

Article

# A Multi-Task Decomposition-Based Evolutionary Algorithm for Tackling High-Dimensional Bi-Objective Feature Selection

Hang Xu <sup>1,\*</sup> , Chaohui Huang <sup>1</sup>, Jianbing Lin <sup>1</sup>, Min Lin <sup>1</sup>, Huahui Zhang <sup>2</sup> and Rongbin Xu <sup>1</sup>

<sup>1</sup> School of Mechanical, Electrical & Information Engineering, Putian University, Putian 351100, China; jxlfhch@ptu.edu.cn (C.H.); linjb@ptu.edu.cn (J.L.); lmmindy@ptu.edu.cn (M.L.); xurongbing@ptu.edu.cn (R.X.)

<sup>2</sup> New Engineering Industry College, Putian University, Putian 351100, China; zhanghuahui@ptu.edu.cn

\* Correspondence: xuhang@ptu.edu.cn

**Abstract:** Evolutionary algorithms have been widely applied for solving multi-objective optimization problems, while the feature selection in classification can also be treated as a discrete bi-objective optimization problem if attempting to minimize both the classification error and the ratio of selected features. However, traditional multi-objective evolutionary algorithms (MOEAs) may have drawbacks for tackling large-scale feature selection, due to the curse of dimensionality in the decision space. Therefore, in this paper, we concentrated on designing an multi-task decomposition-based evolutionary algorithm (abbreviated as MTDEA), especially for handling high-dimensional bi-objective feature selection in classification. To be more specific, multiple subpopulations related to different evolutionary tasks are separately initialized and then adaptively merged into a single integrated population during the evolution. Moreover, the ideal points for these multi-task subpopulations are dynamically adjusted every generation, in order to achieve different search preferences and evolutionary directions. In the experiments, the proposed MTDEA was compared with seven state-of-the-art MOEAs on 20 high-dimensional classification datasets in terms of three performance indicators, along with using comprehensive Wilcoxon and Friedman tests. It was found that the MTDEA performed the best on most datasets, with a significantly better search ability and promising efficiency.

**Keywords:** bi-objective feature selection; evolutionary algorithm; high-dimensional classification data; multi-task decomposition

**MSC:** 68W50



**Citation:** Xu, H.; Huang, C.; Lin, J.; Lin, M.; Zhang, H.; Xu, R. A Multi-Task Decomposition-Based Evolutionary Algorithm for Tackling High-Dimensional Bi-Objective Feature Selection. *Mathematics* **2024**, *12*, 1178. <https://doi.org/10.3390/math12081178>

Academic Editors: Lev Kazakovtsev and Predrag S. Stanimirovic

Received: 18 March 2024

Revised: 6 April 2024

Accepted: 12 April 2024

Published: 14 April 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 widely-used as common tools to solve multi-objective optimization problems (MOPs) [2] during the past decades. In fact, when the number of objectives to be optimized is more than one, they are normally contradictory to each other, and therefore multi-objective evolutionary algorithms (MOEAs) [3] are used for finding a set of nondominated solutions. Compared with other meta-heuristics [4], MOEAs have the advantages of a population-based search mode and no need of domain knowledge. Thus, a huge variety of MOEAs have been proposed all over the world and can be roughly divided into the following categories: dominance-based MOEAs [5–8], decomposition-based MOEAs [9–13], indicator-based MOEAs [14–17], surrogate-based MOEAs [18–20], cooperative coevolutionary MOEAs [21–23], multi-task MOEAs [24–26], and so on. There are also many other kinds of excellent MOEAs [27–29], including the novel multi-objective particle swarm optimization algorithm proposed by Leung et al. [30], which adopted a hybrid global leader selection strategy with two leaders: one for exploration and the other for exploitation. Moreover, MOEAs have also been used to solve many real-world optimization problems [31–33], such as system control [34,35], community detection [36,37], network construction [38–40], task allocation [41,42], and feature selection [43,44]. Generally speaking,

feature selection is normally used to select useful feature subsets for classification [45], while the bi-objective feature selection problem usually seeks to minimize both the classification error and the number of selected features [46].

However, due to the curse of dimensionality in the decision space as the number of features expands to large scale, traditional MOEAs are likely to encounter setbacks when tackling bi-objective feature selection. In fact, the large-scale multi-objective optimization problem (LSMOP) [47] remains challenging for most MOEAs, although many MOEAs have attempted to solve it. For example, Ma et al. [48] and Zhang et al. [49] focused on analyzing different kinds of decision variables before evolution, while Bai et al. [50] and Yang et al. [51] proposed decision-variable-based MOEAs for solving continuous LSMOPs. However, not all large-scale MOEAs can be used for discrete optimization, especially for feature selection that has to face the complex interrelationships among features and a large number of feasible feature combinations. One possible approach is to bring in the multi-task framework [52,53] for cooperative evolution, as this can achieve an overall effect of learning by transferring genetic knowledge between different evolutionary tasks. Many MOEAs have already adopted multi-tasking [54–56] and it seems quite promising for decomposition-based MOEAs to be integrated with the multi-task framework, due to the intrinsic parallelism characteristics inside the weight-vector-related aggregation functions [43].

In fact, the integration of multi-task mechanisms with the decomposition-based evolutionary approaches is theoretically appropriate, mainly because of the following three reasons. First, both multi-task mechanisms and decomposition-based approaches share the same philosophy: to decompose a complex problem (task) into multiple interrelated or independent subproblems (tasks), in order to solve the problem (task) more efficiently and comprehensively. Second, each subproblem in a decomposition-based MOEA can be treated as an independent search task, whose search behaviors can be adjusted by changing the direction of the related weight vector. Third, the global ideal point for a decomposition based MOEA controls the general search areas, while different global ideal points would make the algorithm display quite different search preferences, which can be combined with the multi-task mechanism, allowing different global ideal points to represent different evolutionary search tasks. The abovementioned third point is exactly what is made use of in this paper and also acted as our major motivation.

Therefore, in this paper, we aimed to design an adaptive decomposition-based MOEA framework combined with a dynamic multi-task mechanism, named a multi-task decomposition-based evolutionary algorithm (MTDEA), for tackling the large-scale bi-objective evolutionary feature selection problem in high-dimensional classification datasets. To be more specific, we attempted to organically combine a multi-task mechanism with the decomposition-based approach to self-adaptively adjust the search preferences and evolutionary directions for each task-related subpopulation, in order to improve both the optimization and classification performance, and to achieve cooperative evolution within the population for better diversity and convergence. Overall, our major contributions can be summarized as follows:

- First, a dynamic multi-task mechanism is designed and combined with the decomposition-based MOEA framework, which assigns multiple evolutionary search tasks for different subpopulations within the entire population and then conditionally merges them into a single task or an integrated population as the evolutionary process goes on, for tackling the large-scale bi-objective feature selection in a more effective way.
- Second, an adaptive decomposition-based MOEA framework is set up, which cooperates with the above multi-task mechanism via adaptively adjusting the ideal point for each subpopulation related to different tasks, so that each task has distinct search biases and focuses its computational resources on searching more productive areas in the objective space.
- Third, a series of comprehensive studies were conducted in experiments to analyze the optimization and classification performance of the proposed MTDEA algorithm

against other state-of-the-art MOEAs, in terms of multiple indicators and using a variety of high-dimensional classification datasets.

The remainder of this paper is organized as follows: First of all, the related works are introduced in Section 2. Then, the proposed algorithm MTDEA is comprehensively illustrated in Section 3. The experiment setups are given in Section 4, while the empirical results are studied in Section 5. Last, the conclusions are given in Section 6.

## 2. Related Works

### 2.1. Bi-Objective Feature Selection Problem

Generally, a multi-objective feature selection problem [57] can be defined as a multi-objective optimization problem, 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 total number of objectives to be optimized, and  $D$  is the full number of features that can be selected, i.e., also the dimensionality of the decision space. In this paper,  $M$  is set to 2, and  $\mathbf{F}(\mathbf{x})$  is the objective vector of  $\mathbf{x}$ , while  $f_i(\mathbf{x})$  is the objective value in the  $f_i$  direction.  $\mathbf{x} = (x_1, x_2, \dots, x_D)$  is the decision vector of a certain solution, while the value of 1 means selecting that feature and 0 means not. In addition, the first objective function  $f_1(\mathbf{x})$  can be further defined as follows:

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

where its function values are between 0 and 1, i.e.,  $\in \{0, 1/D, 2/D, \dots, 1\}$ , denoted as the rate of currently selected features. Moreover, the second objective function  $f_2(\mathbf{x})$  denotes the resultant classification error rate for the previously selected features in  $\mathbf{x}$ , whose function values also discretely range from 0 to 1. Given the results of  $TP$  (true positive),  $TN$  (true negative),  $FP$  (false positive), and  $FN$  (false negative),  $f_2(\mathbf{x})$  can be formalized as follows:

$$f_2(\mathbf{x}) = \frac{FP + FN}{TP + TN + FP + FN} \quad (3)$$

### 2.2. Evolutionary Feature Selection Methods

In the past decades, evolutionary feature selection [58] has been roughly categorized into wrapper-based or the filter-based approaches [59,60]. Generally, a wrapper-based approach [61,62] uses a classification model, like SVM (support vector machine) or KNN (K-nearest neighbor) [63], as a “black box” to evaluate the classification accuracy, while the filter-based approach [64,65] is independent of any classifier and ignores the classification results of the currently selected features. Thus, a wrapper-based approach is normally more accurate but may consume a higher computational cost [66–68]. In this paper, the wrapper-based approach is adopted for bi-objective feature selection, and in fact, many other such MOEAs have been proposed in the last few years [69]. For example, in 2020, Tian et al. [70] proposed a large-scale MOEA framework, named SparseEA, based on pre-analyzing each feature’s classification performance, which however, would consume a large number of objective function evaluations for high-dimensional datasets. Subsequently, Xu et al. [71] proposed a duplication-analysis-based MOEA, named DAEA, with an efficient reproduction method to generate more valid and diverse offspring, but its tested feature dimensionality has not yet reached 10000. Following the idea of DAEA, Jiao et al. [72] modified the solution duplication handling method and further designed a problem reformulation mechanism, named PRDH, whose applicability across other MOEA frameworks remains unconfirmed. In 2022, Cheng et al. [73] proposed a steering-matrix-based algorithm for high-dimensional classification, named SM-MOEA, which was quiet efficient in tackling large-scale optimization, but its generalization ability still needs to be verified on more datasets, as only 12 datasets were tested in their work.

### 2.3. Decomposition-Based MOEA Approaches

Since its first introduction in the famous algorithm MOEA/D [74], a great number of decomposition-based MOEAs have been proposed all around the world [75–79]. To be more specific, a decomposition-based approach uses a series of uniformly distributed weight vectors as the aggregation functions to decompose a complex MOP into a set of simpler single-objective optimization problems that are related to each weight vector one by one. Generally speaking, there are three widely-used aggregation functions for decomposition, i.e., the weighted sum (WS) approach [80], the Tchebycheff (TCH) approach [81], and the penalty-based boundary intersection (PBI) approach [82]. The TCH approach is further introduced here, as it was the original decomposition approach adopted in MOEA/D and is also one of the comparison algorithms used in our later experiment. In detail, the TCH approach can be formally defined as follows:

$$\begin{aligned} \text{minimize } g^{tch}(\mathbf{x}|\mathbf{w}, \mathbf{z}) &= \max_{1 \leq i \leq M} \{|f_i(\mathbf{x}) - z_i| * w_i\} \\ \text{subject to } \mathbf{x} &= (x_1, x_2, \dots, x_D), x_i \in \{0, 1\} \end{aligned} \quad (4)$$

where  $\mathbf{w}$  and  $\mathbf{z}$  respectively denote the current weight vector and the ideal point, while the selection principle is based on a *Max-Min* mechanism that first calculates the maximum values and then selects the minimum one.

## 3. Proposed Algorithm

In this section, we first introduce the general framework of the proposed MTDEA, and then further illustrate its essential components, i.e., initialization, reproduction, environmental selection, and task merging processes.

### 3.1. General Framework

The general framework of the MTDEA is shown in Algorithm 1, where the population size  $N$  and the decision space dimension  $D$  (i.e., the total number of features) are input as the primary parameters. In brief, its general framework is similar to traditional MOEAs but brings in multi-task factors and adds a task merging process at the end of every evolutionary generation. In detail, each solution in the population is related with a task number, which are initialized together at the beginning and correlated closely throughout the subsequent evolution. The implicit relationship between solutions and tasks is dynamically maintained by an external array with two rows and  $N$  columns, where the first row stores the solution indexes and the second row stores the corresponding task numbers. More specific relationship and constructions of the initial tasks and population are shown in Figure 1, which will be further illustrated in Section 3.2. As a result, the reproduction process of the MTDEA also uses a multi-task mechanism for cooperatively exchanging genetic information between the different tasks, which will be further illustrated in Section 3.3. Moreover, the duplicated decision vectors should first be removed before separately truncating each task-related subpopulation from the current population and offspring, while the detailed selection process based on decomposition will be shown in Section 3.4. At last, the task merging process dynamically decides whether to merge each two tasks or not and will eventually integrate all the tasks into a single task, details of which will be given in Section 3.5.

### 3.2. Initialization Process

The initialization process of the MTDEA is shown in Algorithm 2, which also invokes Algorithm 3. During initialization, there are a total of three subpopulations to be generated related with different task numbers. The implicit inner relationship between the solutions and tasks is shown in Figure 1, assuming a population size of 100. It can be seen from Figure 1 that the first 25% of solutions are related to task-1, the median 50% to task-2, and the last 25% to task-3, with each solution corresponding one-to-one to a task number at the same index position. Thus, the whole population can be split into three subpopulations,

and the generation of a new subpopulation is shown in Algorithm 3, where the so-called distribution axis parameter is input to roughly control the solution distributions in the objective space.

---

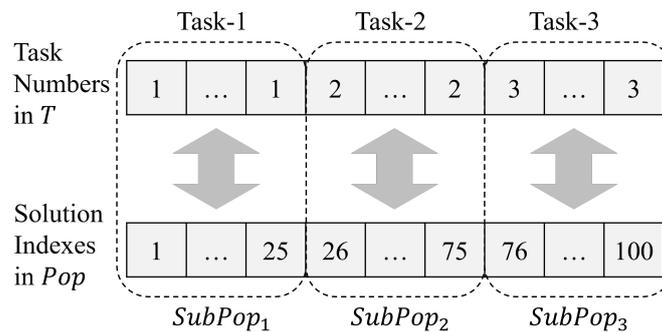
**Algorithm 1** *General Framework* ( $N, D$ )

---

**Input:** population size  $N$ , decision space dimension  $D$ ;  
**Output:** final population  $Pop$ ;

- 1:  $[Pop, T] = Initialize(N, D)$ ; // Algorithm 2
- 2: **while** termination criterion is not reached **do**
- 3:      $[Pop^*, T^*] = Reproduce(Pop, T)$ ; // Algorithm 4
- 4:     **for** each unique task  $i$  from  $T$  **do**
- 5:          $j \leftarrow$  get the quantity of task- $i$  solutions in  $Pop$   
        according to the task numbers in  $T$ ;
- 6:          $\psi \leftarrow$  get all the task- $i$  solutions in  $Pop$  and  $Pop^*$   
        according to the task numbers in  $T$  and  $T^*$ ;
- 7:          $\psi \leftarrow$  remove duplicated decision vectors in  $\psi$ ;
- 8:          $\psi = Select(\psi, i, j)$ ; // Algorithm 5
- 9:         replace all the task- $i$  solutions in  $Pop$  with  $\psi$ ;
- 10:     **end for**
- 11:     **if** more than one task exist **then**
- 12:          $T = Merge(Pop, T)$ ; // Algorithm 6
- 13:     **end if**
- 14: **end while**

---



**Figure 1.** Implicit inner relationship between the solutions and tasks with a population size of 100 and three subpopulations inside.

---

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

---

**Input:** population size  $N$ , decision space dimension  $D$ ;  
**Output:** initial population  $Pop$ , related tasks  $T$ ;

- 1:  $Tasks = [1, 2, 3]$ ; // task numbers
- 2:  $Axes = [0.25, 0.5, 0.75]$ ; // distribution axes
- 3:  $Sizes = [N/4, N/2, N/4]$ ; // subpopulation sizes
- 4:  $Pop = \emptyset, T = \emptyset$ ;
- 5: **for**  $i = 1, 2, \dots, Length(Tasks)$  **do**
- 6:      $Pop = Pop \cup NewPop(Sizes(i), D, Axes(i))$ ;
- 7:      $\psi = Ones(1, Sizes(i))$ ; // create a vector of ones
- 8:      $T = T \cup (\psi * Tasks(i))$ ;
- 9: **end for**

---

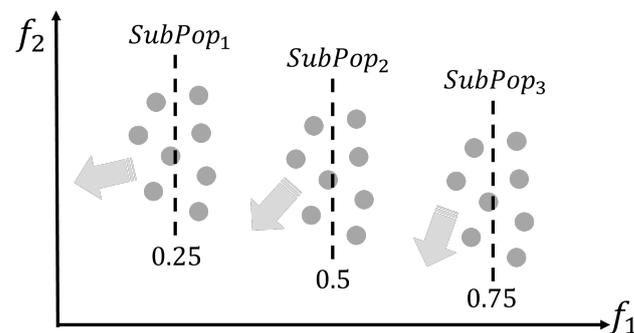
**Algorithm 3** *NewPop*( $K, D, A$ )**Input:** subpopulation size  $K$ , decision space dimension  $D$ , distribution axis  $A$ ;**Output:** new subpopulation *SubPop*;

```

1: SubPop = Zeros( $K, D$ ); // create a matrix of zeros
2: for  $i = 1, 2, \dots, K$  do
3:   for  $j = 1, 2, \dots, D$  do
4:     if  $\rho < A$  then //  $\rho$  is a random probability
5:       SubPop( $i, j$ ) = 1; // select the  $j$ th feature
6:     end if
7:   end for
8: end for

```

As can be seen from Line 4 in Algorithm 3, the distribution axis  $A$  plays a role as the probability threshold, while a larger threshold indicates a higher probability of randomly selecting this feature, which also means that the resultant decision vector is likely to select more features. Reflected in the objective space, this implies that the overall distribution of the newly generated solutions should be more backward in the  $f_1$  direction for a larger distribution axis value, and vice versa. Figure 2 gives an intuitive example of how the three subpopulations related to different tasks are likely to be distributed in the objective space. It is shown in Figure 2 that each newly generated subpopulation is assumed to be distributed around their preset axes in the  $f_1$  direction. Moreover, the gray arrows in Figure 2 point out the expected evolutionary direction of each subpopulation, corresponding to the specific search bias of each related task, which will be further illustrated in Section 3.4.



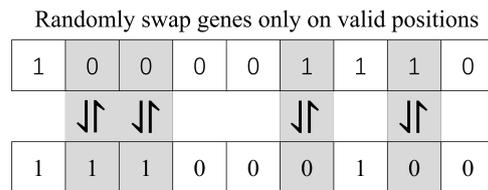
**Figure 2.** An example of how three initial subpopulations are likely to be distributed in the objective space, with different distribution axes input: 0.25, 0.5, and 0.75, respectively, for *SubPop*<sub>1</sub>, *SubPop*<sub>2</sub>, and *SubPop*<sub>3</sub>, while solutions are shown in dots and general search directions are shown in arrows.

### 3.3. Reproduction Process

The reproduction process of the MTDEA is shown in Algorithm 4, where the current population and the related task set are input. First, a total of  $N$  (i.e.,  $Length(Pop)$ ) pairs of parent solutions are randomly selected from the current population. Then, for each pair of parents, a corresponding new offspring will be generated after conducting the so-called valid crossover (from Lines 6 to 10 in Algorithm 4) and bitwise mutation (from Lines 11 to 13 in Algorithm 4) operations. It should be noted that the related task number for a new offspring is the same as its parent  $p_1$ . Moreover, the valid crossover will not occur if the two parents have different tasks and the random probability  $\rho$  fails to reach 0.5, which acts as a key parameter for controlling the genetic information transfer among different tasks (Line 6 in Algorithm 4). Here, it is set to 0.5 for the sake of a relatively uniform random selection, without loss of generality. The so-called valid crossover operation can also be further referred to in our previous work [71], a simple example of which is shown in Figure 3, where two decision vectors are drawn as parents and only the gray parts owning different decision variable values are allowed to swap genes randomly.

**Algorithm 4** *Reproduce*(Pop, T)

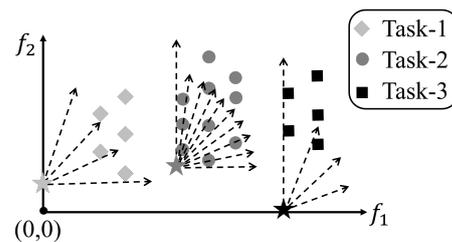
**Input:** current population *Pop*, related tasks *T*;  
**Output:** offspring solutions *Pop\**, offspring tasks *T\**;  
1: *Pairs*  $\leftarrow$  randomly select *Length*(*Pop*) pairs of solutions from *Pop* as parents; // random mating  
2: **for**  $i = 1, 2, \dots, \text{Length}(\text{Pop})$  **do**  
3:      $p_1, p_2 \leftarrow$  get the *i*th pair of parents from *Pairs*;  
4:      $t_1, t_2 \leftarrow$  get the related task numbers of  $p_1, p_2$  according to the corresponding indexes in *T*;  
5:      $T^*(i) = t_1$ ; // assign offspring task number  
6:     **if**  $(t_1 = t_2) \parallel (t_1 \neq t_2 \cap \rho < 0.5)$  **then**  
7:          $j \leftarrow$  find the indexes of different decision variable values between  $p_1$  and  $p_2$ ;  
8:          $j \leftarrow$  uniformly randomly select indexes from  $j$ ;  
9:          $p_1(j) = p_2(j)$ ; // valid crossover  
10:     **end if**  
11:      $\rho \leftarrow$  get a set of *Length*( $p_1$ ) random probabilities;  
12:      $j \leftarrow$  get the indexes satisfying  $\rho < 1/\text{Length}(p_1)$ ;  
13:      $p_1(j) = \neg p_1(j)$ ; // bitwise mutation  
14:      $\text{Pop}^*(i) = p_1$ ; // get the new offspring  
15: **end for**



**Figure 3.** A simple example of the adopted valid crossover operation between two decision vectors during reproduction.

3.4. Environmental Selection Process

The environmental selection process of the MTDEA is shown in Algorithm 5, where a certain task-related union population (subpopulation + offspring) is input for further truncation into the required size of subpopulation. The whole selection process is based on decomposition with a set of uniformly distributed weight vectors in the objective space, which selects the best qualified solution for each weight-vector-related aggregation function (the so-called subproblem) one by one, until the required number of solutions have been selected. Combined with the multi-task mechanism, the major differences of the proposed MTDEA from the traditional-decomposition based MOEAs are that it uses different ideal points (i.e., *Zmin* in Algorithm 5) for different related tasks (also shown in Lines 5 to 11 in Algorithm 5) and uses a normalized modified inverse Tchebycheff (abbreviated as I-TCH) approach as the aggregation function for decomposition. A simple example of how to adaptively adjust the ideal point is given in Figure 4 for a more intuitive explanation.



**Figure 4.** An example of how the ideal point *Zmin* (drawn as stars in this figure) is adaptively adjusted according to the related task numbers for each subpopulation: task-1, task-2, and task-3, while weight vectors are roughly shown in arrows.

**Algorithm 5** *Select*( $\psi, \tau, K$ )

---

**Input:** union subpopulation  $\psi$ , related task number  $\tau$ , required subpopulation size  $K$ ;  
**Output:** required subpopulation  $\psi$ ;

- 1:  $W \leftarrow$  get a set of  $K$  normalized uniformly distributed weight vectors in the objective space for decomposition;
- 2:  $S \leftarrow$  get a  $K$ -dimensional vector of false boolean values;
- 3:  $Zmin \leftarrow$  get the best ideal point from  $\psi$ ;
- 4:  $Zmax \leftarrow$  get the worst ideal point from  $\psi$ ;
- 5: **if**  $\tau == 1$  **then**
- 6:     set the  $f_1$  objective value of  $Zmin$  to zero;
- 7: **else if**  $\tau == 2$  **then**
- 8:     normalize objective values of  $\psi$  by  $Zmax-Zmin$ ;
- 9: **else if**  $\tau == 3$  **then**
- 10:    set the  $f_2$  objective value of  $Zmin$  to zero;
- 11: **end if**
- 12: **for**  $i = 1, 2, \dots, K$  **do**
- 13:     $j \leftarrow$  get indexes of all the false boolean values in  $S$ ;
- 14:     $fit \leftarrow$  get the fitness of all the solutions in  $\psi(j)$  by Equation (5), along with  $Zmin$  and  $W(i)$  input;
- 15:     $best \leftarrow$  get the index of the smallest value in  $fit$ ;
- 16:    set the boolean value of  $S(j(best))$  to be true;
- 17: **end for**
- 18:  $\psi = \psi(S)$ ; // get the final selected solutions

---

In Figure 4, the ideal point  $Zmin$  is drawn as stars for the three different task-related subpopulations. As previously shown in Lines 6 and 10 of Algorithm 5,  $Zmin$  for the task-1 subpopulation is transformed by changing its  $f_1$  objective value to zero, thereby moving the ideal point to the  $f_2$  axis, as shown in Figure 4. The same goes for the ideal point of the task-3 subpopulation, which is moved to the  $f_1$  axis by contrast, while that of the task-2 subpopulation has no transformation. In this way, the three different task-related subpopulations can have distinct evolutionary search behaviors, with their exploring preferences adjusted in different directions in the objective space, while task-1 prefers to search more of the sparse areas in the  $f_1$  direction, task-3 prefers to search more of the sparse areas in the  $f_2$  direction, and task-2 makes a balanced compromise between the other two tasks. It should also be noted that the weight vectors for each subpopulation are uniformly generated and distributed in the objective space, which uses the same method recommended in the classic MOEA/D algorithm [80]. In brief, they are uniformly sampled from a normalized hyperplane in the objective space; more details can be referred to in the literature for the MOEA/D [80].

However, this also requires that the final optimal solution for a weight vector should be obtained as closely as possible to both the weight vector itself and the ideal point  $Zmin$ . Thus, to better achieve this, the previously mentioned normalized modified I-TCH approach is then adopted in this paper as the aggregation function for decomposition, which can be formally defined as follows:

$$\begin{aligned} \text{minimize } g^{i-tch}(\mathbf{x}|\mathbf{w}, \mathbf{z}) &= \max_{1 \leq i \leq M} \{ |f_i(\mathbf{x}) - z_i| / w_i \} \\ \text{subject to } \mathbf{x} &= (x_1, x_2, \dots, x_D), \quad x_i = \{0, 1\} \end{aligned} \quad (5)$$

where the number of objectives  $M$  is set to two, and  $\mathbf{w}$  or  $\mathbf{z}$  respectively denotes the current weight vector  $W(i)$  or the ideal point  $Zmin$  in Algorithm 5, while  $\mathbf{x}$  denotes the  $D$ -dimensional decision vector of a certain solution. It can be seen that the above Equation (5) is almost the same as the previously introduced Equation (4) but uses  $w_i$  as a denominator instead of a multiplier. In fact, one of the comparison algorithms (i.e., MOEA/AWA [83]) in our experiments also adopted the I-TCH approach for decomposition, which also proved its effectiveness, while more details can be found in the following reference [83]. It is also

worth noting that Equation (5) can be treated as the final fitness function in this paper, while a smaller value of Equation (5) is preferred for environmental selection.

### 3.5. Task Merging Process

The task merging process of the MTDEA is shown in Algorithm 6, where the current population and the related task numbers are input. First, it is seen from Algorithm 6 that the merging of tasks is divided into the merging of task-1 and task-2, and the merging of task-3 and task-2. Nevertheless, they share the same merging rule based on the coverage of unique  $f_1$  objective values from two different tasks. As adopted in Lines 5 and 12 of Algorithm 6, if the coverage rate is beyond half, or if task-1 or task-3 is totally covered by task-2 in the  $f_1$  direction, then the other two tasks will be merged into task-2, and eventually, there will be only a single task-2 remaining for evolution. Moreover, an intuitive picture of the so-called coverage in the  $f_1$  direction (also shown as  $\psi_1$  and  $\psi_2$  in Algorithm 6) is given in Figure 5 to further explain the above process, which takes the merging of task-1 and task-2 as an example. As can be seen from Figure 5, it is actually quite common in bi-objective feature selection that two different solutions may share the same  $f_1$  value, meaning they own the same number of selected features (but probably have different feature combinations). Thus, the coverage of solutions in the  $f_1$  direction is naturally suitable for use as a criterion for determining the distribution relationship between two subpopulations during evolution.

---

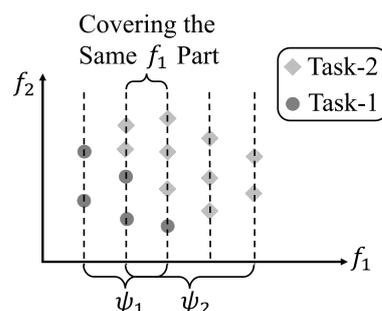
#### Algorithm 6 Merge( $Pop, T$ )

---

**Input:** current population  $Pop$ , related tasks  $T$ ;

**Output:** updated tasks  $T$ ;

- 1:  $\psi_2 \leftarrow$  get unique  $f_1$  objective values of all the task-2 solutions in  $Pop$  according to the task numbers in  $T$ ;
  - 2: **if** task-1 still exists **then**
  - 3:      $\psi_1 \leftarrow$  get unique  $f_1$  objective values of all the task-1 solutions in  $Pop$  according to the task numbers in  $T$ ;
  - 4:      $k \leftarrow$  get the number of all the duplicated elements between  $\psi_1$  and  $\psi_2$ ;
  - 5:     **if**  $(k/Length(\psi_2)) > 0.5 \parallel k == Length(\psi_1)$  **then**
  - 6:         set all the task-1 numbers in  $T$  to be task-2;
  - 7:     **end if**
  - 8: **end if**
  - 9: **if** task-3 still exists **then**
  - 10:      $\psi_3 \leftarrow$  get unique  $f_1$  objective values of all the task-3 solutions in  $Pop$  according to the task numbers in  $T$ ;
  - 11:      $k \leftarrow$  get the number of all the duplicated elements between  $\psi_3$  and  $\psi_2$ ;
  - 12:     **if**  $(k/Length(\psi_2)) > 0.5 \parallel k == Length(\psi_3)$  **then**
  - 13:         set all the task-3 numbers in  $T$  to be task-2;
  - 14:     **end if**
  - 15: **end if**
- 



**Figure 5.** An example of how solutions cover the  $f_1$  direction for the task-1 and task-2.

## 4. Experimental Setups

### 4.1. Classification Datasets

A total of 20 open-source classification datasets [84] were used as test problems to evaluate the search performance of the compared MOEAs in tackling bi-objective feature selection. More detailed information of these datasets is given in Table 1, with the number of total features sorted in ascending order. It can be seen that the number of total features for each dataset varied from 1024 to 10,509, which covered a wide variety of feature dimensions and concentrated on high dimensionality. In addition, the number of samples varied from 50 to 400, and the classes ranged from 2 to 40, also suggesting the comprehensiveness of the test problems.

**Table 1.** Attributes of each dataset used for test problems.

No.	Dataset	Feature	Sample	Class
1	ORL	1024	400	40
2	Yale	1024	165	15
3	Colon	2000	62	2
4	SRBCT	2308	83	4
5	AR10P	2400	130	10
6	PIE10P	2420	210	10
7	Lymphoma	4026	96	9
9	DLBCL	5469	77	2
8	TOX171	5748	171	4
10	Brain1	5920	90	5
11	Leukemia	7070	72	2
12	CNS	7129	60	2
13	ALLAML	7129	72	2
14	Carcinom	9182	174	11
15	Nci9	9712	60	9
16	Arcene	10,000	200	2
17	Pixraw10P	10,000	100	10
18	Orlraws10P	10,304	100	10
19	Brain2	10,367	50	4
20	Prostate	10,509	102	2

### 4.2. Comparison Algorithms

Seven state-of-the-art MOEAs were adopted in this paper as comparison algorithms against the proposed MTDEA, i.e., the NSGA-II (nondominated sorting-based genetic algorithm) [85], MOEA/D (MOEA based on decomposition) [80], MOEA/AWA (MOEA with adaptive weight adjustment) [83], MOEA/HD (MOEA based on hierarchical decomposition) [11], SparseEA (sparse evolutionary algorithm) [70], DAEA (duplication-analysis-based evolutionary algorithm) [71], and PRDH (problem reformation and duplication handling based algorithm) [72]. Among them, the NSGA-II and MOEA/D are among the most classic and well-known MOEAs based on dominance and decomposition, respectively. MOEA/AWA is also a decomposition-based MOEA but further modifies the adjustment of weight vectors in order to handle convex or disconnected Pareto fronts. Moreover, the MOEA/AWA also uses the same inverse Tchebycheff approach as the MTDEA, and thereby having comparative values. The MOEA/HD is a recently published MOEA based on dominance and decomposition, and especially designed for solving MOPs with complex Pareto fronts. The SparseEA, DAEA and PRDH are recently published MOEAs based on dominance and specifically designed for tackling large-scale and discrete MOPs, like bi-objective feature selection in classification.

### 4.3. Performance Indicators

In this work, multiple performance indicators (i.e., hypervolume [86], minimum classification error, and number of selected features [71]) were used to measure the com-

prehensive performance of the compared algorithms, in terms of both optimization and classification. More specifically, the hypervolume (HV) indicator was used as the main metric to measure the MOEAs' general optimization performance, whose reference point was set to (1, 1). In addition, the minimum classification error (MCE) indicator and the number of selected features (NSF) indicator were used to measure the MOEAs' classification performance, related to the best converged solution found in the  $f_2$  direction (i.e., the solution with the best classification accuracy). Here, MCE denotes the classification error rate of the solution and NSF denotes the corresponding number of selected features. Normally, a greater HV value means better performance, while smaller MCE and NSF values are preferred by contrast. Finally, Wilcoxon's test with a 5% significance level was adopted to identify the significant differences between each pair of algorithms, while Friedman's test was utilized to calculate the overall mean ranks among all algorithms.

#### 4.4. Parameter Settings

In the experiment, all the compared algorithms adopted the same traditional initialization method, for the sake of fairness, while the reproduction methods and other parameter settings taken from the algorithms' references. All algorithms were coded and run on an open-source MATLAB platform called PlatEMO [87]. For classification, each dataset was randomly divided into training and test subsets, with a proportion of 70/30, according to the stratified split process [71]. Moreover, a KNN ( $K = 5$ ) classification model was utilized with 10-fold cross-validation on the training data, so as to avoid feature selection bias [88]. Last, each experiment was independently run 20 times with randomly preset starting seeds, while the population size for each algorithm was set to 100 and the termination criterion (i.e., the number of objective function evaluations) was set to 10,000 (about 100 generations).

### 5. Experimental Studies

In this section, we first study the general empirical results of the algorithms on each dataset, in terms of the three performance metrics. Then, we analyze the nondominated solution distributions obtained by each algorithm in the objective space during optimization. Moreover, the proposed MTDEA is further analyzed by comparing it with a modified benchmark algorithm to verify the contribution of its essential components. Finally, the computational time of each algorithm was recorded and compared to analyze the computational efficiency of the proposed algorithm.

#### 5.1. General Performance Studies

The general performances of each algorithm run on all classification datasets are shown in Tables 2–5. First of all, it is suggested from the overall mean ranks of each algorithm on all datasets, in terms of the Friedman's test as shown in Table 2, that the proposed MTDEA always ranked first for all three performance metrics (HV, MCE, and NSF). In contrast, the DAEA generally fell behind the MTDEA and ranked second in terms of all three metrics, while the SparseEA seemed to perform much better in terms of MCE than for the other two metrics. Generally, the highest HV rank being for the MTDEA implied its comprehensive advantages in both diversity and convergence in multi-objective optimization, while the highest MCE rank being for MTDEA implied its overall superiority in finding solutions with the best classification accuracy. In addition, the highest NSF rank being for the MTDEA suggests its promising computational efficiency, because the number of selected features generally affects the complexity of classification models for learning (normally a larger number of selected features leads to a higher computational cost of classification). The computational time of each algorithm will be further analyzed in Section 5.4.

**Table 2.** General mean ranks calculated using Friedman’s test for each algorithm in terms of each metric, with best performances marked in gray.

Metric	MTDEA	NSGA-II	MOEA/D	MOEA/AWA	MOEA/HD	SparseEA	DAEA	PRDH
HV	1.012	5.215	3.917	5.290	6.275	7.530	2.192	4.567
MCE	3.615	4.688	5.124	4.787	5.030	4.303	3.740	4.714
NSF	1.000	5.270	3.112	5.640	6.461	7.949	2.110	4.457

**Table 3.** Mean HV performances of each algorithm on each dataset, with the best performance marked in gray, and insignificant differences prefixed by \*.

Dataset	MTDEA	NSGA-II	MOEA/D	MOEA/AWA	MOEA/HD	SparseEA	DAEA	PRDH
ORL	7.928e-01 ±1.80e-02	6.103e-01 ±1.08e-02	6.642e-01 ±1.53e-02	6.576e-01 ±1.79e-02	5.996e-01 ±1.25e-02	5.859e-01 ±9.94e-03	7.249e-01 ±1.75e-02	6.350e-01 ±1.18e-02
Yale	6.597e-01 ±3.18e-02	4.878e-01 ±1.34e-02	5.127e-01 ±2.62e-02	5.199e-01 ±2.40e-02	4.920e-01 ±3.39e-02	4.811e-01 ±1.64e-02	6.050e-01 ±2.48e-02	5.071e-01 ±2.15e-02
Colon	7.746e-01 ±3.84e-02	5.500e-01 ±2.65e-02	6.078e-01 ±4.78e-02	5.466e-01 ±3.55e-02	5.239e-01 ±3.30e-02	4.987e-01 ±2.01e-02	6.680e-01 ±3.95e-02	5.554e-01 ±3.10e-02
SRBCT	6.654e-01 ±1.70e-01	2.841e-01 ±2.05e-03	3.178e-01 ±2.03e-03	2.983e-01 ±2.74e-03	2.554e-01 ±1.81e-03	2.479e-01 ±1.69e-03	3.028e-01 ±1.96e-02	2.768e-01 ±3.17e-03
AR10P	4.972e-01 ±2.74e-02	3.631e-01 ±2.01e-02	3.709e-01 ±2.26e-02	3.544e-01 ±1.35e-02	3.449e-01 ±1.75e-02	3.223e-01 ±1.11e-02	4.314e-01 ±2.05e-02	3.603e-01 ±1.86e-02
PIE10P	8.127e-01 ±2.00e-02	6.023e-01 ±1.06e-02	6.458e-01 ±1.13e-02	6.003e-01 ±1.29e-02	5.883e-01 ±1.22e-02	5.434e-01 ±6.44e-03	6.982e-01 ±1.57e-02	6.056e-01 ±9.92e-03
Lymphoma	7.645e-01 ±2.18e-02	5.603e-01 ±9.91e-03	6.007e-01 ±9.76e-03	5.598e-01 ±1.59e-02	5.456e-01 ±8.85e-03	5.067e-01 ±1.07e-02	6.434e-01 ±1.57e-02	5.626e-01 ±6.99e-03
TOX171	6.749e-01 ±2.21e-02	4.829e-01 ±8.58e-03	4.876e-01 ±1.97e-02	4.697e-01 ±1.13e-02	4.776e-01 ±1.65e-02	4.575e-01 ±1.19e-02	5.423e-01 ±1.92e-02	4.907e-01 ±1.23e-02
DLBCL	8.054e-01 ±3.06e-02	5.852e-01 ±2.02e-02	5.999e-01 ±1.82e-02	5.722e-01 ±2.17e-02	5.735e-01 ±1.67e-02	5.469e-01 ±1.83e-02	6.703e-01 ±2.12e-02	5.982e-01 ±1.77e-02
Brain1	6.496e-01 ±1.38e-02	4.718e-01 ±3.11e-03	4.906e-01 ±1.09e-02	4.657e-01 ±1.06e-02	4.535e-01 ±1.00e-02	4.312e-01 ±1.78e-03	5.127e-01 ±8.53e-03	4.744e-01 ±2.57e-03
Leukemia	7.626e-01 ±2.70e-02	5.360e-01 ±8.95e-03	5.450e-01 ±1.94e-02	5.295e-01 ±1.89e-02	5.126e-01 ±1.79e-02	4.972e-01 ±9.84e-03	6.022e-01 ±1.67e-02	5.454e-01 ±2.04e-02
CNS	5.255e-01 ±5.88e-02	3.781e-01 ±3.28e-02	3.743e-01 ±3.32e-02	3.690e-01 ±3.09e-02	3.707e-01 ±3.25e-02	3.669e-01 ±1.57e-02	4.408e-01 ±3.03e-02	3.844e-01 ±2.94e-02
ALLAML	7.278e-01 ±2.93e-02	5.205e-01 ±1.52e-02	5.358e-01 ±1.34e-02	5.084e-01 ±1.60e-02	5.060e-01 ±1.44e-02	4.853e-01 ±1.52e-02	5.826e-01 ±1.83e-02	5.280e-01 ±1.32e-02
Carcinom	7.199e-01 ±1.40e-02	5.180e-01 ±1.09e-02	5.233e-01 ±1.55e-02	5.098e-01 ±1.00e-02	5.091e-01 ±1.10e-02	4.871e-01 ±8.18e-03	5.809e-01 ±1.18e-02	5.193e-01 ±7.20e-03
Nci9	3.570e-01 ±2.33e-02	2.406e-01 ±2.54e-02	2.616e-01 ±2.94e-02	2.544e-01 ±2.80e-02	2.370e-01 ±2.21e-02	2.254e-01 ±2.00e-02	2.707e-01 ±2.75e-02	2.451e-01 ±2.71e-02
Arcene	5.132e-01 ±4.15e-03	3.625e-01 ±1.10e-03	3.724e-01 ±1.85e-03	3.649e-01 ±2.23e-03	3.445e-01 ±1.24e-03	3.374e-01 ±1.29e-03	3.859e-01 ±2.75e-03	3.647e-01 ±1.71e-03
Pixraw10P	8.104e-01 ±6.33e-03	5.795e-01 ±2.22e-03	5.911e-01 ±9.88e-03	5.773e-01 ±7.09e-03	5.640e-01 ±7.64e-03	5.407e-01 ±1.57e-03	6.327e-01 ±9.37e-03	5.846e-01 ±2.65e-03
Orlraws10P	7.491e-01 ±1.43e-02	5.390e-01 ±7.53e-03	5.447e-01 ±9.58e-03	5.328e-01 ±1.13e-02	5.297e-01 ±8.00e-03	5.057e-01 ±3.88e-03	5.951e-01 ±8.65e-03	5.444e-01 ±4.77e-03
Brain2	5.603e-01 ±2.84e-02	3.903e-01 ±2.15e-02	3.824e-01 ±2.43e-02	3.816e-01 ±2.91e-02	3.782e-01 ±2.12e-02	3.687e-01 ±1.66e-02	4.335e-01 ±2.82e-02	3.898e-01 ±1.83e-02
Prostate	6.456e-01 ±2.38e-02	4.629e-01 ±1.29e-02	4.599e-01 ±1.51e-02	4.574e-01 ±1.58e-02	4.559e-01 ±1.16e-02	4.419e-01 ±8.71e-03	5.204e-01 ±1.53e-02	4.693e-01 ±1.52e-02



Table 4. Cont.

Dataset	MTDEA	NSGA-II	MOEA/D	MOEA/AWA	MOEA/HD	SparseEA	DAEA	PRDH
Orlraws10P	1.033e-01	* 1.050e-01	* 1.117e-01	* 1.067e-01	* 1.050e-01	* 1.017e-01	* 1.033e-01	* 1.017e-01
	±1.84e-02	±1.22e-02	±1.63e-02	±1.37e-02	±1.22e-02	±7.45e-03	±1.03e-02	±7.45e-03
Brain2	3.500e-01	3.767e-01	3.967e-01	3.833e-01	3.900e-01	* 3.700e-01	* 3.733e-01	3.833e-01
	±3.67e-02	±3.91e-02	±4.03e-02	±5.24e-02	±3.91e-02	±3.40e-02	±4.54e-02	±2.96e-02
Prostate	2.355e-01	* 2.419e-01	2.613e-01	* 2.387e-01	* 2.435e-01	* 2.274e-01	* 2.258e-01	* 2.371e-01
	±3.33e-02	±2.45e-02	±2.94e-02	±2.65e-02	±1.95e-02	±1.65e-02	±2.56e-02	±2.82e-02

Table 5. Mean NSF performance of each algorithm on each dataset, with best performances marked in gray, and insignificant differences prefixed by \*.

Dataset	MTDEA	NSGA-II	MOEA/D	MOEA/AWA	MOEA/HD	SparseEA	DAEA	PRDH
ORL	1.301e+02	3.498e+02	2.771e+02	3.257e+02	3.516e+02	4.001e+02	2.272e+02	3.054e+02
	±2.37e+01	±2.55e+01	±1.95e+01	±8.52e+01	±1.56e+01	±3.34e+01	±3.23e+01	±1.35e+01
Yale	1.368e+02	3.340e+02	2.682e+02	3.149e+02	3.291e+02	3.848e+02	2.311e+02	2.966e+02
	±2.19e+01	±2.99e+01	±1.46e+01	±5.93e+01	±1.68e+01	±2.03e+01	±4.03e+01	±1.94e+01
Colon	2.518e+02	6.990e+02	5.509e+02	7.224e+02	7.374e+02	8.752e+02	4.804e+02	6.866e+02
	±2.74e+01	±1.50e+01	±5.59e+01	±4.67e+01	±2.43e+01	±3.37e+01	±3.17e+01	±1.98e+01
SRBCT	2.712e+02	8.142e+02	6.096e+02	7.275e+02	9.884e+02	1.034e+03	7.004e+02	8.581e+02
	±5.61e+01	±1.24e+01	±1.23e+01	±1.66e+01	±1.10e+01	±1.03e+01	±1.19e+02	±1.93e+01
AR10P	3.889e+02	9.154e+02	7.826e+02	9.657e+02	9.311e+02	1.081e+03	6.920e+02	8.866e+02
	±4.58e+01	±1.77e+01	±2.61e+01	±5.67e+01	±2.85e+01	±3.27e+01	±7.48e+01	±1.65e+01
PIE10P	3.518e+02	9.076e+02	7.648e+02	9.489e+02	9.400e+02	1.085e+03	6.721e+02	8.814e+02
	±3.11e+01	±2.51e+01	±2.78e+01	±8.51e+01	±2.50e+01	±2.95e+01	±3.75e+01	±1.65e+01
Lymphoma	6.319e+02	1.600e+03	1.412e+03	1.607e+03	1.688e+03	1.890e+03	1.239e+03	1.598e+03
	±4.38e+01	±2.11e+01	±2.43e+01	±7.12e+01	±4.38e+01	±1.56e+01	±7.37e+01	±2.49e+01
TOX171	1.126e+03	2.510e+03	2.376e+03	2.594e+03	2.530e+03	2.768e+03	2.064e+03	2.465e+03
	±7.98e+01	±6.23e+01	±3.48e+01	±8.09e+01	±3.03e+01	±3.54e+01	±6.63e+01	±4.24e+01
DLBCL	8.355e+02	2.300e+03	2.154e+03	2.342e+03	2.361e+03	2.626e+03	1.838e+03	2.287e+03
	±5.62e+01	±2.54e+01	±9.42e+01	±7.18e+01	±4.26e+01	±2.31e+01	±7.16e+01	±2.78e+01
Brain1	9.804e+02	2.492e+03	2.332e+03	2.544e+03	2.648e+03	2.838e+03	2.143e+03	2.470e+03
	±9.27e+01	±2.65e+01	±9.29e+01	±9.00e+01	±8.53e+01	±1.52e+01	±7.27e+01	±2.19e+01
Leukemia	1.201e+03	3.041e+03	2.893e+03	3.081e+03	3.184e+03	3.423e+03	2.541e+03	3.010e+03
	±5.41e+01	±3.44e+01	±1.06e+02	±9.67e+01	±5.71e+01	±3.21e+01	±8.82e+01	±4.22e+01
CNS	1.375e+03	3.102e+03	2.939e+03	3.193e+03	3.172e+03	3.449e+03	2.550e+03	3.064e+03
	±8.44e+01	±7.67e+01	±6.44e+01	±8.03e+01	±4.51e+01	±4.15e+01	±5.91e+01	±8.10e+01
ALLAML	1.273e+03	3.085e+03	2.930e+03	3.112e+03	3.182e+03	3.450e+03	2.566e+03	3.043e+03
	±7.06e+01	±2.61e+01	±4.67e+01	±7.95e+01	±4.63e+01	±3.33e+01	±5.94e+01	±4.01e+01
Carcinom	1.853e+03	4.097e+03	3.981e+03	4.240e+03	4.149e+03	4.493e+03	3.480e+03	4.079e+03
	±9.64e+01	±3.82e+01	±6.53e+01	±7.39e+01	±3.06e+01	±3.26e+01	±1.01e+02	±8.56e+01
Nci9	1.777e+03	4.288e+03	4.083e+03	4.230e+03	4.600e+03	4.727e+03	3.890e+03	4.245e+03
	±1.07e+02	±3.25e+01	±3.29e+01	±2.61e+01	±2.86e+01	±1.76e+01	±4.30e+01	±1.69e+01
Arcene	1.686e+03	4.421e+03	4.240e+03	4.377e+03	4.748e+03	4.875e+03	3.996e+03	4.380e+03
	±7.54e+01	±1.99e+01	±3.35e+01	±4.04e+01	±2.25e+01	±2.34e+01	±5.00e+01	±3.10e+01
Pixraw10P	1.807e+03	4.426e+03	4.295e+03	4.452e+03	4.603e+03	4.866e+03	3.823e+03	4.368e+03
	±7.19e+01	±2.52e+01	±1.12e+02	±8.04e+01	±8.66e+01	±1.79e+01	±1.06e+02	±3.00e+01
Orlraws10P	1.971e+03	4.581e+03	4.463e+03	4.648e+03	4.697e+03	5.023e+03	3.890e+03	4.536e+03
	±6.24e+01	±3.63e+01	±4.45e+01	±8.97e+01	±5.78e+01	±2.67e+01	±7.47e+01	±3.49e+01
Brain2	2.038e+03	4.642e+03	4.584e+03	4.734e+03	4.729e+03	5.084e+03	3.935e+03	4.591e+03
	±7.39e+01	±4.55e+01	±1.21e+02	±9.91e+01	±4.01e+01	±3.71e+01	±8.19e+01	±4.58e+01
Prostate	2.081e+03	4.713e+03	4.587e+03	4.820e+03	4.796e+03	5.154e+03	4.024e+03	4.670e+03
	±9.25e+01	±5.72e+01	±5.63e+01	±8.07e+01	±4.31e+01	±4.61e+01	±1.16e+02	±6.92e+01

5.2. Nondominated Solution Distributions

For a more intuitive view of the performances, Figure 6 illustrates the nondominated solution distributions of each algorithm, in terms of Pareto curves in the objective space, with the median HV performances run on each dataset. We only show the nondominated solutions instead of the whole population, for a clearer representation of the diversity and convergence states. It is also worth noting that the nondominated solutions shown in each subfigure of Figure 6 look quite sparse for most datasets, due to the limited number of training samples and the large number of features. General speaking, it is suggested from Figure 6 that the proposed MTDEA performed the best on the majority of the classification datasets, with a promising overall diversity and convergence.

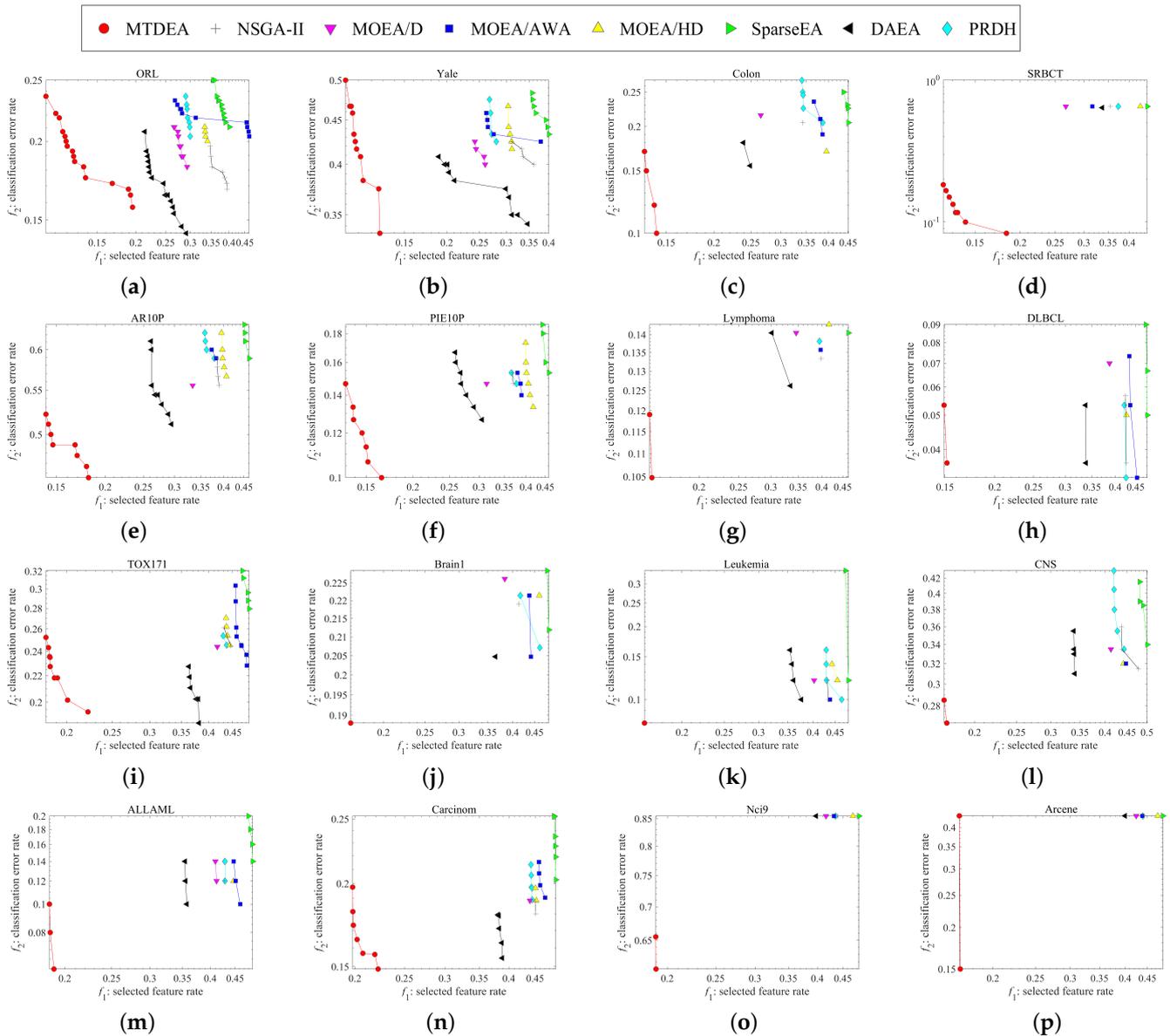
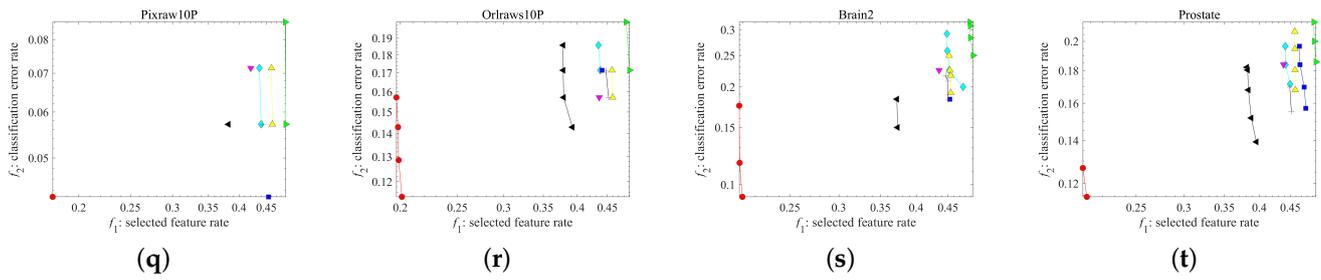


Figure 6. Cont.



**Figure 6.** Nondominated solution distributions in objective space, with median HV performances obtained by each algorithm. (a) ORL. (b) Yale. (c) Colon. (d) SRBCT. (e) AR10P. (f) PIE10P. (g) Lymphoma. (h) DLBCL. (i) TOX171. (j) Brain1. (k) Leukemia. (l) CNS. (m) ALLAML. (n) Carcinom. (o) Nci9. (p) Arcene. (q) Pixraw10P. (r) Orlraws10P. (s) Brain2. (t) Prostate.

In detail, the MTDEA generally obtained the largest number of nondominated solutions with best diversity, while the solution distributions in the  $f_1$  and  $f_2$  directions of objective space were overall the most converged. However, there are still a few unsatisfactory performances shown in some subfigures, such as Figure 6a,h,i, where the MTDEA was not the best (but still good) of all the algorithms in the  $f_2$  objective direction (i.e., the classification error rate). Nevertheless, the few flaws of the MTDEA are generally insignificant and it was generally close to the best algorithms in the  $f_2$  objective direction, while the MTDEA always exhibited the best performance in the  $f_1$  objective direction (i.e., the selected feature rate). It should also be noted that although we choose the performance for the median HV values obtained by each algorithm for the sake of fairness, they have a certain degree of random fluctuation and can only be seen as a supplement to the previously studied statistical empirical results in Section 5.1.

### 5.3. Component Contribution Analyses

To further confirm the effectiveness of the proposed MTDEA, especially to verify the contributions of its essential component (i.e., the dynamic multi-task framework with adaptive ideal points for decomposition), rather than adopting other existing methods (such as the I-TCH based decomposition approach and normalization), the MTDEA and its Baseline algorithm were compared in terms of all three performance metrics, as shown in Table 6. In this paper, the so-called Baseline algorithm removed all the multi-task contents from Algorithm 1, replacing Algorithms 2 and 4 respectively with traditional initialization and reproduction methods, which are commonly used in most MOEAs [80,85], while its environmental selection process still used the modified and normalized I-TCH based decomposition approach as previously introduced. It is suggested from Table 6 that the MTDEA generally showed outstanding performance, with significant advantages over the Baseline algorithm on almost all datasets. The only insignificant loss for the MTDEA took place on the Carcinom dataset in terms of the MCE metric, which was slightly worse than the Baseline algorithm. By contrast, the performance of the MTDEA in terms of the HV and NSF metrics were the best on all datasets compared with the Benchmark algorithm. Thus, it was demonstrated that the essential component of the MTDEA made an overall positive contribution in improving the performance in finding optimal or near-optimal solutions in the large-scale search space of the high-dimensional datasets.

### 5.4. Computational Time Analyses

The mean computational time for each algorithm run on each dataset was recorded and calculated in seconds, as shown in Table 7. In Table 7, we not only mark the best performance in gray but also mark the second best performance in a lighter gray color, in order to provide more comprehensive analyses. First, it can be seen from Table 7 that the MTDEA performed the best on 9 out of 20 datasets, and it also performed second best on the remaining 11 datasets. Thus, it can be concluded that the MTDEA had a generally high computational efficiency, spending much less time than the traditional

algorithms in tackling feature selection. Moreover, when looking into the 11 second best performances marked in the lighter gray color, it is seen that the MTDEA only just lost to SparseEA in terms of computational time, while most of the losses were on very high-dimensional datasets. In fact, the reason why the SparseEA could spend even less time than the proposed MTDEA running on those datasets was mainly due to its pre-analysis process before the normal evolution. In brief, the SparseEA pre-analyzed each feature’s classification performance, which actually consumed much less time than the subsequent normal evolution, because only a single feature was selected for classification.

**Table 6.** Mean HV, MCE, and NSF performances of the MTDEA against the Baseline algorithm, with best performances marked in gray and insignificant differences prefixed by \*.

Dataset	HV Metric		MCE Metric		NSF Metric	
	MTDEA	Baseline	MTDEA	Baseline	MTDEA	Baseline
ORL	7.9280e-01 ±1.797e-02	6.2015e-01 ±2.020e-02	1.3750e-01 ±1.493e-02	* 1.4250e-01 ±1.402e-02	1.3005e+02 ±2.373e+01	3.3955e+02 ±2.964e+01
Yale	6.5971e-01 ±3.181e-02	4.9766e-01 ±2.479e-02	2.9444e-01 ±3.596e-02	3.3556e-01 ±3.808e-02	1.3675e+02 ±2.190e+01	3.3045e+02 ±1.961e+01
Colon	7.7459e-01 ±3.839e-02	5.4779e-01 ±2.682e-02	1.4211e-01 ±4.860e-02	2.1316e-01 ±4.345e-02	2.5180e+02 ±2.739e+01	7.0630e+02 ±1.718e+01
SRBCT	6.6537e-01 ±1.696e-01	2.8661e-01 ±1.828e-03	2.8400e-01 ±2.152e-01	6.4000e-01 ±1.139e-16	2.7120e+02 ±5.614e+01	7.9880e+02 ±1.110e+01
AR10P	4.9725e-01 ±2.744e-02	3.5777e-01 ±1.439e-02	4.6875e-01 ±3.524e-02	4.9750e-01 ±2.420e-02	3.8895e+02 ±4.576e+01	9.2410e+02 ±2.474e+01
PIE10P	8.1273e-01 ±2.001e-02	5.9713e-01 ±9.652e-03	7.7500e-02 ±2.247e-02	1.0083e-01 ±1.144e-02	3.5175e+02 ±3.107e+01	9.1285e+02 ±2.042e+01
Lymphoma	7.6449e-01 ±2.178e-02	5.6584e-01 ±8.253e-03	1.2000e-01 ±2.269e-02	* 1.3167e-01 ±1.701e-02	6.3190e+02 ±4.382e+01	1.5825e+03 ±2.530e+01
TOX171	6.7490e-01 ±2.214e-02	4.8274e-01 ±1.347e-02	2.0377e-01 ±2.974e-02	2.2736e-01 ±2.247e-02	1.1264e+03 ±7.979e+01	2.5053e+03 ±8.390e+01
DLBCL	8.0543e-01 ±3.059e-02	5.7887e-01 ±2.118e-02	7.1739e-02 ±3.805e-02	* 7.6087e-02 ±3.419e-02	8.3550e+02 ±5.622e+01	2.2779e+03 ±3.618e+01
Brain1	6.4956e-01 ±1.384e-02	4.7324e-01 ±5.628e-03	2.5926e-01 ±1.202e-02	* 2.5926e-01 ±0.000e+00	9.8040e+02 ±9.274e+01	2.4799e+03 ±4.795e+01
Leukemia	7.6257e-01 ±2.699e-02	5.4152e-01 ±1.669e-02	1.0909e-01 ±3.731e-02	* 1.3182e-01 ±2.912e-02	1.2007e+03 ±5.407e+01	2.9922e+03 ±3.736e+01
CNS	5.2552e-01 ±5.878e-02	3.7420e-01 ±2.658e-02	4.0000e-01 ±7.774e-02	* 4.2500e-01 ±4.862e-02	1.3754e+03 ±8.443e+01	3.0789e+03 ±6.460e+01
ALLAML	7.2778e-01 ±2.933e-02	5.2572e-01 ±1.071e-02	1.4545e-01 ±3.789e-02	* 1.5227e-01 ±2.224e-02	1.2732e+03 ±7.057e+01	3.0622e+03 ±4.950e+01
Carcinom	7.1986e-01 ±1.404e-02	5.2429e-01 ±8.217e-03	1.3558e-01 ±1.588e-02	* 1.3462e-01 ±1.395e-02	1.8533e+03 ±9.638e+01	4.0858e+03 ±6.919e+01
Nci9	3.5698e-01 ±2.326e-02	2.4945e-01 ±2.105e-02	6.2895e-01 ±3.183e-02	* 6.4737e-01 ±3.856e-02	1.7768e+03 ±1.071e+02	4.2062e+03 ±4.303e+01
Arcene	5.1318e-01 ±4.154e-03	3.6631e-01 ±1.934e-03	4.3333e-01 ±1.139e-16	* 4.3333e-01 ±1.139e-16	1.6858e+03 ±7.540e+01	4.3514e+03 ±3.511e+01
Pixraw10P	8.1039e-01 ±6.335e-03	5.8583e-01 ±4.056e-03	3.3333e-02 ±0.000e+00	* 3.3333e-02 ±0.000e+00	1.8071e+03 ±7.186e+01	4.3545e+03 ±4.601e+01
Orlraws10P	7.4909e-01 ±1.430e-02	5.4469e-01 ±6.752e-03	1.0333e-01 ±1.842e-02	* 1.0333e-01 ±1.026e-02	1.9710e+03 ±6.243e+01	4.5222e+03 ±3.812e+01
Brain2	5.6026e-01 ±2.838e-02	3.8877e-01 ±1.692e-02	3.5000e-01 ±3.667e-02	3.8333e-01 ±2.962e-02	2.0377e+03 ±7.393e+01	4.6294e+03 ±8.246e+01
Prostate	6.4555e-01 ±2.378e-02	4.6144e-01 ±1.361e-02	2.3548e-01 ±3.326e-02	* 2.4839e-01 ±2.585e-02	2.0812e+03 ±9.247e+01	4.6716e+03 ±4.519e+01

**Table 7.** Mean computational time in seconds for each algorithm run on each dataset, with best times marked in gray, and insignificant differences prefixed by \*. Moreover, if MTDEA performed second best of all, it is also marked in a lighter gray color.

Dataset	MTDEA	NSGA-II	MOEA/D	MOEA/AWA	MOEA/HD	SparseEA	DAEA	PRDH
ORL	7.289e+02 ±2.43e+01	1.071e+03 ±3.74e+01	9.947e+02 ±3.94e+01	9.512e+02 ±3.00e+01	1.084e+03 ±4.59e+01	1.065e+03 ±2.32e+01	8.724e+02 ±3.81e+01	1.053e+03 ±2.42e+01
Yale	1.434e+02 ±4.69e+00	1.942e+02 ±4.97e+00	1.860e+02 ±3.54e+00	1.910e+02 ±3.57e+00	2.032e+02 ±6.39e+00	2.028e+02 ±4.44e+00	1.790e+02 ±4.73e+00	2.022e+02 ±4.74e+00
Colon	6.170e+01 ±1.06e+00	8.630e+01 ±1.55e+00	7.946e+01 ±1.58e+00	8.715e+01 ±1.85e+00	8.672e+01 ±2.60e+00	8.419e+01 ±1.84e+00	8.861e+01 ±1.74e+00	9.450e+01 ±1.65e+00
SRBCT	1.653e+02 ±5.26e+00	2.327e+02 ±5.64e+00	2.078e+02 ±2.87e+00	2.154e+02 ±5.34e+00	2.618e+02 ±7.01e+00	2.254e+02 ±5.98e+00	2.078e+02 ±1.03e+01	2.558e+02 ±3.74e+00
AR10P	4.898e+02 ±2.02e+01	6.246e+02 ±9.67e+00	6.008e+02 ±1.52e+01	6.226e+02 ±1.60e+01	6.349e+02 ±1.30e+01	5.974e+02 ±1.74e+01	5.872e+02 ±1.11e+01	6.314e+02 ±1.02e+01
PIE10P	9.958e+02 ±3.44e+01	1.462e+03 ±9.16e+01	1.382e+03 ±6.56e+01	1.469e+03 ±1.03e+02	1.539e+03 ±9.56e+01	1.411e+03 ±6.73e+01	1.198e+03 ±5.89e+01	1.415e+03 ±5.14e+01
Lymphoma	6.800e+02 ±1.93e+01	9.464e+02 ±5.54e+01	8.585e+02 ±2.14e+01	9.201e+02 ±5.38e+01	9.785e+02 ±7.62e+01	8.230e+02 ±5.04e+01	8.358e+02 ±3.06e+01	8.834e+02 ±1.62e+01
TOX171	2.417e+03 ±1.96e+02	3.248e+03 ±2.29e+02	3.065e+03 ±2.00e+02	3.128e+03 ±2.37e+02	3.300e+03 ±2.51e+02	2.301e+03 ±1.68e+02	2.826e+03 ±3.77e+01	3.013e+03 ±1.10e+02
DLBCL	8.218e+02 ±3.99e+01	1.135e+03 ±7.51e+01	1.083e+03 ±7.73e+01	1.109e+03 ±7.66e+01	1.159e+03 ±8.80e+01	9.196e+02 ±7.54e+01	1.032e+03 ±2.52e+01	1.064e+03 ±1.94e+01
Brain1	1.224e+03 ±1.08e+02	1.731e+03 ±1.44e+02	1.715e+03 ±1.56e+02	1.726e+03 ±1.42e+02	1.789e+03 ±1.59e+02	1.266e+03 ±1.08e+02	1.498e+03 ±8.45e+01	1.593e+03 ±6.91e+01
Leukemia	1.150e+03 ±9.20e+01	1.648e+03 ±1.36e+02	1.584e+03 ±1.28e+02	1.705e+03 ±1.89e+02	1.708e+03 ±1.28e+02	1.080e+03 ±9.02e+01	1.379e+03 ±5.30e+01	1.537e+03 ±8.69e+01
CNS	9.885e+02 ±8.98e+01	1.199e+03 ±8.43e+01	1.145e+03 ±8.29e+01	1.161e+03 ±8.87e+01	1.217e+03 ±9.35e+01	8.565e+02 ±6.03e+01	1.077e+03 ±2.52e+01	1.122e+03 ±2.03e+01
ALLAML	1.175e+03 ±9.54e+01	1.677e+03 ±1.43e+02	1.648e+03 ±1.51e+02	1.705e+03 ±1.90e+02	1.731e+03 ±1.42e+02	1.086e+03 ±8.30e+01	1.407e+03 ±5.50e+01	1.558e+03 ±8.15e+01
Carcinom	4.527e+03 ±4.36e+02	5.767e+03 ±5.05e+02	5.719e+03 ±4.51e+02	5.692e+03 ±4.65e+02	5.860e+03 ±6.06e+02	3.349e+03 ±3.03e+01	4.934e+03 ±1.35e+02	5.280e+03 ±2.00e+02
Nci9	1.348e+03 ±9.02e+01	1.940e+03 ±1.58e+02	1.891e+03 ±9.31e+01	1.926e+03 ±8.43e+01	2.276e+03 ±3.22e+02	1.228e+03 ±1.69e+02	1.818e+03 ±4.41e+01	1.867e+03 ±1.14e+02
Arcene	5.231e+03 ±1.97e+02	7.230e+03 ±5.42e+02	6.937e+03 ±4.00e+02	7.047e+03 ±3.64e+02	7.745e+03 ±9.04e+02	4.329e+03 ±4.43e+02	8.062e+03 ±1.38e+03	6.646e+03 ±3.93e+02
Pixraw10P	2.564e+03 ±1.83e+02	3.430e+03 ±1.94e+02	3.382e+03 ±2.31e+02	3.429e+03 ±2.45e+02	3.694e+03 ±3.92e+02	1.989e+03 ±2.61e+02	3.845e+03 ±5.98e+02	3.276e+03 ±2.56e+02
Orlraws10P	2.693e+03 ±1.63e+02	3.628e+03 ±3.12e+02	3.547e+03 ±2.06e+02	3.598e+03 ±2.57e+02	4.208e+03 ±8.63e+02	* 2.602e+03 ±2.36e+02	3.944e+03 ±6.29e+02	3.400e+03 ±2.76e+02
Brain2	1.230e+03 ±6.56e+01	1.789e+03 ±1.49e+02	1.708e+03 ±1.66e+02	1.723e+03 ±1.44e+02	1.978e+03 ±3.51e+02	1.170e+03 ±6.17e+01	1.775e+03 ±2.42e+02	1.661e+03 ±1.03e+02
Prostate	2.857e+03 ±1.36e+02	3.906e+03 ±1.63e+02	3.762e+03 ±1.51e+02	3.775e+03 ±2.99e+02	4.505e+03 ±1.02e+03	* 2.676e+03 ±7.03e+02	3.546e+03 ±1.71e+02	3.426e+03 ±1.46e+02

## 6. Conclusions

This paper proposed a multi-task decomposition-based evolutionary algorithm (named MTDEA) for solving the discrete bi-objective feature selection problem with binary-coding, especially for large-scale classification data. The proposed MTDEA not only adopts an adaptive decomposition-based MOEA framework with transformed ideal points for more efficient evolution, but also has a multi-task mechanism with dynamically merged tasks for

more diverse cooperation. In this work, we focused on studying the adaptive integration of the multi-task framework with the decomposition-based evolutionary approach. To be more specific, first, a dynamic multi-task framework was designed, which initializes different search tasks for three separate subpopulations and then eventually merges them into a single-task population at the later evolutionary stage. Furthermore, an adaptive decomposition-based evolutionary approach was also designed, which cooperates with the above multi-task framework and adaptively adjusts the global ideal point for each multi-task subpopulation. In this way, the search performance of the MTDEA was significantly enhanced, while the population diversity and the classification convergence could also be delicately balanced. It is suggested from the comprehensive empirical studies that the MTDEA significantly outperformed the seven other state-of-the-art MOEAs on a total of 20 high-dimensional classification datasets, in terms of three different performance metrics, with a better distribution diversity and classification convergence. Moreover, it was also shown that the most essential component of the MTDEA, i.e., the dynamic multi-task framework with adaptive ideal points for decomposition, made an overall positive contribution to the improvement in the algorithm performance.

In our future work, it is planned to study the adaptive setting of initial parameters for the multi-task framework, probably based on the dynamic analysis of the optimization environment of different tested datasets. Moreover, we would also like to further study the performance of the proposed MTDEA on more high-dimensional datasets, as well as its applicability to other kinds of discrete optimization problems such as neural network construction and community node detection.

**Author Contributions:** Conceptualization, H.X.; Methodology, H.X.; Software, H.X., J.L., M.L. and H.Z.; Validation, C.H., J.L. and R.X.; Formal analysis, R.X.; Investigation, C.H., J.L., M.L. and H.Z.; Resources, C.H. and M.L.; Data curation, H.Z. and R.X.; Writing—original draft, H.X.; Project administration, H.X.; Funding acquisition, 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 number 62103209, by the Natural Science Foundation of Fujian Province grant number 2020J05213, by the Scientific Research Project of Putian Science and Technology Bureau grant number 2021ZP07, by the Startup Fund for Advanced Talents of Putian University grant number 2019002, and by the Research Projects of Putian University grant number JG202306.

**Data Availability Statement:** The data will be made available by the authors on request.

**Acknowledgments:** This work was supported in part by the National Natural Science Foundation of China under Grant 62103209, by the Natural Science Foundation of Fujian Province under Grant 2020J05213, by the Scientific Research Project of Putian Science and Technology Bureau under Grant 2021ZP07, 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.; 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](#)]
7. Castillo, J.C.; Segura, C.; Coello, C.A.C. VSD-MOEA: A Dominance-Based Multiobjective Evolutionary Algorithm with Explicit Variable Space Diversity Management. *Evol. Comput.* **2022**, *30*, 195–219. [[CrossRef](#)] [[PubMed](#)]
8. 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](#)]
9. 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](#)]
10. Ganesh, N.; Shankar, R.; Kalita, K.; Jangir, P.; Oliva, D.; Pérez-Cisneros, M. A Novel Decomposition-Based Multi-Objective Symbiotic Organism Search Optimization Algorithm. *Mathematics* **2023**, *11*, 1898. [[CrossRef](#)]
11. 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](#)] [[PubMed](#)]
12. Montero, E.; Zapotecas-Martínez, S. An Analysis of Parameters of Decomposition-Based MOEAs on Many-Objective Optimization. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [[CrossRef](#)]
13. Nojima, Y.; Arahari, K.; Takemura, S.; Ishibuchi, H. Multiobjective fuzzy genetics-based machine learning based on MOEA/D with its modifications. In Proceedings of the 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy, 9–12 July 2017; pp. 1–6. [[CrossRef](#)]
14. 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](#)]
15. Menchaca-Mendez, A.; Coello, C.A.C. GDE-MOEA: A new MOEA based on the generational distance indicator and  $\epsilon$ -dominance. In Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 25–28 May 2015; pp. 947–955. [[CrossRef](#)]
16. Xu, H.; Zeng, W.; Zeng, X.; Yen, G.G. A Polar-Metric-Based Evolutionary Algorithm. *IEEE Trans. Cybern.* **2021**, *51*, 3429–3440. [[CrossRef](#)]
17. Lopes, C.L.d.V.; Martins, F.V.C.; Wanner, E.F.; Deb, K. Analyzing Dominance Move (MIP-DoM) Indicator for Multiobjective and Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2022**, *26*, 476–489. [[CrossRef](#)]
18. 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](#)]
19. 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](#)]
20. 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](#)]
21. Goh, C.K.; Tan, K.C. A competitive-cooperative coevolutionary paradigm for dynamic multiobjective optimization. *IEEE Trans. Evol. Comput.* **2008**, *13*, 103–127.
22. 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](#)]
23. 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](#)]
24. 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.
25. Gupta, A.; Ong, Y.S.; Feng, L.; Tan, K.C. Multiobjective Multifactorial Optimization in Evolutionary Multitasking. *IEEE Trans. Cybern.* **2017**, *47*, 1652–1665. [[CrossRef](#)]
26. Rauniar, 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](#)]
27. Palakonda, V.; Mallipeddi, R. An Evolutionary Algorithm for Multi and Many-Objective Optimization with Adaptive Mating and Environmental Selection. *IEEE Access* **2020**, *8*, 82781–82796. [[CrossRef](#)]
28. Park, J.; Ajani, O.S.; Mallipeddi, R. Optimization-Based Energy Disaggregation: A Constrained Multi-Objective Approach. *Mathematics* **2023**, *11*, 563. [[CrossRef](#)]
29. Ríos, A.; Hernández, E.E.; Valdez, S.I. A Two-Stage Mono- and Multi-Objective Method for the Optimization of General UPS Parallel Manipulators. *Mathematics* **2021**, *9*, 543. [[CrossRef](#)]
30. Leung, M.F.; Coello, C.A.C.; Cheung, C.C.; Ng, S.C.; Lui, A.K.F. A Hybrid Leader Selection Strategy for Many-Objective Particle Swarm Optimization. *IEEE Access* **2020**, *8*, 189527–189545. [[CrossRef](#)]
31. 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](#)]

32. Garces-Jimenez, A.; Gomez-Pulido, J.M.; Gallego-Salvador, N.; Garcia-Tejedor, A.J. Genetic and Swarm Algorithms for Optimizing the Control of Building HVAC Systems Using Real Data: A Comparative Study. *Mathematics* **2021**, *9*, 2181. [[CrossRef](#)]
33. Ramos-Pérez, J.M.; Miranda, G.; Segredo, E.; León, C.; Rodríguez-León, C. Application of Multi-Objective Evolutionary Algorithms for Planning Healthy and Balanced School Lunches. *Mathematics* **2021**, *9*, 80. [[CrossRef](#)]
34. 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](#)]
35. 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](#)]
36. Zhu, W.; Li, H.; Wei, W. A Two-Stage Multi-Objective Evolutionary Algorithm for Community Detection in Complex Networks. *Mathematics* **2023**, *11*, 2702. [[CrossRef](#)]
37. 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](#)]
38. Xue, Y.; Chen, C.; S?owik, A. Neural Architecture Search Based on a Multi-Objective Evolutionary Algorithm with Probability Stack. *IEEE Trans. Evol. Comput.* **2023**, *27*, 778–786. [[CrossRef](#)]
39. Ponti, A.; Candelieri, A.; Giordani, I.; Archetti, F. Intrusion Detection in Networks by Wasserstein Enabled Many-Objective Evolutionary Algorithms. *Mathematics* **2023**, *11*, 2342. [[CrossRef](#)]
40. Othman, R.A.; Darwish, S.M.; Abd El-Moghith, I.A. A Multi-Objective Crowding Optimization Solution for Efficient Sensing as a Service in Virtualized Wireless Sensor Networks. *Mathematics* **2023**, *11*, 1128. [[CrossRef](#)]
41. 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](#)]
42. 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](#)]
43. 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](#)]
44. 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. [[CrossRef](#)]
45. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [[CrossRef](#)]
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; pp. 444–452.
47. 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.
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. Bai, H.; Cheng, R.; Yazdani, D.; Tan, K.C.; Jin, Y. Evolutionary Large-Scale Dynamic Optimization Using Bilevel Variable Grouping. *IEEE Trans. Cybern.* **2022**, *53*, 6937–6950. [[CrossRef](#)] [[PubMed](#)]
51. Yang, X.; Zou, J.; Yang, S.; Zheng, J.; Liu, Y. A Fuzzy Decision Variables Framework for Large-Scale Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **2023**, *27*, 445–459. [[CrossRef](#)]
52. Ma, X.; Zheng, Y.; Zhu, Z.; Li, X.; Wang, L.; Qi, Y.; Yang, J. Improving Evolutionary Multitasking Optimization by Leveraging Inter-Task Gene Similarity and Mirror Transformation. *IEEE Comput. Intell. Mag.* **2021**, *16*, 38–53. [[CrossRef](#)]
53. Ming, F.; Gong, W.; Gao, L. Adaptive Auxiliary Task Selection for Multitasking-Assisted Constrained Multi-Objective Optimization [Feature]. *IEEE Comput. Intell. Mag.* **2023**, *18*, 18–30. [[CrossRef](#)]
54. Liang, Z.; Liang, W.; Wang, Z.; Ma, X.; Liu, L.; Zhu, Z. Multiobjective Evolutionary Multitasking with Two-Stage Adaptive Knowledge Transfer Based on Population Distribution. *IEEE Trans. Syst. Man, Cybern. Syst.* **2022**, *52*, 4457–4469. [[CrossRef](#)]
55. Feng, Y.; Feng, L.; Kwong, S.; Tan, K.C. A Multivariation Multifactorial Evolutionary Algorithm for Large-Scale Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **2022**, *26*, 248–262. [[CrossRef](#)]
56. Gao, W.; Cheng, J.; Gong, M.; Li, H.; Xie, J. Multiobjective Multitasking Optimization with Subspace Distribution Alignment and Decision Variable Transfer. *IEEE Trans. Emerg. Top. Comput. Intell.* **2022**, *6*, 818–827. [[CrossRef](#)]
57. 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](#)]
58. De La Iglesia, B. Evolutionary computation for feature selection in classification problems. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 381–407. [[CrossRef](#)]
59. 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](#)]

60. Dokeroglu, T.; Deniz, A.; Kiziloz, H.E. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing* **2022**, *494*, 269–296. [[CrossRef](#)]
61. 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](#)] [[PubMed](#)]
62. 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](#)]
63. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
64. 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](#)]
65. 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](#)]
66. 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](#)]
67. 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](#)]
68. 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](#)]
69. 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](#)]
70. 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](#)]
71. 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](#)]
72. Jiao, R.; Xue, B.; Zhang, M. Solving Multi-objective Feature Selection Problems in Classification via Problem Reformulation and Duplication Handling. *IEEE Trans. Evol. Comput.* **2022**, *1*, 1. [[CrossRef](#)]
73. 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](#)] [[PubMed](#)]
74. Li, H.; Zhang, Q. A multiobjective differential evolution based on decomposition for multiobjective optimization with variable linkages. In Proceedings of the International Conference on Parallel Problem Solving from Nature, Reykjavik, Iceland, 9–13 September 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 583–592.
75. Yang, S.; Huang, H.; Luo, F.; Xu, Y.; Hao, Z. Local-Diversity Evaluation Assignment Strategy for Decomposition-Based Multiobjective Evolutionary Algorithm. *IEEE Trans. Syst. Man, Cybern. Syst.* **2023**, *53*, 1697–1709. [[CrossRef](#)]
76. He, L.; Shang, K.; Nan, Y.; Ishibuchi, H.; Srinivasan, D. Relation Between Objective Space Normalization and Weight Vector Scaling in Decomposition-Based Multiobjective Evolutionary Algorithms. *IEEE Trans. Evol. Comput.* **2023**, *27*, 1177–1191. [[CrossRef](#)]
77. Zhao, Q.; Guo, Y.; Yao, X.; Gong, D. Decomposition-Based Multiobjective Optimization Algorithms with Adaptively Adjusting Weight Vectors and Neighborhoods. *IEEE Trans. Evol. Comput.* **2023**, *27*, 1485–1497. [[CrossRef](#)]
78. Pang, L.M.; Ishibuchi, H.; Shang, K. Use of Two Penalty Values in Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Cybern.* **2023**, *53*, 7174–7186. [[CrossRef](#)]
79. Wang, R.; Zhang, Q.; Zhang, T. Decomposition-Based Algorithms Using Pareto Adaptive Scalarizing Methods. *IEEE Trans. Evol. Comput.* **2016**, *20*, 821–837. [[CrossRef](#)]
80. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Comput.* **2007**, *11*, 712–731. [[CrossRef](#)]
81. Zhang, Q.; Liu, W.; Li, H. The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances. In Proceedings of the 2009 IEEE Congress on Evolutionary Computation, Trondheim, Norway, 18–21 May 2009; pp. 203–208. [[CrossRef](#)]
82. Li, K.; Deb, K.; Zhang, Q.; Kwong, S. An Evolutionary Many-Objective Optimization Algorithm Based on Dominance and Decomposition. *IEEE Trans. Evol. Comput.* **2015**, *19*, 694–716. [[CrossRef](#)]
83. Qi, Y.; Ma, X.; Liu, F.; Jiao, L.; Sun, J.; Wu, J. MOEA/D with Adaptive Weight Adjustment. *Evol. Comput.* **2014**, *22*, 231–264. [[CrossRef](#)]
84. Dua, D.; Graff, C. UCI Machine Learning Repository. 2017. Available online: <https://archive.ics.uci.edu/datasets> (accessed on 12 April 2024).
85. 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](#)]
86. While, L.; Hingston, P.; Barone, L.; Huband, S. A faster algorithm for calculating Hypervolume. *IEEE Trans. Evol. Comput.* **2006**, *10*, 29–38. [[CrossRef](#)]

- 
87. 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](#)]
  88. 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.