Adaptive Bi-Directional Asymmetric Flip and Multi-Criteria Knowledge Transfer Strategy in Multitask Framework for High-Dimensional Feature Selection

Yikang Zhou
School of Computer and Information Technology
Shanxi University
Taiyuan, China
zhouyikang@sxu.edu.cn

Nan Li
School of Computer and Information Technology
Shanxi University
Taiyuan, China
linan10@sxu.edu.cn

Ruwang Jiao
School of Future Science and Engineering
Soochow University
Suzhou, China
ruwangjiao@gmail.com

***Abstract*—Evolutionary multi-task optimization (EMTO) has been widely applied to address the challenges of high-dimensional feature selection. However, existing EMTO algorithms still suffer from two key deficiencies: in the feature search process, the flipping strategy uses fixed probabilities and thresholds, failing to adapt to population dynamics during evolution; and in the knowledge transfer stage, the adaptive aggregation strategy randomly selects sources, ignoring particle value and task correlation, which may lead to negative transfer. To overcome these limitations, this paper proposes AAMCSO-AP, a dynamic adaptive aggregative multi-task competitive particle swarm optimization algorithm. It introduces two novel mechanisms: a population entropy and feature distribution-aware dynamic update strategy for winners, and a value assessment and relationship-aware knowledge transfer strategy for losers, enabling adaptive parameter adjustment and intelligent knowledge transfer. Experiments on seven real high-dimensional datasets (with up to 10,509 features) demonstrate that AAMCSO-AP achieves significantly superior classification accuracy and feature reduction compared to the original AAMCSO and other state-of-the-art algorithms, confirming the effectiveness of the proposed dynamic mechanisms.**

***Index Terms*—EMTO, High-dimensional Feature Selection, Dynamic parameter adaptation, Knowledge transfer, Particle swarm optimization**

I. Introduction

Feature selection (FS) is a crucial task in pattern recognition, machine learning, and data mining [1]. Its objective is to extract a compact and informative subset from the original feature space to reduce computational costs and mitigate the curse of dimensionality. Since a dataset with N features yields 2^N possible subsets, FS is inherently an NP-hard optimization problem.

FS methods are generally categorized into filter, wrapper, and embedded approaches. Filter methods are computationally efficient, while wrapper and embedded methods typically achieve higher predictive accuracy at the cost of increased complexity [2]. Evolutionary computation (EC) algorithms [3]–[7] have demonstrated strong capability in solving NP-hard FS problems, but high-dimensional settings remain challenging due to susceptibility to local optima and high computational overhead.

Evolutionary multitask optimization (EMTO) has emerged as a promising paradigm for FS by decomposing a high-dimensional problem into multiple related low-dimensional tasks and sharing knowledge among them. However, existing EMTO-based FS methods still suffer from deficiencies. Yang et al. considered only linear correlations in task generation, Li et al. [8] used non-adaptive weights for knowledge transfer, and Chen et al. [9] struggled to balance global search and computational efficiency. To address these issues, Zhang et al. [10] proposed the Adaptive Aggregated Multi-Task Competitive Swarm Optimization (AAMCSO) algorithm, which incorporates three innovative strategies: linear-and-nonlinear task generation (LTG), adaptive aggregative knowledge transfer (AAKT), and bidirectional asymmetric flip (BAF). AAMCSO achieved superior performance on eight high-dimensional datasets.

Nevertheless, our in-depth analysis reveals two critical limitations of the original AAMCSO:

- **Lack of adaptivity in BAF:** The flipping threshold

Corresponding author: Nan Li (linan10@sxu.edu.cn).

This work was supported in part by the National Natural Science Foundation of China (62406212), in part by the Science and Technology Program of Jiangsu Province (BK20251895), in part by the Suzhou Planning Project of Science and Technology (ZXP2025060), and in part by the China Postdoctoral Science Foundation (2025M781499).

$\delta = 0.6$ and flipping probability $P_f = 0.7$ are fixed throughout evolution, failing to adapt to the dynamic distribution of feature subsets across different stages. Relaxed thresholds are desirable in early exploration, while stricter criteria are needed in later exploitation.
- **Lack of transfer-awareness in AAKT:** Knowledge is aggregated by randomly selecting winner particles from other tasks, without considering task similarity or inter-task relevance. This may lead to negative transfer due to weakly related or misleading knowledge.

To overcome these limitations, this paper proposes a Dynamic Adaptive Aggregative Multitask Competitive Particle Swarm Optimization with Adaptive Parameters (**AAMCSO-AP**), which introduces two novel strategies:

1) A population entropy and feature distribution-aware **PEBAF** strategy that quantifies population diversity via Shannon entropy and dynamically adjusts flipping thresholds, enabling adaptive, feature-aware flipping.
2) A value assessment and relationship-aware knowledge transfer (**VARKT**) strategy that constructs a multi-criteria particle evaluation framework (potential, diversity, relevance) and selects high-value sources for adaptive aggregation, shifting knowledge transfer from random selection to value-guided learning.

Extensive experiments on eight high-dimensional datasets demonstrate that AAMCSO-AP consistently outperforms the original AAMCSO and other state-of-the-art methods in both classification accuracy and subset compactness, while exhibiting strong adaptability through dynamic parameter adjustment.

II. Background

The purpose of FS is to select the most relevant and informative subset from the original set of features, thereby reducing computational costs and addressing the curse of dimensionality. The FS problem for classification can be formalized as follows. Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ with N samples, where $\mathbf{x}_i \in \mathbb{R}^D$ is a D-dimensional feature vector and $y_i \in \{1, \dots, C\}$ is the class label, the objective is to find an optimal binary selection vector $\mathbf{s} \in \{0, 1\}^D$ that minimizes a predefined objective function:

$$\min_{\mathbf{s}} \mathcal{F}(\mathbf{s}) = \alpha \cdot \mathcal{L}(\mathbf{s}) + (1 - \alpha) \cdot \frac{\|\mathbf{s}\|_0}{D} \tag{1}$$

where $\|\mathbf{s}\|_0$ denotes the number of selected features, $\mathcal{L}(\mathbf{s})$ represents the classification error using only the selected features, and $\alpha \in [0, 1]$ balances accuracy and sparsity.

Evolutionary Multitask Optimization (EMTO) extends population-based evolutionary algorithms to simultaneously solve multiple optimization tasks. In FS, EMTO decomposes the high-dimensional problem into several low-dimensional related subtasks, enabling knowledge transfer across tasks to improve efficiency. Early methods generated tasks using statistical filters (e.g., variance, correlation), but knowledge transfer was often temporary and performance depended heavily on filter quality. More advanced approaches (e.g., MFCSO [8] and MBDPSO [11]) adopt a wrapper-based EMTO framework: they partition features based on relevance metrics, conduct population-based searches in each task, and allow cross-task knowledge sharing. However, these methods typically use fixed or empirically tuned transfer rates and weights that cannot adapt to changing task relationships during optimization. Although some recent works introduced adaptive elements, the adaptation was often limited to task creation or resource allocation, while the core search operators within tasks remained static.

III. AAMCSO-AP

A. Overview Framework

AAMCSO-AP is an enhanced version of the original AAMCSO algorithm, including two core dynamic mechanisms: a Population Entropy-based Bi-directional Asymmetric Flip (PEBAF) strategy and a Value Assessment and Relationship-aware Knowledge Transfer (VARKT) strategy. These mechanisms replace the static components of the original AAMCSO with adaptive, state-aware alternatives, enabling the algorithm to self-adjust during the evolutionary process based on real-time feedback from the population and task characteristics.

The LTG strategy decomposes the original high-dimensional FS problem into four low-dimensional, related subtasks. It employs two filters, PCC and MIC, and performs elbow point selection before initializing the subtasks. During the evolutionary loop phase, each particle is evaluated through the objective function and ranked according to fitness. Losers are updated via dynamic VARKT, while winners are updated through dynamic PEBAF. The loop will continue to execute until the maximum number of iterations is reached. Upon termination, each task will generate an optimal feature subset. The final solution is determined based on the highest classification accuracy; if there is a tie, the solution with the fewest features is selected.

B. Linear-and-Nonlinear-Correlation-based Task Generation Strategy

The Linear-and-Nonlinear-Correlation-based Task Generation (LTG) strategy serves as the foundation for decomposing the original high-dimensional FS problem into multiple tractable subtasks. As illustrated in Fig. 1, LTG constructs several low-dimensional search spaces that preserve diverse feature–label relationships, thereby facilitating effective multi-task optimization and knowledge transfer. Although LTG was originally proposed in AAMCSO [10], its core structure is retained in this work due to its proven capability in capturing heterogeneous relevance patterns. LTG employs two complementary statistical filters to evaluate feature relevance from both linear and nonlinear perspectives.

Pearson Correlation Coefficient (PCC) measures the strength of linear correlation between a feature vector $\mathbf{f} =$

$[f_1, \ldots, f_N]^T$ and the class label vector $\mathbf{y} = [y_1, \ldots, y_N]^T$, defined as

$$W_{\text{PCC}}(\mathbf{f}) = \left| \frac{\sum_{i=1}^{N}(f_i - \bar{f})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(f_i - \bar{f})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \right|, \tag{2}$$

where $\bar{f}$ and $\bar{y}$ denote the sample means.

Maximal Information Coefficient (MIC) is employed to characterize general associations, including nonlinear dependencies, between features and labels. It is defined as

$$W_{\text{MIC}}(\mathbf{f}) = \max_{|G|<B(N)} \frac{I[\mathbf{f}, \mathbf{y} \mid G]}{\log_2 \min(|G_f|, |G_y|)}, \tag{3}$$

where $I[\cdot \mid \cdot]$ denotes mutual information under grid partition G, G_f and G_y are the numbers of bins, and $B(N)$ is a sample-size-dependent bound (typically $N^{0.6}$ [12]).

Given a feature matrix $X \in \mathbb{R}^{N \times D}$ and label vector y, LTG proceeds as follows.

Step 1: Feature weighting. Compute W_{PCC} and W_{MIC} for all D features.

Step 2: Adaptive partitioning. For each weight set independently, features are sorted in descending order and divided into a high-relevance subset S_H and a low-relevance subset S_L using knee-point detection [13].

Step 3: Task construction. Four optimization tasks are generated:

- **PCC-High:** search over S_H^{PCC};
- **PCC-Mixed:** search over $S_H^{\text{PCC}} \cup R$, where R is a small random sample from S_L^{PCC};
- **MIC-High:** search over S_H^{MIC};
- **MIC-Mixed:** search over $S_H^{\text{MIC}} \cup R'$, where R' is randomly sampled from S_L^{MIC}.

C. VARKT Strategy

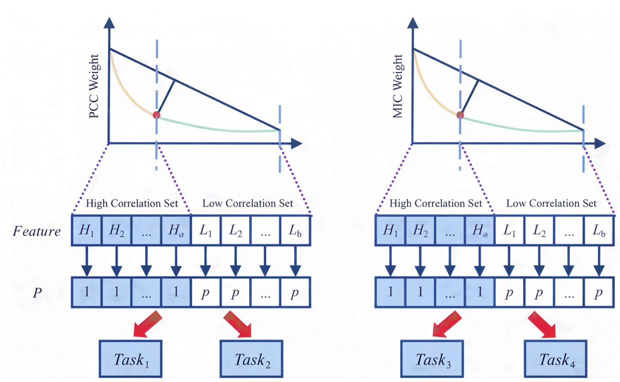

Fig. 1: LTG Strategic Work Plan

The Value Assessment and Relationship-aware Knowledge Transfer (VARKT) strategy addresses the limitation of random source selection in the original AAKT [10]. Instead of blindly aggregating winners from other tasks based solely on fitness, VARKT evaluates each candidate particle from a multi-criteria perspective, then selects the most valuable ones for knowledge transfer.

For a candidate particle c and the current loser l, we define a composite value score

$$V(c) = \lambda_1 \cdot \text{Potential}(c) + \lambda_2 \cdot \text{Diversity}(c, l) + \lambda_3 \cdot \text{Relevance}(c, l) \tag{4}$$

where $\lambda_1, \lambda_2, \lambda_3$ sum to 1, and

$$\text{Potential}(c) = \frac{f(l) - f(c)}{f(l)} \tag{5}$$

$$\text{Diversity}(c, l) = \frac{1}{D}\sum_{j=1}^{D} \left| b(c)_j - b(l)_j \right| \tag{6}$$

$$\text{Relevance}(c, l) = \frac{\mathbf{w}_t \cdot \mathbf{w}_{t(c)}}{\|\mathbf{w}_t\| \|\mathbf{w}_{t(c)}\|} \tag{7}$$

Here $f(\cdot)$ is the fitness function, $b(\cdot)$ the binary conversion, and $\mathbf{w}_t$ the feature weight vector of the current task. Higher values indicate that the candidate is more beneficial for updating the loser [1].

After computing $V(c)$ for all candidate winners, we select the top-K candidates (typically $K = 2$ or 3) while ensuring at least one from the current task and one from other tasks. If the top candidates are too similar, we replace some with lower-scoring but more diverse ones to maintain diversity [2].

The weight for each selected candidate c_i is determined by a softmax function

$$w_i = \frac{\exp(\beta\, V(c_i))}{\sum_{j=1}^{K} \exp(\beta\, V(c_j))} \tag{8}$$

and the aggregated knowledge is $A = \sum_{i=1}^{K} w_i\, c_i$. The learning rate η adapts as

$$\eta = \eta_{\text{base}} + \gamma \cdot \left(\frac{iter}{T} \right) \cdot \sigma\big(\{V(c_i)\}\big) \tag{9}$$

with η_{base} the base rate, γ a scaling factor, and $\sigma(\cdot)$ the standard deviation of the selected value scores. This encourages exploration early or when values are diverse, and exploitation later or when values are consistent [3].

Finally, the loser is updated following a PSO-like rule:

$$\begin{aligned} v_l' &= r_1 v_l + r_2(\text{best\_local} - l) + \eta\, r_3 (A - l) \\ l' &= l + v_l' \end{aligned} \tag{10}$$

where r_1, r_2, r_3 are random numbers in $[0, 1]$, and best_local is the best particle in the current task's population.

D. PEBAF Strategy

The Population Entropy-based Bi-directional Asymmetric Flip (PEBAF) strategy is proposed to overcome the limitation of fixed parameters in the original BAF strategy [10]. Inspired by Li et al.([14]–[21]), unlike the original BAF, which adopts constant flipping probability ($P_f = 0.7$) and binarization threshold ($\delta = 0.6$), PEBAF introduces a dynamic

[1]Detailed explanations of the three components are provided in the Supporting Materials.

[2]More details on the selection threshold and diversity maintenance are given in the Supporting Materials.

[3]The rationale behind the adaptive learning rate and its parameter settings are elaborated in the Supporting Materials.

adaptation mechanism that adjusts these parameters according to the evolutionary state of the population and the importance of individual features. Specifically, PEBAF consists of three tightly coupled components: (1) population entropy estimation for diversity characterization, (2) entropy- and importance-aware threshold adaptation, and (3) feature importance-guided flip probability adjustment.

Population entropy estimation. To quantify the diversity of FS patterns, Shannon entropy is employed under a binary selection framework. Given a population $\mathcal{P}$ containing N particles, each particle is first transformed into a binary vector using the current adaptive thresholds. For each feature j, the selection probability p_j is computed as

$$p_j = \frac{1}{N}\sum_{i=1}^{N} b_{ij}, \tag{11}$$

where $b_{ij} \in \{0,1\}$ indicates whether feature j is selected by particle i. The overall population entropy is then defined as

$$H = -\frac{1}{D}\sum_{j=1}^{D}\left[p_j \log_2(p_j+\epsilon) + (1-p_j)\log_2(1-p_j+\epsilon)\right], \tag{12}$$

where D denotes the number of features and ϵ is a small constant to avoid numerical instability. The normalization by D ensures that the entropy measure remains comparable across datasets with different dimensionalities.

The entropy value H provides a concise characterization of the evolutionary state of the population. A high entropy ($H \approx 1$) indicates a diverse population with balanced FS probabilities, suggesting that exploration should be emphasized. Conversely, a low entropy ($H \approx 0$) reflects population convergence, where exploitation and feature refinement become more desirable. Notably, this entropy-based indicator is obtained directly from population statistics, without relying on manually defined heuristics or stage-dependent thresholds.

In the original BAF strategy, all features share a fixed binarization threshold $\delta = 0.6$. In contrast, PEBAF introduces a feature-specific dynamic threshold δ_j, which adaptively varies according to the population entropy H and the feature importance weight fw_j. The threshold is defined as

$$\delta_j = 0.5 + \alpha_H(0.5-H) + \alpha_w(0.5-fw_j), \tag{13}$$

where 0.5 denotes the neutral midpoint of the binarization interval, α_H controls the influence of population diversity, and α_w regulates the impact of feature importance. The importance weight $fw_j \in [0,1]$ is computed as the normalized average of PCC and MIC scores.

When the population entropy is high ($H > 0.5$), δ_j decreases, encouraging feature exploration; when entropy is low ($H < 0.5$), δ_j increases, promoting feature pruning. Similarly, important features ($fw_j > 0.5$) are assigned lower thresholds to preserve discriminative information, whereas unimportant features receive higher thresholds and are more likely to be eliminated.

The original BAF employs a fixed flip probability ($P_f = 0.7$), which biases the search toward feature reduction. PEBAF replaces this with an adaptive flip probability that depends on population entropy and evolutionary progress:

$$P_f = P_{f,\text{base}} + \beta(0.5-H) + \gamma\left(\frac{iter}{T}\right), \tag{14}$$

where $P_{f,\text{base}}$ denotes the baseline probability, β adjusts entropy sensitivity, and γ controls iteration-dependent adaptation.

At early stages or under high entropy, P_f remains close to 0.5, balancing feature addition and removal; at later stages or under low entropy, P_f increases, favoring feature elimination. Moreover, PEBAF adopts an importance-guided flipping mechanism: when removing features, the least important selected feature is flipped ($\arg\min fw_j$), whereas when adding features, the most important unselected feature is activated ($\arg\max fw_j$).

E. Objective Function and Fitness Evaluation

The optimization process in AAMCSO-AP is guided by a carefully designed objective function that balances the dual objectives of FS: maximizing classification performance while minimizing the number of selected features.Formally, for a candidate solution represented by a binary selection vector $\mathbf{s} \in \{0,1\}^D$, the fitness value $\mathcal{F}(\mathbf{s})$ is computed as (15)

$$\mathcal{F}(\mathbf{s}) = \alpha \cdot (1 - \text{AC}(\mathbf{s})) + (1-\alpha)\cdot\frac{\text{NS}(\mathbf{s})}{D} \tag{15}$$

Where $\text{AC}(\mathbf{s})$ denotes the classification accuracy achieved by a classifier trained using only the features selected by $\mathbf{s}$.$\text{NS}(\mathbf{s}) = \|\mathbf{s}\|_0 = \sum_{j=1}^{D} s_j$ is the number of selected features. D is the total number of features in the original dataset.$\alpha \in [0,1]$ is a user-defined trade-off parameter that controls the relative importance of classification accuracy versus feature sparsity. Following established practices in high-dimensional FS, we set $\alpha = 0.9$ for all experiments, prioritizing accuracy while maintaining a moderate pressure for compactness. By simultaneously optimizing for accuracy and parsimony, this objective function guides the dynamic search strategies (VARKT and PEBAF) toward discovering compact, high-performing feature subsets—the ultimate goal of any FS task.

IV. Experimental Studies

A. Experimental Settings

TABLE I: Basic information of 7 datasets.

Dataset	Features	Instances	Classes
Colon	2000	62	2
SRBCT	2308	83	4
9_Tumor	5726	60	9
Leukemia_1	5327	72	3
Leukemia_2	7129	72	3
DLBCL	5469	77	2
Lung_Cancer	12600	203	5

To comprehensively evaluate the proposed AAMCSO-AP, extensive experiments are conducted on several publicly available high-dimensional datasets. The experimental design focuses on three key aspects: (1) comparing AAMCSO-AP with state-of-the-art FS algorithms in terms of classification accuracy and subset compactness; (2) investigating the contribution of the proposed dynamic mechanisms (VARKT and PEBAF) through ablation studies; and (3) examining the adaptive behavior of dynamic parameters (e.g., P_f, δ, and learning rate η) during the evolutionary process.

Seven benchmark datasets from bioinformatics and image processing domains are employed, characterized by extremely high dimensionality and limited sample sizes. These datasets present substantial challenges due to the curse of dimensionality, class imbalance, and complex nonlinear feature interactions. Detailed dataset statistics are reported in Table I.

For fair comparison, AAMCSO-AP is evaluated against representative single-task FS algorithms (CSO [22], LDW-PSO [23]), state-of-the-art multitask FS algorithms (MTPSO, MFCSO [8], MBDPSO [11], and the original AAMCSO [10]), as well as several classical filter methods, including ReliefF [24], TV [25], PCC [26], and MIC [12]. The performance obtained using all features (FULL) is also reported as a reference.

Each algorithm is independently executed 10 times on each dataset with 10-fold stratified cross-validation. Statistical significance is assessed using the Wilcoxon signed-rank test at the 0.05 level. A k-NN classifier ($k = 5$) is adopted as the fitness evaluator for all methods. The population size per task is set to 200 with 100 iterations, resulting in 20,000 fitness evaluations. Parameters α_H, α_w, β, and γ are determined via preliminary sensitivity analysis, while other baseline algorithms follow their original parameter settings.

B. Comparison Studies

1) Comparison of Classification Accuracy: Table II reports the average classification accuracies. AAMCSO-AP achieves the best accuracy on four datasets (Colon, 9_Tumor, Leukemia 1, DLBCL), with improvements of 4–10% over the original AAMCSO. It attains the best mean rank (3.3) and shows statistically significant gains on four datasets, with no significant performance degradation elsewhere. The slightly lower accuracy on SRBCT and Leukemia 2 correlates with higher feature counts, reflecting the inherent accuracy–sparsity trade-off[4].

2) Comparison of Selected Feature Numbers: AAMCSO-AP obtains the best mean rank (tied with MFCSO at 2.8) and identifies the most compact subsets on 9_Tumor and Leukemia 2. It selects fewer features than the original AAMCSO on most datasets, confirming the effectiveness of the PEBAF strategy. Slightly higher feature counts on SRBCT and Leukemia 1 align with the accuracy observations[5].

[4]Detailed per-dataset analysis is provided in the Supporting Materials.

[5]Detailed feature count comparisons and case-specific analysis are available in the Supporting Materials.

3) Comparison with Filter Methods: To further assess the performance of our wrapper-based approach, we compare AAMCSO-AP against four classical filter methods. The results, presented in the Supporting Materials, show that AAMCSO-AP achieves the best mean rank (1.8) in classification accuracy while selecting an order-of-magnitude fewer features than any filter method. Notably, on the 9_Tumor dataset, it improves accuracy by over 36 percentage points compared to the best filter method (ReliefF), underscoring the advantage of wrapper-based multitask optimization in capturing complex feature interactions.

C. Ablation Analysis

Ablation experiments confirm the complementary roles of PEBAF and VARKT. The full AAMCSO-AP consistently achieves superior or competitive accuracy across all datasets, especially on the challenging Leukemia_1 dataset. While the variants occasionally select fewer features, this often comes at the cost of reduced accuracy, demonstrating that the dynamic mechanisms effectively balance sparsity and accuracy[6].

V. Conclusion and Future Work

This paper presents AAMCSO-AP, an enhanced EMTO algorithm for high-dimensional FS. By integrating the LTG, PEBAF, and VARKT strategies, the proposed method alleviates the limitations of static parameter control and indiscriminate knowledge transfer in the original AAMCSO. Experimental results on seven high-dimensional datasets demonstrate superior classification accuracy and competitive feature sparsity compared with state-of-the-art methods, while ablation studies confirm the effectiveness of the proposed dynamic mechanisms. Future work will focus on structured knowledge transfer beyond the particle level and improving scalability for ultra-high-dimensional feature spaces.

The supplementary materials are available at: https://github.com/zhou2006-max/IJCNN--Supporting_materials

References

[1] T. Yang *et al.*, "Local feature selection for large-scale data sets with limited labels," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 7152–7163, 2023.

[2] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[3] Y. Jiang *et al.*, "Knowledge learning for evolutionary computation," *IEEE Transactions on Evolutionary Computation*, 2023. [Online]. Available: https://doi.org/10.1109/TEVC.2023.3278132

[4] J.-Y. Li *et al.*, "Multiple tasks for multiple objectives: A new multiobjective optimization method via multitask optimization," *IEEE Transactions on Evolutionary Computation*, Jul. 2023. [Online]. Available: https://doi.org/10.1109/TEVC.2023.3294307

[5] X.-F. Liu *et al.*, "Interaction-based prediction for dynamic multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 6, pp. 1881–1895, 2023.

[6] Z.-H. Zhan *et al.*, "Matrix-based evolutionary computation," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 2, pp. 315–328, 2022.

[6]Additional ablation results and statistical significance tests are provided in the Supporting Materials.

TABLE II: Classification Accuracy Comparison of Different Algorithms on 6 Datasets

Dataset	FULL	LDW-PSO	CSO	MTPSO	MFCSO	MBDPSO	AAMCSO	AAMCSO-AP
Colon	74.76(+)	75.43(+)	77.40(+)	81.67(+)	82.29(+)	77.88(+)	83.03(+)	**93.65**
SRBCT	88.96(-)	93.56(-)	**98.57**(-)	95.37(-)	95.77(-)	95.41(-)	96.55(-)	83.33
9_Tumor	41.05(+)	39.22(+)	44.17(+)	39.44(+)	44.67(+)	38.00(+)	43.72(+)	**78.17**
Leukemia 1	88.14(+)	89.55(+)	93.46(+)	87.76(+)	91.67(+)	88.69(+)	94.05(+)	**98.15**
Leukemia 2	84.40(-)	84.20(-)	86.99(-)	83.45(-)	86.63(-)	82.36(-)	**87.36**(-)	75.05
DLBCL	88.43(+)	89.30(+)	92.55(+)	92.20(+)	91.65(+)	89.52(+)	92.57(+)	**99.17**
+/≈/-	4/0/2	4/0/2	4/0/2	4/0/2	4/0/2	4/0/2	4/0/2	NA
Mean Rank	6.1	5.1	3.7	5.1	4.0	4.5	3.9	**3.3**

TABLE III: Average Feature Number Selected by Different Algorithms on 6 Datasets

Dataset	FULL	LDW-PSO	CSO	MTPSO	MFCSO	MBDPSO	AAMCSO	AAMCSO-AP
Colon	2000(+)	893.68(+)	428.36(+)	144.66(+)	235.86(+)	**43.58**(-)	108.75(+)	49.55
SRBCT	2308(+)	1024.17(+)	329.16(+)	195.81(-)	**35.84**(-)	124.30(-)	71.83(-)	233.58
9_Tumor	5726(+)	2709.54(+)	1752.24(+)	285.17(+)	125.08(+)	703.53(+)	323.00(+)	**114.10**
Leukemia 1	5327(+)	2444.32(+)	1216.84(+)	333.74(+)	296.59(+)	410.16(+)	**205.30**(-)	231.85
Leukemia 2	7129(+)	3272.79(+)	1832.09(+)	353.11(+)	530.50(+)	612.91(+)	328.50(+)	**162.00**
DLBCL	5469(+)	2431.29(+)	1208.11(+)	137.00(+)	**90.36**(-)	228.60(+)	285.92(+)	102.63
+/≈/-	6/0/0	6/0/0	6/0/0	5/0/1	4/0/2	4/0/2	4/0/2	NA
Mean Rank	8.0	7.0	5.8	4.8	2.8	3.8	4.3	**2.8**

TABLE IV: Comparisons between AAMCSO-AP and its two variants

Dataset	Dynamic_noPEBAF		Dynamic_noVARKT		AAMCSO-AP	
	Accuracy	Features	Accuracy	Features	Accuracy	Features
Colon	0.9310(≈)	44.00(-)	0.9220(+)	**42.00**(-)	**0.9365**	49.55
Leukemia_1	0.9122(+)	**197.00**(-)	0.9360(+)	206.00(-)	**0.9815**	231.85
Lung_Cancer	0.9390(+)	48.00(+)	0.9600(≈)	44.00(+)	**0.9670**	**43.00**
+/≈/-	2/1/0	1/0/2	2/1/0	1/0/2	NA	NA

[7] Z.-G. Chen *et al.*, "Evolutionary computation for intelligent transportation in smart cities: A survey," *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 83–102, 2022.

[8] L.-J. Li *et al.*, "An evolutionary multitasking algorithm with multiple filtering for high-dimensional feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 4, pp. 802–816, 2023.

[9] K. Chen *et al.*, "An evolutionary multitasking-based feature selection method for high-dimensional classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 7172–7186, 2022.

[10] Y. Zhang *et al.*, "Adaptive aggregative multitask competitive particle swarm optimization with bi-directional asymmetric flip strategy for high-dimensional feature selection," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '24)*, 2024, pp. 1515–1523.

[11] J.-Q. Yang *et al.*, "Evolutionary multitasking bi-directional particle swarm optimization for high-dimensional feature selection," in *Proceedings of the 2023 IEEE Congress on Evolutionary Computation (CEC)*, 2023, pp. 1–8.

[12] D. N. Reshef *et al.*, "Detecting novel associations in large data sets," *Science*, vol. 334, pp. 1518–1524, 2011.

[13] I. Das, "On characterizing the 'knee' of the pareto curve based on normal-boundary intersection," *Structural Optimization*, vol. 18, pp. 107–115, 1999.

[14] Y. Zhang, R. Jiao, B. Xue, and M. Zhang, "Evolutionary streaming feature selection via incremental feature clustering," *IEEE Transactions on Evolutionary Computation*, 2026, early Access.

[15] Z. Chen, Q. Fan, R. Jiao, B. Xue, H. Huang, Y. Dai, and M. Zhang, "Multi-expert genetic programming based ensemble for long-tailed image classification," *IEEE Transactions on Evolutionary Computation*, 2026.

[16] N. Li, L. Ma, G. Yu, B. Xue, M. Zhang, and Y. Jin, "Survey on evolutionary deep learning: Principles, algorithms, applications, and open issues," *ACM Computing Surveys*, vol. 56, no. 2, pp. 41:1–41:34, 2024.

[17] N. Li, B. Xue, M. Zhang, and L. Ma, "Automatic fuzzy architecture design for defect detection via classifier-assisted multiobjective optimization approach," *IEEE Transactions on Evolutionary Computation*, 2024, major revision.

[18] T. Zhang, L. Ma, S. Cheng, Y. Liu, N. Li, and H. Wang, "Automatic prompt design via particle swarm optimization driven llm for efficient medical information extraction," *Swarm and Evolutionary Computation*, vol. 95, p. 101922, 2025.

[19] N. Li, B. Xue, L. Ma, and M. Zhang, "Transferable relativistic predictor: Mitigating cross-task cold-start issue in nas," in *Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI 2025)*, 2025, pp. 5672–5680.

[20] A. Mei, N. Li, L. Ma, K. Gao, M. Huang, B. Xue, and Y. Jin, "Binary-to-decimal encoding-based performance predictor for evolutionary graph fusion architecture search and applications to medical data," *IEEE Transactions on Evolutionary Computation*, 2024, accepted.

[21] T. Zhang, J. Cheng, L. Miao, H. Chen, Q. Li, and Q. He, "Multi-hop reasoning with relation based node quality evaluation for sparse medical knowledge graph," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 1130–1142, 2025.

[22] S.-K. Gu, R. Cheng, and Y.-C. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," *Soft Comput*, vol. 22, no. 3, pp. 811–822, 2018.

[23] Y.-H. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, Anchorage, AK, USA, 1998, pp. 69–73.

[24] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine Learning*, vol. 53, pp. 23–69, 2003.

[25] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112–123, 2014.

[26] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.