

# Supervoxel Segmentation Based Evolutionary Multitasking Algorithm for Feature Selection of Hyperspectral Images

Lingjie Li, *Member, IEEE*, Yuze Zhang, Qiuzhen Lin, *Member, IEEE*, Zhong Ming, Carlos A. Coello Coello, *Fellow, IEEE*, and Victor C. M. Leung, *Life Fellow, IEEE*

**Abstract**—Feature selection (FS) is a very important technique for hyperspectral image (HSI) classification, as successfully selecting informative features can significantly increase the learning performance while reducing the computational cost. However, most of the existing FS methods tend to treat the HSI as a whole for FS, which does not fully consider the unique characteristics of HSIs and disregards the fact that different feature classes possess varying preferences for features. Thus, this paper proposes a supervoxel segmentation based evolutionary multitasking algorithm for FS of HSIs, called SS-EMT. First, the supervoxel segmentation method is used to partition the original HSI into several supervoxel blocks, which can preserve well the information of different classes of the original image. Second, in order to explore each supervoxel block efficiently, an evolutionary multitasking algorithm using particle swarm optimization is designed, which treats each supervoxel block as a subtask and then optimizes these subtasks collaboratively by transferring useful knowledge among related subtasks. In addition, a new individual evaluation mechanism is devised to obtain multiple high-quality feature subsets with different numbers of features simultaneously in a single run, thus reducing the computational cost. Finally, extensive experimental results on four common HSI datasets under three classifiers validate that our proposed method outperforms several state-of-the-art FS methods.

**Index Terms**—Feature selection, hyperspectral images, evolutionary multitasking, supervoxel segmentation.

## I. INTRODUCTION

A HYPERPECTRAL image (HSI) contains a reflected spectrum with hundreds to thousands of elements from each pixel lying on the Earth's surface that are acquired by airborne/spaceborne sensors [1]. Over the last two decades, HSIs have been widely used in several areas, such as medical imaging [2], robotics and autonomous systems [3], urban

planning [4], and environmental monitoring [5]. However, with the increased spectral information in HSIs, there are several new challenges for data processing and analysis. First, the dramatic increase in data volume poses a huge challenge to data processing (e.g., storage, transmission) and greatly increases the computational burden of learning models. Second, as the dimensionality of HSIs increases, the Hughes phenomenon [6] in classification may appear, i.e., the accuracy of classification will significantly degrade when the data dimensionality is too high under limited training data. Therefore, the dimensionality reduction of HSI is a crucial technique to improve classification accuracy and reduce computational cost [7]. Feature selection (FS) is considered as a representative and effective technique for processing HSIs. Note that FS in the field of hyperspectral remote sensing images also refers to band selection [8].

According to the availability of label information, existing FS methods can be mainly classified into three categories: (1) supervised [9], [10], (2) semi-supervised [11], [12] and (3) unsupervised [13], [14]. In general, due to the fact that manual labeling of HSI is prohibitively expensive, unsupervised FS methods are often preferred over the first two types of FS methods because they do not require labels for HSI processing. Furthermore, existing unsupervised FS methods can be broadly divided into three categories, including ranking-based [15], [16], [17], clustering-based [18], [19], [20], and search-based methods [21], [22], [23]. Specifically, ranking-based methods [15], [16] quantify the importance of features and then select the top-ranked features to form the optimal subset of features. Although this type of ranking-based methods has low complexity and high execution efficiency, most of them only consider the importance of each feature and ignore the correlation between the features, which may lead to a subset of features containing much redundant information [24]. Clustering-based methods [18], [19], [20] divide all features into multiple clusters through certain metrics, and then select representative features from each cluster to form the optimal subset of features. These clustering-based methods can eliminate redundant information of the features subset, but selecting representative features from each cluster may not guarantee the overall performance of the selected features. Moreover, it is also difficult to determine the appropriate number of clusters in advance [25]. In contrast, search-based methods first transform the FS problem into a discrete optimization problem and then employ some heuristic search strategy to explore the optimal

This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 62376163 and 62272315; in part by the Guangdong Regional Joint Foundation Key Project under Grant 2022B1515120076. (Corresponding Author: *Qiuzhen Lin*)

L. Li are affiliated with Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University, Shenzhen, 518060 China.

Y. Zhang, Q. Lin, and V. Leung are affiliated with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060 China (email: qiuazhlin@szu.edu.cn).

Z. Ming is affiliated with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China. He is also affiliated with the College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China.

C.A. Coello Coello is affiliated with the Department of Computer Science, CINVESTAV-IPN (Evolutionary Computation Group), México, D.F. 07300, MÉXICO. He is also (as part of a sabbatical leave) a member of the Faculty of Excellence at the School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, N.L., Mexico. (email: carlos.coellocoello@cinvestav.mx)

subset of features. Particularly, several types of evolutionary algorithms (EAs) [26], [27], [28] have become the mainstay of search-based FS methods for HSI processing due to their superior search capability, ease of implementation, and scalability [29], [30]. More recently, inspired by the outstanding performance of evolutionary multitasking (EMT) in numerical optimization [31], [32] and high-dimensional FS [33], [34], [35], a few studies have successfully applied the idea of EMT to deal with the FS problem of HSI and have shown favorable performance [25], [36]. Specifically, He et al. [25] applied a variable-size clustering strategy to model the FS of HSI as a multitasking optimization problem and then employed a multitasking multi-micro-group bee colony algorithm to explore optimal feature subsets. Shi et al. [36] proposed a semi-supervised FS model consisting of two tasks, which are constructed from labelled and unlabelled samples, respectively, and then adopted an EMT algorithm to cooperatively evolve these two tasks.

Although the above FS methods have achieved impressive performance in FS of HSI, most of them still face several major challenges. First, most existing FS methods treat the HSI as a whole and perform the FS process for all feature classes simultaneously, which ignores the spatial and spectral information of HSIs and fails to fully consider the characteristics of HSIs, i.e., different feature classes have different preferences for features (bands). Consequently, these FS methods cannot properly identify the optimal subset of features for different feature classes. Second, most existing methods tend to conduct FS directly on raw HSIs that consist of a vast amount of data. However, the enormous search space inherent to HSIs poses a great challenge to the exploration capabilities of these methods. Third, most existing FS methods are typically designed to search for the optimal subset of features with a fixed size, which is not well suited and is time-consuming when the decision maker needs to acquire multiple subsets of features with different sizes simultaneously.

To tackle the above challenges, this paper proposes a superpixel segmentation based evolutionary multitasking algorithm for FS of HSIs, called SS-EMT. Specifically, a superpixel segmentation method is used to partition the original HSI into multiple superpixel blocks. Then, an EMT algorithm is devised based on particle swarm optimization (PSO), where the superpixel blocks considered as subtasks are optimized through knowledge transfer among related subtasks. This way, SS-EMT can explore different optimal feature subsets for each superpixel block. In addition, a new individual evaluation mechanism is designed so that SS-EMT can simultaneously obtain multiple optimal feature subsets with different numbers of features. In summary, the main contributions of this work are the following:

- 1) A superpixel segmentation technique is employed in SS-EMT to partition HSI into multiple superpixel blocks, which can be considered as several subtasks. This way, the original FS problem of HSI is transformed into a multitasking optimization problem that considers the fact that different feature classes possess varying preferences for features, which is a significant difference with respect to existing work.

- 2) An EMT algorithm via PSO is designed in SS-EMT, which can search corresponding optimal feature subsets for different feature classes. In addition, a knowledge transfer strategy based on the similarity of subtasks is devised to achieve efficient knowledge transfer between subtasks that share the same or similar feature classes.
- 3) A new individual evaluation mechanism is proposed in SS-EMT, which integrates the individual performance of small feature subsets, medium feature subsets and large feature subsets, and then uses their average value as the evaluation metric of the individuals. This way, SS-EMT can find multiple optimal feature subsets with different sizes simultaneously in a single run, rather than requiring several independent runs as in most of the existing work.

The performance of SS-EMT is evaluated on four common HSI datasets, including the Indian Pines, Pavia University, Salinas, and Botswana datasets. Extensive experimental results under three widely used classifiers (i.e., k-nearest neighborhood (KNN) [37], support vector machine (SVM) [38] and random forest (RF) [11]) validate that SS-EMT can achieve higher classification accuracy than several state-of-the-art FS methods, including four non-EA methods (i.e., MVPCA [39], E-FDPC [15], ASPS-MN [40] and FNGBS [41]), two EA methods (i.e., MOBS [21] and SSGA [42]), and two EMT methods (i.e., MTPSO [34] and MBBS-VC [25]).

The remainder of this paper is organized as follows. **Section II** introduces some preliminaries. **Section III** presents the details of SS-EMT. **Section IV** provides our experimental results and their corresponding discussion. Finally, **Section V** provides our conclusions and some paths for future research.

## II. BACKGROUND AND PREVIOUS RELATED WORK

### A. Feature Selection of Hyperspectral Images

Feature selection (FS) is a key technique for data processing, especially when dealing with high-dimensional datasets consisting of a large number of irrelevant or redundant features. The main goal of FS is to improve the performance of learning models by reducing the dimensionality of the input space, thus reducing the complexity of the model and enhancing its interpretability [43]. It is worth noting that FS for HSI is also known as hyperspectral band selection (BS) because the selection target are the bands contained in the HSI.

During the FS process, a solution  $S$  can be expressed as a subset of  $L$  bands ( $L < I$ ) selected from the original dataset of  $I$  bands, where each band is represented as a binary value  $s$  in  $S$ . Specifically,  $S = (s_1, s_2, \dots, s_L)$ , where  $s_l = 1, l \in \{1, 2, \dots, L\}$ , indicates that the band in the  $l^{th}$  dimension is selected; otherwise, this band is abandoned with  $s_l = 0$ . Mathematically, an FS problem with  $L$  bands can be formulated as follows:

$$\begin{aligned} \min \quad & F(S), \\ \text{s.t.} \quad & S = (s_1, s_2, \dots, s_L), \\ & s_l \in \{0, 1\}, \quad l \in \{1, 2, \dots, L\}, \end{aligned} \quad (1)$$

where  $F(S)$  represents the performance evaluation function (e.g., classification error rate) to be optimized. This way,

the subset of bands contains fewer bands, which can reduce the amount of input data and improve the efficiency of the model [44]. At the same time, the subset of bands with better correlation can reduce the impact of redundant and noisy information, which also improves the accuracy and generalization ability of the model.

### B. Previous Related Work

In this section, a literature review of FS methods for HSIs is presented, including clustering-based, ranking-based, and search-based FS methods.

Ranking-based FS methods quantify the importance of each spectral band based on predefined band prioritization criteria and then select the top-ranked bands in the ranking sequence. For example, Rodriguez et al. [45] proposed a fast density-peak-based clustering method (FDPC), which identifies cluster centers by estimating the local density and intra-cluster distance of each point. To make FDPC more applicable to the FS problem of HSIs, Jia et al. [15] proposed an improved variant of FDPC that automatically determines the appropriate number of clusters to be selected by using a ranked sample component analysis method. In addition, Xu et al. [17] proposed a similarity-based ranking strategy, which uses structural similarity to measure the relationship between bands and then selects a subset of bands based on their ability to become cluster centers. This type of FS methods has low complexity and high execution efficiency, but most of them tend to consider only the importance of each band and ignore the correlation between bands; thus, the subset of bands may contain a large amount of redundancy [25].

Clustering-based FS methods divide the original bands into several different clusters based on some specific metrics, and then select representative bands from each cluster to form the final subset of bands. For example, Sun et al. [46] proposed a sparse spectral clustering method based on correlation entropy to cluster the bands and then selected high-quality bands from each cluster on the connected graph. Ji et al. [47] proposed a divisive hierarchical clustering approach, which aims to obtain an arbitrary number of subsets of bands and capture the intrinsic hierarchical nature of hyperspectral bands. This method first suppresses outlier clustering by introducing average dissimilarity through local density, and then selects representative frequency bands in each cluster. Sun et al. [48] proposed a fast and latent low-rank subspace clustering method, which conducts spectral clustering on the connected graph denoted by the affinity matrix, and then selects the bands closest to their corresponding cluster centroids to form the final band subset. This type of FS methods shows efficient performance in terms of redundancy elimination, but they select the representative bands from each cluster, which may ignore the overall performance after all the selected bands are combined. In addition, it is also difficult to determine an appropriate number of clusters in advance [25].

Search-based FS methods treat the FS problem as a target optimization problem and then apply some nature-inspired heuristic search method to explore the optimal subset of features. Among them, EAs have become one of the mainstream

methods in this category due to their good search capability, ease of implementation, and scalability. For example, Sawant et al. [49] proposed a hybrid EA by using wind-driven optimization and cuckoo search to co-evolve two subpopulations, which can avoid premature convergence and obtain a good subset of features. Phaneendra et al. [30] proposed a whale optimization-based algorithm with a hybrid filter to extract intrinsic and edge-preserving spatial features, which applies a nonlinear support vector machine trained on smoothed bands to achieve effective classification. Chen et al. [28] proposed a decomposition-based multiobjective clone selection algorithm. This method first decomposes the target problem into several subproblems by a set of weight vectors, and then employs a clone selection operator to perform an efficient local search on each subproblem to obtain a high-quality subset of features.

Recently, EMT, as an emerging paradigm of EAs, has received much attention because it exhibits good performance with low computational cost. Therefore, a few research studies have implemented the idea of EMT for FS of HSIs. For example, Shi et al. [36] proposed a multi-criteria semi-supervised EMT-based FS method. This method accelerates the search for promising bands by merging the band information, and then selects bands with high information content, high discrimination and low redundancy based on fully mining the numerical properties of labelled and unlabelled samples. He et al. [25] proposed an unsupervised multitasking FS algorithm based on artificial bee colony and variable-size clustering. This method uses a variable-size band clustering method to transform the BS problem into a multitasking optimization problem. Then, it applies a multitasking multi-micro-group bee colony algorithm with variable encoding length to search multiple optimal subsets of bands with different sizes simultaneously.

### C. Our Motivation

Although the above existing work shows good performance and the EMT technique has been successfully applied to FS of HSIs, most of them still face several challenges:

- 1) As summarized in the last column of Table I, most existing work treats HSIs as a whole and then performs the FS process for all feature classes simultaneously, which does not sufficiently take into account the fact that different feature classes have different preferences for features, thus limiting the classification performance to some extent. Therefore, it is necessary to design an effective FS method that can fully consider the feature preferences of different feature classes.
- 2) As shown in the penultimate column of Table I, most existing methods are usually designed to search for an optimal subset of features with a fixed size. That is, these methods all need to be run independently multiple times to obtain different numbers of feature subsets, leading to significant time costs. Although He et al. [25] devised a variable-size clustering method to alleviate this problem, it is still computationally expensive as it requires repeated evaluation of each feature subset with different numbers of features. Therefore, it is worthwhile further investigating a more efficient mechanism for feature subset evaluation.

TABLE I: Summary of several EA-based FS methods for HSIs regarding the method type, task type, fitness function, their capability to obtain multiple results in one run and their consideration of class preference

Algorithms	Year	Type	Task type	Fitness function	Obtaining multiple results in one run	Considering class preference
WDOMCS[49]	2021	Supervised	Single task	Accuracy	No	No
BLNPK2021[30]	2021	Supervised	Single task	Accuracy	No	No
MABC-BS[14]	2021	Unsupervised	Single task	Variation coefficient and correlation measure	No	No
HGWO[9]	2022	Supervised	Single task	Inter-class dispersion matrix and intra-class dispersion matrix	No	No
MOCSA/D_FS[28]	2022	Unsupervised	Single task	Information entropy, mutual information and the ratio of the relative scatter value	No	No
LFGA[12]	2022	Semi-supervised	Single task	Accuracy	No	No
MTBS[36]	2022	Semi-supervised	Multitask	Information entropy, mutual information and accuracy	No	No
MBBS-VC[25]	2022	Unsupervised	Multitask	Variation coefficient	Yes	No
SS-EMT (ours)	-	Unsupervised	Multitask	Information entropy and mutual information	<b>Yes</b>	<b>Yes</b>

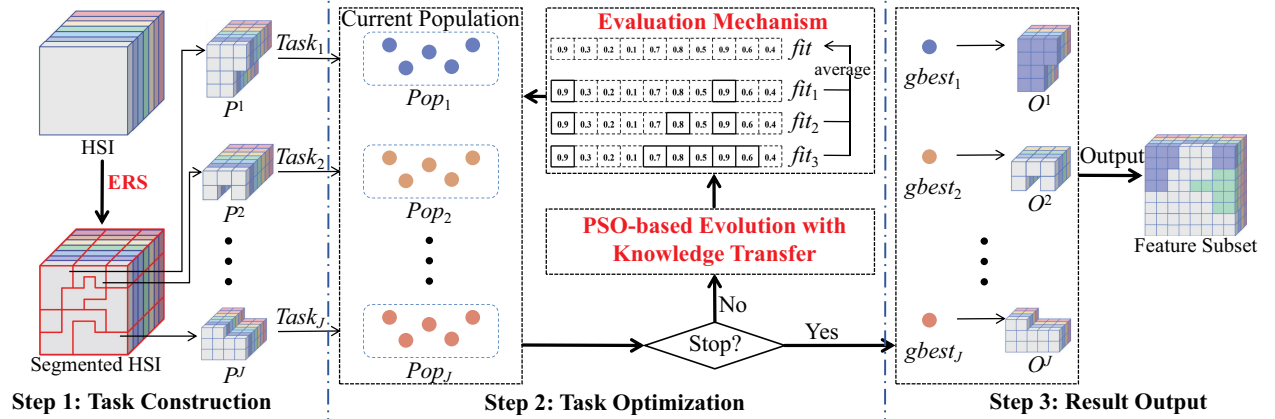


Fig. 1: Flowchart of our proposed SS-EMT.

Based on the above analysis, this paper proposes a superpixel segmentation based EMT algorithm for FS of HSIs, called SS-EMT. Specifically, for addressing the first challenge, SS-EMT adopts a superpixel segmentation technique to partition the original HSI into multiple superpixel blocks, and then treats each superpixel block as an FS subtask. This way, the original FS problem of HSI is transformed into a multitasking optimization problem, where each subtask is responsible for selecting the most appropriate subset of features for a particular feature class. To address the second challenge, SS-EMT proposes a novel individual evaluation mechanism that can simultaneously evaluate different subsets of features with different numbers of features, thus significantly reducing the computational cost. The details of our SS-EMT are provided in the next section.

### III. OUR PROPOSED APPROACH

#### A. The General Framework

With the aim of selecting different optimal feature subsets for several feature classes, our proposed SS-EMT first transforms the original FS problem of HSI into a multitasking optimization problem by using the superpixel segmentation method, and then designs a PSO-based EMT algorithm to optimize the previously constructed problem. **Algorithm 1** provides the pseudo-code of SS-EMT and Fig. 1 provides the flowchart of the proposed SS-EMT, which consists of three main steps, including multitasking construction, multitasking optimization, and result output. Specifically, mul-

#### Algorithm 1 The framework of SS-EMT

**Input:** Original HSI data  $\mathbf{H}$ .

**Output:** The selected feature subset  $\mathbf{O}$ .

- 1:  $\{P^1, P^2, \dots, P^J\} \leftarrow \text{Multitasking Construction } (\mathbf{H})$ ;
- 2: Treat each superpixel block as an FS task and initialize  $J$  populations for each task, i.e.,  $(Pop_1, Pop_2, \dots, Pop_J)$ ;
- 3: **while** the termination condition is not met **do**
- 4:   **Multitasking Optimization**  $(Pop_1, Pop_2, \dots, Pop_J)$ ;
- 5: **end while**
- 6: Select the best particle for each task and output its corresponding subset of features, i.e.,  $\mathbf{O} = \{O^1, O^2, \dots, O^J\}$ .

titasking construction is performed by using the superpixel segmentation method, which segments the original HSI into several superpixel blocks and then treats each superpixel block as a subtask to be optimized. This way, the original FS problem of HSI is easily transformed into a multitasking optimization problem. The details of multitasking construction are introduced in **Section III-B**. After that, multitasking optimization is carried out, where a PSO-based EMT algorithm is designed to optimize the multiple subtasks indicated before, in a cooperative manner by transferring useful knowledge among related subtasks. In addition, a novel individual evaluation mechanism is designed to search for various optimal feature subsets with different sizes simultaneously, thus reducing the time complexity. The details of multitasking optimization (i.e., PSO-based evolution with knowledge transfer and individual



evaluation mechanism) are introduced in **Section III-C** and **Section III-D**, respectively. Finally, the best individual for each subtask is selected and its corresponding features are produced as the final output.

### B. Multitasking Construction

As mentioned above, selecting different feature subsets for various feature classes according to their corresponding feature preferences can lead to favorable classification results. To this end, this paper introduces an entropy rate superpixels (ERS) segmentation method [50], an unsupervised technique without any class information, which partitions the original HSI data space into multiple superpixel blocks and then treats each superpixel block as an FS subtask. Thus, the original FS problem of HSI can be easily transformed into a multitasking optimization problem, where each subtask is responsible for selecting the most appropriate subset of features for a particular feature class. The details of multitasking construction are introduced below.

First, given an HSI dataset  $\mathbf{H}$ , it contains a total of  $I$  bands (features), each of which consists of  $M \times N$  pixels ( $M$  and  $N$  are the height and width of the HSI, respectively). The mathematical model can be expressed as follows:

$$\begin{aligned} \mathbf{H} &= \{H_1, H_2, \dots, H_I\}, \\ H_i &= \{p_1, p_2, \dots, p_{M \times N}\}, \end{aligned} \quad (2)$$

where  $H_i$  denotes the set of pixel points in the  $i^{th}$  band and  $p$  refers to the pixel points. Then, the ERS method is adopted to construct homogeneous regions. Note that the ERS method is a graph-based clustering approach that generates superpixels by performing graph partitioning. Generally, there are four main steps when generating superpixels from images, including graph construction, entropy rate, balancing function, and optimization, as introduced below.

**Graph construction:** An image is initially mapped to a graph  $G$ , described as follows:

$$G = (V, E), \quad (3)$$

where the vertex  $V$  represents the pixels in an image and the edge set  $E$  indicates the corresponding set of edges between  $V$ . Then, the graph is partitioned into a connected subgraph by choosing a subset of edges  $E' \in E$  such that the resulting graph  $G' = (V, E')$  consists of smaller connected subgraphs.

**Entropy rate:** To create the compact and homogeneous superpixels, the criterion adopted is the entropy rate of the random walk on the constructed graph  $G' = (V, E')$ . The function  $ER(\cdot)$  of the random walk on  $G'$  is defined as follows:

$$ER(E') = - \sum_i \mu_i \sum_j P_{i,j}(E') \log(P_{i,j}(E')), \quad (4)$$

where  $P_{i,j}(\cdot)$  is the transition probability for the random walks from the  $i^{th}$  vertex to the  $j^{th}$  vertex.  $\mu_i$  is defined by the stationary distribution, calculated by

$$\mu_i = w_i / w_T, \quad (5)$$

---

### Algorithm 2 Multitasking construction

---

**Input:** Original HSI data  $\mathbf{H}$ .

**Output:** Multiple superpixel blocks  $\{P^1, P^2, \dots, P^J\}$ .

- 1: Map  $\mathbf{H}$  to a graph  $G$ ;
  - 2: Select a subset of edges in  $G$  to construct a subgraph  $G'$ ;
  - 3: Calculate the entropy rate of random walk on  $G'$  by equation (4);
  - 4: Obtain multiple clusters of similar size by equation (6);
  - 5: Optimize the clustering objective function in equation (8);
  - 6: Obtain multiple superpixel blocks  $\{P^1, P^2, \dots, P^J\}$ .
- 

where  $w_i$  is the sum of the weights of edges passing the  $i^{th}$  vertex and  $w_T$  is the total number of edges in  $G'$  (i.e.,  $w_T = \sum_{i=1}^{|V|} w_i$ ).

**Balancing function:** To obtain clusters of similar size, a balancing function is introduced, defined as follows:

$$\begin{aligned} B(E') &= H(Z_{E'}) - N_{E'} \\ &= - \sum_i PZ_{E'}(i) \log(PZ_{E'}(i)) - N_{E'}, \end{aligned} \quad (6)$$

where  $N_{E'}$  denotes the number of connected components. The cluster membership distribution is represented by  $Z_{E'}$  and the graph partitioning for the edge set  $E'$  is denoted as  $S_{E'} = \{S_1, S_2, \dots, S_{N_{E'}}\}$ . The distribution of  $Z_{E'}$  can be represented as follows:

$$PZ_{E'}(i) = \frac{|S_i|}{|V|}, \quad i = \{1, \dots, N_{E'}\}. \quad (7)$$

**Optimization:** ERS proposes a clustering objective function based on graph theory, where the entropy rate function  $ER(\cdot)$  and the balancing function  $B(\cdot)$  are introduced to optimize the superpixel segmentation. The objective function for ERS is formulated as follows:

$$\arg \max_{E'} \text{Tr}\{ER(E') + \alpha B(E')\} \quad s.t. E' \subseteq E, \quad (8)$$

where  $\alpha$  is used to balance the contributions of the entropy rate function and the balancing function,  $\text{Tr}\{\cdot\}$  means the trace operation of the matrix. Note that the value of  $\alpha$  is set according to the suggestion of the original reference [50]. The function  $ER(\cdot)$  is used to form homogeneous and compact clusters, and the function  $B(\cdot)$  is used to form clusters with similar sizes.

After the above steps, the original data  $\mathbf{H}$  is segmented into multiple superpixel blocks by using the ERS method. This way, the original mathematical model in equation (2) can be transformed as follows:

$$\begin{aligned} \mathbf{H} &= \{P^1, P^2, \dots, P^J\}, \\ P^j &= \{H_1^j, H_2^j, \dots, H_I^j\}, \\ H_i^j &= \{p_1, p_2, \dots, p_k\}, \end{aligned} \quad (9)$$

where  $P^j$  indicates the  $j^{th}$  superpixel block,  $j \in \{1, 2, \dots, J\}$  and  $J$  is the number of superpixel blocks.  $H_i^j$  means the  $i^{th}$  band of the  $j^{th}$  superpixel block,  $i \in \{1, 2, \dots, I\}$ , consisting of  $k$  pixel points (i.e.,  $\{p_1, p_2, \dots, p_k\}$ ). Note that  $k$  is a variable, meaning that each superpixel block may contain a different number of pixel points.

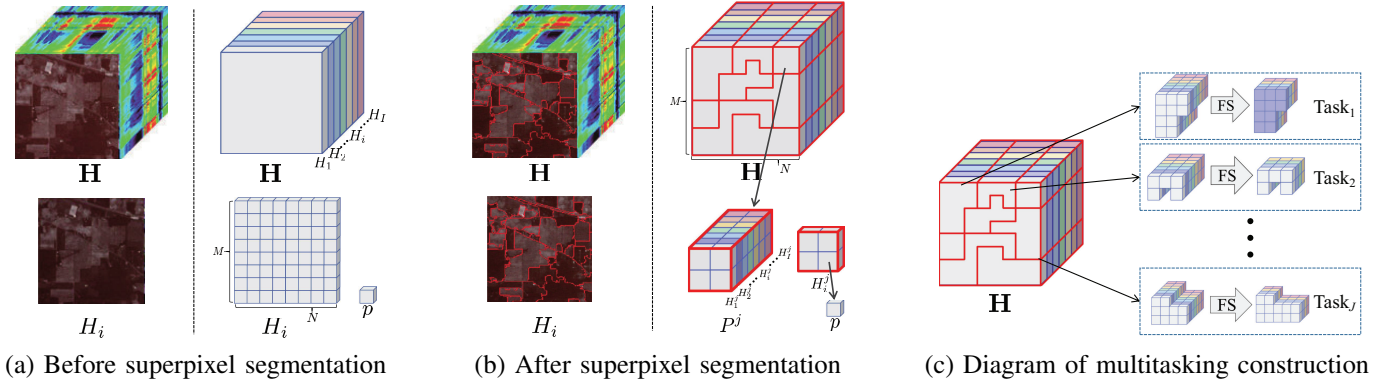


Fig. 2: Illustration of the ERS method and its use for multitasking construction

For a more intuitive observation, Fig. 2 is provided here, where Figs. 2(a)-(b) show the visualization of the transformed mathematical model in equation (9), and Fig. 2(c) shows the diagram of multitasking construction after using the above ERS method. Finally, these segmented superpixel blocks are considered as several FS subtasks, i.e.,  $\{Task_1, \dots, Task_J\}$ , which are then optimized using the EMT algorithm. This way, the purpose of selecting different optimal feature subsets for different feature classes is achieved. **Algorithm 2** provides the pseudo-code of the **Multitasking Construction**.

### C. Multitasking Optimization

After the completion of the above multitasking construction, multitasking optimization is performed, which optimizes the  $J$  subtasks constructed above cooperatively. At the beginning,  $J$  subpopulations, denoted as  $Pop_1, Pop_2, \dots, Pop_J$ , are initialized to solve their corresponding subtasks simultaneously. Note that each subpopulation contains  $SP_N$  individuals, and each individual adopts the real encoding approach with a length equal to the number of features (bands).

To effectively optimize the multitasking optimization problem constructed above, this paper devises an EMT to achieve efficient knowledge transfer between subtasks belonging to the same or similar feature classes by using a PSO algorithm [51]. The reasons for adopting PSO as the solver in EMT are because of its good search capability, ease of implementation, and scalability [34], [52], [53]. In addition, the empirical results in [34], [54] also show its advantages in solving high-dimensional FS problems. Therefore, it is reasonable to adopt PSO as the solver of EMT in this paper. Mathematically, the velocity  $V_i(t)$  and position  $X_i(t)$  of each particle in the original PSO [51] are updated by

$$V_i(t+1) = \omega \times V_i(t) + c_1 \times r_1 \times (X_{pbest_i} - X_i(t)) + c_2 \times r_2 \times (X_{gbest_i} - X_i(t)), \quad (10)$$

$$X_i(t+1) = X_i(t) + V_i(t+1), \quad (11)$$

where  $\omega$  is the inertia weight,  $t$  is the current iteration,  $c_1$  and  $c_2$  are two learning factors, and  $r_1$  and  $r_2$  are two random numbers in the range  $[0,1]$ .  $X_{pbest_i}$  and  $X_{gbest_i}$  represent the personal best and global best positions at the current iteration for the  $i^{th}$  particle, respectively.

Typically, due to the inherent correlation among subtasks, utilizing the valid knowledge (experience) from the subtasks

that are similar to the target subtask can help to better solve the target subtask, thus significantly improving the overall performance of EMT. However, there are three key research questions (RQs) to consider when designing an efficient knowledge transfer strategy in the EMT, namely, **When to transfer? What to transfer? How to transfer?** Next, these three RQs will be discussed.

**RQ1: When to transfer?** As suggested in [52], this paper introduces a random mating probability, denoted as  $rm_p$ , to decide whether to transfer knowledge between subtasks. Specifically, for each particle in one iteration, a randomly generated real-valued number in the range  $[0,1]$  is compared to the predefined transfer probability  $rm_p$ . If this random number is larger than  $rm_p$ , knowledge transfer does not occur, and conventional PSO is executed to evolve the corresponding particle whose velocity and position are updated by using equation (10) and equation (11), respectively. Otherwise, the knowledge transfer between subtasks is considered, expecting that valid knowledge from other related subtasks is utilized to assist in the optimization of the target subtask. Apparently, an appropriate transfer probability has a huge impact on the EMT performance. In this paper,  $rm_p$  is set to 0.5 according to the parameter-sensitive analysis provided in **Section IV-D**.

**RQ2: What to transfer?** This is a core part of knowledge transfer in EMT because only valid knowledge from highly related subtasks can facilitate the optimization of the target subtask. In contrast, some knowledge acquired from unrelated subtasks may mislead the optimization of subtasks, which eventually leads to the deterioration of the optimization performance. Unfortunately, most existing studies [25], [36] have ignored the differences between subtasks, leading to potential negative knowledge transfer. The key to solving this problem is to achieve effective knowledge transfer between some highly relevant subtasks as much as possible. To this end, this paper adopts the mutual information (MI) [55] to measure the similarity between two subtasks. In particular, two important definitions are introduced below:

**Definition** (Task Similarity):  $\Psi_q^j$  is the similarity between subtask  $Task_j$  and other subtasks  $Task_q$  according to the MI values, calculated as follows:

$$\Psi_q^j = \sum_{i=1}^I \text{MI}(H_i^q; H_i^j), \quad j, q \in \{1, 2, \dots, J\}, j \neq q, \quad (12)$$

**Algorithm 3 Calculation of the similarity factor**


---

**Input:** The set of superpixel blocks  $\{P^1, P^2, \dots, P^J\}$ .  
**Output:** The similarity factor of each subtask  $\{\tau_1, \tau_2, \dots, \tau_J\}$ .

- 1: **for**  $j = 1 : J$  **do**
- 2:   **for**  $q = 1 : J$  **do**
- 3:      $\Psi_q^j \leftarrow$  Calculate the subspace similarity between  $Task_q$  and  $Task_j$  by equation (12);
- 4:   **end for**
- 5:    $\tau_j = \{\tau_j^1, \tau_j^2, \dots, \tau_j^J\}$ ;
- 6: **end for**

---

where  $H_i^j$  and  $H_i^q$  are the sets of pixel points in the  $i^{th}$  band of the  $j^{th}$  and  $q^{th}$  superpixel blocks, respectively.  $MI(X; Y)$  means the amount of  $X$  contained in  $Y$ , computed by:

$$MI(X; Y) = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (13)$$

where  $x$  and  $y$  represent the samples of variables  $X$  and  $Y$ , respectively.  $p(x)$  and  $p(y)$  refer to the edge probabilities of  $X$  and  $Y$ , respectively.  $p(x, y)$  is the joint probability density.

**Definition** (Similarity Factor):  $\tau_j = \{\tau_j^1, \tau_j^2, \dots, \tau_j^J\}$  is the set of similarity rankings of all other subtasks to  $Task_j$ , i.e., the similarity factor. Note that a higher ranking indicates a higher similarity of a task to others.

It should be noted that the number of subtasks ( $J$ ) in this work is set as twice the number of feature classes contained in the HSI dataset, which is based on the parameter-sensitive analysis provided in **Section IV-D**. That is,  $\tau_j^1$  is the most similar subtask of  $Task_j$ , which belongs to the same feature class, and  $\tau_j^2$  is the second similar subtask of  $Task_j$ , which belongs to a different feature class but has a high similarity. Therefore, it is reasonable to conclude that the valid information in these two most similar subtasks (i.e.,  $\tau_j^1$  and  $\tau_j^2$ ) has a high potential to help the evolution of the target subtask  $Task_j$ . **Algorithm 3** presents the pseudo-code of the calculation of the similarity factor.

**RQ3: How to transfer?** After the above process, the two most similar subtasks (i.e.,  $\tau_j^1$  and  $\tau_j^2$ ) of subtask  $Task_j$  are identified, which belong to the same or similar feature classes. Then, the valid knowledge from these two highly relevant subtasks is transferred to assist the evolution of the target subtask  $Task_j$ . Specifically, the particles with the best fitness values in subtasks  $\tau_i^1$  and  $\tau_i^2$  are selected. Note that the calculation of the fitness value is introduced in **Section III-E**. This way, the velocities of the particles of  $Task_j$  are updated by using the knowledge transfer strategy, defined as follows.

$$\begin{aligned} V_i(t+1) = & \omega \times V_i(t) + c_1 \times r_1 \times (X_{pbest_i} - X_i(t)) \\ & + c_2 \times r_2 \times (X_{gbest_i} - X_i(t)) \\ & + c_3 \times r_3 \times (X_{ts} - X_i(t)), \end{aligned} \quad (14)$$

where  $X_{ts}$  is the average position of these two transferred particles from subtasks  $\tau_j^1$  and  $\tau_j^2$ . Finally, the positions of the particles are updated using equation (11). **Algorithm 4** provides the pseudo-code of the **Multitasking Optimization**.

**Algorithm 4 Multitasking optimization**


---

**Input:** The set of current populations  $\{Pop_1, \dots, Pop_J\}$ .  
**Output:** The set of evolved populations  $\{Pop'_1, \dots, Pop'_J\}$ .

- 1: **for** each subtask  $Task_j$  **do**
- 2:   **if**  $rand < rmp$  **then**
- 3:     Calculate the task similarity factor between  $Task_j$  and other subtasks; // See **Algorithm 3**
- 4:     Evolve the velocity of particles in  $Pop_j$  using equation (14);
- 5:   **else**
- 6:     Evolve the velocity of particles in  $Pop_j$  using equation (10);
- 7:   **end if**
- 8:   Update the position of particles in  $Pop_j$  using equation (11);
- 9:   Update the  $X_{pbest}^j, X_{gbest}^j$  according to the proposed individual evaluation mechanism; // See **Section III-D**
- 10: **end for**

---

**D. A Novel Individual Evaluation Mechanism**

In real-world scenarios, decision makers may prefer to obtain multiple high-quality feature subsets with different numbers of features simultaneously, so that they can choose the most suitable one according to different requirements [25]. However, most existing work requires a single independent run to obtain a feature subset with a specific number of features, yet this process is tedious and time-consuming [56], [49]. To alleviate these issues, this paper proposes a novel individual evaluation mechanism that simultaneously evaluates the overall quality of an individual by comprehensively considering the performance (i.e., fitness value, as detailed in **Section III-E**) of its three corresponding feature subsets of different sizes, including the small-size feature subset, the medium-size feature subset, and the large-size feature subset. That is, a high-quality individual means that it can obtain good classification performance on multiple feature subsets with a varying numbers of features. This way, multiple high-quality feature subsets with different sizes can be easily obtained in a single run, thus reducing the time complexity of the algorithm.

Fig. 3 provides a simple example to illustrate the schematic of the proposed individual evaluation mechanism. Assuming that an individual, namely *ind*, with 10 dimensions is to be evaluated and the maximum number of selected bands is 6, where the gene position in the solid line frame means that the corresponding feature is selected. The fitness value (*fit*) of *ind* is averaged by that of its three feature subsets of different sizes, including a small feature subset of 2 dimensions, a medium feature subset of 3 dimensions and a large feature subset of 6 dimensions, i.e.,  $fit = (fit_1 + fit_2 + fit_3)/3$ . This way, a good fitness value implies that this individual exhibits high quality of feature subsets with different sizes.

**E. Fitness Function**

The design of the fitness function plays a crucial role in the performance of FS, because a reasonable fitness function can effectively guide the FS process to select high-quality features.

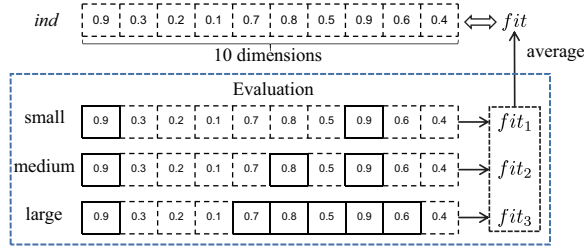


Fig. 3: An example of the individual evaluation mechanism

Typically, when performing FS on HSIs, the main goal is to select features that are informative and have low redundancy. Note that in HSI-specific scenarios, FS is also referred to as band selection (BS) contained in HSIs. Therefore, this paper designs a novel fitness function consisting of two objectives, defined as follows:

$$\min F(S^j) = \min[f_1(S^j), f_2(S^j)],$$

$$= \begin{cases} f_1(S^j) = 1 / \sum_i^I \text{IE}(H_i^j), \\ f_2(S^j) = 1 / \sum_i^I \sum_q^I \text{MI}(H_i^j, H_q^j), \end{cases} \quad (15)$$

where  $S^j$  is the solution of  $\text{Task}_j$  and  $I$  is the number of selected bands.  $H_i^j$  and  $H_q^j$  are the sets of pixel points in the  $i^{\text{th}}$  and  $q^{\text{th}}$  bands of the  $j^{\text{th}}$  superpixel blocks, respectively.  $\text{IE}(\cdot)$  and  $\text{MI}(\cdot)$  refer to the amount of information and the correlation of features, respectively. Specifically, this paper adopts information entropy to measure the amount of information in bands and uses mutual information to reflect the correlation among the features, introduced as follows:

- 1) Information entropy: Information entropy is used to measure the amount of information contained in each feature (band). According to the definition of Shannon [21], the **entropy** of a discrete random variable can be expressed as follows:

$$\text{IE}(H_i^j) = - \sum p(H_i^j\_g) \log p(H_i^j\_g),$$

$$\text{s.t. } \sum p(H_i^j\_g) = 1, \quad (16)$$

where  $H_i^j\_g$  denotes the grayscale level in the  $i^{\text{th}}$  band of the  $j^{\text{th}}$  superpixel block.  $p(H_i^j\_g)$  denotes the proportion of  $H_i^j\_g$  in all grayscale levels. For HSIs, each band is considered as the set of outputs of a random variable, and the grayscale histogram of this band is the probability distribution.

- 2) Mutual information: Mutual information is used to measure the correlation between the features (bands), where  $\text{MI}(H_i^j, H_q^j)$  means the amount of  $H_i^j$  contained in  $H_q^j$ , calculated using equation (13).

#### F. Complexity Analysis

As presented in Fig. 1, the proposed SS-EMT consists of three main steps: multitasking construction, multitasking

TABLE II: Basic information of the four adopted HSI datasets

Dataset	Bands	Classes	Size	Samples
Indian Pines	200	16	$145 \times 145$	10249
Pavia University	103	9	$610 \times 340$	42776
Salinas	204	16	$512 \times 217$	54129
Botswana	145	14	$1476 \times 256$	3248

optimization (i.e., PSO-based evolution and evaluation mechanism), and results output. Specifically, the ERS method is used to segment the original HSI dataset into several superpixel blocks, which requires a total computational complexity of  $O(M \times N \times \log(M \times N))$ , where  $M$  and  $N$  are the height and width of the HSI, respectively. Second, as presented in **Algorithm 4**, the PSO-based evolution is performed on several subtasks, which requires a computational complexity of  $O(I \times J^2 \times SP_N)$ , where  $I$  is the total number of bands in the data,  $J$  is the number of superpixel blocks, and  $SP_N$  is the size of each subpopulation. Moreover, the process of individual evaluation has a computational complexity of  $O(3 \times \text{Cof} \times J \times SP_N)$ , where  $\text{Cof}$  is the cost of the objective function. Last, the computational complexity of the results output is  $O(J \times SP_N)$ . Overall, the worst computational complexity of SS-EMT in one iteration is  $\max\{O(I \times J^2 \times SP_N), O(\text{Cof} \times J \times SP_N)\}$ .

#### IV. EXPERIMENTAL ANALYSIS

##### A. Datasets

In our experiments, four widely used HSI datasets were selected to assess the performance of the proposed SS-EMT, including the Indian Pines, Pavia University, Salinas, and Botswana datasets<sup>1</sup>. Table II summarizes the basic information of these four adopted datasets. In addition, due to page limitations, more details of each dataset are presented in Section S-I of the **Supplementary Material**.

##### B. Comparison Methods

Eight representative feature/band selection methods are adopted for performance comparisons in our experiments, including four non-EA-based FS methods (i.e., MVPCA [39], E-FDPC [15], ASPS-MN [40] and FNGBS [41]), two EA-based FS methods (i.e., MOBS [21] and SSGA [42]), and two EMT-based FS methods (i.e., MTPSO [34] and MBBS-VC [25]). Note that all the simulations were conducted using MATLAB R2020a on a 64-bit Windows 10 personal computer with 24 GBytes of RAM and an Intel Core-i7 3.6 GHz processor. The average results of 20 independent runs obtained by each algorithm were recorded and considered as the final classification results. Note that the source code<sup>2</sup> of SS-EMT is also provided so that those interested can reproduce our experimental results.

##### C. Experimental Setup

- **Classification settings:** Three commonly used classifiers, including k-nearest neighborhood (KNN) [37], support

<sup>1</sup>The sources of the four adopted datasets are available at: [www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)

<sup>2</sup>The source code of SS-EMT is available at: <https://github.com/zhangyuzecn/SS-EMT>

TABLE III: Parameters settings for SS-EMT

Parameters	Values
The maximum number of selected bands ( $L$ )	30
The number of superpixel blocks ( $J$ )	$2 \times \text{Classes}$
The population size for each subtask ( $SP_N$ )	5
The probability of knowledge transfer ( $rm_p$ )	0.5
The maximum number of iterations ( $G_{max}$ )	100
The learning factors ( $c_1, c_2, c_3$ )	1.49445
The inertia weight ( $\omega$ )	$0.9 - 0.5 * (t/G_{max})$

vector machine (SVM) [38] and random forest (RF) [11], were adopted to classify samples on four public HSI datasets and verify the quality of the obtained feature subset. In this paper, as suggested in [41],  $k$  is set to 5 for the KNN classifier. For the SVM classifier, the RBF kernel is used and the penalties  $C$  and gamma are initialized to  $1 \times 10^4$  and 0.5, respectively. For the RF classifier, the number of trees is set to 10. In addition, the context of each dataset is not considered to simplify the classification problem. The setting for the ratio of training set used for the tests is the same as the one used in MBBS-VC [25] to ensure fairness. Specifically, 20% of the samples are randomly selected for each experiment to form the training set, and the remaining 80% of the samples are used for testing.

- **Parameters settings:** Table III summarizes the parameters settings of the proposed SS-EMT. Note that the settings of the inertia weight and the learning factors were suggested by [34]. To ensure fairness, for non-EA-based FS methods, the parameters settings were set according to their original references. For EA-based and EMT-based FS methods, the settings of the maximum number of iterations and total population size were the same as those of SS-EMT, while other unique parameters and fitness functions were set according to their original references. Note that the parameters settings of all the methods are listed in Table S-1 of the **Supplementary Material**.
- **Performance metrics:** Two well-known classification accuracy metrics were chosen to evaluate the performance of the algorithm, including the overall accuracy (OA) and the KAPPA coefficient [57], where the larger values for both metrics indicate better performance. More details of these two metrics are introduced in Section S-II of the **Supplementary Material**.

#### D. Parametric Analysis

As summarized in Table III, the proposed SS-EMT consists of three unique and key parameters, including the number of superpixel blocks segmented by using the ERS method ( $J$ ), the population size for each subtask ( $SP_N$ ), and the probability of knowledge transfer ( $rm_p$ ). This section conducts parameter sensitivity analysis of the above parameters to clarify the reasons for the chosen parameters settings.

1) Analysis of  $J$  and  $SP_N$ : The parameter analyses of  $J$  and  $SP_N$  were conducted on a representative dataset, i.e., the Indian Pines dataset, and the average classification results under the SVM classifier were collected after 20 independent runs with different parameters settings. Specifically, three different values for  $J$  (i.e.,  $1 \times$ ,  $2 \times$  and  $3 \times$ ) and three

TABLE IV: Parametric analysis on  $J$  and  $SP_N$ 

$J$	$1 \times$		$2 \times$		$3 \times$	
$SP_N$	TIME	OA	TIME	OA	TIME	OA
3	180.79	86.32	319.85	89.01	454.31	90.14
5	182.76	86.34	321.37	90.21	456.60	91.26
7	185.13	86.54	323.37	88.88	457.52	90.79

different values for  $SP_N$  (i.e., 3, 5 and 7) were considered for these experiments. Note that the number of superpixel blocks ' $1 \times$ ' refers to the multiple of the number of feature classes contained in the dataset. For example, there are 16 feature classes in the Indian Pines dataset, then ' $1 \times$ ', ' $2 \times$ ' and ' $3 \times$ ' represent that  $J$  is set to 16, 32 and 48, respectively. Thus, there are 9 different combinations of these two parameters. Table IV provides the average classification accuracy (%) and average execution time for each combination, where the number of selected bands ranges from 2 to 30. Clearly, the classification performance improves as  $J$  and  $SP_N$  increase. The combination of  $\{3 \times, 5\}$  achieves the highest OA result, while it has a slight performance improvement compared to the combination of  $\{2 \times, 5\}$  with additional computational overhead. Therefore, considering both the classification performance and the computational cost, it is reasonable to use the combination of  $\{2 \times, 5\}$  for SS-EMT in this paper.

2) Analysis of  $rm_p$ : The probability of knowledge transfer  $rm_p$  is a key parameter in EMT, as it determines whether knowledge is transferred among subtasks. This experiment aims to obtain an appropriate setting for  $rm_p$ . Due to page limitations, Table S-2 of the **Supplementary Material** provides the average classification accuracy (%) of SS-EMT for several values of  $rm_p$  (i.e.,  $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ ) on the Indian Pines dataset with the number of selected bands ranging from 4 to 30. As shown in the last column of Table S-1, SS-EMT with  $rm_p = 0.5$  achieves the best classification performance. Therefore,  $rm_p = 0.5$  is adopted for SS-EMT in all experiments.

#### E. Comparisons with Non-EA-based FS Methods

In this section, four non-EA-based FS methods are selected for performance comparisons, including two ranking-based methods (MVP-CA [39], E-FDPC [15]) and two clustering-based methods (ASPS-MN [40], FNGBS [41]). The results of OA and of the KAPPA coefficient obtained by each compared algorithm on four common HSI datasets (i.e., Indian Pines, Pavia University, Salinas, and Botswana datasets) under three classifiers (i.e., KNN, SVM and RF) are discussed next.

1) Results on the Indian Pines dataset: Fig. 4 provides the OA results of each compared algorithm on the Indian Pines dataset. Some conclusions can be drawn from these figures. First, as shown in Fig. 4(a), which provides the OA results under the KNN classifier, SS-EMT achieves the best results on all subsets of band sizes 2-30, which are significantly better than those of its competitors. In particular, SS-EMT outperforms the baseline approach (i.e., All Bands) in most cases except for the 2-band case, while most of the compared algorithms fail to achieve better classification results than the baseline approach, with only FNGBS and ASPS-MN outperforming the baseline approach in a few cases. Second, the classification results of SS-EMT can exceed 91% on just



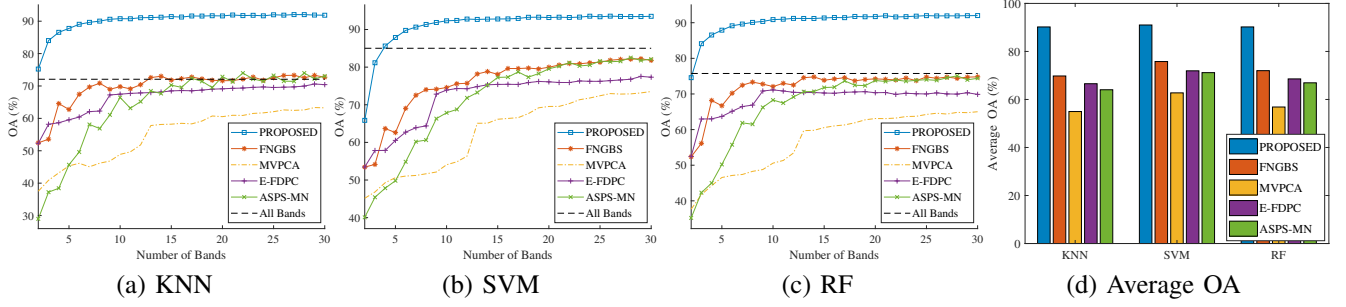


Fig. 4: The OA results of SS-EMT and four non-EA-based FS methods on the Indian Pines dataset

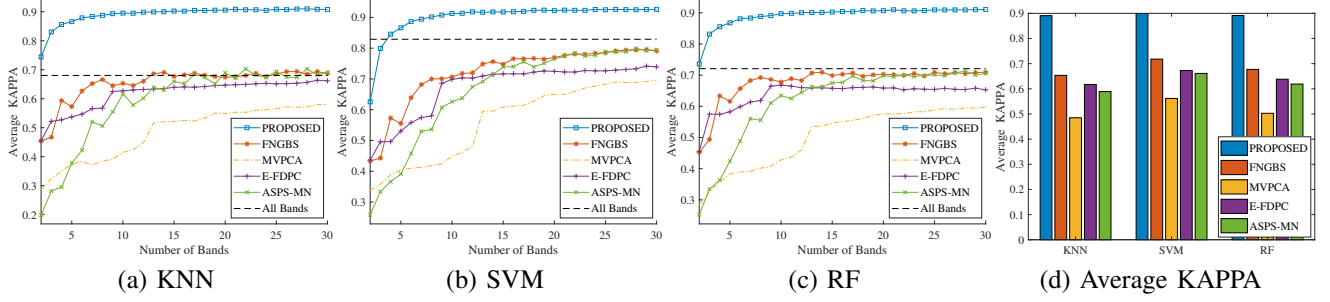


Fig. 5: The KAPPA coefficient results of SS-EMT and four non-EA-based FS methods on the Indian Pines dataset

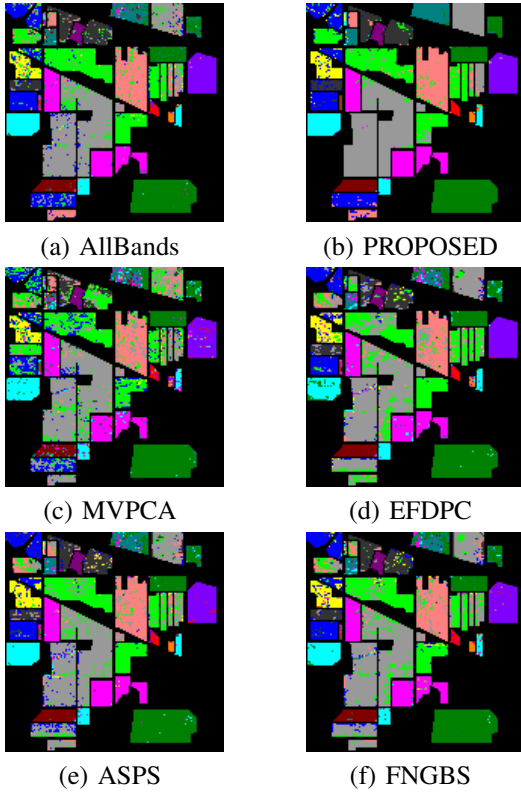


Fig. 6: The classification maps of SS-EMT and four non-EA-based FS methods on the Indian Pines dataset

a subset of 15 bands, while the classification results of the compared algorithms are only between 58% and 72% under the same situation. In addition, Figs. 4(b)-(c) provide the OA results of the compared algorithms under the SVM and RF classifiers, respectively, which also show similar results. Finally, as shown in Fig. 4(d), the average OA of SS-EMT

is much higher than that of the others. In summary, Fig. 4 confirms that SS-EMT outperforms these four competitors in terms of OA results under KNN, SVM and RF classifiers.

Furthermore, Fig. 5 provides the results of the KAPPA coefficients for each compared algorithm on the Indian Pines dataset under three different classifiers, which further validates the superiority of our proposed SS-EMT compared to these four adopted non-EA-based FS methods.

2) *Results on the Pavia University dataset:* Due to page limitations, the OA results of each algorithm on the Pavia University dataset are provided in Fig. S-2 of the **Supplementary Material**. From these figures, some conclusions can be drawn as follows. First, as shown in Fig. S-2 (a), which provides the OA results under the KNN classifier, SS-EMT achieves the best classification results in 25 out of all 29 cases compared to the four non-EA-based FS methods. SS-EMT shows better classification results than the baseline approach in all cases, except for the case where the numbers of selected bands are set from 2 to 5. However, the other compared algorithms failed to achieve higher classification accuracy than the baseline approach in most cases, among which only E-FDPC, ASPS-MN and FNGBS outperformed the baseline approach for the selected number of bands of 10, 24 and 24, respectively. Third, the classification results of SS-EMT can exceed 95% with only 15 bands, while the classification results of the other compared algorithms are only between 85% and 89% in the same case. Moreover, Figs. S-2 (b)-(c) provide the OA results of each compared algorithm under the SVM and RF classifiers, which also show the superiority of our proposed SS-EMT. Finally, as shown in Fig. S-2 (d), the average OA of SS-EMT is much higher than that of the others. In summary, Fig. S-2 verifies that the OA results of our proposed SS-EMT under three classifiers are significantly better than those of its four competitors.

Furthermore, Fig. S-3 of the **Supplementary Material** provides the KAPPA coefficient results for each compared



algorithm on the Pavia University dataset, which further validates the superiority of our proposed SS-EMT compared to these four adopted non-EA-based FS methods.

3) *Results on the Salinas dataset:* Due to page limitations, the OA results of each algorithm on the Salinas dataset are provided in Fig. S-4 of the **Supplementary Material**. Some conclusions can be drawn from these figures. First, as shown in Figs. S-4 (a)–(c) with the OA results obtained under three classifiers, SS-EMT achieves the best results on all subsets of band sizes 2 to 30, which are significantly better than these competitors. Second, SS-EMT achieves almost 100% satisfactory classification results on subsets of band sizes from 4 to 30, while the compared algorithms fail to outperform the baseline approach (All Bands) even on the 30-band subset. Unlike the Pavia University dataset, SS-EMT also shows good performance on smaller band subsets. Finally, as shown in Fig. S-4 (d), the average OA of SS-EMT is much higher than that of the others. In summary, Fig. S-4 verifies that the proposed SS-EMT outperforms these four competitors in terms of the OA results obtained under the KNN, SVM and RF classifiers.

In addition, Fig. S-5 of the **Supplementary Material** shows the results of the KAPPA coefficients for each of the compared algorithms on the Salinas dataset under three different classifiers, which further confirms the superiority of our proposed approach compared to these four non-EA-based FS methods.

4) *Results on the Botswana dataset:* Due to page limitations, the experimental results and corresponding discussions on the Botswana dataset are presented in Section S-III.B of the **Supplementary Material**.

Furthermore, for a more intuitive understanding of the classification results on the four datasets, classification maps for each compared algorithm are provided in Fig. 6 and Figs. S-14–S-16 of the **Supplementary Material** under the SVM classifier using a feature subset of size 20 bands. As shown in Fig. 6, the classification results of the baseline approach and of the other compared algorithms have pixels with incorrect colors in each class block. Our proposed SS-EMT presents correct colors perfectly in most of the class blocks except for three class blocks (i.e., Corn-Notill, Corn-Min and Corn). The possible reason for this is that these three classes are very similar and difficult to distinguish from each other. Moreover, a similar phenomenon can be observed in Figs. S-14–S-16. In conclusion, these classification results on different HSI datasets show that the classification maps of SS-EMT are much clearer than those of the other compared algorithms.

Overall, the above extensive experimental results on the four adopted HSI datasets under three different classifiers validate the clear advantages of our proposed SS-EMT compared to four non-EA-based FS methods.

#### F. Comparisons with EA-based FS Methods

In this section, four EA-based FS methods were selected for performance comparisons, including MOBS [21], SSGA [42], MTPSO [34] and MBBS-VC [25]. It should be noted that MTPSO was originally proposed as a supervised method for high-dimensional medical FS problems, which is modified in this paper to solve unsupervised FS problems of HSIs, where

information entropy is used to measure the importance of each band, while all the other strategies remain the same as in their corresponding original reference [34]. In addition, in order to ensure the fairness of the experiments, the population size and the maximum number of iterations of the other compared algorithms were set the same as those of SS-EMT.

1) *Results on the Indian Pines dataset:* The OA results for the Indian Pines dataset are shown in Fig. 7. From these figures, some conclusions can be drawn. First, SS-EMT achieves the best results in all 29 cases (i.e., on the subset of 2–30 bands) under three different classifiers, and its classification accuracy is significantly higher than that of the other compared algorithms. Second, SS-EMT outperforms the baseline approach in most cases, while all of the compared algorithms failed to outperform the baseline approach even on the subset of 30 bands. Third, the classification results of SS-EMT increase smoothly with the increased subset band size and tend to converge. However, most of the classification results of the other compared algorithms fluctuate and even decrease with an increasing subset band size. The reason behind this is that the compared algorithms conduct FS on the raw HSI or treat the original HSI as a whole, which makes it difficult to obtain a high-quality subset of bands with high information and low redundancy. In contrast, our proposed SS-EMT adopts the superpixel segmentation technique to partition the raw HSI into several superpixel blocks, which can greatly reduce the complexity of the FS problem. More importantly, SS-EMT uses the EMT algorithm to search the optimal subset of bands for different feature classes, thus greatly improving the classification performance. In addition, the average OA results in Fig. 7 (d) further validate that SS-EMT has a strong competitive advantage over the compared algorithms.

Furthermore, Fig. 8 provides the KAPPA coefficient results for each compared algorithm on the Indian Pines dataset, which also verifies that our proposed SS-EMT is superior to these four competitive EA-based FS methods.

2) *Results on the Pavia University dataset:* Fig. S-8 of the **Supplementary Material** provides the OA results of the compared algorithms on the Pavia University dataset. Some conclusions can be drawn from these figures. First, Fig. S-8 (a) provides the OA results of the compared algorithms under the KNN classifier, showing that SS-EMT obtains the best results in 24 out of 29 cases, while it shows poor classification results on a subset of 2–3 bands and is slightly inferior to SSGA on a subset of 4–5 bands. Second, Figs. S-8 (b) and (c) provide the OA results of the compared algorithms under SVM and RF classifiers, respectively, which show similar phenomena. That is, the advantage of SS-EMT becomes more obvious as the size of the band subset increases. Third, as shown in Fig. S-8 (d), the average OA results obtained by SS-EMT are much higher than those of the other compared algorithms.

Furthermore, Fig. S-9 of the **Supplementary Material** provides the results of the KAPPA coefficients for each compared algorithm on the Pavia University dataset under three different classifiers, which further validates the superiority of our proposed SS-EMT over these four EA-based FS methods.

3) *Results on the Salinas dataset:* The results of the OA for the Salinas dataset are shown in Fig. S-10 of the **Sup-**

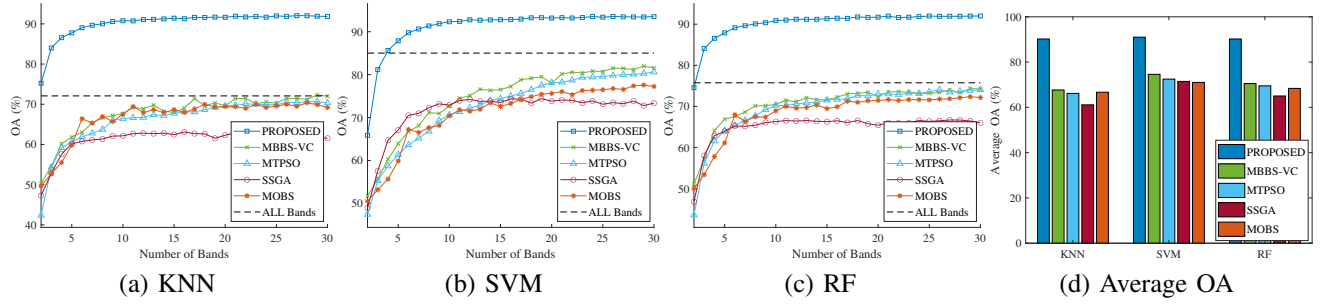


Fig. 7: The OA results of SS-EMT and four EA-based FS methods on the Indian Pines dataset

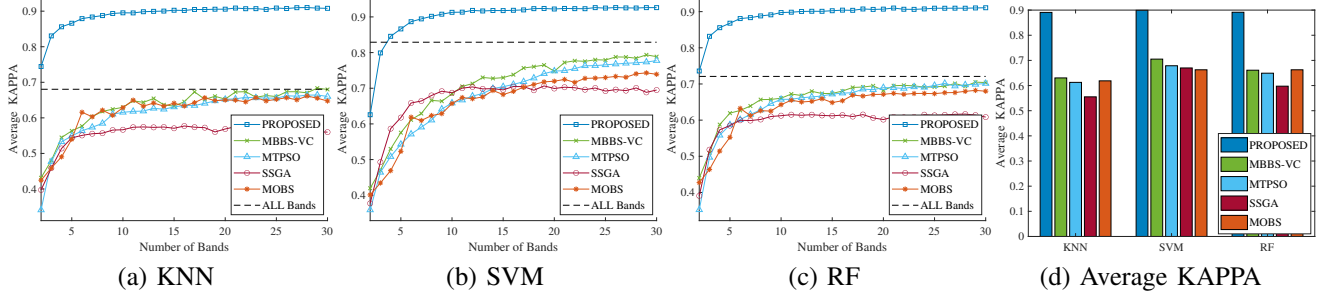


Fig. 8: The KAPPA coefficient results of SS-EMT and four EA-based FS methods on the Indian Pines dataset

**plementary Material.** Some conclusions can be drawn from these figures. First, compared to the other four EA-based FS methods, SS-EMT achieves the best results under all three classifiers on subsets of band sizes 2-30. Second, SS-EMT outperforms the baseline on subsets of band sizes 3 to 30, while all of the compared algorithms fail to outperform the baseline even on a subset of 30 bands. In addition, the average OA results are shown in Fig. S-10, which clearly shows that SS-EMT has a strong competitive advantage over these compared algorithms.

Furthermore, Fig. S-11 of the **Supplementary Material** shows the results of the KAPPA coefficients for each of the compared algorithms on the Salinas dataset under three different classifiers, further confirming the superiority of our proposed SS-EMT over these four EA-based FS methods.

4) *Results on the Botswana dataset:* Due to page limitations, the detailed experimental results and corresponding discussions on the Botswana dataset are provided in Section S-III.C of the **Supplementary Material**.

Additionally, in order to provide a more intuitive view of the classification results on the four datasets, the classification maps of each EA-based FS methods are provided in Figs. S-17 to S-20 of the **Supplementary Material**, which are the classification results under the SVM classifier using a feature subset of 20-band size. As shown in these figures, for these compared algorithms, there are many misclassified pixel points in the region of the feature class, which rarely occurs in the classification map of SS-EMT. The reason behind this is that these compared algorithms tend to select bands for all feature classes, which results in poor differentiation between feature classes, so it is easy to classify some pixel points into similar but incorrect feature classes. In contrast, our proposed SS-EMT selects bands for different feature classes separately, which can effectively reduce this phenomenon and can significantly improve the classification accuracy. Therefore, Figs.

S-17 to S-20 show that the classification map of SS-EMT is much clearer than that of other compared algorithms.

In conclusion, SS-EMT achieves good classification results across different HSI datasets and significantly outperforms other state-of-the-art EA-based FS methods.

#### G. Comparisons of Execution Efficiency

The execution time is also an important metric for evaluating the efficiency of FS methods. Thus, we further investigated the execution efficiency of SS-EMT and other FS methods included in our experimental study. Table V provides the average CPU running time of SS-EMT and other compared methods over 20 independent runs. Some conclusions can be derived from the results listed in Table V. First, the non-EA-based FS methods have a significantly lower computational cost with respect to the EA-based and EMT-based FS methods on the four adopted HSI datasets, with FNGBS showing the lowest running time. The advantage of the execution efficiency of non-EA-based FS methods is because they do not need to iteratively explore the optimal subset of features. Second, our proposed SS-EMT achieves acceptable execution efficiency compared to EA-based and EMT-based FS methods, as the running time of SS-EMT is much lower than that of the other compared methods except for MOBS. Specifically, MOBS has the lowest execution time among all EA-based and EMT-based FS methods because it uses only a single metric (i.e., information entropy) to evaluate the quality of features, yet this also restricts its classification performance. In addition, our proposed SS-EMT shows higher execution efficiency compared to two EMT-based FS methods (i.e., MTPSO and MBBS-VC), which is mainly attributed to the designed efficient individual evaluation mechanism. In conclusion, SS-EMT not only has obvious advantages in classification accuracy, but is also competitive in terms of execution efficiency.

TABLE V: Comparison of average running time (s)

Types	Algorithm	IP	PU	SA	BW
Non-EA-based	MVPCA	0.64	9.27	9.11	0.19
	EFDPC	0.70	3.69	3.76	0.16
	ASPS_MN	1.91	7.74	7.45	11.29
	FNGBS	<b>0.15</b>	<b>0.05</b>	<b>0.11</b>	<b>0.07</b>
EA-based	MOBS	33.20	31.83	42.19	64.35
	SSGA	1678.68	5073.16	6778.37	522.72
EMT-based	MTPSO	90.77	676.05	2461.78	674.70
	MBBS-VC	341.31	2104.63	2061.47	8081.37
	SS-EMT	321.37	105.50	442.64	336.21

### H. Ablation Experiments and Analysis

In this section, and in order to validate the effectiveness of two important components contained in SS-EMT (i.e., knowledge transfer and similarity factor), two variants of SS-EMT are implemented in our ablation experiments, introduced as follows:

- **SS-EMT-NKT**: To validate the effectiveness of the knowledge transfer strategy in SS-EMT, we designed the variant SS-EMT-NKT without performing knowledge transfer among different subtasks.
- **SS-EMT-RKT**: To validate the effectiveness of the similarity factors used in SS-EMT, we designed the variant SS-EMT-RKT, which randomly transfers the knowledge among different subtasks rather than adopting the knowledge transfer strategy based on the similarity factor.

The ablation experimental results on the IP dataset are summarized in Table VI. The results and their corresponding discussion are as follows:

- 1) **The effect of knowledge transfer in SS-EMT**: As shown in Table VI, the classification accuracy of SS-EMT is higher than that of SS-EMT-NKT under all the three classifiers on the IP dataset. Particularly under the KNN classifier, the accuracy of SS-EMT is about 1.27% higher than that of SS-EMT-NKT. In conclusion, the comparison between SS-EMT and SS-EMT-NKT shows that knowledge transfer is effective by sharing useful knowledge between different tasks.
- 2) **The effect of the similarity factor in SS-EMT**: Some observations can be derived from the results reported in Table VI. First, the accuracy of SS-EMT-RKT is slightly higher than that of SS-EMT-NKT, which illustrates the effectiveness and the necessity of knowledge transfer. Second, the accuracy of SS-EMT-RKT is much lower than that of our proposed SS-EMT that performs knowledge transfer based on the similarity factor, which shows the effectiveness of this mechanism. Our experimental results show that the selection of source knowledge domains has a great influence on the effectiveness of knowledge transfer, and that knowledge transfer between similar subtasks can greatly promote the optimization of target subtasks and vice versa.

To conclude, the above ablation experimental results further validate the effectiveness of two important components contained in our proposed SS-EMT, i.e., the knowledge transfer and the similarity factor.

## V. CONCLUSIONS AND FUTURE WORK

In this study, a superpixel segmentation based EMT algorithm was proposed for FS of HSIs. Unlike most existing proposals that treat HSI as a whole and perform FS on all feature classes, our proposed approach considers the fact that different feature classes have different feature preferences, which adopts the ERS method to partition the original HSI into several superpixel blocks and then optimizes them as multiple subtasks. This way, the FS problem of HSIs is transformed into a multitasking optimization problem. To effectively solve the previously formulated problem, a PSO-based EMT algorithm is designed, which can explore the optimal subset of features for each subtask by transferring useful knowledge among subtasks belonging to the same or similar feature classes. In addition, a novel individual evaluation mechanism is proposed to obtain multiple high quality subsets of features with different numbers of features in a single run. Finally, extensive experiments validate that our proposed method outperforms several state-of-the-art non-EA-based and EA-based FS methods on four common HSI datasets under three classifiers.

It should be noted that, to the best of the authors' knowledge, this work is the first attempt to combine the superpixel segmentation approach and EMT technique to solve the FS problem of HSIs. The work described in this article makes it possible to obtain smaller, higher quality subsets of bands. This allows the model to capture more accurate spectral information at a lower computational cost, helping to optimize real-world scenarios such as crop management, precise medical diagnostics and accurate geological resource detection. Although our approach greatly improves classification performance, there are still many issues that deserve further investigation. First, it is necessary to investigate how to reduce the computational cost of band subset evaluation. Second, we envision a general and robust superpixel segmentation-based EMT framework for FS of HSI, in which most existing superpixel segmentation techniques and EAs can be embedded and this is also part of our future research work.

## REFERENCES

- [1] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens Environ.*, vol. 113, pp. S110–S122, 2009.
- [2] S. Karim, A. Qadir, U. Farooq, M. Shakir, and A. A. Laghari, "Hyperspectral imaging: A review and trends towards medical imaging," *Curr Med Imaging*, 2022, doi:10.2174/1573405618666220519144358.
- [3] N. Hanson, T. Kelestemur, J. Berman, D. Ritzenhoff, and T. Padir, "Hyperbot-a benchmarking testbed for acquisition of robot-centric hyperspectral scene and in-hand object data," in *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2022, pp. 1–5, doi:10.1109/WHISPERS56178.2022.9955133.
- [4] A. Nisha and A. Anitha, "Current advances in hyperspectral remote sensing in urban planning," in *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*. IEEE, 2022, pp. 94–98.
- [5] M. B. Stuart, A. J. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: a review of recent developments and technological advances in compact field deployable systems," *Sensors*, vol. 19, no. 14, p. 3071, 2019.

TABLE VI: Comparisons between SS-EMT and its two variants on the IP dataset.

Classifier	Algorithm	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	MEAN
KNN	SS-EMT	<b>75.22</b>	<b>86.57</b>	<b>89.05</b>	<b>90.02</b>	90.77	<b>91.07</b>	<b>91.23</b>	<b>91.28</b>	<b>91.58</b>	<b>91.63</b>	<b>91.73</b>	<b>91.66</b>	91.87	<b>92.03</b>	91.84	<b>89.84</b>
	SS-EMT-NKT	70.67	85.54	87.48	88.99	89.19	89.63	90.20	90.10	90.69	91.35	90.86	90.44	91.69	90.65	91.01	88.57
	SS-EMT-RKT	62.51	85.23	88.05	89.97	<b>90.96</b>	91.03	90.96	91.23	90.90	91.47	91.54	91.15	<b>91.88</b>	91.90	<b>92.07</b>	88.72
SVM	SS-EMT	65.86	<b>85.61</b>	<b>89.76</b>	<b>91.34</b>	<b>92.32</b>	<b>92.78</b>	92.75	<b>92.82</b>	<b>93.25</b>	<b>93.18</b>	93.20	<b>93.54</b>	<b>93.55</b>	<b>93.44</b>	<b>93.51</b>	<b>90.46</b>
	SS-EMT-NKT	<b>70.67</b>	83.00	88.76	90.20	91.49	92.21	92.10	92.20	92.62	91.91	92.16	92.59	92.96	93.29	92.96	89.94
	SS-EMT-RKT	65.44	85.38	89.66	90.14	91.91	92.43	<b>92.80</b>	92.12	92.41	92.86	<b>93.46</b>	93.10	92.92	93.24	93.06	90.06
RF	SS-EMT	<b>74.58</b>	<b>86.55</b>	<b>89.13</b>	90.09	90.87	<b>91.21</b>	91.19	<b>91.46</b>	<b>91.76</b>	<b>91.71</b>	<b>91.62</b>	<b>91.85</b>	<b>91.91</b>	91.92	92.01	<b>89.86</b>
	SS-EMT-NKT	71.92	84.26	87.71	88.97	89.98	89.15	90.42	90.93	91.05	91.35	90.97	91.74	91.75	91.47	91.65	88.89
	SS-EMT-RKT	64.57	85.90	88.94	<b>90.19</b>	<b>90.96</b>	91.03	<b>91.35</b>	91.23	91.16	91.57	91.53	91.46	91.86	<b>92.07</b>	<b>92.02</b>	89.06

- [6] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–16, 2021.
- [7] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans Geosci Remote Sens*, vol. 58, no. 8, pp. 5336–5353, 2020.
- [8] Y. Wan, A. Ma, Y. Zhong, X. Hu, and L. Zhang, "Multiobjective hyperspectral feature selection based on discrete sine cosine algorithm," *IEEE Trans Geosci Remote Sens*, vol. 58, no. 5, pp. 3601–3618, 2020.
- [9] Y. Wang, Q. Zhu, H. Ma, and H. Yu, "A hybrid gray wolf optimizer for hyperspectral image band selection," *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–13, 2022.
- [10] A. Paul and N. Chaki, "Supervised data-driven approach for hyperspectral band selection using quantization," *Geocarto Int*, vol. 37, no. 8, pp. 2312–2322, 2022.
- [11] W. Feng, Y. Quan, G. Dauphin, Q. Li, L. Gao, W. Huang, J. Xia, W. Zhu, and M. Xing, "Semi-supervised rotation forest based on ensemble margin theory for the classification of hyperspectral image with limited training data," *Inf Sci*, vol. 575, pp. 611–638, 2021.
- [12] R. Aghaee, M. Momeni, and P. Moallem, "Semisupervised band selection from hyperspectral images using levy flight-based genetic algorithm," *IEEE Geosci Remote Sens Lett*, vol. 19, pp. 1–5, 2022.
- [13] W. Chen, Z. Yang, J. Ren, J. Cao, N. Cai, H. Zhao, and P. Yuen, "Mimn-dpp: Maximum-information and minimum-noise determinantal point processes for unsupervised hyperspectral band selection," *Pattern Recognit*, vol. 102, p. 107213, 2020.
- [14] Z. Yong, H. Chun-lin, S. Xian-fang, and S. Xiao-yan, "A multi-strategy integrated multi-objective artificial bee colony for unsupervised band selection of hyperspectral images," *Swarm Evol. Comput.*, vol. 60, p. 100806, 2021.
- [15] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 54, no. 1, pp. 88–102, 2015.
- [16] D. Varade, A. K. Maurya, and O. Dikshit, "Unsupervised hyperspectral band selection using ranking based on a denoising error matching approach," *Int J Remote Sens*, vol. 40, no. 20, pp. 8031–8053, 2019.
- [17] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei, "A similarity-based ranking method for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 59, no. 11, pp. 9585–9599, 2021.
- [18] Y. Liu, H. Xie, L. Wang, and K. Tan, "Hyperspectral band selection based on a variable precision neighborhood rough set," *Applied Optics*, vol. 55, no. 3, pp. 462–472, 2016.
- [19] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 56, no. 10, pp. 5910–5922, 2018.
- [20] W. Sun, G. Yang, J. Peng, X. Meng, K. He, W. Li, H.-C. Li, and Q. Du, "A multiscale spectral features graph fusion method for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–12, 2021.
- [21] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans Geosci Remote Sens*, vol. 54, no. 1, pp. 544–557, 2015.
- [22] M. Zhang, M. Gong, and Y. Chan, "Hyperspectral band selection based on multi-objective optimization with high information and low redundancy," *Appl Soft Comput*, vol. 70, pp. 604–621, 2018.
- [23] J. Tschannerl, J. Ren, P. Yuen, G. Sun, H. Zhao, Z. Yang, Z. Wang, and S. Marshall, "Mimn-dgsa: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm," *Inf Fusion*, vol. 51, pp. 189–200, 2019.
- [24] K. He, W. Sun, G. Yang, X. Meng, K. Ren, J. Peng, and Q. Du, "A dual global-local attention network for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–13, 2022.
- [25] C. He, Y. Zhang, D. Gong, X. Song, and X. Sun, "A multi-task bee colony band selection algorithm with variable-size clustering for hyperspectral images," *IEEE Trans Evol Comput*, vol. 26, no. 6, pp. 1566–1580, 2022.
- [26] A. Paul and N. Chaki, "Band selection using spectral and spatial information in particle swarm optimization for hyperspectral image classification," *Soft Computing*, vol. 26, no. 6, pp. 2819–2834, 2022.
- [27] A. R. Chowdhury, J. Hazra, K. Dasgupta, and P. Dutta, "Fuzzy rule-based hyperspectral band selection algorithm with ant colony optimization," *Innov Syst Softw Eng*, pp. 1–14, 2022.
- [28] C. Chen, Y. Wan, A. Ma, L. Zhang, and Y. Zhong, "A decomposition-based multiobjective clonal selection algorithm for hyperspectral image feature selection," *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–16, 2022, doi:10.1109/TGRS.2022.3216685.
- [29] S. Sawant and P. Manoharan, "Hyperspectral band selection based on metaheuristic optimization approach," *Infrared Physics & Technology*, vol. 107, p. 103295, 2020.
- [30] B. L. Phaneendra Kumar and P. Manoharan, "Whale optimization-based band selection technique for hyperspectral image classification," *Int J Remote Sens*, vol. 42, no. 13, pp. 5105–5143, 2021.
- [31] S. Liu, Q. Lin, L. Feng, K.-C. Wong, and K. C. Tan, "Evolutionary multitasking for large-scale multiobjective optimization," *IEEE Trans Evol Comput*, 2022, doi:10.1109/TEVC.2022.3166482.
- [32] K. Qiao, K. Yu, B. Qu, J. Liang, H. Song, C. Yue, H. Lin, and K. C. Tan, "Dynamic auxiliary task-based evolutionary multitasking for constrained multi-objective optimization," *IEEE Trans Evol Comput*, 2022, doi:10.1109/TEVC.2022.3175065.
- [33] L. Li, M. Xuan, Q. Lin, M. Jiang, Z. Ming, and K. C. Tan, "An evolutionary multitasking algorithm with multiple filtering for high-dimensional feature selection," *IEEE Trans Evol Comput*, 2023, doi:10.1109/TEVC.2023.3254155.
- [34] K. Chen, B. Xue, M. Zhang, and F. Zhou, "Evolutionary multitasking for feature selection in high-dimensional classification via particle swarm optimization," *IEEE Trans Evol Comput*, vol. 26, no. 3, pp. 446–460, 2021.
- [35] E. Kalhor and B. Bakhtiari, "Speaker independent feature selection for speech emotion recognition: A multi-task approach," *Multimed Tools Appl*, vol. 80, no. 6, pp. 8127–8146, 2021.
- [36] J. Shi, X. Zhang, X. Liu, Y. Lei, and G. Jeon, "Multicriteria semi-supervised hyperspectral band selection based on evolutionary multitask optimization," *Knowl Based Syst*, vol. 240, p. 107934, 2022.
- [37] Y. Zhang, Y. Ma, X. Dai, H. Li, X. Mei, and J. Ma, "Locality-constrained sparse representation for hyperspectral image classification," *Inf Sci*, vol. 546, pp. 858–870, 2021.
- [38] W. Ma, H. Ma, H. Zhu, Y. Li, L. Li, L. Jiao, and B. Hou, "Hyperspectral image classification based on spatial and spectral kernels generation network," *Inf Sci*, vol. 578, pp. 435–456, 2021.
- [39] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans Geosci Remote Sens*, vol. 37, no. 6, pp. 2631–2641, 1999.
- [40] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 12, no. 12, pp. 4940–4950, 2019.
- [41] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans Geosci Remote Sens*, vol. 59, no. 6, pp. 5028–5039, 2020.
- [42] H. Zhao, L. Bruzzone, R. Guan, F. Zhou, and C. Yang, "Spectral-spatial genetic algorithm-based unsupervised band selection for hyperspectral

image classification,” *IEEE Trans Geosci Remote Sens*, vol. 59, no. 11, pp. 9616–9632, 2021.

- [43] S. Liu, H. Wang, W. Peng, and W. Yao, “A surrogate-assisted evolutionary feature selection algorithm with parallel random grouping for high-dimensional classification,” *IEEE Trans Evol Comput*, vol. 26, no. 5, pp. 1087–1101, 2022.
- [44] W. Sun, G. Yang, J. Peng, and Q. Du, “Hyperspectral band selection using weighted kernel regularization,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 12, no. 9, pp. 3665–3676, 2019.
- [45] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [46] W. Sun, J. Peng, G. Yang, and Q. Du, “Correntropy-based sparse spectral clustering for hyperspectral band selection,” *IEEE Geosci Remote Sens Lett*, vol. 17, no. 3, pp. 484–488, 2019.
- [47] H. Ji, Z. Zuo, and Q.-L. Han, “A divisive hierarchical clustering approach to hyperspectral band selection,” *IEEE Trans Instrum Meas*, vol. 71, pp. 1–12, 2022.
- [48] W. Sun, J. Peng, G. Yang, and Q. Du, “Fast and latent low-rank subspace clustering for hyperspectral band selection,” *IEEE Trans Geosci Remote Sens*, vol. 58, no. 6, pp. 3906–3915, 2020.
- [49] S. Sawant and P. Manoharan, “A hybrid optimization approach for hyperspectral band selection based on wind driven optimization and modified cuckoo search optimization,” *Multimed Tools Appl*, vol. 80, pp. 1725–1748, 2021.
- [50] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, “Entropy rate superpixel segmentation,” in *CVPR 2011*. IEEE, 2011, pp. 2097–2104.
- [51] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95-international conference on neural networks*, vol. 4. IEEE, 1995, pp. 1942–1948.
- [52] K. Chen, B. Xue, M. Zhang, and F. Zhou, “An evolutionary multitasking-based feature selection method for high-dimensional classification,” *IEEE Trans Cybern*, vol. 52, no. 7, pp. 7172–7186, 2020.
- [53] X. Wu, W. Wang, H. Yang, H. Han, and J. Qiao, “Diversified knowledge transfer strategy for multitasking particle swarm optimization,” *IEEE Trans Cybern*, 2023, doi:10.1109/TCYB.2022.3232113.
- [54] X. Song, Y. Zhang, D. Gong, H. Liu, and W. Zhang, “Surrogate sample-assisted particle swarm optimization for feature selection on high-dimensional data,” *IEEE Trans Evol Comput*, 2022, doi:10.1109/TEVC.2022.3175226.
- [55] X.-f. Song, Y. Zhang, D.-w. Gong, and X.-y. Sun, “Feature selection using bare-bones particle swarm optimization with mutual information,” *Pattern Recognit*, vol. 112, p. 107804, 2021.
- [56] M. Song, S. Liu, D. Xu, and H. Yu, “Multiobjective optimization-based hyperspectral band selection for target detection,” *IEEE Trans Geosci Remote Sens*, vol. 60, pp. 1–22, 2022.
- [57] J. Cohen, “A coefficient of agreement for nominal scales,” *Educ Psychol Meas*, vol. 20, no. 1, pp. 37–46, 1960.



**Lingjie Li** received his B.S. degree from Shandong Technology and Business University, Yantai, China in 2017, and his M.S. degree and Ph.D. degree from the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China in 2020 and 2023, respectively.

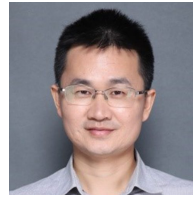
He is currently an Associate Researcher at the Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University. His current research interests are in evolutionary optimization and its applications, including multi-

objective optimization, large-scale optimization, hyperspectral remote sensing images, feature selection and cloud computing.



**Yuze Zhang** received the B.S. degree from Hebei University of Engineering, Handan, China in 2021. He is currently pursuing his Ph.D degree in College of Computer Science and Software Engineering, Shenzhen University.

His current research interests include evolutionary computation and its application to computer vision, feature selection and hyperspectral remote sensing images.



**Qiuzhen Lin** (Member IEEE) received the B.S. degree from Zhaoqing University and the M.S. degree from Shenzhen University, China, in 2007 and 2010, respectively. He received the Ph.D. degree from Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, in 2014.

He is currently an associate professor in College of Computer Science and Software Engineering, Shenzhen University. He has published over sixty research papers since 2008. His current research

interests include artificial immune system, multi-objective optimization, and dynamic system.



**Zhong Ming** received the Ph.D. degree in Computer Science and Technology from Sun Yat-sen University, Guangzhou, China, in 2003.

He is currently the Executive Director of the Graduate School of Shenzhen University, and a Professor with the National Engineering Laboratory for Big Data System Computing Technology and the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include software engineering and artificial intelligence. He has published more

than 200 refereed international conference and journal papers (including 40+ ACM/IEEE Transactions papers). He was the recipient of the ACM TiS 2016 Best Paper Award and some other best paper awards. His research interests include software engineering and artificial intelligence.



**Carlos A. Coello Coello** (Fellow, IEEE) received the Ph.D. degree in computer science from Tulane University, New Orleans, LA, USA, in 1996.

He is a Professor (CINVESTAV-3F Researcher) with the Department of Computer Science of CINVESTAV-IPN, Mexico City, Mexico. He has authored and coauthored over 470 technical papers and book chapters. He has also coauthored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Second Edition, Springer, 2007). His publications currently report over 72 3800 citations in

Google Scholar (his H-index is 104). His research interests include evolutionary multiobjective optimization and constraint-handling techniques for evolutionary algorithms.



**Victor C. M. Leung** (Life Fellow, IEEE) received the B.A.Sc. degree (Hons.) in electrical engineering and the Ph.D. degree in electrical engineering (through the Canadian Natural Sciences and Engineering Research Council Postgraduate Scholarship) from the University of British Columbia, Vancouver, BC, Canada, in 1977 and 1982, respectively.

He is currently a Distinguished Professor of Computer Science and Software Engineering with Shenzhen University, Shenzhen, China. He is also an Emeritus Professor of Electrical and Computer

Engineering and the Director of the Laboratory for Wireless Networks and Mobile Systems, The University of British Columbia (UBC), Vancouver, BC, Canada. He is named in the current Clarivate Analytics list of “Highly Cited Researchers.” His research interests include wireless networks and mobile systems. He has published widely in these areas.