Multiscale-adaptive and Size-adaptive PSO-based Feature Selection for Gene Expression Analysis

Weihao Deng[1] , **Lingyun Zhao**[1] , **Fei Han**[1*]

[1]School of Computer Science and Communication Engineering, JiangSu University, ZhenJiang, China
hanfei@ujs.edu.cn

Abstract

In gene expression analysis, high dimension low sample size data limits the applicability of deep learning, motivating increasing interest in variable-length evolutionary algorithm–based feature selection methods (VLEAs). However, existing VLEAs suffer from unreliable single-metric discrimination under dynamic search-space variation, mismatches between population size and search space dimensionality, and particle performance degradation after search space changes. To this end, a Multiscale-adaptive and Size-adaptive Particle Swarm Optimization (MASA-PSO) is proposed for gene expression analysis. MASA-PSO adopts a multiscale-adaptive weighting framework to explore feature subsets that distinguish between-class sample distributions during search spaces changes, and theoretically proves it enables the collaborative evaluation of multiple metrics. Meanwhile, it proposes an adaptive population division that explicitly models the functional relationship between the population size and the search space to resolve the mismatch. Furthermore, a particle degradation phenomenon impairing the performance of VLEAs is observed and alleviated through a hybrid elite strategy. Experiments on ten gene expression datasets verify that MASA-PSO outperforms state-of-the-art methods in classification accuracy while capturing smaller feature subsets.

1 Introduction

Gene expression analysis and cancer diagnosis are representative and widely studied problems in the bioinformatics and medical research [Ravindran and Gunavathi, 2023]. These tasks typically involve numerous features, and the high cost acquisition severely limits the sample sizes, resulting in high-dimension low-sample size (HDLSS) datasets[Liu *et al.*, 2017] and the curse of dimensionality [Chattopadhyay and Lu, 2019]. Despite the success of deep learning (DL) in vision and language domains, significant challenges remain in this fields [Noor and Ige, 2025]. The characteristics of HDLSS datasets make it difficult for most DL approaches based on manifold assumptions to learn informative and interpretable representations [Loaiza-Ganem *et al.*, 2024]. Moreover, the class imbalance in biological data further increases the risk of overfitting [Min *et al.*, 2017].

In this context, feature selection (FS), particularly evolutionary algorithm (EA) methods, receives considerable attention due to its ability to significantly reduce dimensionality, enhance interpretability, and avoid manifold assumptions [Xue *et al.*, 2015]. Among EAs, particle swarm optimization (PSO) has been widely applied to optimization tasks as its fast convergence and robustness in high-dimensional spaces [Bonyadi and Michalewicz, 2017]. However, EAs often suffer from high computational and memory costs in HDLSS biological datasets. To address this, Tran et al. [Tran *et al.*, 2018] proposed VLPSO, which dynamically adjusts the search space via length-adaptive and mutual learning mechanisms, enabling more efficient discovery of compact feature subsets with better classification performance. It has inspired subsequent variable-length mechanisms EAs (VLEAs) in bioinformatics and medical applications. For instance, Qu et al. [Qu *et al.*, 2023] designed a PSO with the explicit feature representation (ESAPSO) to further reduces computational cost and memory usage. Nonetheless, existing VLEAs [Tran *et al.*, 2018; Qu *et al.*, 2023; Zhou *et al.*, 2022] typically construct different search space sizes through feature grouping, while overlooking the mismatch between population size and space size. A subgroup with the excessively small search space may impair collaborative search and miss the global optimum, while a subgroup with too large search space may lead to redundant exploration of similar regions, resulting in wasted search resources and increased computational cost [Antonio and Coello, 2013].

In addition, due to the curse of dimensionality, most existing EAs including VLEAs commonly adopt Manhattan distance to measure sample distances to capture differences in distributions among different labels [Tran *et al.*, 2018; Qu *et al.*, 2023; Antonio and Coello, 2013]. However, the effect of a single distance metric is less desirable for HDLSS FS problems [Happy *et al.*, 2017]. During the variable-length process, the sample distribution dynamically changes with the search space, making single distance metrics like Manhattan distance potentially unreliable for consistent identifi-

*Corresponding author: Fei Han; Code and datasets are available at https://anonymous.4open.science/r/MASA-PSO-15FE/.

cation. [Huang *et al.*, 2023].

Remarkably, in the study of VLEAs, we observed that a certain number of particles experience a sudden drop in performance after search space changes. This phenomenon has not been explicitly reported in prior VLEAs.

To this end, a novel multiscale adaptive and size-adaptive PSO algorithm (MASA-PSO) for gene expression analysis is proposed. It not only includes an adaptive weighting discrimination framework but also resolves the mismatch between population size and search space. Additionally, a hybrid elite strategy is proposed to mitigate the degradation of particles. The main contributions are as follows:

(1) A novel multiscale-adaptive weighting framework is creatively proposed, which integrates three distance metrics to measure sample distributions for fitness evaluation, aiming at enhancing the algorithm's discrimination and stability.

(2) Designing an adaptive division mechanism that initializes population based on a function measuring the matching degree between the population and the search space, addressing the imbalance of search capabilities among groups.

(3) Discovering and analyzing the degradation of particles induced by search space variation, and proposing a hybrid strategy to alleviate this problem.

(4) Extensive experiments on ten HDLSS gene expression datasets show that MASA-PSO outperforms compared state-of-the-art (SOTA) methods.

2 Related Work

2.1 Variable-Length Particle Swarm Optimization

As an important branch of EAs [Bonyadi and Michalewicz, 2017], PSO iteratively updates a swarm of particles in a D-dimensional search space, where each particle represents a candidate solution characterized by a position and a velocity. The quality of each particle's current candidate solution is evaluated by a fitness function. Guided by its local best ($pbest$) and the global best ($gbest$) among the swarm, particles move toward the promising regions and find the optimal solution based on the following equations,

$$v_{i,d}^{t+1} = \omega v_{i,d}^{t} + c_1 r_1 (pbest_{i,d}^{t} - x_{i,d}^{t}) + c_2 r_2 (gbest_{d}^{t} - x_{i,d}^{t}) \tag{1}$$

$$x_{i,d}^{t+1} = x_{i,d}^{t} + v_{i,d}^{t} \tag{2}$$

where $x_{i,d}^{t}$ and $v_{i,d}^{t}$ are the d^{th} position and velocity of the i^{th} particle at time t, respectively. $pbest_{i,d}^{t}$ and $gbest_{d}^{t}$ are the local best position of the i^{th} particle and global best position for the whole swarm in dimension d at time t. c_1 and c_2 are two positive constants. ω is the inertia weight. r_1 and r_2 are two random values uniformly distributed ranging from 0 to 1.

VLPSO [Tran *et al.*, 2018] is the first length-adaptive PSO in high-dimensional FS, adopting the feature grouping from cooperative coevolution and a comprehensive learning strategy [Yu and Zhang, 2014] to accelerate convergence. Specifically, VLPSO relies on $pbest$ alone for particle guidance and introduces a self-adaptive probability. It achieves a smaller feature subset and better classification accuracy in less time. However, VLEAs[Tran *et al.*, 2018; Qu *et al.*, 2023] ignored the population size in division which led to inefficient utilization of search resources. Yazdani et al. [Yazdani *et al.*, 2023] suggested that there is a lack of a functional relationship between population and the search.

2.2 Distance metrics in high-dimensional FS

In high-dimensional FS, some EAs aim for higher accuracy and smaller feature subsets that introduced distance metrics. Most VLEAs [Tran *et al.*, 2018; Qu *et al.*, 2023] utilized the Manhattan distance as in Eq.(3).

$$Dst_{Man}(s_i, s_j) = \sum_{k=1}^{d} |s_i^k - s_j^k| \tag{3}$$

Yuan et al. [Dubey and Saxena, 2016] applied cosine distance for FS as Eq.(4), which is also noteworthy.

$$Dst_{Cos}(s_i, s_j) = 1 - \frac{\sum_{k=1}^{d} s_i^k \cdot s_j^k}{\sqrt{\sum_{k=1}^{d} (s_i^k)^2} \sqrt{\sum_{k=1}^{d} (s_j^k)^2}} \tag{4}$$

Moreover, Huang et al. [Huang *et al.*, 2023] showed that Euclidean distance as Eq.(5) may not be entirely ineffective in high-dimensional spaces.

$$Dst_{Eul}(s_i, s_j) = \sqrt{\sum_{k=1}^{d} |s_i^k - s_j^k|} \tag{5}$$

2.3 Maximal Information Coefficient

Maximal Information Coefficient (MIC) is a correlation measure that features generality and equitability [Reshef *et al.*, 2011]. It finds partitions schemes via grid search to maximize the mutual information, assessing the association between two variables. The formula is as follows,

$$MIC(F; C) = \max_{ab < B(n)} \frac{I_{a,b}(F; C)}{\log(\min\{a, b\})} \tag{6}$$

where a, b are the partition numbers in directions of features F and labels C, with the constraint $ab \leq B(n)$. n is the dataset size. $I_{a,b}(F; C)$ is the maximum mutual information under constraint a, b. $MIC(F; C) \in [0, 1]$, where a higher value means a stronger correlation between F and C, and vice-versa. Typically, $B(n) = n^{0.6}$.

3 The Proposed Method

This section elaborates the proposed MASA-PSO. As shown in Figure.1, the MASA-PSO is divided into three main parts. The initial part covers a filtering step to eliminate useless features and an adaptive division mechanism for initializing population that make it suitable for corresponding search spaces. In the second part, we propose a fitness with multiscale adaptive discrimination framework based on adaptive weighting of multiple metrics. The last part proposes a hybrid elite strategy to mitigate the particles degradation phenomenon in most existing VLEAs when search space changes.

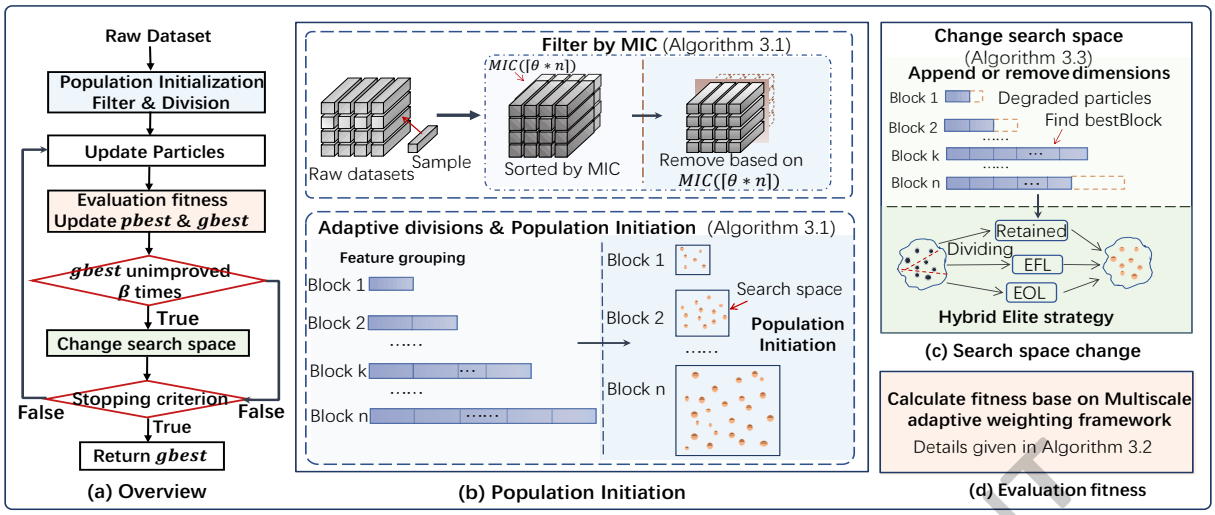

Figure 1: The overview of MASA-PSO. The major components in (a) includes the filtering part in (b), the adaptive division mechanism in (b), the hybrid elite strategy in (c) and the multiscale adaptive discrimination framework in (d).

3.1 Particle Initialization

Feature Ranking and Filtering In this study, the filtering threshold is $MIC(\lceil \theta_r \cdot m \rceil)$, where m is the number of raw features. It denotes the value ranked in descending order at position $\lceil \theta_r \cdot m \rceil$. Features with weak correlations to labels are removed. The remaining features form a set F in descending order based on MIC values. Note that $\theta_r = 0.7$.

Adaptive Division and Population Initialization Existing VLEAs divides the population into several equal-size subpopulations for cooperation learning, which may lead to inefficient resource utilization when search space sizes vary significantly across different subpopulation [Yazdani *et al.*, 2023]. Inspired by the S-shaped growth curve of species in constrained environments [Kucharavy and De Guio, 2011; Rangasamy *et al.*, 2021], we design a novel adaptive division mechanism that measures the relationship between population and search space based on Eqs.(7)(8). By approximating the S-curve growth pattern, this mechanism dynamically adjusts initial population sizes according to each block's search space dimensions, thereby enhancing search efficiency.

$$\mathrm{ParLen_i} = |F| * \frac{i}{M} \tag{7}$$

$$\mathrm{popSize_i} = \mathrm{Pop}_{min} + \frac{\mathrm{Pop}_{up}}{2}(1 + \tanh\frac{(i-\tilde{\imath})^3}{10}) \tag{8}$$

where i, $\tilde{\imath}$ is the block number and the median number of blocks, F is the number of features, M is the total number of blocks. Pop_{min} is the minimum population size, Pop_{up} is the maximum increase number of populations. This function allows the search space to expand as the number of groups increases, with the corresponding population size growing accordingly until it reaches $Pop_{min} + Pop_{up}$. (The whole operations see Algorithm 1)

3.2 Multiscale Adaptive Discrimination

VLEAs typically evaluate candidate solutions via a fitness function that combines classification accuracy with single-distance-based discrimination scores [Tran *et al.*, 2018; Qu *et al.*, 2023]. However, these single-distance metrics are difficult to maintain desired identification during search-space variation. To this end, we propose a fitness with the classification accuracy and a multiscale adaptive discrimination framework as Eq.(9). This framework dynamically weights Manhattan, Euclidean, and cosine distances to calculate an aggregation score. A logistic function is then constructed to maximize within-class aggregation while minimizing between-class aggregation.(Algorithm 2)

$$fitness = \gamma \cdot BACC + (1-\gamma) \cdot Score_{agg} \tag{9}$$

where γ is a weight set to 0.8, $BACC$ is the balanced accuracy [Tran *et al.*, 2016] for the imbalanced high-dimensional datasets as Eq.(10). $Score_{agg}$ is the aggregation score based on the adaptive weighting of three distances.

$$BACC = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{N_c} \tag{10}$$

where C is the number of classes, TP_c and N_c are the number of true positive samples and the total number of samples for the c^{th} class, respectively.

Multiscale Adaptive Weighting Mechanism For a given feature subset, the distances between samples are calculated upon Eqs.(3-5), as distance sets $\{Dst_l(s_i, s_j) \mid 1 \le i,j \le n, i \ne j\}$, where l only denotes Manhattan, Euclidean, and cosine distance in this paper.

For each class c, the sample set is divided into the c-class set $S_c = \{s_i^c | i = 1, 2, \ldots, n_c\}$ and the non-c class set $S_{\neg c} = \{s_i^{\neg c} | i = 1, 2, \ldots, n_{\neg c}\}$, where n_c and $n_{\neg c}$ are the number of samples of S_c and $S_{\neg c}$. The distances of within-class sample pairs in S_c and of between-class sample pairs between S_c and $S_{\neg c}$ are formulated as distance sets $Dst_l(S_c) = \{Dst_l(s_i^c, s_j^c) \mid 1 \le i,j \le n_c\}$ and $Dst_l(S_c, S_{\neg c}) = \{Dst_l(s_i^c, s_j^{\neg c}) | \, 1 \le j \le n_{\neg c}\}$.

Algorithm 1 The particles initialization.

Input: Raw training set D, the number of blocks M, minimum population size Pop_{min}, maximum population growth size Pop_{up}, the total number of features m, select threshold θ_r.
Output: $\{Block_i|Block_i = \{X_{i,1},X_{i,2},\cdots,X_{i,pSize_i}\}, i=1,\cdots,M\}$. ▷ Swarm

Calculate MIC_j for each feature f_j on D. ▷ Eq.(6)
Sort $\{f_j\}$ in descending order of $\{MIC_j\}$.
$\tau \leftarrow MIC(\lceil \theta_r \cdot m \rceil)$.
$D \leftarrow \{f_j \in D \mid MIC_j \geq \tau\}$. ▷ Filter step
for $i = 1$ to M **do** ▷ Adaptive division mechanism
  Calculate feature space $ParLen_i$ of block i. ▷ Eq.(7)
  Calculate population size $pSize_i$ of block i. ▷ Eq.(8)
  for $p = 1$ to $pSize_i$ **do** ▷ Initializing swarm
    for $d = 1$ to $ParLen_i$ **do**
      $pos^d_{i,p} \leftarrow rand(0,1)$. ▷ Position of particle $X_{i,p}$
      if $pos^d_{i,p} > \theta_r$ **then**
        $f_d \leftarrow 1$
      else
        $f_d \leftarrow 0$
      end if
    end for
    $pbest_p \leftarrow X_p$, $fitness(pbest_p) \leftarrow fitness(X_p)$.
  end for
end for
$gbest \leftarrow \arg\max_{X_{i,p}}\{fitness(X_{i,p})\}$. ▷ Initialize $gbest$
return $\{Block_i|Block_i = \{X_{i,1},\cdots,X_{i,pSize_i}\}, i = 1,\cdots,M\}$.

Then the within-class and between-class joint distance vectors for each sample s_i^c is constructed as follows,

$$Dst_w(s_i^c) = \begin{pmatrix} \max Dst_{Man}(s_i^c, S_c) \\ \max Dst_{Eul}(s_i^c, S_c) \\ \max Dst_{Cos}(s_i^c, S_c) \end{pmatrix} \quad (11)$$

$$Dst_b(s_i^c) = \begin{pmatrix} \min Dst_{Man}(s_i^c, S_{\neg c}) \\ \min Dst_{Eul}(s_i^c, S_{\neg c}) \\ \min Dst_{Cos}(s_i^c, S_{\neg c}) \end{pmatrix} \quad (12)$$

where $Dst_l(s_i^c,S_c)$ and $Dst_l(s_i^c,S_{\neg c})$ are selected from $Dst_l(S_c)$ and $Dst_l(S_c, S_{\neg c})$, representing distances between s_i^c and samples of S_c and of $S_{\neg c}$, respectively. Within-class vector $Dst_w(s_i^c)$ is defined as the three maximum distance metrics in $Dst_l(s_i^c, S_c)$, reflecting the clustering degree of class c centered on s_i^c, while between-class vector $Dst_b(s_i^c)$ is the three minimum distance metrics in $Dst_l(s_i^c, S_{\neg c})$, indicating the distance from $\neg c$ samples.

For all c class samples, the within-class matrix $X_w(c)$ and the between-class matrix $X_b(c)$ are constructed as follows,

$$X_w(c) = \begin{pmatrix} Dst_w(s_1^c) & Dst_w(s_2^c) & \cdots & Dst_w(s_{n_c}^c) \end{pmatrix} \quad (13)$$

$$X_b(c) = \begin{pmatrix} Dst_b(s_1^c) & Dst_b(s_2^c) & \cdots & Dst_b(s_{n_c}^c) \end{pmatrix} \quad (14)$$

where $X_w(c)$ and $X_b(c)$ are $3\times n_c$ matrices, with columns denoting the maximum within-class and the minimum between-class joint distance vectors, respectively.

Algorithm 2 The fitness calculation procedure.

Input: Training set $S = \{s_i|1 \leq i \leq n\}$, Classes of training set C, particle X_p, and outlier threshold θ_z.
Output: $fitness(X_p)$.

for each sample $s_i \in S$ **do** ▷ Select features
  Remove features not selected by particle X_p in s_i
end for
for $l \in \{Eul, Man, Cos\}$ **do** ▷ Calculate three distances
  Calculate $\{Dst_l(s_i,s_j)|\,\forall(s_i,s_j)\in S\}$ via Eqs.(3-5).
end for
for class $c \in C$ **do** ▷ Calculate the aggregation score
  $S_c \leftarrow \{s_i^c \mid i = 1,\ldots,n_c\}$, $S_{\neg c} \leftarrow S \setminus S_c$.
  $Dst_l(S_c) \leftarrow \{Dst_l(s_i^c, s_j^c) \mid 1 \leq i,j \leq n_c\}$
  $Dst_l(S_c, S_{\neg c}) \leftarrow \{Dst_l(s_i^c, s_j^{\neg c}) \mid 1 \leq j \leq n_{\neg c}\}$.
  for each sample $s_i^c \in S_c$ **do**
    Obtain $Dst_w(s_i^c), Dst_b(s_i^c)$ based on Eqs.(11)(12).
  end for
  $X_w(c) \leftarrow \{Dst_w(s_i^c) \mid i = 1,2,\ldots,n_c\}$ ▷ Eq.(13)
  $X_b(c) \leftarrow \{Dst_b(s_i^c) \mid i = 1,2,\ldots,n_c\}$ ▷ Eq.(14)
  Calculate $\delta_w(c)$ and $\delta_b(c)$ via Eqs.(15)(16).
end for
$\delta_w \leftarrow \frac{1}{C}\sum_{c=1}^{C}\delta_w(c)$, $\delta_b \leftarrow \frac{1}{C}\sum_{c=1}^{C}\delta_b(c)$. ▷ Average score
$f \leftarrow \frac{\delta_w - \delta_b}{len}$, $Score_{agg} \leftarrow \frac{1}{1+e^f}$. ▷ Normalize scores
$fitness(X_p) \leftarrow \gamma \cdot BACC + (1-\gamma)\cdot Score_{agg}$. ▷ Eq.(9)
return $fitness(X_p)$.

To integrate three distance metrics for collaboratively discriminating the dynamically changing sample distributions, while eliminating interference from correlations and dimensional inconsistency among these distance metrics, this paper develops a novel adaptive weighting framework based on precision matrix [Horn and Johnson, 2012] (**Proof of decorrelation and dimension unity is in Supplementary A**). Specifically, for each class c, the aggregation score of within-class $\delta_w(c)$ and the aggregation score of between score $\delta_b(c)$ are carried out via weighting operation as follows,

$$\delta_k(c) = \sqrt{(X_k(c) - \mu_k(c))^T P_k(c)(X_k(c) - \mu_k(c))} \quad (15)$$

where $k = \{w, b\}$, $\mu_k(c)$ is the average joint distance vector for all samples in $X_k(c)$. The precision matrix $P_k(c)$ is the inverse of covariance matrix $\Sigma_k(c)$ calculated by Eq.(16).

$$\Sigma_k(c) = \frac{1}{n_c}(X_k(c) - \mu_k(c))^T(X_k(c) - \mu_k(c)) \quad (16)$$

where the constrain $n_c > 3$ ensures that the precision matrix can be obtained, and the main diagonal elements of the precision matrix measure the guidance ability of three distance metrics in the given feature space [Horn and Johnson, 2012].

The average within-class score δ_w and the average between-class score δ_b are calculated, based on $\delta_k = \frac{1}{C}\sum_{c=1}^{C}\delta_k(c)$, where C is the number of classes. Inspired by the work [Tran *et al.*, 2017], the fitness is integrated by maximizing δ_w, minimizing δ_b, and normalizing the score to $[0, 1]$. Since the precision matrix implies the information of features

Algorithm 3 The procedure of Elite hybrid strategy.

Input: Elite grouping threshold θ_e, $Swarm = \{Block_i | Block_i = \{X_{i,1}, X_{i,2}, \cdots, X_{i,pSize_i}\}, i = 1, \cdots, M\}$, select threshold θ_s.
Output: $\{Block_i | Block_i = \{X_{i,1}, X_{i,2}, \cdots, X_{i,pSize_i}\}, i = 1, \cdots, M\}$. ▷ New Swarm

$E \leftarrow \arg\max_{Block_i \in Swarm} \sum fitness(X_{i,j}), j = 1, ..., pSize_i$. ▷ Elite
for $Block_i$ in $Swarm$ **do**
  for each $X_{i,j} \in Block_i$ **do** ▷ Compute forgotten value
    Calculate the difference of $fitness(X_{i,j})$ before and after the search space changes, as fg_j.
    if $fg_j < 0$ **then** ▷ Collect the degraded particles
      $S_F \leftarrow S_F \cup \{X_{i,j}\}$.
    end if
  end for
  Sort S_F by fg_j in descending order.
  $EOL \leftarrow$ The bottom $\lceil \theta_e \cdot |S_F| \rceil$ particles in S_F.
  $EFL \leftarrow$ The next $\lceil \theta_e \cdot |S_F| \rceil$ particles in S_F.
  $R \leftarrow$ The rest particles in S_F.
  $EOL \leftarrow EOLS(EOL, \theta_s)$. ▷ EOL strategy
  $EFL \leftarrow EFLS(EFL, E, \theta_s)$. ▷ EFL strategy
  $Block_i \leftarrow EOL \cup EFL \cup R$. ▷ Update block
end for
return $Swarm$.

[Horn and Johnson, 2012], a larger feature subset contains more information but hinder compact exploration, a regularization term len (the size of feature subset) is introduced to constrain the subset size.

$$Score_{agg} = \frac{1}{1 + \mathrm{e}^f} \tag{17}$$

$$f = -\frac{\delta_{\mathrm{w}} - \delta_{\mathrm{b}}}{len} \tag{18}$$

3.3 Hybrid Elite Strategy for Degraded Particles

In experiments, we observed that many particles in VLEAs experience performance degradation as the search space shifts toward more promising regions. VLEAs alter particle lengths by truncating or appending features, causing deviations from the search direction. The search efficiency of these particles is impaired. Although it is inherently inevitable, a hybrid strategy based on elite forward learning (EFL) [Deng *et al.*, 2020] and elite opposition learning (EOL) [Zhou *et al.*, 2012] is proposed to alleviate it. Specifically, degraded particles of each block are divided among EOL, EFL and retention group R based on their degradation degree, where the most severely degraded $\lceil \theta_e \cdot |S_F| \rceil$ particles are assigned to EOL, the next $\lceil \theta_e \cdot |S_F| \rceil$ particles undergo EFL, and the remaining particles form R. The reinitialized particles from both elite strategies, combined with R, form a new population for each block. Note that θ_e is the elite select threshold, S_F is the degraded particles set (Algorithm 3). The pseudocode of EFL and EOL strategy are in Supplementary B.

Elite Forward Learning Strategy For EFL group, the block with the highest average fitness is selected as the elite population $E = \{E_1, E_2, \cdots, E_n\}$. The selection probability of each feature is calculated as the ratio of times each feature is selected to the total number of occurrences of all features in E, the probability of each feature is as follows,

$$p_i = \frac{Fnum_i}{\sum_{i=1}^{d} Fnum_i}, i = 1, 2, 3, \cdots, d \tag{19}$$

where $Fnum_i$ represents the times of feature F_i selected by E, d is the given subset size. A roulette wheel selection [Wang *et al.*, 2020] is constructed to re-initialize particles based on these probabilities. The fitness of particles is evaluated before and after re-initialization, with only the higher fitness particles being retained.

Elite Opposition Learning Strategy For EOL group, since severe degradation after space changes indicates that the development value of their opposition space may exceed that of their original space [Zhou *et al.*, 2012]. Hence, an elite opposition learning strategy is introduced to tackle variable-length forgetting. Let $P_i = (p_{i,1}, p_{i,2}, \cdots, p_{i,d}), i = 1, 2, \cdots, n.$ represents the position of the i^{th} degraded particles of $Block_i$. The opposition solution $\check{P}_i$ is defined as follows,

$$\check{P}_\iota = (\check{p}_{i,1}, \check{p}_{i,2}, \cdots, \check{p}_{i,n}), i = 1, 2, \cdots, n. \tag{20}$$

$$\check{p}_{i,j} = r \cdot (da_j + db_j) - p_{i,j} \tag{21}$$

where $k \in U(0, 1)$ is a coefficient representing the magnitude of opposition, da_j and db_j are the upper and lower bounds of $Block_i$ in feature dimension j, as follows,

$$da_j = \max\{p_{i,j} | i = 1, 2, \cdots, n\} \tag{22}$$

$$db_j = \min\{p_{i,j} | i = 1, 2, \cdots, n\} \tag{23}$$

These particles are re-initialized based on the opposition solutions, and their fitness before and after initialization are compared. Particles with higher fitness are retained.

4 Experiments and Results Analysis

4.1 Experimental Setup

Benchmark Datasets 10 widely used gene expression datasets are selected to verify the performance of MASA-PSO. These datasets are described in Table 1, which includes the ratio of the largest and smallest classes. Most of them are imbalanced, high-dimensional, and low-sample size that make classification particularly challenging and representative of many real-world bioinformatics and medical diagnosis problems. Experimental results on these datasets effectively demonstrate the performance of the proposed algorithm.

Baseline Methods To evaluate the performance of MASA-PSO, this paper compares it with six SOTA methods in gene expression analysis, including VLPSO, ESAPSO, Many-Objective Jaccard-Based Evolutionary algorithm (JSEMO) [Saadatmand and Akbarzadeh-T, 2024], a Simple, Fast, and Efficient FS Algorithm (SFE) [Ahadzadeh *et al.*, 2023], and an Improved Conditional Wasserstein GAN With Gradient Penalty (CWGANGP) [Han *et al.*, 2025]. These methods involve VLEAs, other FS methods, and deep learning-based models, all of which have demonstrated competitive performance in gene expression analysis and are therefore representative baselines for comparison.(The comparative experiments with other SOTA methods and hyperparameter optimization are shown in Supplementary D and E)

Dataset	Feature	Sample	Class	Min(%)	Max(%)
SRBCT	2,308	83	4	13	35
DLBCL	5,469	77	2	25	75
9_Tumors	5,726	60	9	3	15
Leukemia1	5,327	72	3	13	53
Leukemia2	11,225	72	3	28	39
Brain_Tumor1	5,920	90	5	4	67
Brain_Tumor2	10,367	50	4	14	30
Prostate	10,509	102	2	49	51
Lung_Cancer	12,600	203	5	3	68
11_Tumors	12,533	174	11	4	16

Table 1: Descriptions of datasets.

Parameters	Setting
Maximum iterations	100
c1, c2	1.49445
w	$0.9 - 0.5 \cdot \frac{Iter_{curr}}{Iter_{max}}$
Threshold for selected feature(θ_s)	0.6
Max iterations for renew exemplars (α)	7
Max iterations for length changing(β)	9
Minimum population size (Pop_{min})	10
Maximum population growth size (Pop_{up})	20
The number of block (M)	5
Feature filtering threshold (θ_r)	0.7
Particles grouping threshold (θ_e)	0.2

Table 2: Parameter setting.

Experiment Settings ALL baselines use the authors' recommended settings, while MASA-PSO parameters are detailed in Table 2. To ensure fairness, all FS methods employed K-nearest neighbor (KNN) with K = 1 classifier. Each dataset undergoes 30 independent runs using ten-fold cross validation. Leave-one-out validation on the training set evaluates candidate solutions. Performance is measured by the mean and standard deviation of BACC and the size of selected features. The Wilcoxon significance test [Derrac *et al.*, 2011] with a 5% significance level is applied. All experiments are conducted on PC with Intel Core I7-14700KF @3.4GHz.

4.2 Experimental Results and Discussion

This section comprehensively presents the performance of MASA-PSO and five SOTA methods. Table 3 displays the results on ten gene expression datasets. " " denotes the size of final feature subset, "Best" represents the best result over 30 runs. "S" indicates the Wilcoxon significance test results, where "+" signifies a significant improvement of our algorithm over the baseline, "-" indicates that our method's BACC is significantly lower than the baseline, and "=" means no significant difference. (The computational and time cost analysis is detailed in Supplementary C)

(1) Comapred with CWGAN-GP, MASA-PSO achieves significantly better classification performance on most datasets, especially on highly imbalanced datasets (e.g., DLBCL, Bran_Tumor1,and Lung_Cancer).It suggests that the deep models are vulnerable to noise and prone to overfitting in HDLSS scenarios, while MASA-PSO effectively identifies informative feature subsets and offers robustness and interpretability, which is beneficial for subsequent gene analysis.

Dataset	Method	Size	Best	Mean±Std(%)	S
SRBCT	SFE	76.50	98.75	96.26±2.06	+
	VLPSO	49.10	**100**	**99.67±0.52**	=
	ESAPSO	47.80	**100**	99.50±0.76	=
	CWGAN-GP	128	**100**	99.17±0.65	=
	JSEMO	53.96	**100**	99.43±0.62	=
	MASA-PSO	42.99	**100**	99.62±0.54	
DLBCL	SFE	52.17	93.12	89.92±2.72	+
	VLPSO	73.46	93.33	86.51±3.52	+
	ESAPSO	53.67	99.17	92.37±3.09	+
	CWGAN-GP	128	95.12	92.93±2.17	+
	JSEMO	38.24	99.57	93.72±3.98	+
	MASA-PSO	56.13	**100**	**94.45±3.73**	
9_Tumors	SFE	71.10	53.97	49.26±3.20	+
	VLPSO	44.20	61.67	55.11±4.71	+
	ESAPSO	82.90	61.67	55.28±4.97	+
	CWGAN-GP	128	63.50	55.16±4.09	+
	JSEMO	1671.4	58.35	55.04±5.38	+
	MASA-PSO	78.16	**63.40**	**56.50±4.61**	
Leukemia1	SFE	132.20	94.32	87.04±3.81	+
	VLPSO	54.70	97.92	93.31±2.34	+
	ESAPSO	50.30	98.75	96.15±1.87	=
	CWGAN-GP	128	97.89	95.86±1.69	+
	JSEMO	218.27	98.40	95.61±1.91	+
	MASA-PSO	47.90	**99.17**	**96.59±1.79**	
Leukemia2	SFE	129.33	90.12	89.63±0.42	+
	VLPSO	35.20	94.44	91.56±1.84	+
	ESAPSO	86.53	**96.67**	**94.35±1.23**	=
	CWGAN-GP	128	96.99	93.12±1.92	+
	JSEMO	191.99	99.26	94.02±2.31	=
	MASA-PSO	73.40	98.02	93.94±1.97	
Brain_Tumor1	SFE	110.43	80.83	76.17±2.13	+
	VLPSO	26.80	79.17	71.19±3.52	+
	ESAPSO	66.41	83.75	76.17±3.30	+
	CWGAN-GP	128	79.48	76.18±3.42	+
	JSEMO	102.13	81.13	76.36±3.62	+
	MASA-PSO	33.51	**84.17**	**77.15±3.60**	
Brain_Tumor2	SFE	116.17	**82.41**	**74.51±4.23**	=
	VLPSO	81.5	73.33	66.78±4.10	+
	ESAPSO	106.45	80.00	74.21±3.94	=
	CWGAN-GP	128	77.08	73.51±2.36	+
	JSEMO	204.24	78.89	73.33±2.33	+
	MASA-PSO	80.95	79.62	74.27±3.60	
Prostate	SFE	106.40	90.74	87.56±1.39	+
	VLPSO	26.40	94.17	89.82±2.28	+
	ESAPSO	48.17	94.33	91.85±1.67	=
	CWGAN-GP	128	94.17	91.69±1.65	+
	JSEMO	112.95	94.86	91.18±2.53	+
	MASA-PSO	30.30	**95.17**	**92.65±1.79**	
Lung_Cancer	SFE	140.25	89.07	84.69±2.77	+
	VLPSO	176.10	94.08	89.47±2.18	+
	ESAPSO	100.35	94.44	90.11±2.18	+
	CWGAN-GP	128	95.24	91.80±1.81	+
	JSEMO	131.41	94.51	91.78±1.90	+
	MASA-PSO	73.75	**95.75**	**92.95±1.96**	
11_Tumors	SFE	95.67	90.10	84.09±2.26	+
	VLPSO	246.70	85.16	80.81±2.32	+
	ESAPSO	116.60	89.16	84.70±2.16	+
	CWGAN-GP	128	88.50	84.01±2.70	+
	JSEMO	126.20	88.57	84.48±2.93	+
	MASA-PSO	104.74	**89.98**	**85.70±2.29**	

Table 3: Experimental results.

(2) Compared with SFE and JSEMO, MASA-PSO not only achieves superior classification accuracy on most datasets but also captures fewer features than these methods. For instance, on Leukemia1, MASA-PSO collects approximately 80 fewer features than SFE, while on 9_Tumors, it reduces the features by nearly 1,500 compared with JSEMO. These results indicate that the variable-length mechanism enables MASA-PSO to identify more compact feature subsets while maintaining or improving performance. (3) Compared with VLPSO and ESAPSO, MASA-PSO consistently achieves higher classification accuracy on most datasets (e.g., more than 2.18% increase on DLBCL, over 2.84% increase on Lung_Cancer). Even on datasets where comparable accuracy is obtained, MASA-PSO selects significantly fewer features than these VLEAs, demonstrating its effectiveness in addressing the limitations of existing VLEAs. (A detailed comparison of optimization efficiency with VLEAs is in Supplementary E).

Overall, MASA-PSO demonstrates superior search capability over baseline methods, rapidly identifying better candidate solutions and finding more promising and smaller search space, thereby improving classification accuracy and capturing smaller feature subset. Meanwhile, it ensures the interpretability of the selected features for subsequent biological significance research.

4.3 Ablation Study

Experiments have verified that MASA-PSO outperforms the compared methods in classification accuracy, feature subset size and convergence speed. To further assess its effectiveness, ablation studies are conducted by adding or removing components. Figure.2 present the ablation results on six representative gene datasets (Results on more datasets are in the Supplementary of Figure.F4), where ”D”,”M”, and ”H” denote the adaptive division, the mulitiscale framework and the hybrid elite mechanism, respectively. ”+” means the inclusion of corresponding components. VLPSO is adopted as the basic algorithm, denoted as ”V”.

As Figure.2 shown, the introduction of **D** leads to varying degrees of BACC improvement. Combining the mechanism with other modules results in better BACC than when tested in isolation. For effect of **M**, it demonstrates the significant improvements for those applying the framework on most datasets, particularly on high-dimensional datasets (Prostate, and Lung_Cancer). It indicates that **M** effectively adapts to dynamic search spaces, improving BACC and capturing smaller feature set. As for **H**, Figure.2 reveals significant BACC improvements across all datasets with the strategy, especially on DLBCL and Brain_Tumor1, with increases of 7% and 4%.

4.4 Analysis of Selected Features

MASA-PSO is further examined through visualization and theoretical analysis. Figure.3 shows the distribution of selected features on DLBCL via t-SNE [Hinton and Roweis, 2002]. It shows that MASA-PSO successfully turns high-dimensional features into a compact subset, enhancing between-class separability. Figure.4 reveals that more than 75% of features selected by MASA-PSO are in the significant region, with few representation from non-significant regions. It demonstrates the algorithm’s capability to effectively identify biologically relevant differentially expressed features, rather than merely optimizing classification performance. (More analysis are in Supplementary G)

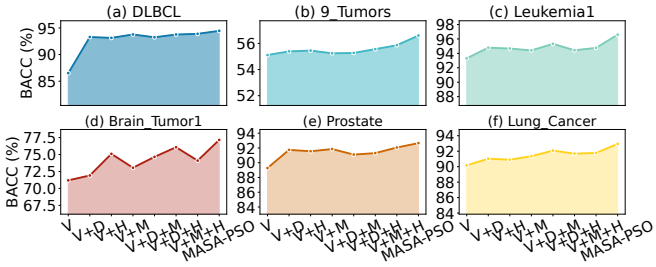

Figure 2: Visualization of BACC(%) in ablation experiments.

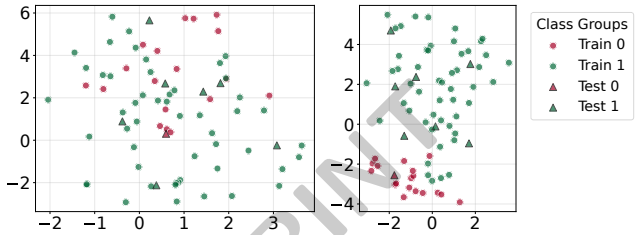

Figure 3: Sample distributions before and after FS on DLBCL.

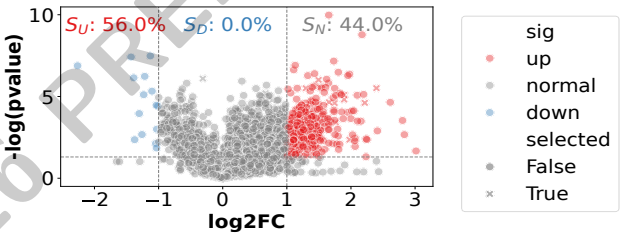

Figure 4: The volcano plot after feature selection on DLBCL.

5 Conclusions

In this study, we propose a novel size-adaptive PSO algorithm for HDLSS gene expression analysis (MASA-PSO). a multiscale adaptive discrimination framework, an adaptive division mechanism, and a hybrid elite strategy are designed to enhance classification performance, improve search efficiency, and mitigate particle degradations. Experiments on ten gene expression datasets show MASA-PSO outperforms SOTA methods. Moreover, the selected feature subsets exhibit statistical and biological significance, facilitating the subsequent research and diagnosis. Future work will explore combining multiscale discrimination and variable-length mechanisms with deep learning models to advance bioinformatics and medical research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant nos. 62576154 and

62102002.

References

[Ahadzadeh *et al.*, 2023] Behrouz Ahadzadeh, Moloud Abdar, Fatemeh Safara, Abbas Khosravi, Mohammad Bagher Menhaj, and Ponnuthurai Nagaratnam Suganthan. Sfe: A simple, fast, and efficient feature selection algorithm for high-dimensional data. *IEEE Transactions on Evolutionary Computation*, 27(6):1896–1911, 2023.

[Antonio and Coello, 2013] Luis Miguel Antonio and Carlos A. Coello Coello. Use of cooperative coevolution for solving large scale multiobjective optimization problems. In *2013 IEEE Congress on Evolutionary Computation*, pages 2758–2765, 2013.

[Bonyadi and Michalewicz, 2017] Mohammad Reza Bonyadi and Zbigniew Michalewicz. Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary computation*, 25(1):1–54, 2017.

[Chattopadhyay and Lu, 2019] Amrita Chattopadhyay and Tzu-Pin Lu. Gene-gene interaction: the curse of dimensionality. *Annals of translational medicine*, 7(24):813, 2019.

[Deng *et al.*, 2020] Li-Bao Deng, Li-Li Zhang, Ning Fu, Hai-li Sun, and Li-Yan Qiao. Erg-de: An elites regeneration framework for differential evolution. *Information Sciences*, 539:81–103, 2020.

[Derrac *et al.*, 2011] Joaquín Derrac, Salvador García, Daniel Molina, and Francisco Herrera. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3–18, 2011.

[Dubey and Saxena, 2016] Vimal Kumar Dubey and Amit Kumar Saxena. Cosine similarity based filter technique for feature selection. In *2016 International Conference on Control, Computing, Communication and Materials (ICCCCM)*, pages 1–6. IEEE, 2016.

[Han *et al.*, 2025] Fei Han, Yutao Liang, Qinghua Ling, and Henry Han. An improved conditional wasserstein gan with gradient penalty for gene expression profiling data augmentation based on data segmentation and depth feature constraint. *IEEE Transactions on Computational Biology and Bioinformatics*, 2025.

[Happy *et al.*, 2017] S L Happy, Ramanarayan Mohanty, and Aurobinda Routray. An effective feature selection method based on pair-wise feature proximity for high dimensional low sample size data. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1574–1578, 2017.

[Hinton and Roweis, 2002] Geoffrey E Hinton and Sam Roweis. Stochastic neighbor embedding. *Advances in neural information processing systems*, 15, 2002.

[Horn and Johnson, 2012] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[Huang *et al.*, 2023] Siquan Huang, Yijiang Li, Chong Chen, Leyu Shi, and Ying Gao. Multi-metrics adaptively identifies backdoors in federated learning. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4629–4639, 2023.

[Kucharavy and De Guio, 2011] Dmitry Kucharavy and Roland De Guio. Application of s-shaped curves. *Procedia Engineering*, 9:559–572, 2011.

[Liu *et al.*, 2017] Bo Liu, Ying Wei, Yu Zhang, and Qiang Yang. Deep neural networks for high dimension, low sample size data. In *IJCAI*, volume 2017, pages 2287–2293, 2017.

[Loaiza-Ganem *et al.*, 2024] Gabriel Loaiza-Ganem, Brendan Leigh Ross, Rasa Hosseinzadeh, Anthony L Caterini, and Jesse C Cresswell. Deep generative models through the lens of the manifold hypothesis: A survey and new connections. *arXiv preprint arXiv:2404.02954*, 2024.

[Min *et al.*, 2017] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.

[Noor and Ige, 2025] Mohd Halim Mohd Noor and Ayokunle Olalekan Ige. A survey on state-of-the-art deep learning applications and challenges. *Engineering Applications of Artificial Intelligence*, 159:111225, 2025.

[Qu *et al.*, 2023] Litao Qu, Weibin He, Jianfei Li, Hua Zhang, Cheng Yang, and Bo Xie. Explicit and size-adaptive pso-based feature selection for classification. *Swarm and Evolutionary Computation*, 77:101249, 2023.

[Rangasamy *et al.*, 2021] Devi Priya Rangasamy, Sivaraj Rajappan, Anitha Natarajan, Rajadevi Ramasamy, and Devisurya Vijayakumar. Variable population-sized particle swarm optimization for highly imbalanced dataset classification. *Computational Intelligence*, 37(2):873–890, 2021.

[Ravindran and Gunavathi, 2023] U Ravindran and C Gunavathi. A survey on gene expression data analysis using deep learning methods for cancer diagnosis. *Progress in Biophysics and Molecular Biology*, 177:1–13, 2023.

[Reshef *et al.*, 2011] David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.

[Saadatmand and Akbarzadeh-T, 2024] Hassan Saadatmand and Mohammad-R Akbarzadeh-T. Many-objective jaccard-based evolutionary feature selection for high-dimensional imbalanced data classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):8820–8835, 2024.

[Tran *et al.*, 2016] Binh Tran, Mengjie Zhang, and Bing Xue. A pso based hybrid feature selection algorithm for high-dimensional classification. In *2016 IEEE congress on evolutionary computation (CEC)*, pages 3801–3808. IEEE, 2016.

[Tran *et al.*, 2017] Binh Tran, Bing Xue, and Mengjie Zhang. A new representation in pso for discretization-based feature selection. *IEEE Transactions on Cybernetics*, 48(6):1733–1746, 2017.

[Tran *et al.*, 2018] Binh Tran, Bing Xue, and Mengjie Zhang. Variable-length particle swarm optimization for feature selection on high-dimensional classification. *IEEE Transactions on Evolutionary Computation*, 23(3):473–487, 2018.

[Wang *et al.*, 2020] Hongzhi Wang, Chengquan He, and Zhuping Li. A new ensemble feature selection approach based on genetic algorithm. *Soft Computing*, 24(20):15811–15820, 2020.

[Xue *et al.*, 2015] Bing Xue, Mengjie Zhang, Will N Browne, and Xin Yao. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on evolutionary computation*, 20(4):606–626, 2015.

[Yazdani *et al.*, 2023] Delaram Yazdani, Danial Yazdani, Donya Yazdani, Mohammad Nabi Omidvar, Amir H Gandomi, and Xin Yao. A species-based particle swarm optimization with adaptive population size and deactivation of species for dynamic optimization problems. *ACM Transactions on Evolutionary Learning and Optimization*, 3(4):1–25, 2023.

[Yu and Zhang, 2014] Xiang Yu and Xueqing Zhang. Enhanced comprehensive learning particle swarm optimization. *Applied Mathematics and Computation*, 242:265–276, 2014.

[Zhou *et al.*, 2012] Xinyu Zhou, Zhijian Wu, and Hui Wang. Elite opposition-based differential evolution for solving large-scale optimization problems and its implementation on gpu. In *2012 13th international conference on parallel and distributed computing, applications and technologies*, pages 727–732. IEEE, 2012.

[Zhou *et al.*, 2022] Junhai Zhou, Quanwang Wu, Mengchu Zhou, Junhao Wen, Yusuf Al-Turki, and Abdullah Abusorrah. Lagam: A length-adaptive genetic algorithm with markov blanket for high-dimensional feature selection in classification. *IEEE Transactions on Cybernetics*, 53(11):6858–6869, 2022.