# Self-Representation difference matrix graph convolutional network for hyperspectral image classification

Shengnan Ding[1], Fuguo Liu[2], Yufeng Shi[1,3]
{1519439152@qq.com[1], lfg53880@cjc.edu.cn[2], yfshi@sdu.edu.cn[1,3]}
Institute for Financial Studies & School of Mathematics, Shandong University, Jinan 250100, P. R. China[1]
School of Mathematics and Data Sciences, Changji University, Changji 831100, Xinjiang, China[2]
Institute for Financial Studies & School of Mathematics, Shandong University, Jinan 250100, P. R. China[3]

**Abstract.** Hyperspectral Image (HSIs) classification is a research problem that utilizes spectral information across different bands to identify surface objects. Due to the continuity of spectral bands, the correlation between adjacent spectral is significantly higher than non-adjacent bands. To avoid the redundancy of similar information, this paper proposes a methodology to selection spectral bands through the construction of a graph convolutional network with a band difference matrix and a sparse self-representation model. Firstly, for homogeneous areas, entropy rate segmentation is employed to obtain the profile information of the target, and segment HSIs into several patches based on profile information. Within each patch, a spectral difference matrix is constructed using the spectral reflectance of different targets. A small difference matrix means that the band is not conducive for distinguishing different targets. Secondly, the difference matrix is regarded as nodes in the graph convolution, and the relationship between nodes represent the correlation degree between different bands. Lastly, an optimization model is constructed using the band self representation model and sparse constraints on weights to determine the optimal subset of spectral bands. The effectiveness of the model is validated on multiple data sets. Experimental results show that the proposed method surpasses other comparative methods. Moreover, the classification maps derived from the experiments can more effectively discern the differences in surface coverage.

**Keywords:** Hyperspectral image, band selection, sparse constraint, self representation, difference matrix.

## 1 Introduction

Hyperspectral imaging (HSI) provides hundreds of contiguous spectral bands per pixel, enabling fine-grained analysis for agriculture, mineral exploration, and urban mapping. Yet the high dimensionality coupled with scarce labels often leads to the well-known Hughes phenomenon, i.e., performance peaking and then degrading as dimensionality grows, which motivates dimensionality reduction as a principled step in HSI pipelines [1, 2]. Recent surveys further show a clear shift from

IIKI 2025, December 18-21, Chengdu, People's Republic of China
Copyright © 2026 EAI
DOI 10.4108/eai.18-12-2025.2365302

traditional machine learning to deep architectures (CNNs, Transformers), with growing interest in efficient state-space models—while concerns remain about computation, interpretability, and cross-scene robustness [3, 4]. In parallel, sustainability has become a priority: methods that reduce input bands and parameters align with *Green AI* goals [5].

Two families dominate dimensionality reduction: feature extraction (e.g., PCA, LDA, manifold learning such as LPP) and band selection (BS). The former constructs compact latent spaces but sacrifices the physical meaning of spectral channels; the latter retains original bands, benefiting interpretability and downstream physical analysis [6, 7, 8]. Deep BS has advanced via attention/reconstruction (e.g., BS-Nets), reinforcement learning (DRL), and graph neural networks (GCN), alongside classical unsupervised ranking/clustering and wrapper/filter/embedding strategies [9, 10, 11, 12]. Nevertheless, key issues persist: (i) large models demand heavy computation and labels; (ii) region-level homogeneity is often under-modeled or handled at a single granularity; and (iii) the interpretability of selection mechanisms lags behind black-box classifiers [4]. Meanwhile, Transformer-style HSI classifiers (e.g., SpectralFormer and efficient variants) achieve strong accuracy but further amplify (i) and (iii), reinforcing the value of compact, interpretable BS [13, 14].

To address these gaps, we propose a Self-Representation Difference Matrix Graph Convolutional Network (SDM-GCN) for interpretable band selection and HSI classification. First, we partition images into homogeneous regions via entropy-rate superpixel segmentation (ERS) [15] to capture spatially consistent spectra. A Grouping Strategy via inter-Band Differences (GSBD) then coarsely screens bands with high information and low redundancy. Next, a sparse self-representation reconstructs inter-band relations, and a GCN—treating bands as nodes—refines the similarity structure [16]. This region-aware + sparsity-driven design contrasts with prior graph/self-representation BS approaches such as EGCSR, MGSR, and TGSR by jointly leveraging (a) region-level cues from ERS, (b) explicit inter-band difference priors, and (c) lightweight graph reasoning for selection, yielding a more interpretable and efficient pipeline [17, 18, 19]. Compared with attention-based or DRL BS, our framework reduces dependency on heavy parametric modules and reward design, while remaining compatible with modern classifiers (e.g., CNN/Transformer backbones) as a front-end selector [9, 10, 11].

**Contributions.**

- **Interpretable, region-aware BS**: We integrate ERS-based homogeneous region modeling with inter-band difference screening and sparse self-representation + GCN selection, improving the physical interpretability of chosen bands [15].
- **Efficiency & sustainability**: By selecting compact, non-redundant bands and a light GCN head, we cut data/parameter budgets in line with Green AI, without altering downstream classifiers [5].
- **Robust benchmarks & generalization**: On standard HSI benchmarks (Indian Pines, Pavia University, Salinas) and newer datasets (e.g., WHU-Hi, Houston 2018 GRSS DFC), we achieve competitive accuracy with fewer bands, highlighting cross-scene and few-shot potential [20, 21, 22].

## 2 Method

This section introduces the proposed SDM-GCN for HSI band selection, whose workflow mainly includes two modules: (i) *Grouping Strategy via Band Difference* (GSBD) to obtain discriminative candidate bands, and (ii) *Band Self-Representation* (BSR) to learn sparse band weights under a band-similarity graph. Given an HSI cube $X \in \mathbb{R}^{W\times H\times B}$ with $N = W \times H$ pixels and $B$ bands, the objective is to select a compact band subset that preserves inter-class discriminability while reducing redundancy.

### 2.1 Grouping Strategy via Band Difference (GSBD)

We first reshape the cube into a spectral matrix $X \in \mathbb{R}^{N\times B}$ and apply Entropy Rate Superpixel Segmentation (ERS) to partition the scene into $S$ regions:

$$X = \bigcup_{s=1}^{S} \mathscr{H}_s, \quad \text{s.t. } \mathscr{H}_s \cap \mathscr{H}_g = \emptyset \ (s \neq g), \tag{1}$$

where $\mathscr{H}_s$ denotes the $s$-th superpixel. With the label map $Y \in \mathbb{R}^{W\times H}$ containing $K$ classes, pixels in each region are further grouped by class to form class-wise spectral patches. To handle size inconsistency across regions/classes, all patches are resized by bilinear interpolation to a unified spatial size, yielding matrices $H_{s,k} \in \mathbb{R}^{N_k\times B}$ $(k = 1,\ldots,K)$.

**Band difference statistics.** To roughly locate bands that best distinguish different targets, we construct difference matrices between two classes. Let $\mathscr{S}_i = \{H_i^{(p)}\}_{p=1}^{m}$ and $\mathscr{S}_j = \{H_j^{(q)}\}_{q=1}^{n}$ be the patch sets of classes $i$ and $j$. For each pair, we compute

$$DM_{p,q} = H_i^{(p)} - H_j^{(q)} \in \mathbb{R}^{N_k\times B}, \tag{2}$$

and evaluate the per-band variance of each $DM_{p,q}$. Averaging over all $m \times n$ pairs produces an average variance vector $DB_{i,j} \in \mathbb{R}^{1\times B}$, where larger entries indicate bands with larger reflectance differences between the two classes. Repeating for all class pairs yields a collection $DB_S$. According to the required number of bands $c$, we select the top-ranked bands from these vectors and merge them to form a candidate band set $B_S = \{B_1,\ldots,B_c\}$.

### 2.2 Band Similarity with Graph Convolutional Networks

To capture nonlinear correlations among bands, we build a band graph where each band corresponds to a node. The adjacency matrix $A$ is defined by an RBF kernel:

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{\|x_i - x_j\|^2}{2\sigma^2}\right), & x_i, x_j \in \mathscr{H}_s, \\ 0, & x_i \in \mathscr{H}_s,\ x_j \in \mathscr{H}_g,\ s \neq g, \end{cases} \tag{3}$$

where $x_i$ is the feature vector of the $i$-th band. With self-loops $\tilde{A} = A + I$ and $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, we use the normalized matrix

$$\hat{A} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}. \tag{4}$$

A GCN layer updates band embeddings as

$$H^{(l+1)} = \sigma(\hat{A}H^{(l)}W^{(l)}), \tag{5}$$

where $W^{(l)}$ are trainable parameters. Since band-level labels are unavailable, pseudo-labels are assigned according to the GSBD-derived band groups (e.g., bands selected from different $DB$ vectors are treated as different pseudo-classes). The two-layer GCN output is then used to obtain a band similarity representation.

### 2.3 Band Self-Representation (BSR)

To enhance robustness and integrate spectral and graph information, we adopt a graph self-representation model. Given $X$ and $\hat{A}$, we learn a band-to-band coefficient matrix $W \in \mathbb{R}^{B\times B}$ by

$$\min_{W} \ \|X - \hat{A}XW\|_F^2 + \lambda \mathscr{R}(W), \tag{6}$$

where $\lambda$ controls the regularization strength. For band selection, we employ an $\ell_{2,1}$ norm to induce row sparsity:

$$\min_{W} \ \frac{1}{M}\sum_{m=1}^{M} \|X - \hat{A}\tilde{X}_m W\|_F^2 + \lambda \|W\|_{2,1}, \tag{7}$$

where $\tilde{X}_m$ denotes the $m$-th sample/patch and $M$ is the number of samples. After optimization, bands are ranked by row norms $\|\omega_i\|_2$ ($i = 1,\ldots,B$), and the top-$c$ bands are selected.

### 2.4 Complexity

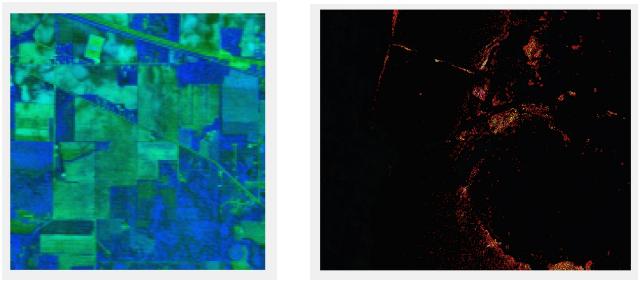
Let $S$ be the number of superpixels, $K$ the number of classes, and $B$ the number of bands. GSBD costs $O(SKB)$ to construct difference statistics. The BSR optimization on $X \in \mathbb{R}^{M\times N\times B}$ has dominant complexity $O(\max\{M,N\}^2 B)$. Therefore, the overall complexity is $O(SKB) + O(\max\{M,N\}^2 B)$, which is lower than $O(\max\{S,K,B\}^3)$ in typical settings.

## 3 Results

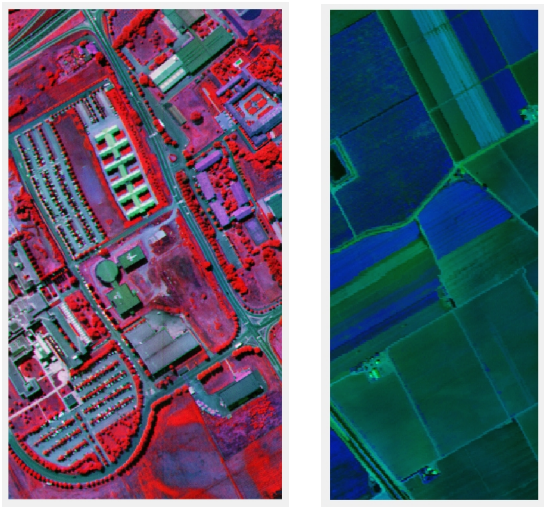
To verify the effectiveness of our method, this paper employs four HSIs data sets. These data sets originate from diverse hyperspectral applications (i.e., crop area mapping, vegetables in Salinas Valley, land, etc.) to assess the generalization capability of our method. In the following section, the data sets used are first described. Subsequently, the experimental setup is presented. Finally, the experimental results are discussed.

### 3.1 Datasets

In our experiment, five distinct HSIs data sets are employed to evaluate the effectiveness of the proposed model for band selection. Below, we provide a very comprehensive description of these specific data sets.



**(a)** Indian Pines **(b)** KSC



**(c)** PaviaU **(d)** Salinas

**Fig. 1.** False color renderings of diverse geographical regions.

(1) Indian Pines: This data set is a widely used benchmark HSIs captured in the Indian Pines test site in northwestern Indiana, USA. It consists of $145 \times 145$ pixels and contains 16 different classes, including various types of vegetation and crops. The image has 220 bands in the spectral range from 0.4 to 2.45 $\mu m$. After removing 20 water absorption bands, 200 bands were used in the experiments.

(2) Kennedy Space Center (KSC): This data set consists of images captured in 1996 at the Kennedy Space Center in Florida, USA. It includes $512 \times 614$ pixels with a spatial resolution of 18 meters per pixel and comprises 224 bands covering the spectral range from 0.4 to 2.5 $\mu m$. After eliminating bands affected by noise and atmospheric absorption, 176 bands were used in the experiments. It includes 13 different categories.

(3) University of Pavia (PaviaU): This data set was acquired using Reflective Optics System

Imaging Spectrometer (ROSIS) sensor. It consists of $610 \times 340$ pixels with a spatial resolution of 1.3 meters per pixel and comprises 103 bands covering spectral range from 0.43 to 0.86 $\mu m$. In experimental process, all bands were used, including 9 different categories.

(4) Salinas: This data set provides images covering the Salinas Valley area in California, USA. The Salinas images were captured using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It includes $512 \times 217$ pixels with a spatial resolution of 3.7 meters per pixel and contains 16 different sample classes. After removing 20 bands related to water absorption, 224 bands with different spectral resolutions were used in subsequent analysis.

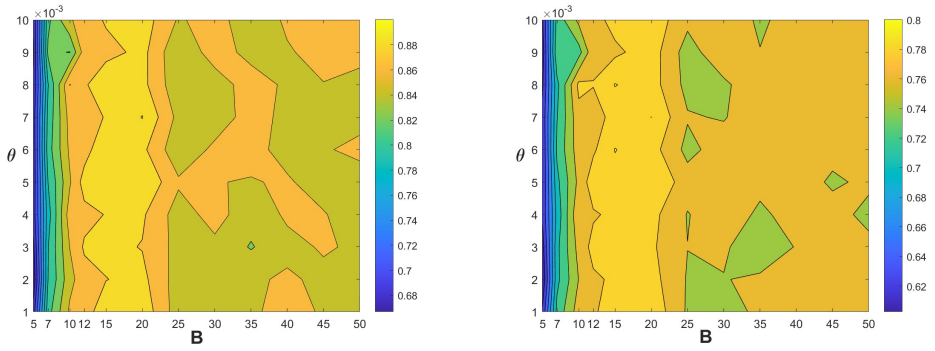
### 3.2 Experimental Setup

(1) Evaluation Metrics: In order to quantitatively assess classification performance of all research methods, three popular and widely used metrics were adopted, namely Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient. OA measures percentage of total correctly identified pixels. AA represents the average percentage of correctly identified pixels for each land cover class. Kappa coefficient calculates the percentage of identified pixels corrected by the number of agreements that would be expected purely by chance.

(2) Competitive Approaches: In this study, six state-of-the-art band selection approaches are adopted for comparison, including: MGSR [23], RDGSR [24], HLFC [25], GRSC [26], ASPS-MN [27], and LvaHAI [28]. MGSR [23] is a method based on edge-preserving graph self-representation. This method first represents the boundary information of each region, constructs a self-representation structural graph using segmented boundaries, and performs graph convolution. bands are then sorted based on the convolution results, retaining the top-ranked bands. RDGSR [24] is an optimization of MGSR [23]. It enhances MGSR [23] by incorporating a structural graph about bands, merging the band structural graph with the pixel structural graph to obtain a Dual Graph Convolution structure. bands are sorted based on the coefficient matrix concerning bands, yielding the required number of bands. GRSC [26] is a graph convolution and clustering-based algorithm. Initially, HIS is segmented into several sub-regions based on target positions using ERS and PCA. Then, a region similarity graph is constructed based on pixel relationships, and optimal bands are selected through graph convolution. The data is then reconstructed based on the selected bands, and clustering is performed to obtain the final required number of bands. HLFC [25] is a clustering algorithm based solely on spatial positional relationships. Similar to GRSC, it first segments HIS into several sub-regions based on target positions, recording the position information of each region. The position information is combined with spectral information, and clustering is performed on the mixed data, selecting bands with high discrimination. ASPS-MN [27] presents a new Adaptive Subspace Partition Strategy for band selection in HSIs. LvaHAI [28] are multi-objective optimization methods. Because the reflection degrees of ground targets for similar spectra are close, these methods partition the HSIs into several subparts according to bands and establish optimization functions within these subparts. These optimization problems are then integrated to obtain the required number of bands.

To ensure fairness, all band selection methods maintain consistent band numbers, starting from 5 bands and incrementally increasing to 7, 10, 12, 15, 20, 25, 30, 35, 40, 45, and 50 bands. The effectiveness and superiority of the methods are comprehensively compared under different band numbers. Additionally, all experimental data are divided into 30% testing and 70% training sets.

SVM and KNN classifiers are employed to evaluate classification accuracy.




**(a)** Effect to Kappa coefficients **(b)** Effect to classification accuracy

**Fig. 2.** Effect of different number bands and variance on model's kappa coefficient and classification accuracy.

**Table 1:** Classification accuracy (%) based on Indian Pines data set: using 10% of samples from each class as training set. Black bold font denotes optimal classification accuracy, while underlining indicates a slightly worse accuracy.

| | 5 | 7 | 10 | 12 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MGSR [23] | 0.1429 | 0.1575 | 0.1725 | 0.1786 | 0.1980 | 0.2700 | 0.2970 | 0.3593 | 0.4220 | 0.4464 | 0.4492 | 0.4748 |
| RDGSR [24] | 0.4692 | 0.5782 | 0.5941 | 0.5908 | 0.6161 | 0.6380 | 0.6345 | 0.6408 | 0.6356 | 0.6308 | 0.6455 | 0.6471 |
| HLFC [25] | 0.1910 | 0.1869 | 0.1688 | 0.2083 | 0.2153 | 0.1870 | 0.2199 | 0.2180 | 0.2028 | 0.2237 | 0.2031 | 0.2152 |
| GRSC [26] | 0.1953 | 0.1831 | 0.1553 | 0.1293 | 0.1806 | 0.1618 | 0.1749 | 0.1928 | 0.1740 | 0.1389 | 0.1705 | 0.1685 |
| ASPS-MN [27] | 0.6322 | **0.6893** | 0.6255 | 0.6382 | 0.6505 | 0.6404 | 0.6411 | 0.6560 | 0.6450 | 0.6328 | 0.6522 | 0.6391 |
| LvaHAl [28] | 0.4727 | 0.5699 | 0.5072 | 0.6290 | 0.5962 | 0.6036 | 0.6315 | 0.6255 | 0.6252 | 0.6540 | 0.6411 | 0.6304 |
| FNGBS [29] | 0.3857 | 0.5542 | 0.6045 | 0.5877 | 0.6035 | 0.5463 | 0.6044 | 0.5861 | 0.6120 | 0.6240 | 0.6054 | 0.5832 |
| SDM-GCN | **0.6310** | 0.6360 | **0.6990** | **0.7152** | **0.7073** | **0.6982** | **0.7148** | **0.7124** | **0.7111** | **0.7084** | **0.7151** | **0.7227** |

**Table 2:** Classification accuracy (%) based on KSC data set: using 10% of samples from each class as training set. Black bold font denotes optimal classification accuracy, while underlining indicates a slightly worse accuracy.

| | 5 | 7 | 10 | 12 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MGSR [23] | 0.4790 | 0.5223 | 0.5385 | 0.5558 | 0.5960 | 0.6052 | 0.6987 | 0.7117 | 0.7254 | 0.7301 | 0.7397 | 0.7363 |
| RDGSR [24] | 0.7023 | 0.7282 | 0.7485 | 0.7538 | 0.7538 | 0.7455 | 0.7833 | 0.7707 | 0.7760 | 0.7974 | 0.7922 | 0.7865 |
| HLFC [25] | 0.2336 | 0.2287 | 0.2464 | 0.2349 | 0.2228 | 0.2249 | 0.2316 | 0.2280 | 0.2499 | 0.2291 | 0.2268 | 0.2282 |
| GRSC [26] | 0.1685 | 0.2068 | 0.2605 | 0.2139 | 0.1831 | 0.2449 | 0.1980 | 0.1901 | 0.2051 | 0.1957 | 0.2047 | 0.2289 |
| ASPS-MN [27] | 0.7239 | 0.5336 | 0.5373 | 0.6857 | 0.7794 | 0.7805 | 0.7826 | 0.7897 | 0.7886 | 0.7901 | 0.7963 | 0.7713 |
| LvaHAl [28] | **0.8048** | 0.7732 | **0.8309** | 0.7895 | **0.8405** | 0.8121 | **0.8371** | 0.8087 | **0.8260** | 0.8061 | 0.8234 | 0.7948 |
| FNGBS [29] | 0.7523 | 0.7422 | 0.6432 | 0.7654 | 0.7135 | 0.7536 | 0.7630 | 0.7752 | 0.7512 | 0.7821 | 0.7922 | 0.8071 |
| SDM-GCN | 0.7549 | **0.7837** | 0.7860 | **0.8213** | 0.8198 | **0.8298** | 0.8255 | **0.8319** | 0.8247 | **0.8247** | **0.8240** | **0.8328** |

**Table 3:** Classification accuracy (%) based on PaviaU data set: using 10% of samples from each class as training set. Black bold font denotes optimal classification accuracy, while underlining indicates a slightly worse accuracy.

| | 5 | 7 | 10 | 12 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MGSR [23] | 0.7410 | 0.7568 | 0.7822 | 0.8138 | 0.8337 | 0.8414 | 0.8669 | 0.8646 | 0.8402 | 0.8389 | 0.8348 | 0.8442 |
| RDGSR [24] | 0.7388 | 0.7592 | 0.8207 | 0.8295 | 0.8554 | 0.8521 | 0.8549 | 0.8610 | **0.8654** | 0.8705 | 0.8652 | **0.8735** |
| HLFC [25] | 0.5724 | 0.5587 | 0.5862 | 0.5429 | 0.5643 | 0.5374 | 0.5222 | 0.5807 | 0.5169 | 0.5083 | 0.5671 | 0.5458 |
| GRSC [26] | 0.4614 | 0.5820 | 0.4416 | 0.5709 | 0.5516 | 0.2417 | 0.3760 | 0.5885 | 0.3771 | 0.5864 | 0.5151 | 0.1756 |
| ASPS-MN [27] | 0.8044 | 0.8422 | 0.8513 | 0.8324 | 0.8571 | 0.8603 | 0.8628 | 0.8490 | 0.8484 | 0.8572 | 0.8564 | 0.8594 |
| LvaHAl [28] | 0.8025 | 0.8178 | 0.8463 | 0.8280 | 0.8357 | 0.8403 | 0.8462 | 0.8562 | 0.8573 | 0.8657 | 0.8635 | 0.8733 |
| FNGBS [29] | 0.8122 | 0.8142 | 0.8245 | 0.7977 | 0.8135 | 0.8263 | 0.8347 | 0.8214 | 0.8371 | 0.8065 | 0.8201 | 0.8532 |
| SDM-GCN | **0.8338** | **0.8330** | **0.8604** | **0.8598** | **0.8609** | **0.8532** | **0.8695** | **0.8773** | 0.8626 | **0.8717** | **0.8725** | 0.8695 |

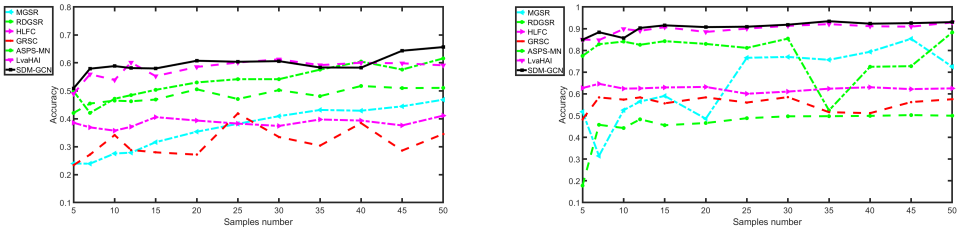
**Table 4:** Classification accuracy (%) based on Salinas pines data set: using 10% of samples from each class as training set. Black bold font denotes optimal classification accuracy, while underlining indicates a slightly worse accuracy.

| | 5 | 7 | 10 | 12 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MGSR [23] | 0.8001 | 0.8100 | 0.8221 | 0.8314 | 0.8361 | 0.8478 | 0.8494 | 0.8672 | 0.8672 | 0.8659 | 0.8688 | 0.8707 |
| RDGSR [24] | **0.8548** | 0.8584 | 0.8573 | 0.8630 | 0.8631 | 0.8715 | 0.8781 | 0.8801 | 0.8785 | 0.8784 | 0.8830 | 0.8830 |
| HLFC [25] | 0.3563 | 0.3676 | 0.3589 | 0.3609 | 0.3632 | 0.3701 | 0.3877 | 0.3939 | 0.3724 | 0.3653 | 0.3660 | 0.3622 |
| GRSC [26] | 0.3506 | 0.3720 | 0.3481 | 0.3254 | 0.3959 | 0.3438 | 0.3482 | 0.4425 | 0.3482 | 0.3590 | 0.3667 | 0.3779 |
| ASPS-MN [27] | 0.8391 | 0.8485 | 0.8592 | 0.8752 | 0.8679 | 0.8762 | 0.8782 | 0.8807 | 0.8819 | 0.8821 | 0.8834 | 0.8821 |
| LvaHAl [28] | 0.7037 | 0.7320 | 0.8254 | 0.8334 | 0.8648 | 0.8617 | 0.8595 | 0.8642 | 0.8625 | 0.8675 | 0.8709 | 0.8775 |
| FNGBS [29] | 0.7689 | 0.7856 | 0.7896 | 0.8024 | 0.8128 | 0.7936 | 0.7863 | 0.7928 | 0.8072 | 0.8110 | 0.7936 | 0.8230 |
| SDM-GCN | 0.8224 | **0.8664** | **0.8744** | **0.8767** | **0.8807** | **0.8769** | **0.8792** | **0.8817** | **0.8837** | **0.8862** | **0.8865** | **0.8835** |

### 3.3 Parameter Analysis

There are four parameters that need to be fixed, variance threshold between bands, required number of bands, smoothing parameter $\lambda$ when constructing the optimization objective function, and learning rate $\mu$ during the optimization process. This paper uses Indian Pines data set to verify impact of these four parameters on model. When analyzing selected number of bands and variance threshold, both $\lambda$ and $\mu$ are fixed at constant values of 0.1 and 0.9.

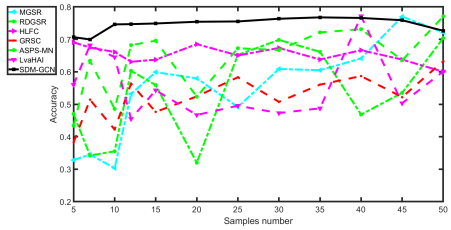
The effect of different number of bands and variance thresholds on the kappa coefficient of the model is clearly shown in Fig. 2. The kappa coefficient of the model is low when the number of bands is low. From the Fig. 2a, it can be found that when the number of bands is low, the kappa coefficient is low and the model does not show very good consistency, but when the number of bands is gradually increased to more than 10, the kappa coefficient appears to increase significantly. From Fig. 2b, the model was not significantly affected by the variance threshold when the number of bands was between 10 and 20 and had the highest consistency of all experiments. Taken together, the experimental results may indicate that the number of bands is less than 10, which results in insufficient model consistency due to the special nature of the hyperspectral spectrum and other reasons resulting in the inability to adequately capture information between targets. When the number of bands is between 10 and 20, the model can already adequately mine the relationship between targets. When the number of bands exceeds 20, redundancy between bands may have been generated, which is not



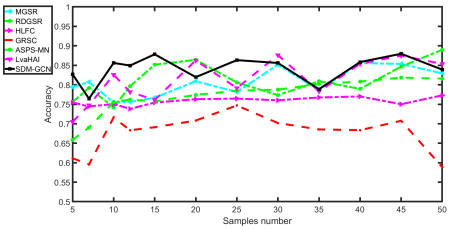

**(a)** Classification accuracy on Indian Pines




**(b)** Classification accuracy on KSC




**(c)** Classification accuracy on PaviaU




**(d)** Classification accuracy on Salinas

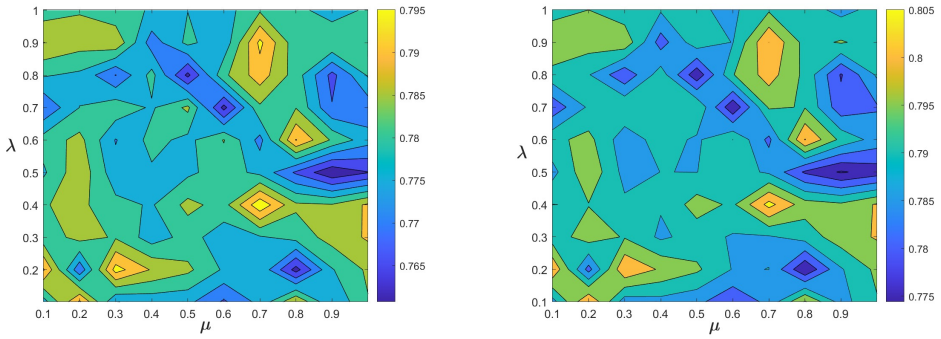
**Fig. 3.** Classification accuracy of SVM based on SDM-GCN and other state or art methods when different numbers of bands are selected on four data sets.

conducive to the improvement of model accuracy.

Similarly, when analyzing smoothing parameter $\Lambda$ and learning rate $\mu$, number of bands is fixed at 30 and variance threshold is fixed at 0.005. The following figure exemplifies impact that smoothing parameter $\lambda$ and learning rate $\mu$ have on kappa coefficients of model as well as the OA. Fig. 4 demonstrates effect of smoothing parameter $\lambda$ and learning rate $\mu$ on kappa coefficient. From Fig. 4a, it can be found that values of Kappa coefficient are on the whole relatively average, and OA values remain between 0.76 and 0.78, except for a few regions. This indicates that $\lambda$ and learning rate $\mu$ have less impact on consistency of model results in this paper, and OA values are kept at a relatively high level with good consistency. Similarly, it can be found in Fig. 4b that OA are relatively close to each other and remain between 0.77-0.8 except for a small part of the region. This can also indicate that model's accuracy is not drastically affected by parameter changes and has a relatively stable nature. Comparing two graphs again can also be found that some regions in the graphs have relatively similar locations, indicating that there is a high degree of consistency while model achieves higher accuracy. Therefore, in order to achieve better accuracy, parameter settings of $\mu = 0.7$ and $\lambda = 0.9$ are used in this paper.

### 3.4 Classification Results

This paper will illustrate the classification performance of the model from two aspects: classification accuracy and the selected number of bands. In order to demonstrate the robustness of the




**(a)** Effect to Kappa coefficients

**(b)** Effect to classification accuracy

**Fig. 4.** Effect of different parameter $\lambda$ and $\mu$ on model's kappa coefficient and classification accuracy.

model and eliminate the influence of different classifiers on classification accuracy, a comparative experiment was conducted using two classifiers, SVM and KNN.

In summary, the SDM-GCN algorithm presented in this paper has achieved commendable classification performance across various data sets. Particularly in scenarios with a limited number of bands, SDM-GCN demonstrated higher classification accuracy compared to other competing methods, further affirming the overall effectiveness of the algorithm. However, it should also be noted that the performance of the algorithm may exhibit some instability in situations with small samples and a high number of bands. Future research should focus on addressing these issues to enhance the robustness and applicability of the algorithm.

## 4 Discussion

The results provide meaningful insights into hyperspectral band selection and HSI classification. Overall, SDM-GCN shows consistently strong performance across different scenes, with particularly clear advantages when only a small number of bands are retained. This suggests that explicitly exploiting inter-class spectral differences is a practical and effective principle for selecting informative and non-redundant bands.

The performance gain can be explained by the complementary contributions of each module. Entropy rate segmentation first partitions HSIs into homogeneous regions, which stabilizes local spectral statistics and enables patch-level spectral–spatial modeling. Based on these regions, the proposed intra-group difference matrix highlights bands where reflectance discrepancies between targets are more pronounced, providing a discriminative prior before deeper modeling. The subsequent graph convolution further captures nonlinear relationships among bands beyond simple distance measures, refining correlation modeling under complex spectral structures. Finally, sparse self-representation learning improves compactness and robustness: the $\ell_{2,1}$-induced row sparsity naturally suppresses redundant bands and favors a compact subset that preserves discriminability.

We also observe that scene characteristics influence the downstream classification behavior. When targets are spatially concentrated and class structures are relatively regular, distance-based decision rules tend to benefit more from well-selected bands, and performance typically saturates quickly as band number increases. In scenes with sparse targets, complex background, or weak class boundaries, classification becomes more sensitive to noise and decision margins, and the benefit of band selection mainly comes from improved separability under moderate band numbers rather than simply increasing dimensionality.

Despite these promising outcomes, SDM-GCN may exhibit fluctuations in settings with limited labeled samples, strong noise, or highly sparse targets. This is likely due to less reliable estimation of band relationships and the re-emergence of redundancy as the selected band number grows. Moreover, the current framework focuses on patch-level band selection and does not explicitly incorporate fine-grained pixel-level discriminative representations, which may limit performance in highly mixed regions or near class boundaries.

Future work will focus on improving stability under small-sample and noisy conditions, for example by incorporating uncertainty-aware pseudo-labeling or robust losses. Integrating lightweight pixel-level spatial encoders (e.g., CNN-based modules) may further enhance boundary discrimination and reduce sensitivity to complex backgrounds. In summary, SDM-GCN offers an effective band selection strategy by combining difference-guided discriminability, graph-based correlation modeling, and sparse self-representation, and it is particularly suitable for low-band HSI classification scenarios.

## 5 Conclusions

This paper proposes SDM-GCN, a band selection framework for hyperspectral image (HSI) classification that integrates intra-group spectral differences, graph convolution, and sparse self-representation. First, the entropy rate superpixel (ERS) method partitions the HSI into $S$ homogeneous regions $\{\mathscr{H}_s\}_{s=1}^{S}$, and each region is further separated into class-specific sub-regions. Based on these regions, we construct an intra-group difference matrix to emphasize inter-class spectral discrepancies. Treating bands (or difference-guided band representations) as nodes, a graph convolutional network (GCN) is used to model nonlinear band correlations and produce a similarity representation. Finally, a sparse self-representation model with row-sparsity regularization identifies a compact subset of informative bands: rows with larger (non-zero) weights correspond to selected bands, while near-zero rows indicate redundant bands to be removed.

Experiments on multiple benchmark datasets show that classification performance generally improves as more bands are selected, confirming that multi-band spectral information provides richer discriminative cues. More importantly, SDM-GCN achieves competitive or superior accuracy under low-band settings, demonstrating its ability to retain highly discriminative bands while reducing redundancy.

The main advantages of SDM-GCN are summarized as follows:

1. **Difference-guided discriminability:** The intra-group difference matrix explicitly amplifies spectral separability among targets, facilitating subsequent similarity learning and reducing the impact of redundant adjacent bands.

2. **Unsupervised sparse selection:** Sparse self-representation enables band selection without requiring band-level labels, producing a compact subset and maintaining strong performance even with relatively few bands.

3. **Nonlinear correlation modeling:** GCN effectively captures non-Euclidean and nonlinear relationships among bands, providing a principled way to learn band similarity from difference-aware representations.

Despite these benefits, two limitations remain:

1. **Stability under small samples and noise:** Performance may fluctuate when training data are limited or when targets are sparse and background/noise is strong. Improving robustness in such settings is an important direction for future work.

2. **Limited pixel-level modeling:** The current design emphasizes band selection with region-level cues and does not explicitly incorporate fine-grained pixel-level features. Future work will integrate pixel-level representations to further improve discrimination, especially near class boundaries.

## Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2023YFA1008903), the Major Fundamental Research Project of Shandong Province of China (Grant No. ZR2023ZD33), Taishan Scholar Project of Shandong Province of China (Grant tstp20240803), the Natural Science Foundation of Shandong Province (Grant No. ZR2023LLZ012), the Key Project of Shandong Provincial Key R&D Program (Soft Science) (Grant No. 2024RZB0204).

## Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

## References

[1] Hughes GF. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory. 1968;14(1):55-63.

[2] Li J, Bioucas-Dias J, Plaza A. Dimensionality reduction of hyperspectral images with graph-based methods. IEEE Transactions on Geoscience and Remote Sensing. 2020;58(3):1969-85.

[3] Ahmad A, Zhang Y, Li C, et al. Hyperspectral image analysis: From shallow learning to Transformers and Mamba. Neurocomputing. 2025;574:127103.

[4] Wu X, Zhang L, Li W, et al. Graph neural networks for hyperspectral image classification: A comprehensive review. ISPRS Journal of Photogrammetry and Remote Sensing. 2024;209:181-200.

[5] Chen Y, Yang J, Xu K. Toward sustainable and green AI in remote sensing. Remote Sensing of Environment. 2024;300:113977.

[6] Jolliffe IT. Principal Component Analysis. 2nd ed. Springer; 2002.

[7] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):711-20.

[8] He X, Niyogi P. Locality Preserving Projections. In: Proceedings of the 17th International Conference on Neural Information Processing Systems (NeurIPS); 2004. p. 153-60.

[9] Zhong Z, Li J, Luo W, Chapman M. Band selection by end-to-end BS-Nets. IEEE Transactions on Geoscience and Remote Sensing. 2018;56(9):5376-89.

[10] Mou L, Gao F, Zhu XX. Reinforcement learning for band selection in hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing. 2021;59(8):7081-94.

[11] Mou L, Zhang X, Zhu XX. Multi-agent reinforcement learning for interpretable band selection. IEEE Transactions on Geoscience and Remote Sensing. 2024. Early Access.

[12] Wang Y, Li W, Du Q. Graph-based band selection for hyperspectral image classification. Pattern Recognition. 2023;137:109281.

[13] Hong D, Gao L, Yao J, Zhang B, Plaza A, Chanussot J. SpectralFormer: Transformer-based hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2021;59(9):7839-55.

[14] Cao J, et al. EfficientFormer for hyperspectral image classification. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2022. p. 3012-5.

[15] Liu MY, Tuzel O, Ramalingam S, Chellappa R. Entropy Rate Superpixel Segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2011. p. 2097-104.

[16] Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. In: International Conference on Learning Representations (ICLR); 2017. Available from: `https://arxiv.org/abs/1609.02907`.

[17] Fu Y, Zheng X, Zhang Y. Efficient graph-constrained self-representation for hyperspectral band selection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2021;14:6432-45.

[18] Li X, Zhou Y, Wang R. Tensor graph band selection for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1-13.

[19] Zhang H, Huang J, Li W. Multi-graph self-representation for hyperspectral band selection. Information Sciences. 2023;619:401-15.

[20] Zhu L, Li W, Tao R, Plaza A. Domain generalization for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2023;61:1-13.

[21] Yokoya N, Ghamisi P, Xia J, et al. 2018 IEEE GRSS Data Fusion Contest: Multisource and hyperspectral Houston data. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2018. p. 1-4.

[22] Zhou J, Xu Y, Zhang J, et al. WHU-Hi: UAV-borne hyperspectral benchmark dataset. ISPRS Journal of Photogrammetry and Remote Sensing. 2021;175:236-48.

[23] Zhang Y, Wang X, Jiang X, Zhou Y. Marginalized Graph Self-Representation for Unsupervised Hyperspectral Band Selection. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1-12.

[24] Zhang Y, Wang X, Jiang X, Zhou Y. Robust Dual Graph Self-Representation for Hyperspectral Band Selection. IEEE Transactions on Geoscience and Remote Sensing. 2022;60:1-13.

[25] Wang J, Tang C, Liu X, Zhang W, Li W, Zhu X, et al. Region-Aware Hierarchical Latent Feature Representation Learning-Guided Clustering for Hyperspectral Band Selection. IEEE Transactions on Cybernetics. 2022. Early Access.

[26] Wang J, Tang C, Zheng X, Liu X, Zhang W, Zhu E. Graph regularized spatial–spectral subspace clustering for hyperspectral band selection. Neural Networks. 2022;153:292-302.

[27] Wang Q, Li Q, Li X. Adaptive Subspace Partition Strategy for Hyperspectral Band Selection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2019;12(12):4940-50.

[28] Wei F, Cai L, Liao B, Lu T. Local-View-Assisted Discriminative Band Selection With Hypergraph Autolearning for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing. 2020;58(3):2042-55.

[29] Wang Q, Li Q, Li X. A Fast Neighborhood Grouping Method for Hyperspectral Band Selection. IEEE Transactions on Geoscience and Remote Sensing. 2021;59(6):5028-39.