electronics

Article

Enhanced Wild Horse Optimizer with Cauchy Mutation and Dynamic Random Search for Hyperspectral Image Band Selection

Tao Chen [1], Yue Sun [1], Huayue Chen [1,2,*] and Wu Deng [3,4,*]

1 School of Computer, China West Normal University, Nanchong 637002, China; sunnyue@stu.cwnu.edu.cn
2 Key Laboratory of Optimization Theory and Applications, China West Normal University of Sichuan Province, Nanchong 637002, China
3 School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China
4 Traction Power State Key Laboratory, Southwest Jiaotong University, Chengdu 610031, China
* Correspondence: sunnyxiaoyue20@cwnu.edu.cn (H.C.); wdeng@cauc.edu.cn (W.D.)

Abstract: The high dimensionality of hyperspectral images (HSIs) brings significant redundancy to data processing. Band selection (BS) is one of the most commonly used dimensionality reduction (DR) techniques, which eliminates redundant information between bands while retaining a subset of bands with a high information content and low noise. The wild horse optimizer (WHO) is a novel metaheuristic algorithm widely used for its efficient search performance, yet it tends to become trapped in local optima during later iterations. To address these issues, an enhanced wild horse optimizer (IBSWHO) is proposed for HSI band selection in this paper. IBSWHO utilizes Sobol sequences to initialize the population, thereby increasing population diversity. It incorporates Cauchy mutation to perturb the population with a certain probability, enhancing the global search capability and avoiding local optima. Additionally, dynamic random search techniques are introduced to improve the algorithm search efficiency and expand the search space. The convergence of IBSWHO is verified on commonly used nonlinear test functions and compared with state-of-the-art optimization algorithms. Finally, experiments on three classic HSI datasets are conducted for HSI classification. The experimental results demonstrate that the band subset selected by IBSWHO achieves the best classification accuracy compared to conventional and state-of-the-art band selection methods, confirming the superiority of the proposed BS method.

Keywords: band selection (BS); global optimization; wild horse optimizer (WHO); hyperspectral image (HSI) classification

check for updates

Citation: Chen, T.; Sun, Y.; Chen, H.; Deng, W. Enhanced Wild Horse Optimizer with Cauchy Mutation and Dynamic Random Search for Hyperspectral Image Band Selection. *Electronics* **2024**, *13*, 1930. https://doi.org/10.3390/electronics13101930

Academic Editor: Manohar Das

Received: 12 April 2024
Revised: 10 May 2024
Accepted: 12 May 2024
Published: 15 May 2024

1. Introduction

Hyperspectral remote-sensing images (HSIs) have rich spectral and spatial information and the unity of map and spectrum. Their applications have yielded important value in the fields of mineral exploration [1], environmental monitoring [2], precision agriculture and forestry [3] and national defense and the military [4]. However, the large volume and high dimension of HSI data can easily lead to a dimensional disaster [5]. On the other hand, due to the numerous HSI bands, the strong correlation between adjacent bands and a lot of redundant information, information recognition and feature extraction are difficult and the accuracy is not high. Therefore, when using the original HSI, the band with the highest separability should be selected.

In recent years, scholars have conducted a large amount of research on hyperspectral BS methods. Jiang et al. [6] proposed a BS method based on the minimum redundancy maximum relevance (MRMR), which computed the correlation between each band and the labels and then calculated the redundancy between each band and the other bands. The band subsets were selected by maximizing the relevance and minimizing the redundancy.

Xu et al. [7] used structural similarity to measure the relationship between bands and ranked bands according to their similarity and significant differences to select representative band subsets. In order to speed up the efficiency of band selection, a strategy called band grouping was introduced in BS. Wang et al. [8] grouped bands followed by the clone selection algorithm to select representative bands for each group as a subset. Wang et al. [9] proposed a fast neighborhood grouping BS method (FNGBS), which divided the HSI into several groups using a coarse-to-fine strategy and simultaneously selected the most relevant and information-rich bands for each group based on local density and information entropy factors. Although ranking-based BS methods can quickly obtain feature subsets, they often overlook the intrinsic structure of HSI data, resulting in high correlations among the selected bands. In contrast, filtering-based methods are designed based on certain mathematical models or principles, allowing them to utilize the structure and characteristics of HSIs to select bands relevant to the task, which can effectively remove redundant bands. Kira et al. [10] proposed a statistical approach Relief, which had advantages in terms of time and accuracy. Fu et al. [11] proposed a novel adjacent band grouping and normalized matching filter for BS (NGNMF), which preserved spatial–spectral information while reducing data dimensionality. However, limitations in the type of filters and parameter settings may restrict its applicability and performance. Cluster-based methods offer flexibility in selecting bands based on the actual characteristics of the data, thereby reducing the correlation among band subsets. Wang et al. [12] proposed a hyperspectral band selection method based on an adaptive subspace partitioning strategy (ASPS), which divided HSI into multiple sub-cubes with the maximum ratio of inter-class distance to intra-class distance and selected the least noisy bands within each sub-cube. Zhang [13] selected representative bands based on the similarity metric, calculated applicable similarity metric weights using the coefficient of variation and then put the similarity metric into K-means as the kernel with weights. Wang et al. [14] raised an area-aware hierarchical latent feature representation learning-guided clustering (HLFC) method, which reflected band similarity by constructing a similarity map, learning latent features hierarchically and then clustering with k-means. However, cluster-based methods do not consider the overall performance after merging representative bands, and they are not guided by the classifier when searching for band subsets, leading to relatively low accuracy. In recent years, deep learning has also been applied to hyperspectral band selection and it has achieved good results [15–24], but it requires a large number of training samples, while labeled samples for HSI are limited.

To overcome the aforementioned issue, the wrapper method based on a heuristic search, with its strong ability to search feature space and full consideration of classification effect, has been extensively used in HSI BS. Su et al. [25] used particle swarm optimization (PSO) to select hyperspectral bands while automatically identifying the optimum number of selected bands. It used two particle swarms: an external swarm to determine the appropriate band numbers and an internal swarm for band selection. It was proven that automatically selecting a variable number of bands had a better classification result than a fixed number of bands, but using two PSOs to search greatly increased the complexity. Medjahed et al. [26] used the grey wolf optimization algorithm (GWO) for solving the BS problem and selected the subset of bands well by optimizing the objective function. Medjahed et al. [27] proposed a BS method based on the sine cosine algorithm (SCA), which used SCA in combination with KNN to select band subsets based on classification accuracy. The results showed that IMACA produced accurate classification results, but the classification accuracy of selected bands was not high. Most existing optimization algorithms have many parameters, tend to fall into local optimality and cannot solve complex optimization problems. As a result, newly proposed algorithms are increasing, such as greylag goose optimization (GGO) [28], Coati optimization algorithm (COA) [29], parrot optimizer (PO) [30], reptile search algorithm (RSA) [31], MOQEA/D [32], machine learning (ML) [33], deep neural network (DNN) [34], image segmentation [35] and so on [36]. In addition, there are also winners of CEC competitions such as LSHADE [37], COLSHADE [38], IMODE [39], KGE [40] and SASS [41].

The wild horse optimizer (WHO) [42] is a new meta-heuristic algorithm that has been employed successfully for the optimization of practical problems due to its limited parameters, great optimum capability and relatively low time complexity. For example, it has been employed in extracting model parameters in photovoltaic systems [43], solving nonlinear multi-objective optimization problems in energy management [44] and solving link failure problems in underwater channels [45]. Although WHO can achieve satisfactory results on some practical issues, there are still some problems, such as a limited exploitation capability and stagnation of locally optimal solutions. Therefore, it is necessary to improve WHO according to practical problems. Ewees et al. [46] proposed an improved version of WHO (WHOW) by using the spiral position update strategy of the whale optimization algorithm (WOA) for updating positions in WHO, and the experimental consequences indicated the advantages of WHOW in solving different optimization problems and its outstanding feature selection ability for most benchmark datasets. Zheng et al. [47] proposed an improved WHO (IWHO), which utilized the random running strategy and a competition mechanism with waterhole to enhance development capabilities and then used a dynamic inertia weight strategy to optimize the global optimal solution. Simulation tests and application experiments demonstrated the best optimization ability of the improved algorithm. However, according to the NFL theory [48], no single optimization algorithm can address all optimization difficulties, and algorithms need to be improved to fit actual problems. BS is essentially an NP-hard problem [49] and as the number of bands grows, the above algorithms' optimization processes may prematurely converge or even stagnate.

To solve these problems, this paper proposes a hyperspectral BS algorithm based on an enhanced WHO (IBSWHO), which can overcome the shortcomings of the original WHO and automatically select informative bands while maintaining excellent classification precision. The main contributions of this paper are described as follows.

1. The Sobol sequence, Cauchy mutation and dynamic random search technique are helpful for IBSWHO to step out of local optimality and increase search efficiency and exploitation ability.
2. A binary-coded version of WHO is proposed and applied to image feature selection.
3. HSI BS is constructed as a binary optimization problem, and IBSWHO is applied to HSI BS to select the optimal number of bands automatically, and its performance is verified on typical HSI datasets.

The rest of this paper is organized as follows: Section 2 introduces the proposed BS algorithm (IBSWHO) and its specific steps. Experimental results and a comparative study are presented in Section 3. Finally, the full paper and future work are summarized in Section 4.

2. Proposed Method

To deal with the poor quality of band subsets, the need to preset the expected number of bands and the stagnation of local optimal solutions of WHO, a binary enhanced WHO (IBSWHO) for HSI BS is proposed in this paper.

2.1. Binary Encoding

The goal of BS is to determine whether a band is selected in the band subset, so the value assigned to the band should reflect whether it is selected or not. Here are two possible values for each band: one indicates that the band is selected, and the other indicates that the band is not selected. Thus, binary encoding is the best choice [50]. A band is reserved or removed, and its weight is set to 1 or 0. The details are as follows:

Let the size of the hyperspectral image D be $m \times n \times h$, where $m \times n$ is the number of pixels, h is the total number of bands, $b = \{b_1, b_2, \ldots, b_h\}$ is the band set of D, $i = \{1, 2, \ldots, h\}$ and Xp represents a candidate band found by IBSWHO. To determine whether a band is selected, a static threshold of 0.5 is used, as per the expression is given in (1).

$$b_i = \left\{ \begin{array}{l} 1,\ if\ Xp > 0.5 \\ 0,\ otherwise \end{array} \right\}, \tag{1}$$

2.2. Band Selection Method Based on Enhanced Wild Horse Optimizer (IBSWHO)

The overall flow of IBSWHO for HSI BS is shown in Figure 1. According to the group behavior of the wild horses, IBSWHO can be divided into the following seven steps.

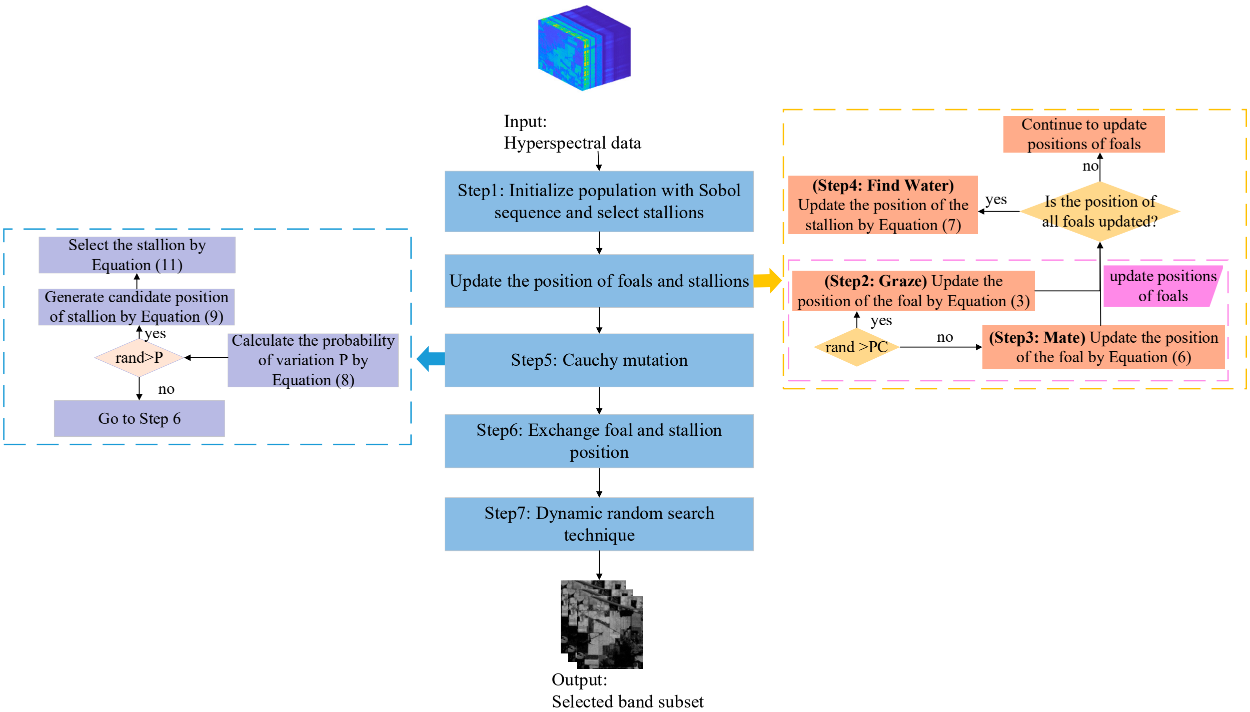


Figure 1. The flow of the IBSWHO.

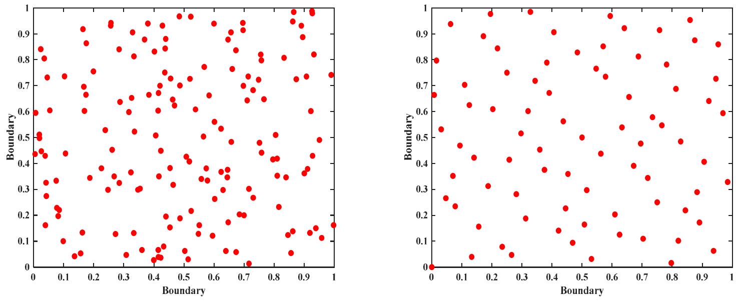
2.2.1. The Sobol Sequence Initializes the Population

WHO uses the rand function to randomly initialize the population, resulting in a high randomness but uneven distribution across the entire solution space, which results in a sluggish population search speed, poor diversity and low solution accuracy. To address these issues, this paper introduces the Sobol sequence to map the initialized populations. The Sobol sequence, as opposed to pseudo-random sequences, is a low-discrepancy sequence that replaces a pseudo-random sequence with a deterministic low-discrepancy sequence, which allows for shorter calculation periods, faster sampling speeds and higher efficiency in handling high-dimensional sequences [51].

Let ub and lb represent the upper and lower boundaries of the optimal solution. $K_n \in [0,1]$ is a random number generated by the Sobol sequence, and the mathematical expression for the initial population x_n is as follows:

$$x_n = lb + K_n \times (ub - lb), \tag{2}$$

Assuming that the upper and lower bounds of the population are 0 and 1, respectively, the search space dimension is 2, and the population size N is 100. When the initial population distribution of the Sobol sequence is compared to that of the random function, as shown in Figure 2, it can be seen that the distribution initialized by the Sobol sequence is more uniform.



(**a**) (**b**)

Figure 2. Distribution of initialization generated by different methods. (**a**) Random initialization, (**b**) Sobol sequence initialization.

2.2.2. Grazing Behavior of Foals

The foals graze around the stallions; that is, they search around the locally optimal band subset. The behavior is represented by a mathematical model as follows:

$$\overline{BX}_{i,G}^{j} = 2Z\cos(2\pi RZ) \times \left(BS^{j} - BX_{i,G}^{j}\right) + BS^{j}, \quad (3)$$

where $\overline{BX}_{i,G}^{j}$ is the band subset updated by the foal, BS^{j} is the band subset represented by the stallion, i.e., the optimal band subset in this group and $BX_{i,G}^{j}$ is the current band subset. Z is the adaptable mechanism calculated by (4), R is a uniform random number in the range of $[-2, 2]$, π equals 3.14 and the cos function makes foals update the band subset around the center with different angles and radius.

$$W = \vec{R_1} < TDR,\ IDX = (W == 0),\ Z = R_2 \Theta IDX + \vec{R_3} \Theta (\sim IDX), \quad (4)$$

where W is a vector composed of 0 and 1, whose size is equal to the total number of bands. The values 0 and 1 in W represent logical values of false and true. TDR is an adaptive parameter that gradually decreases from 1 to 0 with iteration, and its expression is given in (5). $\vec{R_1}$ and $\vec{R_3}$ are random vectors uniformly distributed in the range of [0,1], and R_2 is a random number uniformly distributed in the range of [0,1]. Θ is the dot multiplication of elements and $\sim$ is the reverse operation. IDX indexes $\vec{R_1}$ and returns the value satisfying the condition $W == 0$.

$$TDR = 1 - \frac{iter}{maxiter}, \quad (5)$$

where *iter* represents the current number of iterations and *maxiter* is the maximum number of iterations.

2.2.3. Mating Behavior of Foals

Foals leave the original group before puberty, and females and males join different groups. Upon reaching puberty, they join a temporary group for mating, then their offspring leave the temporary group to join another group. The mathematical model is expressed as follows.

$$BX_{G,k}^{p} = Crossover\left(BX_{G,i}^{q}, BX_{G,j}^{z}\right),\ i \neq j \neq k,\ q = z = end,\ Crossover = Mean, \quad (6)$$

where $BX_{G,k}^{p}$ is the offspring of foal q and foal z, i.e., the updated band subset. $BX_{G,i}^{q}$ is the location of foal q from group i and $BX_{G,j}^{z}$ is the location of foal z from group j. The average operation is used to represent mating, and we take the average value of the position of the two mating foals as the position of their offspring.

2.2.4. Leading Groups to Find Water by Leaders

Stallions must lead the respective group to the appropriate water source, and each group moves towards this source. If the source is empty, it can be used; otherwise, groups must leave. This process can be described as follows:

$$\overline{BS_{G_i}} = \left\{ \begin{array}{l} 2Z\cos(2\pi RZ) \times \left(BG - BS_{G_i}\right) + BG,\ if\ R_3 > 0.5 \\ 2Z\cos(2\pi RZ) \times \left(BG - BS_{G_i}\right) - BG,\ if\ R_3 \leq 0.5 \end{array} \right\}, \quad (7)$$

where $\overline{BS_{G_i}}$ is the updated locally optimal band subset, Z is the adaptive mechanism calculated by (3), R is a uniform random number in the range of [−2,2], BG is the globally optimal band subset (i.e., the location of water source) and BS_{G_i} is the current subset of the locally optimal band.

2.2.5. Cauchy Mutation

As the number of iterations increases, WHO converges gradually, and once it converges to a local optimum, it will find it difficult to jump out. To solve this problem, this paper uses Cauchy mutation when updating the local optimal band subset so that the range of the updated band subset becomes wider, and it is easier to step out of the local optimum. In the early stage, the variance probability should be small because the algorithm needs to find the global optimal subset quickly. In the later stage, to avoid converging prematurely, the population diversity needs to be increased, so the variance probability should be increased. In this paper, a simple and effective formula for variation probability is designed as follows.

$$P = \frac{\exp\left(\frac{iter}{maxiter}\right)}{3}, \quad (8)$$

where P is the probability of variation, $iter$ is the current number of iterations, $maxiter$ is the maximum iterations and the value of P grows gradually with the increase in iterations. To visualize the change in the P-value under different iterations, (8) is plotted as in Figure 3.

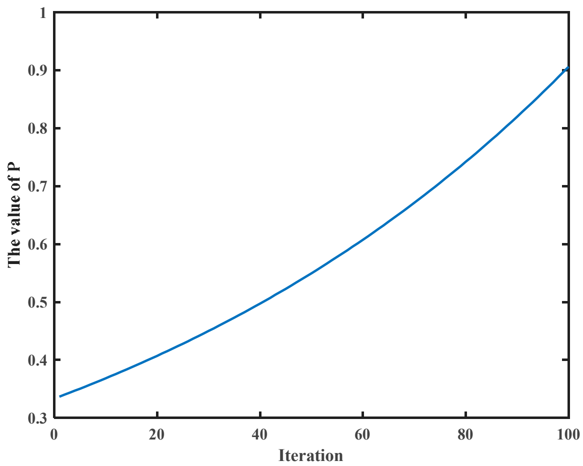


Figure 3. Iterative curve of the P-value.

The mathematical description of Cauchy mutation for the band subset update is described as follows.

$$\overline{BS_{G_i new}} = \overline{BS_{G_i}} + \overline{BS_{G_i}} \times Cauchy(0,1), \quad (9)$$

where $\overline{BS_{G_i new}}$ is a new band subset for which the leader of the group i has been updated by the Cauchy variation, and $\overline{BS_{G_i}}$ is the new band subset for which the leader of the group i has been updated by (7). *Cauchy* is the Cauchy operator, which is calculated.

$$Cauchy = \tan\left(original\_x - \frac{1}{2}\right) \times \pi, \tag{10}$$

where $original\_x$ is the initial random number uniformly distributed in the range of (0,1), and π equals 3.14.

At the same time, the greedy strategy is employed here, and if the new band subset generated by the variation is better than the current locally optimal band subset, the mutant individual is retained; otherwise, it is discarded.

$$\overline{BS_{G_i}}' = \left\{ \begin{array}{l} \overline{BS_{G_i}}, if\ cost\left(\overline{BS_{G_i}}\right) \leq cost\left(\overline{BS_{G_i new}}\right) \\ BS_{G_i new},\ if\ cost\left(\overline{BS_{G_i}}\right) > cost\left(\overline{BS_{G_i new}}\right) \end{array} \right\}, \tag{11}$$

where $\overline{BS_{G_i}}'$ is the final updated subset of locally optimal bands, *cost* represents the classification error rate obtained by sending this band subset to SVM classifier; that is, $cost = 1 - \mathrm{OA}$, a smaller value indicates a better classification performance of the subset. OA is the overall classification accuracy, and its concept is in Appendix A.

2.2.6. Exchange and Selection of Group Leaders

The algorithm randomly selects the leader to maintain randomness at first, and in the late stage, the leader is selected according to the classification error ratio for the band subset. If the error rate of the member is lower than that of the leader, the positions of the member and the leader are exchanged.

$$BS_{G_i} = \left\{ \begin{array}{l} BX_{G,i},\ if\ cost(BX_{G,i}) < cost\left(BS_{G_i}\right) \\ BS_{G_i}, if\ cost(BX_{G,i}) \geq cost\left(BS_{G_i}\right) \end{array} \right\}, \tag{12}$$

2.2.7. Dynamic Random Search Technique

When searching for the global optimum, WHO suffers from poor convergence and blindness in the search process. In this paper, we adopt the local search phase in the dynamic random search technique (DRASET) [52] to save the relevant information of the current optimal band subset and search on this basis, to increase IBSWHO's search efficiency and exploitation capacity.

In the local search phase, the search process centers on the current global best solution *gBest* and investigates whether a better subset exists by smaller search steps. After the search iteration, if *gBest*'s cost becomes smaller, the new *gBest* is adopted, and this process continues until the stopping criterion is reached. The parameter T determines whether the technique is used or not, and the value of T is determined according to the experience [53].

3. Experiments

In this section, two sets of experiments are conducted to test the optimization and convergence performance of IBSWHO on benchmark functions, and its effectiveness on HSI BS. All experiments were conducted on a simulation experimental platform with Intel (R) Core (TM) i7-8750H CPU @ 2.20 GHz, 8 GB RAM, the Windows 11 operating system and MATLAB R2021a.

3.1. Optimization Performance Test

For testing the optimization and convergence performance of IBSWHO, nine commonly used nonlinear benchmark functions are selected and PSO, GWO, WHO and a recently proposed improved WHO (IWHO) are used for comparison. The details of the test functions and the parameters of the optimization algorithms are shown in Tables 1–3, respectively. The results of the comparative experiments are shown in Table 4, and we have highlighted the best results in bold. Among the benchmark functions, F1–F3 are unimodal functions, F4–F6 are multimodal functions and F7–F9 are fixed-dimension multimodal functions. To verify the optimization ability of IBSWHO for complex problems, the CEC2019 test functions, which are more complex and difficult than other benchmark test functions, are introduced and compared with the three recent state-of-the-art top algorithms, i.e., SASS, COLSHADE and KGE, and basic optimization algorithms, i.e., PSO, GWO and WHO. The experimental results are shown in Table 5.

To ensure the fairness and objectivity of the experiment, the population size of all algorithms is set to 30, the maximum iteration is set to 500, each algorithm runs 30 times independently and the average and standard deviation of 30 experiments are counted under the dimensions of 30, 200 and 500, respectively. The accuracy and quality of each algorithm's solutions are evaluated by comparing their average values, while the algorithm's stability is indicated by the standard deviation value.

Table 1. Information about test functions.

Function	Function Name	Dimension	Range	Optimum Value
F1	Sphere	30/200/500	[−100,100]	0
F2	Step Function	30/200/500	[−100,100]	0
F3	Quartic Function	30/200/500	[−1.28,1.28]	0
F4	Generalized Rastrigin's Function	30/200/500	[−5.12,5.12]	0
F5	Ackley's Function	30/200/500	[−32,32]	0
F6	Generalized Penalized Function 1	3	[−50,50]	0
F7	Shekell's Foxholes Function	2	[−65,65]	0.998004
F8	Six-Hump Camel-Back Function	2	[−5,5]	−1.031630
F9	Hatman's Function 2	6	[0,1]	−3.322000

Table 2. Information about the CEC2019 test functions.

Function	Function Name	Dimension	Range	Optimum Value
CEC01	Storn's Chebyshev polynomial fitting problem	9	[−8192,8192]	1
CEC02	Inverse Hilbert matrix problem	16	[−16,384,16,384]	1
CEC03	Lennard-Jones minimum energy cluster	18	[−4,4]	1
CEC04	Rastrigin's function	10	[−100,100]	1
CEC05	Griewank's function	10	[−100,100]	1
CEC06	Weierstrass function	10	[−100,100]	1
CEC07	Modified Schwefel's function	10	[−100,100]	1
CEC08	Expanded Schaffer's F6 function	10	[−100,100]	1
CEC09	Happy cat function	10	[−100,100]	1
CEC10	Ackley function	10	[−100,100]	1

Table 3. Parameter values for IBSWHO and other algorithms.

Algorithm	Parameters
IBSWHO	PS = 0.2; PC = 0.13
IWHO	PS = 0.2; PC = 0.13; PRR = 0.1; w ∈ [0.01,0.99]
WHO	PS = 0.2; PC = 0.13
GWO	a = [2,0]
PSO	$c_1 = 2; c_2 = 2; W[0.2, 0.9]; vMax = 6$
GA	CR = 0.4; MR = 0.01

It can be seen from Table 4 that the proposed IBSWHO achieves the greatest advantages in accuracy and stability compared with the other four algorithms in most functions, regardless of unimodal, multimodal or fixed-dimension multimodal functions. Therefore, when solving optimization problems, IBSWHO finds it the easiest to exit out of the local optimum and has the best exploration and development ability.

From Table 5, it can be seen that compared to the other six algorithms, IBSWHO has the closest average value to the theoretical optimum in the eight functions, demonstrating the best convergence performance and search success rate. Therefore, IBSWHO possesses the capability to solve complex problems.

Table 4. Comparison results of the nine test functions.

Function	Dim	Metric	IBSWHO	IWHO	WHO	GWO	PSO
F1	30	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	1.00×10^{-43}	7.64×10^{-28}	1.01×10^{-04}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	4.14×10^{-43}	1.05×10^{-27}	1.07×10^{-04}
	200	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	1.30×10^{-33}	1.03×10^{-07}	3.29×10^{02}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	4.36×10^{-33}	6.03×10^{-08}	4.51×10^{01}
	500	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	1.02×10^{-29}	1.68×10^{-03}	5.88×10^{03}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	5.28×10^{-29}	7.19×10^{-04}	3.62×10^{02}
F2	30	Mean	$\mathbf{5.10 \times 10^{-06}}$	5.09×10^{-05}	2.46×10^{-02}	8.27×10^{-01}	1.98×10^{-04}
		Std	$\mathbf{1.03 \times 10^{-05}}$	4.64×10^{-05}	8.29×10^{-02}	3.23×10^{-01}	5.77×10^{-04}
	200	Mean	2.15×10^{01}	$\mathbf{2.19 \times 10^{00}}$	2.98×10^{01}	2.89×10^{01}	3.26×10^{02}
		Std	$\mathbf{1.73 \times 10^{00}}$	3.26×10^{00}	1.20×10^{01}	1.45×10^{00}	4.66×10^{01}
	500	Mean	8.63×10^{01}	$\mathbf{1.07 \times 10^{01}}$	1.26×10^{02}	9.14×10^{01}	5.95×10^{03}
		Std	$\mathbf{1.88 \times 10^{00}}$	1.34×10^{01}	1.22×10^{02}	1.88×10^{00}	4.31×10^{02}
F3	30	Mean	$\mathbf{1.98 \times 10^{-04}}$	2.23×10^{-04}	1.02×10^{-03}	2.06×10^{-03}	1.87×10^{-01}
		Std	$\mathbf{1.91 \times 10^{-04}}$	2.05×10^{-04}	6.73×10^{-04}	1.09×10^{-03}	6.71×10^{-02}
	200	Mean	$\mathbf{2.45 \times 10^{-04}}$	4.58×10^{-04}	1.56×10^{-03}	1.47×10^{-02}	7.84×10^{03}
		Std	$\mathbf{2.25 \times 10^{-04}}$	4.58×10^{-04}	1.21×10^{-03}	5.52×10^{-03}	7.82×10^{02}
	500	Mean	$\mathbf{3.19 \times 10^{-04}}$	5.06×10^{-04}	2.05×10^{-03}	5.08×10^{-02}	5.75×10^{04}
		Std	$\mathbf{2.64 \times 10^{-04}}$	6.25×10^{-04}	1.97×10^{-03}	1.24×10^{-02}	2.46×10^{03}
F4	30	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	6.72×10^{-10}	2.68×10^{00}	5.92×10^{01}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	3.61×10^{-09}	4.55×10^{00}	1.43×10^{01}
	200	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	0.00×10^{00}	1.92×10^{01}	1.89×10^{03}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	0.00×10^{00}	1.56×10^{01}	1.23×10^{02}
	500	Mean	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	0.00×10^{00}	7.70×10^{01}	6.54×10^{03}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	0.00×10^{00}	2.90×10^{01}	4.02×10^{02}
F5	30	Mean	$\mathbf{8.88 \times 10^{-16}}$	8.88×10^{-16}	1.84×10^{-15}	1.04×10^{-13}	1.15×10^{-01}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	1.60×10^{-15}	1.86×10^{-14}	3.05×10^{-01}
	200	Mean	$\mathbf{8.88 \times 10^{-16}}$	8.88×10^{-16}	1.60×10^{-15}	2.49×10^{-05}	6.63×10^{00}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	1.45×10^{-15}	7.99×10^{-06}	3.18×10^{-01}
	500	Mean	$\mathbf{8.88 \times 10^{-16}}$	8.88×10^{-16}	1.95×10^{-15}	1.85×10^{-03}	1.18×10^{01}
		Std	$\mathbf{0.00 \times 10^{00}}$	0.00×10^{00}	2.12×10^{-15}	3.62×10^{-04}	3.18×10^{-01}
F6	30	Mean	$\mathbf{3.35 \times 10^{-08}}$	6.24×10^{-07}	2.10×10^{-02}	3.81×10^{-02}	6.91×10^{-03}
		Std	$\mathbf{2.51 \times 10^{-08}}$	1.16×10^{-06}	4.21×10^{-02}	1.56×10^{-02}	2.63×10^{-2}
	200	Mean	1.52×10^{-01}	$\mathbf{9.49 \times 10^{-04}}$	4.09×10^{-01}	5.28×10^{-01}	3.86×10^{01}
		Std	4.32×10^{-02}	$\mathbf{1.56 \times 10^{-03}}$	2.57×10^{-01}	5.97×10^{-02}	1.41×10^{01}
	500	Mean	4.97×10^{-01}	$\mathbf{1.44 \times 10^{-03}}$	7.29×10^{-01}	7.67×10^{-01}	2.47×10^{-05}
		Std	4.89×10^{-02}	$\mathbf{2.08 \times 10^{-03}}$	2.18×10^{-01}	5.97×10^{-02}	8.28×10^{04}
F7	2	Mean	$\mathbf{9.98 \times 10^{-01}}$	9.98×10^{-01}	1.69×10^{00}	4.85×10^{00}	3.24×10^{00}
		Std	$\mathbf{8.25 \times 10^{-17}}$	1.30×10^{-16}	1.08×10^{00}	4.47×10^{00}	2.28×10^{00}
F8	2	Mean	$\mathbf{-1.03 \times 10^{00}}$	-1.03×10^{00}	-1.03×10^{00}	-1.03×10^{00}	-1.03×10^{00}
		Std	$\mathbf{4.52 \times 10^{-16}}$	6.05×10^{-16}	5.13×10^{-16}	2.99×10^{-05}	6.39×10^{-16}
F9	6	Mean	$\mathbf{-3.32 \times 10^{00}}$	-3.26×10^{00}	-3.27×10^{00}	-3.28×10^{00}	-3.29×10^{00}
		Std	$\mathbf{2.17 \times 10^{-02}}$	6.05×10^{-02}	6.81×10^{-02}	6.13×10^{-02}	5.54×10^{-02}

Table 5. Comparison results of the CEC2019 test functions.

Function	Metric	IBSWHO	SASS [54]	COLSHADE [54]	KGE [40]	WHO	GWO	PSO
CEC01	Mean	$\mathbf{3.89 \times 10^{04}}$	4.05×10^{04}	5.43×10^{04}	6.19×10^{04}	4.06×10^{04}	4.30×10^{08}	2.16×10^{12}
	Std	$\mathbf{1.03 \times 10^{03}}$	2.72×10^{03}	3.52×10^{03}	6.57×10^{03}	2.82×10^{03}	8.73×10^{08}	1.87×10^{12}
CEC02	Mean	$\mathbf{1.73 \times 10^{01}}$	1.73×10^{01}	1.73×10^{01}	1.75×10^{01}	1.73×10^{01}	1.74×10^{01}	1.34×10^{04}
	Std	$\mathbf{9.14 \times 10^{-15}}$	8.72×10^{-09}	2.13×10^{-08}	1.10×10^{-01}	7.23×10^{-15}	5.87×10^{-02}	3.73×10^{03}
CEC03	Mean	$\mathbf{1.27 \times 10^{01}}$	1.27×10^{01}	1.27×10^{01}	1.27×10^{01}	1.27×10^{01}	1.27×10^{01}	1.27×10^{01}
	Std	4.41×10^{-09}	$\mathbf{0.00 \times 10^{00}}$	7.14×10^{-09}	5.67×10^{-07}	3.61×10^{-15}	2.39×10^{-04}	3.61×10^{-15}
CEC04	Mean	1.63×10^{01}	$\mathbf{8.42 \times 10^{00}}$	9.85×10^{00}	1.87×10^{01}	9.85×10^{00}	1.42×10^{02}	1.67×10^{01}
	Std	8.67×10^{00}	4.29×10^{00}	$\mathbf{2.66 \times 10^{00}}$	8.90×10^{00}	3.57×10^{00}	3.16×10^{02}	6.20×10^{00}
CEC05	Mean	$\mathbf{1.06 \times 10^{00}}$	1.08×10^{00}	1.23×10^{03}	2.31×10^{00}	1.07×10^{00}	1.45×10^{00}	1.15×10^{00}
	Std	$\mathbf{3.62 \times 10^{-02}}$	6.23×10^{-02}	5.49×10^{-02}	2.96×10^{-01}	1.71×10^{-02}	2.39×10^{-01}	7.99×10^{-02}
CEC06	Mean	$\mathbf{5.35 \times 10^{00}}$	5.76×10^{00}	7.33×10^{00}	8.95×10^{00}	5.86×10^{00}	1.67×10^{01}	9.06×10^{00}
	Std	7.60×10^{-01}	$\mathbf{2.15 \times 10^{-01}}$	3.19×10^{-01}	9.96×10^{-01}	8.45×10^{-01}	9.28×10^{-01}	1.47×10^{00}
CEC07	Mean	$\mathbf{5.28 \times 10^{01}}$	7.66×10^{01}	9.85×10^{01}	4.38×10^{02}	5.42×10^{01}	3.05×10^{02}	1.47×10^{02}
	Std	9.11×10^{01}	$\mathbf{6.43 \times 10^{01}}$	7.21×10^{01}	2.24×10^{02}	1.02×10^{02}	2.31×10^{02}	1.12×10^{02}
CEC08	Mean	4.65×10^{00}	$\mathbf{4.39 \times 10^{00}}$	4.89×10^{03}	5.42×10^{00}	4.96×10^{00}	5.17×10^{00}	5.28×10^{00}
	Std	5.48×10^{-01}	8.35×10^{-01}	6.83×10^{-01}	$\mathbf{4.82 \times 10^{-01}}$	5.74×10^{-01}	1.07×10^{00}	7.50×10^{00}
CEC09	Mean	$\mathbf{2.38 \times 10^{00}}$	2.44×10^{00}	2.42×10^{00}	2.44×10^{00}	2.45×10^{00}	1.36×10^{01}	2.37×10^{00}
	Std	2.95×10^{-02}	$\mathbf{4.19 \times 10^{-04}}$	7.03×10^{-01}	6.63×10^{-02}	1.23×10^{-02}	5.12×10^{01}	1.88×10^{-02}
CEC10	Mean	$\mathbf{1.94 \times 10^{01}}$	1.99×10^{01}	2.00×10^{01}	1.99×10^{01}	1.98×10^{01}	2.05×10^{01}	1.96×10^{01}
	Std	$\mathbf{1.12 \times 10^{-02}}$	1.17×10^{-01}	8.53×10^{-02}	1.31×10^{-02}	5.78×10^{00}	7.91×10^{-02}	3.71×10^{00}

However, this is not the case for all test functions, such as F2, F6, CEC04 and CEC08, which is normal, as the NFL theorem suggests that one optimization algorithm cannot solve all problems.

3.2. Experimental Results and Analysis of Band Selection

To certify the effect of IBSWHO in HSI BS, it is compared with several meta-heuristic optimization algorithms and common and advanced band selection methods. To find the band subset that minimizes the classification error rate, all algorithms employ the classification error rate under the SVM classifier as the objective function, given that it stands out as one of the most competitive classifiers in small-sample problems. The LIBSVM library is utilized to implement the SVM classifier, with the radial basis function (RBF) serving as the chosen kernel. During the training phase, the two parameters of the SVM (c and γ) are selected through 5-fold cross-validation.

3.2.1. Description of Datasets

The experiment is conducted using three commonly used hyperspectral datasets, which were obtained by different sensors, namely the Indian Pines, Pavia University, and Salinas datasets.

(1) Indian Pines: The Indian Pines dataset is imaged by the airborne visual infrared imaging spectrometer (AVIRIS) in 1992 on a patch of Indian Pine trees in Indiana, USA, acquiring an image size of 145 × 145 pixels, a spectrum range of 0.4–2.5 μm and a spatial resolution of 20 m, containing 220 bands and 16 ground truth classes. In this paper, we use the bands that contain the absorbing region removed for a total of 200 bands. The pseudo-color image and the ground truth image are shown in Figure 4.

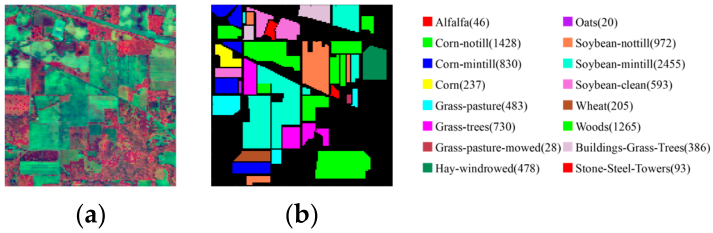


Figure 4. Indian Pines dataset. (**a**) Pseudo-color image; (**b**) Color ground truth images with class labels.

(2) Pavia University: The Pavia University dataset is a portion of hyperspectral data imaged by the German airborne reflection optical spectral imager (ROSIS) in Pavia, Italy, in 2003. The size of the images is 610 × 340 pixels, the spectrum range is 0.43–0.86 mm and the spatial resolution is about 1.3 m. It contains 115 bands and 9 classes of ground truth. In this paper, 103 available bands are used for subsequent research after removing 12 noise bands. The pseudo-color image and real image of the ground are shown in Figure 5.

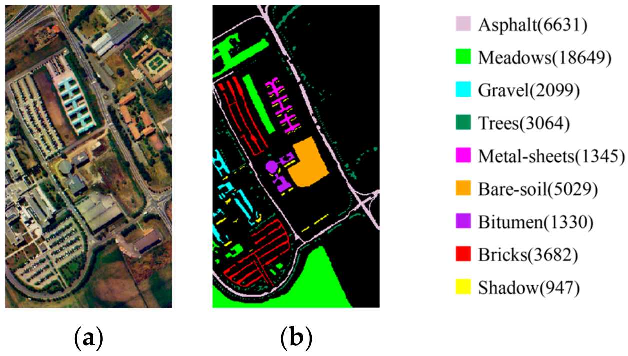


Figure 5. Pavia University dataset. (**a**) Pseudo-color image; (**b**) Color ground truth images with class labels.

(3) Salinas: The Salinas dataset is also imaged by AVIRIS, which is an image of the Salinas Valley in California, USA. The size of the image is 512 × 217 pixels, the spectrum range is 0.4–2.5 µm and the spatial resolution is 3.7 m. It contains 224 bands and 16 kinds of ground objects. The pseudo-color image and the real image of the ground are shown in Figure 6.

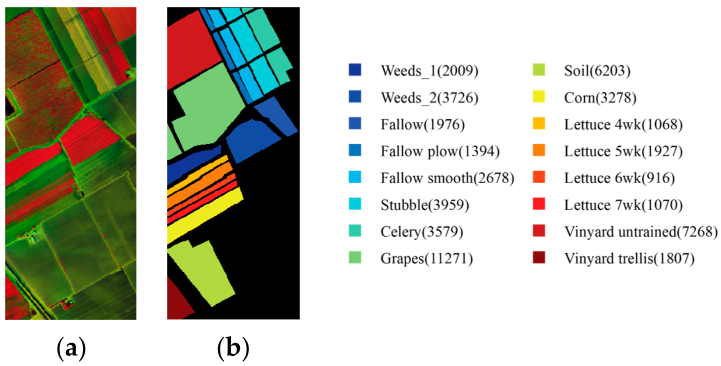


Figure 6. Salinas dataset. (**a**) Pseudo-color image; (**b**) Color ground truth images with class labels.

3.2.2. Experimental Parameter Settings

Firstly, the proposed IBSWHO is compared with two basic and selection methods (MRMR and Relief), and three recently proposed effective BS methods (ASPS, FNGBS and NGNMF). Additionally, other optimization algorithms, including PSO, GWO, GA and

WHO, are also used for further comparison with the parameters set as shown in Table 3. Selected band subsets are classified by SVM, and the classification accuracy is used to assess the discrimination ability of the band subsets. For each dataset, 20% of samples are randomly selected as training data, and the remaining 80% are used as testing data. To ensure the fairness of the experimental results, each algorithm is repeated 10 times on each dataset with the initial population of 20, and the maximum number of iterations is 20. There are numerous evaluation metrics to assess the classification performance in HSI classification. In this paper, we utilize the averages of OA (overall accuracy), AA (average accuracy), Kappa (Kappa coefficient) and each category accuracy of the band subset for evaluation. The concepts of OA, AA and Kappa are described in Appendix A.

3.2.3. Analysis of Experimental Results

Table 6 lists the average results of IBSWHO and other competing methods running on the Indian Pines dataset. As shown in the table, IBSWHO has the highest classification accuracy and the best performance in all methods. Class 3 (Corn-mintil) increased from 40.06% to 77.56%. Class 12 (Soybean-clean) improved its classification accuracy by 7.98% compared with the best-performing NGNMF. However, it can also be observed that IBSWHO performs poorly in some categories, possibly because similar spectral land-cover categories make it more difficult to distinguish between categories. The classification effect of Class 9 (Oats) is not ideal owing to the small sample size in this category. Classification plots for all methods on the Indian Pines dataset are shown in Figure 7. The black background represents unlabeled pixels, colors consistent with the true sample color represent correctly classified pixels and inconsistent colors represent misclassified pixels. The more correctly classified pixels, the smoother the classification plot is. From the figure, for example, we can see that the classification map obtained by IBSWHO is the smoothest, compared to that obtained by other methods that produce the Soybean-clean region.

Table 6. Experimental results on the Indian Pines dataset.

Class	MRMR	Relief	GA	PSO	GWO	WHO	ASPS	FNGBS	NGNMF	IBSWHO
1	22.22%	58.33%	69.44%	58.33%	58.33%	63.89%	63.89%	75.61%	93.02%	75.00%
2	65.32%	55.95%	73.82%	74.08%	71.10%	72.94%	67.78%	80.47%	79.79%	83.28%
3	40.81%	40.06%	57.38%	56.17%	58.58%	53.61%	57.98%	64.79%	61.68%	77.56%
4	20.11%	20.11%	45.50%	50.26%	45.50%	52.91%	68.78%	63.85%	42.22%	71.43%
5	91.71%	81.35%	93.26%	93.26%	93.01%	93.00%	87.23%	91.24%	89.52%	94.56%
6	97.26%	97.77%	97.43%	96.75%	98.63%	97.26%	91.95%	95.13%	95.09%	95.21%
7	54.55%	0.00%	86.36%	72.73%	72.72%	81.82%	68.18%	84.00%	92.31%	77.27%
8	99.21%	99.48%	97.91%	98.69%	97.38%	98.43%	98.43%	97.44%	99.34%	97.12%
9	0.00%	0.00%	0.00%	0.00%	0.00%	18.75%	37.50%	22.22%	63.15%	43.75%
10	49.42%	42.86%	69.24%	71.04%	67.05%	72.72%	69.50%	70.61%	82.01%	78.38%
11	83.60%	84.93%	88.09%	87.98%	88.75%	88.75%	84.98%	81.89%	86.41%	84.32%
12	22.15%	25.74%	74.26%	67.93%	72.57%	70.46%	67.09%	78.24%	72.82%	80.80%
13	98.78%	95.73%	98.78%	99.39%	98.78%	94.51%	98.78%	96.20%	96.91%	98.17%
14	96.54%	97.13%	96.74%	97.23%	97.13%	94.37%	96.34%	95.61%	94.50%	91.90%
15	50.32%	43.83%	61.69%	57.79%	60.39%	59.74%	50.97%	48.42%	68.31%	65.91%
16	98.65%	93.24%	97.30%	95.95%	97.30%	81.08%	93.24%	90.36%	89.78%	97.30%
AA	61.92%	58.53%	75.45%	73.60%	73.58%	74.29%	75.20%	77.25%	81.68%	82.00%
OA	71.69%	69.51%	81.52%	81.17%	81.06%	80.94%	79.02%	81.37%	83.18%	84.92%
Kappa	67.07%	64.32%	78.72%	78.32%	78.17%	78.23%	75.87%	78.71%	80.76%	82.80%

The average of the experimental results on the Pavia University dataset is shown Table 7.

It can be seen from Table 7 that the classification result of IBSWHO is the highest in all algorithms, especially the increase from 9.96% to 86.94% for Class 7 (Bitumen), and the individual accuracy of Class 6 (Bare-soil) which also increased by 18.37% compared to ASPS. IBSWHO's Kappa coefficient is greater than 0.93, indicating that the predicted labels are generally consistent with the true labels, so IBSWHO has a strong optimization ability on the Pavia University dataset. The classification graphs on this dataset are shown in Figure 8, which show that the classification image of IBSWHO is the smoothest. For example, the Bricks region generated by IBSWHO has the fewest misclassified pixels which is the smoothest.

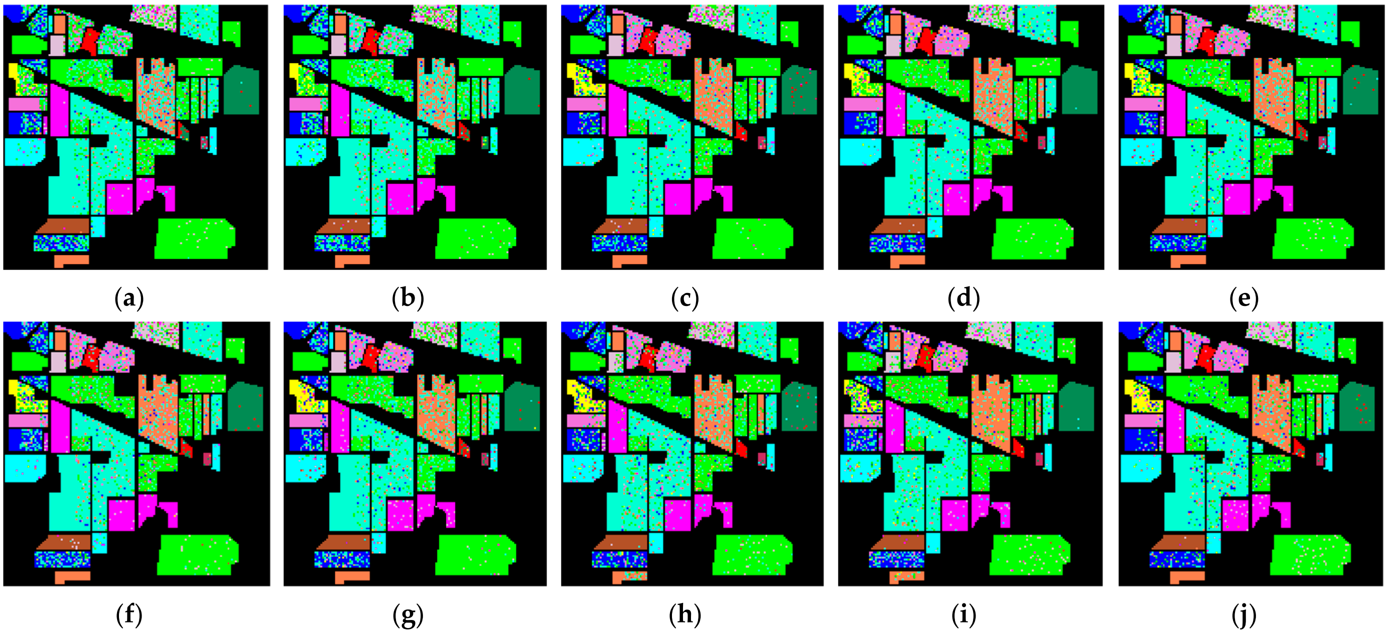


Figure 7. Plots of classification results on the Indian Pines dataset. (**a**–**j**) are the classification plots after band selection by MRMR, Relief, GA, PSO, GWO, WHO, ASPS, FNGBS, NGNMF and IBSWHO, respectively.

Table 7. Experimental results on the Pavia University dataset.

Class	MRMR	Relief	GA	PSO	GWO	WHO	ASPS	FNGBS	NGNMF	IBSWHO
1	91.14%	93.63%	94.27%	94.21%	93.72%	94.44%	93.93%	93.58%	96.40%	94.59%
2	98.30%	97.96%	98.15%	98.37%	98.45%	98.49%	96.84%	97.16%	96.90%	98.46%
3	50.74%	65.28%	75.82%	76.12%	77.01%	77.96%	73.50%	75.54%	85.61%	80.03%
4	90.00%	88.70%	95.84%	94.94%	94.78%	94.82%	95.84%	95.18%	91.48%	95.23%
5	99.54%	99.63%	99.44%	99.63%	99.63%	99.63%	99.81%	99.75%	99.45%	99.91%
6	39.75%	49.39%	86.75%	86.68%	85.58%	84.61%	71.12%	89.59%	83.76%	89.49%
7	53.95%	9.96%	84.12%	81.58%	86.28%	83.65%	76.88%	84.04%	90.34%	86.94%
8	87.40%	91.38%	90.25%	91.10%	90.97%	90.63%	89.41%	87.99%	90.02%	91.58%
9	100.00%	99.74%	100.00%	99.87%	100.00%	99.60%	100.00%	100.00%	99.89%	100.00%
AA	78.98%	77.30%	91.63%	91.38%	91.82%	91.54%	88.59%	91.43%	92.64%	92.73%
OA	85.14%	86.10%	93.91%	93.92%	93.94%	93.88%	91.05%	93.46%	93.64%	94.75%
Kappa	79.63%	81.01%	91.89%	91.89%	91.92%	91.83%	88.01%	91.32%	91.54%	93.02%

The average result on the Salinas dataset is shown Table 8.

It is observed from Table 8 that the OAs of the BS methods using optimization algorithms are all above 90%, indicating that the optimization algorithms have a good band selection effect on the Salinas dataset. Furthermore, the results of IBSWHO are still the best. For instance, the accuracy of Class 14 (Lettuce7 wk) was 3.85% higher than that of PSO, and Class 15 (Vinyarduntrained) improved by 16.33% over PSO. Compared to the performance of the three advanced methods, IBSWHO demonstrates the best performance, indicating that IBSWHO is the most feasible BS method for the Salinas dataset. Classification plots are shown in Figure 9, which show that the plot of IBSWHO is the smoothest and clearest.

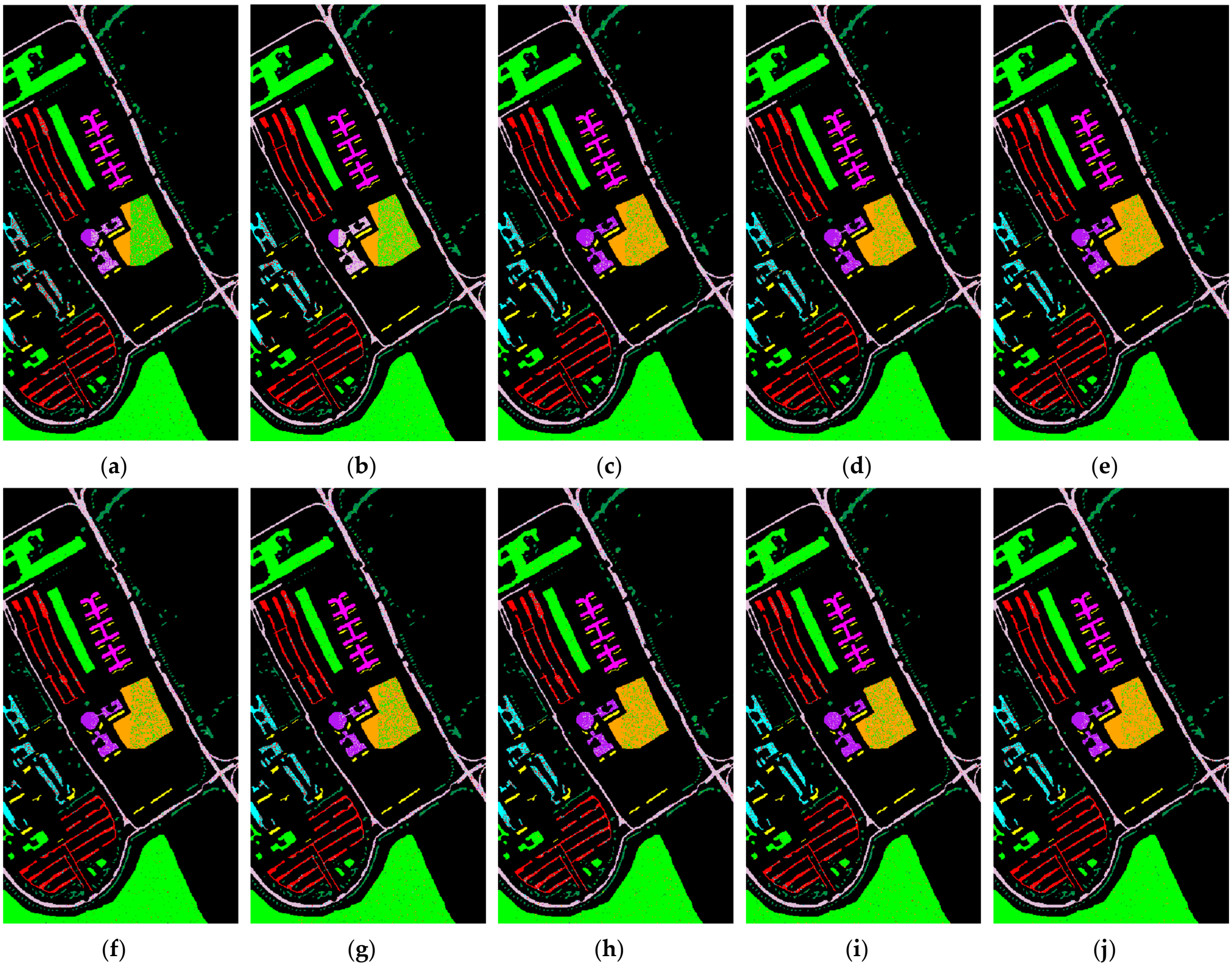


Figure 8. Plots of classification results on the Pavia University dataset. (**a**–**j**) are the classification plots after band selection by MRMR, Relief, GA, PSO, GWO, WHO, ASPS, FNGBS, NGNMF and IBSWHO, respectively.

Table 8. Experimental results on the Salinas dataset.

Class	MRMR	Relief	GA	PSO	GWO	WHO	ASPS	FNGBS	NGNMF	IBSWHO
1	98.01%	98.13%	99.07%	98.94%	99.32%	99.32%	99.32%	99.61%	99.00%	99.56%
2	99.77%	99.93%	100.00%	99.90%	99.83%	99.80%	99.97%	99.88%	100.00%	100.00%
3	95.32%	93.54%	99.11%	99.18%	99.24%	99.30%	99.43%	99.94%	99.43%	99.94%
4	99.55%	99.64%	99.55%	99.46%	99.64%	99.55%	99.78%	99.44%	98.83%	99.82%
5	94.58%	94.91%	97.57%	97.85%	97.43%	97.29%	99.49%	98.71%	99.35%	99.44%
6	99.46%	99.84%	99.81%	99.87%	99.87%	99.91%	99.78%	99.94%	99.84%	99.94%
7	99.55%	99.76%	99.69%	99.72%	99.41%	99.51%	99.69%	99.25%	99.79%	99.93%
8	91.39%	90.59%	91.60%	92.20%	91.34%	91.65%	88.78%	90.20%	90.88%	90.24%
9	97.42%	99.19%	99.88%	99.52%	99.96%	99.94%	99.92%	99.96%	99.97%	99.92%
10	92.49%	94.16%	94.24%	94.97%	93.94%	94.47%	95.61%	97.73%	97.36%	97.75%
11	79.39%	89.70%	92.51%	94.03%	93.01%	94.03%	92.15%	98.54%	98.71%	99.30%
12	99.94%	100.00%	99.94%	99.81%	99.94%	99.94%	99.94%	99.65%	100.00%	99.94%
13	97.13%	97.54%	98.09%	97.81%	98.09%	97.13%	99.73%	99.15%	98.77%	99.73%
14	95.21%	93.34%	95.79%	94.51%	95.56%	95.56%	98.13%	97.20%	94.86%	98.36%
15	49.09%	51.34%	58.69%	57.96%	59.44%	59.37%	60.60%	63.43%	70.57%	74.29%
16	98.69%	98.41%	98.82%	99.03%	98.89%	98.75%	99.38%	96.62%	99.58%	99.52%
AA	92.94%	93.89%	95.27%	95.30%	95.31%	95.35%	95.70%	96.20%	96.69%	97.34%
OA	89.41%	90.06%	91.80%	91.84%	91.89%	91.91%	91.75%	92.53%	93.68%	94.23%
Kappa	88.17%	88.90%	90.85%	90.89%	90.95%	90.97%	90.79%	91.66%	92.95%	93.57%

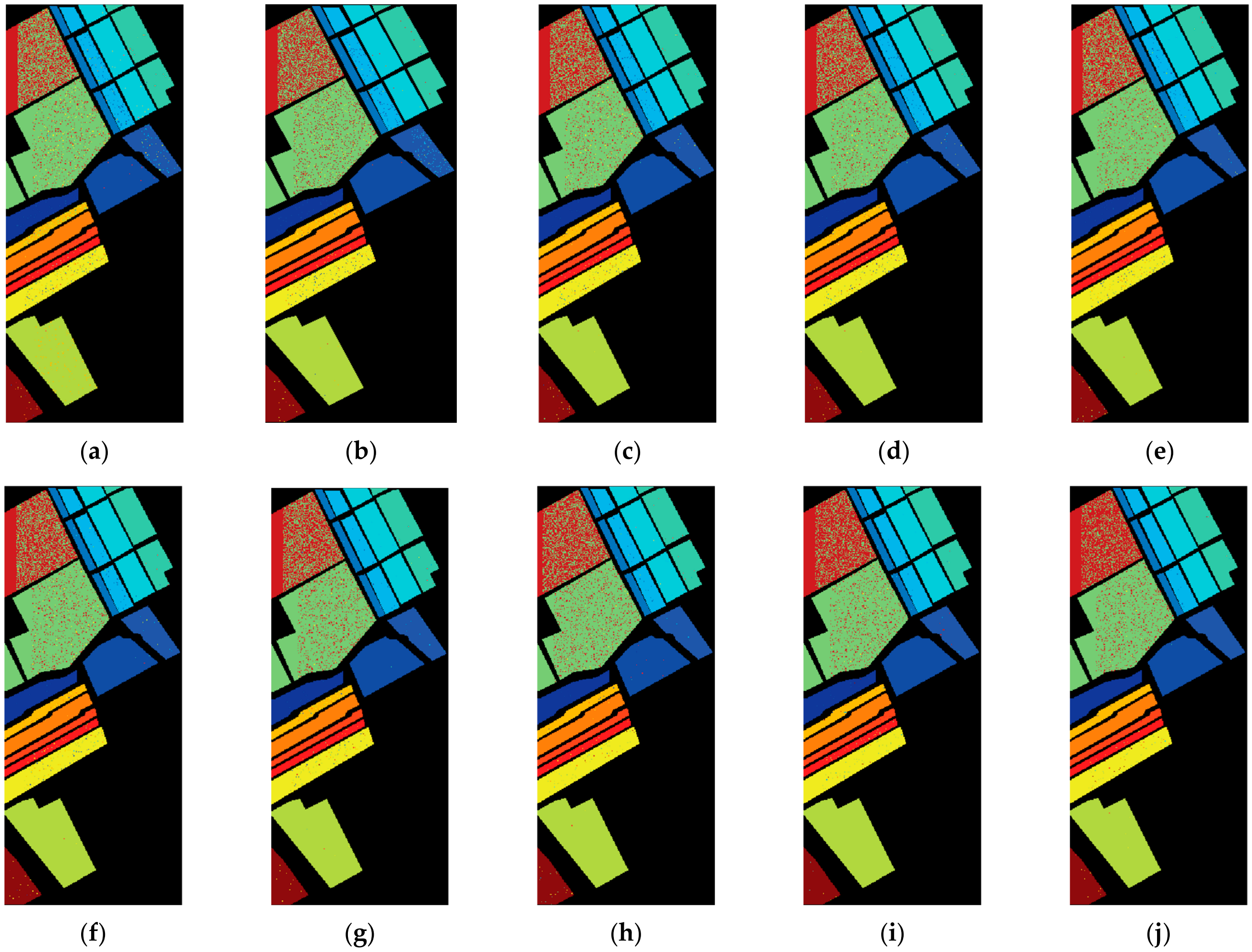
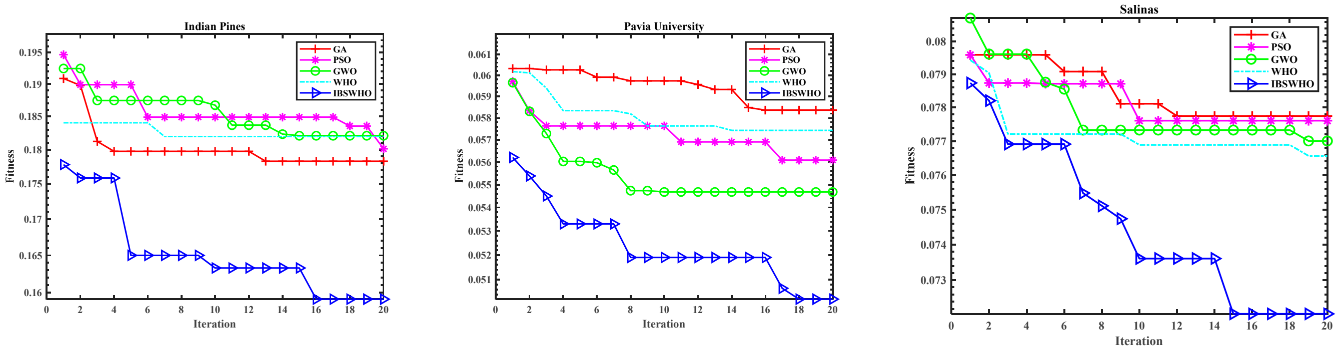


Figure 9. Plots of classification results on the Salinas dataset. (**a**–**j**) are the classification plots after band selection by MRMR, Relief, GA, PSO, GWO, WHO, ASPS, FNGBS, NGNMF and IBSWHO, respectively.

In summary, in terms of classification, compared with the basic ranking and filtering methods, the wrapper based on an optimization algorithm is the most effective. The reason is that the basic ranking and filtering methods use mutual information as the indicator of selecting band subset rather than a classifier system, so the filter has a short time but a low accuracy. Compared with other optimization algorithms and advanced band selection techniques, although the accuracy of IBSWHO is lower than that of the comparison algorithm in some classes, the overall accuracy is the highest, which can reflect the effectiveness of the added modification strategies. Moreover, the performance of IBSWHO in the three datasets is the best, which also shows the robustness of IBSWHO.

In terms of convergence performance, curves for the variation in fitness with iterations are displayed in Figure 10. As can be observed, the initial fitness of IBSWHO is the lowest in all datasets, indicating that the Sobol sequence used in the initialization phase can enhance population diversity and provide a better solution. With the increase in iterations, the iterative curves of GA, PSO, GWO and WHO tend to stabilize, while IBSWHO maintains a declining trend, which indicates that the mutation strategy and dynamic random search technique are helpful to improve the exploration and development ability of IBSWHO and help to find a better band subset. Thus, IBSWHO has an excellent optimization ability to find the best band subset with the best classification accuracy.



(**a**) (**b**) (**c**)

Figure 10. Convergence plot of optimization algorithms for different datasets: (**a**) Indian Pines, (**b**) Pavia University, (**c**) Salinas.

3.2.4. Comparison of the Number of Selected Bands

Due to the adaptive selection of bands based on the task, both the selected bands and the number of bands are not fixed. In the above band selection experiments, we present the average number of bands selected by the optimal band subset in Table 9.

Table 9. The number of bands in the optimal band subsets selected by different optimization algorithms.

Dataset	IBSWHO	GA	PSO	GWO	WHO
Indian pines	98	112	108	96	115
Pavia University	61	63	66	59	65
Salinas	110	131	124	111	118

From the table, it can be observed that GWO selects the smallest number of band subsets. This is because GWO is an efficient search algorithm with diverse search strategies, allowing it to effectively explore different band subsets and find the optimal solution with the minimum number of bands. However, although GWO selects the smallest number of bands, as shown in Tables 6–8, its classification accuracy is not high. This indicates that GWO may lose bands with high information content, and the features are not well preserved. In addition to GWO, the proposed IBSWHO selects the smallest number of bands, and its classification accuracy is the highest. This suggests that IBSWHO can select the most critical and effective features, and the small number of bands indicates the strong generalization of the algorithm.

3.2.5. Statistical Significance Evaluation

In order to evaluate the statistical significance of the differences between IBSWHO and other comparison algorithms for classification, we carried out a nonparametric McNemar test [55]. It is based on a standardized normal test statistic:

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}, \tag{13}$$

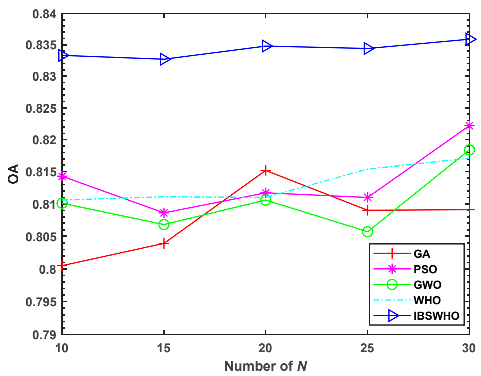
where f_{12} indicates the number of samples correctly classified by Method 1 but incorrectly classified by Method 2, and f_{21} is the number of samples correctly classified by Method 2 but incorrectly classified by Method 1. For a 5% significance level, if $|z| > 1.96$, there is a significant performance difference between the two methods. Table 10 presents the value of $|z|$ comparing IBSWHO with the other comparison methods. The results of the McNemar test show that the performance of IBSWHO is statistically different from other methods, and only has no significant difference from the NGNMF method on the Salinas dataset. The reason for this is that the Salinas dataset contains a large area of land cover and a sufficient number of samples. The strategy of selecting similar pixels in NGNMF can adaptively extract land-cover features, thus demonstrating a stronger robustness for complex land-cover scenarios. Therefore, NGNMF can achieve better classification results, but the proposed IBSWHO method still outperforms it.

Table 10. Values for each optimization algorithm at the 5% significance level, considered significant if $|z| > 1.96$.

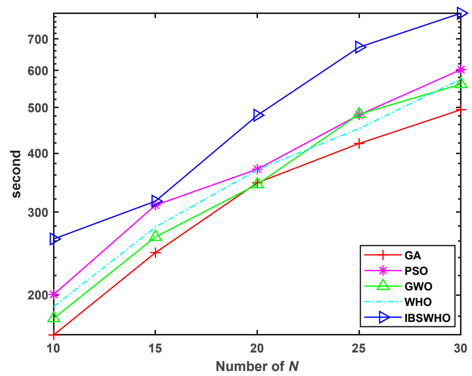
	Indian Pines		**Pavia University**		**Salinas**	
	IBSWHO		**IBSWHO**		**IBSWHO**	
	$\|z\|$	significant?	$\|z\|$	significant?	$\|z\|$	significant?
MRMR	16.08	yes	25.16	yes	15.75	yes
Relief	19.92	yes	20.41	yes	12.76	yes
GA	4.48	yes	22.52	yes	7.12	yes
PSO	5.02	yes	2.20	yes	6.99	yes
GWO	5.07	yes	2.18	yes	6.76	yes
WHO	5.53	yes	2.13	yes	6.72	yes
ASPS	7.75	yes	12.09	yes	7.27	yes
FNGBS	4.89	yes	4.46	yes	5.10	yes
NGNMF	2.59	yes	2.29	yes	1.59	no

3.2.6. Effect of Hyperparametric Population Size N on Accuracy

The population size N is defined at the initial stage of the algorithm. In this paper, N is set to 20, the value of which has a large impact on the convergence speed and solution accuracy of the algorithm. In this subsection, we discuss the effect of N on the accuracy of all algorithms on the Indian Pines dataset. The relationship among different numbers of N, OA, running time on the Indian Pines dataset is shown in Figure 11. According to the graph, when the number of populations is 20 and 30, the classification accuracy reaches the maximum.



(**a**)



(**b**)

Figure 11. Effects of different populations sizes on the Indian Pines dataset. (**a**) OA curves, (**b**) running time curves.

However, when population size is 30, the time complexity increases a lot, but there is little difference in classification accuracy, so the population size of 20 was selected in this experiment.

3.2.7. Ablation Analysis of IBSWHO

For the sake of further proving the effectiveness of incorporating different strategies, we conducted ablation experiments on three datasets. The three strategies are denoted by Sobol, Cauchy and dynamic, respectively. Each experiment was performed ten times and then the results were averaged. The results of experiments are shown in Table 11.

Table 11. Classification performance in ablation experiments on the three datasets.

Dataset	Metric	WHO	SobolWHO	CauchyWHO	DynamicWHO	SobolCauchyWHO	SobolDynamicWHO	CauchyDynamicWHO	IBSWHO
Indian Pines	OA	80.94	81.82	81.29	81.09	82.11	81.06	81.01	84.92
	AA	74.64	76.12	73.50	74.04	78.05	74.86	72.76	82.00
	Kappa	78.05	79.04	78.46	78.26	79.42	78.23	78.15	82.80
Pavia University	OA	93.88	94.11	94.58	94.29	94.60	93.87	94.39	94.75
	AA	91.54	91.97	92.49	92.16	92.69	91.77	92.35	92.73
	Kappa	91.83	92.15	92.71	92.39	92.75	91.82	92.52	93.02
Salinas	OA	91.91	92.17	91.90	92.54	92.30	91.95	92.54	94.23
	AA	95.35	95.53	95.23	95.85	95.74	95.47	96.12	97.34
	Kappa	90.97	91.26	90.95	91.52	91.40	91.01	91.64	93.57

The Indian Pines dataset suffers from limited training data and unbalanced category distribution. As shown in Table 11, the inclusion of the Sobol strategy resulted in a relatively significant improvement, with OA improving by 0.88% and AA improving by 1.48% over WHO. This is due to the fact that the Sobol sequence is a non-overlapping, low-bias random number generation method that allows for more accurate results with small sample sizes.

The Pavia University and Salinas datasets have a large scale and high band correlation, making it difficult to remove redundant information. The Cauchy variation and the dynamic random search technique can speed up the algorithm to jump out of the local optimum and expand the search range to select band subsets more efficiently. The Cauchy variation is more effective for the Pavia University dataset, which improves the OA by 0.7% compared to WHO. The dynamic random search technique is more effective for the Salinas dataset, which improves the OA by 0.63% compared to WHO. Moreover, different combinations of these strategies have different effects for different datasets, and only when the three strategies are added at the same time are the classification results the best in different datasets, which shows that the combination of the three strategies increases the robustness of the proposed IBSWHO.

In summary, the combination of three strategies allows IBSWHO to achieve excellent results in both band selection and classification, and the absence of strategy reduces classification accuracy.

4. Summary and Prospects

BS is a non-transformational feature selection method, which is important for improving the classification accuracy of HSI, removing redundant bands with high correlation and reducing computational complexity. In this paper, a new HSI BS method based on an enhanced wild horse optimizer (IBSWHO) is proposed. IBSWHO improves the performance of jumping out of local optimum by increasing population diversity and mutation to expand search ranges, and it automatically selects the most appropriate band subset for HSI classification task. To verify the effectiveness of IBSWHO, we use SVM as a classifier to compare IBSWHO with advanced band selection methods and other optimization algorithms on three commonly used hyperspectral datasets, and we use overall accuracy, average accuracy, Kappa coefficient and individual class accuracy as evaluation indicators. In accordance with experimental results, IBSWHO's classification accuracy is satisfactory, and for some classes with complex spectral features, it is also the best in comparative methods. Therefore, IBSWHO can select excellent bands for classification tasks, separate classes well and improve classification accuracy. Moreover, IBSWHO has a small number of parameters, which do not easily fall into the local optimum and converge stably to the global optimum solution. With fixed parameters, IBSWHO achieves good results both for the benchmarking function and for the hyperspectral band selection task, so this algorithm is a generally applicable and robust algorithm that does not involve a large number of hyperparameter adjustments.

However, as only classification accuracy is used as an objective function, the relationship between classification accuracy and the number of bands selected is not well balanced. At the same time, improved strategies increase the time complexity of the algorithm to some extent. Therefore, in future work, we will explore higher-quality objective functions and improve the execution efficiency of the algorithm.

Author Contributions: Conceptualization, Y.S. and H.C.; methodology, Y.S. and H.C.; software, T.C.: validation, Y.S.; resources, H.C.; data curation, Y.S.; writing—original draft preparation, Y.S. and H.C.; writing—review and editing, T.C. and W.D.; visualization, Y.S.; supervision, Y.S.; project administration, T.C.; funding acquisition, H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of Sichuan Province under Grant 2022NSFSC0536, and the Traction Power State Key Laboratory of Southwest Jiaotong University under Grant TPL2203.

Data Availability Statement: The original contributions presented in the study are public evaluation data and free satellite data, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

OA, AA and Kappa are commonly used indicators for evaluating the classification performance of HSI, and their values are calculated based on the confusion matrix. The confusion matrix C is a square matrix of size $N \times N$, where N represents the number of classes in the given dataset. C_{ii} represents the number of pixels correctly classified as class i, while C_{ij} represents the number of pixels of class i classified as class j. The columns of the matrix represent the number of true pixels in each class, while the rows represent the number of predicted pixels in each class.

OA represents the proportion of correctly classified samples to the total number of samples, calculated as follows:

$$\mathrm{OA} = \frac{\sum_{i=1}^{N} C_{ii}}{\sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij}} \tag{A1}$$

AA represents the average accuracy for each class, calculated as follows:

$$\mathrm{AA} = \frac{\sum_{i=1}^{N} CA_i}{\sum_{i=1}^{N} T_i} \tag{A2}$$

where T_i is the total number of test samples for class i and $CA_i = \frac{\sum_i^N C_{ii}}{T_i}$.

Kappa provides mutual information about the strong consistency between the ground truth image and the classification image. The calculation method is as follows:

$$\mathrm{Kappa} = \frac{\mathrm{OA} - F_e}{1 - F_e} \tag{A3}$$

where $F_e = \frac{\sum_{i=1}^{N} F_{i.} F_{.i}}{\left(\sum_{i=1}^{N} T_i\right)^2}$, $F_{i.}$ and $F_{.i}$ represent the sum of the ith row and column of the confusion matrix, respectively.

References

1. Yu, C.; Zhao, X.; Gong, B.Y.; Hu, Y.; Song, M.; Yu, H.; Chang, C.-I. Distillation-Constrained Prototype Representation Network for Hyperspectral Image Incremental Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 3359629. [CrossRef]
2. Chen, H.Y.; Long, H.Y.; Chen, T.; Song, Y.; Chen, H.; Zhou, X.; Deng, W. M^{3}FuNet: An Unsupervised Multivariate Feature Fusion Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5513015.
3. Lu, Y.; Wang, Y.; Parikh, D.; Khan, A.; Lu, G. Simultaneous Direct Depth Estimation and Synthesis Stereo for Single Image Plant Root Reconstruction. *IEEE Trans. Image Process.* **2021**, *30*, 4883–4893. [CrossRef] [PubMed]
4. Long, H.; Chen, T.; Chen, H.; Zhou, X.; Deng, W. Principal Space Approximation Ensemble Discriminative Marginalized Least-Squares Regression for Hyperspectral Image Classification. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108031. [CrossRef]

5. Chen, H.; Wang, T.; Chen, T.; Deng, W. Hyperspectral Image Classification Based on Fusing S3-PCA, 2D-SSA and Random Patch Network. *Remote Sens.* **2023**, *15*, 3402. [CrossRef]
6. Jiang, Y.; Li, C. mRMR-Based Feature Selection for Classification of Cotton Foreign Matter Using Hyperspectral Imaging. *Comput. Electron. Agric.* **2015**, *119*, 191–200. [CrossRef]
7. Xu, B.; Li, X.; Hou, W.; Wang, Y.; Wei, Y. A Similarity-Based Ranking Method for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9585–9959. [CrossRef]
8. Wang, Q.; Lin, J.; Yuan, Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1279–1289. [CrossRef]
9. Wang, Q.; Li, Q.; Li, X. A Fast Neighborhood Grouping Method for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5028–5039. [CrossRef]
10. Kira, K.; Rendell, L.A. A Practical Approach to Feature Selection. In *Machine Learning Proceedings 1992*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 249–256.
11. Fu, H.; Zhang, A.; Sun, G.; Ren, J.; Jia, X.; Pan, Z.; Ma, H. A Novel Band Selection and Spatial Noise Reduction Method for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
12. Wang, Q.; Li, Q.; Li, X. Hyperspectral Band Selection via Adaptive Subspace Partition Strategy. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4940–4950. [CrossRef]
13. Zhang, J. A Hybrid Clustering Method with a Filter Feature Selection for Hyperspectral Image Classification. *J. Imaging* **2022**, *8*, 180. [CrossRef]
14. Wang, J.; Tang, C.; Liu, X.; Zhang, W.; Li, W.; Zhu, X.; Wang, L.; Zomaya, A.Y. Region-Aware Hierarchical Latent Feature Representation Learning-Guided Clustering for Hyperspectral Band Selection. *IEEE Trans. Cybern.* **2023**, *53*, 5250–5263. [CrossRef] [PubMed]
15. Feng, J.; Bai, G.; Li, D.; Zhang, X.; Shang, R.; Jiao, L. MR-Selection: A Meta-Reinforcement Learning Approach for Zero-Shot Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–20. [CrossRef]
16. Yao, Z.; Wang, Z.; Wang, D.; Wu, J.; Chen, L. An ensemble CNN-LSTM and GRU adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as input. *J. Hydrol.* **2023**, *625*, 129977. [CrossRef]
17. Deng, W.; Chen, X.; Li, X.; Zhao, H. Adaptive Federated Learning with Negative Inner Product Aggregation. *IEEE Internet Things J.* **2023**, *11*, 6570–6581. [CrossRef]
18. Zhang, Y.; Wang, X.; Jiang, X.; Zhou, Y. Robust Dual Graph Self-Representation for Unsupervised Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
19. Li, M.; Wang, Y.Q.; Yang, C.; Lu, Z.; Chen, J. Automatic Diagnosis of Depression Based on Facial Expression Information and Deep Convolutional Neural Network. *IEEE Trans. Comput. Soc. Syst.* **2024**. [CrossRef]
20. Li, T.; Shu, X.; Wu, J.; Zheng, Q.; Lv, X.; Xu, J. Adaptive Weighted Ensemble Clustering via Kernel Learning and Local Information Preservation. *Knowl.-Based Syst.* **2024**, *294*, 111793. [CrossRef]
21. Xie, C.; Zhou, L.; Ding, S.F.; Lu, M.; Zhou, X. Research on self-propulsion simulation of a polar ship in a brash ice channel based on body force model. *Int. J. Nav. Archit. Ocean Eng.* **2023**, *15*, 100557. [CrossRef]
22. Li, X.; Zhao, H.; Deng, W. IOFL: Intelligent-Optimization-Based Federated Learning for Non-IID Data. *IEEE Internet Things J.* **2024**, *11*, 16693–16699. [CrossRef]
23. Wang, Z.; Wang, Q.; Liu, Z.; Wu, T. A deep learning interpretable model for river dissolved oxygen multi-step and interval prediction based on multi-source data fusion. *J. Hydrol.* **2024**, *629*, 130637. [CrossRef]
24. Xu, J.; Li, T.; Zhang, D.; Wu, J. Ensemble clustering via fusing global and local structure information. *Expert Syst. Appl.* **2024**, *237*, 121557. [CrossRef]
25. Su, H.; Du, Q.; Chen, G.; Du, P. Optimized Hyperspectral Band Selection Using Particle Swarm Optimization. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **2014**, *7*, 2659–2670. [CrossRef]
26. Medjahed, S.A.; Ait Saadi, T.; Benyettou, A.; Ouali, M. Gray Wolf Optimizer for Hyperspectral Band Selection. *Appl. Soft Comput.* **2016**, *40*, 178–186. [CrossRef]
27. Medjahed, S.A.; Ouali, M. A Hybrid Approach for Supervised Spectral Band Selection in Hyperspectral Images Classification. *Comp. y Sist.* **2020**, *24*, 213–219. [CrossRef]
28. El-kenawy, E.-S.M.; Khodadadi, N.; Mirjalili, S.; Abdelhamid, A.A.; Eid, M.M.; Ibrahim, A. Greylag Goose Optimization: Nature-Inspired Optimization Algorithm. *Expert Syst. Appl.* **2024**, *238*, 122147. [CrossRef]
29. Dehghani, M.; Montazeri, Z.; Trojovská, E.; Trojovský, P. Coati Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems. *Knowl.-Based Syst.* **2023**, *259*, 110011. [CrossRef]
30. Lian, J.; Hui, G.; Ma, L.; Zhu, T.; Wu, X.; Heidari, A.A.; Chen, Y.; Chen, H. Parrot Optimizer: Algorithm and Applications to Medical Problems. *Comput. Biol. Med.* **2024**, *172*, 108064. [CrossRef]
31. Abualigah, L.; Elaziz, M.A.; Sumari, P.; Geem, Z.W.; Gandomi, A.H. Reptile Search Algorithm (RSA): A Nature-Inspired Meta-Heuristic Optimizer. *Expert Syst. Appl.* **2022**, *191*, 116158. [CrossRef]
32. Deng, W.; Cai, X.; Wu, D.Q.; Song, Y.; Chen, H.; Ran, X.; Zhou, X.; Zhao, H. MOQEA/D: Multi-objective QEA with decomposition mechanism and excellent global search and its application. *IEEE Trans. Intell. Transp. Syst.* **2024**. [CrossRef]

33. Zhou, L.; Sun, Q.; Ding, S.; Han, S.; Wang, A. A machine-learning-based method for ship propulsion power prediction in ice. *J. Mar. Sci. Eng.* **2023**, *11*, 1381. [CrossRef]
34. Deng, W.; Li, K.; Zhao, H. A flight arrival time prediction method based on cluster clustering-based modular with deep neural network. *IEEE Trans. Intell. Transp. Syst.* **2023**. [CrossRef]
35. Dong, W.B.; Zhou, L.; Ding, S.F.; Wang, A.-M.; Cai, J.-Y. Two-staged method for ice channel identification based on image segmentation and corner point regression. *China Ocean Eng.* **2024**, *38*, 313–325. [CrossRef]
36. Zhao, H.; Wu, Y.; Deng, W. An Interpretable Dynamic Inference System Based on Fuzzy Broad Learning. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2527412. [CrossRef]
37. Tanabe, R.; Fukunaga, A.S. Improving the Search Performance of SHADE Using Linear Population Size Reduction. In Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, China, 6–11 July 2014; pp. 1658–1665.
38. Gurrola-Ramos, J.; Hernandez-Aguirre, A.; Dalmau-Cedeno, O. COLSHADE for Real-World Single-Objective Constrained Optimization Problems. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020; pp. 1–8.
39. Sallam, K.M.; Elsayed, S.M.; Chakrabortty, R.K.; Ryan, M.J. Improved Multi-Operator Differential Evolution Algorithm for Solving Unconstrained Problems. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020; pp. 1–8.
40. Flor-Sánchez, C.O.; Reséndiz-Flores, E.O.; Altamirano-Guerrero, G. Kernel-Based Gradient Evolution Optimization Method. *Inf. Sci.* **2022**, *602*, 313–332. [CrossRef]
41. Kumar, A.; Das, S.; Zelinka, I. A Self-Adaptive Spherical Search Algorithm for Real-World Constrained Optimization Problems. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion, Cancún, Mexico, 8–12 July 2020; pp. 13–14.
42. Naruei, I.; Keynia, F. Wild Horse Optimizer: A New Meta-Heuristic Algorithm for Solving Engineering Optimization Problems. *Eng. Comput.* **2022**, *38*, 3025–3056. [CrossRef]
43. Ramadan, A.; Kamel, S.; Taha, I.B.M.; Tostado-Véliz, M. Parameter Estimation of Modified Double-Diode and Triple-Diode Photovoltaic Models Based on Wild Horse Optimizer. *Electronics* **2021**, *10*, 2308. [CrossRef]
44. Milovanović, M.; Klimenta, D.; Panić, M.; Klimenta, J.; Perović, B. An Application of Wild Horse Optimizer to Multi-Objective Energy Management in a Micro-Grid. *Electr. Eng.* **2022**, *104*, 4521–4541. [CrossRef]
45. Saha, S.; Arya, R. Improved Hybrid Node Localization Using the Wild Horse Optimization in the Underwater Environment. *Int. J. Syst. Assur. Eng. Manag.* **2023**, *14*, 865–885. [CrossRef]
46. Ewees, A.A.; Ismail, F.H.; Ghoniem, R.M. Wild Horse Optimizer-Based Spiral Updating for Feature Selection. *IEEE Access* **2022**, *10*, 106258–106274. [CrossRef]
47. Zheng, R.; Hussien, A.G.; Jia, H.-M.; Abualigah, L.; Wang, S.; Wu, D. An Improved Wild Horse Optimizer for Solving Optimization Problems. *Mathematics* **2022**, *10*, 1311. [CrossRef]
48. Wolpert, D.H.; Macready, W.G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Computat.* **1997**, *1*, 67–82. [CrossRef]
49. Wang, J.; Ye, M.; Xiong, F.; Qian, Y. Cross-Scene Hyperspectral Feature Selection via Hybrid Whale Optimization Algorithm With Simulated Annealing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 2473–2483. [CrossRef]
50. Wang, M.; Wu, C.; Wang, L.; Xiang, D.; Huang, X. A Feature Selection Approach for Hyperspectral Image Based on Modified Ant Lion Optimizer. *Knowl.-Based Syst.* **2019**, *168*, 39–48. [CrossRef]
51. Wong, W.C.; Chung, C.Y.; Chan, K.W.; Chen, H. Quasi-Monte Carlo Based Probabilistic Small Signal Stability Analysis for Power Systems With Plug-In Electric Vehicle and Wind Power Integration. *IEEE Trans. Power Syst.* **2013**, *28*, 3335–3343. [CrossRef]
52. Hamzaçebi, C.; Kutay, F. Continuous Functions Minimization by Dynamic Random Search Technique. *Appl. Math. Model.* **2007**, *31*, 2189–2198. [CrossRef]
53. Deng, L.; Liu, S. A Multi-Strategy Improved Slime Mould Algorithm for Global Optimization and Engineering Design Problems. *Comput. Methods Appl. Mech. Eng.* **2023**, *404*, 115764. [CrossRef]
54. Zivkovic, T.; Nikolic, B.; Simic, V.; Pamucar, D.; Bacanin, N. Software Defects Prediction by Metaheuristics Tuned Extreme Gradient Boosting and Analysis Based on Shapley Additive Explanations. *Appl. Soft Comput.* **2023**, *146*, 110659. [CrossRef]
55. Foody, G.M. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogramm. Eng.* **2004**, *70*, 627–634. [CrossRef]