TSMS-HRO: A Two-Stage Multi-Strategy Hybrid Rice Optimisation Algorithm for High-Dimensional Feature Selection

Zhiwei Ye[1,2], Jie Sun[1,2], Wen Zhou[1,2], Bogdan Adamyk[3,*], Jixin Zhang[1,2], Ting Cai[1,2], Jun Shen[4], Mengya Lei[1,2], Jing Zhou[1,2], and Ruihan Li[1,2]

[1]School of Computer Science, Hubei University of Technology, No.28, Nanli Road, Hongshan District, Wuhan, 430068, Hubei, China, https://orcid.org/0000-0002-1218-0681

[2]Hubei Provincial Key Laboratory of Green Intelligent Computing Power Network, Wuhan, 430068, Hubei, China

[3]Aston Business School, Aston University, Birmingham B4 7ET, UK, https://orcid.org/0000-0001-5136-3854

[4]School of Computing and Information Technology, University of Wollongong, Wollongong, 2500, New South Wales, Australia, https://orcid.org/0000-0002-9403-7140

[*]Corresponding author. b.adamyk@aston.ac.uk

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Abstract

High-dimensional feature selection remains a challenging and active topic in machine learning. Swarm intelligence and evolutionary computation have demonstrated promising results for high-dimensional feature selection, such as ant colony optimisation algorithm, particle swarm optimisation algorithm, and hybrid rice optimisation algorithm, etc. However, these algorithms still face two major challenges: The first is the presence of excessive redundant features in the selected subset, which degrades classification performance; the second is the long runtime of existing methods, which hampers efficient search and timely solution. To address these challenges, the paper proposes a novel two-stage algorithm, termed the two-stage multi-strategy hybrid rice optimisation algorithm (TSMS-HRO), specifically designed for high-dimensional feature selection. In the first stage, the minimum redundancy maximum relevance method is used to compute prior information to enhance the guidance of the feature subset search in the second stage. In the second stage, the hybrid rice optimisation algorithm is enhanced through four mechanisms: enhancing the quality and diversity of the initial population with good point set and elite opposition-based learning strategies; increasing the utilisation rate of maintainer line individuals with multiple adaptive differential operator selection strategies; improving the global and local search capabilities of the hybridisation process with a

© The Author(s) 2025. Published by Oxford University Press on behalf of the Society for Computational Design and Engineering. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

t-distribution mutation perturbation strategy; and enhancing the flexibility and diversity of the selfing process of restorer line individuals by introducing an improved adaptive crossover strategy. To evaluate the performance of the proposed method, extensive numerical experiments were conducted using benchmark functions from CEC2022. Results are compared with other well-known algorithms, such as the whale optimisation algorithm and grey wolf optimiser. Furthermore, TSMS-HRO is applied to 12 high-dimensional biomedical datasets. The experimental results show that TSMS-HRO outperforms other two-stage and metaheuristic algorithms based feature selection methods in terms of accuracy and convergence speed. For example, on the CLL_SUB_111 dataset with 11,340 dimensions, TSMS-HRO achieved an average accuracy of 95.25% with a 98.86% reduction in features, clearly surpassing other methods in both effectiveness and stability. These findings confirm that TSMS-HRO is an efficient and reliable algorithm not only for the optimisation of functions with different characteristics but also for real-world optimisation problems.

Keywords: High-dimensional feature selection, Hybrid Rice Optimisation algorithm, Minimum redundancy maximum relevance, Good point set, Elite opposition-based learning.

Nomenclature

X_m	The individuals of maintainer line
X_s	The individuals of sterile line
X_r	The individuals of restorer line
S	The number or ratio of feature selection
V_{max}^d, V_{min}^d	The maximum and minimum values of the d-th dimension
F	The smoothing factor
s_{max}, s_{min}	The boundary adjustment parameters
$\sigma_{max}, \sigma_{min}$	The maximum and minimum mutation scales
g_r	The growth rate to control σ
c_r	The crossover rate in the differential evolution stage
SC_{max}, SC_{min}	The values of the selfing upper and lower limit
SC	The maximum number of selfing in HRO
α	The step-size control factor
λ	The weight of error rate
μ	The weight of the feature selection rate

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

1. Introduction

In today's era of information explosion, massive amounts of data are being generated (Badshah et al., 2024), characterised by large volumes and high dimensionality. While those features reflect the richness of data, they have also included some redundant features and noise (Barbieri et al., 2024). In a higher-dimensional data environment, redundant features and noise have significantly impacted the effectiveness of intelligent systems, leading to biased analysis results (G. Li et al., 2024). For example, in gene expression data analysis, these issues significantly hinder the development of intelligent systems (Liang et al., 2024). Among tens of thousands of genes, only a small subset of the gene expression levels has been closely associated with the disease diagnosis, resulting in the data sparsity (Y.-C. Wang et al., 2024). In the context of high-dimensional data, feature engineering is key to improving the generalisation ability of models, as it enables models to better understand and interpret complex data (Lameesa et al., 2024). Consequently, effective feature reduction enhances both classification accuracy and computational efficiency.

Dimensionality reduction techniques are used to reduce the number of data features by retaining important ones while eliminating redundant features. This not only conserves storage but also aids in uncovering meaningful information and mitigates the potential risk of model overfitting. Dimensionality reduction techniques can be broadly classified into two categories: Feature Extraction (Kapoor et al., 2024) and Feature Selection (FS) (Moslemi, 2023; X. Song, Zhang et al., 2024). Feature extraction methods, such as Principal Components Analysis (PCA) (Zheng et al., 2024), Linear Discriminant Analysis (LDA) (J. Zhou et al., 2024), and Independent Component Analysis (ICA) (Buchaiah & Shakya, 2022), reduce data dimensionality by projecting the original feature space onto a new lower-dimensional feature space through a specified mapping process. These methods have been successfully applied in various fields, including disease classification (Yu et al., 2022) and hyperspectral image classification (C. Wang et al., 2024).

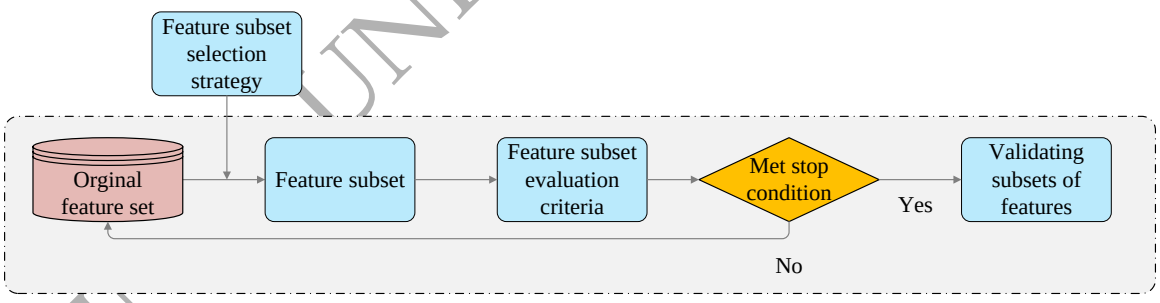

Figure 1: Basic process of feature selection.

FS has been one of the key steps in pattern recognition and classification (Nematzadeh et al., 2024). Its goal has been to eliminate some redundant and irrelevant features from the data, enabling the extraction of valuable information from the higher-dimensional data. The basic flowchart of FS is shown in Figure 1. The application of FS in high-

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

dimensional data environments is crucial. First, it effectively addresses the curse of dimensionality, where distances between data points become smaller in high-dimensional spaces, making it difficult for classifiers to distinguish different samples (Anuragi et al., 2024). Second, minimising the feature space can save storage space, facilitate information discovery, and reduce the likelihood of model overfitting (Telikani et al., 2021). Moreover, a key advantage of FS lies in its direct selection of optimal feature subsets from the original feature space. Traditional FS methods rely solely on the information inherent to the samples themselves, without directly depending on learning algorithms. These methods are based on various evaluation metrics, including statistical (BinSaeedan & Alramlawi, 2021), information-theoretic (Gao et al., 2023; He et al., 2024), similarity-based (Ouadfel & Abd Elaziz, 2022; B. Zhang et al., 2022), and sparsity-based (R. Zhou et al., 2023) approaches. In contrast, another category of FS methods evaluates the performance of feature subsets through learning algorithms, assisting in the selection of the optimal feature subset. Currently, mainstream approaches in this category utilise metaheuristic algorithms with strong search capabilities to explore the entire feature space. These algorithms validate subsets using a predefined learning algorithm until the optimal solution is found. Common methods include the Genetic Algorithm (GA) (Fang et al., 2024), Grey Wolf Optimiser (GWO) (Y. Wang et al., 2024), Whale Optimisation Algorithm (WOA) (Miao et al., 2024), Harris Hawks Optimisation (HHO) (Peng et al., 2023; Zhao et al., 2024), Salp Swarm Algorithm (SSA) (Qaraad et al., 2022), Equilibrium Optimiser (EO) (Ahmed et al., 2021) and Particle Swarm Optimisation (PSO) (Xue & Zhang, 2024). FS avoids the potential loss of semantic information that may occur during the transformation process in feature extraction techniques, making it a more extensively researched field (Rajammal et al., 2022).

With the rapid increase in dataset dimensionality, traditional FS techniques face challenges such as high computational costs and degraded model performance. To address these issues, researchers have proposed FS methods specifically designed for high-dimensional data scenarios (Osama et al., 2023). Manikandan & Abirami (2021) proposed a high-dimensional FS method based on mutual information (MI) and Monte Carlo techniques. It employs an approximate Markov blanket, MI, and a novel strategy based on Monte Carlo Tree Search (MCTS) technology. In the first stage, primary features are selected from high-dimensional data, and in the second stage, redundant features identified in the first stage are eliminated. This approach achieves dimensionality reduction and effectively enhances feature interaction. J. Zhang et al. (2021) proposed FS–GBDT, a hybrid FS framework combining Fisher score and Gradient Boosting Decision Tree (GBDT), to identify robust cancer-related gene subsets from high-dimensional expression data. By jointly analysing 11 cancer types, FS–GBDT effectively discovered overlapping risk gene modules. Zuo et al. (2025) proposed an unsupervised FS method (MRMGRFS) that combines spectral clustering-based relevance evaluation with global redundancy minimisation using Jensen–Shannon divergence, effectively selecting informative and non-redundant

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

features from high-dimensional data. Chamlal et al. (2024) proposed a supervised FS method ISClique, which first uses a Kendall's tau-based filter to evaluate feature interactions, and then applies a maximal clique strategy to construct optimal subsets, achieving superior performance in high-dimensional regression tasks.

Although the aforementioned FS methods have made adaptive improvements for high-dimensional FS tasks, they still face challenges such as low search efficiency and suboptimal classification accuracy of the selected feature subsets. Moreover, many of these methods require prior knowledge to guide the search process of FS. This limitation not only restricts their general applicability but also necessitates expert knowledge to assist in the search process, making their usage less straightforward. Metaheuristic algorithms are a class of strategies designed to solve optimisation problems. They generally do not rely on specific assumptions about data distribution or require prior knowledge to guide the search process. Instead, they utilise general search mechanisms, combining randomness and heuristic information to iteratively seek near-optimal solutions. This makes them particularly suitable for tackling large-scale, nonlinear, high-dimensional problems with challenging gradient-based solutions. Metaheuristic algorithms have shown great potential in handling high-dimensional FS tasks (Nssibi et al., 2023; X. Song et al., 2022), and their applications in this domain have significantly increased in recent years. When applying metaheuristic algorithms to high-dimensional FS problems, the initial feature subsets are typically generated randomly. The algorithms rely on their search strategies to explore the feature space for the optimal feature combinations. Consequently, the search efficiency and update mechanisms for candidate solutions directly influence the quality of the selected feature subsets. This has become a primary focus for researchers in the field. Currently, various high-performance metaheuristic methods have been successfully applied to different FS tasks. Hussain et al. (2021) proposed a hybrid optimisation method called Sine-Cosine Harris Hawk Optimisation (SCHHO), which combines the Sine Cosine Algorithm (SCA) and HHO. This method aims to enhance the performance of numerical optimisation and FS. To effectively select the optimal gene combination from microarray data, Pashaei (2022) utilised Minimum Redundancy Maximum Relevance (mRMR) in the initial stage to filter the top m promising genes and reduce the feature space. Subsequently, the Aquila Optimiser (AO) with a mutation mechanism and a Time-Varying Mirrored S-shaped (TVMS) transfer function was applied to search for the optimal feature subset. Nssibi et al. (2024) proposed a hybrid binary FS method called iBABC-CGO, combining the island model of the Binary Artificial Bee Colony (BABC) with Chaos Game Optimisation (CGO). To enhance the search process in binary space, two transfer functions were used. The method integrates SVM for evaluating gene subsets and was tested on 15 biological datasets. Experimental results showed that iBABC-CGO achieved competitive classification accuracy, efficient gene selection, and fast convergence, with meaningful biological interpretations of selected genes. Jiang et al. (2024) proposed the Dynamic Crow Search Algorithm (DCSA) to enhance FS for high-dimensional bio-

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

medical data classification. To improve exploration and exploitation balance, a dynamic bi-level awareness probability was introduced. Additionally, Lévy flight with an adaptive step length replaced the original random search mechanism, and a dynamic flight length strategy was employed to accelerate convergence. Experimental results on seven high-dimensional biomedical datasets demonstrated that DCSA achieved superior classification accuracy and selected the smallest feature subsets, with 100% accuracy on SRBCT data.

To further enhance the performance and efficiency of FS methods based on metaheuristic algorithms in high-dimensional FS problems, researchers have combined filter-based FS methods with metaheuristic algorithms to form a multi-stage FS approach (X. Song, Ma et al., 2024). These approaches aim to leverage the high search efficiency of filter methods and the strong search capability of metaheuristic algorithms. For example, X. Song, Ma et al. (2024) introduced a three-stage Streaming Feature Selection method based on Dynamic feature clustering and Particle Swarm Optimisation (SFS-DPSO). The method combines online relevance analysis to quickly eliminate irrelevant features, dynamic clustering to handle redundancy, and an integer PSO to search for the optimal subset. Experiments on 12 benchmark datasets and a real-world case demonstrated that SFS-DPSO achieves superior classification performance within reasonable time compared to existing algorithms. Got et al. (2021) developed a novel multi-objective hybrid filter-wrapper method to address the FS problem. The method leverages the WOA to explore promising regions in the feature space. Additionally, two objective functions were considered during the optimisation process: the first objective uses MI as a filter fitness function to assess the relevance and redundancy among features, aiming to identify a non-dominated subset of features with minimal redundancy and maximum relevance to the target class. The second objective employs a learning classifier as a wrapper fitness function to evaluate classification accuracy. Experimental results demonstrated that the proposed algorithm could achieve multiple feature subsets with fewer features while maintaining excellent classification accuracy. X.-F. Song et al. (2021) proposed a novel three-stage hybrid FS algorithm (Hybrid FS algorithm based on correlation-guided Clustering and Particle Swarm Optimisation, HFS-C-P) to address the challenge of high computational costs in high-dimensional data. In the first and second stages, a filter-based FS method and a correlation-guided clustering approach were designed to reduce the search space for the third stage. Subsequently, in the third stage, an evolutionary algorithm with global search capabilities was employed to identify the optimal feature subset. To enhance the performance of all three stages, the authors developed a feature elimination method based on symmetrical uncertainty, a fast correlation-guided feature clustering strategy, and an improved integer PSO algorithm. Experimental results on 18 datasets demonstrated that this algorithm could obtain a high-quality feature subset with minimal computational cost. Thirumoorthy et al. (2023) employed a two-stage FS strategy combining filter and wrapper-based methods. In the first stage, four filtering techniques (MI, ReliefF, Information Gain (IG), and F-Score) were used to select a reduced feature subset. In the

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

second stage, a wrapper-based hybrid Coati Optimisation Algorithm (COA) was applied. This hybrid strategy incorporated opposition-based learning, adaptive population size adjustment, elite learning, and differential evolution operations to enhance search efficiency and accuracy. The proposed method was evaluated on three benchmark datasets. The comprehensive results demonstrated that the method outperformed other compared approaches, significantly improving breast cancer classification success rates while reducing the FS ratio. Moustafa et al. (2024) proposed an innovative hybrid FS approach to derive an optimal feature subset for high-precision crop mapping. In the first stage, MI and ReliefF filtering methods were employed to rank the spectral-temporal remote sensing features in the dataset. The most relevant features identified by these filtering methods were then combined into a unified subset. Subsequently, the GWO was applied to refine the initial feature set. Finally, a Random Forest classifier was used with the optimised feature subset to predict crop types accurately. Performance evaluation conducted in the Behiera province of Egypt demonstrated that the proposed method outperformed existing crop mapping approaches, achieving an accuracy of 82%. Agrawal et al. (2022) proposed a Normalised MI-based Equilibrium Optimiser (NMIEO) to enhance the efficiency of FS in high-dimensional datasets. This method integrates a novel local search strategy based on Normalised Mutual Information (NMI) to improve the algorithm's local exploitation capabilities. Additionally, chaotic mapping is employed to enhance the diversity of the initial population. Before searching for the optimal feature subset, NMIEO reduces the feature space using a filtering method. Four common filtering methods were compared experimentally, and the most suitable one was selected as the first-stage filter. Results demonstrated that NMIEO outperformed eight well-known metaheuristic algorithms from recent literature in handling high-dimensional datasets. Askr et al. (2024) proposed the Binary Enhanced Golden Jackal Optimisation (BEGJO) algorithm to improve FS for high-dimensional data. To overcome the local optima, enhancement strategies were introduced, and Copula Entropy (CE) was integrated for dimensionality reduction while maintaining classification accuracy. The sigmoid transfer function transformed BEGJO into a binary form suited for FS tasks. Experimental results showed that BEGJO outperformed existing algorithms in classification accuracy and FS efficiency, with statistical validation confirming its effectiveness.

In addition to the above-mentioned well-known algorithms, a novel population-based metaheuristic algorithm named Hybrid Rice Optimisation Algorithm (HRO) is proposed (Z. Ye et al., 2016), which is inspired by the hybrid breeding process of three-line hybrid rice. According to heterosis theory, first-generation hybrid offspring often exhibit superior traits in growth, reproduction, and behavioural characteristics compared to their parents. As a result, HRO shows strong search capability, high efficiency, and adaptability. Because of these advantages, coupled with its high flexibility and ease of implementation, HRO has been applied by researchers to problems such as disease diagnosis (Mei et al., 2025; A. Z. Ye et al., 2023) and intrusion detection (Z. Ye et al., 2024). Compared to tradi-

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

tional metaheuristic algorithms, HRO stands out by emphasizing the utilisation of hybrid breeding mechanisms and heterosis to enhance the population's evolution and iteration. Therefore, the paper intends to take advantage of the HRO algorithm and combine it with the aforementioned two-stage method, which might exhibit even better performance.

The newly proposed FS technique combines the mRMR filtering method with the HRO enhanced by multiple strategies, specifically targeting the classification of ultra-high-dimensional biomedical gene expression data. Experimental results validate the effectiveness of the approach, achieving superior performance compared to other related methods. The main contributions of the study can be summarised as follows:

1. A Two-Stage Multi-Strategy Hybrid Rice Optimisation Algorithm (TSMS-HRO), is proposed based on the improved HRO algorithm. Unlike existing two-stage approaches that often rely on simple filter–wrapper combinations, our method establishes a tighter coupling between filtering and metaheuristic search, improving both search efficiency and feature subset quality.

2. Multi-strategy enhancements are introduced into HRO. Compared with single-strategy improvements in existing HRO variants, these four mechanisms work collaboratively to achieve a better balance between global exploration and local exploitation.

3. The method's effectiveness is demonstrated by employing an SVM classifier on 12 biomedical datasets. In terms of classification accuracy, feature reduction rate, and convergence speed, our method consistently outperforms state-of-the-art two-stage feature selection algorithms and advanced metaheuristic-based algorithms.

4. A series of auxiliary experiments is conducted to demonstrate the effectiveness of the proposed method, such as ablation experiments. By isolating different strategies, we verify that the joint design of the two-stage framework and multi-strategy enhancements is essential for achieving superior stability and robustness.

The remainder of this study is organised as follows: Section 2 introduces the basic HRO algorithm and the mRMR filtering method. Section 3 details the methodology proposed in this study. Section 4 presents the experimental setup and discusses the results. Finally, the conclusions of this work and future research are provided in Section 5.

2. Preliminaries

Before detailing the TSMS-HRO algorithm, we first review the basic components of the HRO algorithm, which serves as the core metaheuristic in this work.

2.1 The hybrid rice optimisation algorithm

HRO is a novel population-based metaheuristic algorithm, proposed by Z. Ye et al. (2016), that boasts strong search capabilities and high computational efficiency. The algorithm's

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

main process includes four stages: Three-line Division, Hybridisation, Selfing, and Renewal. The core idea is shown in Figure 2.

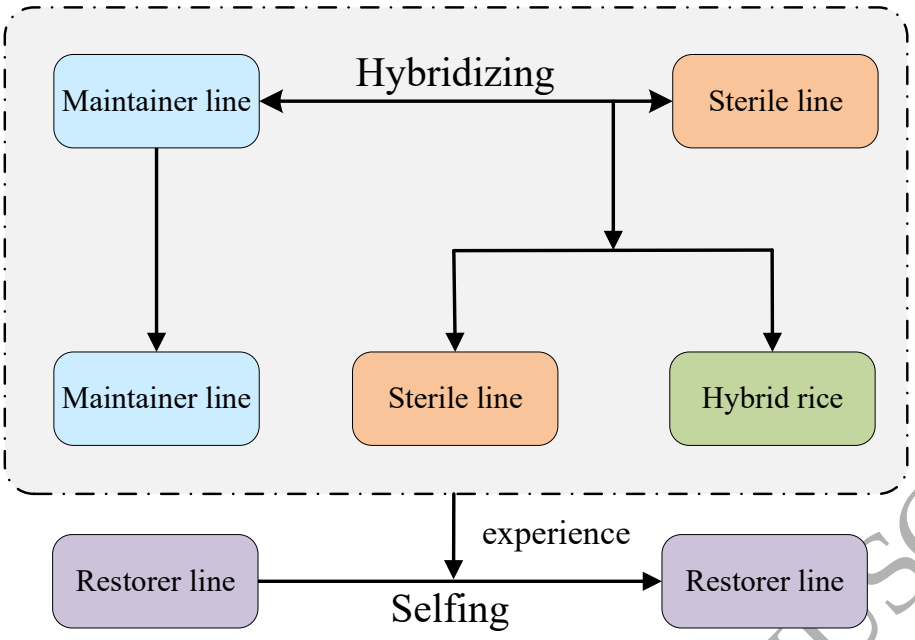

Figure 2: Basic HRO algorithm.

(i) ***Three-line Division***: In the breeding process, the initial populations $X = \{X_1, X_2, \cdots, X_n\}$ are sorted in each iteration according to the fitness values, where n demonstrates the size of populations. The maintainer line represents a subgroup of the population with the best fitness values and is denoted as $X_m = \{X_1, X_2, \cdots, X_p\}$, where $p = \lfloor n/3 \rfloor$. A group of seeds with poorer fitness values, which need to hybridise with the maintainer line to improve the quality of individuals, forms the sterile line, denoted as $X_s = \{X_{2p+1}, X_{2p+2}, \cdots, X_n\}$. The remaining subgroup is the restorer line, represented as $X_r = \{X_p, X_{p+1}, \cdots, X_{2p}\}$, which attempts to update its position toward the maintainer line through selfing.

(ii) ***Hybridisation***: The hybridisation process involves crossing the maintainer line and the sterile line, which exhibit the greatest difference in fitness values. This process is designed to enhance the genetic quality of sterile line individuals via heterosis-inspired recombination with superior hybrid individuals. The procedure for generating new individuals through hybridisation is described in Equation 1.

$$\begin{cases} X^d_{new(i)}(t+1) = r_1 \cdot X^d_s(t) + (1 - r_1) \cdot X^d_m(t), \\ m \in \{1, 2, \cdots, p\}; i, s \in \{2p+1, 2p+2, \cdots, n\}. \end{cases} \tag{1}$$

where $X^d_{new(i)}$ denotes the d-th gene of i-th hybrid in the sterile line during the $(t+1)$-th iteration. $X^d_s(t)$ and $X^d_m(t)$ denote the d-th gene of randomly selected individuals from

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

the sterile line and maintainer line populations, respectively. r_1 is a random number in the range $[0, 1]$.

(iii)*Selfing*: The selfing stage is a critical step in optimising the individuals within the restorer line. During this, the seed individuals in the restorer line exchange genetic information through crossover and recombination, enabling subpopulations to evolve toward the optimal solution. Equation 2 models the selfing process.

$$\begin{cases} X^d_{new(i)}(t+1) = r_2(X^d_{best}(t) - X^d_{r(j)}(t)) + X^d_{r(i)}(t), \\ i, j \in \{p+1, p+2, \cdots, 2p\}; i \neq j. \end{cases} \tag{2}$$

where $X^d_{new(i)}$ denotes the new gene generated through selfing between the i-th and j-th individuals ($i \neq j$) of the restorer line. $X^d_{best}(t)$ denotes the d-th gene of the best individual found so far, while $X^d_{r(j)}(t)$ denotes the d-th gene of the i-th individual randomly selected from the restorer line. r_2 is a random number in the range $[0, 1]$.

After generating new individuals through hybridisation and selfing, they are compared with the original candidate individuals. If the fitness value of the new individual is superior to that of the original candidate, replacement is performed according to Equation 3.

$$X_i(t+1) = \begin{cases} X_{new(i)}(t+1), & iff(X_{new(i)}(t)) > f(X_i(t)), \\ X_i(t), & otherwise. \end{cases} \tag{3}$$

(iv)*Renewal*: In the HRO algorithm, the Self Crossing (SC) count is used to measure the cumulative number of iterations in which an individual from the restorer line has not been updated. When the self-crossing count for a restorer line individual reaches the preset maximum value ($SCmax$), it indicates that the individual has not been effectively updated over multiple consecutive iterations. At this point, a reset operation is performed, as described by the following Equation 4.

$$X^d_{r(i)}(t+1) = r_3(V^d_{max} - V^d_{min}) + X^d_{r(i)}(t) + V^d_{min}. \tag{4}$$

where $X^d_{r(i)}(t)$ denotes the d-th gene of the i-th restorer line individual that has not been updated. V^d_{max} and V^d_{min} denote the maximum and minimum values of the d-th dimension. r_3 is a random number selected from the range $[0, 1]$.

2.2 Minimum redundancy - maximum relevance (mRMR)

The mRMR algorithm aims to select a subset of features that are highly relevant to the target variable while maintaining minimal redundancy among the features. This approach is based on an intuitive concept: a good feature set should contain features that are closely related to the target variable while ensuring independence among the features to avoid redundant information. This balance is achieved by jointly considering the mutual information between each feature and the target variable, as well as the average mutual

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

information among the features. Formally, let X represent the feature set and Y represent the target variable. The mRMR selection criterion is shown in Equation 5.

$$max\left(\frac{1}{|S|}\sum_{x_i\in S} I(x_i;Y) - \frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i;x_j)\right). \tag{5}$$

where S denotes the set of selected features (the default value is 300, confirmed in subsequent experiments). $I(x_i;Y)$ denotes the mutual information between feature x_i and the target variable Y, which measures their relevance. On the other hand, $I(x_i;x_j)$ denotes the mutual information between features, which measures their redundancy. The mRMR method aims to maximise the difference between these two quantities, thereby ensuring that the selected feature set contains features that are highly relevant to the target variable while being mutually independent within the set.

3. The Proposed Approach

In this paper, a two-stage improved FS approach is proposed. A filter-based method is employed in the first stage to preliminarily filter high-dimensional features, with the features subset selected by the filter serving as the initial search space. In the second stage, the multi-strategy integrated HRO algorithm searches within this refined feature space and outputs the final feature subset.

3.1 mRMR-based filter

In the first stage of this method, the mRMR-based technique is used to select a feature set that is both highly relevant to the target variable and minimally redundant among features, as defined in Equation 5. The selected features subset serves as the initial search space for the subsequent metaheuristic algorithm.

The pseudo code is in Algorithm 1. After we input the feature subset size $|S|$ and the dataset, for each feature x_i in the dataset, we need to calculate the mutual information between it and the feature label Y and all other features x_j, and then use Equation 5 to calculate the mRMR score of each feature. According to the size of $|S|$, the feature with the highest score is taken as the feature subset S_1, and S_1 is used as the input of the second stage.

3.2 MS-HRO: multi-strategy integrated hybrid rice optimisation

In the second stage of this method, four different strategies are employed to optimise the basic HRO algorithm. As shown in the Figure 3, in the MS-HRO algorithm, the population is initialised using the good point set and elite opposition-based learning to initialise the population. This is very important, as it effectively improves the quality and

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Algorithm 1 Preliminary Filtering with mRMR.

Require: Desired subset size $|S|$, dataset.
Ensure: Initial feature subset S_1 for Stage 2.

```
for each feature x_i in the dataset do
    Calculate mutual information I(x_i; Y) between feature x_i and target variable Y;
end for
for each pair of features (x_i, x_j) in the dataset do
    Calculate mutual information I(x_i, x_j) between features x_i and x_j;
end for
for each feature x_i do
    Compute mRMR score using Equation 5;
end for
Select the top-ranked features based on mRMR scores to form the initial feature subset S_1;
return the initial feature subset S_1.
```

diversity of our initialised population. Then it enters the iterative optimisation until the specified maximum number of iterations is reached and the optimal solution is output. In the iteration, the fitness of all individuals in the population must be measured first, and then the three-line population is sorted and divided according to the fitness. Subsequently, the maintainer line is updated by the aptive difference operator selection strategy, the sterile line is hybridised with the maintainer line to obtain its excellent genes while using the t-distribution mutation perturbation strategy to improve the search performance at different stages, and the restorer line is updated by the enhanced adaptive selfing strategy and the Lévy flight strategy. Finally, the global optimal solution is updated, and the next iteration is entered. And Algorithm 2 for the pseudocode.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

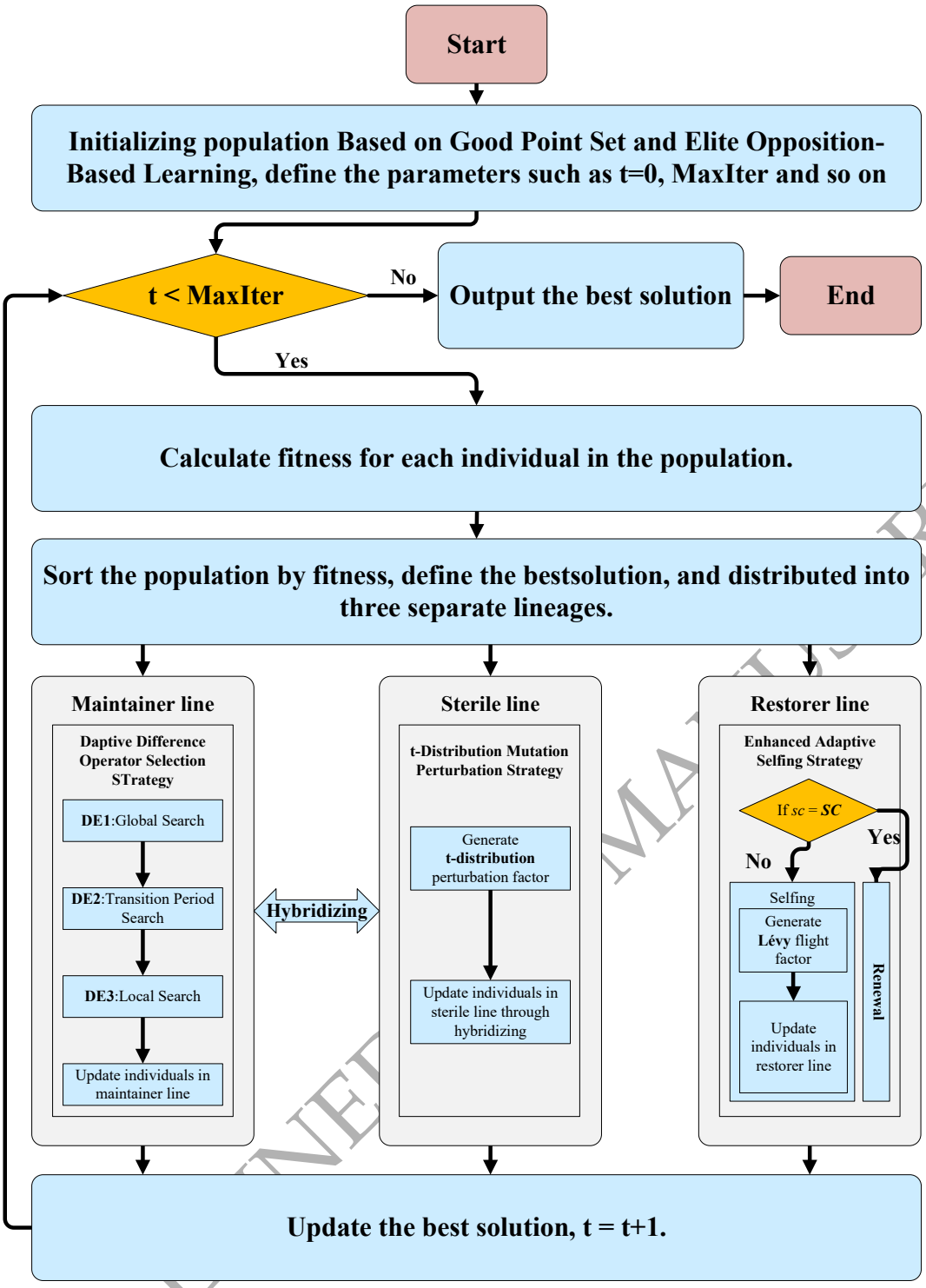

Figure 3: The Flowchart of MS-HRO.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Algorithm 2 MS-HRO-based optimisation.

Require: Maximum iterations *MaxIter*, population size n, $F \Leftarrow 0.5$, $s_{max} \Leftarrow 0.92$, $s_{min} \Leftarrow 0.01$, $\sigma_{max} \Leftarrow 1.0$, $\sigma_{min} \Leftarrow 0.1$, $g_r \Leftarrow 2$, $cr \Leftarrow 0.9$, $SC_{max} \Leftarrow 10$, $SC_{min} \Leftarrow 4$, $\alpha \Leftarrow 0.1$.

Ensure: Best solution S.

Initialise population $X \Leftarrow \{X_1, X_2, \ldots, X_n\}$ using good point set and elite reverse learning for diversity (Equations 6 – 9);
for $t = 1$ to MaxIter **do**
▷ Three-line Division
    Calculate fitness for each individual in the population;
    Sort the population by fitness:
        Maintainer line $X_m \Leftarrow \{X_1, X_2, \ldots, X_p\}$;
        Restorer line $X_r \Leftarrow \{X_{p+1}, \ldots, X_{2p}\}$;
        Sterile line $X_s \Leftarrow \{X_{2p+1}, \ldots, X_n\}$;
▷ Adaptive Differential Operator Selection
    for each individual in maintainer line X_m **do**
        Select differential operator based on adaptive probabilities (Equation 19);
        Apply DE1, DE2, or DE3 (Equations 10 – 12);
        Update individual positions in X_m;
    end for
▷ Hybridisation
    for each individual $X_s(i)$ in sterile line **do**
        Select a random individual X_m and $X_s(j)$;
        Apply t-distribution-based mutation (Equations 20 – 23);
        Update $X_s(i)$;
    end for
▷ Selfing
    for each individual $X_r(i)$ in restorer line **do**
        Select a neighboring individual $X_r(j)$;
        Perform self-crossing (Equations 25 – 27);
        Update $X_r(i)$ and increment self-crossing counter SC
    end for
▷ Renewal
    for each individual in X_r with $sc = SC$ (Equation 24) **do**
        Perform reset operation;
        Reset SC for $X_r(i)$;
    end for
    Update the best solution found so far;
end for
return The best solution S.

3.2.1 Initialisation strategy based on good point set and elite opposition-based learning (INIT)

To address the issue of insufficient diversity in the initial population of HRO, a population initialisation strategy based on good point set mapping and elite reverse learning is proposed to optimise the generation of the initial population. The mapping of the good point set to the initial search vectors of the population individuals is expressed as follows

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

in Equation 6.

$$X_i^d = (UB^d - LB^d) \times \{r_d^i \times k\} + LB^d. \tag{6}$$

where X_i^d denotes the value of the d-th dimension of the i-th individual. UB^d and LB^d denote the upper and lower bounds of the d-th dimension of the search space. r_d^i denotes the proportional factor of the i-th individual in the d-th dimension. k is a scaling factor used to adjust the search range.

Figure 4 illustrates the two-dimensional initial population distributions generated by uniform distribution, good point set initialisation, logistic chaotic mapping, and Gaussian chaotic mapping. It could be observed that when the population sizes are 30, 45, and 60, the initial populations generated by the good point set are more evenly distributed. This effectively avoids the clustering and dispersion of individuals in specific regions, significantly enhancing the diversity of the population. As a result, the global search capability of the algorithm is improved, facilitating the discovery of the global optimal solution. In contrast, the other three strategies tend to generate populations with clustering or uncovered regions in the search space. For instance, logistic chaotic mapping often leads to individuals aggregating near the boundary while leaving large gaps in the center, which not only slows down convergence but also increases the risk of overlooking promising areas, thus hampering the search for high-quality individuals.

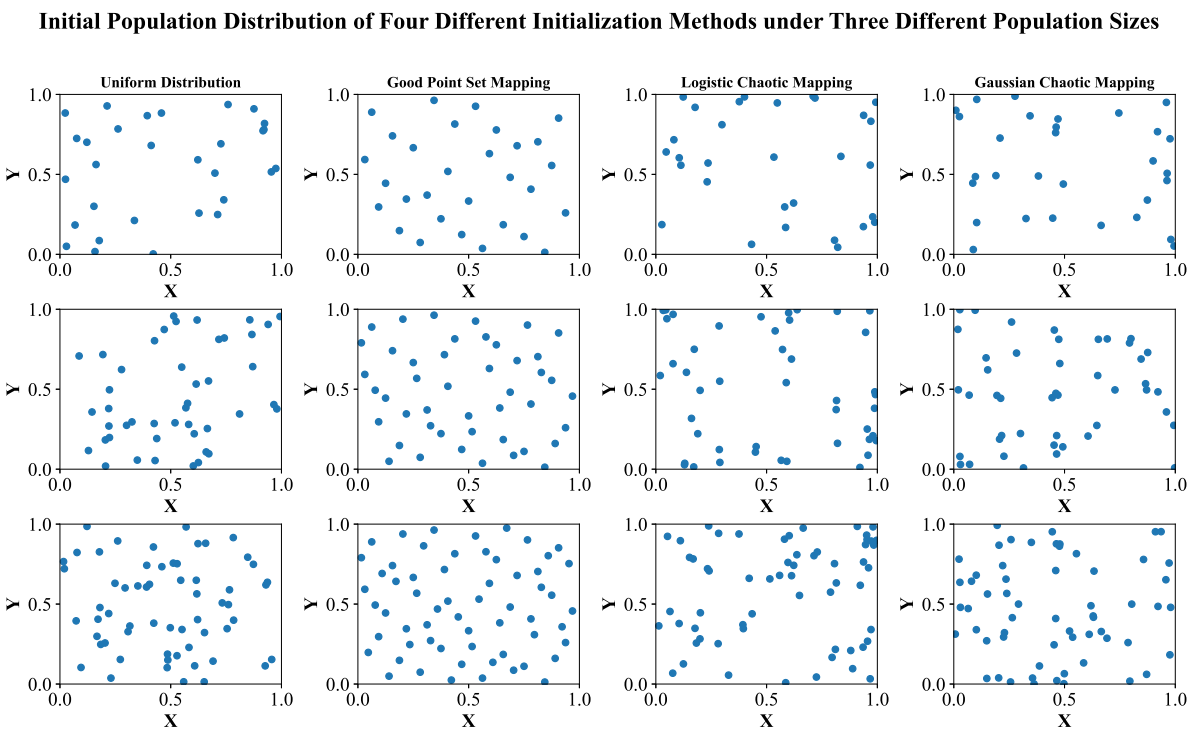

Figure 4: The initial populations generated by different initialisation strategies.

Additionally, since elite individuals often carry more effective search information, this study applies elite opposition-based learning to the population initialised by the good point set. Let $X_i = \left(x_i^1, x_i^2, \cdots, x_i^D \right), (i = 1, 2, \cdots, n)$ represent elite individuals in a D-dimensional search space. Their opposition-based individuals could be expressed by

ORIGINAL MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Equations 7 and 8.

$$\widetilde{X_i} = (\widetilde{x_i^1}, \widetilde{x_i^2}, \cdots, \widetilde{x_i^D}). \tag{7}$$

$$\widetilde{x_i^j} = K \cdot (LB^j + UB^j) - x_i^j, \quad j = 1, 2, \cdots, D. \tag{8}$$

where K is a dynamic coefficient within the range $[0, 1]$. UB^j and LB^j are the upper and lower bounds of the j-th dimension.

If the dynamic coefficient causes the opposition-based solution to exceed the search boundaries, rendering it an infeasible solution, it is corrected using a uniform distribution, as shown in Equation 9.

$$\widetilde{x_i^j} = \text{Uniform}(LB^j, UB^j). \tag{9}$$

In summary, this study constructs elite opposition-based individuals for the population initialised by the good point set. The optimal individuals are selected from the combination of the initial solutions from the good point set and their opposition-based individuals, forming the final initial solution set.

3.2.2 Adaptive difference operator selection strategy (DE)

To address the absence of an effective strategy for updating the maintainer line in the basic HRO algorithm, this study proposes an adaptive difference operator-based dynamic selection strategy to improve the quality of maintainer line. This includes the global difference operator DE_1 for the global search phase, the transitional difference operator DE_2 for transitioning from the global search phase to the local search phase, and the local difference operator DE_3 for the local exploitation phase, as described in Equations 10 - 12.

$$DE_1 : X_i^d(t+1) = X_i^d(t) + r_1 \cdot (X_{p_1}^d(t) - X_{p_2}^d(t)) + (1 - r_1) \cdot (X_{p_3}^d(t) - X_i^d(t)). \tag{10}$$

$$DE_2 : X_i^d(t+1) = X_i^d(t) + F \cdot (X_{p_1}^d(t) - X_{p_2}^d(t)) + F \cdot (X_{\text{best}}^d(t) - X_i^d(t)). \tag{11}$$

$$DE_3 : X_i^d(t+1) = X_{\text{best}}^d(t) + r_2 \cdot (X_{p_1}^d(t) - X_{p_2}^d(t)) + (1 - r_2) \cdot (X_{p_3}^d(t) - X_{p_4}^d(t)). \tag{12}$$

where $X_i^d(t+1)$ denotes the updated value of the d-th dimension of the i-th maintainer population individual in the $(t+1)$-th iteration. $X_{best}^d(t)$ denotes the value of the d-th dimension of the best solution found in the t-th iteration. X_{p_m} denotes a randomly selected individual from the maintainer line $(p_m \neq i)$, and $(p_i \neq p_j, \quad i \neq j, \quad 1 \leq i, j \leq 4)$. F is a smoothing factor controlling the transition of the maintainer population from global search to local search, with values in the range $[0, 1]$. r_1 and r_2 are random numbers in

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

the range of $[0, 1]$.

To enable the algorithm to select different difference operators with varying probabilities during different optimisation phases, this study defines an adaptive probability generation method to dynamically generate selection probabilities for the three types of difference operators. The selection process utilises a roulette wheel algorithm. The formulation is as follows, Equations 13 - 19.

$$s_1 = \frac{peak}{1 + \exp\left((t - \frac{T}{6})/\frac{T}{25}\right)} + s_{min}. \tag{13}$$

$$s_2 = peak \cdot \exp\left(-(t - \frac{T}{2})^2/(10 \cdot T)\right) + s_{min}. \tag{14}$$

$$s_3 = \frac{peak}{1 + \exp\left(-(t - 5 \cdot \frac{T}{6})/\frac{T}{25}\right)} + s_{min}. \tag{15}$$

$$peak = s_{max} - s_{min}. \tag{16}$$

$$S = s_1 + s_2 + s_3. \tag{17}$$

$$p_i = s_i/S, i = 1, 2, 3. \tag{18}$$

$$DE_s = roulette\_wheele\_selection(p). \tag{19}$$

where s_{min} and s_{max} are boundary adjustment parameters to prevent excessively small or large values before normalisation. $p_i (i = 1, 2, 3)$ denotes the normalised probability of selecting the i-th difference operator. t and T denote the current iteration count and the maximum number of iterations, respectively. DE_s denotes the difference operator ultimately selected using the roulette wheel strategy.

Figure 5 illustrates the adaptive dynamic adjustment of selection probabilities for the three different operators as the number of iterations increases. At the early stage of optimisation, the algorithm selects the global difference operator DE_1 with a higher probability to emphasize global exploration. During the transition phase from global to local search, the transitional difference operator DE_2 is chosen with a higher probability of balancing exploration and exploitation. In the later stages of optimisation, the local search operator DE_3 is selected with a higher probability to refine and enhance the optimal solution obtained.

3.2.3 t-distribution mutation perturbation strategy (TD)

The paper addresses the lack of differentiation between the early global exploration phase and the later local exploitation phase in the original HRO update strategy by introducing a

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

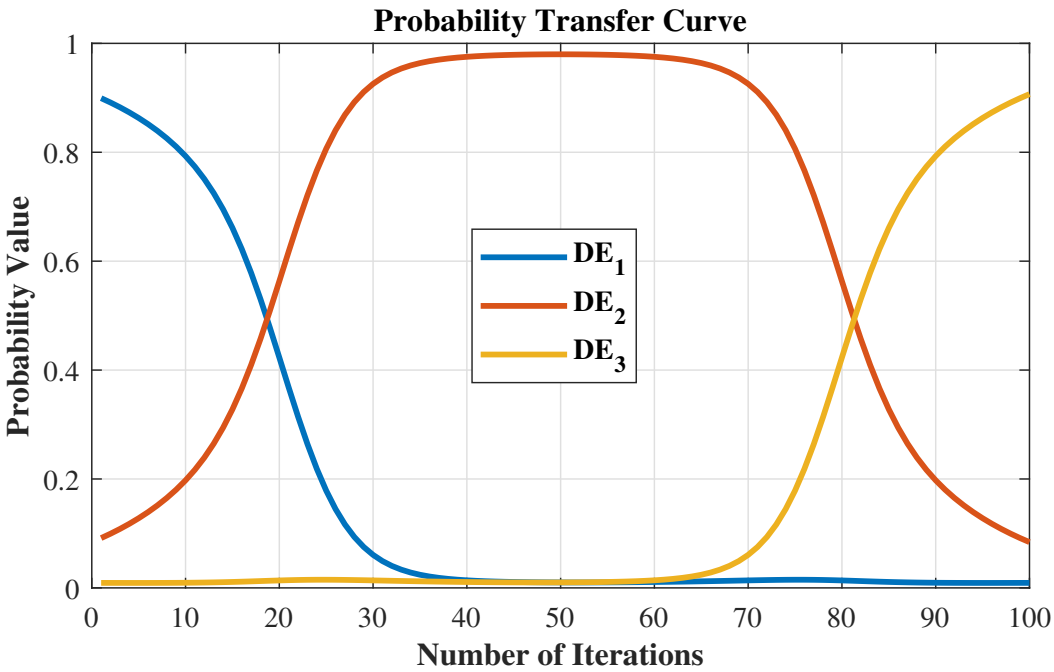

Figure 5: Differential operator selection probability transfer curve.

t-distribution-based mutation perturbation strategy to improve the hybridisation process. By leveraging t-distribution characteristics, the hybridisation phase in HRO is improved (Equations 20 - 23), where the original random number r_1 is replaced by t-distribution sampling.

$$X^d_{new(i)}(t) = \mathbf{t} \cdot X^d_{s(j)}(t) + (1 - \mathbf{t}) \cdot X^d_m(t). \tag{20}$$

$$f(t|df, \sigma) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\Gamma\left(\frac{df}{2}\right)\sqrt{df\pi}\sigma}\left(1 + \frac{1}{df}\left(\frac{t}{\sigma}\right)^2\right)^{-\frac{df+1}{2}}. \tag{21}$$

$$df = 2 + \frac{t}{T} \cdot 28. \tag{22}$$

$$\sigma = (\sigma_{max} - \sigma_{min}) \cdot \left(1 - \left(\frac{t}{T}\right)^{gr}\right) + \sigma_{min}. \tag{23}$$

where $X^d_{new(i)}(t)$ denotes the updated value of the d-th gene of the i-th sterile line individual at iteration t. $X^d_{s(j)}(j \neq i)$ and $X^d_m(t)$ denote the individuals selected randomly from sterile line and maintainer line, respectively. Each dimension of a newly generated sterile line individual is perturbed by a random variable generated from the t-distribution. Γ denotes the Gamma function, σ_{max} and σ_{min} are the maximum and minimum mutation scales, controlling the range of generated random numbers. Larger scales result in broader mutation ranges, and smaller scales result in narrower ranges. gr denotes the growth rate that controls the convexity of the σ variation curve.

Figure 6 illustrates the shapes of the t-distribution under different degrees of freedom and mutation scales. This study leverages the characteristics of this distribution

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

by combining various degrees of freedom and mutation scales to control the shape of the t-distribution, to dynamically adjust between global exploration and local exploitation stages for individuals. In the initial stages of iteration, the degree of freedom is relatively small while the mutation scale is large. At this point, the t-distribution approximates a Cauchy distribution, with data being more dispersed, resulting in larger perturbations that drive the updates of individuals toward global exploration. As the iterations progress, the degree of freedom gradually increases, and the mutation scale decreases. The t-distribution then transitions towards a normal distribution with a smaller standard deviation, generating smaller perturbations that make individuals more inclined to search within local regions.

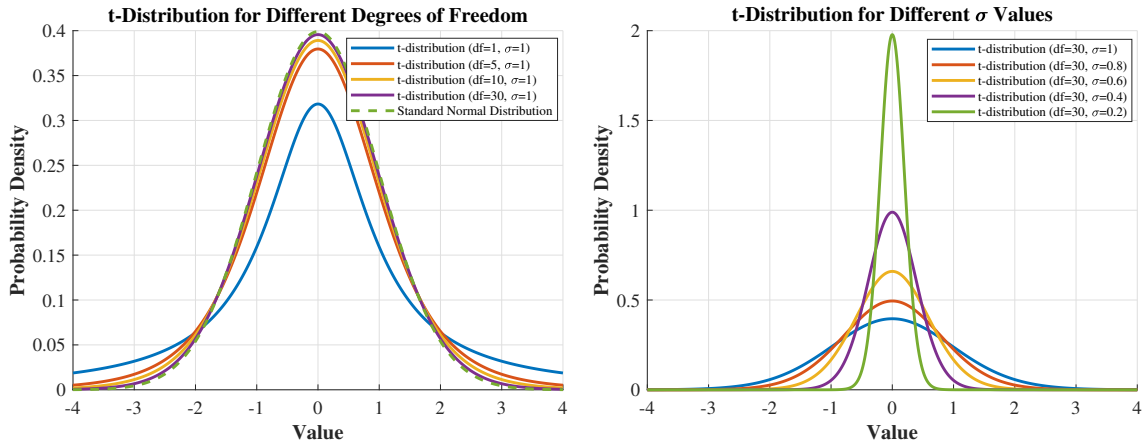

Figure 6: The shapes of t-distributions corresponding to different degrees of freedom and mutation scales.

3.2.4 Enhanced adaptive selfing strategy (SC)

To address the limitation of simply setting the selfing upper limit parameter (SC) as a constant during the selfing process, this study proposes an adaptive dynamic adjustment mechanism based on the iteration count and introduces a Lévy flight mechanism to improve the update of recovery system individuals.

To overcome the deficiency of SC being fixed as a constant in the basic HRO, the value of SC is improved by making it adaptively adjustable with iterations, as shown in Equation 24.

$$SC = SC_{\min} + (SC_{max} - SC_{min}) \cdot \left(1 - \left(\frac{t}{T}\right)^2\right). \tag{24}$$

where SC_{max} and SC_{min} denote the values of the selfing upper and lower limit, respectively.

In the early stages of the algorithm, when the algorithm is in a global search state, most individuals have a higher probability of performing effective updates within a small number of iterations. At this point, SC is set to a relatively high value. If a seed individual reaches this value in the early stage, it indicates that the individual has already fallen into

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

a local optimum. In such cases, a forced reset operation is executed. As the iterations progress to later stages, the likelihood of individuals being trapped in the local optima increases. Therefore, setting SC to a smaller value at this stage can help individuals quickly escape from the local optima and improve the overall search efficiency.

Additionally, this study also improves the selfing strategy for individuals from restorer line, replacing the original gene update formula with the following Equations 25 - 27.

$$X_{new(i)}^{d}(t) = r_3 \cdot (X_{best}^{d} - X_{r(j)}^{d} + c_1 \cdot (LB^d + L)). \tag{25}$$

$$L = \alpha \cdot \text{Lévy}(\beta) \cdot (UB^d - LB^d). \tag{26}$$

$$c_1 = 2 \cdot \exp\left(-\left(\frac{4 \cdot t}{T}\right)^2\right). \tag{27}$$

where α is the step-size control factor, and c_1 is an adaptive parameter that nonlinearly decreases from 2 to nearly 0 as iterations progress, providing dynamic global and local search capabilities for the recovery system individuals. $X_{r(j)}^{d}$ denotes the d-th dimensional gene value of the j-th seed individual randomly selected from the recovery system ($j \neq i$). X_{best}^{d} denotes the d-th dimensional gene value of the best solution found up to the current iteration. r_3 is a random number in the range $[0, 1]$. Lévy(β) refers to the Lévy distribution with parameter, characterised by alternating short-distance searches and random long-distance searches. This property enhances the algorithm's global and local search capabilities. The specific form of the Lévy distribution is given in Equation 28.

$$\text{Lévy}(\beta) \sim \mu = t^{-\mu}, \quad (1 < \beta \leq 3). \tag{28}$$

where β controls the step length; smaller β promotes global exploration, larger β favours local exploitation. μ is the Lévy exponent related to β, influencing the probability of long jumps.

3.3 TSMS-HRO: a high-dimensional feature selection algorithm

Based on MS-HRO, the first-stage filter-based algorithm, along with transformation functions and classifiers, is integrated to adapt the algorithm for high-dimensional FS tasks (refer to Figure 7 for the TSMS-HRO flowchart, and Algorithm 3 for the pseudocode).

3.3.1 Binary encoding strategy and classifier

In common, most of metaheuristic algorithms were originally designed for continuous optimisation problems and cannot be directly applied to discrete optimisation problems like FS. The solutions obtained by HRO are continuous, and they need to be mapped to the FS solution space using a transfer function. The transfer function used in this study

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

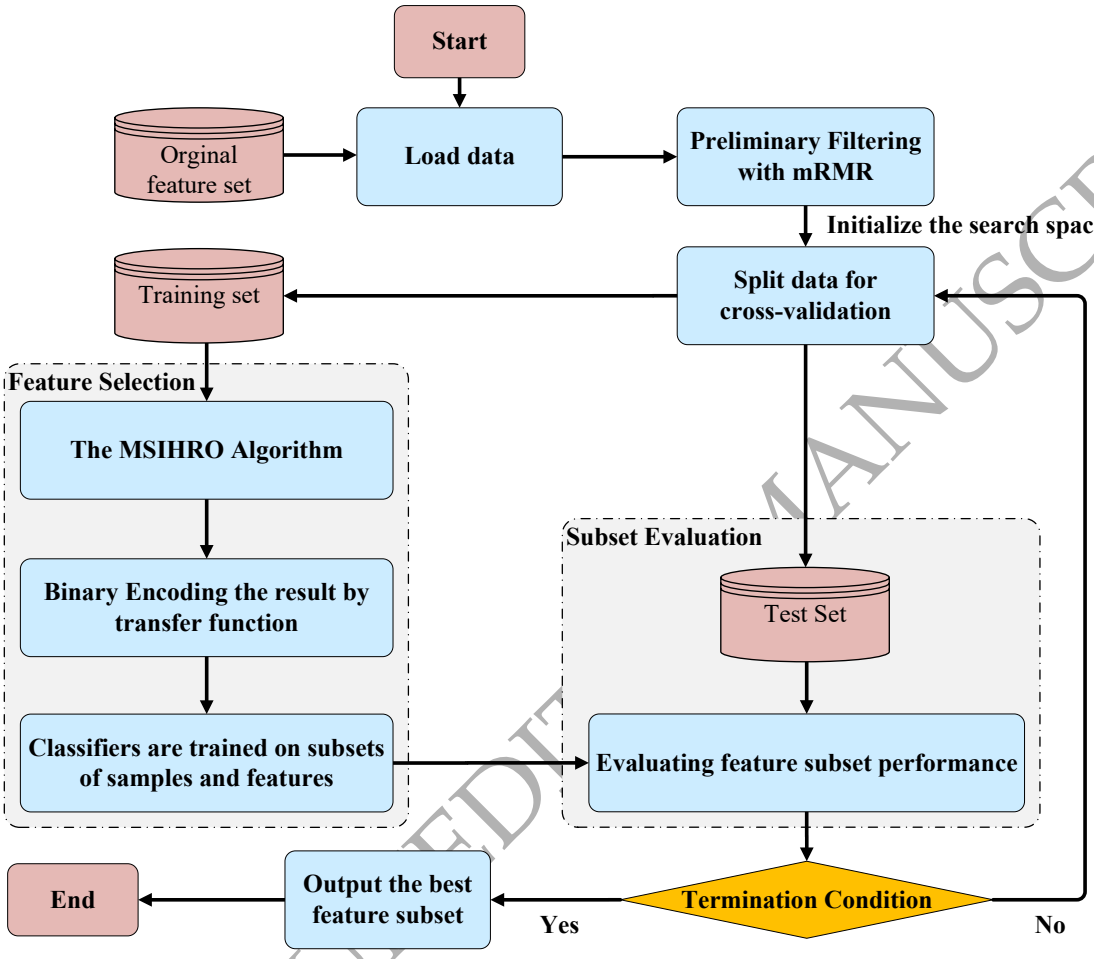

Figure 7: The Flowchart of TSMS-HRO.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

is defined as Equations 29 - 30.

$$S(x) = \frac{1}{1 + \exp(-x/2)}. \tag{29}$$

$$X_i^d = \begin{cases} 1, & if S(X_i^d) > rand, \\ 0, & otherwise. \end{cases} \tag{30}$$

In addition, to validate the effectiveness of the proposed algorithm, this study uses SVM as the classifier.

3.3.2 The fitness function

The fitness function used in this study is defined as Equation 31.

$$Fitness = \lambda \cdot error + \mu \cdot \frac{n}{N}. \tag{31}$$

where error denotes the classification error rate. n and N are the sizes of the selected feature subset and the total number of features, respectively. λ and μ are weighting factors used to balance the influence of the classification error rate and the feature subset size. The weights λ and μ satisfy the condition $\lambda + \mu = 1$, ensuring that the fitness values of all algorithms range between 0 and 1. This normalisation facilitates a fair comparison of the performance across different algorithms.

Algorithm 3 TSMS-HRO for Feature Selection.

Require: Subset S_1, maximum iterations $MaxIter$, population size n, $F \Leftarrow 0.5$, $s_{max} \Leftarrow 0.92$, $s_{min} \Leftarrow 0.01$, $\sigma_{max} \Leftarrow 1.0$, $\sigma_{min} \Leftarrow 0.1$, $g_r \Leftarrow 2$, $cr \Leftarrow 0.9$, $SC_{max} \Leftarrow 10$, $SC_{min} \Leftarrow 4$, $\alpha \Leftarrow 0.1$, $\lambda \Leftarrow 0.99$, $\mu \Leftarrow 0.01$.

Ensure: Best feature subset.

Initialise population $X \Leftarrow \{X_1, X_2, \ldots, X_n\}$ using:
  Use subset S_1 from Stage 1 as the initial search space;
  Good point set and elite reverse learning for diversity (Equations 6 – 9);
for $t = 1$ to MaxIter **do**
    Calculate fitness for each individual in the population;
    Sort the population by fitness:
      Maintainer line $X_m \Leftarrow \{X_1, X_2, \ldots, X_p\}$;
      Restorer line $X_r \Leftarrow \{X_{p+1}, \ldots, X_{2p}\}$;
      Sterile line $X_s \Leftarrow \{X_{2p+1}, \ldots, X_n\}$;
    Apply MS-HRO's improvement strategies;
    ▷ Discretization
    Apply S-shaped transfer function to convert continuous solutions to binary values (Equations 29 – 30);
    ▷ Fitness Evaluation
    Calculate fitness using error rate and feature subset size (Equation 31);
    Update the best solution found so far if the new solution is better;
end for
return Best feature subset.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

4. Experiment and Discussion

To assess the effectiveness of the proposed method, this section presents a series of benchmark experiments. It begins with a description of the experimental framework, including datasets and baseline algorithms, followed by a detailed analysis of the results.

The experiments were conducted on a PC running the Windows 10 operating system. The hardware specifications include a 13th Gen Intel(R) Core(TM) i9-13900K CPU @ 3.00 GHz and 64GB of memory. All algorithms were implemented using the Python programming language.

4.1 Experimental settings

To evaluate the proposed algorithm, five groups of experiments were organised:

1. Validation of MS-HRO's performance in continuous space: To demonstrate the optimisation capability of MS-HRO in a continuous space, this section uses the CEC2022 benchmark functions as benchmarks.

2. Integration of filter methods into MS-HRO: To enhance the search efficiency and the quality of selected feature subsets, the filter method was incorporated into MS-HRO. Several mainstream filter methods, including Fisher, Laplacian, Maximal Information Coefficient (MIC), MI, mRMR, ReliefF, and Trace Ratio Criterion (TR), were compared and evaluated.

3. Verification of classifier-independence: To examine whether the proposed method is adaptable to different classifiers, additional experiments were conducted using KNN, DecisionTree (DT), Naive Bayes (NB), XGBoost, and Random Forest (RF) while keeping the selected feature subsets unchanged. This allows us to evaluate the generalisation capability of the algorithm and confirm that its effectiveness is not limited to SVM.

4. Evaluation of TSMS-HRO on high-dimensional biomedical datasets: The classification performance of the TSMS-HRO algorithm was assessed (mRMR + SVM as the baseline algorithm) and analyzed on 12 high-dimensional gene expression datasets derived from biomedical microarray databases.

5. Ablation study of TSMS-HRO components: To evaluate the effectiveness of each mechanism in TSMS-HRO, we conducted an ablation study on multiple high-dimensional biomedical datasets by selectively incorporating different strategies into the algorithm.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

4.2 Dataset and parameter settings

To validate the performance of MS-HRO, the study utilised the CEC2022 benchmark functions, which include: 1 Unimodal Function, 4 Basic Functions, 3 Hybrid Functions, and 4 Composition Functions. For comparative experiments, the study selected several state-of-the-art optimisation algorithms as baselines, including HRO, GA, Honey Badger Algorithm (HBA), HHO, JAYA, Rime Optimisation Algorithm (RIME), SSA, WOA, AO, Clonal Selection Algorithm (CSA), GWO. The parameters of these comparison algorithms are provided in Table 1. All algorithms shared a consistent configuration: an initial population size of 42, a maximum of 1000 iterations, and 30 independent runs for each 20-dimensional test function, ensuring statistical reliability and fairness.

Table 1: Parameter Settings.

Algorithm	Parameter Setting
HRO	$r_1, r_2, r_3 \in [0,1]$, $SC = 8$
MS-HRO	$F = 0.5$, $s_{\max} = 0.92$, $s_{\min} = 0.01$, $\sigma_{\max} = 1.0$, $\sigma_{\min} = 0.1$, $g_r = 2$, $cr = 0.9$, $SC_{\max} = 10$, $SC_{\min} = 4$, $\alpha = 0.1$
GA	$CR = 1$, $MR = 0.01$
HBA	$C = 2$, $\beta = 6$, $F \in \{-1, 1\}$
HHO	$E_1 = 2(1 - \frac{t}{T})$, $E_0 \sim \mathcal{U}(-1,1)$, $E = E_1 \cdot E_0$, $q, r \sim \mathcal{U}(0,1)$, $J = 2(1 - \text{rand})$, Lévy($\beta = 1.5$)
JAYA	–
RIME	$W = 5$
SSA	$c_1 = 2e^{-(\frac{4t}{T})^2}$, $c_2 \sim \mathcal{U}(0,1)^n$, $c_3 \sim \mathcal{U}(0,1)^n$
WOA	$a \in [0,2]$, $A = 2ar_1 - a$, $C = 2r_2$, $b = 1$, $l \in [1, a_2]$, $p \in [0,1]$
AO	$\alpha = 0.1$, $\delta = 0.1$, $a = 2(1 - \frac{t}{T})$, $r_1 \sim \mathcal{U}(0,1)$, Lévy($\beta = 1.5$)
CSA	CR = 0.1, MR = 0.1, SR = 0.2
GWO	$\alpha \in [0,2]$

This study aimed to select a more effective filter method by evaluating mainstream filter techniques, including Fisher, Laplacian, MIC, MI, mRMR, ReliefF, and TR. Each filter method was compared and validated under varying fixed dimensions (ranging from 100 to 500, with an increment of 50) and feature ratios (ranging from 5% to 30%, with an increment of 5%). The evaluation criterion was based on calculating the average classification accuracy achieved by each filter method across all datasets in Table 2 when selecting the corresponding number of features.

To verify the classifier-independence of the proposed feature selection algorithm, multiple classifiers, including SVM, KNN, DT, NB, XGBoost, and RF, were employed for evaluation. Each classifier was executed 30 times on the 12 datasets listed in Table 2, and their performance was assessed under the same feature subsets generated by the proposed algorithm. The parameter settings for each classifier are summarised in Table 3.

To validate the performance of the proposed FS method, it is compared with the latest two-stage FS methods, such as MBAO (Pashaei, 2022), HFSIA (Zhu et al., 2023), MG-

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Table 2: Datasets used for feature selection by TSMS-HRO.

ID	Dataset	Instances	Number of Features	Number of Classes
D1	colon	62	2000	2
D2	lung	203	3312	5
D3	GLIOMA	50	4434	4
D4	leukemia_1	72	5327	3
D5	DLBCL	77	5469	2
D6	TOX_171	171	5748	4
D7	ALLAML	72	7129	2
D8	Brain_Tumor_2	50	10367	4
D9	Prostate_Tumors	102	10509	2
D10	CLL_SUB_111	111	11340	3
D11	SMK_CAN_187	187	19993	2
D12	GLI_85	85	22283	2

Note: These datasets vary significantly in the number of features, ranging from $2,000$ to $22,283$. Specifically, datasets $D1-D7$ are high-dimensional datasets with feature counts between $2,000$ and $7,129$, while datasets $D8$-$D12$ have even higher feature dimensions, ranging from $10,367$ to $22,283$. This diversity in feature numbers provides a broad spectrum of data perspectives and analytical layers for in-depth research.

Table 3: Classifier Parameter Settings.

Algorithm	Parameter Setting
SVM	$C = 4$
KNN	$K = 5$
DT	$criterion = gini, max\_depth = None$
NB	–
XGBoost	$max\_depth = 3, learning\_rate = 0.01, n\_estimators = 100, subsample = 1.0, colsample\_bytree = 1.0$
RF	$n\_estimators = 100, max\_depth = None, criterion = gini$

Note: The parameters are based on commonly used or default settings in the respective classifiers, without further hyperparameter tuning, to ensure a fair comparison under the same feature subsets.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

WOR (Pan et al., 2023), BIGWO (Moustafa et al., 2024), and improved two-stage methods, TS-GA, TS-GWO, TS-HBA, TS-HHO, TS-JAYA, TS-RIME, TS-SSA, TS-WOA, in 12 datasets listed in Table 2. These datasets were derived from gene expression profiles in high-dimensional biomedical microarray databases (Ghosh et al., 2021; J. Li et al., 2017). The collection of such datasets typically relies on high-throughput technologies, such as microarray techniques or next-generation sequencing, which enable the measurement of thousands of gene expression levels in a single experiment. These datasets are widely used in disease diagnosis, biomarker discovery, and drug development. By analysing differences in gene expression patterns, researchers can uncover molecular mechanisms underlying specific diseases and support the development of novel therapeutic approaches. The experiments were conducted under a unified standard. Each algorithm was run 30 times on each data set. The population size was set to 30, the minimum number of iterations was 100, and the filters used by the improved TS-GA, TS-GWO and other algorithms were all mRMR (the parameters were consistent with TSMS-HRO), and the other algorithm parameters were consistent with those in Table 1.

To evaluate the effectiveness of each mechanism within TSMS-HRO, an ablation study was conducted by selectively incorporating different strategies into the algorithm. The impact of each component on performance was assessed across multiple high-dimensional biomedical datasets in Table 2 to determine its contribution to the overall effectiveness of the algorithm. The parameters of all comparison algorithms are consistent with TSMS-HRO. The comparison algorithm list and its explanation are shown in Table 4.

Table 4: Comparison of algorithms in ablation experiments.

Algorithm	Description
HRO	Original HRO algorithm.
TS-HRO	Two-stage HRO algorithm with mRMR-based filter.
MS-HRO	Multi-strategy improved HRO algorithm.
TSMS-HRO	Two-Stage Multi-Strategy HRO algorithm with mRMR-based filter.
TS-HRO-DE	Two-stage DE-improved HRO algorithm with mRMR-based filter.
TS-HRO-INIT	Two-stage INIT-improved HRO algorithm with mRMR-based filter.
TS-HRO-SC	Two-stage SC-improved HRO algorithm with mRMR-based filter.
TS-HRO-TD	Two-stage TD-improved HRO algorithm with mRMR-based filter.

Note: The parameters of all variant algorithms are consistent with TSMS-HRO.

4.3 Performance metrics

4.3.1 The result of the CEC2022 benchmark tests

MS-HRO demonstrates significant advantages in most benchmark functions. Specifically, in test functions $F5$, $F6$, $F8$, $F9$, $F10$, $F11$, and $F12$, MS-HRO outperforms other algorithms, achieving solutions that are closer to the optimal. Additionally, for functions $F1$, $F2$, $F3$, and $F7$, the algorithm consistently converges to competitive solutions, albeit

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

slightly behind the top-performing algorithm in those cases. Moreover, we can observe that MS-HRO performs better in the Hybrid and Composition functions. Among the 7 functions from $F6$ to $F12$, MS-HRO achieved the best average fitness values in 6 cases and demonstrated excellent performance in terms of variance, indicating a certain level of stability. This suggests that MS-HRO possesses strong search capabilities, allowing it to explore the search space and identify the optimal values thoroughly. For example, in $F8$, MS-HRO outperformed other methods by several orders of magnitude, reaching 2.36E+3, and achieved the lowest variance of 2.66E+2 compared to other algorithms in the same group. However, at the same time, we can also observe that MS-HRO performs worse than HBA on Basic and Simple Multimodal functions.

In Figure 8 (mainly Unimodal Functions and Basic Functions), we first observe that the convergence speed of MS-HRO is not as fast as that of a few other algorithms. This is because, in order to avoid premature convergence to a local optimum, MS-HRO continues to explore unknown regions in search of better solutions. This exploratory behaviour also lays the groundwork for its excellent performance on subsequent Hybrid and Composite Functions. Secondly, although its convergence speed is relatively slow, MS-HRO still converges to the optimal value on certain functions, such as $F5$ and $F6$. On $F5$, the convergence of MS-HRO in the early iterations is not as strong as that of the original HRO and GWO algorithms. However, after 100 iterations, it surpasses all algorithms and achieves the optimal value. On $F6$, after nearly 400 iterations, MS-HRO escapes from the local optimum, outperforms the AO algorithm, which remains trapped, and reaches the best convergence value. In addition, MS-HRO performs slightly worse than the HBA algorithm on $F1$, $F2$, and $F3$, but the performance gap is minor. Notably, the convergence speed of MS-HRO is much faster than that of HBA, indicating a better trade-off between optimisation quality and convergence efficiency. Finally, on $F4$, MS-HRO fails to obtain a competitive solution before the end of the iteration, likely due to being trapped in a local optimum. The rugged and multimodal nature of $F4$ may not align well with MS-HRO's search dynamics, leading to reduced exploration and premature convergence. Nevertheless, its convergence trend still indicates potential for further improvement.

In Figure 9 (Hybrid Functions and Composition Functions), we observe that, in terms of convergence value, MS-HRO achieves the optimal results on 5 out of 6 functions, with the exception of $F7$, where its performance is slightly worse than that of the original HRO. Overall, MS-HRO shows a significant advantage. Although the original HRO obtains relatively competitive results on $F8$, $F9$, and $F12$, its performance still falls short of that of MS-HRO. This demonstrates the effectiveness of our multi-strategy improvement approach. Regarding convergence speed, MS-HRO remains ahead of most algorithms even when dealing with more complex functions. This highlights the efficiency of the proposed exploration strategy in finding high-quality solutions more rapidly. For instance, MS-HRO converges to the optimal value on $F10$ within approximately 300 iterations, while other algorithms continue to search without reaching optimal convergence under the same

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

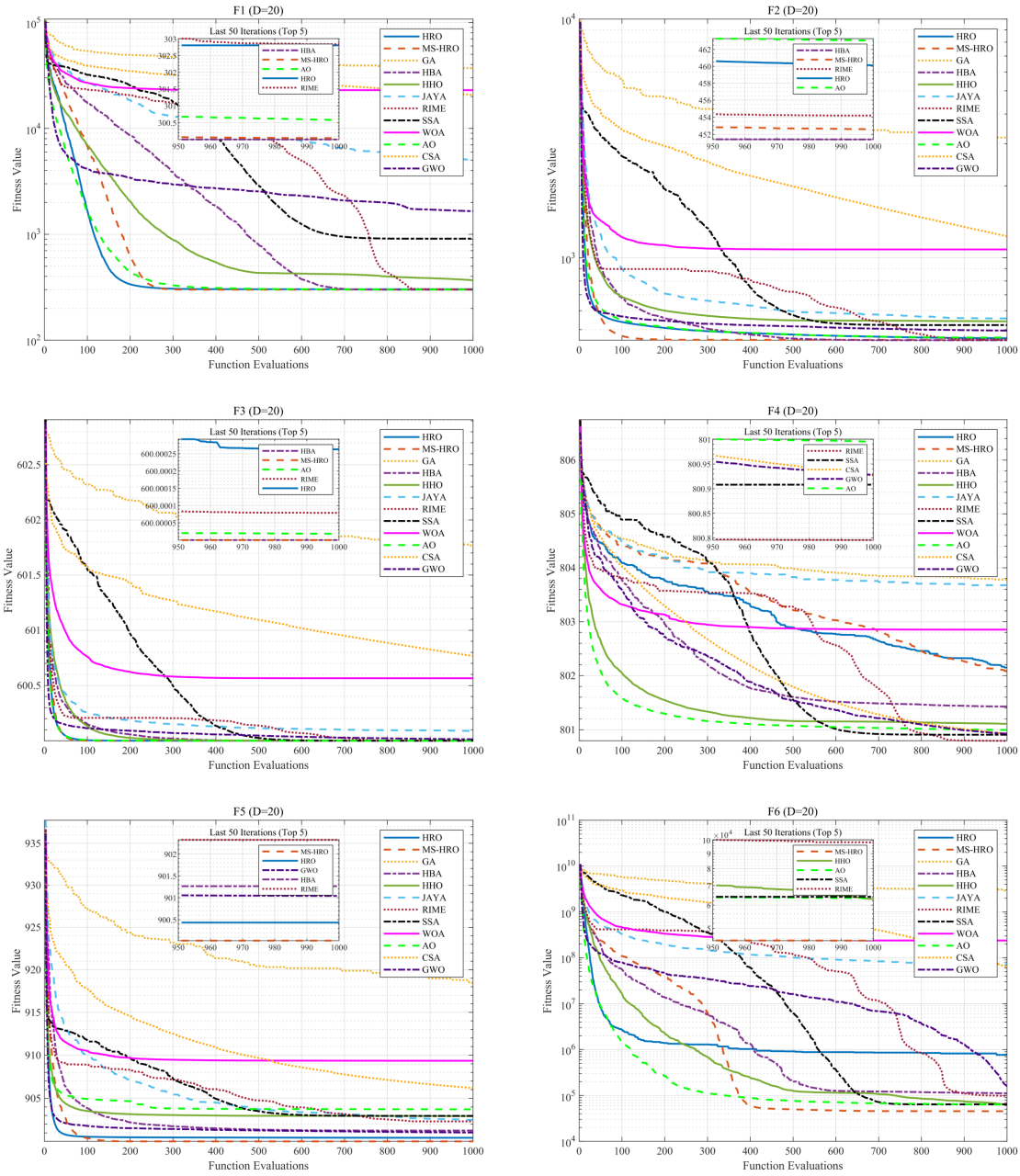

Figure 8: Convergence curves of different algorithms on CEC2022, (F1 - F6). The proposed method MS-HRO is represented by an orange dotted line.

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

ORIGINAL ARTICLE

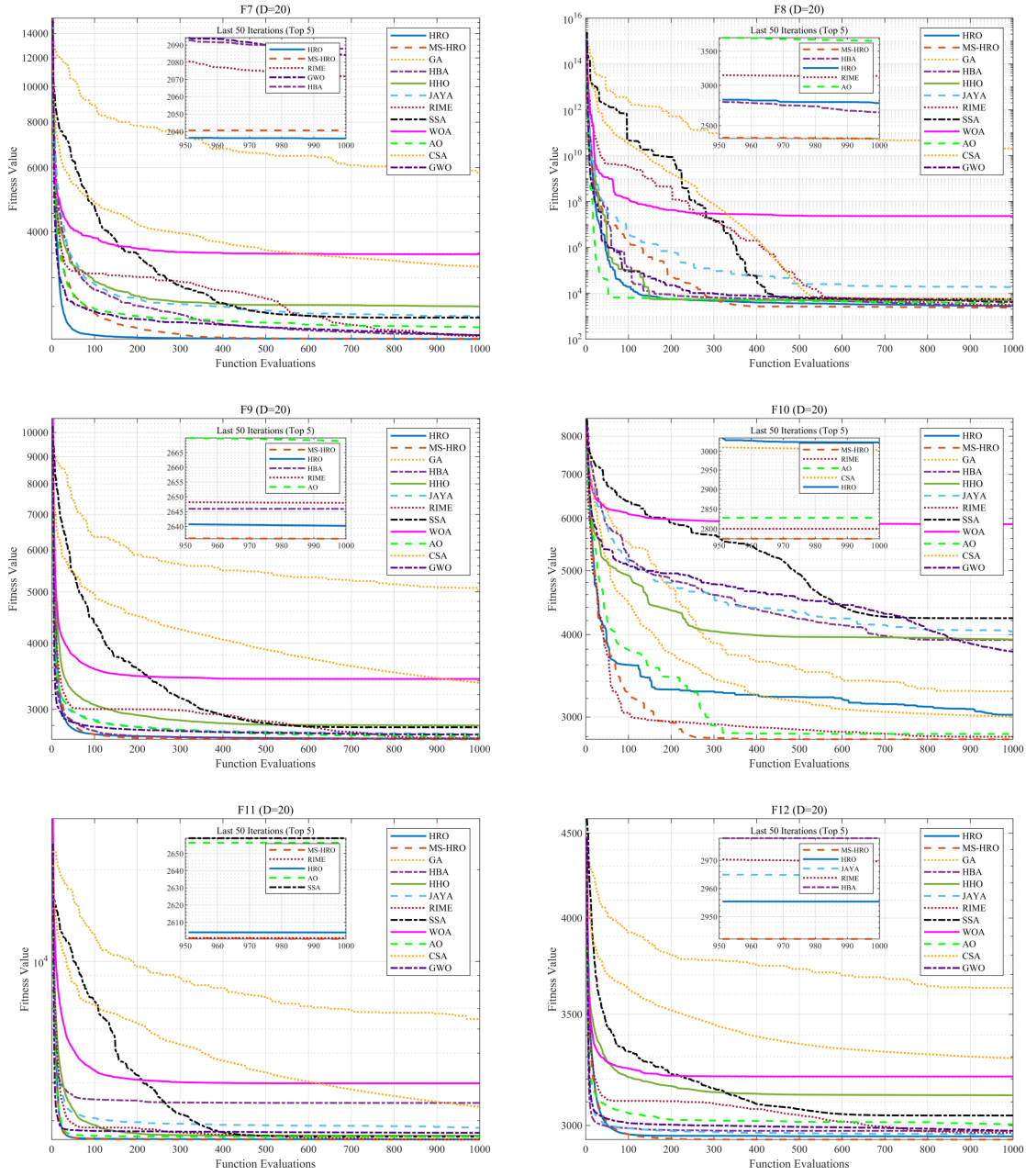

Figure 9: Continued (F7 - F12).

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

ORIGINAL ARTICLE

iteration count.

To further compare the performance differences between MS-HRO and other algorithms in terms of convergence value, variance, and runtime, we present a bar chart. Additionally, we take the logarithm of the convergence values to make the comparisons more visually apparent. In Figure 10 (mainly Unimodal Functions and Basic Functions), on $F1$, $F2$, and $F3$, although the HBA algorithm achieves slightly better convergence values, the differences between HBA and MS-HRO are marginal. MS-HRO still converges to highly competitive solutions. For instance, on $F1$, the difference between MS-HRO and HBA is only 3.00E−3, which is negligible. On $F2$, HBA achieves 4.51E+2, and MS-HRO achieves 4.53E+2, with a gap of only 1.86, while also obtaining the lowest variance, which highlights the stability of MS-HRO. On $F3$, the difference is merely 8.30E−12, which is practically insignificant. On $F4$, MS-HRO ranks eighth, likely due to being trapped in a local optimum. However, most algorithms—including MS-HRO—approach the optimal value of approximately 800, as shown in the figure. On both $F5$ and $F6$, MS-HRO achieves the best convergence values. Notably, it also achieves the lowest variance (0.08) on $F5$. On $F6$, MS-HRO performs particularly well, reaching 4.53E+4, significantly outperforming the second-best AO algorithm at 6.34E+4.

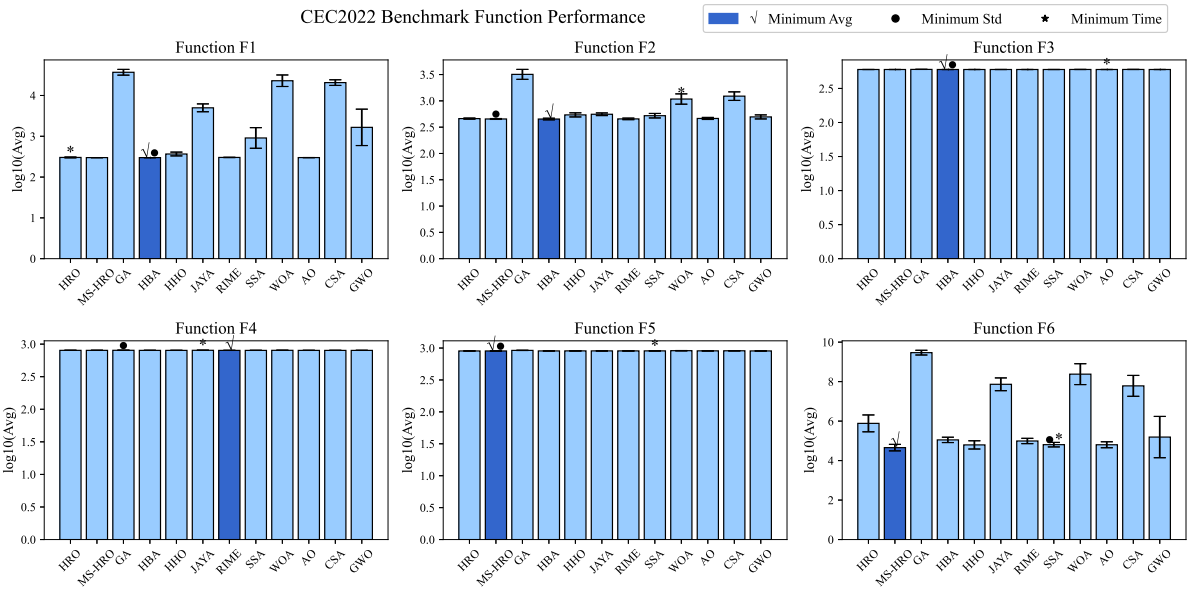

Figure 10: Bar chart comparing the performance of different algorithms on CEC2022. The algorithm with the best convergence value is marked with dark blue and a check mark; the one with the smallest variance is marked with a circle and an error bar; and the asterisk indicates the algorithm with the shortest running time (F1 – F6).

In Figure 11 (Hybrid Functions and Composition Functions), the performance from $F7$ to $F12$ is even more impressive. MS-HRO achieves both the best convergence values and the lowest variances on $F8$, $F9$, $F10$, $F11$, and $F12$, demonstrating strong optimisation ability and excellent stability when dealing with complex functions. Although the original HRO algorithm performs best on $F7$, the gap with MS-HRO is only 4.07—relatively small compared to the differences observed with other algorithms. Additionally, MS-

ORIGINAL MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

HRO demonstrates outstanding convergence values on other functions. For example, on $F8$, its result exceeds the second-best HBA algorithm by 2.93E+2. On $F12$, MS-HRO achieves 2.94E+3, whereas the second-best HRO algorithm only reaches 2.96E+3.

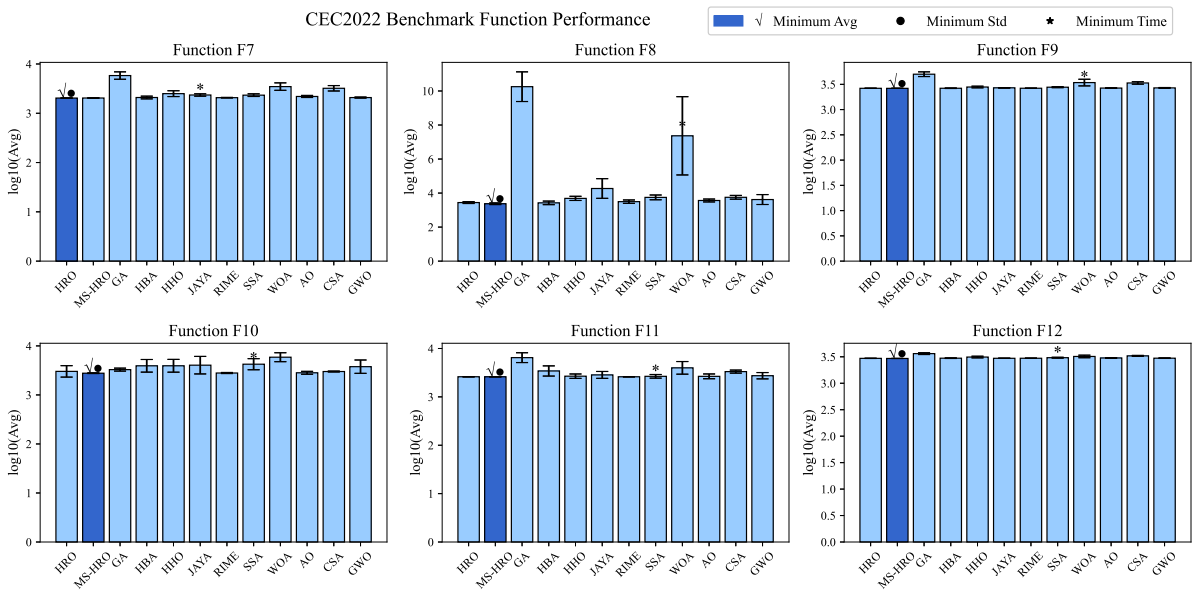

Figure 11: Continued. (F7 - F12)

In summary, after multiple iterations, MS-HRO demonstrates exceptional optimisation capabilities on the CEC2022 benchmark suite. These results can be attributed to the solid foundation of the original HRO algorithm and the effectiveness of the proposed enhancements, making MS-HRO a highly competitive optimisation method.

4.3.2 The result of the filter experiment

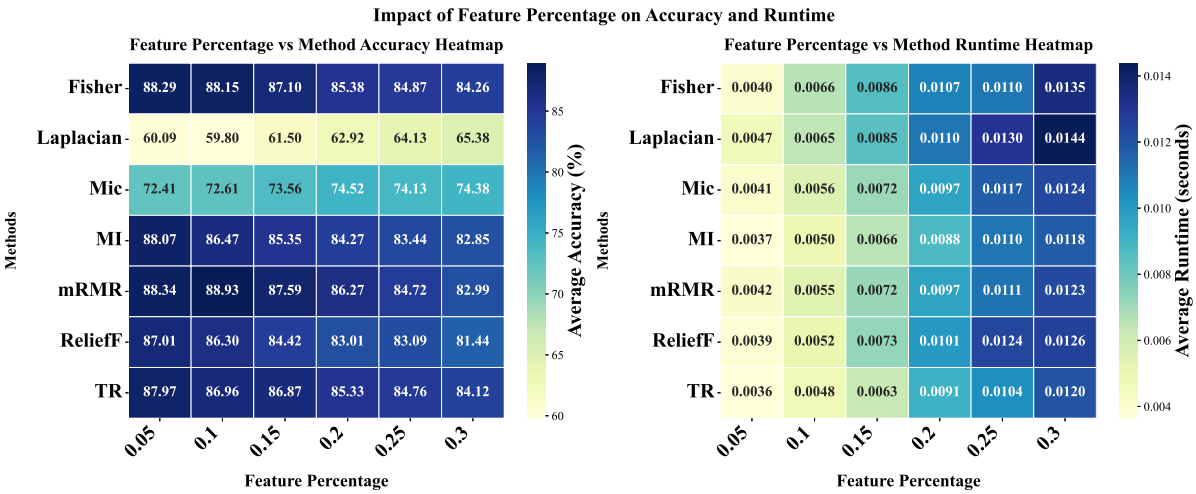

Figure 12: Heat map showing the average accuracy and average running time of each filter on 12 datasets at different feature percentages. Darker colors on the left indicate higher accuracy, while lighter colors on the right indicate shorter running times.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

As shown in Figure 12, within the feature ratio range of 0.05 to 0.30, the mRMR, Fisher, and MI methods achieved the highest classification accuracy. Among them, the mRMR method exhibited remarkable stability across all ratios, achieving a peak accuracy of 88.93% at a feature ratio of 0.1, thereby outperforming most other methods. In contrast, Laplacian yielded the lowest accuracy, which increased slightly with the feature ratio but remained significantly lower than that of other methods. Regarding runtime, all methods exhibited increased execution time as the feature ratio rose, which is an expected trend. Although TR had the shortest runtimes at lower feature ratios, the runtime of mRMR was comparable to that of other mainstream methods (such as reliefF and MI) and remained within an acceptable range.

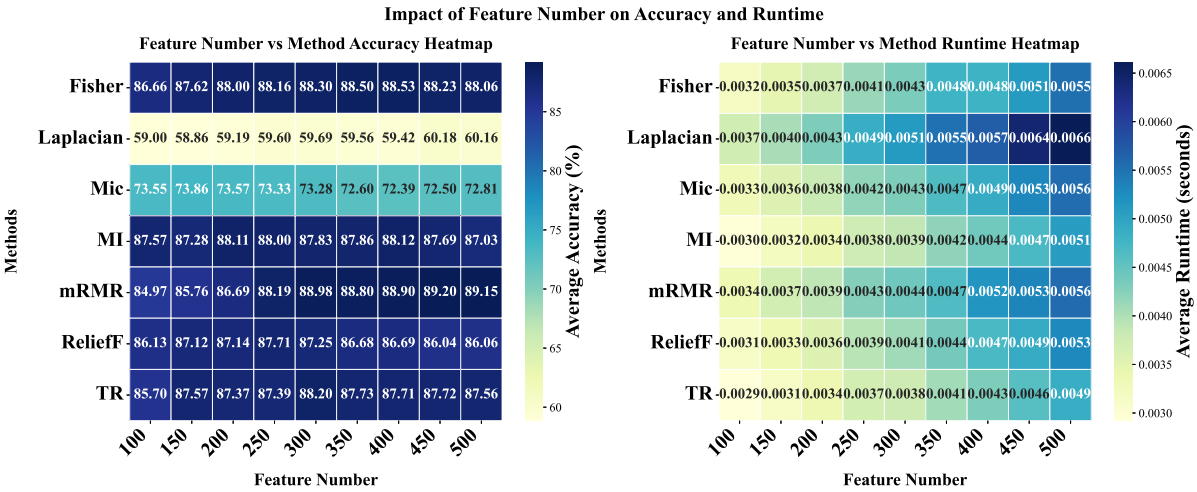

Figure 13: Heat map showing the average accuracy and average running time of each filter on 12 datasets with different numbers of features. Darker colors on the left indicate higher accuracy, while lighter colors on the right indicate shorter running times.

As shown in Figure 13, the classification accuracy of the mRMR method steadily increases as the number of features grows from 100 to 500, reaching peak values of 89.20% at 450 features and 89.15% at 500 features. Notably, at 300 features, the accuracy already reaches 88.98%, only 0.22% below the maximum, indicating it is very close to optimal. Compared to other methods (such as Fisher, MI, and ReliefF), mRMR consistently maintains a leading performance across medium and high-dimensional settings. In terms of runtime, mRMR shows a near-linear increase with the number of features, reaching approximately 0.0043 seconds at 300 features—substantially faster than the 0.0056 seconds at the maximum feature count. Given the marginal gain in accuracy versus the notable increase in computational cost, selecting 300 features represents a more cost-effective trade-off.

In conclusion, by comprehensively evaluating classification accuracy and computational efficiency, we selected the mRMR method with 300 features as the first-stage filter in our FS method.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

4.3.3 The result of the classifier experiment

As shown in Figure 14 and Table 5, the results demonstrate that the proposed feature selection method maintains high classification performance across different classifiers, which verifies the generalisation capability and classifier-independence of the selected features. Overall, all six classifiers (SVM, KNN, DT, NB, XGBoost, and RF) achieve consistently high accuracy values on most datasets, suggesting that the selected features preserve discriminative information independent of the specific classifier applied.

From an overall perspective, SVM consistently yields the best or near-best results, achieving the highest accuracy on 11 out of 12 datasets. For instance, it reaches perfect classification (100%) on four datasets (D4, D5, D7, D12), and above 95% accuracy on most others, with the exception of D11, where the accuracy drops to 85.93%. KNN also performs strongly, matching or surpassing SVM in several datasets, such as D2 (98.67%) and D12 (100%).

Other classifiers also exhibit competitive performance. NB achieves 100% accuracy on three datasets (D4, D5, D7), while RF attains 100% on D4 and nearly perfect accuracy on D5 and D12. XGBoost maintains balanced and stable results, often close to SVM (e.g., 97.65% on D12, compared with SVM's 100%). In contrast, DT shows relatively lower performance, particularly on D6 (76.10%) and D11 (77.09%), indicating higher sensitivity to dataset characteristics.

In summary, although SVM achieves the most stable and superior results overall, the fact that other classifiers—especially KNN, RF, and NB—also achieve competitive or even identical accuracy on several datasets demonstrates that the proposed method is not limited to a single classifier. This confirms that the selected features possess strong generalisation ability and classifier independence.

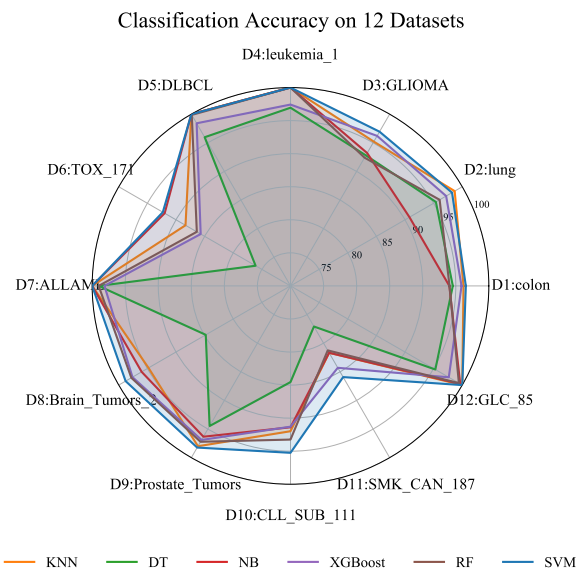

Figure 14: Radar chart of the average accuracy of each classifier running on 12 data sets.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Table 5: The average accuracy of each classifier running on 12 data sets.

Dataset	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
SVM	**96.52**	98.21	**96.93**	**100**	**100**	**92.30**	**100**	**98.80**	**98.23**	**95.25**	**85.93**	**100**
KNN	96.17	**98.67**	95.20	**100**	**100**	88.34	**100**	94.80	98.00	91.98	81.69	**100**
DT	94.55	95.39	92.80	96.95	95.97	76.10	99.02	84.80	94.47	84.55	77.09	95.29
NB	94.03	90.78	93.20	**100**	**100**	91.98	**100**	96.00	96.33	91.38	81.67	99.65
XGBoost	95.87	97.19	96.20	97.43	98.40	85.69	98.14	97.60	96.90	91.29	84.30	97.65
RF	94.05	96.05	92.40	**100**	99.87	86.36	99.02	97.80	97.22	93.24	81.26	99.41

4.3.4 TSMS-HRO two-stage high-dimensional feature selection results

From the convergence curves, it can be observed that TSMS-HRO achieved the highest accuracy on 10 out of the 12 datasets, demonstrating leading performance particularly on D1, D3, D8, and D11. As shown in Figure 15, for the first seven datasets (with feature dimensions ranging from 2,000 to 7,129), TSMS-HRO achieved the top accuracy in five cases and exhibited clear advantages on the D1 and D3 datasets. It not only outperformed the second-best algorithms, HFSIA and MGWOR, by nearly 2 percentage points but also demonstrated the fastest convergence to the optimal solution. Although the convergence speed of TSMS-HRO was relatively slower on the D2 and D6 datasets, it ultimately achieved strong results, ranking third and second, respectively. On the D4, D5, and D7 datasets, TSMS-HRO—as well as several other improved two-stage algorithms (including TS-GA, TS-HBA, TS-RIME, and HFSIA)—reached the maximum convergence accuracy of 100%. Furthermore, it is worth noting that these algorithms already exhibited high accuracy in the early stages of iteration, which can be largely attributed to the mRMR filter employed during the first selection stage.

As illustrated in Figure 16, on the final five high-dimensional datasets (with feature dimensions ranging from 10,367 to 22,283), TSMS-HRO delivered even stronger performance, consistently converging to the optimal accuracy. On D8 and D9, TSMS-HRO maintained superior convergence curves throughout the entire optimisation process, reflecting its excellent global and local search capabilities. Although it initially lagged behind HFSIA on D10, TSMS-HRO surpassed it after 40 iterations and eventually reached the best convergence value. Particularly on D11—the second most high-dimensional dataset—TSMS-HRO outperformed the second-ranked MGWOR algorithm by a margin of 5 percentage points, demonstrating a clear and significant advantage. Moreover, on D12, the dataset with the highest dimensionality, TSMS-HRO was the only algorithm to achieve 100% accuracy, further underscoring its superior capability in high-dimensional FS tasks.

In general, TSMS-HRO demonstrates excellent convergence ability, particularly on high-dimensional datasets. Its superior initial solutions can be attributed to the effectiveness of the first stage, which significantly reduces noise and redundancy. Meanwhile, the algorithm's ability to reach better final convergence values is largely due to the strengths of the original HRO algorithm and the enhancements introduced in the second-stage optimisation. This synergy between the two stages highlights the overall robustness and

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

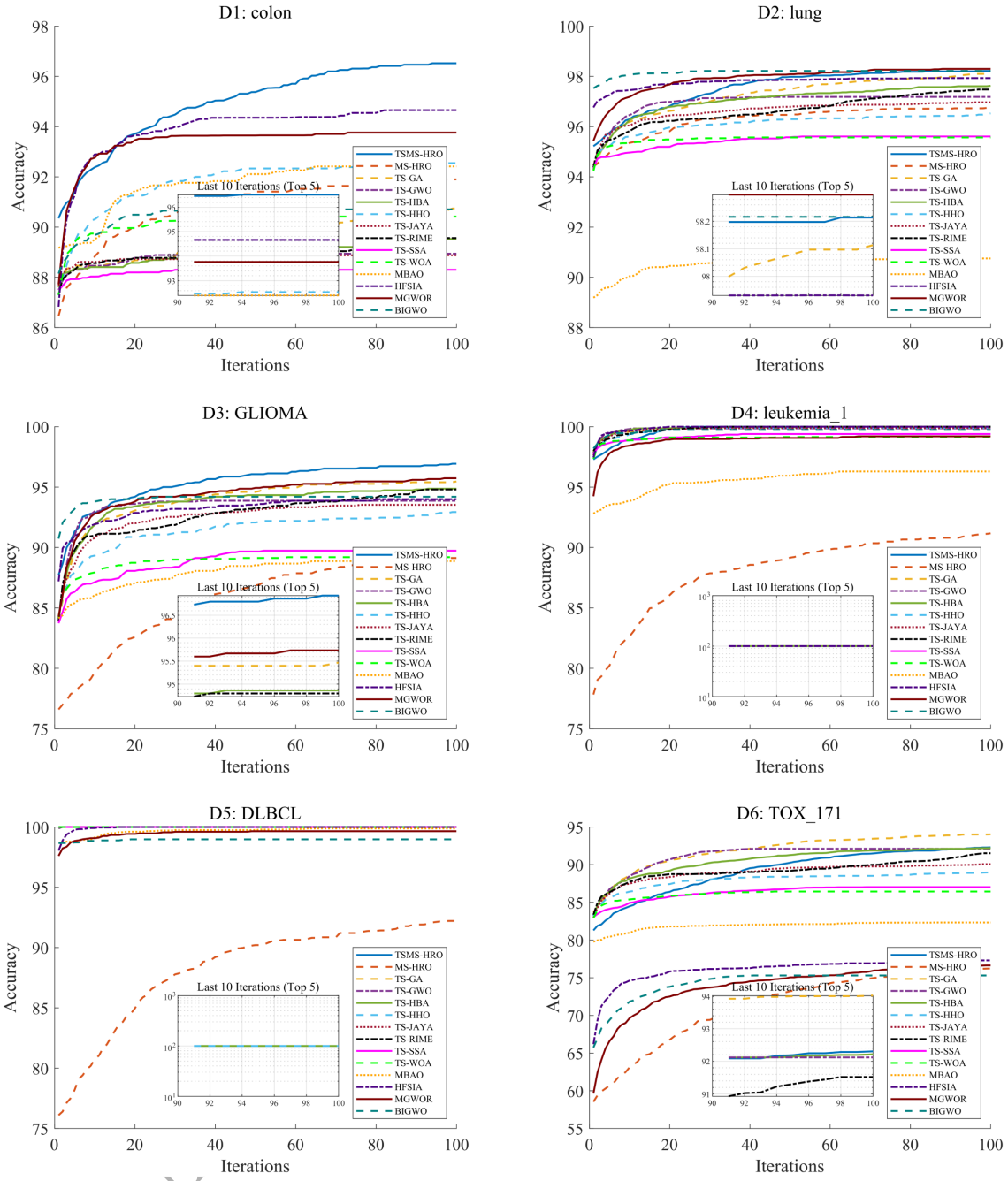

Figure 15: Convergence curves of TSMS-HRO and other algorithms on 12 datasets. The dotted line represents the convergence curve of TSMS-HRO.

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

ORIGINAL

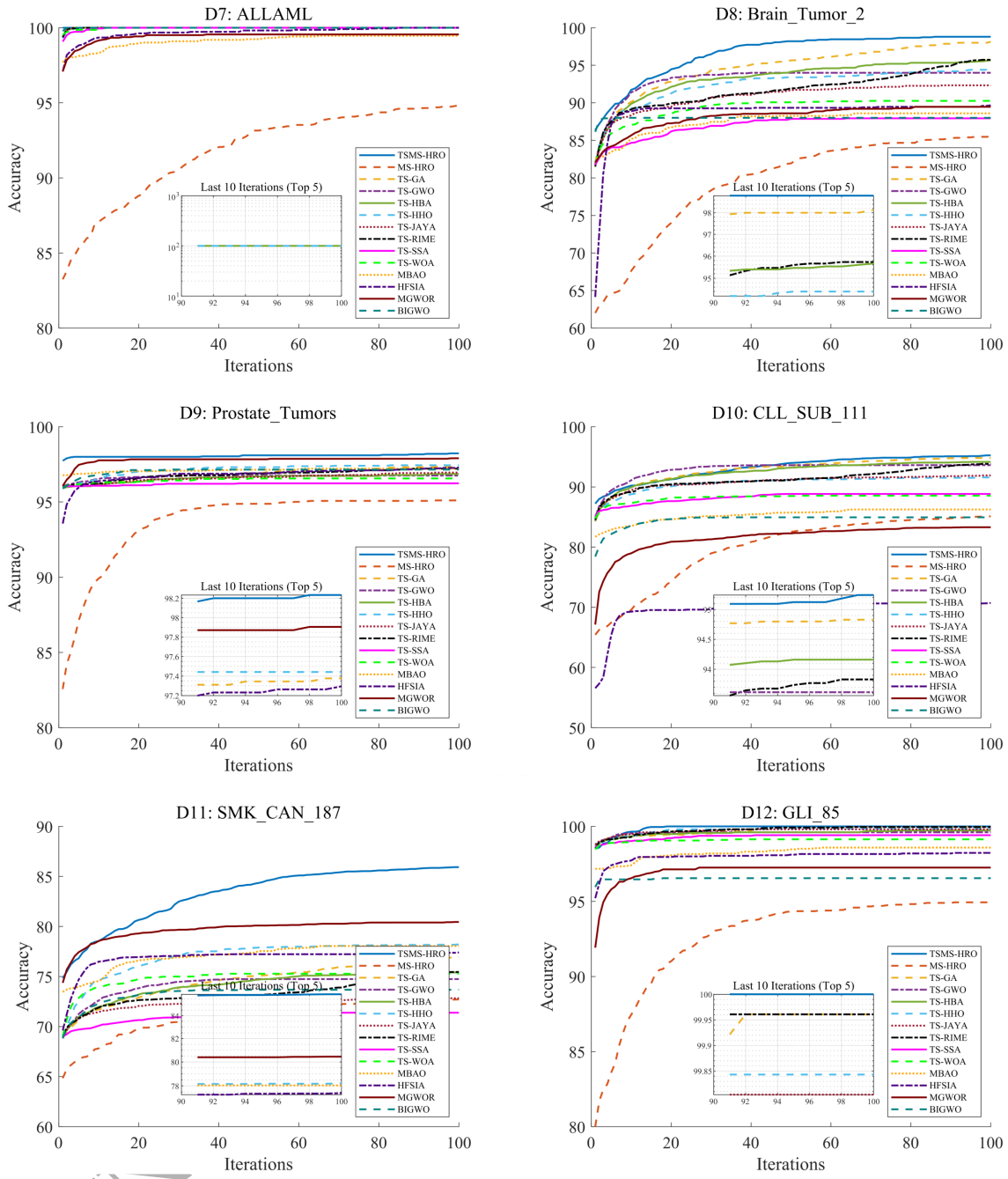

Figure 16: Continued.

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

ORIGINAL

effectiveness of the proposed TSMS-HRO method.

In order to further observe the indicators of TSMA-HRO in the high-dimensional FS task, we drew the bar graph and provided the detailed data, and the yellow horizontal line is used to represent the results of the baseline experiment. And Table 6 shows the accuracy of the baseline methods on 12 datasets, along with the improvements achieved by TSMS-HRO. We can see that after the improved HRO algorithm is used for feature selection, the accuracy rate is mostly improved by more than 20%.

Table 6: The average accuracy of baseline algorithm (mRMR + SVM) on 12 data sets.

Dataset	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
baseline	82.18	92.61	70.00	74.86	75.33	51.5	76.38	58.00	71.24	55.81	63.09	69.41
Improvements	+14.34	+5.60	+26.93	+25.14	+24.67	+40.80	+23.62	+40.80	+26.99	+39.44	+22.84	+30.59

As shown in Figure 17 and Table 7, the best accuracy was achieved on the datasets D1, D3, D4, and D5, reaching 96.52%, 96.93%, 100%, and 100% respectively. And the variance was 0 on the datasets D4 and D5, and TSMS-HRO showed strong stability. But at the same time, we can also observe that most of the other methods that added mRMR filters also achieved the same indicators, which shows that the mRMR method played a key role in the first stage, and the algorithm's search in the second stage was only to select fewer features and achieve the same effect. On the dataset D2, although TSMS-HRO did not reach the optimal value, the gap with the optimal algorithm MGWOR was only 0.07%, but TSMS-HRO had a smaller variance and was more stable. On the D6 dataset, TSMS-HRO ranked second with an accuracy of 92.30%, which is lower than the 94.03% of the TS-GA algorithm. However, it is worth noting that TSMS-HRO only selected 71.57 features to achieve the effect of TS-GA selecting 136.40 features.

As shown in Figure 18 and Table 8, TSMS-HRO performed better in the following six datasets with higher dimensions, all of which achieved the best accuracy. On the D7, D8, D9, D10, and D11, they achieved the best 100%, 98.80%, 98.23%, 95.25%, and 85.93%, respectively. At the same time, the variance on the D7 datasets was 0. In particular, TSMS-HRO achieves 100% accuracy and 0 variance on the dataset D12 with the highest data dimension (dimension is 22283), which shows the excellence of TSMS-HRO in high-dimensional FS.

At the same time, we also noticed that TSMS-HRO achieved 100% accuracy on some datasets, such as D4, D5, D7, and D12. This impressive performance is primarily due to the mRMR filtering method in the first stage of TSMS-HRO. Figure 15 and Figure 16 show that algorithms using the mRMR method in the first stage (including TSME-HRO, TS-GA, and TS-GWO etc.) consistently achieve excellent solutions in the early stages of iteration. On datasets that achieve 100% accuracy, the average accuracy of the initial population is above 95%. Furthermore, the convergence to 100% accuracy is closely related to the search performed by the improved HRO algorithm. During the iteration process, the algorithm further reduces data dimensionality while maintaining accuracy, thereby improving accuracy. Subsequent ablation experiments further validate

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

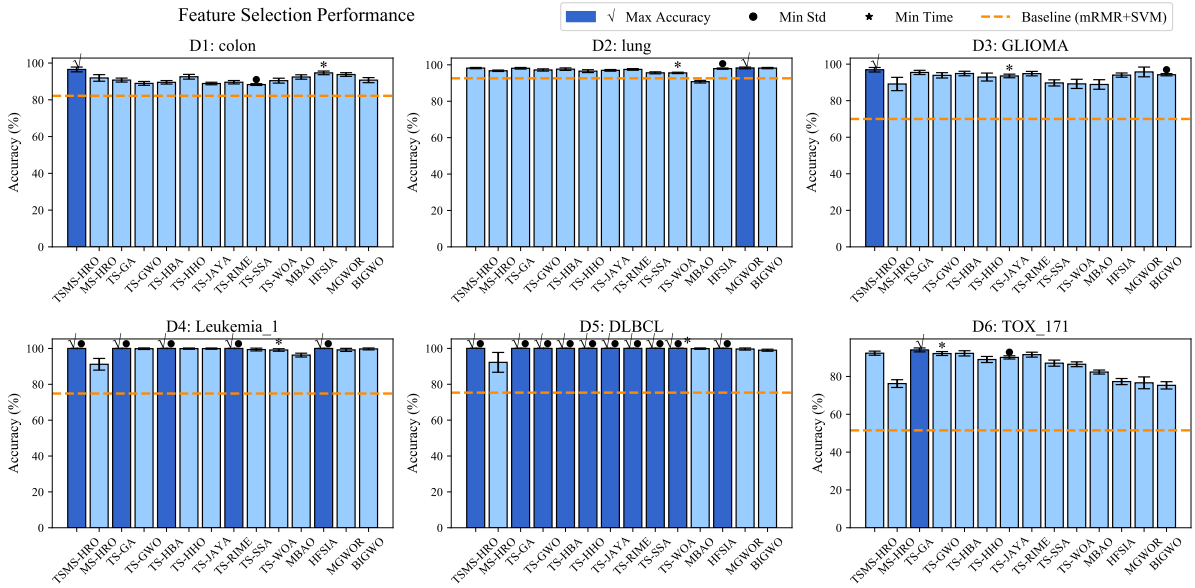

Figure 17: Bar chart comparing the performance of different algorithms on the FS problem. The algorithm with the highest accuracy is marked with a dark blue check mark; the one with the lowest variance is marked with a circle and error bars; and the asterisk indicates the algorithm with the shortest running time.

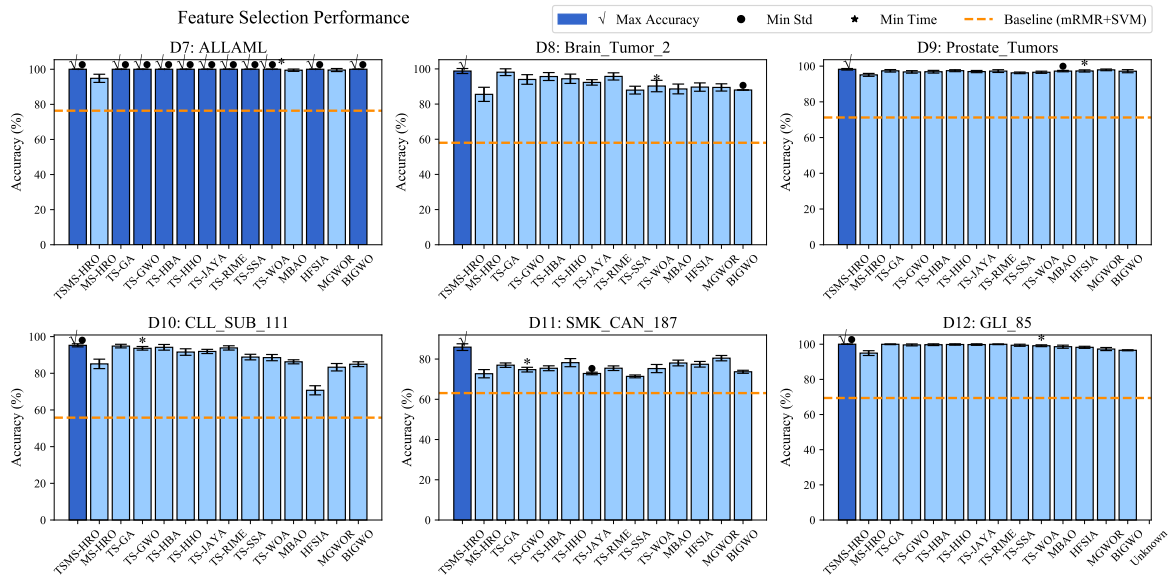

Figure 18: Continued.

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Table 7: Comparison Results of the Two-stage Improved Metaheuristic Algorithm on Datasets.

Dataset	Metric	TSMS-HRO	MS-HRO	TS-GA	TS-GWO	TS-HBA	TS-HHO	TS-JAYA	TS-RIME	TS-SSA	TS-WOA	MBAO	HFSIA	MGWOR	BIGWO
D1	Acc	**96.52**	91.90	90.74	88.94	89.53	92.55	88.88	89.56	88.30	90.42	92.41	94.65	93.76	90.71
	Std	1.30	1.79	1.08	1.04	0.95	1.29	0.69	0.93	**0.49**	1.335	1.20	0.99	0.99	1.38
	Num	15.97	74.97	81.20	90.83	76.40	12.20	130.83	78.83	128.83	26.87	8.03	6.30	**5.47**	87.87
	Time	5.35	20.71	6.06	4.81	19.54	13.49	4.61	13.66	11.75	4.53	40.40	**4.23**	8.03	4.74
D2	Acc	98.21	96.74	98.11	97.18	97.62	96.52	96.97	97.48	95.61	95.57	90.75	97.93	**98.30**	98.22
	Std	0.35	0.39	0.44	0.61	0.68	0.78	0.46	0.45	0.54	0.38	0.73	**0.35**	0.48	0.35
	Num	115.33	840.87	113.70	123.13	109.30	97.47	158.90	110.03	141.07	176.27	40.37	40.17	**22.93**	102.93
	Time	16.74	147.68	13.44	7.44	56.46	43.70	6.43	39.09	38.60	**5.97**	146.11	9.13	16.66	9.30
D3	Acc	**96.93**	89.13	95.47	93.87	94.87	92.93	93.53	94.80	89.73	89.20	88.87	94.00	95.73	94.20
	Std	1.24	3.64	1.15	1.45	1.23	2.17	0.99	1.22	1.69	2.51	2.62	1.15	2.67	**0.60**
	Num	47.67	187.07	102.33	112.60	96.70	59.70	154.13	100.27	141.53	101.73	30.73	17.13	**13.03**	90.47
	Time	4.94	52.86	5.20	4.04	16.84	11.75	**3.82**	11.68	11.32	3.91	40.02	6.22	8.41	4.53
D4	Acc	**100**	91.17	**100**	99.86	**100**	99.91	99.91	**100**	99.39	99.15	96.29	**100**	99.20	99.73
	Std	**0.00**	3.26	**0.00**	0.43	**0.00**	0.34	0.34	**0.00**	0.78	0.70	1.06	**0.00**	0.86	0.53
	Num	41.70	58.57	83.00	104.90	81.97	48.87	136.40	84.93	137.47	111.27	19.13	**8.87**	10.10	85.17
	Time	4.89	80.73	5.09	4.07	18.26	12.64	4.06	11.21	11.26	**3.47**	43.87	5.25	10.07	5.20
D5	Acc	**100**	92.21	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.88	**100**	99.65	98.98
	Std	**0.00**	5.53	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.00	**0.00**	0.38	**0.00**	0.58	0.56
	Num	23.97	8.73	67.57	82.30	72.87	18.20	122.50	75.07	123.77	55.03	24.83	**5.10**	5.67	75.40
	Time	4.83	97.91	6.23	4.81	21.46	13.33	3.97	14.06	13.95	**3.80**	46.53	4.95	9.72	4.55
D6	Acc	92.30	76.23	**94.03**	92.12	92.21	88.96	90.07	91.52	87.03	86.44	82.32	77.30	76.67	75.29
	Std	1.08	2.06	1.04	1.04	1.38	1.61	**0.85**	1.30	1.61	1.28	1.07	1.61	3.13	1.93
	Num	71.57	132.73	136.40	138.70	137.47	131.77	170.97	138.73	150.97	158.20	41.73	**10.07**	23.40	129.17
	Time	11.63	331.86	10.30	**6.53**	54.86	40.47	6.75	47.46	45.86	7.48	133.20	7.16	14.90	7.40

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Table 8: Continued.

Dataset	Metric	TSMS-HRO	MS-HRO	TS-GA	TS-GWO	TS-HBA	TS-HHO	TS-JAYA	TS-RIME	TS-SSA	TS-WOA	MBAO	HFSIA	MGWOR	BIGWO
D7	Acc	**100**	94.83	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.46	**100**	99.56	**100**
	Std	**0.00**	2.33	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.66	**0.00**	0.87	**0.00**
	Num	38.57	145.20	73.37	91.03	76.57	28.03	126.90	76.97	131.63	81.00	29.90	10.13	**8.10**	80.70
	Time	4.79	112.08	5.39	4.03	17.73	12.55	4.00	12.08	11.98	**3.95**	41.55	4.76	8.43	4.67
D8	Acc	**98.80**	85.53	98.13	94.00	95.67	94.40	92.33	95.73	87.93	90.27	88.60	89.67	89.47	88.00
	Std	1.42	4.02	1.93	2.73	2.26	2.65	1.56	2.05	2.22	3.26	2.79	2.37	2.06	**0.00**
	Num	60.77	26.17	104.23	116.90	102.80	38.57	155.20	103.27	139.13	71.73	24.07	**6.20**	14.13	83.93
	Time	5.60	137.57	5.24	4.00	17.39	11.79	3.69	10.93	11.84	**3.55**	40.43	4.12	8.35	4.33
D9	Acc	**98.23**	95.10	97.38	96.76	96.86	97.44	96.95	97.21	96.24	96.57	97.21	97.30	97.90	97.15
	Std	0.42	0.81	0.69	0.71	0.76	0.55	0.61	0.77	0.46	0.60	**0.39**	0.60	0.47	0.86
	Num	30.00	74.67	79.63	94.93	79.00	23.43	133.67	84.63	128.77	57.57	24.70	8.33	**5.57**	84.13
	Time	4.69	327.10	5.41	4.15	20.12	12.82	4.00	13.92	14.54	4.91	47.29	**3.74**	7.04	3.81
D10	Acc	**95.25**	85.12	94.83	93.62	94.16	91.58	91.91	93.83	88.84	88.55	86.26	70.73	83.30	84.95
	Std	**0.84**	2.58	0.99	0.99	1.55	1.78	1.14	1.17	1.53	1.70	1.11	2.49	2.02	1.32
	Num	128.73	83.90	128.80	129.60	120.43	92.47	166.93	121.80	141.87	113.90	25.50	**4.03**	15.17	124.03
	Time	7.72	339.49	5.20	**3.53**	21.92	16.85	4.46	23.05	26.62	6.13	88.49	6.29	13.05	5.60
D11	Acc	**85.93**	72.71	76.96	74.74	75.42	78.21	72.83	75.45	71.38	75.28	78.04	77.41	80.45	73.68
	Std	1.66	2.05	1.10	1.16	1.21	2.02	**0.64**	1.15	0.71	2.03	1.43	1.40	1.33	0.75
	Num	26.50	58.10	121.77	123.77	118.13	15.73	156.57	119.03	140.80	20.30	10.37	**5.17**	13.20	115.27
	Time	7.74	1226.14	6.16	**4.21**	13.40	12.82	5.82	13.66	25.20	5.05	82.09	4.89	9.93	6.48
D12	Acc	**100**	94.94	99.96	99.61	99.73	99.84	99.80	99.96	99.41	99.14	98.59	98.24	97.25	96.55
	Std	**0.00**	1.36	0.21	0.55	0.50	0.40	0.44	0.21	0.59	0.52	0.88	0.59	0.88	0.29
	Num	35.80	35.27	78.80	93.33	81.17	29.47	133.87	82.33	131.50	60.00	24.10	6.77	**6.57**	85.67
	Time	3.92	373.73	4.13	3.03	14.55	9.89	2.96	10.10	10.11	**2.88**	35.12	3.50	6.62	3.53

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

this observation by comparing MS-HRO and TSMS-HRO.

In terms of the number of selected features, TSMS-HRO has a smaller number of selected features than other traditional algorithms while ensuring leading accuracy, and there is no extreme number of selected features. Regarding the maximum number of features, TSMS-HRO limits the initial search space to 300 features filtered by mRMR. This algorithm directly removes many obvious redundant features, effectively reducing the data dimensionality. Furthermore, there is no case of too few features. If TSMS-HRO selects too few features, or even excludes core features, the final accuracy will certainly not be as good as it is now (for example, on the D10, HFSIA selected only 4.03 features, resulting in the lowest accuracy). The specific performance of TSMS-HRO in terms of feature count is as follows: TSMS-HRO selects 47.67 and 26.50 features on D3 and D11, respectively. Despite selecting more features than MGWOR (13.03 features on D3) and HFSIA (5.17 features on D11), its accuracy improves by 1.20 and 8.52 percentage points, respectively, compared to MGWOR and HFSIA. On the other hand, TS-GA selected 102.33 and 121.77 features on D3 and D11, respectively, far exceeding the number of features selected by TSMS-HRO. However, its accuracy dropped by 1.46 and 8.97 percentage points, respectively, as some redundant features interfered with classifier performance. This demonstrates that TSMS-HRO effectively balances the number of selected features, retaining important features while removing redundant ones to achieve higher accuracy.

In terms of algorithm running time, it still maintains a low running time on high-dimensional datasets and has a high overall efficiency. Compared with the MS-HRO method without filters, its running time has been greatly reduced, and the rate of running time reduction is accelerating as the dimension of the dataset increases. As shown in Figure 19, on the D1 dataset with the lowest dimension, TSMS-HRO reduces the running time by more than 15 seconds compared to MS-HRO, with a reduction ratio of nearly 75%. As the dimension increases, the reduction ratio continues to increase. On the D12 dataset with the highest dimension, the running time is reduced by 98.95%, which once again proves the effectiveness of the first stage improvements.

In order to comprehensively evaluate the performance difference between the proposed algorithm, TSMS-HRO, and other comparison algorithms, this paper uses the Friedman test and the Wilcoxon signed rank test with the Holm multiple comparison correction method for statistical analysis. Table 9 shows the average ranking and final ranking of each algorithm on all experimental tasks. The results show that TSMS-HRO ranks first with an average ranking of 2.21 and the narrowest 95% confidence interval $[94.21, 99.49]$, significantly better than other algorithms. The Friedman test statistic is 73.37, and the corresponding p-value is 1.91×10^{-10}, which is significantly less than 0.05, indicating that there are significant differences in overall performance between different algorithms.

Further, to verify the pairwise significant differences between TSMS-HRO and other algorithms, this paper conducts paired comparisons based on the Wilcoxon signed rank test and uses the Holm method to correct for multiple comparisons. The results are shown

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

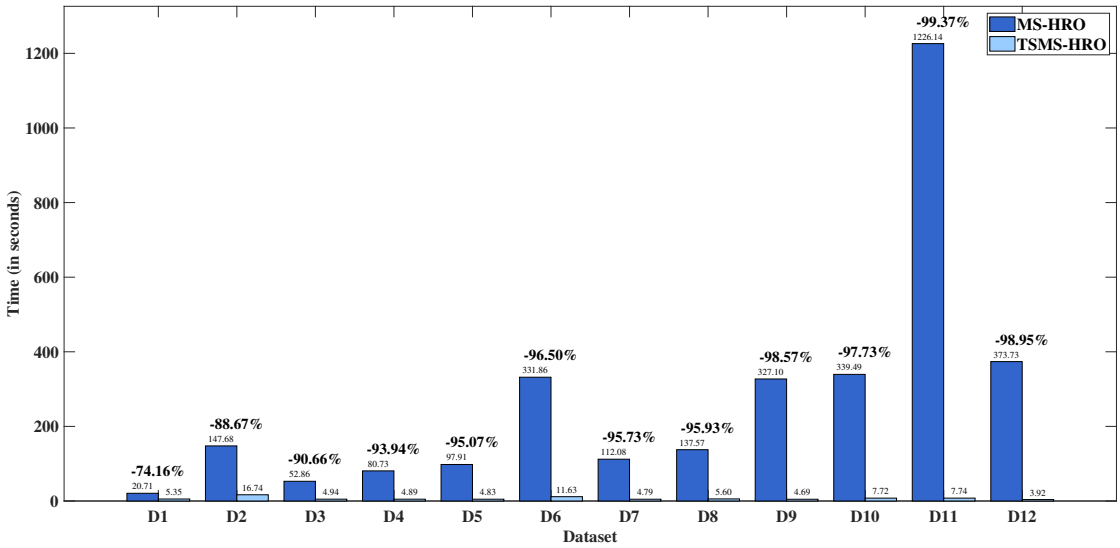

Figure 19: Comparison of running time between the algorithm with and without the filter (MS-HRO vs. TSMS-HRO). TSMS-HRO greatly improves the running efficiency of the algorithm.

Table 9: Comparison of Algorithms: Friedman Rankings and 95% Confidence Intervals

Algorithm Name	Average Rank	Final Rank	95% Confidence Intervals
TSMS-HRO	2.21	1	$[94.21, 99.49]$
TS-GA	3.83	2	$[91.33, 99.61]$
TS-RIME	5.33	3	$[90.20, 99.06]$
TS-HBA	5.79	4	$[90.27, 99.08]$
TS-HHO	6.00	5	$[90.34, 98.39]$
HFSIA	6.96	6	$[84.83, 98.04]$
TS-GWO	7.54	7	$[89.71, 98.74]$
MGWOR	7.67	8	$[87.41, 97.79]$
TS-JAYA	7.83	9	$[88.70, 98.50]$
TS-WOA	9.46	10	$[87.87, 97.22]$
BIGWO	9.50	11	$[85.58, 97.33]$
MBAO	10.17	12	$[87.07, 96.05]$
TS-SSA	10.21	13	$[86.69, 97.28]$
MS-HRO	12.50	14	$[83.94, 93.67]$

Note: The Friedman test yields a test statistic of 73.37 with a corresponding p-value of 1.91×10^{-10}, indicating statistically significant differences among the algorithms at the 0.05 level.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

in Table 10. The comparison results show that TSMS-HRO has a significant advantage over all other 13 algorithms (the corrected p-values are all less than 0.05), and all the null hypotheses are rejected. Especially in comparison with poorly performing algorithms such as MS-HRO and MBAO, the p-value is as low as 0.0063, indicating that TSMS-HRO has a significant statistical advantage over these algorithms in multiple tasks.

Table 10: Wilcoxon Test Results with Holm Correction (TSMS-HRO vs. other algorithms)

Comparison	R+	R-	Stat	p-value	Corrected p	Hypothesis
MS-HRO	78.0	0.0	0.0	0.0005	0.0063	*Rejected*
TS-GA	65.0	13.0	13.0	0.0411	0.0411	*Rejected*
TS-GWO	76.5	1.5	1.5	0.0033	0.0325	*Rejected*
TS-HBA	75.0	3.0	3.0	0.0047	0.0325	*Rejected*
TS-HHO	76.5	1.5	1.5	0.0033	0.0325	*Rejected*
TS-JAYA	76.5	1.5	1.5	0.0033	0.0325	*Rejected*
TS-RIME	75.0	3.0	3.0	0.0047	0.0325	*Rejected*
TS-SSA	76.5	1.5	1.5	0.0033	0.0325	*Rejected*
TS-WOA	76.5	1.5	1.5	0.0033	0.0325	*Rejected*
MBAO	78.0	0.0	0.0	0.0005	0.0063	*Rejected*
HFSIA	75.0	3.0	3.0	0.0047	0.0325	*Rejected*
MGWOR	77.0	1.0	1.0	0.0010	0.0107	*Rejected*
BIGWO	75.5	2.5	2.5	0.0042	0.0325	*Rejected*

In summary, the statistical test results fully verify that TSMS-HRO has the best comprehensive performance under the selected test sets and tasks, and its improved strategy has shown significant advantages in improving search efficiency and solution quality.

4.3.5 The results of ablation experiments

It is clear from the heatmap (Figure 20) that the improved algorithm generally has higher accuracy than the baseline HRO on all datasets in the high-dimensional FS task, and the average number of selected features is generally less than the baseline HRO. Notably, TSMS-HRO presents the darkest color on all datasets, with an accuracy of 100% on D4, D7, D5, and D12, demonstrating its strong generalisation ability and stability. Other variants, such as TS-HRO and its extensions (DE, INIT, SC, TD), also show similarly excellent performance, highlighting the effectiveness of the proposed two-stage feature selection method. For example, after adding the two-stage mechanism, the accuracy on multiple datasets reached 100%.

In contrast, HRO has a relatively low accuracy while selecting a very large number of features. In particular, on some datasets, features with more than 1000 dimensions were selected, but at the same time, their accuracy was not high enough: D8: (num: 2102.3, acc: 65.53%), D10: (num: 1347.7, acc: 69.76%), D11: (num: 1432.4, acc: 66.16%). These results indicate that HRO selects numerous redundant or irrelevant features, leading to reduced classification accuracy and limited adaptability to heterogeneous high-dimensional

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

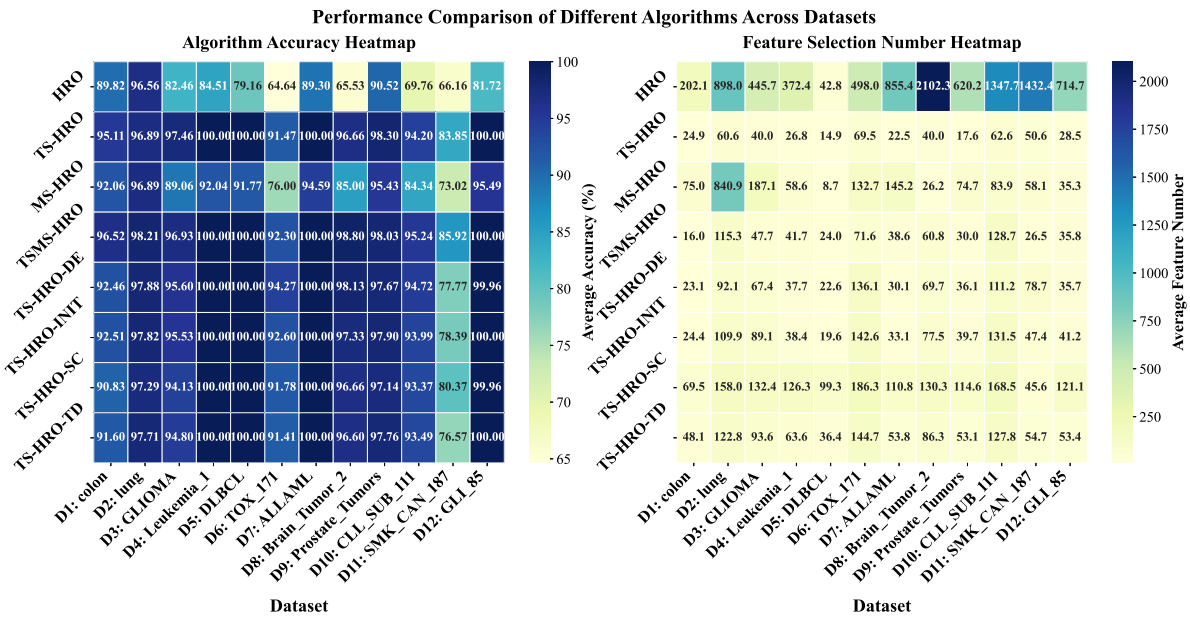

Figure 20: Heat map showing the average accuracy and feature count. Darker shades on the left indicate higher accuracy, while darker shades on the right indicate more selected features.

data. In addition, its performance on high-dimensional feature selection tasks varies greatly, up to 96.56% on the D2 dataset, but only 64.65% and 65.53% on D6 and D8, respectively, indicating that it is highly sensitive to feature redundancy or noise. By adopting a two-stage strategy, TS-HRO significantly alleviates these problems, achieving 91.47% accuracy on D6, an improvement of nearly 27% over HRO, while selecting only 69.5 features on average. Similar trends are observed across multiple datasets, validating the effectiveness of the mRMR feature filter. Multi-strategy enhancement (DE, INIT, SC, TD) further improves performance, albeit to varying degrees. For example, TS-HRO-DE performs particularly well on D6 (94.27%), outperforming other sub-strategies. TS-HRO-SC performs slightly better than INIT and TD on D3 and D10. While TS-HRO-TD performs slightly worse than INIT on D11 (76.57% vs. 78.39%), the difference is not significant.

Despite TSMS-HRO achieving or approaching 100% accuracy on most datasets, it performs slightly lower on D6 (92.30%) and D11 (85.93%). This may suggest that complex nonlinear correlations among features have not been fully exploited, and that there is still room for improvement in the joint optimisation of strategies under high-dimensional sparse conditions.

As shown in Table 11, HRO achieves an average accuracy of 80.02% with 794.30 selected features, serving as the baseline for comparison. By introducing the mRMR feature filter and adopting a two-stage selection mechanism, TS-HRO significantly improves performance, reaching 96.17% accuracy while reducing the average number of selected features to only 38.20. This highlights the strong advantage of the two-stage framework in simultaneously enhancing predictive accuracy and promoting feature dimension reduc-

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

tion. When only multi-strategy optimisation (DE, INIT, SC, TD) is applied, MS-HRO improves accuracy to 88.81% with an average of 143.85 features. Although its accuracy is lower than that of TS-HRO, the considerable reduction in feature count (81.89% fewer than HRO) still confirms the effectiveness of the individual strategies in driving feature selection. This trade-off also suggests that excessive feature reduction may risk eliminating informative attributes and thereby lower accuracy, while insufficient reduction may retain redundant features and compromise classification performance.

Importantly, the proposed TSMS-HRO, which integrates both the two-stage mechanism and multi-strategy optimisation, achieves the best overall performance with 96.83% accuracy and 53.05 selected features. This result demonstrates that the two components are highly complementary: the two-stage mechanism ensures effective feature filtering, while the multiple strategies enhance search robustness, together leading to superior generalisation and stability.

Further examination of the four TS-HRO variants, each augmented with a single strategy, reveals that all strategies consistently contribute to performance gains, though with different emphases. DE and INIT tend to yield relatively higher accuracy, SC retains more features and thus favours stability, while TD provides a balanced trade-off between accuracy and feature reduction. Although none of these variants surpasses the integrated TSMS-HRO, their complementary strengths explain why the combined framework achieves the most robust and well-rounded results

Table 11: Comparison of average accuracy and feature selection results on 12 datasets.

Algorithm	Avg. Acc.	Lift vs. HRO	Avg. Features	Red. Ratio vs. HRO
HRO	80.02%	–	794.30	–
TS-HRO	96.17%	+16.15%	**38.20**	**−95.19**%
MS-HRO	88.81%	+8.83%	143.85	−81.89%
TSMS-HRO	**96.83**%	**+16.81**%	53.05	−93.32%
TS-HRO-DE	95.71%	+15.69%	61.70	−92.23%
TS-HRO-INIT	95.51%	+15.49%	66.21	−91.66%
TS-HRO-SC	95.13%	+15.11%	121.90	−84.65%
TS-HRO-TD	95.00%	+14.98%	78.19	−90.16%

5. Conclusion and Future Work

In this paper, we propose a two-stage high-dimensional FS algorithm based on a modified HRO algorithm to enhance the classification performance of feature subsets while reducing the computational time required to search for the optimal subset. The study found that the filtering method in the first stage eliminated some redundant features, significantly reducing the search space and greatly shortening the algorithm's runtime. Furthermore, the introduction of four mechanisms has significantly improved the original HRO algorithm. The good point set and elite opposition-based learning strategy effectively improve the

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

quality and diversity of the initial population. The adaptive differential operator strategy enhances the utilisation rate of maintainer line individuals. The t-distribution mutation strategy balances global and local search capabilities. The improved adaptive crossover strategy increases the flexibility and diversity of the selfing process in restorer line individuals. The proposed algorithm was compared with recent two-stage and improved metaheuristic-based FS methods, such as MBAO, MGWOR, HFSIA, BIGWO, and improved two-stage methods, such as TS-GA, TS-GWO, TS-HBA, TS-HHO, TS-JAYA, TS-RIME, TS-SSA, and TS-WOA. The results show that TSMS-HRO achieves better initial solutions in the early iterations and converges to more promising values in the later iterations in the field of high-dimensional FS.

Despite its advantages, TSMS-HRO still presents several limitations. First, while it performs well on benchmark functions and 12 high-dimensional biomedical datasets, its applicability to non-biomedical domains such as text or image data may face new challenges, including data noise and class imbalance. Second, although the mRMR method is effective, its assumption of linear relationships based on mutual information limits its ability to capture high-order nonlinear interactions—such as gene co-regulation—in biological data, potentially leading to the omission of critical feature combinations. Lastly, the integration of multiple strategies inevitably increases algorithmic complexity compared to simpler methods.

In future work, more effective optimisation strategies can be explored to ensure the algorithm's generalisation ability across problems of varying scale and dimension. Additionally, investigating different transfer functions, filters, and advanced classifiers would be highly beneficial, as these components have the potential to further enhance the algorithm's performance in various optimisation scenarios. In particular, integrating deep learning–based classifiers with the feature subsets generated by TSMS-HRO could improve its adaptability to complex biomedical data. Moreover, further optimising the operators of HRO—such as developing adaptive hybridisation or crossover mechanisms—may significantly boost both convergence efficiency and solution quality.

Conflicts of Interest

We declare that there are no conflicts of interest regarding the publication of this paper.

Author Contributions

Zhiwei Ye: Writing – review & editing, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Jie Sun**: Writing – original draft, Visualisation, Validation, Methodology, Software. **Wen Zhou**: Formal analysis, Data curation, Conceptualization. **Bogdan Adamyk**: Writing – review & editing. **Jixin Zhang**: Investigation. **Ting Cai**: Formal analysis, Data curation, Conceptualization. **Jun Shen**:

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Data curation. **Mengya Lei**: Supervision. **Jing Zhou**: Data curation. **Ruihan Li**: Conceptualization, Methodology, Software.

Data Availability

The datasets used in this paper can be accessed via the following link: UCI Machine Learning Repository.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 62376089, 62302153, 62302154, U23A20318), the Key Research and Development Program of Hubei Province, China (Grant No.2023BEB024), the Young and Middle-aged Scientific and Technological Innovation Team Plan in Higher Education Institutions in Hubei Province, China (Grant No. T2023007).

References

Agrawal, U., Rohatgi, V., & Katarya, R. (2022). Normalized mutual information-based equilibrium optimizer with chaotic maps for wrapper-filter feature selection. *Expert Systems with Applications*, 207, 118107. https://doi.org/10.1016/j.eswa.2022.118107

Ahmed, S., Ghosh, K. K., Mirjalili, S., & Sarkar, R. (2021). Aieou: Automata-based improved equilibrium optimizer with u-shaped transfer function for feature selection. *Knowledge-Based Systems*, 228, 107283. https://doi.org/10.1016/j.knosys.2021.107283

Anuragi, A., Sisodia, D. S., & Pachori, R. B. (2024). Mitigating the curse of dimensionality using feature projection techniques on electroencephalography datasets: An empirical review. *Artificial Intelligence Review*, 57(3), 75. https://doi.org/10.1007/s10462-024-10711-8

Askr, H., Abdel-Salam, M., & Hassanien, A. E. (2024). Copula entropy-based golden jackal optimization algorithm for high-dimensional feature selection problems. *Expert Systems with Applications*, 238, 121582. https://doi.org/10.1016/j.eswa.2023.121582

Badshah, A., Daud, A., Alharbey, R., Banjar, A., Bukhari, A., & Alshemaimri, B. (2024). Big data applications: Overview, challenges and future. *Artificial Intelligence Review*, 57(11), 290. https://doi.org/10.1007/s10462-024-10938-5

Barbieri, M. C., Grisci, B. I., & Dorn, M. (2024). Analysis and comparison of feature selection methods towards performance and stability. *Expert Systems with Applications*, 249, 123667. https://doi.org/10.1016/j.eswa.2024.123667

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

BinSaeedan, W., & Alramlawi, S. (2021). Cs-bpso: Hybrid feature selection based on chi-square and binary pso algorithm for arabic email authorship analysis. *Knowledge-Based Systems*, 227, 107224. https://doi.org/10.1016/j.knosys.2021.107224

Buchaiah, S., & Shakya, P. (2022). Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. *Measurement*, 188, 110506. https://doi.org/10.1016/j.measurement.2021.110506

Chamlal, H., Benzmane, A., & Ouaderhman, T. (2024). Maximal cliques-based hybrid high-dimensional feature selection with interaction screening for regression. *Neurocomputing*, 607, 128361. https://doi.org/10.1016/j.neucom.2024.128361

Fang, Y., Yao, Y., Lin, X., Wang, J., & Zhai, H. (2024). A feature selection based on genetic algorithm for intrusion detection of industrial control systems. *Computers & Security*, 139, 103675. https://doi.org/10.1016/j.cose.2023.103675

Gao, W., Hao, P., Wu, Y., & Zhang, P. (2023). A unified low-order information-theoretic feature selection framework for multi-label learning. *Pattern Recognition*, 134, 109111. https://doi.org/10.1016/j.patcog.2022.109111

Ghosh, K. K., Begum, S., Sardar, A., Adhikary, S., Ghosh, M., Kumar, M., & Sarkar, R. (2021). Theoretical and empirical analysis of filter ranking methods: Experimental study on benchmark dna microarray data. *Expert Systems with Applications*, 169, 114485. https://doi.org/10.1016/j.eswa.2020.114485

Got, A., Moussaoui, A., & Zouache, D. (2021). Hybrid filter-wrapper feature selection using whale optimization algorithm: A multi-objective approach. *Expert Systems with Applications*, 183, 115312. https://doi.org/10.1016/j.eswa.2021.115312

He, J., Qu, L., Wang, P., & Li, Z. (2024). An oscillatory particle swarm optimization feature selection algorithm for hybrid data based on mutual information entropy. *Applied Soft Computing*, 152, 111261. https://doi.org/10.1016/j.asoc.2024.111261

Hussain, K., Neggaz, N., Zhu, W., & Houssein, E. H. (2021). An efficient hybrid sine-cosine harris hawks optimization for low and high-dimensional feature selection. *Expert Systems with Applications*, 176, 114778. https://doi.org/10.1016/j.eswa.2021.114778

Jiang, H., Yang, Y., Wan, Q., & Dong, Y. (2024). Feature selection based on dynamic crow search algorithm for high-dimensional data classification. *Expert Systems with Applications*, 250, 123871. https://doi.org/10.1016/j.eswa.2024.123871

Kapoor, P., Kaushal, S., Kumar, H., & Kanwar, K. (2024). A survey on feature extraction and learning techniques for link prediction in homogeneous and heterogeneous complex networks. *Artificial Intelligence Review*, 57(12), 348. https://doi.org/10.1007/s10462-024-10998-7

Lameesa, A., Hoque, M., Alam, M. S. B., Ahmed, S. F., & Gandomi, A. H. (2024). Role of metaheuristic algorithms in healthcare: A comprehensive investigation across clinical diagnosis, medical imaging, operations management, and public health.

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Journal of Computational Design and Engineering, 11(3), 223–247. https://doi.org/10.1093/jcde/qwae046

Li, G., Yu, Z., Yang, K., Lin, M., & Chen, C. P. (2024). Exploring feature selection with limited labels: A comprehensive survey of semi-supervised and unsupervised approaches. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2024.3397878

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6), 1–45. https://doi.org/10.1145/3136625

Liang, J., Hu, Z., Bi, Y., Cheng, H., Yu, K., Yue, C.-T., Wang, X., & Guo, W.-F. (2024). A survey on evolutionary computation for identifying biomarkers of complex disease. *IEEE Transactions on Evolutionary Computation*. https://doi.org/10.1109/TEVC.2024.3414442

Manikandan, G., & Abirami, S. (2021). An efficient feature selection framework based on information theory for high dimensional data. *Applied Soft Computing*, 111, 107729. https://doi.org/10.1016/j.asoc.2021.107729

Mei, M., Zhang, S., Ye, Z., Wang, M., Zhou, W., Yang, J., Zhang, J., Yan, L., & Shen, J. (2025). A cooperative hybrid breeding swarm intelligence algorithm for feature selection. *Pattern Recognition*, 111901. https://doi.org/10.1016/j.patcog.2025.111901

Miao, F., Wu, Y., Yan, G., & Si, X. (2024). A memory interaction quadratic interpolation whale optimization algorithm based on reverse information correction for high-dimensional feature selection. *Applied Soft Computing*, 164, 111979. https://doi.org/10.1016/j.asoc.2024.111979

Moslemi, A. (2023). A tutorial-based survey on feature selection: Recent advancements on feature selection. *Engineering Applications of Artificial Intelligence*, 126, 107136. https://doi.org/10.1016/j.engappai.2023.107136

Moustafa, M. S., Mahmoud, A. S., Farg, E., Nabil, M., & Arafat, S. M. (2024). Bi-stage feature selection for crop mapping using grey wolf metaheuristic optimization. *Advances in Space Research*, 73(10), 5005–5016. https://doi.org/10.1016/j.asr.2024.02.037

Nematzadeh, H., García-Nieto, J., Aldana-Montes, J. F., & Navas-Delgado, I. (2024). Pattern recognition frequency-based feature selection with multi-objective discrete evolution strategy for high-dimensional medical datasets. *Expert Systems with Applications*, 249, 123521. https://doi.org/10.1016/j.eswa.2024.123521

Nssibi, M., Manita, G., Chhabra, A., Mirjalili, S., & Korbaa, O. (2024). Gene selection for high dimensional biological datasets using hybrid island binary artificial bee colony with chaos game optimization. *Artificial Intelligence Review*, 57(3), 51. https://doi.org/10.1007/s10462-023-10675-1

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Nssibi, M., Manita, G., & Korbaa, O. (2023). Advances in nature-inspired metaheuristic optimization for feature selection problem: A comprehensive survey. *Computer Science Review*, 49, 100559. https://doi.org/10.1016/j.cosrev.2023.100559

Osama, S., Shaban, H., & Ali, A. A. (2023). Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review. *Expert Systems with Applications*, 213, 118946. https://doi.org/10.1016/j.eswa.2022.118946

Ouadfel, S., & Abd Elaziz, M. (2022). Efficient high-dimension feature selection based on enhanced equilibrium optimizer. *Expert Systems with Applications*, 187, 115882. https://doi.org/10.1016/j.eswa.2021.115882

Pan, H., Chen, S., & Xiong, H. (2023). A high-dimensional feature selection method based on modified gray wolf optimization. *Applied Soft Computing*, 135, 110031. https://doi.org/10.1016/j.asoc.2023.110031

Pashaei, E. (2022). Mutation-based binary aquila optimizer for gene selection in cancer classification. *Computational Biology and Chemistry*, 101, 107767. https://doi.org/10.1016/j.compbiolchem.2022.107767

Peng, L., Cai, Z., Heidari, A. A., Zhang, L., & Chen, H. (2023). Hierarchical harris hawks optimizer for feature selection. *Journal of Advanced Research*, 53, 261–278. https://doi.org/10.1016/j.jare.2023.01.014

Qaraad, M., Amjad, S., Hussein, N. K., & Elhosseini, M. A. (2022). Addressing constrained engineering problems and feature selection with a time-based leadership salp-based algorithm with competitive learning. *Journal of Computational Design and Engineering*, 9(6), 2235–2270. https://doi.org/10.1093/jcde/qwac095

Rajammal, R. R., Mirjalili, S., Ekambaram, G., & Palanisamy, N. (2022). Binary grey wolf optimizer with mutation and adaptive k-nearest neighbour for feature selection in parkinson's disease diagnosis. *Knowledge-Based Systems*, 246, 108701. https://doi.org/10.1016/j.knosys.2022.108701

Song, X., Ma, H., Zhang, Y., Gong, D., Guo, Y., & Hu, Y. (2024). A streaming feature selection method based on dynamic feature clustering and particle swarm optimization. *IEEE Transactions on Evolutionary Computation*. https://doi.org/10.1109/TEVC.2024.3451688

Song, X.-F., Zhang, Y., Gong, D.-W., & Gao, X.-Z. (2021). A fast hybrid feature selection based on correlation-guided clustering and particle swarm optimization for high-dimensional data. *IEEE Transactions on Cybernetics*, 52(9), 9573–9586. https://doi.org/10.1109/TCYB.2021.3061152

Song, X., Zhang, Y., Gong, D., Liu, H., & Zhang, W. (2022). Surrogate sample-assisted particle swarm optimization for feature selection on high-dimensional data. *IEEE Transactions on Evolutionary Computation*, 27(3), 595–609. https://doi.org/10.1109/TEVC.2022.3175226

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

Song, X., Zhang, Y., Zhang, W., He, C., Hu, Y., Wang, J., & Gong, D. (2024). Evolutionary computation for feature selection in classification: A comprehensive survey of solutions, applications and challenges. *Swarm and Evolutionary Computation*, 90, 101661. https://doi.org/10.1016/j.swevo.2024.101661

Telikani, A., Tahmassebi, A., Banzhaf, W., & Gandomi, A. H. (2021). Evolutionary machine learning: A survey. *ACM Comput. Surv.*, 54(8). https://doi.org/10.1145/3467477

Thirumoorthy, K., et al. (2023). A two-stage feature selection approach using hybrid quasi-opposition self-adaptive coati optimization algorithm for breast cancer classification. *Applied Soft Computing*, 146, 110704. https://doi.org/10.1016/j.asoc.2023.110704

Wang, C., Zhan, C., Lu, B., Yang, W., Zhang, Y., Wang, G., & Zhao, Z. (2024). Ssfan: A compact and efficient spectral-spatial feature extraction and attention-based neural network for hyperspectral image classification. *Remote Sensing*, 16(22), 4202. https://doi.org/10.3390/rs16224202

Wang, Y., Ran, S., & Wang, G.-G. (2024). Role-oriented binary grey wolf optimizer using foraging-following and lévy flight for feature selection. *Applied Mathematical Modelling*, 126, 310–326. https://doi.org/10.1016/j.apm.2023.08.043

Wang, Y.-C., Song, H.-M., Wang, J.-S., Song, Y.-W., Qi, Y.-L., & Ma, X.-R. (2024). Gogmbsho: Multi-strategy fusion binary sea-horse optimizer with gaussian transfer function for feature selection of cancer gene expression data. *Artificial Intelligence Review*, 57(12), 347. https://doi.org/10.1007/s10462-024-10954-5

Xue, Y., & Zhang, C. (2024). A novel importance-guided particle swarm optimization based on mlp for solving large-scale feature selection problems. *Swarm and Evolutionary Computation*, 91, 101760. https://doi.org/10.1016/j.swevo.2024.101760

Ye, A. Z., Li, B. R., Zhou, C. W., Wang, D. M., Mei, E. M., Shu, F. Z., & Shen, G. J. (2023). High-dimensional feature selection based on improved binary ant colony optimization combined with hybrid rice optimization algorithm. *International Journal of Intelligent Systems*, 2023(1), 1444938. https://doi.org/10.1155/2023/1444938

Ye, Z., Luo, J., Zhou, W., Wang, M., & He, Q. (2024). An ensemble framework with improved hybrid breeding optimization-based feature selection for intrusion detection. *Future Generation Computer Systems*, 151, 124–136. https://doi.org/10.1016/j.future.2023.09.035

Ye, Z., Ma, L., & Chen, H. (2016). A hybrid rice optimization algorithm. *2016 11th International Conference on Computer Science & Education (ICCSE)*, 169–174. https://doi.org/10.1109/ICCSE.2016.7581575

Yu, W., Kang, H., Sun, G., Liang, S., & Li, J. (2022). Bio-inspired feature selection in brain disease detection via an improved sparrow search algorithm. *IEEE Transactions*

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025

on Instrumentation and Measurement, 72, 1–15. https://doi.org/10.1109/TIM.2022.3228003

Zhang, B., Li, Y., & Chai, Z. (2022). A novel random multi-subspace based relieff for feature selection. *Knowledge-Based Systems*, 252, 109400. https://doi.org/10.1016/j.knosys.2022.109400

Zhang, J., Xu, D., Hao, K., Zhang, Y., Chen, W., Liu, J., Gao, R., Wu, C., & De Marinis, Y. (2021). Fs–gbdt: Identification multicancer-risk module via a feature selection algorithm by integrating fisher score and gbdt. *Briefings in bioinformatics*, 22(3), bbaa189. https://doi.org/10.1093/bib/bbaa189

Zhao, Z., Yu, H., Guo, H., & Chen, H. (2024). Multi-strategy augmented harris hawks optimization for feature selection. *Journal of Computational Design and Engineering*, 11(3), 111–136. https://doi.org/10.1093/jcde/qwae030

Zheng, J., Yang, Z., & Ge, Z. (2024). Deep residual principal component analysis as feature engineering for industrial data analytics. *IEEE Transactions on Instrumentation and Measurement*, 73, 1–10. https://doi.org/10.1109/TIM.2024.3420267

Zhou, J., Zhang, Q., Zeng, S., Zhang, B., & Fang, L. (2024). Latent linear discriminant analysis for feature extraction via isometric structural learning. *Pattern Recognition*, 149, 110218. https://doi.org/10.1016/j.patcog.2023.110218

Zhou, R., Zhang, Y., & He, K. (2023). A novel hybrid binary whale optimization algorithm with chameleon hunting mechanism for wrapper feature selection in qsar classification model: A drug-induced liver injury case study. *Expert Systems with Applications*, 234, 121015. https://doi.org/10.1016/j.eswa.2023.121015

Zhu, Y., Li, W., & Li, T. (2023). A hybrid artificial immune optimization for high-dimensional feature selection. *Knowledge-Based Systems*, 260, 110111. https://doi.org/10.1016/j.knosys.2022.110111

Zuo, X., Zhang, W., Wang, X., Dang, L., Qiao, B., & Wang, Y. (2025). Unsupervised feature selection via maximum relevance and minimum global redundancy. *Pattern Recognition*, 164, 111483. https://doi.org/10.1016/j.patcog.2025.111483

ORIGINAL UNEDITED MANUSCRIPT

Downloaded from https://academic.oup.com/jcde/advance-article/doi/10.1093/jcde/qwaf113/8297130 by guest on 05 November 2025