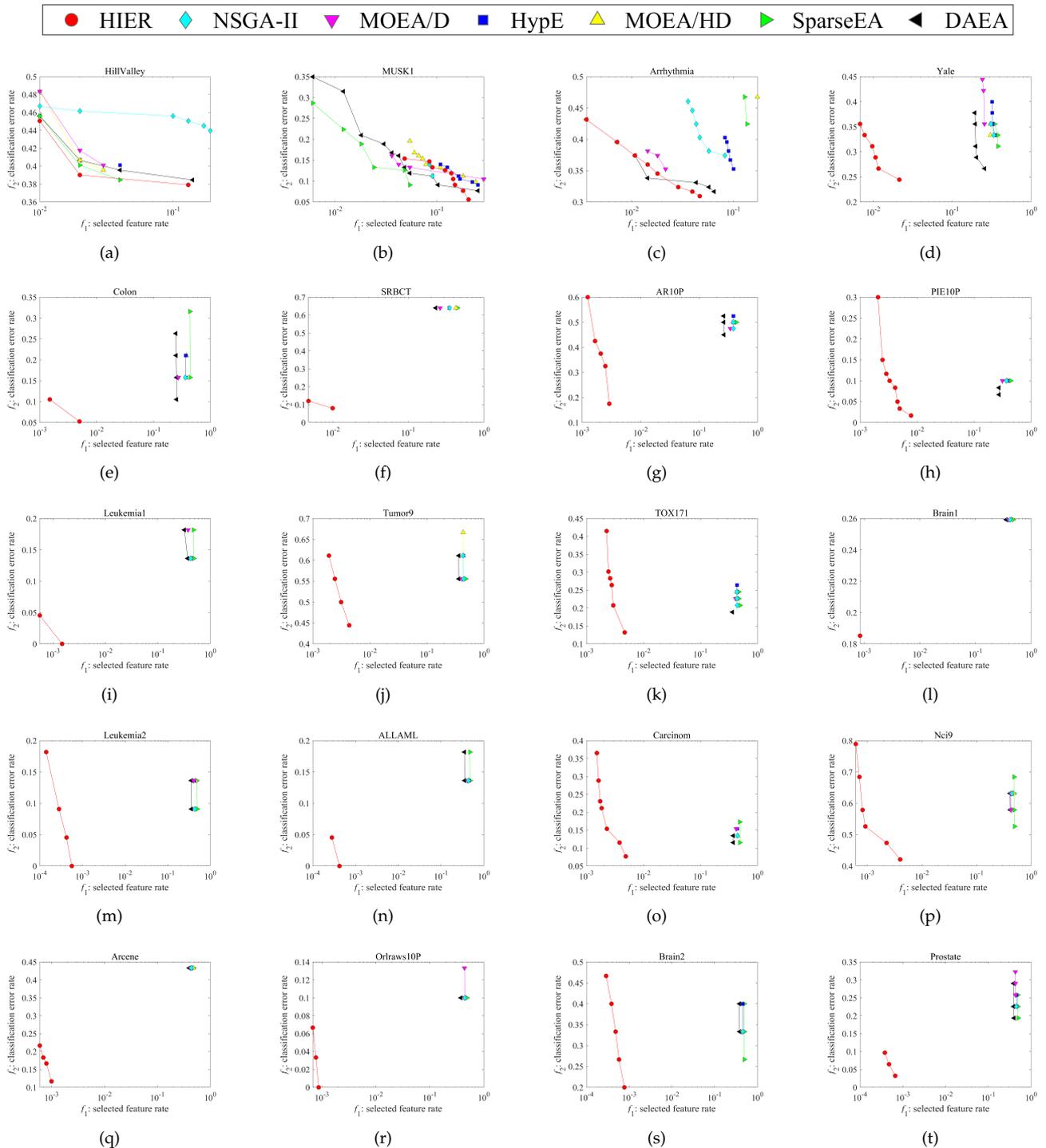
Table 5. Cont.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
Colon	3.9000e+00 ±4.01e+00	6.9905e+02 ±1.50e+01	5.5090e+02 ±5.59e+01	7.2470e+02 ±3.36e+01	7.3735e+02 ±2.43e+01	8.7520e+02 ±3.37e+01	4.8045e+02 ±3.17e+01
SRBCT	1.1700e+01 ±1.27e+01	8.1420e+02 ±1.24e+01	6.0965e+02 ±1.23e+01	8.0820e+02 ±1.39e+01	9.8835e+02 ±1.10e+01	1.0337e+03 ±1.03e+01	7.0040e+02 ±1.19e+02
AR10P	1.4850e+01 ±5.97e+00	9.1535e+02 ±1.77e+01	7.8260e+02 ±2.61e+01	9.4745e+02 ±5.17e+01	9.3110e+02 ±2.85e+01	1.0808e+03 ±3.27e+01	6.9200e+02 ±7.48e+01
PIE10P	1.4800e+01 ±3.72e+00	9.0765e+02 ±2.51e+01	7.6480e+02 ±2.78e+01	9.3650e+02 ±1.91e+01	9.4000e+02 ±2.50e+01	1.0852e+03 ±2.95e+01	6.7210e+02 ±3.75e+01
Leukemia1	3.5500e+00 ±1.54e+00	2.2284e+03 ±2.11e+01	2.0482e+03 ±8.40e+01	2.2672e+03 ±3.81e+01	2.3012e+03 ±3.01e+01	2.5398e+03 ±3.41e+01	1.7904e+03 ±7.59e+01
Tumor9	2.0750e+01 ±1.13e+01	2.4570e+03 ±3.33e+01	2.3302e+03 ±4.08e+01	2.5034e+03 ±2.67e+01	2.5214e+03 ±4.34e+01	2.7432e+03 ±2.99e+01	1.9975e+03 ±6.94e+01
TOX171	2.9450e+01 ±1.27e+01	2.5100e+03 ±6.23e+01	2.3762e+03 ±3.48e+01	2.5703e+03 ±6.66e+01	2.5303e+03 ±3.03e+01	2.7681e+03 ±3.54e+01	2.0640e+03 ±6.63e+01
Brain1	6.3500e+00 ±4.44e+00	2.4922e+03 ±2.65e+01	2.3319e+03 ±9.29e+01	2.5038e+03 ±3.08e+01	2.6484e+03 ±8.53e+01	2.8378e+03 ±1.52e+01	2.1434e+03 ±7.27e+01
Leukemia2	2.0500e+00 ±1.39e+00	3.0410e+03 ±3.44e+01	2.8929e+03 ±1.06e+02	3.0675e+03 ±3.57e+01	3.1837e+03 ±5.71e+01	3.4228e+03 ±3.21e+01	2.5413e+03 ±8.82e+01
ALLAML	2.1000e+00 ±1.07e+00	3.0848e+03 ±2.61e+01	2.9301e+03 ±4.67e+01	3.1164e+03 ±3.68e+01	3.1823e+03 ±4.63e+01	3.4504e+03 ±3.33e+01	2.5657e+03 ±5.94e+01
Carcinom	2.3350e+01 ±7.86e+00	4.0971e+03 ±3.82e+01	3.9811e+03 ±6.53e+01	4.1602e+03 ±4.30e+01	4.1489e+03 ±3.06e+01	4.4926e+03 ±3.26e+01	3.4802e+03 ±1.01e+02
Nci9	1.3250e+01 ±1.20e+01	4.2878e+03 ±3.25e+01	4.0833e+03 ±3.29e+01	4.2805e+03 ±3.16e+01	4.6002e+03 ±2.86e+01	4.7272e+03 ±1.76e+01	3.8900e+03 ±4.30e+01
Arcene	1.3350e+01 ±6.95e+00	4.4209e+03 ±1.99e+01	4.2405e+03 ±3.35e+01	4.4179e+03 ±3.68e+01	4.7479e+03 ±2.25e+01	4.8754e+03 ±2.34e+01	3.9959e+03 ±5.00e+01
Orlraws10P	1.3250e+01 ±5.37e+00	4.5808e+03 ±3.63e+01	4.4630e+03 ±4.45e+01	4.6233e+03 ±3.48e+01	4.6971e+03 ±5.78e+01	5.0229e+03 ±2.67e+01	3.8903e+03 ±7.47e+01
Brain2	7.2500e+00 ±3.70e+00	4.6416e+03 ±4.55e+01	4.5841e+03 ±1.21e+02	4.7026e+03 ±2.84e+01	4.7289e+03 ±4.01e+01	5.0841e+03 ±3.71e+01	3.9348e+03 ±8.19e+01
Prostate	6.5000e+00 ±3.46e+00	4.7128e+03 ±5.72e+01	4.5869e+03 ±5.63e+01	4.7604e+03 ±5.75e+01	4.7962e+03 ±4.31e+01	5.1540e+03 ±4.61e+01	4.0243e+03 ±1.16e+02

## 5.2. Nondominated Solution Distributions

For more intuitive observations of performance, Figure 4 illustrates the nondominated solution distributions of each algorithm in the objective space in terms of the Pareto curves, always choosing their median HV performance runs for the sake of fairness. In Figure 4, it can be seen that the proposed HIER generally performs the best on almost every dataset, except for MUSK1 in Figure 4b, which is also consistent with the previously introduced HV performance shown in Table 3. Nevertheless, HIER still obtains the smallest classification error rate on MUSK1 in Figure 4b in the  $f_2$  objective direction compared with the other algorithms. Generally speaking, HIER can obtain significantly better population diversity and outstanding convergence as the dimensionality of features grows to the large-scale level. In fact, even on relatively low-dimensional datasets, such as HillValley and Arrhythmia in Figure 4a,c, respectively, HIER can still obtain better diversity and convergence of population distributions than the other algorithms. It is also worth noting that in this paper the  $f_1$  objective direction (i.e., the x-axis) in Figure 4 is illustrated on the logarithmic (base 10) scale, so that even the distributions of other algorithms far apart from that of the proposed algorithm can be presented in the same picture for clearer observations. Overall,

the proposed HIER can generally achieve the most diverse and converged nondominated solutions when compared to all the other comparison algorithms in terms of the Pareto curves drawn in the objective space.



**Figure 4.** Nondominated solution distributions in terms of the Pareto curves drawn in the objective space, with the run of median HV performance on the final test data, obtained by each algorithm. (a) HillValley. (b) MUSK1. (c) Arrhythmia. (d) Yale. (e) Colon. (f) SRBCT. (g) AR10P. (h) PIE10P. (i) Leukemia1. (j) Tumor9. (k) TOX171. (l) Brain1. (m) Leukemia2. (n) ALLAML. (o) Carcinom. (p) Nci9. (q) Arcene. (r) Orlraws10P. (s) Brain2. (t) Prostate.

Moreover, it can be observed from Figure 4 that the final obtained nondominated solutions for each algorithm are generally sparse on the test data, which is mainly because of the following reasons. First, as can be seen from Table 1, the number of samples used for training and testing on some datasets is actually not sufficient, such as the ALLAML dataset with only 72 samples but having a large-scale number of 7129 features to explore. This unbalanced ratio between samples and features is quite common for large-scale discrete optimization, which makes the feasible decision space rather sparse. As a result, the nondominated solutions obtained by each algorithm are restricted to a much smaller number compared with the cases in continuous optimization. Second, the nondominated solutions shown in Figure 4 are obtained from the test data, which have already been filtered by the conversion process from training to testing. Thus, what used to be nondominated on the training data may become dominated on the test data, which makes the number of nondominated solutions obtained on the test data become even smaller.

### 5.3. Component Contribution Analyses

For more comprehensive studies, the component contribution performances of the proposed hybrid initialization (abbreviated as HI) and effective reproduction (abbreviated as ER) methods are shown in Tables 6–9. To be more specific, three variant algorithms, i.e., Base/HI, Base/ER, and Base, are created to be compared with the proposed HIER; Base denotes the baseline algorithm deleting both the HI and ER methods from HIER. Base/HI and Base/ER denote the variant algorithms adding the HI and ER methods to Base, respectively. Overall, it is suggested that HIER performs the best on most of the datasets in Table 7–9, and always ranks the first in Table 6, proving that the proposed HI and ER methods together can make the greatest contribution. Furthermore, it is also implied from the marking situation of ✓ shown in Tables 7–9 that either Base/HI or Base/ER can generally perform better than the Base algorithm on most of the datasets, in terms of all three metrics, and always ranks better than Base in Table 6, which also implies that each component of HIER (i.e., HI or ER) can have a positive effect on improving the algorithm’s performance but their combination (i.e., the entire HIER) makes the greatest contribution. This is mainly owing to the complementary effects made by combining both hybrid initialization and effective reproduction in HIER, which has already been discussed in the last paragraph of Section 3.4, for improving both the exploration and exploitation factors in addressing high-dimensional feature selection.

**Table 6.** Mean ranks calculated by Friedman’s test on both training and test data, for each algorithm, with best ranks marked in gray for component contribution analyses.

Metric	Data	HIER	Base/HI	Base/ER	Base
HV	Train	1.1300	2.1125	2.8075	3.9500
	Test	1.3350	1.9225	2.8250	3.9175
MCE	Train	1.3375	1.9038	3.0688	3.6900
	Test	1.7475	2.0975	2.8687	3.2862
NSF	Train	1.2463	2.0275	2.8862	3.8400
	Test	1.2337	2.0500	2.8862	3.8300

**Table 7.** Mean HV performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †, also a ✓ means a better performance than that of the corresponding Base algorithm.

Dataset	HIER		Base/HI	Base/ER	Base
HillValley	6.2892e-01	✓	6.0206e-01	† 6.2892e-01	6.0206e-01
	±8.46e-03		±2.58e-02	±8.46e-03	±2.58e-02
MUSK1	8.8189e-01	✓	8.3186e-01	†8.8189e-01	8.3186e-01
	±1.84e-02		±2.82e-02	±1.84e-02	±2.82e-02

Table 7. Cont.

Dataset	HIER		Base/HI		Base/ER		Base
Arrhythmia	6.9949e-01 ±1.49e-02	✓	6.7456e-01 ±1.70e-02	✓	† 6.9239e-01 ±1.28e-02	✓	6.1565e-01 ±3.45e-02
Yale	7.2988e-01 ±3.46e-02	✓	† 7.1830e-01 ±4.03e-02	✓	6.1279e-01 ±2.52e-02	✓	4.9729e-01 ±2.18e-02
Colon	8.8458e-01 ±5.47e-02	✓	8.8048e-01 ±4.48e-02	✓	6.9894e-01 ±2.99e-02	✓	5.5372e-01 ±2.85e-02
SRBCT	8.8158e-01 ±7.70e-02	✓	8.4499e-01 ±6.29e-02	✓	4.6735e-01 ±1.67e-01	✓	2.8731e-01 ±2.35e-03
AR10P	7.9190e-01 ±3.81e-02	✓	7.0132e-01 ±4.87e-02	✓	4.4533e-01 ±2.39e-02	✓	3.5733e-01 ±1.93e-02
PIE10P	9.5463e-01 ±2.49e-02	✓	† 9.7024e-01 ±1.86e-02	✓	7.3321e-01 ±1.55e-02	✓	6.0007e-01 ±9.66e-03
Leukemia1	9.4797e-01 ±3.25e-02	✓	9.2420e-01 ±5.80e-02	✓	6.3440e-01 ±2.17e-02	✓	5.3239e-01 ±1.68e-02
Tumor9	5.0677e-01 ±5.87e-02	✓	4.8747e-01 ±5.37e-02	✓	3.1102e-01 ±2.29e-02	✓	2.8046e-01 ±2.54e-02
TOX171	8.3099e-01 ±3.79e-02	✓	† 8.3398e-01 ±3.01e-02	✓	5.5152e-01 ±1.21e-02	✓	4.8962e-01 ±1.27e-02
Brain1	7.8591e-01 ±3.82e-02	✓	† 7.7663e-01 ±3.34e-02	✓	5.5687e-01 ±5.32e-03	✓	4.7613e-01 ±3.95e-03
Leukemia2	9.4408e-01 ±5.57e-02	✓	9.2057e-01 ±5.67e-02	✓	6.2342e-01 ±2.32e-02	✓	5.3959e-01 ±1.43e-02
ALLAML	9.5646e-01 ±4.54e-02	✓	9.4528e-01 ±4.46e-02	✓	5.9713e-01 ±1.73e-02	✓	5.2342e-01 ±1.73e-02
Carcinom	8.8720e-01 ±2.73e-02	✓	8.7518e-01 ±2.53e-02	✓	5.8885e-01 ±1.02e-02	✓	5.2509e-01 ±1.08e-02
Nci9	5.0449e-01 ±7.48e-02	✓	† 4.9181e-01 ±8.25e-02	✓	2.8777e-01 ±2.90e-02	✓	2.4995e-01 ±2.88e-02
Arcene	8.6704e-01 ±2.45e-02	✓	8.5578e-01 ±2.96e-02	✓	4.2071e-01 ±1.86e-03	✓	3.6608e-01 ±2.63e-03
Orlraws10P	9.6479e-01 ±2.83e-02	✓	9.5124e-01 ±3.47e-02	✓	6.1629e-01 ±1.12e-02	✓	5.4507e-01 ±6.24e-03
Brain2	7.2102e-01 ±7.46e-02	✓	6.7439e-01 ±8.16e-02	✓	4.3591e-01 ±3.54e-02	✓	3.9461e-01 ±1.79e-02
Prostate	9.4399e-01 ±4.03e-02	✓	9.1163e-01 ±5.66e-02	✓	5.2506e-01 ±1.55e-02	✓	4.6475e-01 ±1.37e-02

Table 8. Mean MCE performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †, also a ✓ means a better performance than that of the corresponding Base algorithm.

Dataset	HIER		Base/HI		Base/ER		Base
HillValley	4.0055e-01 ±9.75e-03	✓	4.1099e-01 ±2.08e-02		† 4.0055e-01 ±9.75e-03	✓	4.1099e-01 ±2.08e-02
MUSK1	8.5315e-02 ±1.74e-02	✓	9.4755e-02 ±1.78e-02		† 8.5315e-02 ±1.74e-02	✓	9.4755e-02 ±1.78e-02
Arrhythmia	3.2554e-01 ±1.68e-02	✓	3.4137e-01 ±1.77e-02	✓	† 3.2950e-01 ±1.56e-02	✓	3.9137e-01 ±4.20e-02

Table 8. Cont.

Dataset	HIER		Base/HI		Base/ER		Base
Yale	2.9333e-01 ±3.86e-02	✓	† 2.9778e-01 ±4.53e-02	✓	† 3.0111e-01 ±3.26e-02	✓	3.4222e-01 ±3.26e-02
Colon	1.2632e-01 ±6.01e-02	✓	† 1.2632e-01 ±4.95e-02	✓	1.5789e-01 ±4.52e-02	✓	2.0789e-01 ±4.97e-02
SRBCT	1.2800e-01 ±8.47e-02	✓	† 1.6200e-01 ±6.93e-02	✓	4.7600e-01 ±2.32e-01	✓	6.4000e-01 ±1.14e-16
AR10P	2.2750e-01 ±4.21e-02	✓	3.2250e-01 ±5.37e-02	✓	4.6750e-01 ±3.15e-02	✓	5.0375e-01 ±3.37e-02
PIE10P	4.8333e-02 ±2.75e-02	✓	2.5000e-02 ±2.06e-02	✓	7.4167e-02 ±2.26e-02	✓	1.0083e-01 ±1.38e-02
Leukemia1	5.6818e-02 ±3.57e-02	✓	† 7.9545e-02 ±6.41e-02	✓	1.4318e-01 ±3.05e-02	✓	1.6364e-01 ±3.09e-02
Tumor9	5.4167e-01 ±6.47e-02	✓	† 5.6111e-01 ±5.95e-02	✓	6.1111e-01 ±3.60e-02		6.0000e-01 ±4.63e-02
TOX171	1.8396e-01 ±4.19e-02	✓	† 1.7736e-01 ±3.32e-02	✓	2.1415e-01 ±1.96e-02	✓	2.1887e-01 ±2.24e-02
Brain1	2.3519e-01 ±4.21e-02	✓	† 2.4259e-01 ±3.70e-02	✓	2.5926e-01 ±0.00e+00		2.5926e-01 ±0.00e+00
Leukemia2	6.1364e-02 ±6.13e-02	✓	† 8.6364e-02 ±6.24e-02	✓	1.3636e-01 ±3.30e-02		1.3636e-01 ±2.55e-02
ALLAML	4.7727e-02 ±5.00e-02	✓	† 5.9091e-02 ±4.91e-02	✓	1.6364e-01 ±2.72e-02		1.6136e-01 ±3.12e-02
Carcinom	1.2308e-01 ±3.02e-02	✓	† 1.3462e-01 ±2.79e-02	✓	† 1.3269e-01 ±1.52e-02	✓	† 1.3558e-01 ±1.82e-02
Nci9	5.4474e-01 ±8.24e-02	✓	† 5.5789e-01 ±9.10e-02	✓	6.4474e-01 ±4.48e-02	✓	6.4737e-01 ±5.15e-02
Arcene	1.4583e-01 ±2.70e-02	✓	† 1.5667e-01 ±3.26e-02	✓	4.3333e-01 ±1.14e-16		4.3333e-01 ±1.14e-16
Orlraws10P	3.8333e-02 ±3.11e-02	✓	† 5.1667e-02 ±3.82e-02	✓	1.0667e-01 ±1.37e-02		1.0333e-01 ±1.03e-02
Brain2	3.0667e-01 ±8.21e-02	✓	† 3.5667e-01 ±8.99e-02	✓	3.9000e-01 ±5.83e-02		3.7667e-01 ±3.26e-02
Prostate	6.1290e-02 ±4.43e-02	✓	9.5161e-02 ±6.23e-02	✓	2.4194e-01 ±2.67e-02	✓	2.4516e-01 ±2.20e-02

Table 9. Mean NSF performance on the final test data, with best results marked in gray and those with insignificant differences prefixed by †, also a ✓ means a better performance than that of the corresponding Base algorithm.

Dataset	HIER		Base/HI		Base/ER		Base
HillValley	4.3500e+00 ±3.01e+00	✓	9.3000e+00 ±5.10e+00		† 4.3500e+00 ±3.01e+00	✓	9.3000e+00 ±5.10e+00
MUSK1	2.8650e+01 ±1.12e+01		† 2.7550e+01 ±7.44e+00		† 2.8650e+01 ±1.12e+01		† 2.7550e+01 ±7.44e+00
Arrhythmia	9.4500e+00 ±2.86e+00	✓	1.3250e+01 ±4.70e+00	✓	† 1.0100e+01 ±4.36e+00	✓	1.9000e+01 ±6.83e+00
Yale	1.9000e+01 ±8.35e+00	✓	2.9850e+01 ±1.10e+01	✓	1.8300e+02 ±1.15e+01	✓	3.2705e+02 ±3.23e+01

Table 9. Cont.

Dataset	HIER		Base/HI		Base/ER		Base
Colon	3.9000e+00 ±4.01e+00	✓	1.3750e+01 ±4.92e+00	✓	4.0850e+02 ±3.07e+01	✓	7.0025e+02 ±1.44e+01
SRBCT	1.1700e+01 ±1.27e+01	✓	3.2550e+01 ±1.22e+01	✓	4.4115e+02 ±4.02e+01	✓	7.9455e+02 ±1.43e+01
AR10P	1.4850e+01 ±5.97e+00	✓	2.8450e+01 ±5.45e+00	✓	6.1860e+02 ±4.64e+01	✓	9.3145e+02 ±5.75e+01
PIE10P	1.4800e+01 ±3.72e+00	✓	3.2850e+01 ±1.52e+01	✓	5.9990e+02 ±4.80e+01	✓	9.0505e+02 ±1.69e+01
Leukemia1	3.5500e+00 ±1.54e+00	✓	2.4300e+01 ±5.25e+00	✓	1.5901e+03 ±4.96e+01	✓	2.2003e+03 ±4.57e+01
Tumor9	2.0750e+01 ±1.13e+01	✓	4.0000e+01 ±1.21e+01	✓	1.8952e+03 ±5.26e+01	✓	2.4187e+03 ±5.87e+01
TOX171	2.9450e+01 ±1.27e+01	✓	5.3700e+01 ±1.57e+01	✓	2.0159e+03 ±6.33e+01	✓	2.4728e+03 ±4.90e+01
Brain1	6.3500e+00 ±4.44e+00	✓	2.6750e+01 ±8.84e+00	✓	1.7675e+03 ±4.53e+01	✓	2.4553e+03 ±3.36e+01
Leukemia2	2.0500e+00 ±1.39e+00	✓	9.2500e+00 ±2.75e+00	✓	2.2483e+03 ±6.90e+01	✓	2.9917e+03 ±3.79e+01
ALLAML	2.1000e+00 ±1.07e+00	✓	9.7500e+00 ±2.86e+00	✓	2.3678e+03 ±1.09e+02	✓	3.0438e+03 ±4.66e+01
Carcinom	2.3350e+01 ±7.86e+00	✓	5.2000e+01 ±1.35e+01	✓	3.3603e+03 ±8.42e+01	✓	4.0616e+03 ±5.96e+01
Nci9	1.3250e+01 ±1.20e+01	✓	2.9150e+01 ±1.01e+01	✓	3.2576e+03 ±8.85e+01	✓	4.1949e+03 ±3.19e+01
Arcene	1.3350e+01 ±6.95e+00	✓	3.1550e+01 ±1.06e+01	✓	3.3642e+03 ±3.38e+01	✓	4.3557e+03 ±4.77e+01
Orlraws10P	1.3250e+01 ±5.37e+00	✓	2.4550e+01 ±4.20e+00	✓	3.5998e+03 ±8.68e+01	✓	4.5159e+03 ±3.09e+01
Brain2	7.2500e+00 ±3.70e+00	✓	2.5250e+01 ±7.18e+00	✓	3.7093e+03 ±1.02e+02	✓	4.5658e+03 ±4.33e+01
Prostate	6.5000e+00 ±3.46e+00	✓	2.8750e+01 ±9.90e+00	✓	3.7854e+03 ±6.45e+01	✓	4.6622e+03 ±5.72e+01

#### 5.4. Computational Time Complexity

The computational time complexity of the proposed HIER algorithm is first estimated by counting the probable time consumption of each algorithm step, and then specifically counted in seconds for the general running time under the same computational environment with all the other comparison algorithms. First of all, the greatest time complexity of Algorithm 2 lies in two parts: the generation of new populations and the final truncation based on nondominated sorting. The former costs  $O(ND)$  according to the two-layer nested for loop in Algorithm 3, while the latter costs  $O(MN^2)$  according to the common complexity for fast nondominated sorting [6]. Then, the greatest time complexity of Algorithm 4 is also  $O(ND)$ , according to its major for loop with crossover and mutation operations. Finally, the greatest time complexity of all the other operations in Algorithm 1 is  $O(MN^2)$  for the nondominated sorting and crowding distance-based environmental selection. Therefore, by comparing the above estimations, the computational time complexity of HIER is theoretically estimated as  $O(MN^2)$ .

However, if also considering the evaluation of objective values, which is actually the most time-consuming part for evolutionary feature selection, it is hard to accurately

estimate the theoretical complexity, but the real consumption is closely related to the number of selected features (i.e., the NSF metric performance shown in Tables 5 and 9); where normally a smaller NSF value leads to a smaller time consumption for classification. As already discussed before, HIER generally performs the best in Tables 5 and 9 in terms of the NSF metric, suggesting its promising efficiency in the classification process for evolutionary feature selection. In fact, this is also verified by counting each comparison algorithm’s general computational time in seconds, shown by Tables 10 and 11. It is implied from these two tables that the computational time complexity of HIER is generally not high at all, consuming the smallest time in a majority of test instances in Table 10, and also ranking the first in terms of the general Friedman’s test in Table 11.

**Table 10.** Mean running time counted in seconds for each algorithm, with best results marked in gray and those with insignificant differences prefixed by †.

Dataset	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
HillValley	1.6542e+02 ±2.29e+00	1.3886e+02 ±5.31e+00	1.1863e+02 ±1.30e+00	1.3268e+02 ±2.11e+00	1.3296e+02 ±3.03e+00	1.2853e+02 ±4.11e+00	1.3046e+02 ±2.42e+00
MUSK1	1.5299e+02 ±3.12e+00	1.2916e+02 ±3.75e+00	1.0806e+02 ±2.35e+00	1.2888e+02 ±2.70e+00	1.2909e+02 ±5.20e+00	1.2506e+02 ±4.18e+00	1.1893e+02 ±2.17e+00
Arrhythmia	1.2357e+02 ±2.96e+00	† 1.2683e+02 ±8.77e+00	† 1.2521e+02 ±6.33e+00	1.1269e+02 ±1.30e+01	1.4024e+02 ±3.60e+00	1.5856e+02 ±4.70e+00	1.1801e+02 ±5.54e+00
Yale	7.1823e+01 ±8.93e-01	1.3820e+02 ±2.61e+00	1.3998e+02 ±2.37e+00	1.4007e+02 ±1.93e+00	1.4154e+02 ±2.96e+00	1.4353e+02 ±1.92e+00	1.3358e+02 ±2.81e+00
Colon	4.1085e+01 ±4.34e-01	6.5188e+01 ±8.52e-01	7.4923e+01 ±1.54e+00	6.7189e+01 ±9.17e-01	6.8834e+01 ±9.49e-01	6.8019e+01 ±7.63e-01	7.4603e+01 ±9.65e-01
SRBCT	6.3116e+01 ±6.48e-01	1.2351e+02 ±9.46e+00	1.2505e+02 ±1.40e+00	1.1532e+02 ±1.17e+00	1.2714e+02 ±1.23e+00	1.1498e+02 ±1.16e+00	1.2557e+02 ±4.77e+00
AR10P	1.4548e+02 ±2.21e+00	3.1300e+02 ±4.12e+01	2.6741e+02 ±7.94e+00	2.7840e+02 ±9.79e+00	2.8346e+02 ±7.91e+00	2.6225e+02 ±6.51e+00	2.7489e+02 ±1.06e+01
PIE10P	2.5280e+02 ±2.73e+00	1.0896e+04 ±1.11e+04	6.7676e+02 ±1.93e+01	7.0708e+02 ±2.34e+01	7.1007e+02 ±2.47e+01	6.5288e+02 ±1.28e+01	6.1220e+02 ±1.69e+01
Leukemia1	1.9317e+02 ±1.92e+00	4.5790e+02 ±6.68e+00	3.3060e+02 ±9.08e+00	3.4796e+02 ±3.73e+00	3.4160e+02 ±3.17e+00	2.6899e+02 ±4.54e+00	3.7272e+02 ±5.57e+00
Tumor9	1.6337e+02 ±2.98e+00	4.4825e+02 ±2.07e+01	3.0571e+02 ±3.95e+00	3.1021e+02 ±2.93e+00	3.0621e+02 ±3.18e+00	2.3338e+02 ±5.95e+00	3.4640e+02 ±5.31e+00
TOX171	5.3810e+02 ±3.02e+01	1.2920e+03 ±2.58e+01	1.4180e+03 ±2.91e+01	1.4445e+03 ±2.96e+01	1.4350e+03 ±3.31e+01	8.9183e+02 ±8.35e+00	1.3980e+03 ±2.98e+01
Brain1	2.6227e+02 ±4.27e+00	9.4599e+02 ±5.02e+01	6.3031e+02 ±1.09e+01	6.4571e+02 ±2.65e+01	6.9449e+02 ±1.13e+01	4.2454e+02 ±1.32e+01	6.8087e+02 ±1.81e+01
Leukemia2	2.4579e+02 ±3.70e+00	6.4440e+02 ±1.01e+01	6.3381e+02 ±1.84e+01	6.5402e+02 ±1.25e+01	6.5913e+02 ±1.61e+01	3.4785e+02 ±8.51e+00	6.0041e+02 ±1.06e+01
ALLAML	2.5027e+02 ±4.92e+00	8.9622e+02 ±3.85e+01	6.4559e+02 ±1.46e+01	6.6036e+02 ±1.44e+01	6.6043e+02 ±1.45e+01	3.5437e+02 ±8.26e+00	6.1682e+02 ±1.07e+01
Carcinom	8.4547e+02 ±3.53e+01	2.1541e+03 ±7.39e+01	2.3686e+03 ±3.80e+01	2.4095e+03 ±4.63e+01	2.4011e+03 ±4.49e+01	8.9313e+02 ±8.86e+00	2.3300e+03 ±4.14e+01
Nci9	3.1374e+02 ±1.83e+01	1.0316e+03 ±4.64e+01	7.0608e+02 ±1.95e+01	6.9304e+02 ±2.35e+01	6.6720e+02 ±1.47e+01	† 2.9938e+02 ±3.33e+00	7.7338e+02 ±1.28e+01
Arcene	1.0505e+03 ±2.62e+01	2.9155e+03 ±4.38e+01	3.0924e+03 ±6.12e+01	2.9755e+03 ±1.58e+01	3.2501e+03 ±5.47e+01	9.6620e+02 ±1.22e+01	3.0443e+03 ±3.80e+01
Orlraws10P	5.4502e+02 ±4.24e+01	1.3103e+03 ±1.58e+01	1.2668e+03 ±3.51e+01	1.2411e+03 ±2.45e+01	1.3601e+03 ±2.66e+01	† 5.0940e+02 ±6.23e+00	1.2799e+03 ±4.12e+01
Brain2	2.8679e+02 ±2.12e+01	7.0154e+02 ±1.26e+01	6.3982e+02 ±1.03e+01	6.5642e+02 ±1.07e+01	6.4192e+02 ±2.01e+01	† 2.7188e+02 ±6.72e+00	6.8426e+02 ±1.88e+01
Prostate	5.6033e+02 ±5.23e+01	1.9147e+03 ±4.50e+01	1.3660e+03 ±7.34e+01	1.3550e+03 ±5.96e+01	1.3583e+03 ±4.33e+01	† 5.3659e+02 ±4.60e+00	1.4231e+03 ±1.88e+01

Compared with the other algorithms, HIER shows superior efficiency in terms of running time, mainly because of the following two reasons. First, owing to the hybrid initialization method HIER establishes a promising starting state before evolution, with an earlier exploitation in the forward areas of objective space and a vast exploration from different distributed and adaptively generated new populations across the objective space. Second, the evolutionary speed is dynamically controlled by the adaptive mutation scale of the effective reproduction method, which maintains a good balance between diversity and convergence during evolution, and thereby reproduces effective and valid offspring with smaller numbers of selected features. The above two reasons help to improve the HV, MCE, and NSF performance of HIER, which in return boosts the running time performance.

**Table 11.** Mean ranks of general running time calculated by Friedman’s test on both training and test data for each algorithm, with best ranks marked in gray.

Metric	Data	HIER	NSGA-II	MOEA/D	HypE	MOEA/HD	SparseEA	DAEA
Time	Train	1.1975	4.5800	3.4475	4.7475	5.5650	6.4850	1.9775
	Test	1.2138	4.5250	3.6162	5.0713	5.3213	6.0563	2.1963

## 6. Conclusions

This paper proposes an evolutionary algorithm based on hybrid initialization and effective reproduction, termed HIER, specifically addressing bi-objective high-dimensional feature selection in classification, which challenges the search ability of MOEAs in exploring and exploiting the sparse and large-scale decision space. In HIER, the initialization process is improved by truncating the adaptively distributed extra initial populations into a promising hybrid one, while the reproduction process is enhanced by introducing the valid crossover operation and a dynamically set mutation scale. As observed and analyzed from the empirical results, it is suggested that HIER shows significant performance advantages over other algorithms it is compared to in terms of all the metrics on most of the tested datasets. Furthermore, the nondominated solution distribution analyses also support the diversity and convergence advantages of HIER in the objective space, while the component contribution analyses implies that each essential component of HIER can separately have a positive effect and the combination of them makes the greatest contribution. Finally, the computational complexity of HIER is comprehensively studied by considering both theoretical and practical scenarios in terms of the estimated greatest time complexity and the experimental running time, which proves that HIER is not time-consuming theoretically and has an acceptable running time cost.

In our future work, it is planned to study the application of adopting the proposed hybrid initialization and effective reproduction ideas to solve more kinds of complex discrete large-scale optimization problems with binary coding, such as network construction, pattern mining, task offloading, and community detection. Moreover, the proposed algorithm HIER can also be applied to some practical cases such as medical diagnoses, which normally have a large-scale level of features and the cost of obtaining those features is often quite expensive, such as performing a pathological examination or performing nuclear magnetic resonance. In such cases, selecting a subset of features with as small a size as possible is vitally important for both patients and doctors; here, HIER could make use of its efficient and effective search ability on high-dimensional datasets. Finally, for the methodology part, there are also two potential improvements that could be added as part of our future work, shown as follows. First, the total number of extra initial populations, i.e.,  $K$  in line 2 of Algorithm 2, adaptively generated for composing the final hybrid one, could be more adaptively and delicately set. Second, the dynamically set mutation scale  $s$  in line 11 of Algorithm 4 could also be more adaptively and delicately controlled.

**Author Contributions:** Conceptualization, H.X.; Data curation, T.Y. and Y.L.; Formal analysis, H.W.; Funding acquisition, H.X., H.W., T.Y., Y.L. and Y.X.; Investigation, C.H., H.W. and Y.X.; Methodology, H.X. and Y.X.; Project administration, H.X.; Resources, C.H. and Y.L.; Software, H.X. and T.Y.; Supervision, H.X.; Validation, H.X. and C.H.; Writing—original draft, H.X.; Writing—review & editing, H.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant numbers 62103209 and 62276146, by the Natural Science Foundation of Fujian Province, grant numbers 2020J05213, 2023J011009, and 2023J011015, by the Scientific Research Project of Putian Science and Technology Bureau, grant number 2021ZP07, by the Research Project of Fujian Provincial Department of Education, grant number JAT190594, by the Startup Fund for Advanced Talents of Putian University, grant number 2019002, and by Research Projects of Putian University, grant number JG202306.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** <https://archive.ics.uci.edu/> (accessed on 9 February 2024)

**Acknowledgments:** This work was supported in part by the National Natural Science Foundation of China under grants 62103209 and 62276146, by the Natural Science Foundation of Fujian Province under grants 2020J05213, 2023J011009, and 2023J011015, by the Scientific Research Project of Putian Science and Technology Bureau under grant 2021ZP07, by the Research Project of Fujian Provincial Department of Education under grant JAT190594, by the Startup Fund for Advanced Talents of Putian University under grant 2019002, and by the Research Projects of Putian University under grant JG202306.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Eiben, A.E.; Smith, J.E. What is an evolutionary algorithm? In *Introduction to Evolutionary Computing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 25–48.
2. Coello, C.A.C.; Lamont, G.B.; Van Veldhuizen, D.A. *Evolutionary Algorithms for Solving Multi-Objective Problems*; Springer: New York, NY, USA, 2007; Volume 5.
3. Zhou, A.; Qu, B.Y.; Li, H.; Zhao, S.Z.; Suganthan, P.N.; Zhang, Q. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm Evol. Comput.* **2011**, *1*, 32–49. [[CrossRef](#)]
4. Holland, J.H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*; MIT Press: Cambridge, MA, USA, 1992.
5. Srinivas, N.; Deb, K. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evol. Comput.* **1994**, *2*, 221–248. [[CrossRef](#)]
6. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
7. Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [[CrossRef](#)]
8. Jain, H.; Deb, K. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point Based Nondominated Sorting Approach, Part II: Handling Constraints and Extending to an Adaptive Approach. *IEEE Trans. Evol. Comput.* **2014**, *18*, 602–622. [[CrossRef](#)]
9. Yuan, Y.; Xu, H.; Wang, B.; Yao, X. A New Dominance Relation-Based Evolutionary Algorithm for Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2016**, *20*, 16–37. [[CrossRef](#)]
10. Tian, Y.; Cheng, R.; Zhang, X.; Su, Y.; Jin, Y. A Strengthened Dominance Relation Considering Convergence and Diversity for Evolutionary Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2019**, *23*, 331–345. [[CrossRef](#)]
11. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [[CrossRef](#)]
12. Li, H.; Zhang, Q. Multiobjective Optimization Problems With Complicated Pareto Sets, MOEA/D and NSGA-II. *IEEE Trans. Evol. Comput.* **2009**, *13*, 284–302. [[CrossRef](#)]
13. Li, K.; Zhang, Q.; Kwong, S.; Li, M.; Wang, R. Stable Matching-Based Selection in Evolutionary Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **2014**, *18*, 909–923. [[CrossRef](#)]
14. Xu, H.; Zeng, W.; Zhang, D.; Zeng, X. MOEA/HD: A Multiobjective Evolutionary Algorithm Based on Hierarchical Decomposition. *IEEE Trans. Cybern.* **2019**, *49*, 517–526. [[CrossRef](#)]

15. Bader, J.; Zitzler, E. HypE: An Algorithm for Fast Hypervolume-Based Many-Objective Optimization. *Evol. Comput.* **2011**, *19*, 45–76. [[CrossRef](#)]
16. Xu, H.; Zeng, W.; Zeng, X.; Yen, G.G. An Evolutionary Algorithm Based on Minkowski Distance for Many-Objective Optimization. *IEEE Trans. Cybern.* **2019**, *49*, 3968–3979. [[CrossRef](#)]
17. Xu, H.; Zeng, W.; Zeng, X.; Yen, G.G. A Polar-Metric-Based Evolutionary Algorithm. *IEEE Trans. Cybern.* **2021**, *51*, 3429–3440. [[CrossRef](#)]
18. Liang, Z.; Luo, T.; Hu, K.; Ma, X.; Zhu, Z. An Indicator-Based Many-Objective Evolutionary Algorithm With Boundary Protection. *IEEE Trans. Cybern.* **2021**, *51*, 4553–4566. [[CrossRef](#)]
19. Wang, H.; Jin, Y.; Sun, C.; Doherty, J. Offline data-driven evolutionary optimization using selective surrogate ensembles. *IEEE Trans. Evol. Comput.* **2018**, *23*, 203–216. [[CrossRef](#)]
20. Lin, Q.; Wu, X.; Ma, L.; Li, J.; Gong, M.; Coello, C.A.C. An Ensemble Surrogate-Based Framework for Expensive Multiobjective Evolutionary Optimization. *IEEE Trans. Evol. Comput.* **2022**, *26*, 631–645. [[CrossRef](#)]
21. Sonoda, T.; Nakata, M. Multiple Classifiers-Assisted Evolutionary Algorithm Based on Decomposition for High-Dimensional Multiobjective Problems. *IEEE Trans. Evol. Comput.* **2022**, *26*, 1581–1595. [[CrossRef](#)]
22. Goh, C.K.; Tan, K.C. A competitive-cooperative coevolutionary paradigm for dynamic multiobjective optimization. *IEEE Trans. Evol. Comput.* **2008**, *13*, 103–127.
23. Zhan, Z.H.; Li, J.; Cao, J.; Zhang, J.; Chung, H.S.H.; Shi, Y.H. Multiple Populations for Multiple Objectives: A Coevolutionary Technique for Solving Multiobjective Optimization Problems. *IEEE Trans. Cybern.* **2013**, *43*, 445–463. [[CrossRef](#)] [[PubMed](#)]
24. Ma, X.; Li, X.; Zhang, Q.; Tang, K.; Liang, Z.; Xie, W.; Zhu, Z. A survey on cooperative co-evolutionary algorithms. *IEEE Trans. Evol. Comput.* **2018**, *23*, 421–441. [[CrossRef](#)]
25. Da, B.; Gupta, A.; Ong, Y.S.; Feng, L. Evolutionary multitasking across single and multi-objective formulations for improved problem solving. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 1695–1701.
26. Gupta, A.; Ong, Y.S.; Feng, L.; Tan, K.C. Multiobjective Multifactorial Optimization in Evolutionary Multitasking. *IEEE Trans. Cybern.* **2017**, *47*, 1652–1665. [[CrossRef](#)]
27. Rauniyar, A.; Nath, R.; Muhuri, P.K. Multi-factorial evolutionary algorithm based novel solution approach for multi-objective pollution-routing problem. *Comput. Ind. Eng.* **2019**, *130*, 757–771. [[CrossRef](#)]
28. Cai, H.; Lin, Q.; Liu, H.; Li, X.; Xiao, H. A Multi-Objective Optimisation Mathematical Model with Constraints Conducive to the Healthy Rhythm for Lighting Control Strategy. *Mathematics* **2022**, *10*, 3471. [[CrossRef](#)]
29. Alshammari, N.F.; Samy, M.M.; Barakat, S. Comprehensive Analysis of Multi-Objective Optimization Algorithms for Sustainable Hybrid Electric Vehicle Charging Systems. *Mathematics* **2023**, *11*, 1741. [[CrossRef](#)]
30. Zhu, W.; Li, H.; Wei, W. A Two-Stage Multi-Objective Evolutionary Algorithm for Community Detection in Complex Networks. *Mathematics* **2023**, *11*, 2702. [[CrossRef](#)]
31. Chalabi, N.E.; Attia, A.; Alnowibet, K.A.; Zawbaa, H.M.; Masri, H.; Mohamed, A.W. A Multi-Objective Gaining-Sharing Knowledge-Based Optimization Algorithm for Solving Engineering Problems. *Mathematics* **2023**, *11*, 3092. [[CrossRef](#)]
32. Cao, F.; Tang, Z.; Zhu, C.; Zhao, X. An Efficient Hybrid Multi-Objective Optimization Method Coupling Global Evolutionary and Local Gradient Searches for Solving Aerodynamic Optimization Problems. *Mathematics* **2023**, *11*, 3844. [[CrossRef](#)]
33. Gao, C.; Yin, Z.; Wang, Z.; Li, X.; Li, X. Multilayer Network Community Detection: A Novel Multi-Objective Evolutionary Algorithm Based on Consensus Prior Information [Feature]. *IEEE Comput. Intell. Mag.* **2023**, *18*, 46–59. [[CrossRef](#)]
34. Xue, Y.; Chen, C.; Słowik, A. Neural Architecture Search Based on a Multi-Objective Evolutionary Algorithm with Probability Stack. *IEEE Trans. Evol. Comput.* **2023**, *27*, 778–786. [[CrossRef](#)]
35. Long, S.; Zhang, Y.; Deng, Q.; Pei, T.; Ouyang, J.; Xia, Z. An Efficient Task Offloading Approach Based on Multi-Objective Evolutionary Algorithm in Cloud-Edge Collaborative Environment. *IEEE Trans. Netw. Sci. Eng.* **2023**, *10*, 645–657. [[CrossRef](#)]
36. Zhang, Z.; Ma, S.; Jiang, X. Research on Multi-Objective Multi-Robot Task Allocation by Lin-Kernighan-Helsgaun Guided Evolutionary Algorithms. *Mathematics* **2022**, *10*, 4714. [[CrossRef](#)]
37. Nguyen, B.H.; Xue, B.; Andreae, P.; Ishibuchi, H.; Zhang, M. Multiple Reference Points-Based Decomposition for Multiobjective Feature Selection in Classification: Static and Dynamic Mechanisms. *IEEE Trans. Evol. Comput.* **2020**, *24*, 170–184. [[CrossRef](#)]
38. Luo, J.; Zhou, D.; Jiang, L.; Ma, H. A particle swarm optimization based multiobjective memetic algorithm for high-dimensional feature selection. *Memetic Comput.* **2022**, *14*, 77–93. [[CrossRef](#)]
39. Gong, Y.; Zhou, J.; Wu, Q.; Zhou, M.; Wen, J. A Length-Adaptive Non-Dominated Sorting Genetic Algorithm for Bi-Objective High-Dimensional Feature Selection. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1834–1844. [[CrossRef](#)]
40. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [[CrossRef](#)]
41. Jiao, R.; Nguyen, B.H.; Xue, B.; Zhang, M. A Survey on Evolutionary Multiobjective Feature Selection in Classification: Approaches, Applications, and Challenges. *IEEE Trans. Evol. Comput.* **2023**, *1*, Early Access. [[CrossRef](#)]
42. Chen, K.; Xue, B.; Zhang, M.; Zhou, F. An Evolutionary Multitasking-Based Feature Selection Method for High-Dimensional Classification. *IEEE Trans. Cybern.* **2022**, *52*, 7172–7186. [[CrossRef](#)]

43. Bai, H.; Cheng, R.; Yazdani, D.; Tan, K.C.; Jin, Y. Evolutionary Large-Scale Dynamic Optimization Using Bilevel Variable Grouping. *IEEE Trans. Cybern.* **2022**, 1–14. [[CrossRef](#)]
44. He, C.; Cheng, R.; Tian, Y.; Zhang, X.; Tan, K.C.; Jin, Y. Paired Offspring Generation for Constrained Large-Scale Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **2021**, *25*, 448–462. [[CrossRef](#)]
45. Xue, B.; Zhang, M.; Browne, W.N. Novel initialisation and updating mechanisms in PSO for feature selection in classification. In Proceedings of the Applications of Evolutionary Computation: 16th European Conference, EvoApplications 2013, Vienna, Austria, 3–5 April 2013; Proceedings 16; Springer: Berlin/Heidelberg, Germany, 2013; pp. 428–438.
46. Xu, H.; Xue, B.; Zhang, M. Segmented Initialization and Offspring Modification in Evolutionary Algorithms for Bi-Objective Feature Selection. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference, New York, NY, USA, 8–12 July 2020; GECCO '20, pp. 444–452.
47. Xu, H.; Xue, B.; Zhang, M. A Duplication Analysis-Based Evolutionary Algorithm for Biobjective Feature Selection. *IEEE Trans. Evol. Comput.* **2021**, *25*, 205–218. [[CrossRef](#)]
48. Ma, X.; Liu, F.; Qi, Y.; Wang, X.; Li, L.; Jiao, L.; Yin, M.; Gong, M. A Multiobjective Evolutionary Algorithm Based on Decision Variable Analyses for Multiobjective Optimization Problems With Large-Scale Variables. *IEEE Trans. Evol. Comput.* **2016**, *20*, 275–298. [[CrossRef](#)]
49. Zhang, X.; Tian, Y.; Cheng, R.; Jin, Y. A Decision Variable Clustering-Based Evolutionary Algorithm for Large-Scale Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2018**, *22*, 97–112. [[CrossRef](#)]
50. Zille, H.; Mostaghim, S. Comparison study of large-scale optimisation techniques on the LSMOP benchmark functions. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–8.
51. Tian, Y.; Shao, S.; Xie, G.; Zhang, X. A multi-granularity clustering based evolutionary algorithm for large-scale sparse multi-objective optimization. *Swarm Evol. Comput.* **2024**, *84*, 101453. [[CrossRef](#)]
52. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–45. [[CrossRef](#)]
53. De La Iglesia, B. Evolutionary computation for feature selection in classification problems. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 381–407. [[CrossRef](#)]
54. Xue, B.; Zhang, M.; Browne, W.N.; Yao, X. A survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comput.* **2015**, *20*, 606–626. [[CrossRef](#)]
55. Dokeroglu, T.; Deniz, A.; Kiziloz, H.E. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing* **2022**, *494*, 269–296. [[CrossRef](#)]
56. Mukhopadhyay, A.; Maulik, U. An SVM-wrapped multiobjective evolutionary feature selection approach for identifying cancer-microRNA markers. *IEEE Trans. Nanobiosci.* **2013**, *12*, 275–281. [[CrossRef](#)]
57. Vignolo, L.D.; Milone, D.H.; Scharcanski, J. Feature selection for face recognition based on multi-objective evolutionary wrappers. *Expert Syst. Appl.* **2013**, *40*, 5077–5084. [[CrossRef](#)]
58. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
59. Lazar, C.; Taminau, J.; Meganck, S.; Steenhoff, D.; Coletta, A.; Molter, C.; de Schaetzen, V.; Duque, R.; Bersini, H.; Nowe, A. A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB)* **2012**, *9*, 1106–1119. [[CrossRef](#)] [[PubMed](#)]
60. Xue, B.; Cervante, L.; Shang, L.; Browne, W.N.; Zhang, M. Multi-objective evolutionary algorithms for filter based feature selection in classification. *Int. J. Artif. Intell. Tools* **2013**, *22*, 1350024. [[CrossRef](#)]
61. Chen, K.; Xue, B.; Zhang, M.; Zhou, F. Evolutionary Multitasking for Feature Selection in High-Dimensional Classification via Particle Swarm Optimization. *IEEE Trans. Evol. Comput.* **2022**, *26*, 446–460. [[CrossRef](#)]
62. Xue, B.; Zhang, M.; Browne, W.N. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Appl. Soft Comput.* **2014**, *18*, 261–276. [[CrossRef](#)]
63. Tian, Y.; Zhang, X.; Wang, C.; Jin, Y. An Evolutionary Algorithm for Large-Scale Sparse Multiobjective Optimization Problems. *IEEE Trans. Evol. Comput.* **2020**, *24*, 380–393. [[CrossRef](#)]
64. Cheng, F.; Cui, J.; Wang, Q.; Zhang, L. A Variable Granularity Search-Based Multiobjective Feature Selection Algorithm for High-Dimensional Data Classification. *IEEE Trans. Evol. Comput.* **2023**, *27*, 266–280. [[CrossRef](#)]
65. Espinosa, R.; Jimenez, F.; Palma, J. Surrogate-Assisted and Filter-Based Multiobjective Evolutionary Feature Selection for Deep Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–15, Early Access. [[CrossRef](#)]
66. Cheng, F.; Chu, F.; Xu, Y.; Zhang, L. A Steering-Matrix-Based Multiobjective Evolutionary Algorithm for High-Dimensional Feature Selection. *IEEE Trans. Cybern.* **2022**, *52*, 9695–9708. [[CrossRef](#)]
67. Dua, D.; Graff, C. Dua, D.; Graff, C. UCI Machine Learning Repository. 2017. Available online: <https://archive.ics.uci.edu/> (accessed on 9 February 2024).
68. While, L.; Hingston, P.; Barone, L.; Huband, S. A faster algorithm for calculating Hypervolume. *IEEE Trans. Evol. Comput.* **2006**, *10*, 29–38. [[CrossRef](#)]

- 
69. Tian, Y.; Cheng, R.; Zhang, X.; Jin, Y. PlatEMO: A MATLAB Platform for Evolutionary Multi-Objective Optimization. *IEEE Comput. Intell. Mag.* **2017**, *12*, 73–87. [[CrossRef](#)]
  70. Tran, B.; Xue, B.; Zhang, M.; Nguyen, S. Investigation on particle swarm optimisation for feature selection on high-dimensional data: Local search and selection bias. *Connect. Sci.* **2016**, *28*, 270–294. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.