

Article

Unsupervised Hyperspectral Band Selection Using Spectral–Spatial Iterative Greedy Algorithm

Xin Yang  and Wenhong Wang * 

College of Computer Science, Liaocheng University, Liaocheng 252000, China; 2220170220@stu.lcu.edu.cn

* Correspondence: wangwenhong@lcu.edu.cn

Abstract

Hyperspectral band selection (BS) is an important technique to reduce data dimensionality for the classification applications of hyperspectral remote sensing images (HSIs). Recently, searching-based BS methods have received increasing attention for their ability to select the best subset of bands while preserving the essential information of the original data. However, existing searching-based BS methods neglect effective exploitation of the spatial and spectral prior information inherent in the data, thus limiting their performance. To address this problem, in this study, a novel unsupervised BS method called Spectral–Spatial Iterative Greedy Algorithm (SSIGA) is proposed. Specifically, to facilitate efficient local search using spectral information, SSIGA conducts clustering on all the bands by employing a K -means clustering method with balanced cluster size constraints and constructs a K -nearest neighbor graph for each cluster. Based on the nearest neighbor graphs, SSIGA can effectively explore the neighborhood solutions in local search. In addition, to efficiently evaluate the discriminability and information redundancy of the solution given by SSIGA using the spatial and spectral information of HSIs, we designed an effective objective function for SSIGA. The value of the objective function is derived by calculating the Fisher score for each band in the solution based on the results of the superpixel segmentation performed on the target HSI, as well as by computing the average information entropy and mutual information of the bands in the solution. Experimental results on three publicly available real HSI datasets demonstrate that the SSIG algorithm achieves superior performance compared to several state-of-the-art methods.



Received: 16 July 2025

Revised: 25 August 2025

Accepted: 2 September 2025

Published: 10 September 2025

Citation: Yang, X.; Wang, W. Unsupervised Hyperspectral Band Selection Using Spectral–Spatial Iterative Greedy Algorithm. *Sensors* **2025**, *25*, 5638. <https://doi.org/10.3390/s25185638>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: band selection; hyperspectral remote sensing images; iterative greedy algorithm; spectral–spatial information; local search

1. Introduction

Hyperspectral remote sensing is an advanced Earth observation technology. It uses a hyperspectral imaging spectrometer to simultaneously capture images of ground objects across dozens or even hundreds of spectral bands, thereby generating hyperspectral remote sensing images (HSIs) that contain rich spatial and spectral information about land cover. Recently, HSIs have been widely used in various fields such as target detection [1], land cover classification [2,3], environmental monitoring [4,5], and precision agriculture [6,7]. However, the large number of bands included in the HSI results in high data dimensionality. On the other hand, the high correlation between bands leads to a large amount of information redundancy [8–10]. This not only causes a computational burden [11,12] but also leads to the Hughes phenomenon in hyperspectral classification tasks [13,14]. An effective way to address this problem is to reduce the dimensionality of HSIs.

Hyperspectral dimensionality reduction primarily falls into two categories: feature extraction and band selection. In recent years, deep learning-based feature extraction has become a prominent approach. For instance, various convolutional neural network (CNN)-based and transformer-based networks have demonstrated powerful capabilities in learning discriminative spectral-spatial representations. Notable examples include the multiscale neighborhood attention transformer (MSNAT) [15], the 3D spectral-spatial Mamba (3DSS-Mamba) [16], and the Mamba-in-Mamba (MiM) framework [17]. These models leverage hierarchical convolutions, attention mechanisms, or state-space modeling to capture local and global dependencies, thereby achieving superior classification accuracy compared to traditional machine learning methods. The advantage of such feature extraction methods lies in their ability to automatically learn high-level semantic features from original data, reducing the reliance on manual feature engineering. However, they typically convert the original spectral bands into a latent feature space, which weakens the physical interpretability of the results. In addition, they usually require a large amount of labeled data and computational resources [17].

By contrast, band selection (BS) directly identifies a subset of informative spectral bands from the original HSI cube without altering their physical meanings [18–20]. This not only reduces data redundancy and computational cost, but also preserves intrinsic spectral characteristics, making the results more interpretable and physically meaningful for remote sensing applications [21]. Based on the availability of labeled samples during the training phase, band selection methods are generally categorized as supervised [22], semi-supervised [23], and unsupervised methods [24,25]. Given the complexities involved in accurately labeling HSIs, unsupervised methods are often preferred in practical applications. Unsupervised band selection methods can be broadly divided into ranking-based methods [26–28], clustering-based methods [29,30], and searching-based approaches [31–33].

Ranking-based methods assign ranks to spectral bands based on predefined criteria and subsequently select the top-ranked bands to constitute a band subset. Although this method is effective in identifying promising subsets, it tends to ignore correlations between bands, which may lead to redundancy between selected bands [34]. Clustering-based methods treat each spectral band as an individual data point and group spectral bands with similar characteristics into a cluster. Then, from each group, a representative spectral band is selected according to specific evaluation criteria to form the final subset of spectral bands [35,36]. This strategy effectively decreases the correlation between bands, thus reducing the redundancy of the selected bands. However, since the selected bands come from different clusters, clustering-based methods can overemphasize the individual importance of the bands while neglecting the overall performance of the selected band subset [37].

Searching-based methods commonly regard band selection as a discrete combinatorial optimization problem, given their effectiveness in exploring large solution spaces [38–41]. Recently, various meta-heuristic optimization algorithms, including the gray wolf optimization algorithm [42], artificial bee colony algorithm [43], and heterogeneous cuckoo search algorithm [44], have been applied to BS tasks. In [42], the authors proposed a band selection method based on the gray wolf optimization algorithm, utilizing the hunting behavior of wolves as a search mechanism to explore the optimal band subset. In addition, category separability was incorporated into the population initialization process, effectively mitigating the risk of converging to local optima. In [43], the band selection problem is transformed into a multi-task optimization framework where clustering techniques are combined with an artificial bee colony algorithm to identify multiple optimal band subsets of varying sizes. In [44], the researchers proposed an unsupervised BS method via the enhanced heterogeneous cuckoo search algorithm integrated with a matched filter. This method uses a mapping

method based on neighborhood band grouping to reduce the similarity between selected bands. Although existing searching-based BS methods have demonstrated better performance, these methods ignore the spatial–spectral properties in real HSIs, and thus their performance is limited. Therefore, it remains a challenge to use the spatial–spectral prior information in real HSIs to design effective algorithmic search strategies and objective functions that can guide searching-based BS methods to find band combinations with high discriminability and low information redundancy.

To address these problems, a novel unsupervised hyperspectral band selection method called Spectral–Spatial Iterative Greedy Algorithm (SSIGA) is proposed in this paper. SSIGA aims to improve the efficiency of band selection by incorporating clustering and superpixel segmentation techniques into IGA, thereby exploiting the spatial and spectral information of the data. Specifically, to facilitate efficient local search using spectral information, SSIGA employs a K -means clustering method with balanced cluster size constraints to conduct clustering on all the bands. Then, based on a group of K -nearest neighbor graphs (K -NNGs) constructed for each cluster, SSIGA can exploit efficient neighborhood solutions in local search. To efficiently evaluate the discriminability and redundancy of the information of the band subset given by the solution, an effective objective function for SSIGA is designed. To evaluate the value of the objective function, we not only calculate the Fisher score for each band in the solution based on the result of performing superpixel segmentation on the HSI data, but also compute the average information entropy and mutual information of the bands given by the solution. Experimental results on three publicly available real HSI datasets demonstrate that SSIGA achieves superior performance compared to several state-of-the-art methods. The contributions of this article include the following.

- (1) A novel unsupervised Spectral–Spatial Iterative Greedy Algorithm-based band selection method is proposed. To the best of our knowledge, this is the first time that the IG algorithm is used to solve the band selection problem.
- (2) By conducting clustering on all bands and constructing a K -NNG for each cluster, our proposed SSIGA enables efficient neighborhood solution construction, which helps to facilitate efficient local search using spectral information.
- (3) We designed an effective objective function that can evaluate the quality of the solution by calculating the average information entropy of all the bands in the solution and the average mutual information between the bands, as well as the Fisher score of each band. Experimental results show that SSIGA outperforms several state-of-the-art methods.

The rest of this paper is structured as follows: Section 2 details the proposed SSIGA method. Then, we present the experimental results and related analyses in Section 3. Finally, Section 4 concludes this study.

2. Method

2.1. Iterative Greedy Algorithm

IGA was first proposed for solving large ensemble coverage problems in 1995 [45]. As a simple and effective heuristic algorithm, IGA solves an optimization problem based on a single solution [46]. Essentially, IGA can be regarded as a special kind of local search method [47]. It solves the problem by continuously exploring better solutions in the neighborhood of the current solution and adopts some perturbation strategies to avoid local optimal solutions. Unlike traditional local search methods, IGA has two unique perturbation phases, namely the destruction and construction phases.

Specifically, through the destruction strategy candidate solutions are broken and then new solutions are reconstructed via construction heuristics. Through the perturbation stage, IGA improves the quality of the solution by searching the neighborhood of the

current solution to identify the local optimum. Finally, a specific acceptance criterion is employed to determine how the algorithm selects a new solution at each iteration. Due to its simplicity and minimal parameter requirements, IGA is both efficient and easy to implement. Currently, IGA is a widely used heuristic approach for solving combinatorial optimization problems.

2.2. Iterative Greedy Algorithm Based on Simulated Annealing

The original IGA relies on greedy search, accepting only solutions better than the current one at each iteration. This makes it prone to falling into local optima. In contrast, the probabilistic acceptance mechanism of simulated annealing [48] allows for the acceptance of worse solutions with a certain probability, thereby increasing the possibility of escaping local optima. To improve its global optimization performance, IGA has recently been combined with simulated annealing [49]. By introducing perturbations through simulated annealing after each local search stage, the search process of IGA is no longer confined to the current neighborhood, enabling IGA to explore a broader solution space and enhancing its global search capability.

Specifically, according to the simulated annealing algorithm, IGA accepts a new solution following specific rules [50]. If a new solution given by the local search of IGA is superior to the current one, it is accepted unconditionally. Conversely, when the new solution is inferior, its acceptance is probabilistic, determined by a combination of factors. In other words, the acceptance probability decreases with the difference between the objective value of the new solution and that of the current solution, as well as with the current temperature. As the temperature decreases over iterations, IGA becomes less likely to accept inferior solutions. Specifically, the acceptance probability P of a new solution in IGA can be mathematically expressed as follows [51]:

$$P = \exp\left(-\frac{\Delta F}{T}\right) \quad (1)$$

where ΔF denotes the difference between the new solution and the current solution in terms of the objective function value; T represents the current temperature in the simulated annealing process. As the algorithm progresses, the temperature T will gradually decrease, facilitating extensive exploration in the initial stages and fostering convergence towards optimal solutions in later stages. This approach significantly improves the algorithm's capacity to avoid local optima, thereby enhancing the prospects of identifying superior solutions. The basic framework of IGA based on simulated annealing is outlined in Algorithm 1 [52,53].

In this section, we provide a detailed description of the proposed SSIGA method. Figure 1 gives a schematic illustration of SSIGA. As shown in Figure 1, SSIGA first performs superpixel segmentation on the original hyperspectral data, utilizing spatial consistency to aggregate pixels into local regions while preserving spatial structural information. In addition, the three-dimensional hyperspectral data is unfolded into a two-dimensional matrix, where clustering analysis is conducted based on spectral similarity, and a set of K-nearest neighbor graphs is constructed based on the results of clustering analysis according to the similarity between bands. During the optimization process of SSIGA, an initial band subset is first selected based on the clustering results. Then, a destruct–reconstruct strategy is applied, where a portion of the bands is randomly removed from the current solution and a new band subset is reconstructed. Next, a problem-dependent local search is performed on the current band subset to further optimize the band selection scheme. For each candidate solution, the objective function value is calculated based on spectral discriminability, information entropy, and mutual information to evaluate the classification performance, information content, and redundancy of the selected bands. Finally, to

prevent the search from falling into local optima, SSIGA employs simulated annealing as the acceptance criterion, accepting or rejecting new solutions based on probability to achieve global optimization. The algorithm iteratively executes this optimization process and outputs the final optimal band subset once the stopping condition is met. The complete iterative optimization workflow is summarized in Algorithm 2.

Algorithm 1 Framework of IGA.

Input: The number of elements q that need to be removed from the current solution.

Output: The best solution π^* .

```

1:  $\pi_0 \leftarrow$  Initialization;
2:  $\pi \leftarrow$  Local Search( $\pi_0$ );
3:  $\pi^* \leftarrow \pi$ ;
4: while termination criterion is not met do
5:    $\pi' \leftarrow$  Destruction and construction( $\pi, q$ );
6:    $\pi' \leftarrow$  local search( $\pi'$ );
7:   if  $F(\pi') < F(\pi)$  then
8:      $\pi \leftarrow \pi'$ ;
9:     Update the optimal solution;
10:  else
11:    Randomly generate a number  $r$  from 0 to 1;
12:    if  $r < e^{-\frac{\Delta F}{T}}$  then
13:       $\pi \leftarrow \pi'$ ;
14:    end if
15:  end if
16:   $T \leftarrow 0.99 \times T$ ;
17: end while
18: return  $\pi^*$ 
  
```

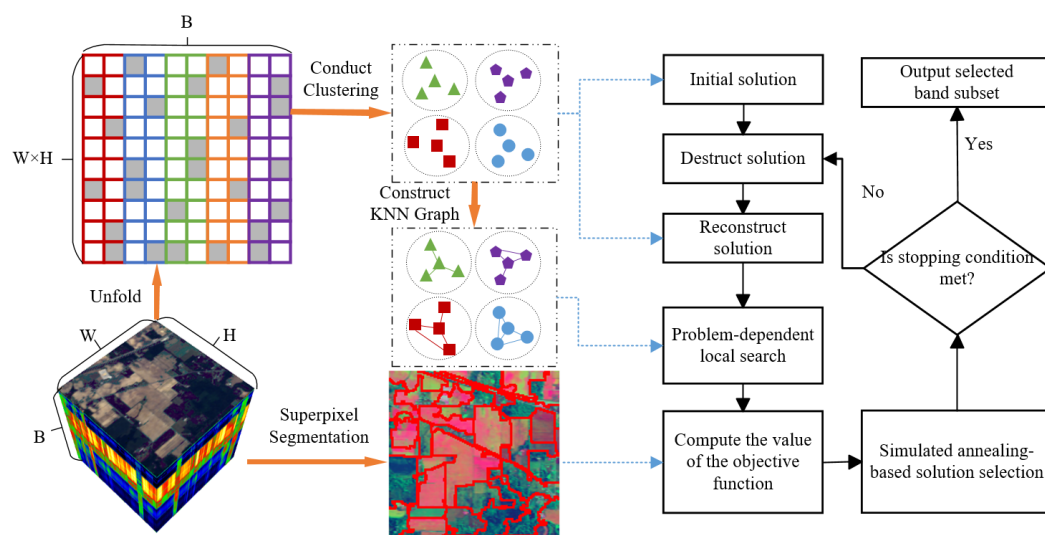


Figure 1. A schematic illustration of SSIGA.

Algorithm 2 The algorithm of the SSIGA method.

Input: Hyperspectral images dataset X
 Number of clusters M
 Number of neighbors K
 Number of superpixels N
 Simulated annealing temperature T
 Maximum number of iterations $Iter$

Output: Selected band subset Y

- 1: Perform clustering and construct K -NNGs;
- 2: Conduct superpixel segmentation on hyperspectral dataset;
- 3: Generate the initial solution $\mathbf{z}$ and calculate its objective function value $F(\mathbf{z})$ using Equation (11);
- 4: $\mathbf{z}^* \leftarrow \mathbf{z}$;
- 5: **while** $t < Iter$ **do**
- 6: Perform destruction reconstruction operations in III-B to generate the solution $\mathbf{z}$;
- 7: Perform local search operations in III-C to generate the solution $\mathbf{z}'$;
- 8: **if** $F(\mathbf{z}') > F(\mathbf{z})$ **then**
- 9: $\mathbf{z} \leftarrow \mathbf{z}'$;
- 10: Update the optimal solution $\mathbf{z}^*$;
- 11: **else**
- 12: Randomly generate a number r from 0 to 1;
- 13: **if** $r < e^{-\frac{\Delta F}{T}}$ **then**
- 14: $\mathbf{z} \leftarrow \mathbf{z}'$;
- 15: **end if**
- 16: **end if**
- 17: $T \leftarrow 0.99 \times T$;
- 18: **end while**
- 19: Generate band subset Y according to the solution $\mathbf{z}^*$;
- 20: **return** Y

2.3. Initialization of the Solution

In SSIGA, the solution is represented as a vector $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$, where z_m indicates an integer denoting the band index. For the initialization of the solution, the simplest way is random initialization, where M different bands are randomly selected from all the bands to form the initial solution. However, random initialization is prone to causing a heuristic algorithm to fall into local minima and can also delay its convergence. To construct a superior initial solution using the spectral information of the target HSI, we first perform clustering on all bands and then select a representative band from each cluster to form the initial solution of SSIGA.

Specifically, to effectively cluster all bands of the target HSI, commonly used clustering methods in the existing literature include subspace clustering [25], fuzzy clustering [54], density peak-based clustering [55], and deep clustering [56]. However, the number of samples contained in each of the clusters produced by these clustering methods can vary significantly. This is not conducive to constructing superior initial solutions by selecting representative bands from each cluster, and it is difficult to provide effective support for the subsequent local search strategy based on clustering results in this study. To make all clusters have a relatively balanced size, we use the K -means algorithm with a balanced cluster size constraint [57] to perform clustering on the target HSI. Based on the set of clusters obtained $C = \{C_1, C_2, \dots, C_m, \dots, C_M\}$, where C_m denotes the set of band indexes, from each group C_m we select a band closest to the centroid of C_m as a representative band and then use the index of the selected band as an element of the initial solution vector $\mathbf{z}$. Since each of the selected bands effectively captures the overall characteristics of the corresponding cluster, the resulting initial solution ensures a certain degree of superiority.

2.4. Destruction and Reconstruction Operator

As previously mentioned, the perturbation operator in IGA is used to destroy the current solution and generate a new one. This allows IGA to expand its exploration of the solution space, thus increasing its ability to get rid of locally optimal solutions and the probability of finding a globally optimal solution. In this study, the perturbation operator of SSIGA consists of two stages: destruction and reconstruction. Let $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$ denote the current solution, where M represents the length of the solution. Consider that in the context of band selection, the sequence of selected bands does not influence the band selection outcome. Hence, during the destruction phase, two bands are randomly removed from the current solution of SSIGA. In the reconstruction stage, we select another two bands to replace the bands that are deleted in the destruction stage. Specifically, assuming that z_m is deleted, we select a new band from the cluster where z_m is located following the Levy flight strategy. This can be expressed as

$$l^{\text{new}} = l^{\text{old}} + \text{Levy}(\lambda), \quad (2)$$

where operator $\text{Levy}(\lambda)$ is used to generate a step and λ denotes the stability index that can affect the value of the step [58]; l^{old} indicates the serial number of the removed z_m in the corresponding cluster, and l^{new} refers to the serial number of the selected band in the same cluster used to replace z_m in the reconstruction stage. To illustrate this process more clearly, we give an example, as shown in Figure 2. In Figure 2, let $z_m = 24$, indicating the removed band, while $l^{\text{old}} = 2$. To select a new band, we execute the Levy flight strategy according to Equation (2) in the cluster that band 24 belongs to (i.e., cluster C_1). In Equation (2), the operator $\text{Levy}(\lambda)$ generates a random step, which represents the position of the target band after moving from the current band according to the size of the step. For example, in Figure 2, if the random steps are 2, 3, and 4, respectively, the corresponding value of l^{new} equals 4, 5, and 1, respectively. It should be noted that if the value of l^{new} exceeds the number of bands in the current cluster, it is wrapped around to fall within the valid range using the modulo operation. Moreover, Figure 2 also illustrates that when l^{new} equals 2, 3, and 4, respectively, we choose bands 37, 41, and 23, respectively, to replace z_m in the reconstruction stage.

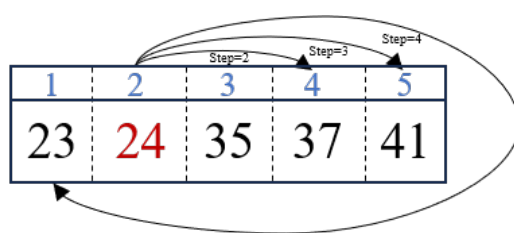


Figure 2. Illustration of the Levy flight strategy.

2.5. Spectral Information-Based Local Search

In heuristic optimization algorithms, a local search strategy can be adopted to efficiently locate improved solutions. In other words, the purpose of local search is to iteratively identify candidate solutions in the vicinity of the current solution to find a superior solution. To this end, we design a structured neighborhood-based local search strategy for SSIGA. This is achieved by first performing clustering on all bands and then constructing a K -NNG for each cluster. Specifically, consider the set of clusters $C = \{C_1, C_2, \dots, C_m, \dots, C_M\}$ given by the K -means with a balanced cluster size constraint. For each cluster C_m , we construct a K -NNG $G_m = (V_m, E_m)$, where $V_m = C_m$ and E_m , respectively, denote the vertex set and the edge set of G_m , according to the following procedure: for each band index x_i in V_m , we generate a group of edges for G_m by finding its K nearest bands within V_m according to

the Euclidean distances between two bands. The constructed set of K -NNGs can act as the structured neighborhoods used to guide the local search in SSIGA.

Given the current solution $\mathbf{z} = (z_1, z_2, \dots, z_m, \dots, z_M)$, the local search starts by randomly selecting q bands that will be replaced by better bands. Then, for each band z_m of the q selected bands, the local search iteratively finds its adjacent bands through a predefined K -NNG where z_m is located to form a new solution. To test the quality of the new solution, the objective function value corresponding to this solution is calculated according to the objective function given in Section 2.6. If the new solution is better than the current solution, the current solution is replaced with the new solution; otherwise, the current solution remains unchanged. To understand this process more intuitively, we give an example, shown in Figure 3. As shown in Figure 3, the current solution is represented as $\mathbf{z} = (24, 64, 100, 141)$. According to the local search process, band 24 is first removed from the current solution; subsequently, each of its neighboring bands is selected to replace band 24 according to the K -NNG G_1 in which band 24 is located, and the resulting new solution is evaluated. The evaluation is based on the objective function value of the new solution. Accordingly, band 25 is identified as the best replacement band. According to this replacing strategy, band 64 is also replaced by its neighboring band 66. This process results in an updated solution $\mathbf{z}' = (25, 66, 100, 141)$. It is worth noting that the proposed local search strategy has three major advantages: (1) it limits the size of the solution space to a more searchable scale; (2) it exploits the inherent spectral correlation between neighboring bands; and (3) it greatly enhances the local search capability of SSIGA, especially in terms of accelerating the convergence speed and improving the quality of the solution.

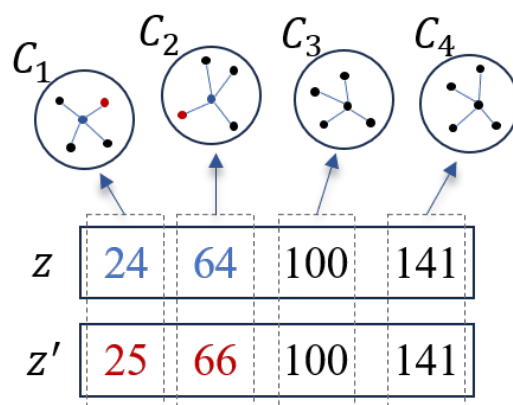


Figure 3. Illustration of the local search strategy.

2.6. Objective Function Based on Spatial–Spectral Prior

Although spectral information in hyperspectral images (such as band-to-band similarity) is often used for band selection, spatial dimension features are still severely neglected. To effectively exploit the spatial information in the data, our method utilizes superpixel segmentation to capture the spatial structural information of the data. Specifically, superpixel segmentation identifies spectrally homogeneous regions to effectively capture spatial information, including texture, edges, and geometric patterns. Based on the spatial regions obtained through superpixel segmentation, we can calculate the inter-class and intra-class scattering matrices within these regions and use Fisher scores to quantitatively analyze the discrimination ability of the selected bands. Aiming to maximize discriminability and information content while minimizing redundancy of the selected band subset, we design an objective function based on the Fisher discriminant score, information entropy, and mutual information. Specifically, to measure the discriminability of the band subset selected by the current solution, we compute its Fisher discriminant score in the objective function. To this end, we first use the entropy rate superpixel (ERS) method to perform superpixel

segmentation on the target HSI. Let $SP = \{sp_1, sp_2, \dots, sp_n, \dots, sp_N\}$ represent the obtained set of superpixels, where $sp_n = \{x_1^n, x_2^n, \dots, x_{u_n}^n\}$ and x_i^n represents the i -th pixel within the n -th superpixel. The Fisher discriminant score for the band subset selected by the current solution is expressed as

$$S_b = \sum_{i=1}^N u_i (w_i - w)(w_i - w)^T \quad (3)$$

$$S_w = \sum_{i=1}^N \sum_{j=1}^{u_i} (x_j^i - w_i)(x_j^i - w_i)^T \quad (4)$$

where u_i indicates the total number of pixels within the i -th superpixel; S_b denotes the between-class scatter matrix, measuring the dispersion of data points across different classes, and S_w represents the within-class scatter matrix, quantifying the dispersion of data points within the same class; and w represents the mean vector computed across all superpixels within the target HSI, whereas w_i signifies the mean vector specific to the i -th superpixel. In the context of classification tasks, a large S_b signifies effective separation between classes, whereas a small S_w implies tight clustering within classes. Thus, based on these metrics, the final discriminability criterion for the selected band subset is defined as

$$J = \frac{\text{trace}(S_b)}{\text{trace}(S_w)} \quad (5)$$

where $\text{trace}(\cdot)$ denotes the trace of the matrix. Note that the discriminability criterion J ensures that the selected band subset has high between-class scatter and low within-class scatter, effectively evaluating the discriminability of the candidate band subset.

In addition, we quantify the information content of the band subset selected by the current solution through information entropy, and we assess redundancy within this subset by utilizing mutual information. In other words, information entropy, which measures the average level of uncertainty or randomness in the outcomes of a random variable, is utilized here to quantify the amount of information in all bands within the selected subset. Moreover, mutual information, which quantifies the amount of information that one random variable contains about another, is used to assess redundancy and dependencies between bands within the selected subset. Specifically, the information entropy $H(x)$ for band x is calculated by [59]

$$H(x) = - \sum_{\alpha \in \Phi_x} p(\alpha) \log p(\alpha) \quad (6)$$

where α denotes the gray value of band x and Φ_x represents the set of gray levels for band x ; $p(\alpha)$ denotes the probability distribution of gray value α in band x .

Given two bands x_1 and x_2 , mutual information $MI(x_1; x_2)$ measures the shared information between them and is calculated by [60]

$$MI(x_1; x_2) = H(x_1) + H(x_2) - H(x_1, x_2) \quad (7)$$

with

$$H(x_1, x_2) = - \sum_{\alpha \in \Phi_{x_1}} \sum_{\beta \in \Phi_{x_2}} p(\alpha, \beta) \log p(\alpha, \beta) \quad (8)$$

where $H(x_1, x_2)$ denotes the joint entropy of bands x_1 and x_2 .

It should be noted that according to Equations (6) and (7), a high value of information entropy means that the uncertainty of the random variable is stronger, which also means that the amount of information contained in all bands within the selected band subset is higher. In addition, a lower mutual information value signifies a weaker dependency

between the variables, signifying a reduced redundancy in the selected band subset for hyperspectral classification. Therefore, maximizing the information entropy of the selected band subset while minimizing its mutual information ensures that all bands in the selected band subset have a higher level of information content and a low level of redundancy. To this end, we calculate the information entropy and mutual information of the selected band subset, respectively, by

$$H^* = \frac{1}{M} \sum_{i=1}^M H(x_i) \quad (9)$$

and

$$MI^* = \frac{2}{M^2 - M} \sum_{1 \leq i < j \leq M} MI(x_i; x_j). \quad (10)$$

Based on Equations (5)–(10), the final objective function of SSIGA is formulated as follows:

$$F = \theta \cdot J + \frac{H^*}{MI^* + \epsilon}, \quad (11)$$

where θ denotes the weight parameter, which is empirically set to 0.002 before training and held fixed throughout optimization, and ϵ is a numerical stability term. It is worth pointing out that by maximizing the objective function in Equation (11), SSIGA intends to identify a subset of bands with high discriminability, a higher level of information content, and low redundancy.

3. Experiments and Discussion

In this section, we perform experimental validation of the proposed method. Specifically, we first introduce the datasets used in the experiments, including the Indian Pines, Botswana, and Kennedy Space Center (KSC) datasets. Next, we describe the experimental setup, including the parameter settings of the algorithm and the classifiers used. Subsequently, we present existing state-of-the-art methods adopted for comparative evaluation. Finally, we carry out experiments to verify the effectiveness of our method, followed by a comprehensive analysis of the experimental results.

3.1. Datasets

In this section, three publicly available datasets, namely Indian, Botswana, and KSC, are used to test the performance of our method (available at https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, accessed on 1 September 2025).

3.1.1. Indian Pines

The Indian Pines dataset was collected by an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in Indiana, USA. The original image contains 224 spectral channels. Each band of this image, consisting of 145×145 pixels, exhibits a spatial resolution of 20 m per pixel, and the wavelengths of all bands range from 400 to 2500 nm.

Before experiments, rigorous preprocessing steps, including the removal of noisy and water absorption bands, reduced the number of spectral channels from 224 to 200 for subsequent analysis. The available ground truth data were comprehensively classified into 16 different land cover classes. The false-color composite image of this dataset is shown in Figure 4a.

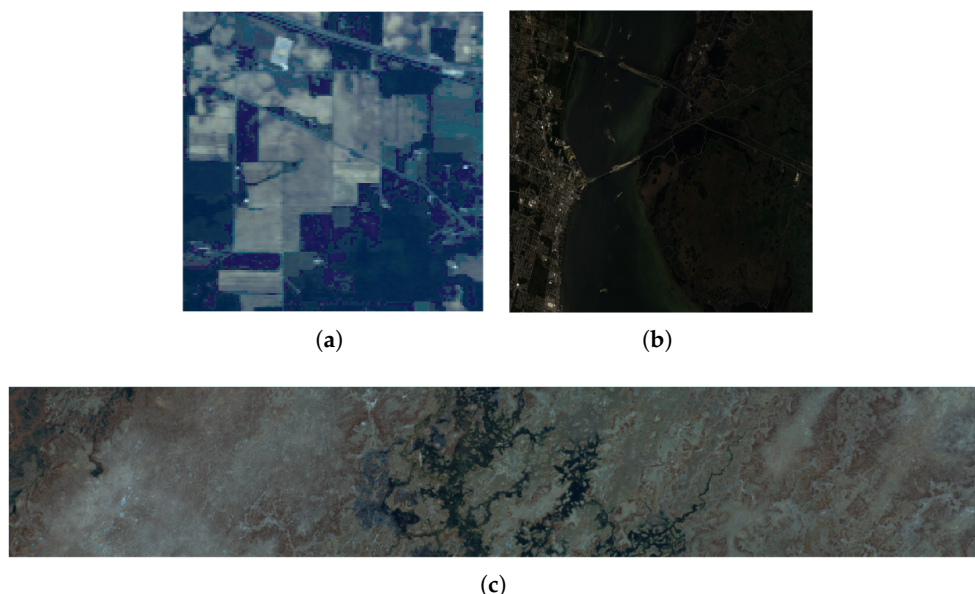


Figure 4. False-color composite images. (a) Indian Pines dataset. (b) KSC dataset. (c) Botswana dataset.

3.1.2. Kennedy Space Center

The Kennedy Space Center (KSC) dataset was captured by NASA in 1996 at the Kennedy Space Center, Florida, USA, using AVIRIS. This dataset comprises 13 land cover classes and originally included 224 spectral bands. Following the removal of water absorption bands, the number of bands used in subsequent experiments was reduced to 176. This dataset spans a wavelength range from 400 nm to 2500 nm, with each pixel representing a spatial resolution of 18 m. The false-color image of this dataset is shown in Figure 4b.

3.1.3. Botswana

The Botswana dataset was acquired in 2001 over the Okavango Delta region in Botswana by NASA's EO-1 satellite, using its hyperion imaging spectrometer. This dataset includes 14 land cover classes and 242 spectral bands collectively covering a wavelength range of 400 to 2500 nm. Each band of this image contains 1476×256 pixels. After removing noisy and water absorption bands, 145 bands were used in our experiments. The false-color image of this dataset is shown in Figure 4c.

3.2. Experiment Setup

To classify the HSI and assess the quality of the selected bands, we employed two classifiers: Support Vector Machine (SVM) and Random Forest (RF). Note that the supervised classifiers (SVM and RF) are only used for quantitative evaluation of the selected band subset. The SSIGA algorithm performs band selection in a completely unsupervised manner, with no access to class labels during selection. For both classifiers, we applied consistent parameter settings across different datasets. Specifically, the Random Forest (RF) classifier was configured with 20 trees, a commonly used setting for balancing performance and computational efficiency. For Support Vector Machine (SVM), the Radial Basis Function (RBF) kernel was employed, with the penalty parameter C set to 5000 and the kernel parameter γ set to 0.5. Since both classifiers operate under supervised learning, we randomly selected 10% of each dataset for training [33,61]; Tables 1–3 present the number of training and test samples per class for the Indian Pines, Botswana, and KSC datasets. We evaluated the classification performance using overall accuracy (OA), average overall accuracy (AOA) [25,38,61], and the kappa coefficient (Kappa), which are widely adopted metrics in hyperspectral image classification. To evaluate the performance of our method,

we compare it with several representative unsupervised band selection methods, including ASPS [25], DSC [54], E-FDPC [28], FNGBS [61], HLFC [62], MBBS-VC [41] and SNEA [33]. For fair comparison, all baseline methods utilize the hyperparameter configurations specified in their original works, while hyperparameters for our method are listed in Table 4. In addition, we also provide experimental results obtained by classifying HSIs using Support Vector Machine (SVM) and Random Forest (RF) classifiers across all original spectral bands (referred to as Baseline-SVM and Baseline-RF, respectively) for performance comparison. The following is a brief introduction to all comparison methods.

Table 1. Number of training and testing samples in the Indian Pines dataset.

Class	Name	Training Samples	Testing Samples
1	Alfalfa	5	41
2	Corn-notill	143	1285
3	Corn-mintill	83	747
4	Grass-pasture	24	213
5	Corn	49	434
6	Grass-trees	73	657
7	Grass-pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybeans-notills	98	874
11	Soybeans-mintills	246	2209
12	Soybeans-clean	60	533
13	Wheat	21	184
14	Woods	127	1138
15	Building-Grass-Tress-Drives	39	347
16	Stone-Steel-Tower	10	83

Table 2. Number of training and testing samples in the Botswana dataset.

Class	Name	Training Samples	Testing Samples
1	Water	27	243
2	Hippo grass	11	90
3	Fp-grassess1	26	225
4	Fp-grassess2	22	193
5	Reeds	27	242
6	Riparian	27	242
7	Firescar	26	233
8	Island interior	21	182
9	Acacia woodlands	32	282
10	Acacia shrublands	25	223
11	Acacia grasslands	31	274
12	Short mopane	19	162
13	Mixed mopane	27	241
14	Exposed soils	10	85

Table 3. Number of training and testing samples in the Kennedy Space Center dataset.

Class	Name	Training Samples	Testing Samples
1	Scrub	77	684
2	Willow swamp	25	218
3	CP hammock	26	230
4	CP/Oak	26	226
5	Slash pine	17	144
6	Oak/Broadleaf	23	206
7	Hardwood swamp	11	94
8	Graminoid marsh	44	387
9	Spartina marsh	52	468
10	Catail marsh	41	363
11	Salt marsh	42	377
12	Mud flats	51	452
13	Water	93	834

Table 4. Values of the hyperparameters.

Parameters	Values
Simulate annealing initial temperature, T	1000
Maximum number of iterations, $iter$	2000
Number of superpixels, N	300

3.2.1. ASPS

As a ranking-based BS method, ASPS achieves band selection by integrating subspace partitioning and band sorting techniques. Initially, the HSI cube is partitioned into multiple subcubes, with the objective of maximizing the ratio of inter-class to intra-class distances within each subcube. Subsequently, within each subcube, the band exhibiting the lowest noise level is selected as a representative band.

3.2.2. DSC

This method belongs to the category of clustering-based BS methods. It leverages deep subspace clustering by seamlessly integrating subspace clustering techniques into a convolutional autoencoder framework. This integration enhances clustering performance and enables efficient end-to-end training.

3.2.3. E-FDPC

This is a ranking-based cluster-driven hybrid band selection method. It calculates the score for each band through a weighted combination of two metrics: the normalized local density and a metric reflecting the compactness within clusters. These metrics are obtained using the density peak-based clustering algorithm.

3.2.4. FNGBS

As a hybrid method integrating clustering and ranking principles, FNGBS begins by dividing the HSI data cube into several spectral groups and then selects spectral bands based on local density and information entropy criteria.

3.2.5. HLFC

This is a clustering-based band selection method that first divides the target HSI into several regions using superpixel segmentation. After constructing a similarity graph for each superpixel to learn low-dimensional hidden features, all latent features are fused to form a unified feature representation. Finally, K -means is used to generate multiple

clusters, from which the representative spectral band with the highest information entropy is selected.

3.2.6. MBBS-VC

This is a searching-based BS method that models the BS problem as a multi-task optimization problem. It employs clustering that allows for clusters of different sizes and adopts a multi-task multi-swarm bee colony strategy, aiming to identify multiple optimal band subsets.

3.2.7. SNEA

This method belongs to the search-based BS methods. It defines the BS problem as a problem of maintaining spatial structure, which is achieved by designing two optimization objectives related to spatial structure. In addition, SNEA proposes a neighborhood grouping pairwise learning strategy for generating high-quality offspring. In this strategy, a neighborhood grouping operation is developed to divide the band space into multiple groups so as to efficiently initialize the population and generate pairwise offspring solutions guided by the grouping.

3.3. Experimental Results and Discussion

3.3.1. Comparison with State-of-the-Art Methods

Table 5 shows the AOA and Kappa values achieved by SSIGA and seven other comparative methods using RF and SVM classifiers on Indian Pines, Botswana, and KSC datasets, where the red bold font indicates the best classification results achieved. Figures 5–7 then show the OA and Kappa values obtained by all methods under different numbers of selected bands. Specifically, as shown in Table 5, the AOA values achieved by our proposed SSIGA on all three experimental datasets are higher than those achieved by the other band selection methods. In addition, SSIGA showed high consistency across all three datasets, which reflects the robustness of our method.

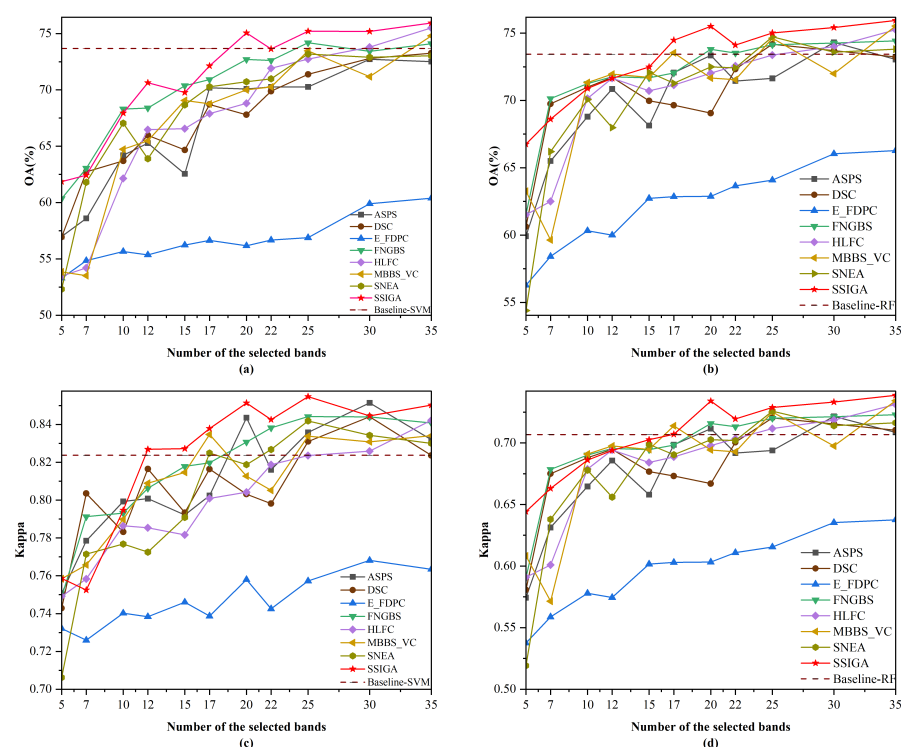


Figure 5. OA and Kappa versus the number of selected bands by different methods on the Indian Pines dataset. (a) OA obtained using the SVM classifier. (b) OA obtained using the RF classifier. (c) Kappa obtained using the SVM classifier. (d) Kappa obtained using the RF classifier.

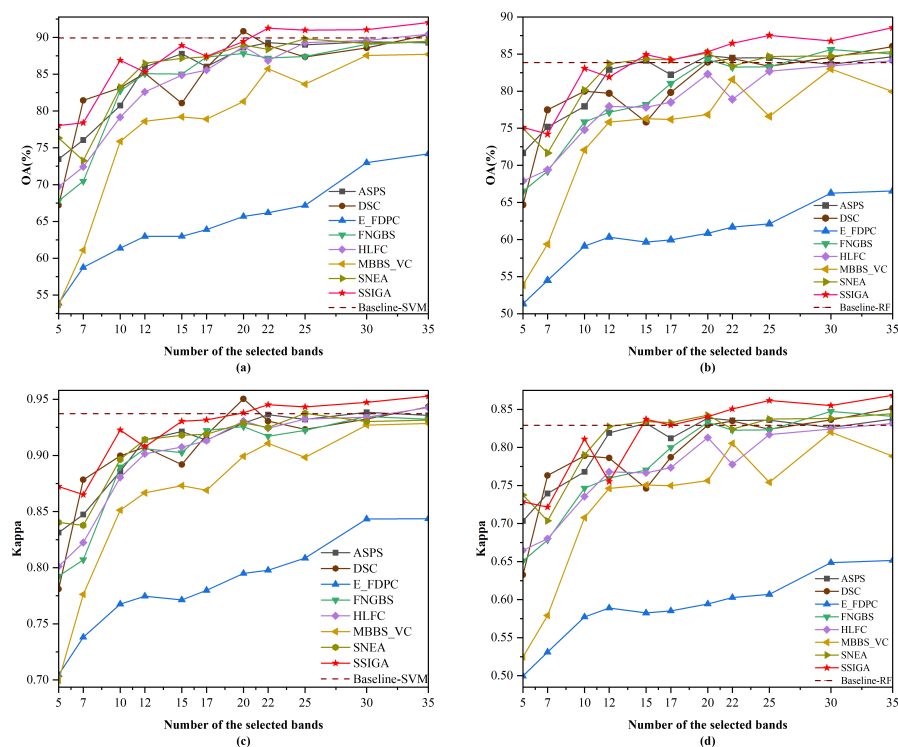


Figure 6. OA and Kappa versus number of selected bands by different methods on the Botswana dataset. (a) OA obtained using the SVM classifier. (b) OA obtained using the RF classifier. (c) Kappa obtained using the SVM classifier. (d) Kappa obtained using the RF classifier.

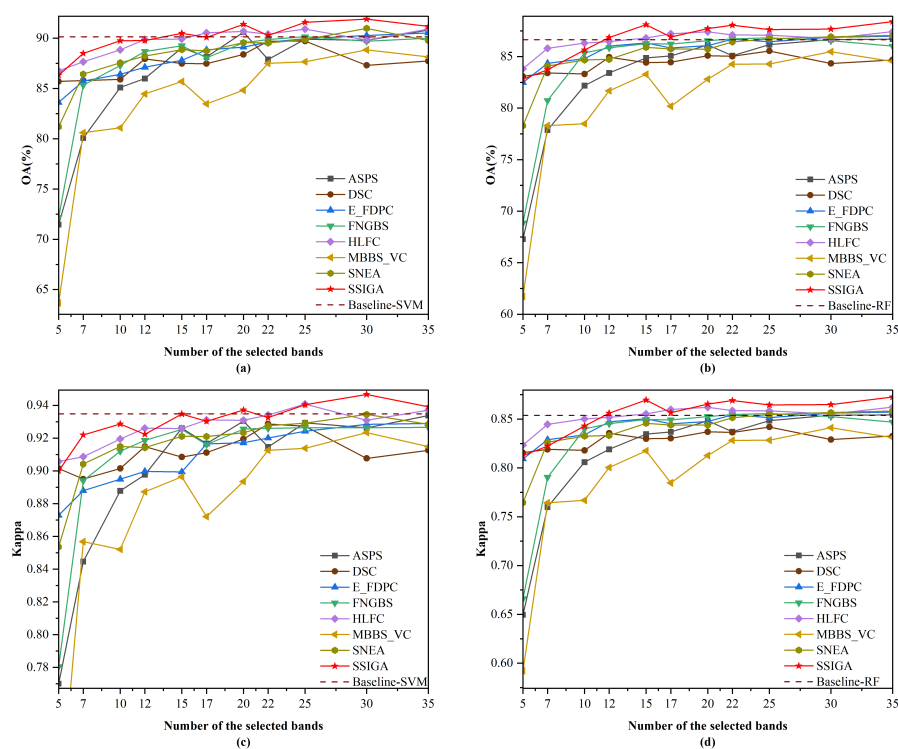


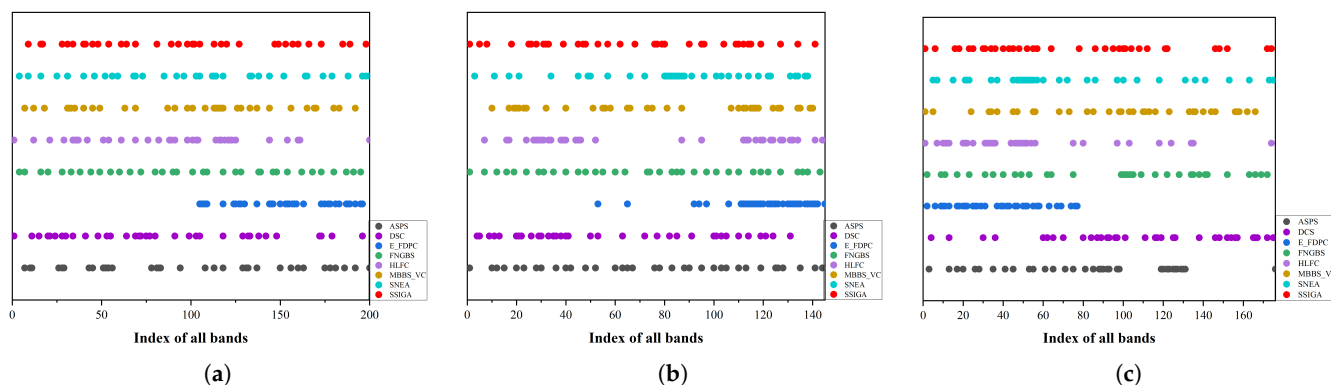
Figure 7. OA and Kappa versus the number of selected bands by different methods on the KSC dataset. (a) OA obtained using the SVM classifier. (b) OA obtained using the RF classifier. (c) Kappa obtained using the SVM classifier. (d) Kappa obtained using the RF classifier.

Table 5. Results obtained by the proposed SSIGA and other methods using RF and SVM classifiers on Indian Pines, Botswana, and KSC datasets (%).

Dataset	Classifier	SSIGA	ASPS [25]	DSC [54]	E-FDPC [28]	FNGBS [61]	HLFC [62]	MBBS-VC [41]	SNEA [33]
Indian Pines	AOA(RF)	72.80	69.91	70.47	62.13	71.68	70.44	70.61	69.91
	Kappa(RF)	70.46	67.63	68.19	59.59	69.45	68.18	68.36	67.64
	AOA(SVM)	70.88	66.69	67.07	56.54	69.85	66.68	66.82	67.70
	Kappa(SVM)	82.19	80.93	80.50	74.64	81.59	79.78	80.80	79.94
Botswana	AOA(RF)	83.45	81.45	79.97	60.21	79.02	77.98	73.78	82.02
	Kappa(RF)	81.44	80.43	78.91	58.81	77.93	76.83	72.57	81.02
	AOA(SVM)	87.25	85.05	84.53	64.55	83.57	83.57	77.57	85.42
	Kappa(SVM)	92.32	90.76	90.50	78.40	89.56	89.91	86.36	90.71
KSC	AOA(RF)	86.68	82.85	84.39	85.71	84.18	86.56	80.56	85.11
	Kappa(RF)	85.42	81.34	82.95	84.37	82.75	85.29	78.79	83.73
	AOA(SVM)	90.11	86.25	87.55	88.09	87.27	89.66	83.26	88.24
	Kappa(SVM)	93.03	89.79	91.16	90.81	90.70	92.64	87.33	91.56

Red bold font indicates the best results.

Figure 8 shows the distribution of the selected band subset by SSIGA and the other seven band selection methods on the three hyperspectral datasets. From Figure 8, it can be seen that the distribution of selected bands using different methods is relatively consistent across different datasets. The distribution of selected bands by SSIGA on Indian Pines, Botswana, and KSC datasets is more uniform in all band ranges, and the other search-based methods such as SNEA also perform well, but the distribution of selected bands for E-FDPC and HLFC is slightly concentrated. From the experimental results, it can be seen that the distribution of the selected band subsets by SSIGA on the three datasets is more uniform, indicating that the selected bands have lower redundancy and more information content.

**Figure 8.** Distribution of the selected subset of bands obtained by the ASPS, DSC, E-FDPC, FNGBS, HLFC, MVPCA, MBBS-VC, and SSIGA methods. (a) Indian Pines. (b) Botswana. (c) KSC.

For the Indian Pines dataset, as shown in Figure 5, SSIGA shows the best performance in most cases. Specifically, it can be seen from Figure 5a that when 5, 12, 17, 20, 22, 25, 30, and 35 bands are selected using the SVM classifier, SSIGA performs best according to the OA values. In the case that 7, 10, and 15 bands are selected, SSIGA performs slightly worse than FNGBS, but it outperforms ASPS, DSC, E-FDPC, HLFC, MBBS-VC, and SNEA. When selecting 25, 30, and 35 bands, SSIGA outperforms Baseline-SVM, further confirming that SSIGA can select a discriminative band subset. According to Figure 5b, SSIGA performs worse than DSC and FNGBS under the condition that 7 and 10 bands are used. However, when selecting 5, 15, and 35 bands with the RF classifier, SSIGA has the best OA values. In addition, SSIGA outperforms Baseline-RF at 20 to 35 bands. As can be seen from Figure 5c, at 12 to 25 and 35 bands, SSIGA gives the best Kappa values, although it shows relatively

poor performance when selecting 7 bands. In addition, we can see from Figure 5d that when 5, 15, 20, and 35 bands are used, SSIGA has the best performance according to the Kappa values. Moreover, when selecting 20 to 35 bands, SSIGA outperforms the Baseline-RF method. As a whole, SSIGA achieves the best classification accuracy in most cases on the Indian Pines dataset with both classifiers.

For the Botswana dataset, as shown in Figure 6, SSIGA demonstrates the best performance in most cases. Specifically, as shown in Figure 6a, when using the SVM classifier with 5, 10, 15, 22, 25, 30, and 35 bands selected, SSIGA achieves the best performance according to the OA values. In particular, when selecting 22, 25, 30, and 35 bands, the performance of SSIGA significantly exceeds that of Baseline-SVM. In the case of selecting 7 and 20 bands, the performance of SSIGA is inferior to that of DSC and Baseline-SVM but superior to that of other methods. According to Figure 6b, when using the RF classifier, SSIGA outperformed all of the comparison methods when selecting 15, 22, 25, 30, and 35 bands. At 17 and 20 bands, SSIGA performs similarly to SNEA and outperforms other methods. It is worth noting that when selecting 15 to 35 bands, the performance of SSIGA exceeds that of the Baseline-RF method. As can be seen from Figure 6c, SSIGA obtains the best Kappa values when selecting 22 to 35 bands using the SVM classifier. At 5, 10, 15, and 17 bands, the performance of SSIGA was lower than that of the Baseline-SVM method but superior to that of other methods. Moreover, from Figure 6d, it can be observed that when 22, 25, 30, and 35 bands are selected, SSIGA significantly outperforms other methods on the RF classifier according to the Kappa values. When selecting 15 and 20 bands, the performance of SSIGA is comparable to that of the SNEA and ASPS methods but superior to other methods. In general, on the Botswana dataset, SSIGA achieves the best classification accuracy in most cases with both classifiers.

For the KSC dataset, Figure 7 shows the performance of SSIGA and all methods with the SVM and RF classifiers, respectively. Specifically, we can see from Figure 7a that when selecting 15, 20, 25, 30, and 35 bands, the OA values of SSIGA are higher than those of other methods. Although SSIGA's performance is lower than Baseline-SVM when using 7 and 10 bands, its performance exceeds that of other methods. At 22 bands, the OA value of SSIGA is comparable to that of HLFC and Baseline-SVM but higher than that of other methods. As shown in Figure 7b, when using the RF classifier to select 12, 15, 20, 22, 25, 30, and 35 bands, SSIGA demonstrates the best performance according to the OA values. At 17 bands, SSIGA performs slightly inferiorly to HLFC yet outperforms the other methods. According to Figure 7c, when 20, 30, and 35 bands are selected, SSIGA achieves the best Kappa values on the SVM classifier. When selecting 7 and 12 bands, SSIGA's Kappa values are lower than those of Baseline-SVM but still superior to the other methods. At 15 bands, the OA value of SSIGA is comparable to that of Baseline-SVM but higher than that of other methods. When using 17 and 22 bands, the performance of SSIGA is comparable to HLFC and slightly lower than Baseline-SVM, but its performance is superior to other methods. Furthermore, as shown in Figure 7d, when selecting 12, 15, and 20 to 35 bands, SSIGA achieves the best performance on the RF classifier according to the Kappa values. At 17 bands, SSIGA performs slightly worse than HLFC but still outperforms the other methods. When 10 bands are chosen, the performance of SSIGA is inferior to that of HLFC and Baseline-RF, yet it still outperforms the other methods. Overall, on the KSC dataset, SSIGA achieves the best performance in most cases with both classifiers.

In addition, to visualize the quality of the bands selected by SSIGA, Figures 9–11 give the classification maps obtained by the RF and SVM classifiers when 30 bands are selected using SSIGA on the Indian Pines, Botswana, and KSC datasets, respectively. By comparing the ground truth maps with the corresponding classification maps provided by RF and SVM classifiers, as shown in Figures 9–11, we can see that SSIGA exhibits satisfactory

results under the condition of removing 85%, 79%, and 83% of the bands from the Indian Pines, Botswana, and KSC datasets, respectively.

3.3.2. Ablation Study

To investigate the individual impact of the local search module and the Fisher score component within the objective function, experiments with three configurations were carried out on the three datasets. Three ablation experimental settings were used in our experiment: SSIGA without local search, SSIGA without Fisher score, and complete SSIGA. Specifically, in the first setting, the local search module in SSIGA was disabled, the band neighborhood structure was removed, and only destruction and reconstruction were applied to the solution during algorithm iteration. In the second setting, the Fisher score was removed from the objective function of SSIGA, while the information entropy and mutual information components were retained. In the third setting, the complete SSIGA was used.

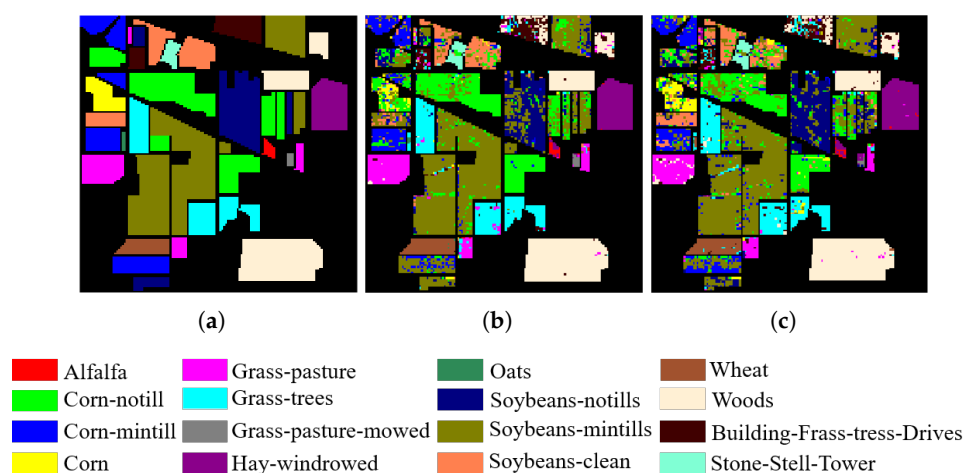


Figure 9. Ground truth map and classification maps of SSIGA on the Indian Pines dataset. (a) Ground truth. (b) Using SVM. (c) Using RF.

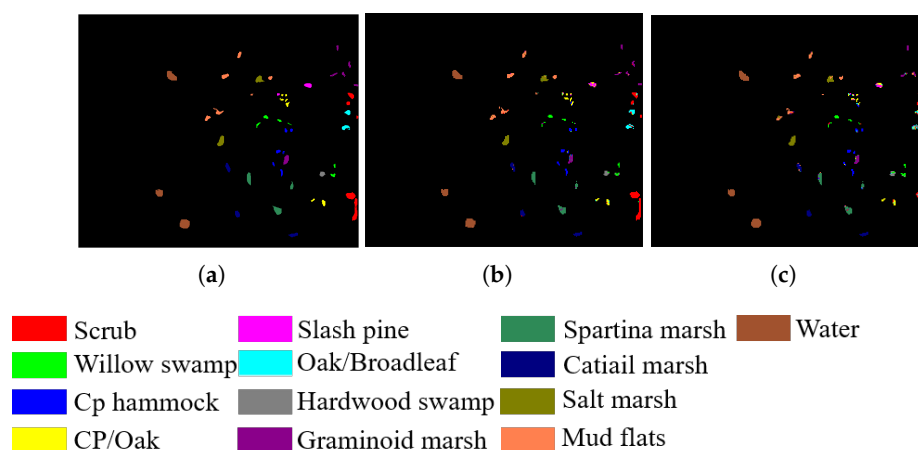


Figure 10. Ground truth map and classification maps of SSIGA on the KSC dataset. (a) Ground truth. (b) Using SVM. (c) Using RF.

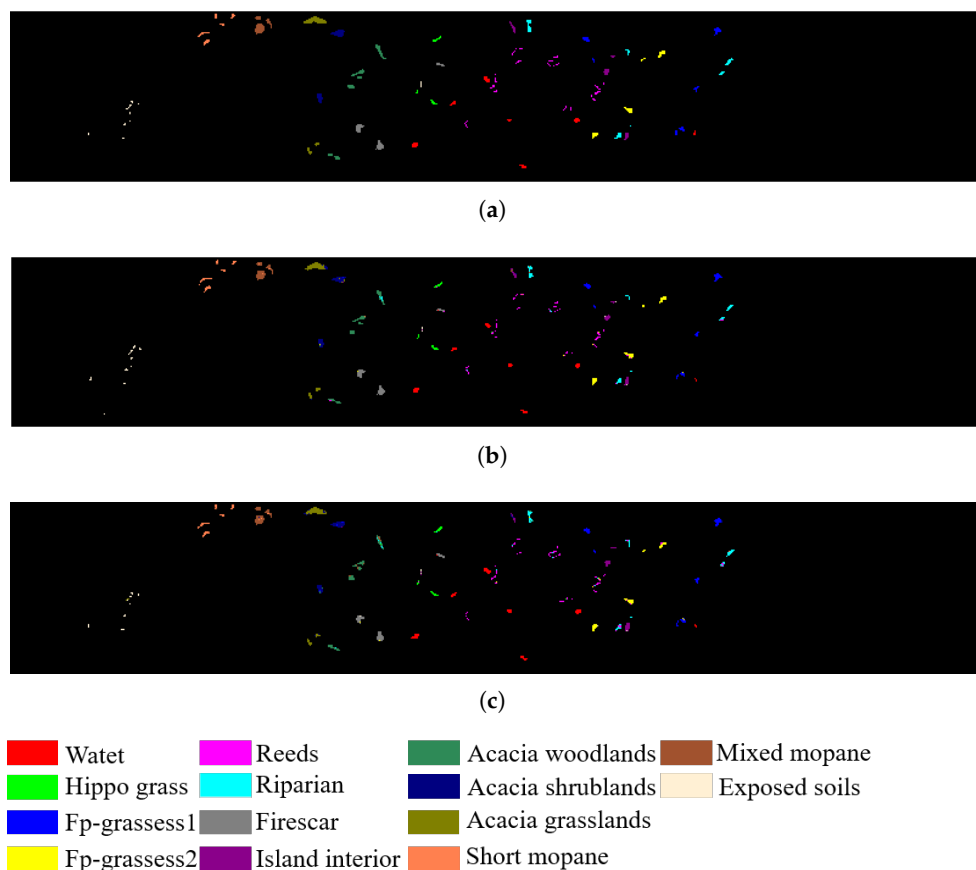


Figure 11. Ground truth map and classification maps of SSIGA on the Botswana dataset. (a) Ground truth. (b) Using SVM. (c) Using RF.

The detailed results are presented in Table 6, where the red bold font indicates the best classification results achieved and the symbol ✓ indicates that the corresponding module is enabled. As shown in Table 6, disabling either the local search module or the Fisher score component degrades the classification performance of SSIGA to varying degrees. Specifically, the absence of local search has a more pronounced impact, likely because the band neighborhood-based local search enhances the optimization capability of SSIGA, whereas destruction and reconstruction alone cannot explore the solution space effectively. Meanwhile, the removal of the Fisher score also led to a performance drop, demonstrating its effectiveness in capturing and utilizing spectral information, even though the remaining components (information entropy and mutual information) still provide a guiding function for the search.

Table 6. Ablation study on the “local search” and “Fisher score” components.

Dataset	Local Search	Fisher score	AOA	Kappa
Indian Pines		✓	64.93	76.98
	✓		68.33	79.27
	✓	✓	70.88	82.19
Botswana		✓	83.06	89.97
	✓		85.41	90.15
	✓	✓	87.25	92.32
KSC		✓	86.39	90.01
	✓		89.08	92.29
	✓	✓	90.11	93.03

Red bold font indicates the best results.

3.3.3. Execution Time

Table 7 reports the executing time of our proposed SSIGA method and other comparative methods, with the best results highlighted in red bold font. From Table 7, we can see that SSIGA runs longer than ASPS, DSC, E-FDPC, FNGBS, and SNEA on the Indian pine dataset but shorter than HLFC and MBBS-VC. On the Botswana and KSC datasets, SSIGA took more time than ASPS, E-FDPC, FNGBS, HLFC, and SNEA but less time than DSC and MBBS-VC. Overall, our proposed SSIGA requires considerable computation time to iteratively search for the optimal band subset. Nevertheless, it can select band subsets that yield significantly more accurate classification performance compared to other methods.

Table 7. The executing time (in seconds) spent by SSIGA and other methods when selecting 10 bands on the Indian Pines, Botswana, and KSC datasets.

Dataset	ASPS	DSC	E-FDPC	FNGBS	HLFC	MBBS-VC	SNEA	SSIGA
Indian Pines	0.11	14.27	0.29	0.05	55.44	62.90	4.80	30.58
Botswana	1.22	206.07	0.58	0.45	41.51	1224.76	5.72	138.93
KSC	1.24	208.97	0.62	0.47	52.56	1257.03	5.88	155.70

Red bold font indicates the best results.

4. Conclusions

In this paper, we have proposed a novel unsupervised BS method, termed Spectral–Spatial Iterative Greedy Algorithm-based band selection (SSIGA). Our method introduces clustering and superpixel segmentation techniques into the iterative greedy algorithm, thereby enhancing the efficiency and effectiveness of band selection by leveraging both spectral and spatial characteristics of the data. By employing a *K*-means clustering method with balanced cluster size constraints, we have facilitated efficient local search using spectral information. The construction of *K*-nearest neighbor graphs for each cluster has further enabled the generation of high-quality neighborhood solutions. In addition, we have designed an effective objective function that evaluates the discriminability and information redundancy of the selected bands. This function leverages superpixel segmentation to calculate the Fisher score for each band, as well as the average information entropy and mutual information among the bands, thus providing a comprehensive assessment of the solution quality. Experimental results on three publicly available real HSI datasets have demonstrated the superiority of the SSIG method compared to several state-of-the-art methods.

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

Funding: This research was supported by Discipline with Strong Characteristics of Liaocheng University—Intelligent Science and Technology under Grant 319462208.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

BS	Band Selection
HSIs	Hyperspectral remote sensing images
SSIGA	Spectral–Spatial Iterative Greedy Algorithm
K-NNs	K-Nearest Neighbor Graphs
IGA	Iterative Greedy Algorithm
ERS	Entropy Rate Superpixel
KSC	Kennedy Space Center
ASPS	Adaptive Subspace Partition Strategy
DSC	Deep Subspace Clustering
E-FDPC	Enhanced Fast Density-Peak-based Clustering
FNGBS	Fast Neighborhood Grouping for Band Selection
HLFC	Hierarchical Latent Feature Clustering
MBBS-VC	Multi-task Bee Band Selection With Variable-Size Clustering
SNEA	Structure-Conserved Neighborhood-Grouped Evolutionary Algorithm
SVM	Support Vector Machine
RF	Random Forest
RBF	Radial Basis Function
OA	Overall Accuracy
AOA	Average Overall Accuracy

References

- Shang, X.; Song, M.; Wang, Y.; Yu, C.; Yu, H.; Li, F.; Chang, C.-I. Target-Constrained Interference-Minimized Band Selection for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6044–6064. [\[CrossRef\]](#)
- Chang, C.-I.; Kuo, Y.-M.; Hu, P.F. Unsupervised Rate Distortion Function-Based Band Subset Selection for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [\[CrossRef\]](#)
- Launeau, P.; Kassouk, Z.; Debaine, F.; Roy, R.; Mestayer, P.G.; Boulet, C.; Rouaud, J.-M.; Giraud, M. Airborne Hyperspectral Mapping of Trees in an Urban Area. *Int. J. Remote Sens.* **2017**, *38*, 1277–1311. [\[CrossRef\]](#)
- An, D.; Zhao, G.; Chang, C.; Wang, Z.; Li, P.; Zhang, T.; Jia, J. Hyperspectral Field Estimation and Remote-Sensing Inversion of Salt Content in Coastal Saline Soils of the Yellow River Delta. *Int. J. Remote Sens.* **2016**, *37*, 455–470. [\[CrossRef\]](#)
- Tan, K.; Zhu, L.; Wang, X. A Hyperspectral Feature Selection Method for Soil Organic Matter Estimation Based on an Improved Weighted Marine Predators Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 1–11. [\[CrossRef\]](#)
- Saputro, A.H.; Handayani, W. Wavelength Selection in Hyperspectral Imaging for Prediction Banana Fruit Quality. In Proceedings of the 2017 International Conference on Electrical Engineering and Informatics (ICELEICs), Banda Aceh, Indonesia, 18–20 October 2017; IEEE: Piscataway, NJ, USA, 2018; pp. 226–230.
- Cheng, E.; Wang, F.; Peng, D.; Zhang, B.; Zhao, B.; Zhang, W.; Hu, J.; Lou, Z.; Yang, S.; Zhang, H.; et al. A GT-LSTM Spatio-Temporal Approach for Winter Wheat Yield Prediction: From the Field Scale to County Scale. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–18. [\[CrossRef\]](#)
- Dong, Y.; Du, B.; Zhang, L.; Zhang, L. Dimensionality Reduction and Classification of Hyperspectral Images Using Ensemble Discriminative Local Metric Learning. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2509–2524. [\[CrossRef\]](#)
- Sun, W.; Tian, L.; Xu, Y.; Zhang, D.; Du, Q. Fast and Robust Self-Representation Method for Hyperspectral Band Selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5087–5098. [\[CrossRef\]](#)
- Zhou, Y.; Peng, J.; Chen, C.L.P. Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095. [\[CrossRef\]](#)
- Li, B.; Zhang, P.; Zhang, J.; Jing, L. Unsupervised Double Weight Graphs Based Discriminant Analysis for Dimensionality Reduction. *Int. J. Remote Sens.* **2020**, *41*, 2209–2238. [\[CrossRef\]](#)
- Paul, A.; Chaki, N. Dimensionality Reduction of Hyperspectral Images Using Pooling. *Pattern Recognit. Image Anal.* **2019**, *29*, 72–78. [\[CrossRef\]](#)
- Li, D.; Shen, Y.; Kong, F.; Liu, J.; Wang, Q. Spectral–Spatial Prototype Learning-Based Nearest Neighbor Classifier for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [\[CrossRef\]](#)
- Zhao, E.; Qu, N.; Wang, Y.; Gao, C.; Duan, S.-B.; Zeng, J.; Zhang, Q. Thermal Infrared Hyperspectral Band Selection via Graph Neural Network for Land Surface Temperature Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [\[CrossRef\]](#)

15. Qiao, X.; Roy, S.K.; Huang, W. Multiscale Neighborhood Attention Transformer With Optimized Spatial Pattern for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [\[CrossRef\]](#)
16. He, Y.; Tu, B.; Liu, B.; Li, J.; Plaza, A. 3DSS-Mamba: 3D-Spectral-Spatial Mamba for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–16. [\[CrossRef\]](#)
17. Zhou, W.; Kamata, S.; Wang, H.; Wong, M.S.; Hou, H. Mamba-in-Mamba: Centralized Mamba-Cross-Scan in Tokenized Mamba Model for Hyperspectral Image Classification. *Neurocomputing* **2025**, *613*, 128751. [\[CrossRef\]](#)
18. Sellami, A.; Farah, M.; Farah, I.R.; Solaiman, B. Hyperspectral Imagery Semantic Interpretation Based on Adaptive Constrained Band Selection and Knowledge Extraction Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1337–1347. [\[CrossRef\]](#)
19. Sun, W.; Peng, J.; Yang, G.; Du, Q. Correntropy-Based Sparse Spectral Clustering for Hyperspectral Band Selection. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 484–488. [\[CrossRef\]](#)
20. Sun, W.; Zhang, L.; Zhang, L.; Lai, Y.M. A Dissimilarity-Weighted Sparse Self-Representation Method for Band Selection in Hyperspectral Imagery Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4374–4388. [\[CrossRef\]](#)
21. Serpico, S.B.; Bruzzone, L. A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1360–1367. [\[CrossRef\]](#)
22. Taşkın, G.; Kaya, H.; Bruzzone, L. Feature Selection Based on High Dimensional Model Representation for Hyperspectral Images. *IEEE Trans. Image Process.* **2017**, *26*, 2918–2928. [\[CrossRef\]](#)
23. Cao, X.; Wei, C.; Ge, Y.; Feng, J.; Zhao, J.; Jiao, L. Semi-Supervised Hyperspectral Band Selection Based on Dynamic Classifier Selection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1289–1298. [\[CrossRef\]](#)
24. Yuan, Y.; Zheng, X.; Lu, X. Discovering Diverse Subset for Unsupervised Hyperspectral Band Selection. *IEEE Trans. Image Process.* **2017**, *26*, 51–64. [\[CrossRef\]](#) [\[PubMed\]](#)
25. 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. [\[CrossRef\]](#)
26. Chang, C.-I.; Du, Q.; Sun, T.-L.; Althouse, M.L.G. A Joint Band Prioritization and Band-Decorrelation Approach to Band Selection for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2631–2641. [\[CrossRef\]](#)
27. Chang, C.-I.; Wang, S. Constrained Band Selection for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1575–1585. [\[CrossRef\]](#)
28. Jia, S.; Tang, G.; Zhu, J.; Li, Q. A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 88–102. [\[CrossRef\]](#)
29. Shahwani, H.; Bui, T.D.; Jeong, J.P.; Shin, J. A Stable Clustering Algorithm Based on Affinity Propagation for VANETs. In Proceedings of the 2017 19th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, Republic of Korea, 19–22 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 501–504.
30. Martinez-Uso, A.; Pla, F.; Sotoca, J.M.; García-Sevilla, P. Clustering-Based Hyperspectral Band Selection Using Information Measures. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4158–4171. [\[CrossRef\]](#)
31. Su, H.; Du, Q.; Chen, G.; Du, P. Optimized Hyperspectral Band Selection Using Particle Swarm Optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2659–2670. [\[CrossRef\]](#)
32. Yang, H.; Du, Q.; Chen, G. Particle Swarm Optimization-Based Hyperspectral Dimensionality Reduction for Urban Land Cover Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 544–554. [\[CrossRef\]](#)
33. Wang, Q.; Song, C.; Dong, Y.; Cheng, F.; Tong, L.; Du, B.; Zhang, X. Unsupervised Hyperspectral Band Selection via Structure-Conserved and Neighborhood-Grouped Evolutionary Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–19. [\[CrossRef\]](#)
34. Li, S.; Peng, B.; Fang, L.; Zhang, Q.; Cheng, L.; Li, Q. Hyperspectral Band Selection via Difference Between Intergroups. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–10. [\[CrossRef\]](#)
35. Wang, Q.; Zhang, F.; Li, X. Optimal Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *16*, 5522319. [\[CrossRef\]](#)
36. Yuan, Y.; Lin, J.; Wang, Q. Dual-Clustering-Based Hyperspectral Band Selection by Contextual Analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [\[CrossRef\]](#)
37. Jia, S.; Yuan, Y.; Li, N.; Liao, J.; Huang, Q.; Jia, X.; Xu, M. A Multiscale Superpixel-Level Group Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [\[CrossRef\]](#)
38. Zhao, H.; Bruzzone, L.; Guan, R.; Zhou, F.; Yang, C. Spectral-Spatial Genetic Algorithm-Based Unsupervised Band Selection for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9616–9632. [\[CrossRef\]](#)
39. Wan, Y.; Chen, C.; Zhang, L.; Gong, X.; Zhong, Y. Adaptive Multistrategy Particle Swarm Optimization for Hyperspectral Remote Sensing Image Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5520115. [\[CrossRef\]](#)
40. Su, H.; Yong, B.; Du, Q. Hyperspectral Band Selection Using Improved Firefly Algorithm. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 68–72. [\[CrossRef\]](#)

41. Su, H.; Cai, Y.; Du, Q. Firefly-Algorithm-Inspired Framework With Band Selection and Extreme Learning Machine for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 309–320. [\[CrossRef\]](#)
42. Wang, Y.; Zhu, Q.; Ma, H.; Yu, H. A Hybrid Gray Wolf Optimizer for Hyperspectral Image Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [\[CrossRef\]](#)
43. He, C.; Zhang, Y.; Gong, D.; Song, X.; Sun, X. A Multitask Bee Colony Band Selection Algorithm With Variable-Size Clustering for Hyperspectral Images. *IEEE Trans. Evol. Computat.* **2022**, *26*, 1566–1580. [\[CrossRef\]](#)
44. Wu, M.; Ou, X.; Lu, Y.; Li, W.; Yu, D.; Liu, Z.; Ji, C. Heterogeneous Cuckoo Search-Based Unsupervised Band Selection for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–16. [\[CrossRef\]](#)
45. Jacobs, L.W.; Brusco, M.J. Note: A Local-Search Heuristic for Large Set-Covering Problems. *Nav. Res. Logist.* **1995**, *42*, 1129–1140. [\[CrossRef\]](#)
46. Zhao, F.; Zhuang, C.; Wang, L.; Dong, C. An Iterative Greedy Algorithm With Q -Learning Mechanism for the Multiobjective Distributed No-Idle Permutation Flowshop Scheduling. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 3207–3219. [\[CrossRef\]](#)
47. Karimi-Mamaghan, M.; Mohammadi, M.; Pasdelpour, B.; Meyer, P. Learning to Select Operators in Meta-Heuristics: An Integration of Q-Learning into the Iterated Greedy Algorithm for the Permutation Flowshop Scheduling Problem. *Eur. J. Oper. Res.* **2023**, *304*, 1296–1330. [\[CrossRef\]](#)
48. Bertsimas, D.; Tsitsiklis, J. Simulated Annealing. *Stat. Sci.* **1993**, *8*, 10–15. [\[CrossRef\]](#)
49. Lee, S.; Kim, S.B. Parallel Simulated Annealing with a Greedy Algorithm for Bayesian Network Structure Learning. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1157–1166. [\[CrossRef\]](#)
50. Ribas, I.; Companys, R.; Tort-Martorell, X. An Iterated Greedy Algorithm for the Flowshop Scheduling Problem with Blocking. *Omega* **2011**, *39*, 293–301. [\[CrossRef\]](#)
51. 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. [\[CrossRef\]](#)
52. Ruiz, R.; Stützle, T. A Simple and Effective Iterated Greedy Algorithm for the Permutation Flowshop Scheduling Problem. *Eur. J. Oper. Res.* **2007**, *177*, 2033–2049. [\[CrossRef\]](#)
53. Huang, Y.-Y.; Pan, Q.-K.; Huang, J.-P.; Suganthan, P.; Gao, L. An Improved Iterated Greedy Algorithm for the Distributed Assembly Permutation Flowshop Scheduling Problem. *Comput. Ind. Eng.* **2021**, *152*, 107021. [\[CrossRef\]](#)
54. Zhang, M.; Ma, J.; Gong, M. Unsupervised Hyperspectral Band Selection by Fuzzy Clustering With Particle Swarm Optimization. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 773–777. [\[CrossRef\]](#)
55. Luo, X.; Xue, R.; Yin, J. Information-Assisted Density Peak Index for Hyperspectral Band Selection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1870–1874. [\[CrossRef\]](#)
56. Zeng, M.; Cai, Y.; Cai, Z.; Liu, X.; Hu, P.; Ku, J. Unsupervised Hyperspectral Image Band Selection Based on Deep Subspace Clustering. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1889–1893. [\[CrossRef\]](#)
57. Zhou, P.; Chen, J.; Fan, M.; Du, L.; Shen, Y.-D.; Li, X. Unsupervised Feature Selection for Balanced Clustering. *Knowl.-Based Syst.* **2020**, *193*, 105417. [\[CrossRef\]](#)
58. Haklı, H.; Uğuz, H. A Novel Particle Swarm Optimization Algorithm with Levy Flight. *Appl. Soft Comput.* **2014**, *23*, 333–345. [\[CrossRef\]](#)
59. Gong, M.; Zhang, M.; Yuan, Y. Unsupervised Band Selection Based on Evolutionary Multiobjective Optimization for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 544–557. [\[CrossRef\]](#)
60. Feng, J.; Jiao, L.C.; Zhang, X.; Sun, T. Hyperspectral Band Selection Based on Trivariate Mutual Information and Clonal Selection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4092–4105. [\[CrossRef\]](#)
61. Wang, Q.; Li, Q.; Li, X. A Fast Neighborhood Grouping Method for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5028–5039. [\[CrossRef\]](#)
62. 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. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.