"Differential Evolution Based Feature Selection: A Niching-based Multi-objective Approach" *Online Supplementary Materials*

Peng Wang, *Graduate Student Member, IEEE*, Bing Xue, *Senior Member, IEEE*, Jing Liang, *Senior Member, IEEE*, and Mengjie Zhang, *Fellow, IEEE*

I. Introduction

This is the Online Supplementary Materials of "Differential Evolution Based Feature Selection: A Niching-based Multi-objective Approach".

First, the basic procedure of multi-objective DE (MODE) algorithm is shown. Second, the computational complexity of the proposed NMDE method is given. Then, the performance of NMDE and MOEA/D-DE is compared in Section S.IV. More results between the proposed NMDE method and other methods can be seen in Section S.V. Next, the effects of major components on the proposed NMDE method are analyzed one by one. Finally, the influence of different learning algorithms on the obtained feature subsets is shown in Section S.VII.

II. Basic Multi-objective Differential Evolution Procedure

MODE is taken as a base, and then it is combined with the proposed strategies in order to test the effectiveness of the proposed strategies. The procedure of MODE is shown in Algorithm 1.

Four main components, initialization, mutation, crossover, and environmental selection are included in MODE, and they are introduced in detail as follows.

*1) **Initialization**:*

$$x_{i,j} = l_j + \mathrm{rand}(0,1) \times (u_j - l_j) \tag{1}$$

where $x_{i,j}$ means the j-th dimension of an individual $\vec{x}_i$ ($i \in \{1,2,..,P\}$ and $j \in \{1,2,..,D\}$); P is the population size, and D is the number of features in a dataset. Meanwhile, $\mathrm{rand}(0,1)$ is a random number uniformly distributed between 0 and 1, and l_j and u_j are the lower and upper bounds of the j-th dimension, respectively.

*2) **Mutation**:* Different mutation operators have been proposed by many researchers. Five commonly used mutation strategies are shown.

DE/rand/1, as shown in Eq. (2)

$$\vec{v}_i = \vec{x}_{r_1} + F \times (\vec{x}_{r_2} - \vec{x}_{r_3}) \tag{2}$$

DE/best/1, as shown in Eq. (3)

$$\vec{v}_i = \vec{x}_{\mathrm{best}} + F \times (\vec{x}_{r_1} - \vec{x}_{r_2}) \tag{3}$$

DE/current-to-rand/1, as shown in Eq. (4).

$$\vec{v}_i = \vec{x}_i + \mathrm{rand}(0,1) \times (\vec{x}_{r_1} - \vec{x}_i) + F \times (\vec{x}_{r_2} - \vec{x}_{r_3}) \tag{4}$$

DE/rand/2, as shown in Eq. (5).

$$\vec{v}_i = \vec{x}_{r_1} + F \times (\vec{x}_{r_2} - \vec{x}_{r_3}) + F \times (\vec{x}_{r_4} - \vec{x}_{r_5}) \tag{5}$$

DE/best/2, as shown in Eq. (6).

$$\vec{v}_i = \vec{x}_{\mathrm{best}} + F \times (\vec{x}_{r1} - \vec{x}_{r_2}) + F \times (\vec{x}_{r_3} - \vec{x}_{r_4}) \tag{6}$$

where F is a scaling factor, and $r_1 \neq r_2 \neq r_3 \in \{1,2,..,P\}$; $\vec{v}_i$ is the mutant vector of $\vec{x}_i$. Meanwhile, $\vec{x}_{\mathrm{best}}$ means the individual in the first Pareto front and has the lowest training error in the whole population. If $v_{i,j}$ (the j-th dimension of the mutant vector $\vec{v}_i$) is smaller than the lower boundary l_j or larger than the upper boundary u_j, it will be reset to l_j or u_j.

*3) **Crossover**:* After the mutation operation, the crossover operator is performed on the mutant vector $\vec{v}_i$ and individual $\vec{x}_i$ to produce a trial vector $\vec{u}_i$.

$$u_{i,j} = \begin{cases} v_{i,j} & \text{if } (\mathrm{rand}(0,1) \leq CR) \text{ or } (j = j_{\mathrm{rand}}) \\ x_{i,j} & \text{otherwise} \end{cases} \tag{7}$$

where j_{rand} is random integer between 1 and D. D is the dimension of individual $\vec{x}_i$, and CR ($CR \in [0,1]$) represents crossover probability.

*4) **Environmental Selection**:* Environmental selection is important in an EMO-based feature selection method. The duplicated feature subsets in the solution space are firstly recognized and removed since they will decrease the population diversity. The method to remove the duplicated solutions is followed by the study from [1]. The individual that has the largest sum of the confident rate (i.e., SCR) will be kept among the multiple duplicated feature subsets, and the others will be removed.

$$CR(j) = \begin{cases} \frac{f_j - \theta}{1-\theta} & \text{if } f_j > \theta \\ \frac{\theta - f_j}{\theta} & \text{if } f_j \leq \theta \end{cases} \tag{8}$$

where $CR(j)$ represents the confidence rate on the j-th feature, and f_j is the position entry in the j-th dimension,

M. Zhang, B. Xue and P. Wang are with the Evolutionary Computation Research Group, Victoria University of Wellington, Wellington 6140, New Zealand.

J. Liang is with the School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China.

```
Algorithm 1: The procedure of MODE
    Step 1: Initialization
        Randomly generate $N$ individuals,
    Step 2: Mutation
        Mutant vector is constructed,
    Step 3: Crossover
        Determine trial vector followed by Eq. (7),
    Step 4: Environmental Selection
        1) Remove the duplicated solutions;
        2) Choose individuals to enter into the next
           generation based on Pareto-dominance
           and crowding estimation,
    Step 5: Termination criteria
        Terminate if preset condition has been reached,
        otherwise go to step 2.
```

$j \in \{1, 2, \ldots, D\}$. Meanwhile, θ is the threshold to determine whether a feature is chosen or not.

$$SCR(\vec{x}) = \sum_{i=1}^{N} CR(j) \tag{9}$$

where $SCR(\vec{x})$ means the sum of the confidence rate of the individual $\vec{x}$. The larger the SCR, the more confidence to choose the corresponding feature(s). After removing the duplicated solutions, the Pareto-dominance is used to select individuals to enter into the next generation. If there are many individuals that are sitting in the same front, the crowding distance calculated in the objective space is used to break a tie. The Pareto-dominance and the calculations of crowding distance are the same as in [2].

III. Computational Complexity of NMDE

NMDE mainly includes six parts: initialization, niching-based mutation, crossover, evaluation, subset repairing scheme, and environmental selection strategy. Table I shows the complexities in the proposed NMDE method. Assume using a population with a size of N to solve a problem with M objectives and D decision variables. The overall complexity of the proposed NMDE method is $O(ND + MN^2)$.

The initialization, crossover, and evaluation operators in NMDE execute $O(N)$ basic operations, which can be finished in a linear time scale. Therefore, their computational complexities are $O(N)$. Since the Hamming distance has the complexity of $O(D)$, and that of the niching-based mutation in NMDE is $O(ND)$. The subset repairing scheme executes $O(m)$ basic operations to generate new individuals in m groups of the ψ-quasi equal feature subsets. For the environmental selection strategy, it uses non-dominated sorting and crowding distance. The complexity of the fast non-dominated sorting is $O(MN^2)$ [2]. The complexity of the calculation of the crowding distance in the objective space is $O(MN\log N)$ [2], and that of the crowding distance in the objective space is $O(ND)$. The crowding distance in the objective space is based on the Hamming distance which has been obtained when performing the niching-based mutation operator. Given the fact that $\log N < N$, the complexity of the environmental selection strategy is $O(MN^2)$.

TABLE I: Computational Complexities in NMDE

Initialization	$O(N)$
Niching-based Mutation	$O(ND)$
Crossover	$O(N)$
Evaluation	$O(N)$
Subset Repairing Scheme	$O(m)$
Environmental Selection	$O(MN^2)$
Whole Complexity	$O(ND + MN^2)$

In this work, $M = 2$, thus the whole complexity of NMDE is $O(ND + 2N^2)$.

IV. NMDE vs MOEA/D-DE

NMDE is compared with MOEA/D-DE since the offspring generation strategy of MOEA/D-DE [3] is also based on niching techniques. The average results of HV between MOEA/D-DE and NMDE on test sets are shown in Table S.II. For the feature subsets with the lowest training error rates from MOEA/D-DE and NMDE (in the 30 independent runs), their test accuracy and the number of selected features are shown in Tables S.III-S.IV, respectively. In each table, the highest HV, the largest classification accuracy, and the smallest number of selected features values obtained on each dataset are in bold. The signs '↑', '↓', and 'o' indicate that MOEA/D-DE is significantly better than, worse than or has no significant difference from NMDE, respectively, where the Wilcoxon test with a significance level of 0.05 is used. In addition, Tables S.II-S.IV give the summaries of '↑', 'o', and '↓' of MOEA/D-DE.

TABLE II: Average HV results between MOEA/D-DE and NMDE

	MOEA/D-DE	NMDE		MOEA/D-DE	NMDE
Zoo	8.043e-01 ±3.298e-02↓	**8.507**e-01 ±1.843e-02	CNAE	7.861e-01 ±1.149e-02↓	**8.759**e-01 ±1.228e-02
SPECT	7.624e-01 ±1.538e-02o	**7.665**e-01 ±8.440e-03	AD	9.336e-01 ±8.350e-03↓	**9.781**e-01 ±1.780e-03
WBCD	9.000e-01 ±2.452e-02↓	**9.155**e-01 ±7.100e-03	SRBCT	8.941e-01 ±2.196e-02↓	**9.781**e-01 ±2.434e-02
Ionosphere	8.609e-01 ±2.888e-02↓	**9.030**e-01 ±1.229e-02	Leukemia1	7.209e-01 ±5.580e-02↓	**9.171**e-01 ±5.117e-02
Sonar	7.961e-01 ±3.397e-02↓	**8.542**e-01 ±3.325e-02	DLBCL	8.759e-01 ±3.199e-02↓	**9.467**e-01 ±5.317e-02
Movement	7.671e-01 ±2.779e-02o	**7.783**e-01 ±1.537e-02	Leukemia2	8.407e-01 ±4.294e-02↓	**9.342**e-01 ±3.449e-02
Hillvally	5.819e-01 ±8.960e-03↓	**5.955**e-01 ±6.660e-03	11Tumor	7.246e-01 ±2.627e-02↓	**8.237**e-01 ±3.757e-02
Musk1	9.614e-01 ±8.220e-03↓	**9.760**e-01 ±2.930e-03	Lung Cancer	8.562e-01 ±2.042e-02↓	**9.272**e-01 ±1.794e-02
Multiple	9.333e-01 ±6.680e-03↓	**9.509**e-01 ±2.830e-03	14Tumor	4.781e-01 ±2.116e-02↓	**5.462**e-01 ±2.558e-02
Madelon	8.608e-01 ±1.420e-02↓	**9.009**e-01 ±4.640e-03	Sum	0 / 2 / 17	N/A

TABLE III: Average classification accuracy (%) between MOEA/D-DE and NMDE

	MOEA/D-DE	NMDE		MOEA/D-DE	NMDE
Zoo	87.47±3.87↓	**89.79**±1.20	CNAE	84.29±1.82↓	**86.26**±2.29
SPECT	74.47±2.44↓	**76.21**±2.34	AD	97.19±0.39↓	**97.46**±0.18
WBCD	93.51±0.96o	**93.87**±0.82	SRBCT	**98.14**±2.40↑	97.20±3.85
Ionosphere	87.09±2.76o	**88.27**±2.37	Leukemia1	77.78±7.23↓	**82.12**±7.38
Sonar	78.75±4.03↓	**83.05**±3.89	DLBCL	**96.44**±3.79↑	89.46±5.27
Movement	**76.88**±2.79o	76.72±2.24	Leukemia2	**91.24**±4.75o	89.01±5.88
Hillvally	54.03±1.20↓	**55.61**±1.13	11Tumor	77.53±3.85↓	**79.28**±4.68
Musk1	97.13±0.80↓	**98.00**±0.44	Lung Cancer	**92.49**±2.36o	91.38±3.03
Multiple	96.28±0.41o	**96.63**±0.54	14Tumor	46.56±2.18o	**47.65**±3.46
Madelon	87.16±1.24↓	**88.58**±1.00	Sum	2 / 7 / 10	N/A

As shown in Table S.II, the proposed NMDE method performs best on 17 datasets out of the used 19 datasets in terms of HV results. On the remaining two datasets: SPECT and Movement, MOEA/D-DE has a similar performance with NMDE. For the average of the classification accuracy in Table S.III, MOEA/D-DE has significantly better results than NMDE only on two datasets, i.e., the SRBCT and DLBCL datasets. However, as indicated by the results in Table S.IV, NMDE selects fewer features on most of the used datasets including the SRBCT and DLBCL datasets. On the Leukemia1 dataset, NMDE selects 150 times fewer features than MOEA/D-DE while still having a higher classification accuracy.

TABLE IV: Average number of selected features between MOEA/D-DE and NMDE

	MOEA/D-DE	NMDE		MOEA/D-DE	NMDE
Zoo	5.1±1.2o	**5.0**±0.7	CNAE	**228.4**±38.9↑	269.0±113.8
SPECT	**6.0**±1.9o	6.7±1.1	AD	173.8±94.2↓	**58.3**±46.9
WBCD	**4.9**±2.4o	5.1±1.3	SRBCT	244.3±34.4↓	**14.1**±16.4
Ionosphere	**3.8**±1.3o	**3.8**±0.7	Leukemia1	630.7±67.6↓	**3.8**±1.0
Sonar	10.3±3.9o	**9.8**±2.7	DLBCL	812.9±100.4↓	**8.9**±15.2
Movement	13.9±5.4o	**12.3**±5.8	Leukemia2	1315.4±195.5↓	**299.5**±462.1
Hillvally	**9.8**±3.4o	10.5±1.9	11Tumor	1933.6±497.2↓	**792.9**±451.8
Musk1	25.2±9.3o	**22.7**±6.7	Lung Cancer	1470.0±352.0↓	**363.9**±337.9
Multiple	**77.8**±17.9o	85.2±17.8	14Tumor	2138.2±282.8↓	**855.2**±459.9
Madelon	30.8±8.6↓	**13.5**±4.0	Sum	1 / 9 / 9	N/A

V. Feature Selection Performance Analysis

The average classification accuracy and the number of selected features from the eight methods, i.e., Omni-optimizer, MO_Ring_PSO_SCD, MMODE_ICD, SparseEA, SM-MOEA, DAEA, GF-NSGAII, and NMDE, are shown in Tables S.V-S.VI, respectively.

From Table S.V, the average accuracies of NMDE are superior to those of Omni-optimizer and MO_Ring_PSO_SCD on most of the used datasets. Furthermore, in Table S.VI, NMDE selects fewer features than the two methods in almost all cases, especially on the high-dimensional datasets. The highest dimensionality reduction can be seen on Leukemia1 where NMDE selects 400 times fewer features than Omni-optimizer and MO_Ring_PSO_SCD and still can improve the accuracy by 0.5% on average. The overall performance on the classification accuracy achieved by MMODE_ICD and NMDE are similar, but NMDE selects fewer features on 13 out of the used 19 datasets. SparseEA and GF-NSGAII show slightly worse accuracy results than NMDE. On several high-dimensional datasets, e.g., the SRBCT, DLBCL, Leukemia2, and Lung Cancer datasets, GF-NSGAII achieves significantly higher accuracy than NMDE. By regulating the granularity of the approximation of fitness values of the feature subsets, the grid-dominance help GF-NSGAII obtain feature subsets with lower classification error rate on the four datasets. However, on the four datasets, NMDE selects fewer features than GF-NSGAII. Although SM-MOEA shows the best size results of the obtained feature subsets, its accuracy results are worse than NMDE on most of the used datasets. On the high-dimensional datasets, DAEA selects fewer features than NMDE but achieves lower accuracy on the SRBCT and 11Tumor datasets. On four datasets, NMDE has significantly better accuracy than DAEA.

In summary, NMDE can generally achieve promising and excellent classification performance for feature selection in classification. More importantly, according to the results in Section V in the paper, NMDE can find more and better feature subsets with very similar or the same classification accuracy.

VI. Major Component Contribution Analysis

A. Algorithm Setting

To test the performance of the proposed mutation operator, the proposed environmental selection strategy, and the proposed subset repairing mechanism, some related details on the formed algorithms are shown as follows:

* **MODE-rand1** uses Eq. (2) as its mutation operator;
* **MODE-best1** uses Eq. (3) as its mutation operator;
* **MODE-current-rand1** uses Eq. (4) as its mutation operator;
* **MODE-rand2** uses Eq. (5) as its mutation operator;
* **MODE-best2** uses Eq. (6) as its mutation operator;
* **MODE-NM** employs the proposed mutation strategy to generate mutant vectors;
* **NMDE-N** means MODE-NM with the proposed environmental selection strategies;
* **NMDE** means NMDE-N with the proposed subset repairing mechanism.

The only difference between the six algorithms, i.e., MODE-rand1, MODE-best1, MODE-current-rand1, MODE-rand2, MODE-best2, and MODE-NM, is that they use different mutation operators.

The only difference between MODE-NM and NMDE-N is that MODE-NM uses the environmental selection strategy from [2], while NMDE-N uses the proposed environmental selection strategy.

The only difference between NMDE and NMDE-N is that NMDE uses the proposed subset repairing mechanism, while NMDE-N does not.

B. Environmental Selection Process in NMDE

In the following, two cases based on ten solutions (S_1-S_{10}) are given to show the process of the environmental selection in NMDE, which are shown in Figs. S.1-S.2, respectively. For the ten solutions, suppose that S_1-S_3 are ψ-quasi equal, S_4-S_6 are ψ-quasi equal, S_7-S_8 are ψ-quasi equal, and S_9-S_{10} are ψ-quasi equal feature subsets.

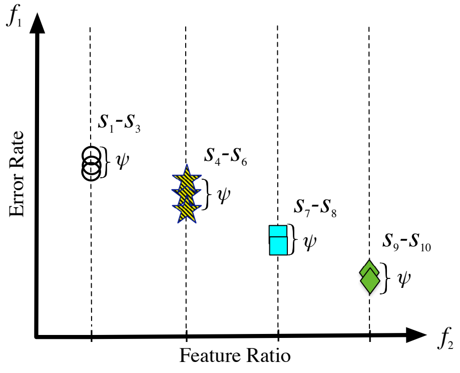

Fig. 1. Two possible situations of the ten solutions.

For the situation in Fig. S.1, all the ten solutions are in the relaxed first front. Under this situation, $|PO| = 10$, and $\text{unique\_number}(f_2(\mathcal{W}_1))$ is 4. Using Eq. (5) in the revised manuscript, we can get $ln = 10//4 = 2$. That means the maximal number of ψ-quasi equal feature subsets allowed to

TABLE V: Average classification accuracy (%) of the eight methods on test sets.

Dataset	Omni-optimizer	MO_Ring_ PSO_SCD	MMODE_ICD	SparseEA	SM-MOEA	DAEA	GF-NSGAII	NMDE
Zoo	**90.43**±5.06o	90.00±2.37o	89.41±1.64o	90.22±1.01o	86.99±5.23↓	**90.43**±0.57o	88.98±1.67↓	89.79±1.20
SPECT	73.27±2.77 ↓	75.60±1.72 o	75.23±2.01 o	74.88±2.22↓	75.31±1.46 o	75.33±0.19 o	75.01±2.85 o	**76.21**±2.34
WBCD	93.48±1.15 o	93.01±1.21↓	92.64±1.22 ↓	93.18±1.38↓	92.66±1.09↓	93.10±1.29↓	93.53±0.49↓	**93.87**±0.82
Ionosphere	81.77±3.01 ↓	89.24±3.00 o	88.95±3.48 o	**90.08**±3.37↑	88.43±4.64 o	89.22±2.12 o	88.14±2.49 o	88.27±2.37
Sonar	81.40±3.59 o	81.85±3.86 o	83.03±4.43 o	82.99±4.15 o	78.02±4.22↓	81.36±3.56 o	80.45±4764↓	**83.05**±3.89
Movement	76.43±1.92 o	75.87±2.03 o	76.20±2.25 o	76.65±2.05 o	71.74±3.59↓	76.39±1.48 o	73.87±2.44↓	**76.72**±2.24
Hillvally	53.14±1.68 ↓	54.18±1.56↓	54.73±1.44 ↓	54.57±1.43↓	53.68±1.00↓	54.11±1.21↓	55.47±1.34 o	**55.61**±1.13
Musk1	96.79±0.52 ↓	97.39±0.48↓	98.07±0.56 o	97.15±0.52↓	96.18±0.74↓	**98.18**±0.39↑	97.18±0.66↓	98.00±0.44
Multiple	96.30±0.44 o	96.53±0.45 o	96.40±0.27 o	96.35±0.41 o	94.77±1.08↓	96.36±0.48 o	91.71±2.82↓	**96.63**±0.54
Madelon	76.53±1.54 ↓	79.08±1.23↓	85.17±1.49 ↓	**88.82**±0.63 o	88.25±1.39 o	88.24±0.69 o	88.65±0.67 o	88.58±1.00
CNAE	79.29±2.26 ↓	83.69±2.12↓	87.52±1.52 ↑	85.27±1.81 o	82.50±3.65↓	**88.25**±1.34↑	86.25±1.34 o	86.26±2.29
AD	96.99±0.41 ↓	97.06 ±0.43↓	97.36±0.33 o	97.51 ±0.30o	96.79±0.29↓	**97.56**±0.17 o	97.53 ±0.31o	97.46±0.18
SRBCT	92.96±3.59 ↓	95.40 ±2.46↓	96.44±2.18 o	88.67 ±3.80↓	88.49 ±5.78↓	91.62 ±3.43↓	**99.20**±1.90↑	97.20 ±3.85
Leukemia1	77.22 ±7.65 ↓	77.80 ±6.00↓	80.83 ±5.49 o	**82.42** ±5.65o	80.18 ±7.12o	80.14 ±4.38o	81.82 ±7.96o	82.12 ±7.38
DLBCL	96.40±3.10 ↑	97.04 ±1.71↑	97.29 ±2.09↑	93.26 ±6.94↑	88.47 ±6.90o	88.79 ±5.26o	**97.92**±2.99↑	89.46 ±5.27
Leukemia2	88.64 ±2.62 o	88.25 ±1.26o	90.15 ±1.47 o	86.67 ±6.86o	86.51 ±6.34o	88.28 ±5.11o	**9136** ±2.71↑	89.01±5.88
11Tumor	74.68 ±3.13 ↓	74.42 ±2.48↓	76.14 ±2.67↓	73.53 ±4.56↓	72.18 ±6.51↓	74.07 ±4.50↓	**80.48** ±4.03o	79.28 ±4.68
Lung Cancer	93.10 ±1.30 ↑	89.34 ±2.32↓	**9362** ±1.18 ↑	88.52 ±3.48↓	88.09 ±3.83↓	90.82 ±2.04o	93.44 ±1.94↑	91.38 ±3.03
14Tumor	46.99 ±2.25 o	45.13 ±1.21↓	47.26 ±2.31 o	43.12 ±2.99↓	41.57 ±5.88↓	45.91 ±4.41o	47.49 ±2.11o	**47.65** ±3.46
Sum	2/7/10	1/7/11	3/12/4	2/9/8	0/6/13	2/13/4	4/9/6	N/A

TABLE VI: Average size of the obtained feature subsets from the eight methods.

Dataset	Omni-optimizer	MO_Ring_ PSO_SCD	MMODE_ICD	SparseEA	SM-MOEA	DAEA	GF-NSGAII	NMDE
Zoo	5.8±1.4↓	5.4±1.0o	4.9±0.6o	4.9±0.6o	**4.0**±1.2↑	5.2±0.4o	4.5±0.5↑	5.0±0.7
SPECT	8.3±1.5↓	7.4±1.1↓	6.5±0.7o	7.4±0.7↓	6.2±1.1o	7.2±0.7↓	**5.2**±1.3↑	6.7±1.1
WBCD	9.8±2.5↓	6.5±2.3↓	4.6±1.4o	5.6±2.7o	**2.1**±0.3↑	6.4±2.2↓	2.3±0.6↑	5.1±1.3
Ionosphere	7.8±2.7↓	3.5±0.6o	3.3±0.5↑	3.3±0.5↑	**2.1**±0.2↑	4.2±0.6↓	3.7±0.5o	3.8±0.7
Sonar	18.8±3.9↓	14.9±2.8↓	11.8±3.0↓	10.8±3.2o	**3.9**±0.9↑	12.3±3.3↓	6.3±1.3↑	9.8±2.7
Movement	29.2±6.3↓	26.1±4.1↓	21.7±5.5↓	10.3±3.1o	**5.6**±0.7↑	14.8±5.4o	6.3±1.1↑	12.3±5.8
Hillvally	33.5±6.3↓	26.8±4.7↓	16.1±6.6↓	9.2±2.0↑	**2.6**±0.5↑	10.3±3.2o	8.7±1.2↑	10.5±1.9
Musk1	60.2±8.4↓	54.5±5.8↓	39.3±6.6↓	27.0±14.1o	**8.2**±3.6↑	32.8±6.5↓	11.0±4.8↑	22.7±6.7
Multiple	88.8±8.4o	98.7±9.1↓	81.9±6.6o	92.7±19.4↓	35.2±10.8↑	81.3±10.2o	**20.1**±10.2↑	85.2±17.8
Madelon	177.2±15.2↓	171.5±10.6↓	71.6±15.5↓	12.0±4.1o	**5.9**±0.9↑	27.4±7.5↓	15.6±3.1↓	13.5±4.0
CNAE	326.9±16.4↓	370.7±26.6↓	284.6±20.8o	329.0±92.6o	120.5±97.7↑	275.0±40.5o	**83.9**±12.6↑	269.0±113.8
AD	574.8±22.1↓	603.3±23.2↓	494.2±38.1↓	54.1±50.1o	**10.5**±6.3↑	44.0±10.6o	171.4±26.5↓	58.3±46.9
SRBCT	866.1±28.3↓	865.7±34.5↓	742.4±61.7↓	8.1±8.1o	**5.0**±2.3↑	7.6±1.8↑	260.1±15.6↓	14.1±16.4
Leukemia1	2035.3±32.7↓	2052.7±73.9↓	1786.6±69.7↓	4.4±3.2o	**2.2**±0.5↑	3.0±1.1↑	1048.9±64.1↓	3.8±1.0
DLBCL	2682.7±54.0↓	2659.3±38.2↓	2320.6±75.9↓	62.3±189.6o	**2.4**±0.6↑	3.5±1.0o	1570.9±31.2↓	8.9±15.2
Leukemia2	4262.4±66.9↓	4014.5.1±99.8↓	3911.1±110.6↓	94.0±480.1o	**3.6**±1.1↑	8.0±14.1↑	2903.6±255.2↓	299.5±462.1
11Tumor	4905.4±57.0↓	4827.2±123.5↓	4696.7±134.5↓	1742.2±1345.1↓	**21.8**±48.2↑	62.7±47.3↑	3607.6±118.3↓	792.9±451.8
Lung Cancer	4935.6±71.8↓	4826.5±118.9↓	4588.9±161.3↓	229.4±348.7o	**5.9**±1.3↑	45.3±39.9↑	3394.9±197.6↓	363.8±337.9
14Tumor	5860.6±87.0↓	5925.7±251.2↓	5651.8±158.0↓	1784.6±2293.1↓	**19.9**±9.5↑	40.6±15.1↑	4520.1±99.8↓	855.2±459.9
Sum	0/1/18	0/2/17	1/5/13	2/13/4	18/1/0	6/7/6	9/1/8	N/A

select from each group is 2. Therefore, two solutions with a larger crowding distance from each group will be first selected. Eq. (6) in the revised manuscript shows the calculation of the crowding distance. Since the population size in this example is five, five solutions from the eight (2 × 4) ones will be selected using the same way to form the population for the next generation.

For the situation in Fig. S.2, some solutions, e.g., S_7-S_8, are dominated by other ψ-quasi equal feature subsets. Since there are still eight solutions ($8 > 5$) in the relaxed first front, the crowding distance of the eight solutions will be calculated by Eq. (6) in the revised manuscript. Then, five solutions will be selected based on their crowding distance in a descending way to form the population for the next generation. Under this situation, no solution is removed from each group (S_1-S_3, S_4-S_6, and S_9-S_{10}) in advance. The reason is under this situation, $\text{unique\_number}(f_2(\mathcal{W}_1)) = 3$ since S_1-S_3, S_4-S_6, and S_9-S_{10} in the relaxed first front have three unique FR (f_2) values. Therefore, $ln = 10//3 = 3$. The size of all the three groups (S_1-S_3, S_4-S_6, and S_9-S_{10}) is not larger than three. Therefore, the eight solutions are all first kept.

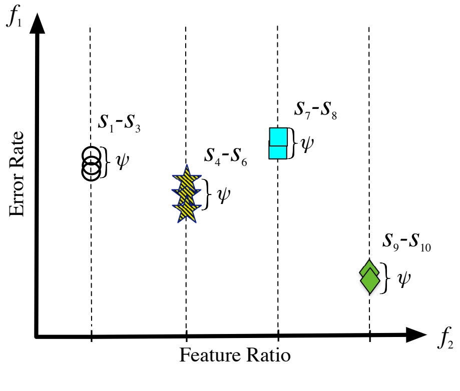

Fig. 2. Two possible situations of the ten solutions.

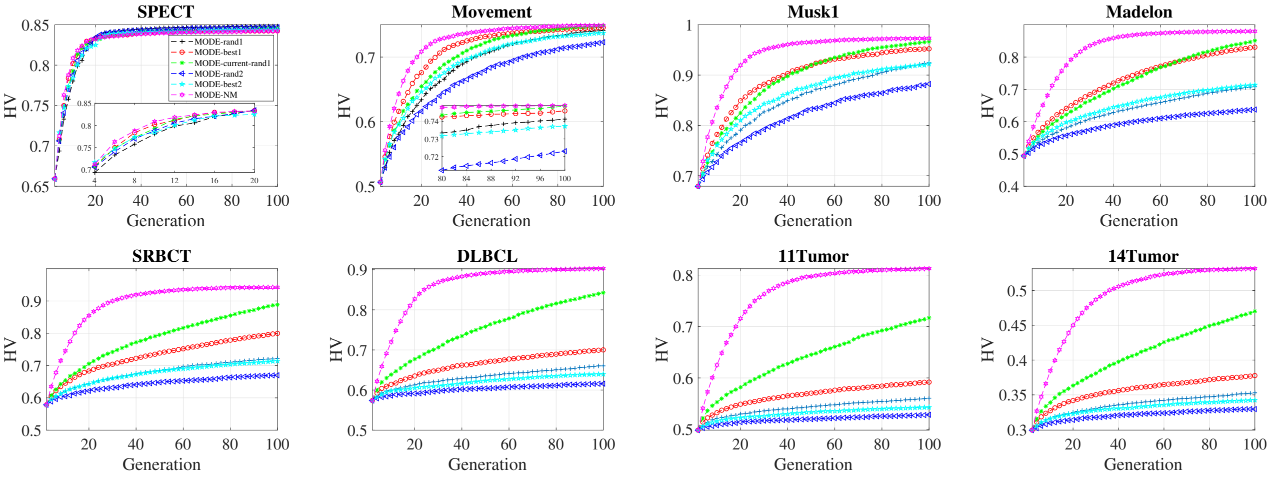

Fig. 3. The plots of the average HV of the population during the evolutionary process.

C. Results

Fig. S.3 gives the average HV plots with generations in different algorithms on eight training sets, including the SPECT, Movement, Musk1, Madelon, SRBCT, DLBCL, 11Tumor, and 14Tumor datasets. These eight datasets with different numbers of features included are selected as representatives, and the remaining datasets show the same patterns.

Tables S.VII and S.VIII show the average HV and IGD results on test sets, respectively. In both tables, '↑', '↓', and 'o' have the same meaning as the symbol in Tables S.III-S.IV. For relative performance rankings among all the compared algorithms, the Freidman test is adopted. In addition, the last row in Tables S.VII and S.VIII gives the summaries of '↑', 'o', and '↓' of each algorithm except for NMDE (reference).

D. Effect of the Proposed Mutation Mechanism

In Fig. S.3, MODE-NM achieves the fast convergence and the largest HV values than MODE-rand1, MODE-best1, MODE-current-rand1, MODE-rand2, and MODE-best2 on almost all the training sets. This trend is particularly obvious as the number of features increases in a dataset. As for the test performance, as shown in Tables S.VII and S.VIII, MODE-NM achieves significantly better rankings than MODE-rand1, MODE-best1, MODE-current-rand1, MODE-rand2, and MODE-best2 on both HV and IGD results. The superiority of MODE-NM in IGD is more obvious than that in HV.

The results show that the proposed mutation operator can produce better feature subsets during evolution and accelerate the convergence of the algorithm.

E. Effect of the Environmental Selection

The performance of the proposed environmental selection strategy can be seen from the comparison between MODE-NM and NMDE-N. As shown in Tables S.VII and S.VIII, on 14 out of the 19 datasets, NMDE-N can achieve larger HV and/or lower IGD performance than MODE-NM on the test sets. Only on the Movement dataset, MODE-NM gets a significantly larger HV value than NMDE-N, while NMDE-N has better IGD performance.

The results show that the proposed environmental selection strategy can help NMDE achieve better HV and IGD results.

F. Effect of the Subset Repairing Mechanism

The performance of the proposed subset repairing mechanism can be seen from the comparison between NMDE-N and NMDE. As shown in Tables S.VII and S.VIII, on ten out of the 19 datasets, NMDE achieves significantly larger HV and significantly lower IGD performance than NMDE-N on the test sets. On the remaining datasets, there is no significant difference between NMDE-N and NMDE. The results show that the proposed subset repairing mechanism by modifying the ψ-quasi equally feature subsets during evolution produces better feature subsets, and therefore achieves higher HV and lower IGD values in most of the 19 datasets.

VII. Effect of Different Learning Algorithms

Similar or the same feature selection performance can be influenced by different training algorithms. To implicitly investigate whether multiple optimal feature subsets still exist regardless of different classification algorithms used, we performed further experiments by using more different classification algorithms. Fig. S.4 gives the frequency matrix results of NMDE using four popular learning algorithms including k-nearest neighbors (KNN), support vector machines (SVM), decision tree (DT), and multi-layer perceptron classifier (MLP) on the SPECT dataset.

As shown in Fig. S.4, by using different learning algorithms, NMDE still can find that different feature subsets with the same size achieve the same classification accuracy. Furthermore, the two feature subsets $\{F_1, F_8, F_{22}\}$ and $\{F_8, F_{10}, F_{22}\}$ can achieve the same accuracy although different learning algorithms are used to evaluate their quality. Another interesting point is that feature F_4 can help the feature subset $\{F_8, F_{10}, F_{22}\}$ achieve better accuracy. Since by adding feature F_4 to $\{F_8, F_{10}, F_{22}\}$, the newly formed feature subset $\{F_4, F_8, F_{10}, F_{22}\}$ achieves better accuracy than $\{F_8, F_{10}, F_{22}\}$ by using SVM, DT, and MLP.

TABLE VII: Average HV results on the test sets (the larger the better).

Dataset	MODE-rand1	MODE-best1	MODE-current -rand1	MODE-rand2	MODE-best2	MODE-NM	NMDE-N	NMDE
Zoo	8.374e-01 ±1.050e-02$^{\downarrow}$	8.362e-01 ±1.070e-02$^{\downarrow}$	8.362e-01 ±9.200e-03 $^{\downarrow}$	8.391e-01 ±5.900e-03$^{\downarrow}$	8.346e-01 ±9.700e-03$^{\downarrow}$	8.324e-01 ±1.030e-02$^{\downarrow}$	8.372e-01 ±9.800e-03^{o}	**8.507e-01** ±1.843e-02
SPECT	7.663e-01 ±3.200e-03^{o}	7.673e-01 ±4.200e-03^{o}	7.663e-01 ±3.300e-03 ^{o}	7.669e-01 ±3.800e-03^{o}	7.671e-01 ±3.800e-03^{o}	7.650e-01 ±0.000e+00^{o}	**7.685e-01** ±5.100e-03^{o}	7.665e-01 ±8.440e-03
WBCD	9.163e-01 ±3.700e-03^{o}	9.139e-01 ±1.020e-02^{o}	9.150e-01 ±6.300e-03 ^{o}	**9.168e-01** ±3.500e-03^{o}	9.162e-01 ±3.700e-03^{o}	9.157e-01 ±5.000e-03^{o}	9.148e-01 ±6.400e-03^{o}	9.155e-01 ±7.100e-03
Ionosphere	8.877e-01 ±1.050e-02$^{\downarrow}$	8.885e-01 ±2.230e-02$^{\downarrow}$	8.890e-01 ±1.790e-02 $^{\downarrow}$	8.874e-01 ±9.600e-03$^{\downarrow}$	8.832e-01 ±1.810e-02$^{\downarrow}$	8.936e-01 ±2.170e-02^{o}	8.869e-01 ±2.020e-02$^{\downarrow}$	**9.030**e-01 ±1.229e-02
Sonar	8.477e-01 ±3.200e-02^{o}	8.423e-01 ±3.590e-02^{o}	8.562e-01 ±2.860e-02 ^{o}	**8.579e-01** ±2.400e-02^{o}	8.449e-01 ±3.020e-02^{o}	8.419e-01 ±2.620e-02^{o}	8.390e-01 ±3.140e-02^{o}	8.542e-01 ±3.325e-02
Movement	**7.811**e-01 ±1.150e-02^{o}	7.775e-01 ±1.380e-02^{o}	7.747e-01 ±1.320e-02 ^{o}	7.766e-01 ±1.770e-02^{o}	7.785e-01 ±1.580e-02^{o}	7.735e-01 ±2.020e-02^{o}	7.810e-01 ±1.010e-02^{o}	7.783e-01 ±1.537e-02
Hillvally	**5.991**e-01 ±8.400e-03^{o}	5.979e-01 ±9.300e-03^{o}	6.015e-01 ±6.500e-03 ^{o}	5.805e-01 ±7.700e-03$^{\downarrow}$	5.958e-01 ±8.500e-03^{o}	5.969e-01 ±9.900e-03^{o}	5.985e-01 ±7.800e-03^{o}	5.955e-01 ±6.660e-03
Musk1	9.280e-01 ±1.000e-02$^{\downarrow}$	9.495e-01 ±1.050e-02$^{\downarrow}$	9.625e-01 ±8.300e-03 $^{\downarrow}$	8.872e-01 ±1.170e-02$^{\downarrow}$	9.262e-01 ±9.500e-03$^{\downarrow}$	9.660e-01 ±7.000e-03$^{\downarrow}$	9.598e-01 ±1.320e-02$^{\downarrow}$	**9.760e-01** ±2.930e-03
Multiple	8.715e-01 ±8.700e-03$^{\downarrow}$	9.020e-01 ±1.160e-02$^{\downarrow}$	9.122e-01 ±9.500e-03 $^{\downarrow}$	8.353e-01 ±1.160e-02$^{\downarrow}$	8.632e-01 ±9.100e-03$^{\downarrow}$	9.347e-01 ±7.500e-03$^{\downarrow}$	9.210e-01 ±1.200e-02$^{\downarrow}$	**9.509e-01** ±2.830e-03
Madelon	7.255e-01 ±7.700e-03$^{\downarrow}$	8.336e-01 ±1.040e-02$^{\downarrow}$	8.562e-01 ±8.500e-03 $^{\downarrow}$	6.608e-01 ±8.300e-03$^{\downarrow}$	7.283e-01 ±1.250e-02$^{\downarrow}$	8.774e-01 ±8.000e-03$^{\downarrow}$	8.683e-01 ±1.800e-02$^{\downarrow}$	**9.009e-01** ±4.640e-03
CNAE	7.017e-01 ±7.700e-03$^{\downarrow}$	7.453e-01 ±9.800e-03$^{\downarrow}$	7.782e-01 ±6.600e-03 $^{\downarrow}$	6.680e-01 ±6.300e-03$^{\downarrow}$	6.931e-01 ±8.600e-03$^{\downarrow}$	8.194e-01 ±1.340e-02$^{\downarrow}$	8.048e-01 ±1.550e-02$^{\downarrow}$	**8.759e-01** ±1.228e-02
AD	7.363e-01 ±5.400e-03$^{\downarrow}$	7.829e-01 ±7.200e-03$^{\downarrow}$	8.387e-01 ±7.300e-03 $^{\downarrow}$	7.013e-01 ±4.100e-03$^{\downarrow}$	7.299e-01 ±6.700e-03$^{\downarrow}$	9.210e-01 ±8.800e-03$^{\downarrow}$	9.082e-01 ±9.100e-03$^{\downarrow}$	**9.781e-01** ±1.780e-03
SRBCT	7.455e-01 ±9.700e-03$^{\downarrow}$	8.099e-01 ±1.600e-02$^{\downarrow}$	8.980e-01 ±1.300e-02 $^{\downarrow}$	6.979e-01 ±1.080e-02$^{\downarrow}$	7.395e-01 ±7.600e-03$^{\downarrow}$	9.387e-01 ±1.630e-02$^{\downarrow}$	9.506e-01 ±2.110e-02$^{\downarrow}$	**9.781e-01** ±2.434e-02
Leukemia1	6.206e-01 ±3.980e-02$^{\downarrow}$	6.319e-01 ±3.590e-02$^{\downarrow}$	7.562e-01 ±5.120e-02 $^{\downarrow}$	5.810e-01 ±2.500e-02$^{\downarrow}$	5.966e-01 ±3.610e-02$^{\downarrow}$	7.512e-01 ±6.860e-02$^{\downarrow}$	8.108e-01 ±6.270e-02$^{\downarrow}$	**9.171e-01** ±5.117e-02
DLBCL	7.033e-01 ±3.400e-03$^{\downarrow}$	7.273e-01 ±1.040e-02$^{\downarrow}$	8.545e-01 ±1.150e-02 $^{\downarrow}$	6.666e-01 ±5.600e-03$^{\downarrow}$	6.809e-01 ±7.200e-03$^{\downarrow}$	8.909e-01 ±3.340e-02$^{\downarrow}$	9.345e-01 ±2.540e-02^{o}	**9.467e-01** ±5.317e-02
leukemia2	6.408e-01 ±1.320e-02$^{\downarrow}$	6.631e-01 ±1.520e-02$^{\downarrow}$	8.142e-01 ±2.010e-02 $^{\downarrow}$	6.064e-01 ±7.200e-03$^{\downarrow}$	6.237e-01 ±1.410e-02$^{\downarrow}$	8.460e-01 ±2.810e-02$^{\downarrow}$	9.247e-01 ±3.800e-02^{o}	**9.342e-01** ±3.449e-02
11Tumor	5.651e-01 ±1.520e-02$^{\downarrow}$	5.915e-01 ±2.160e-02$^{\downarrow}$	7.124e-01 ±2.460e-02 $^{\downarrow}$	5.399e-01 ±1.090e-02$^{\downarrow}$	5.556e-01 ±1.790e-02$^{\downarrow}$	7.377e-01 ±3.390e-02$^{\downarrow}$	7.779e-01 ±3.860e-02$^{\downarrow}$	**8.237e-01** ±3.757e-02
Lung Cancer	6.553e-01 ±6.000e-03$^{\downarrow}$	6.781e-01 ±8.300e-03$^{\downarrow}$	8.041e-01 ±1.180e-02 $^{\downarrow}$	6.259e-01 ±5.800e-03$^{\downarrow}$	6.370e-01 ±4.600e-03$^{\downarrow}$	8.550e-01 ±1.360e-02$^{\downarrow}$	8.937e-01 ±2.290e-02$^{\downarrow}$	**9.272e-01** ±1.794e-02
14Tumor	3.749e-01 ±9.300e-03$^{\downarrow}$	3.825e-01 ±1.350e-02$^{\downarrow}$	4.613e-01 ±1.560e-02 $^{\downarrow}$	3.627e-01 ±9.200e-03$^{\downarrow}$	3.669e-01 ±7.900e-03$^{\downarrow}$	4.805e-01 ±2.900e-02$^{\downarrow}$	5.026e-01 ±1.920e-02$^{\downarrow}$	**5.462e-01** ±2.558e-02
Rank	5.26 (0/5/14)	4.94 (0/5/14)	4.13 (0/5/14)	6.51 (0/4/15)	6.38 (0/5/14)	3.21 (0/6/13)	2.86 (0/8/11)	1.64

REFERENCES

[1] P. Wang, B. Xue, M. Zhang, and J. Liang, "A grid-dominance based multi-objective algorithm for feature selection in classification," in *IEEE Congr. Evol. Comput.*, 2021, pp. 2053–2060.

[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002.

[3] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated pareto sets, MOEA/D and NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 284–302, 2008.

TABLE VIII: Average IGD results on the test sets (the smaller the better).

Dataset	MODE-rand1	MODE-best1	MODE-current-rand1	MODE-rand2	MODE-best2	MODE-NM	NMDE-N	NMDE
Zoo	3.390e-02 ±1.480e-02↓	3.600e-02 ±1.610e-02↓	3.200e-02 ±1.410e-02 ↓	2.710e-02 ±1.030e-02↓	3.380e-02 ±1.460e-02↓	3.690e-02 ±1.280e-02↓	2.290e-02 ±1.610e-02o	**1.880e-02** ±1.630e-02
SPECT	9.980e-02 ±2.510e-02↓	1.014e-01 ±2.040e-02↓	9.750e-02 ±1.760e-02 ↓	9.780e-02 ±1.820e-02↓	9.970e-02 ±2.120e-02↓	1.102e-01 ±2.420e-02↓	9.640e-02 ±2.210e-02↓	**7.550e-02** ±3.380e-02
WBCD	1.670e-02 ±4.400e-03o	1.890e-02 ±5.400e-03↓	1.890e-02 ±3.200e-03 ↓	1.810e-02 ±3.200e-03↓	1.750e-02 ±3.400e-03↓	1.800e-02 ±3.400e-03↓	1.840e-02 ±3.500e-03↓	**1.400e-02** ±5.900e-03
Ionosphere	2.810e-02 ±7.900e-03o	3.050e-02 ±1.420e-02o	**2.730e-02** ±1.130e-02 o	2.850e-02 ±8.500e-03o	3.270e-02 ±1.180e-02o	2.960e-02 ±1.350e-02o	3.010e-02 ±1.310e-02o	3.050e-02 ±1.340e-02
Sonar	5.200e-02 ±1.860e-02o	5.690e-02 ±2.100e-02o	**4.970e-02** ±1.500e-02 o	5.270e-02 ±1.390e-02o	6.080e-02 ±1.910e-02o	5.700e-02 ±1.680e-02o	5.970e-02 ±2.050e-02o	5.520e-02 ±1.810e-02
Movement	3.810e-02 ±8.700e-03↓	3.710e-02 ±1.060e-02o	3.780e-02 ±8.200e-03 ↓	6.410e-02 ±1.840e-02↓	4.430e-02 ±2.040e-02↓	3.800e-02 ±7.900e-03↓	3.870e-02 ±1.440e-02o	**3.290e-02** ±7.300e-03
Hillvally	2.030e-02 ±6.000e-03↓	2.040e-02 ±5.200e-03↓	1.960e-02 ±4.200e-03 ↓	3.330e-02 ±8.100e-03↓	2.150e-02 ±5.900e-03↓	2.080e-02 ±5.900e-03↓	2.110e-02 ±4.900e-03↓	**1.660e-02** ±5.900e-03
Musk1	4.000e-02 ±5.400e-03↓	2.420e-02 ±4.700e-03↓	2.050e-02 ±3.800e-03 ↓	7.300e-02 ±9.200e-03↓	4.100e-02 ±6.500e-03↓	1.630e-02 ±3.100e-03↓	2.190e-02 ±5.800e-03↓	**1.120e-02** ±2.000e-03
Multiple	1.188e-01 ±1.030e-02↓	9.350e-02 ±1.240e-02↓	8.250e-02 ±1.290e-02 ↓	1.507e-01 ±9.600e-03↓	1.248e-01 ±9.400e-03↓	5.200e-02 ±1.740e-02↓	6.960e-02 ±1.750e-02↓	**1.720e-02** ±2.900e-03
Madelon	2.074e-01 ±7.500e-03↓	1.583e-01 ±6.200e-03↓	1.544e-01 ±5.000e-03 ↓	2.523e-01 ±9.500e-03↓	2.043e-01 ±1.330e-02↓	1.446e-01 ±6.600e-03↓	1.203e-01 ±2.970e-02↓	**4.380e-02** ±3.700e-02
CNAE	2.158e-01 ±9.800e-03↓	1.961e-01 ±1.090e-02↓	1.604e-01 ±1.120e-02 ↓	2.375e-01 ±8.200e-03↓	2.222e-01 ±9.200e-03↓	1.391e-01 ±1.290e-02↓	9.840e-02 ±1.920e-02↓	**3.710e-02** ±1.340e-02
AD	2.580e-01 ±5.300e-03↓	2.061e-01 ±7.900e-03↓	1.449e-01 ±7.700e-03 ↓	2.962e-01 ±4.300e-03↓	2.654e-01 ±7.500e-03↓	6.300e-02 ±8.100e-03↓	7.290e-02 ±9.200e-03↓	**9.300e-03** ±3.000e-03
SRBCT	3.155e-01 ±9.400e-03↓	2.611e-01 ±1.730e-02↓	1.960e-01 ±1.560e-02 ↓	3.559e-01 ±6.700e-03↓	3.236e-01 ±9.400e-03↓	1.722e-01 ±1.600e-02↓	1.311e-01 ±3.390e-02↓	**3.510e-02** ±1.820e-02
Leukemia1	3.355e-01 ±1.020e-02↓	2.910e-01 ±1.570e-02↓	1.714e-01 ±1.880e-02 ↓	3.741e-01 ±7.400e-03↓	3.522e-01 ±1.010e-02↓	1.519e-01 ±4.140e-02↓	1.016e-01 ±4.760e-02↓	**4.250e-02** ±3.700e-02
DLBCL	3.287e-01 ±3.000e-03↓	2.967e-01 ±4.900e-03↓	1.612e-01 ±6.100e-03 ↓	3.676e-01 ±3.300e-03↓	3.514e-01 ±3.500e-03↓	1.067e-01 ±9.600e-03↓	6.180e-02 ±1.460e-02↓	**3.760e-02** ±3.380e-02
leukemia2	3.442e-01 ±4.500e-03↓	3.148e-01 ±5.600e-03↓	1.640e-01 ±6.300e-03 ↓	3.805e-01 ±2.900e-03↓	3.641e-01 ±5.100e-03↓	1.257e-01 ±9.300e-03↓	5.240e-02 ±2.110e-02↓	**3.610e-02** ±2.320e-02
11Tumor	3.950e-01 ±7.000e-03↓	3.821e-01 ±9.700e-03↓	2.767e-01 ±1.190e-02 ↓	4.181e-01 ±5.000e-03↓	4.104e-01 ±5.200e-03↓	2.236e-01 ±2.660e-02↓	1.711e-01 ±1.620e-02↓	**5.960e-02** ±1.560e-02
Lung Cancer	3.418e-01 ±4.000e-03↓	3.153e-01 ±3.300e-03↓	1.799e-01 ±5.800e-03 ↓	3.734e-01 ±2.400e-03↓	3.620e-01 ±2.900e-03↓	1.254e-01 ±7.800e-03↓	8.720e-02 ±1.160e-02↓	**3.760e-02** ±1.160e-02
14Tumor	3.735e-01 ±5.500e-03↓	3.566e-01 ±4.600e-03↓	2.337e-01 ±7.300e-03 ↓	4.015e-01 ±4.000e-03↓	3.925e-01 ±5.100e-03↓	1.722e-01 ±1.580e-02↓	1.196e-01 ±1.470e-02↓	**2.480e-02** ±7.700e-03
Rank	5.21 (0/3/16)	5.01 (0/3/16)	3.61 (0/2/17)	6.84 (0/2/17)	6.63 (0/2/17)	2.99 (0/2/17)	2.26 (0/4/15)	1.65

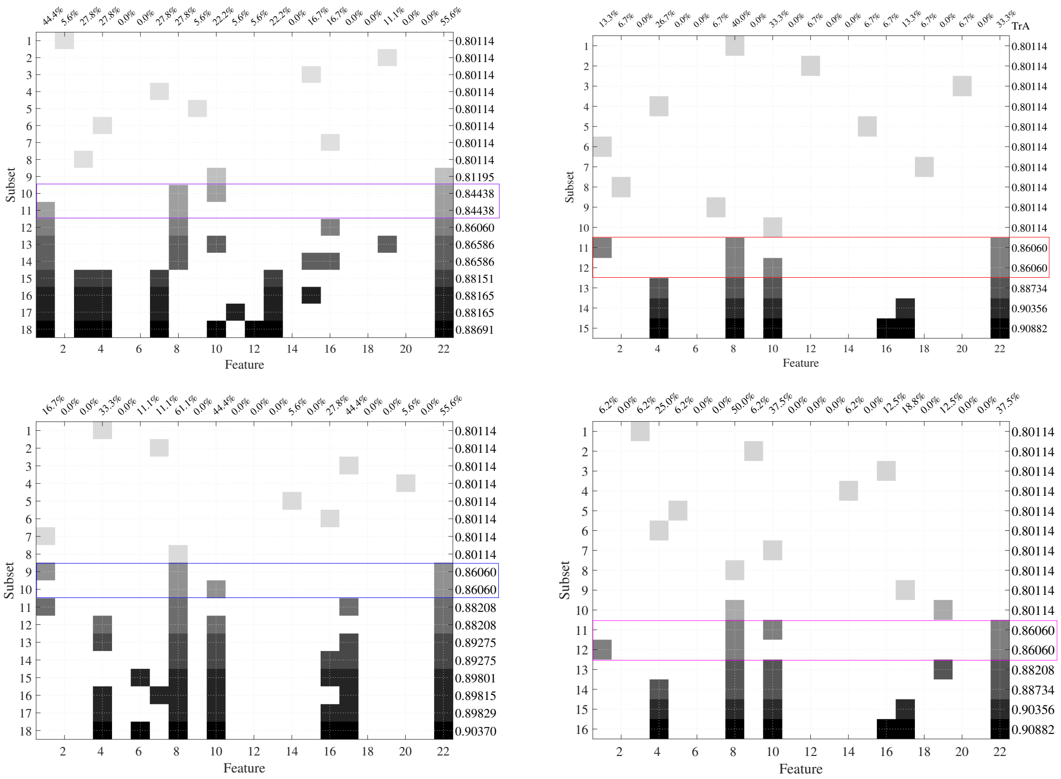

Fig. 4. Frequency matrix from NMDE with 5-NN (top left), SVM (top right), DT (bottom left), and MLP (bottom right) on the SPECT dataset.