

Hyperspectral Band Selection

A review

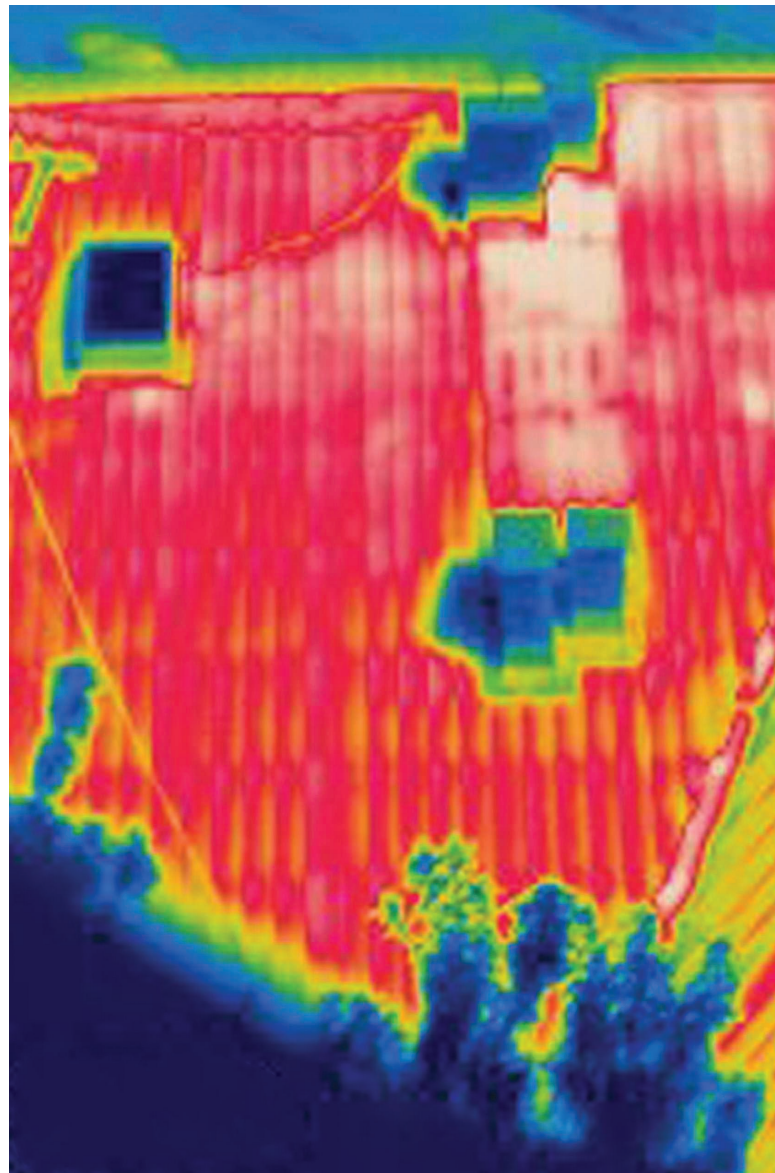
A hyperspectral imaging sensor collects detailed spectral responses from ground objects using hundreds of narrow bands; this technology is used in many real-world applications. Band selection aims to select a small subset of hyperspectral bands to remove spectral redundancy and reduce computational costs while preserving the significant spectral information of ground objects. In this article, we review current hyperspectral band selection methods, which can be classified into six main categories: ranking based, searching based, clustering based, sparsity based, embedding-learning based, and hybrid-scheme based. With two widely used hyperspectral data sets, we illustrate the classification performances of several popular band selection methods. The challenges and research directions of hyperspectral band selection are also discussed.

THE NECESSITY OF HYPERSPECTRAL BAND SELECTION

Due to the advancement of imaging spectrometer technology, a typical hyperspectral sensor now collects the spectral reflectance of ground objects on Earth's surface using hundreds of narrow bands [1], [2]. Because of high spectral resolution, the acquired hyperspectral imagery (HSI) can better differentiate ground objects using subtle spectral discrepancies. It greatly benefits many practical applications, e.g., biophysical parameter retrieval [3], [4], fine classification of complex environments [5], [6], small target detection [7], [8], and so forth. However, HSI produces a very large volume of data. For example, the Hyperspectral Infrared Imager launched by NASA's Jet Propulsion Laboratory (JPL) has the continuous averaged data rate of 65 Mb/s, which produces a data volume of 372 Gb/orbit and 5.2 Tb/day [9]. The resulting high computational cost makes fast on-board processing difficult. Its vast data volume also renders the common image processing algorithms designed for multispectral imagery ineffective. In particular, the high dimensionality of HSI data brings about the "curse of dimensionality" problem, i.e., under a fixed, small number of training samples, the classification accuracy of HSI data decreases when the dimensionality of

HSI data increases. In remote sensing applications, collecting a larger number of training samples is costly, time-consuming, and even impossible. Therefore, it is necessary to develop new dimensionality reduction methods and protocols for hyperspectral image processing [10]–[12].

Adjacent bands in HSI are highly correlated, and some spectral bands may not carry discriminant information in



a specific application [13], [14]. One of the most important preprocessing steps for HSI analysis is dimensionality reduction, which aims to remove spectral redundancy while preserving critical information for subsequent applications [15], [16]. Hyperspectral dimensionality reduction can be achieved through either band selection or feature extraction [17]. Feature extraction transforms the original data into another feature space with certain criteria [18], [19]. Most feature extraction methods linearly combine all of the original bands and may complicate the interpretation of results if the adopted criterion is inappropriate [20]. Band selection selects a desired band subset from the original bands and well preserves the spectral meaning of spectral channels [21]. In this article, we focus on hyperspectral band selection.

A variety of band-selection methods have been presented in the literature. However, to the best of our knowledge, there is no study that reviews the current research status

of hyperspectral band selection. In this article, we survey hyperspectral band selection methods in the literature and point out challenges and potential future directions.

REVIEW OF BAND SELECTION METHODS

Band selection methods can be categorized into six groups: 1) ranking-based methods, 2) searching-based methods, 3) clustering-based methods, 4) sparsity-based methods, 5) embedding learning-based methods, and 6) hybrid scheme-based methods. Let an HSI data $\mathbf{X} = \{\mathbf{x}_i\} \in \mathbf{R}^{N \times D}$ be represented as a collection of band vectors $\mathbf{B} = \{\mathbf{b}_j\} \in \mathbf{R}^{D \times N}$, where \mathbf{x}_i refers to the spectral signature of the i th pixel, D is the number of pixels, N is the number of bands with $N \ll D$, and \mathbf{b}_j corresponds to the j th band. In the supervised methods, a training set with n training samples is given as data set $\mathbf{X}_r = \{(\mathbf{x}_1, \gamma_1), \dots, (\mathbf{x}_n, \gamma_n)\}$, where \mathbf{x} and γ are a training sample and its class label in "1-of- C " encoding of the C classes, respectively. The problem with band selection is selecting m representative bands $\mathbf{M} = \mathbf{B}(:, \kappa) \in \mathbf{R}^{D \times m}$ to achieve better or at least as comparable a performance as using the original bands in a certain application, e.g., classification, target detection, and spectral unmixing. Table 1 shows the taxonomy of the hyperspectral band selection methods discussed in this article.

RANKING-BASED METHODS

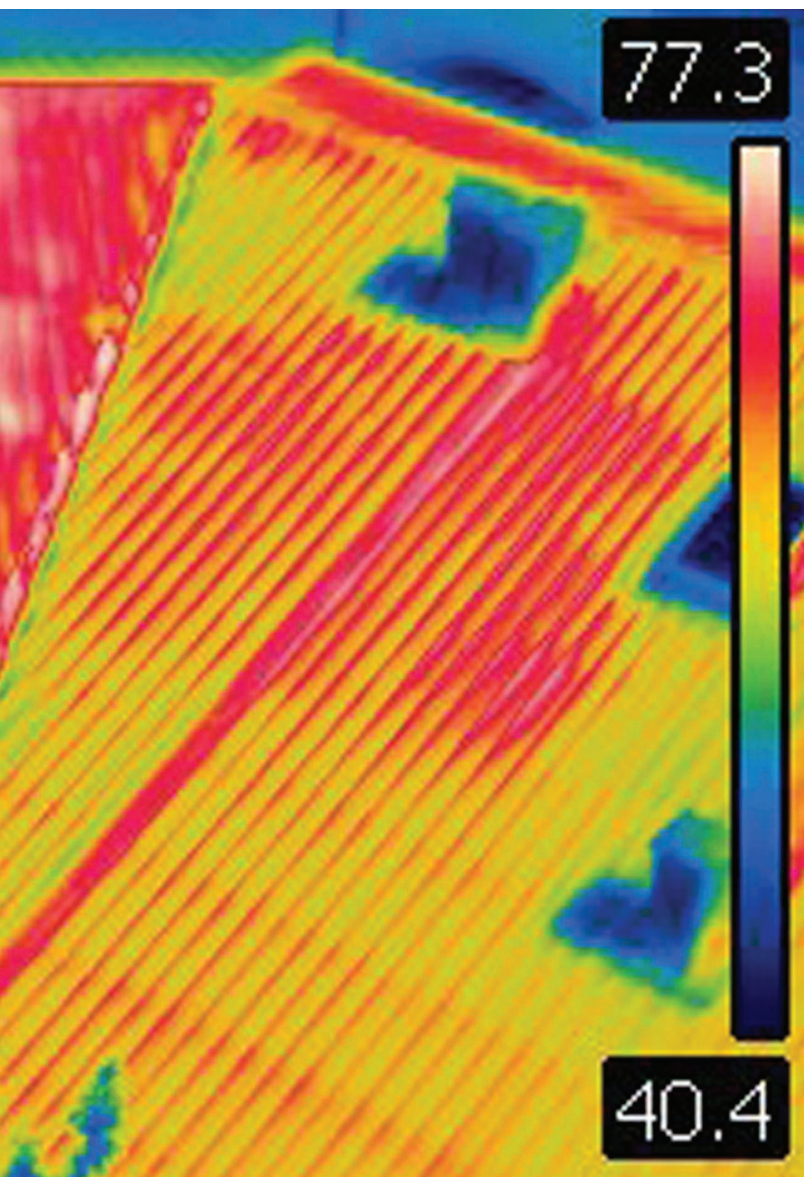
The ranking-based methods quantify the importance of each spectral band according to a pre-defined band-prioritization criterion and select the top-ranked bands in a sorted sequence. The ranking-based methods can be further divided into two types, i.e., unsupervised and supervised, according to whether labeled training samples are used or not.

UNSUPERVISED RANKING-BASED METHODS

Unsupervised criteria consider the information, dissimilarity, or correlation of bands. For instance, metrics such as variance, first spectral derivative, spectral ratio, contrast measurement, signal-to-noise ratio (SNR), third-order statistics (skewness), fourth-order statistics (kurtosis), k th order statistics, negentropy, entropy, and information divergence [22], [23] are often employed to prioritize bands.

HIGH-INFORMATION CRITERIA

The selected bands are required to have large information capacity. In [14], some classical information measurements, such as information entropy and information divergence, were used as ranking criteria to select informative bands. In [23], principal component analysis (PCA) was adopted to prioritize the energy or variances of band images, and two eigen analysis-based methods were presented, i.e., minimum-variance PCA (MVPCA)



©ISTOCKPHOTO.COM/MARCO PHOTO

TABLE 1. THE TAXONOMY OF HYPERSPECTRAL BAND SELECTION METHODS.

CATEGORY	BRIEF DESCRIPTION	SUBCATEGORY	TYPICAL EXAMPLES
A. Ranking based	Ranking-based methods quantify the importance of each band and choose the top-ranked bands.	A.1. Unsupervised ranking based	High-information criteria: MVPCA and MSNRPCA [23], information entropy [14], and the covariance-based method [24] Low-correlation criteria: CBS [25], CMBS [26], optimal band index [27], and curve fitting [28] Large-dissimilarity criteria: ECA [29], and E-FDPC [30]
		A.2. Supervised ranking based	Classification-aimed criteria: MMCA [23], MI [32], feature weighting [33], phase correlation [34], and NHMC [35] Spectral-unmixing-aimed criteria: OSP [23], VNVBS [36], ICA [37], and an alternative separability index [38]
B. Searching based	Searching-based methods search the aimed band subset, which can optimize a given criterion function.	B.1. Incremental searching based	Unsupervised-searching criteria: LP [45], spectral rhythm [46], and a geometry-based algorithm [47] Supervised-searching criteria: BAO, ADM, and MDM [48]; MEAC [49]; NPFS [50]; MMAIQ [51]; a simple, equal-width interval binning [53]; and VPNSR [54]
		B.2. Updated searching based	Classifier-independent criteria: CSFS [69], an immune clonal strategy [70], PSO [71], GA [72], TMI, and STMI [74] Classification-dependent criteria: gravitational search and BA [76], FODPSO [77], a probabilistic memetic algorithm [79], a metaheuristic algorithm [80], and a chaotic binary-coded gravitational search algorithm [80]
C. Clustering based	Clustering-based methods select representative bands from band clusters.	B.3. Eliminating searching based	SBS criteria: BSC CEM [81], VGBS [82], ACMBS [83], a minimum-noise band selection algorithm [84], and a joint-skewness band selection algorithm [85]
		C.1. <i>k</i> -means based	Unsupervised criteria: <i>k</i> -means [90], kernel <i>k</i> -means [92], and dual clustering [93] Supervised and semisupervised criteria: a prototype space algorithm [94], and semisupervised <i>k</i> -means [95]
D. Sparsity based	Sparsity-based methods formulate band selection into a sparsity-constrained optimization problem.	C.2. AP based	Unsupervised criteria: an exemplar-based AP [96], an FMAP [97], and a DFM [98] Supervised and semisupervised criteria: discriminative kernel clustering [99], semisupervised adaptive AP [100], and an NTMI [101]
		C.3. Graph based	Graph-based criteria: spectral clustering [102], an MDPP [104], graph clustering [105], dominant-set extraction [106], and a complex network [107] SNMF criteria: an SNMF [108], and an SNMF-EMD [109]
E. Embedding-learning based	Embedding learning-based methods optimize the objective function of a learning model while selecting proper bands.	D.1. SNMF based	Sparse representation criteria: SpaBS [110], multitask sparsity pursuit [111], a collaborative sparse model [112], ISSC [113], SWLRSC [114], FRSR [115], DWSSR [116], and SSR [117]
		D.2. Sparse representation based	Sparse regression criteria: A LASSO [118], a sparse HSIC and a surrogate-kernel-based algorithm [119], a group-sparsity-based algorithm [121], and a spectral-spatial hypergraph model [122]
F. Hybrid-scheme based	Hybrid scheme-based methods integrate two or more different schemes into select bands.	D.3. Sparse regression based	SVM-classifier criteria: An EFS, [123], an MR-SVM [124], and KFS [125] MLR-classifier criteria: MLR [127], and SMLR [128]
		E.1. Classifier learning based	Application model criteria: sparsity-promoting priors [130] and SCBS [131] Deep-learning model criteria: deep CNNs and distance density [132], segmented autoencoders [133], attention-based CNNs [134], and CM-CNNs [135]
F. Hybrid-scheme based	Hybrid scheme-based methods integrate two or more different schemes into select bands.	E.2. Other learning based	Combined ranking and clustering criteria: an SSI [136], PSO and IFBP [138], and manifold ranking [139] Combined ranking and searching criteria: conditional MI and GA [140], spatial clustering and GA [141], and minimum redundancy/maximum relevance [142]
		F.1. Hybrid scheme based	NOTE: SNR: signal-to-noise ratio; PCA: principal component analysis; MVPCA: minimum-variance PCA; MSNRPCA: maximum SNR PCA; CBS: constrained band selection; CMBS: constrained multiple-band selection; FDPC: fast density peak-based clustering; ECA: exemplar component analysis; E-FDPC: enhanced fast density peak-based clustering; MMCA: minimum misclassification canonical analysis; NHMC: nonhomogeneous hidden Markov chain; VNVBS: variable-number variable-band selection; OSP: orthogonal subspace projection; ICA: independent component analysis; MI: mutual information; TMI: trivariate MI; STMI: semisupervised TMI; NTMI: normalized TMI; LP: linear prediction; BAO: band add-on; ADM: average distance method; MDM: minimum distance method; MEAC: minimum estimation abundance covariance; NPFS: nonlinear parsimonious feature selection; MMAIQ: maximal minimal associated index; VPNSR: variable-precision neighborhood rough set; PSO: particle swarm optimization; GA: genetic algorithm; CSFS: clonal selection feature selection; BA: Bat algorithm; FODPSO: fractional-order Darwinian PSO; SBS: sequential backward selection; CEM: constrained energy minimization; BSC: band selection for CMBS; VGBS: volume-gradient-based band selection; ACMBS: autocorrelation matrix-based band selection; AP: affinity propagation; FMAP: feature-metric-based AP; DFM: discriminative feature metric; MDPP: multigraph determinantal point process; SNMF: sparse nonnegative matrix factorization; EMD: earth's mover distance; SpaBS: sparse representation-based band selection; ISSC: improved sparse subspace clustering; SWLRSC: squared-weighted low-rank subspace clustering; FRSR: fast and robust self-representation; DWSSR: dissimilarity-weighted sparse representation; SSR: symmetric recursive SVM; KFS: kernel-based feature selection; MLR: multinomial linear regression; SMLR: sparse MLR; SCBS: sparse-constrained band selection; CNNs: convolutional neural networks; CM: contribution map; SSI: intraclass separability index; IFBP: impurity-function band prioritization.

NOTE: SNR: signal-to-noise ratio; PCA: principal component analysis; MVPCA: minimum-variance PCA; MSNRPCA: maximum SNR PCA; CBS: constrained band selection; CMBS: constrained multiple-band selection; FDPC: fast density peak-based clustering; ECA: exemplar component analysis; E-FDPC: enhanced fast density peak-based clustering; MMCA: minimum misclassification canonical analysis; NHMC: nonhomogeneous hidden Markov chain; VNVBS: variable-number variable-band selection; OSP: orthogonal subspace projection; ICA: independent component analysis; MI: mutual information; TMI: trivariate MI; STMI: semisupervised TMI; NTMI: normalized TMI; LP: linear prediction; BAO: band add-on; ADM: average distance method; MDM: minimum distance method; MEAC: minimum estimation abundance covariance; NPFS: nonlinear parsimonious feature selection; MMAIQ: maximal minimal associated index; VPNSR: variable-precision neighborhood rough set; PSO: particle swarm optimization; GA: genetic algorithm; CSFS: clonal selection feature selection; BA: Bat algorithm; FODPSO: fractional-order Darwinian PSO; SBS: sequential backward selection; CEM: constrained energy minimization; BSC: band selection for CMBS; VGBS: volume-gradient-based band selection; ACMBS: autocorrelation matrix-based band selection; AP: affinity propagation; FMAP: feature-metric-based AP; DFM: discriminative feature metric; MDPP: multigraph determinantal point process; SNMF: sparse nonnegative matrix factorization; EMD: earth's mover distance; SpaBS: sparse representation-based band selection; ISSC: improved sparse subspace clustering; SWLRSC: squared-weighted low-rank subspace clustering; FRSR: fast and robust self-representation; DWSSR: dissimilarity-weighted sparse representation; SSR: symmetric recursive SVM; KFS: kernel-based feature selection; MLR: multinomial linear regression; SMLR: sparse MLR; SCBS: sparse-constrained band selection; CNNs: convolutional neural networks; CM: contribution map; SSI: intraclass separability index; IFBP: impurity-function band prioritization.

and maximum SNR PCA (MSNRPCA). The MVPCA sorts all of the bands by defining a variance-based band-power ratio, denoted by $R_{\text{var}}(m)$, as

$$R_{\text{var}}(m) = \frac{\sum_{i=1}^m \sigma_i}{E}, \quad (1)$$

where $E = \sum_{i=1}^N \sigma_i$ is the sum of variances of all the band images, σ_i is the variance of the i th band, and m is the number of selected bands. The advantage of MSNRPCA over MVP-CA is that noise is whitened before the PCA to eliminate noise impact. Neither MVPCA nor MSNRPCA took spectral correlation into account, and, accordingly, the divergence-based band-decorrelation scheme [22] based on the Kullback–Leibler distance was implemented to remove either redundant or insignificant bands. If the divergence between two bands was below a predefined threshold, the band with low priority was removed. Later, mutual information (MI) was used to gauge band dissimilarity [23]. In [24], the covariance-based method prioritized all of the spectral bands to minimize their impact on target detection using a matched filter and adaptive coherence estimator.

LOW-CORRELATION CRITERIA

The selected bands are required to have low mutual correlations. In [25], band selection was implemented into a framework of target detection, where a band image was treated as a desired target signature and other bands were unknown signature vectors; the objective was to minimize the correlations of selected bands. The constrained band selection (CBS) method adopted constrained energy minimization (CEM) to linearly constrain a band image while minimizing the band correlation (BC) or dependence provided by other band images. The resulting CEM-CBS was formulated as

$$\min_{\mathbf{w}_i} \{\mathbf{w}_i^T \mathbf{Q} \mathbf{w}_i\} \text{ subject to } \mathbf{b}_i^T \mathbf{w}_i = 1, \quad (2)$$

where \mathbf{b}_i is a column vector of the i th band image, \mathbf{w}_i is the column vector specifying a finite-impulse response filter designed for \mathbf{b}_i , and \mathbf{Q} is the BC matrix. The solutions to the CEM-CBS problem were extended using four different criteria, i.e., band-correlation minimization (BCM), band-correlation constraint, band-dependence constraint, and band-dependence minimization. Unfortunately, CEM-CBS required tremendous computational time because of the enormous dimension of vectors converted from band images. Therefore, [25] developed the linearly constrained MV (LCMV)-CBS to constrain a band image into an image matrix without vector conversion, thereby dramatically reducing computational complexity. Later, the constrained multiple-band selection (CMBS) expanded the CEM-CBS and implemented the LCMV to constrain and select multiple bands simultaneously [26]. In [27], the optimal band index criterion based on fractal dimension was put forward to quantify the correlation of spectral and spatial information among all bands for band

selection. Based on spectral redundancy and band decorrelation, the local curve-fitting technique [28] was employed to select optimal bands by ranking the absolute difference between the average reflectance and smoothed average reflectance generated from curve fitting.

LARGE-DISSIMILARITY CRITERIA

The selected bands are expected to be dissimilar to each other. Inspired by the fast density peak-based clustering (FDPC) algorithm in [29], a ranking-based exemplar component analysis algorithm was proposed to automatically find cluster centers from all the bands. The algorithm required only the measuring of the distance between all pairwise bands and did not need to parametrize a probability density function. A simple criterion was adopted to find the independent density peaks. Each band i was ranked by ES_i , which is the product of its local density ρ_i and the distance δ_i to other bands with higher densities,

$$ES_i = \rho_i \delta_i. \quad (3)$$

A band with an anomalously large ranking score, i.e., ES , was recognized as an exemplar or representative band. Later, enhanced FDPC (E-FDPC) [30] was presented to solve two issues related to the scale problem of ρ_i and δ_i as well as the cutoff threshold. The E-FDPC algorithm computed the ranking score of each band by weighting the local density and the intracluster distance, and an exponential-based learning rule was introduced to properly determine the cutoff threshold. In [31], to benefit the red-green-blue-based hyperspectral visualization, the CBS for visualization (CBSV_z) adopted the second and third orders of normalized information measured to select the triplet of bands containing minimal redundancy and maximal informative content.

SUPERVISED RANKING-BASED METHODS

Different from unsupervised ranking-based methods, supervised methods involve the prior knowledge of HSI data when constructing a band-prioritization criterion. The designed criteria may correlate closely with certain applications, i.e., classification, spectral unmixing, and so on.

1) *Classification-aimed criteria*: The criteria were designed to ensure excellent classification performance using selected bands. In [23], minimum misclassification canonical analysis (MMCA) was proposed to rank bands according to their classification abilities. Derived from Fisher's discriminant function, the MMCA aimed to minimize the misclassification error of selected bands by solving the generalized eigenvalue problem,

$$\mathbf{S}_b \mathbf{v} = \lambda \mathbf{S}_w \mathbf{v}, \quad (4)$$

where \mathbf{S}_b and \mathbf{S}_w are the between-class and within-class scatter matrices, respectively, and λ and \mathbf{v} are the corresponding eigenvalue and eigenvectors, respectively. Similarly, a

band-power ratio can be defined to measure the misclassification rate of bands to be used for classification.

MI provides a framework to measure the similarity between two random variables,

$$I(\mathbf{b}_i, \mathbf{b}_j) = H(\mathbf{b}_i) + H(\mathbf{b}_j) - H(\mathbf{b}_i, \mathbf{b}_j), \quad (5)$$

where $H(\mathbf{b}_i)$ and $H(\mathbf{b}_j)$ are the entropies of variables \mathbf{b}_i and \mathbf{b}_j , respectively, and $H(\mathbf{b}_i, \mathbf{b}_j)$ is their joint entropy. Treating a band image and the corresponding reference map as random vectors, the MI criterion [32] was used to evaluate the relative utility of each band for classification. The reference map was designed using prior knowledge about the spectral signatures of frequently encountered materials. In [33], the feature-weighting band selection algorithm was proposed for HSI classification based on the pairwise separability criterion and matrix coefficients analysis. The algorithm conducted decorrelation through PCA and ranked bands for each class by evaluating the corresponding PCA coefficients and criterion values. The voting techniques were used to determine the weights of all the bands for ranking, and the redundant bands were removed by a correlation threshold. Later, phase correlation [34] was adopted as a correlation measurement to remove redundant bands. In [35], a nonhomogeneous hidden Markov chain (NHMC) model was trained and applied on the wavelet-transformed training and testing samples. The NHMC quantified the importance of each band with the average of all the pairwise correlation coefficients (CCs) for each band.

2) *Spectral-unmixing-aimed criteria*: This criterion was designed for spectral unmixing. Based on the linear mixture model, the orthogonal subspace projection (OSP) [23] made use of subspace projection to eliminate undesired signatures and suppressed noise to maximize the SNR, which is defined as

$$\text{SNR}_{\max} = \frac{\mathbf{A}\mathbf{A}^T}{\sigma^2} [\mathbf{D}^T \mathbf{P}_{\mathbf{U}}^\perp \mathbf{D}], \quad (6)$$

where \mathbf{D} is the signature matrix of endmembers in the image scene, \mathbf{A} is the associated abundance matrix, and $\mathbf{P}_{\mathbf{U}}^\perp = \mathbf{I} - \mathbf{U}\mathbf{U}^\#$ is the orthogonal subspace projector, with \mathbf{U} being the undesired spectral-signature matrix and $\mathbf{U}^\# = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T$ being the pseudoinverse matrix of \mathbf{U} . Later, the algorithm of variable-number, variable-band selection (VNVBS) was presented [36] to select a different number of bands in accordance with the hyperspectral signatures to be processed. The VNVBS used the OSP to form a band-prioritization criterion and assigned a different priority score to each spectral band to capture various features.

Independent component analysis (ICA) is a technique that extracts independent source signals by searching for a linear transform that minimizes statistical dependence among components. In [37], ICA was used to evaluate the weight matrix to assess the contribution from each band to the unmixing procedure. The proposed

ICA-based algorithm sorted the average absolute weight coefficients of individual bands and selected bands, which contained more information. In [38], an alternative separability index was presented to maximize the spectral separability among endmembers by focusing on the unmixing analysis of a band subset, which reduced the plant-detection accuracy of mixture analysis with the selected bands.

SEARCHING-BASED METHODS

The searching-based methods convert band selection into an optimization problem of a given criterion function and search for the best bands to form an optimal solution. Two key issues exist in searching-based methods: 1) the criterion function and 2) the searching strategy. The criterion function can be similarity-based measurements such as Euclidean distance (ED), Bhattacharyya distance, Jeffries–Matusita (JM) distance [39], spectral angle mapping (SAM) [40], structural similarity index measurement [41], or information-based measurements such as spectral information divergence (SID), transformed divergence, MI [42], and spatial entropy-based MI [43]. The searching strategy determines the best way to find an optimal or suboptimal solution. According to the adopted searching strategy, searching-based methods can be grouped into three categories: incremental searching, updated searching, and eliminating searching.

INCREMENTAL SEARCHING-BASED METHODS

Rather than exhaustively testing all band combinations, which is computationally prohibitive, the incremental searching-based methods sequentially add new bands that optimize the criterion into current band subsets until a desired number of bands is selected. The sequential forward selection (SFS) strategy is often implemented. These methods can be unsupervised or supervised, depending on whether labeled training samples are needed during the searching process.

UNSUPERVISED SEARCHING

Unsupervised incremental searching-based methods do not consider prior knowledge of HSI data; rather, they iteratively add dissimilar and informative bands. In [44], a band selection algorithm was proposed to search bands with large skewness or kurtosis values. In [45], linear prediction (LP) and OSP were adopted to jointly evaluate the similarity between a single band and multiple bands, and the selected bands were expected to achieve satisfactory performance in target detection and classification. In [46], the spectral rhythm was adopted to exploit the intermediary representation of HSI data, and the most dissimilar and informative bands were iteratively selected, with the dissimilarity measurement based on bipartite graph matching. Using convex set geometry, the geometry-based algorithm [47] iteratively searched new vertices that maximize the largest simplex in the pixel space, and the selected bands corresponding to these vertices had low correlations with each other.

SUPERVISED SEARCHING

Supervised incremental searching-based methods use prior knowledge of HSI data to maximize class separability. In [48], employing the exact decomposition of SAM, the band add-on (BAO) algorithm was presented to iteratively select bands that increase the angular separation between two spectra in a spectral library. To address more the practical problem of discriminating two classes of spectra, the BAO was extended to two band selection approaches, i.e., the average distance and minimum distance methods (MDMs), and the selected bands increased the angular separability of the two classes. Similarly, in [49], using class spectral signatures, the minimum estimation abundance covariance (MEAC) algorithm incrementally selected dissimilar bands to preserve the desired information for classification through minimization of the trace of abundance covariance matrix as

$$\operatorname{argmin}_{M^S} \left\{ \operatorname{trace}[(\hat{S}^T \hat{S})^{-1}] \right\}, \quad (7)$$

where M^S is the selected band subset and \hat{S} is the matrix containing class spectral signatures in the selected subset. Using fast SFS searching, the nonlinear parsimonious feature selection algorithm in [50] iteratively selected spectral bands that maximized an estimate of the classification rate from the Gaussian mixture model classifier.

Cramer's V (CV) is the most popular nominal association used to measure the strength of the relationship between variables regardless of table size. Based on CV-test theory, the maximal minimal associated index (MMAIQ) [51] was proposed to incrementally select bands that yield statistically maximal associations between target labels/features and minimal associations among selected features. However, the MMAIQ suffers from the discretization problem and does not consider class labels in calculating the CV-test value. Therefore, in [52], the CV-test discretization-based association index and the class-attribution interdependence maximization-based association index were proposed to refine the band selection result of the MMAIQ. The two derived algorithms implemented the CV-test discretization scheme and incrementally selected the optimal spectral bands using the greedy algorithm. Later, a simple equal-width interval-binning approach [53] was implemented in a rough set-based band selection algorithm, which quantified the relevance and significance of each band based on the rough set theory and iteratively selected informative bands with higher relevance and significance values.

Also inspired by the rough set theory, the variable-precision neighborhood rough set (VPNRS)-based algorithm [54] measured the significance of each band using dependency and implemented the forward greedy algorithm to select informative bands. A sequential band subset selection algorithm [55] was put forward to select proper bands for anomaly detection, inspired by the N-FINDER and iteratively added new bands by comparing the performance of anomaly detection with that of all bands using

the averaged least-squares error and 3D receiver operating characteristic curves. A new index was proposed for target detection using selected bands in [56] by comparing the largest outputs in the detection map using all of the original bands.

UPDATED SEARCHING-BASED METHODS

The updated searching-based methods iteratively replace elements of the current band subset with new ones as needed during the searching procedure to ensure that the predefined evaluation criterion is optimized. Aside from simple searching strategies such as the sequential forward-floating search [21], the branch-and-bound search [57], the steep ascent search, and the constrained search [58], [59], evolutionary algorithms have been adopted for band searching, such as particle swarm optimization (PSO) [21], [56], adaptive simulated annealing [60], [61], genetic algorithms (GAs) [62], firefly algorithms (FAs) [63], differential evolutionary algorithms [64], and ant colony optimization [65], [66]. Similar to incremental searching-based methods, similarity or information measurements can be used as an objective function. In [67], a maximum-information and minimum-redundancy (MIMR) criterion was derived as the low-order approximation of maximum joint-MI criterion. The clonal-selection algorithm in an artificial immune system was devised to optimize the objective function of MIMR. In [68], the conflicting relationships between information entropy and the number of selected bands was examined, and a multiobjective optimization band selection algorithm was designed to find the best tradeoff solution. Some objective functions utilize the prior knowledge of HSI data (e.g., labeled training samples and class signatures) to formulate supervised or semisupervised band selection methods for HSI classification. According to whether a classifier is involved during band searching, these approaches can be grouped into classifier-independent or classifier-dependent methods.

CLASSIFIER-INDEPENDENT METHODS

The objective function is independent of classification accuracy from a true-classifier and estimates-only class separability perspective, based on specific metrics. In [69], the clonal-selection, feature-selection algorithm adopted the averaged JM distance among different classes as the criterion function, and the clonal-selection strategy was implemented to search for a band subset that optimizes the criterion function. In [70], band selection was converted into an optimization problem for maximizing class separability using certain criteria, i.e., ED, SAM, spectral correlation mapper, and BC, and the immune clonal-strategy algorithm was adopted for band searching. In [71], the PSO algorithm was used to search for a band subset that optimizes the criterion function of MEAC or JM distance among different classes. In [72], the GA was introduced to search the optimal band combination and coordinate the potential conflicts in the objective function constituted of

information entropy and Bhattacharyya distance among classes. In [73], a greedy coherence-based band selection (GCBS) algorithm was proposed to train supervised unmixing models in the reproducing kernel Hilbert space. Based on the coherence criterion, the GCBS set the largest value allowed for correlations between the basis kernel functions characterizing the selected bands in the unmixing model. In [74], the trivariate MI (TMI), defined as

$$\max \left(\sum_{i=1}^m I(\mathbf{b}_i; \mathbf{Y}) - \frac{2}{m-1} \sum_{1 \leq i < j \leq m} I(\mathbf{b}_i; \mathbf{b}_j; \mathbf{Y}) \right), \quad (8)$$

improved the traditional MI to better measure band redundancy, which was optimized by the clonal-selection algorithm. Here, \mathbf{Y} is the class labels of training samples, and $I(\cdot)$ is the operator of MI. The TMI was also extended into a semisupervised version, STMI, to use unlabeled samples by adding a graph-regularization term. Later, to simultaneously measure the redundancy of labeled and unlabeled training samples, the STMI was incorporated with a semisupervised criterion called *high discrimination, high information, and low redundancy* [75].

CLASSIFICATION-DEPENDENT METHODS

Here, the accuracy produced by an actual classifier is directly adopted to formulate an objective function. In [76], five nature-inspired algorithms, including a gravitational search algorithm, harmony search, PSO, FA, and Bat algorithm, were evaluated to select the best bands that maximize the objective function of the optimum path forest classifier accuracy. In [77], the binary optimization method inspired by fractional-order Darwinian PSO was proposed to select informative bands, and the overall accuracy of a support vector machine (SVM) classifier was utilized as the search criterion. A possible feature selection score based on random forests was used to construct the criterion function, and band selection was performed using the GA. With the linear combination of classification error and the number of selected bands as the search criterion, the probabilistic memetic algorithm was employed to optimize the objective function and identify the most informative bands [79].

In [80], the band selection problem was formulated into a combinatorial problem by modeling a class separability measurement of the Hausdorff distance and the accuracy from SVMs, and the metaheuristic algorithm Gray Wolf optimizer algorithm was adopted to optimize the resulting objective function. Similarly, the band selection problem in [80] was transformed into a combined optimization problem of the number of selected bands and classification accuracy, and the chaotic binary-coded gravitational search algorithm was employed to find the best solution.

ELIMINATING SEARCHING-BASED METHODS

The eliminating searching-based methods are initialized with the full bands and eliminate unnecessary bands at

each iteration until reaching the desired number of selected bands. Sequential backward selection (SBS) was typically adopted. In [81], BS for the CEM algorithm implemented the product of weights and the norms of bands to construct an evaluation criterion and used the SBS to iteratively remove unimportant bands. In [82], the volume gradient-based band selection algorithm iteratively removed the most redundant bands based on the volume gradient of a simplex with respect to HSI data. Here, it was assumed that the redundant bands had the shortest distance to the hyperplane spanned by the other bands, and a criterion function was defined to maximize the volume of subsimplex formed by the selected bands. The autocorrelation matrix-based band selection algorithm [83] employed the minimum relative LP error as the criterion and searched the suboptimal subset of bands by the SBS. In [84], the quality of bands was considered in the minimum-noise band selection algorithm, which designed an index combining the noise-adjusted principal components and the maximum determinant of covariance matrix and adopted the SBS to find the best bands with both high SNRs and low correlation. To better describe the non-Gaussianity of HSI data, the joint-skewness band selection algorithm in [85] defined a criterion function combining the third-order (second-order tensor) and second-order (correlation) statistics of bands and adopted the SBS to remove redundant bands for small target detection. In [86], the proposed band selection algorithm iteratively eliminated one band from the pair of the most correlated neighboring bands depending on their discriminant power.

CLUSTERING-BASED METHODS

The clustering-based methods group the original bands into clusters and select the representative bands from each cluster to constitute the final band subset. The first paper on HSI band clustering describes the hierarchical clustering algorithm [20], based on Ward's linkage, to minimize the intracluster variance and maximize the intercluster variance simultaneously. The representative bands can be determined using information measurements, e.g., MI or Kullback–Leibler divergence, to remove redundant bands. Since then, other clustering-based band selection algorithms were proposed in the literature. With the exception of BandClust [87], fuzzy clustering [88], and automatic band selection algorithms [89], most clustering-based methods are derived from k -means, affinity propagation (AP), and graph clustering, which are discussed in the following sections.

K-MEANS-BASED CLUSTERING METHODS

The k -means algorithm is a widely used clustering technique. It is initialized with a set of randomly selected bands, and the objective function quantifying the sum of distances to a set of putative center candidates is iteratively optimized until the best cluster centers are found. Specifically, it aims to partition

N bands into m clusters $C = \{c_1, \dots, c_j, \dots, c_m\} (1 \leq j \leq m)$ to minimize the objective function

$$\arg\min_C \sum_{j=1}^m \sum_{b_i \in c_j} D(b_i, \mu_j), \quad (9)$$

where μ_j is the cluster center of c_j and $D(\cdot, \cdot)$ is a distance metric gauging the similarity between a band and the center of the cluster it is assigned to. In [90], the k -means algorithm used different characteristics of HSI data for clustering, i.e., interquartile range, geometric mean, median absolute deviation, median, CC, covariance, and mode. In [91], a representative band mining algorithm was presented for band selection in the k -means clustering scheme, where disjointed information was adopted to measure pairwise band difference. As opposed to traditional clustering-based methods that select bands individually from each band cluster, representative band mining selects representative bands simultaneously by exploring all spectral band clusters. Using a nonlinear kernel function, the kernel k -means algorithm [92] was devised to identify nonlinearly separable clusters of spectral bands in the reproducing kernel Hilbert spaces for the nonlinear unmixing problem. The k -means algorithm was also extended into the dual clustering algorithm [93], where the proposed dual-clustering-based band selection by context analysis presented a pairwise hyperspectral angle descriptor to exploit the context information of a specific pixel and implemented the groupwise strategy of representation to select representative bands.

Specific supervised or semisupervised band selection methods were also proposed based on the k -means algorithm. Using only the class spectra, the prototype space band selection algorithm [94] clustered the bands in the prototype space, where all bands were characterized in terms of class signatures to pose spectral properties of classes to bands; the combination of representative bands from each cluster that brought about the best classification accuracy was selected as the final band subset. The class signatures were also input into the k -means algorithm as a semisupervised version in [95], where the clustering result from the semisupervised k -means was purified by removing outlier bands in each cluster before finalizing cluster centers as representative bands. The formulated semisupervised k -means algorithm improved band selection performance and reduced computational cost.

AFFINITY PROPAGATION-BASED CLUSTERING METHODS

The k -means algorithm clustering is sensitive to initial conditions. Therefore, the exemplar-based AP clustering algorithm [96] was proposed to search for an appropriate set of exemplars as representative bands. The AP algorithm considered the correlation or similarity among bands as well as the discriminative capability of each individual band, and the exemplar e was obtained by maximizing the function

$$\mathcal{H}(e; \Theta) = \exp \left(\sum_{i=1}^N s(i, e_i) + \sum_{i=1}^N \log f_i(e) \right), \quad (10)$$

where Θ is the similarity matrix between bands, with each similarity $s(i, e_i)$ indicating the suitability of a band with index e_i to be the exemplar for the i th band; and the second term, $\sum_{i=1}^N \log f_i(e)$, is a coherence constraint of the exemplar, explaining that, if a band is chosen as an exemplar by other bands, it is also forced to be its own self-exemplar. In [97], two feature measurements, i.e., the band-correlation metric (BCM) and band separability metric, were proposed to evaluate the spectral correlation of all bands and the class discrimination capability of single bands, respectively, and the unsupervised version of AP was formulated. Later, the discriminative feature metric (DFM) [98] was utilized with a discriminative constraint that involved chunklet and discriminative information of bands. The proposed DFM-AP can select highly discriminative bands with low redundancy.

The AP clustering algorithm was also formulated into supervised or semisupervised versions. In [99], the AP was combined with discriminative kernel alignment (DKA) to form a discriminative kernel-clustering algorithm for band selection. The DKA measurement compared intraclass and interclass similarities between kernels and provided accurate kernel similarity for AP clustering. In [100], a semisupervised adaptive AP algorithm was proposed in which class signatures were used in the distance measurement to improve band selection performance. In addition, a measurement called *normalized TMI* [101] was designed for the AP clustering; it not only considered band redundancy and band synergy but also quantified the discriminative abilities of bands using both labeled and unlabeled samples.

GRAPH-BASED CLUSTERING METHODS

Graph theory formulates band selection into a graph problem, which states that nodes in a graph represent the HSI bands and edges connecting two nodes correspond to the similarity between the two bands. A graph-based clustering method constructs an affinity matrix \mathbf{A} that correlates with band similarity and clusters the graph into subgraphs to find representative bands. Specifically, \mathbf{A} is a matrix that describes band affinities, and each affinity between pairwise bands is computed as $a_{ij} = \exp(-\|f_i - f_j\|^2 / 2\sigma^2)$, with σ denoting a scaling factor. A typical example of graph-based clustering algorithms is spectral clustering [102], [103], which makes clustering on the stacked eigenvectors of the affinity matrix $\mathbf{\Gamma}$, defined as

$$\mathbf{L} = \mathbf{\Lambda}^{-1/2} \mathbf{\Gamma} \mathbf{\Lambda}^{-1/2}, \quad (11)$$

where \mathbf{L} is the normalized graph Laplacian and $\mathbf{\Lambda}$ is the diagonal matrix computed by the sums over the rows of \mathbf{A} . In [104], the multigraph determinantal point process (DPP) implemented the spectral clustering algorithm to discover the subgraph structures in the high-dimensional band space, and the searching strategy of DPP was then developed to select the most significant subset of bands using the affinity matrix of each subspace. In [105], the

graph clustering utilized Bhattacharyya distance to construct a band adjacency graph, and the Markov clustering algorithm was adopted to cluster the band-adjacent graph into different clusters. The band that has minimal total distances with other bands in each cluster was considered as the representative band.

In [106], the structure-aware measurement exploited the structural information of HSI data by means of spatial-spectral consistency analysis to construct an undirected weighted graph, and the dominant-set extraction was adopted for clustering on graph representation and select dominant vertices as representative bands. In [107], HSI data were transformed into a complex network graph where band selection was carried on by analyzing the network's topological feature corresponding to each band, and the bands that are the most qualified for discriminating the network's demarcation units were selected.

SPARSITY-BASED METHODS

According to the sparsity theory, each band can be sparsely represented using only a few nonzero coefficients associated with atoms in a suitable basis or dictionary. The sparsity-based band selection methods make use of sparse representation or regression to reveal certain underlying structures within HSI data. The representative bands can be found by solving an optimization problem with sparsity constraints. Current sparsity-constrained methods can be categorized into sparse nonnegative matrix factorization (SNMF)-based, sparse representation-based, and sparse regression-based approaches.

SPARSE NONNEGATIVE MATRIX FACTORIZATION-BASED METHODS

The SNMF decomposes an original HSI data matrix into the product of a set of bases (i.e., the basis matrix) and encodings (i.e., the coefficient matrix), where the basis is nonnegative and the encodings are both negative and sparse. The nonnegative constraint in bases and encodings brings about the parts-based feature of SNMF because only additive combinations are allowed. In [108], the SNMF was introduced to solve the HSI band selection problem, where the representative bands were chosen by clustering the sparse coefficients. Specifically, it aims to factorize the HSI band matrix \mathbf{B} into the product of an unknown dictionary matrix $\mathbf{W} \in \mathbb{R}^{D \times r}$ and an unknown sparse coefficient matrix $\mathbf{H} \in \mathbb{R}^{r \times N}$ by optimizing the objective function

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} f_r(\mathbf{W}, \mathbf{H}) &= \frac{1}{2} \|\mathbf{B} - \mathbf{WH}\|_F^2, \\ \text{s.t., } \mathbf{W}, \mathbf{H} &\geq 0, \text{ and } \|\mathbf{h}_i\|_0 \ll r, \quad 1 \leq i \leq N, \end{aligned} \quad (12)$$

where the subscript r in f_r denotes the desired low rank r . The largest entry in each column of \mathbf{H} indicates the cluster or subspace of each band it belongs to. The constraint $\|\mathbf{h}_i\|_0 \ll r$ explains that each column vector \mathbf{h}_i is sparse and the number of nonzero entries is far smaller than

the dimensionality r . Unfortunately, the adopted ED distance measurement in (12) does not perfectly represent the approximate error between \mathbf{X} and its approximation \mathbf{WH} . This is because the Gaussian distribution assumption behind the ED measurement contradicts the nature of HSI data. Therefore, the Earth's mover distance (EMD) was adopted to measure the approximation error and improve the band selection from SNMF [109]. The EMD is a cross-bin distance that mainly resolves the histogram- or image-matching problems and aims to minimize the cost that must be incurred when transforming one feature distribution (i.e., the histogram) into the other. Let X_l represent the source histogram, which corresponds to the spectral responses of pixels in the l th band, and let the column vector \tilde{X}_l be the target normalized histogram, which corresponds to another band. The EMD distance between two histograms b_l and \tilde{b}_l is formulated as a linear programming problem with the objective of minimizing the total cost for transforming the source b_l to the target \tilde{b}_l as

$$\begin{aligned} \text{EMD}(\mathbf{b}_l, \tilde{\mathbf{b}}_l) &= \sum_{i,j} f_i(i,j) D(i,j), \\ \text{s.t., } \begin{cases} f_i(i,j) \geq 0, \sum_{i,j} f_i(i,j) = 1 \\ \sum_i f_i(i,j) = X(i,l), \sum_j f_i(i,j) = \tilde{X}(j,l) \end{cases} \end{aligned} \quad (13)$$

where $f_i(i,j)$ and $d(i,j)$ are the flow amount and flow cost between the i th pixel of the source histogram \mathbf{b}_l and the j th pixel of the target histogram $\tilde{\mathbf{b}}_l$, respectively, and the flow cost measures the L_1 ground distance.

SPARSE REPRESENTATION-BASED METHODS

Unlike SNMF-based methods, sparse representation-based methods learn or manually define a dictionary in advance. They often select the informative bands according to sparse coefficients. In [110], using the sparse representation of HSI bands with a K-SVD-learned dictionary, the sparse representation-based band selection algorithm ranked the bands according to sparse coefficients using majority voting, and the bands with high occurrence in the histograms of sparse coefficients were selected. Band selection [111] was formulated into a multitask sparse learning problem, and an immune clonal strategy was adopted to search for the best band combinations based on sparse coefficients. The collaborative sparse representation model can force all bands to share the same active atoms if possible. In [112], the collaborative sparse model was adopted to refine the preselected bands and minimize the number of selected bands. When the dictionary in a sparse representation model is set to be equal to the HSI data themselves, all of the band vectors are assumed to be sampled from several independent subspaces. The improved sparse subspace clustering (ISSC) model [113] was formulated into the optimization problem

$$\underset{\mathbf{Z}}{\text{argmin}} \|\mathbf{Z}\|_{q^*} \quad \text{s.t., } \mathbf{B} = \mathbf{BZ} \text{ and } \text{diag}(\mathbf{Z}) = \mathbf{0}, \quad (14)$$

where \mathbf{Z} is the coefficient matrix of all band points, and $\|\mathbf{Z}\|_q$ represents the L_q -norm of \mathbf{Z} defined as $\|\mathbf{Z}\|_q = \left(\sum_{i=1}^N \sum_{j=1}^N |Z_{ij}|^q \right)^{1/q}$. Here, the constraint $\text{diag}(\mathbf{Z}) = 0$ eliminates the trivial solution, which states that each band is simply represented by itself. The ISSC algorithm estimates the representative bands using spectral clustering on the similarity matrix constituted using the sparse matrix \mathbf{Z} . In [114], the spatial-spectral information between pairwise pixels was added to weigh the entries in the coefficient matrix \mathbf{Z} and improve the band selection performance. The squared-weighted strategy was adopted to construct a strongly connected adjacent matrix with sparse coefficients, and the ISSC algorithm became the squared-weighted low-rank subspace clustering band selection algorithm [114]. Considering the separability of all the bands, the band selection of separable nonnegative matrix factorization was transformed into the fast and robust self-representation (FRSR) model [115] as

$$\mathbf{B} = \mathbf{B}\mathbf{Z} + \mathbf{E} \text{ s.t., } \begin{cases} \mathbf{Z} \geq 0, \text{Tr}(\mathbf{Z}) = r, 1 \leq i \leq N \\ \mathbf{Z}(i,j) \leq \mathbf{Z}(i,i) \leq 1, 1 \leq j \leq r, \end{cases} \quad (15)$$

where $\text{Tr}(\cdot)$ is the trace operation. The term $\text{Tr}(\mathbf{Z}) = r$ constrains the number of bands to be selected. The constraint $\mathbf{Z}(i,i) \leq 1$ results from the unit L_1 norm of each column in \mathbf{B} . The constraint $\mathbf{Z} \geq 0$ will guarantee the positive weight for constructing each band from the desired band subset. The constraint $\mathbf{Z}(i,j) \leq \mathbf{Z}(i,i)$ comes from the fact that the reconstruction weight for any band point (i.e., a band vector) inside the convex hull cannot exceed the ratio of the L_1 norm of that band to each selected band. The structured random projection was implemented to speed up (15), and the representative bands corresponded to the largest diagonal entries in the estimated factorization localizing matrix \mathbf{C} . The sparse self-representation model in the ISSC algorithm can select informative bands, but it did not consider contributions of individual bands in encoding other bands with sparse coefficients, which may result in similar bands being selected. Therefore, a dissimilarity regularization term was added into the sparse self-representation model, where the dissimilarity information between two bands was quantified by a composite distance metric that integrated the SID and CC. The objective function of dissimilarity-weighted sparse representation model [116] was

$$\begin{aligned} & \underset{\mathbf{Z}}{\text{argmin}} \lambda \|\mathbf{Z}\|_{1,2} + \mu \text{tr}(\mathbf{D}^T \mathbf{Z}) + \frac{1}{2} \|\mathbf{B} - \mathbf{B}\mathbf{Z}\|_F^2 \\ & \text{s.t., } \mathbf{Z} \geq 0, \text{diag}(\mathbf{Z}) = 0, \mathbf{1}^T \mathbf{Z} = \mathbf{1}^T, \end{aligned} \quad (16)$$

where \mathbf{D} is the dissimilarity weighted matrix and $\|\mathbf{Z}\|_{1,2}$ is the sum of the L_2 -norm for the coefficient vector in all rows. The constraint $\mathbf{1}^T \mathbf{Z} = \mathbf{1}^T$ is equivalent to $\sum_{i=1}^N \mathbf{Z}_{ij} = 1$, where $\mathbf{1} \in \mathbf{R}^{N \times 1}$ is a column vector with all entries equal to one.

The sparse self-representation of bands was also extended into a symmetric sparse representation (SSR) model [117],

where the selected and original bands can be symmetrically sparsely represented. The SSR algorithm formulates band selection into the well-known problem of archetypal analysis:

$$\begin{aligned} & \underset{\mathbf{H}, \mathbf{A}}{\text{argmin}} \|\mathbf{B} - \mathbf{B}\mathbf{H}\mathbf{A}\|_F^2 \\ & \text{s.t., } \begin{cases} h_j \geq 0, h_{jj} = 0 \text{ and } |h_j| = 1, \forall j \in \{1, \dots, r\} \\ a_i \geq 0 \text{ and } |a_i| = 1, \forall i \in \{1, \dots, N\} \end{cases} \quad (17) \\ & 1 < r < N \end{aligned}$$

Archetypal analysis assumes that archetypes are convex combinations of all the band points, while all of the band points are approximated in terms of convex combinations of archetypes. Thus, selecting the representative bands can be regarded as finding the archetypes in the minimal convex hull containing all of the high-dimensional band points.

SPARSE REGRESSION-BASED METHOD

Sparse regression-based methods formulate the band selection problem into a sparse linear regression problem using training samples and their class labels. Sparse coefficients from the best solution assist in selecting the bands for better class separability. In [118], the sparsity constraint was imposed on the linear regression between training samples and their class labels, and the least shrinkage and selection operator (LASSO) was formulated as

$$\underset{\mathbf{w}_c}{\text{argmin}} \|\mathbf{y}_c - \mathbf{B}^T \mathbf{w}_c\|_F^2 + \lambda \|\mathbf{w}_c\|_1, \quad (18)$$

where \mathbf{y}_c are class labels according to the c th class (where 1 means belonging to the c th class, and 0 means not belonging), respectively, and λ is a parameter that controls the contribution of the L_1 penalty term. The bands were selected by ranking the coefficient of each band for all classes. The LASSO model and a new separability measurement based on the Hilbert–Schmidt independence criterion [119] were coupled into a unified framework for band selection. The proposed combination selected proper bands, which increased class separability and avoided a computationally intensive searching strategy. In [120], the band selection for target detection using CEM was transformed into the continuous optimization problem of LASSO, and the proposed LASSO band selection algorithm evaluated the importance of all the bands simultaneously and avoided complex subset selection.

Moreover, in [121], the group-sparsity-based algorithm fulfilled the band selection task by imposing the group-sparsity constraint on the regression coefficients, and the bands that held the lowest classification errors were selected. In [122], group sparsity was incorporated into the linear regression classifier to formulate the semisupervised band selection method, where the label propagation was employed to propagate class labels to unlabeled samples using a hypergraph model and the proper bands were selected by ranking the scoring coefficients of the selection matrix.

EMBEDDING LEARNING-BASED METHODS

Embedding learning-based methods incorporate band selection into the optimization of specific application models (e.g., classification, target detection, and spectral unmixing).

CLASSIFIER LEARNING-BASED METHODS

Because of low sensitivity to imbalanced training samples, the SVM classifier is a very popular method used in hyperspectral image analysis. The recursive-feature elimination (RFE)-SVM can provide superior performance in hyperspectral band selection; here, the weight value calculated in the SVM training stage was used as the ranking criterion to remove redundant bands and optimize the SVM classifier. RFE focuses on minimizing the generalization error by eliminating features that maximize the margin. The predictive ability measurement s^{REF} is inversely proportional to the margin and is given by

$$s^{\text{REF}} = \sum_{i=1}^D \sum_{j=1}^D \alpha_i \alpha_j \gamma_i \gamma_j \Phi(\mathbf{x}_i, \mathbf{x}_j), \quad (19)$$

where \mathbf{x}_i and $\gamma_i \in \{-1, 1\}$ are the i th training sample and its class label, respectively, and $\Phi(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function (e.g., polynomial and Gaussian) in the SVM classifier. In [123], the embedded feature-selection algorithm adopted logistic voting to improve the convergence speed of RFE and automatically eliminated redundant bands without having prior knowledge of the fixed band number. In [124], the modified recursive SVM implemented a new ranking criterion in the RFE-SVM training process to achieve higher classification accuracy and reduce processing time. In [125], the RFE-SVM was modified into the kernel-based feature-selection algorithm, which used the magnitude of the SVM coefficients as a ranking criterion to select the important bands and better train the SVM classifier. A convolutional neural network (CNN) was used to exploit deep features of HSI classification and then combined with an AdaBoost SVM classifier to select discriminative and complementary spectral bands [126].

Another typical model widely utilized in classifier-learning based methods is the multinomial linear regression (MLR) classifier [127], which reports the probability p of a given sample \mathbf{x}_i belonging to class c given by

$$p(y_i^c = 1 | \mathbf{x}_i, \mathbf{w}) = \frac{\exp(\mathbf{w}_j^T \mathbf{x}_i)}{\sum_{j=1}^C \exp(\mathbf{w}_j^T \mathbf{x}_i)}, \quad (20)$$

where $\mathbf{w}_j = [\mathbf{w}_j^1, \dots, \mathbf{w}_j^N]$ is the weight vector corresponding to the j th class label, C is the number of classes, and $y_i^c = 1$ refers to the training sample \mathbf{x}_i belonging to class c . To achieve an acceptable level of sparsity, a Laplacian prior was incorporated to estimate the weight \mathbf{w} using the maximum a posteriori (MAP) criterion. The objective function of sparse MLR (SMLR) [128] is defined as

$$\tilde{\mathbf{w}}_{\text{MAP}} = \underset{\mathbf{w}}{\text{argmax}} [L(\mathbf{w}) + \log p(\mathbf{w})], \quad (21)$$

where $p(\mathbf{w})$ is the sparsity-promoting Laplacian prior with $p(\mathbf{w}) \propto \exp(-\lambda \|\mathbf{w}\|_1)$ and $\text{Log}(\mathbf{w})$ is the log-likelihood function

$$\text{Log}(\mathbf{w}) = \sum_{i=1}^n \left[\sum_{j=1}^{C-1} y_j^T \mathbf{w}_j^T \mathbf{x}_i - \log \left(\sum_{j=1}^{C-1} \exp(\mathbf{w}_j^T \mathbf{x}_i) \right) \right]. \quad (22)$$

The SMLR adopts a dynamic training scheme to integrate band selection and SMLR training into a unified framework, and nonzero logistic regression coefficients correspond to the selected bands. Later, the SMLR was tested to improve the accuracy of pine, spruce, and birch tree species classification in the hyperspectral data of a boreal forest [129].

OTHER LEARNING-BASED METHODS

Band selection models can also be embedded in the learning models of target detection and endmember extraction. In [130], a band sparsity term along with a sparsity-promoting prior was incorporated to extend the sparsity-promoting iterated-constrained endmember into the objective function

$$J = \eta \frac{\text{RSS}_B}{N} + \beta \text{SSD}_B + \text{SPT} + \text{BST}, \quad (23)$$

where RSS_B is the residual sum of squares based on the convex geometry model, SSD_B is the term that describes the sum of squared distances, SPT is the band sparsity-promoting term, and the BST term describes the weighted sum of band weights. Here, η and β are regularization parameters that balance RSS_B and SSD_B in the objective function. The iterative-update scheme was implemented to autonomously determine the number of spectral bands required and to select bands while performing endmember determination and unmixing. Using a similar idea, band selection was embedded in the target detection model by adding a sparse regularization item into the CEM detector. The formulated sparse CBS [131] was solved by optimizing a continuous convex quadratic programming problem, and the proper bands and sparse CEM detector were simultaneously obtained.

A few researchers have investigated band selection using deep NNs. In [132], a subspace-partitioning band selection of distance density was proposed to optimize a pretrained CNN model. After that, an autoencoder NN was utilized to replace the CNNs, and the segmented autoencoder method [133] was presented to select the most significant bands. Attention-based CNNs in [134] selected the most informative bands with trained deep networks. The contribution map (CM) in [135], recorded discriminative band locations of each class, was incorporated into CNNs, and the formulated CM-CNNs selected more discriminative bands.

HYBRID SCHEME-BASED METHODS

Hybrid scheme-based methods implement multiple schemes to select appropriate bands. In particular, a clustering scheme was more widely combined with a ranking scheme to form hybrid algorithms. In [136], the spectral separability index (SSI) algorithm integrated clustering, ranking, and searching

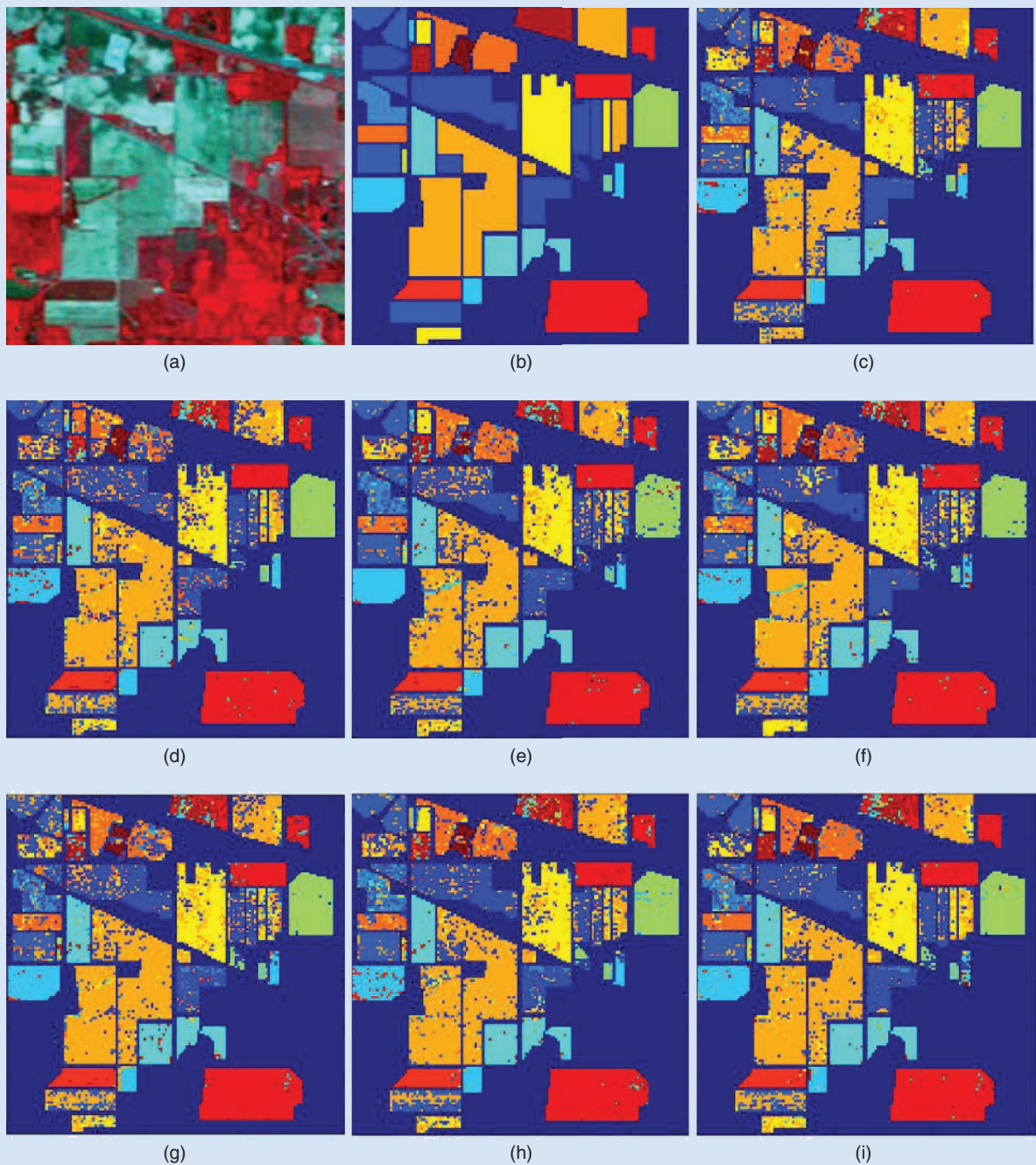


FIGURE 1. Imagery and classification maps using 30 selected bands for the Indian Pines data set. (a) The three-band false color composite; (b) the reference data; (c) all bands (OCA = 79.12%); (d) MVPCA (OCA = 70.18%); (e) MMCA (OCA = 75.79%); (f) WaluDI (OCA = 77.15%); (g) LP (OCA = 73.50%); (h) FDPC (OCA = 77.49%); and (i) ISSC (OCA = 81.61%).

schemes to select the best band combination for classification. The original bands were first grouped into several clusters based on spectral CCs and then a representative band from each subspace was selected by ranking the entropy of the bands in the same subspace. Finally, the best band combination was searched by maximizing SSI separability. A band selection framework in [137] was proposed to integrate both clustering- and ranking-based techniques, where clustering was performed on the band attributes using a density-based spatial clustering of applications with noise to remove redundant bands, and representative bands from all clusters were ranked according to their discriminating capability. In [138], the PSO was used to group all bands into different clusters, and the impurity-function band-prioritization algorithm was adopted to select the most important bands. In [139], the manifold ranking algorithm combined clustering and ranking schemes to select dissimilar bands here, the k -means algorithm was employed to cluster all bands, and representative bands of each cluster were chosen by clone selection and treated as queries. Meanwhile, other bands were ranked according to the data manifold structure of HSI data, and the dissimilar bands were expected to be selected by the query set.

Moreover, the clustering scheme can be combined with a searching scheme. A hybrid band selection strategy [140] integrated clustering and searching schemes to search for the best combination of bands with higher classification accuracy. The band grouping was implemented on a conditional MI between adjacent bands, and then the GA algorithm was adopted to search for the optimal combination in disjointed groups by maximizing classification accuracy. Additionally, the branch-and-bound search algorithm prunes irrelevant bands from the aforementioned result,

and a minimal set of relevant bands was finally obtained. The clustering scheme was combined with an evolutionary searching scheme [141], where the spatial clustering generated overall signature variation and determined cluster averages that represented different signature patterns in the image. The GA algorithm was then applied on the cluster signatures to extract the bands with the most discriminating information.

The ranking and searching schemes can also be combined. For instance, in [142], the ranking and eliminating searching schemes were combined to select the optimal bands for foreign matter classification, where all bands were first ranked according to minimum redundancy and maximum relevance. Sequential backward elimination was then applied to remove bands that contributed the least to classification.

EXPERIMENTAL RESULTS

This section details some experimental results of band selection for hyperspectral classification using two data sets: Indian Pines and the University of Pavia, Italy.

The Indian Pines data set was obtained from the Multispectral Image Data Analysis System Group at Purdue University, West Lafayette, Indiana (https://engineering.purdue.edu/~biehl/MultiSpec/aviris_documentation.html). The data set was acquired by NASA on 12 June 1992, using the airborne visible/infrared imaging spectrometer sensor from JPL. It has 20-m spatial resolutions and 10-nm spectral resolutions covering a spectrum range of 200–2,400 nm. A subset of the image scene, sized 145 × 145 pixels and covering an area of 6 mi west of West Lafayette, is used in our experiment. The data set was preprocessed with radiometric corrections and bad band removal; 200 bands remained. Figure 1(a) and (b) show the three-band false-color images and the corresponding reference samples. The 16 classes of ground objects in the image scene are detailed in Table 2.

The Pavia University (PaviaU) data set was taken from the Computational Intelligence Group at the University of the Basque Country, Spain (http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes). It was obtained from a reflective optics spectrographic imaging system (ROSIS)-3 sensor by the Deutschen Zentrum für Luft- und Raumfahrt, the German Aerospace Agency. The ROSIS-3 sensor has 115 data channels and 1.3-m spatial resolutions and covers a spectral range from 0.43 to 0.86 μm . After removing low SNR bands, 103 bands remained and were employed in our experiments. The data set contains 640 × 340 pixels and covers the University of Pavia area. Figure 2(a) and (b) illustrate the three-band false-color images of PaviaU data and its corresponding reference data. The image scene has nine classes of ground objects, as listed in Table 3.

We compared the classification performance of several typical band selection methods by changing the band subset's size. The comparison methods include unsupervised ranking-based methods, e.g., MVPCA [23] and FDPC [30],

TABLE 2. CLASSES AND SAMPLES IN THE INDIAN PINES DATA SET.

LABEL	NAME	SAMPLES
1	Alfalfa	46
2	No-till corn	1,428
3	Minimal-till corn	830
4	Corn	237
5	Grass/pasture	483
6	Grass/trees	730
7	Mowed grass/pasture	28
8	Windowed hay	478
9	Oats	20
10	No-till soybeans	972
11	Minimal-till soybeans	2,455
12	Clean soybeans	593
13	Wheat	205
14	Woods	1,265
15	Building/grass/tree drives	386
16	Stone/steel towers	93
Total		10,249

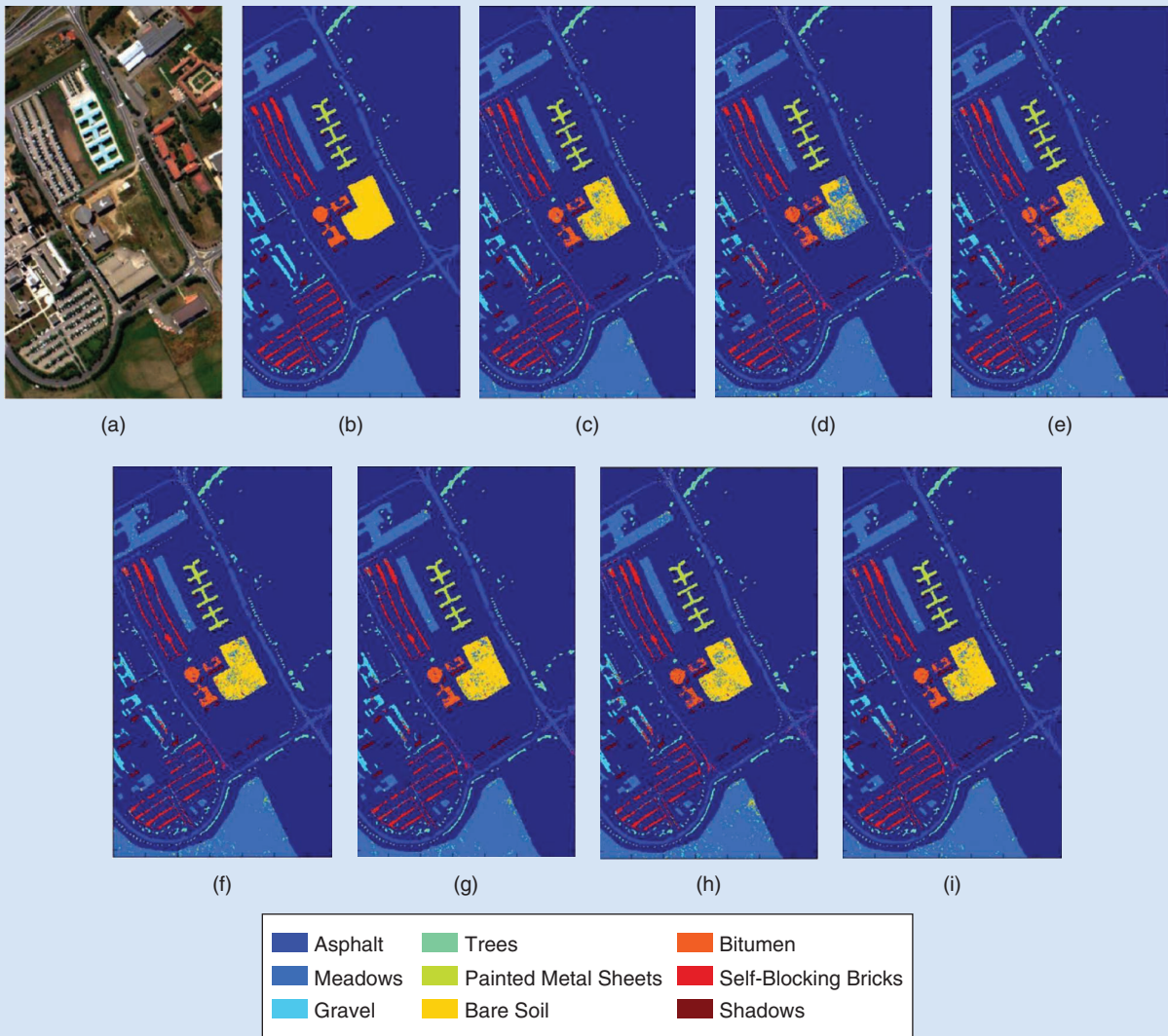


FIGURE 2. Imagery and classification maps using 30 selected bands for the Pavia University (PaviaU) data set. (a) The three-band false color composite; (b) the reference data; (c) all bands (OCA = 93.56%); (d) MVPCA (OCA = 86.04%); (e) MMCA (OCA = 92.42%); (f) WaluDI (OCA = 92.70%); (g) LP (OCA = 93.76%); (h) FDPC (OCA = 91.32%); and (i) ISSC (OCA = 94.14%).

a supervised ranking-based method (e.g., MMCA) [23], an unsupervised clustering-based method (e.g., WaluDI) [20], an unsupervised searching-based method (e.g., LP) [44], and an unsupervised sparsity-based method (e.g., ISSC) [113]. A standard SVM classifier [143] was used during the experiment and adopts the radial basis function as the kernel function; the variance parameter and penalization factor are estimated via cross-validation. For both data sets, 10% of the labeled samples are randomly selected for training, and the rest are used for testing. The overall classification accuracy (OCA), average classification accuracy, and Kappa coefficient are used to quantify classification accuracies. The following classification results are the average of 10 independent runs.

Figure 3 plots the OCA curves from different band selection methods. For the Indian Pines and PaviaU data sets, the numbers of selected bands vary from 5 to 60 with a

TABLE 3. CLASSES AND SAMPLES IN THE PAVIAU DATA SET.

LABEL	NAME	SAMPLES
1	Asphalt	6,631
2	Meadows	18,649
3	Gravel	2,099
4	Trees	3,064
5	Painted metal sheets	1,345
6	Bare soil	5,029
7	Bitumen	1,330
8	Self-blocking bricks	3,682
9	Shadows	947
Total		42,776

step interval of 5. Using cross-validation, the regularization parameters of the ISSC algorithm on the Indian Pines and PaviaU data sets are set to 0.1 and 0.05, respectively. In Figure 3(a), the ISSC algorithm outperforms other methods and obtains a higher OCA compared to using all the bands when the band subset size is over 15. The MVPCA behaves worse than other methods. ISSC and LP achieve the best OCA results, as shown in Figure 3(b). The FDPC and MVPCA methods perform worse than other methods. Moreover, Tables 4 and 5 and Figures 1 and 2 present the

detailed classification information and classification maps. The band subset sizes of both data sets are manually set to be 30. The observation coincides with that of Figure 3 and supports the performance description of all the methods. In addition, Tables 6 and 7 list the computing time of six different methods on the two data sets.

All methods are implemented in MATLAB 2014a and run on a Windows 7 computer with an Intel i5-4570 Quad Core processor and 8 GB of random-access memory. Every method takes more time when the number of selected

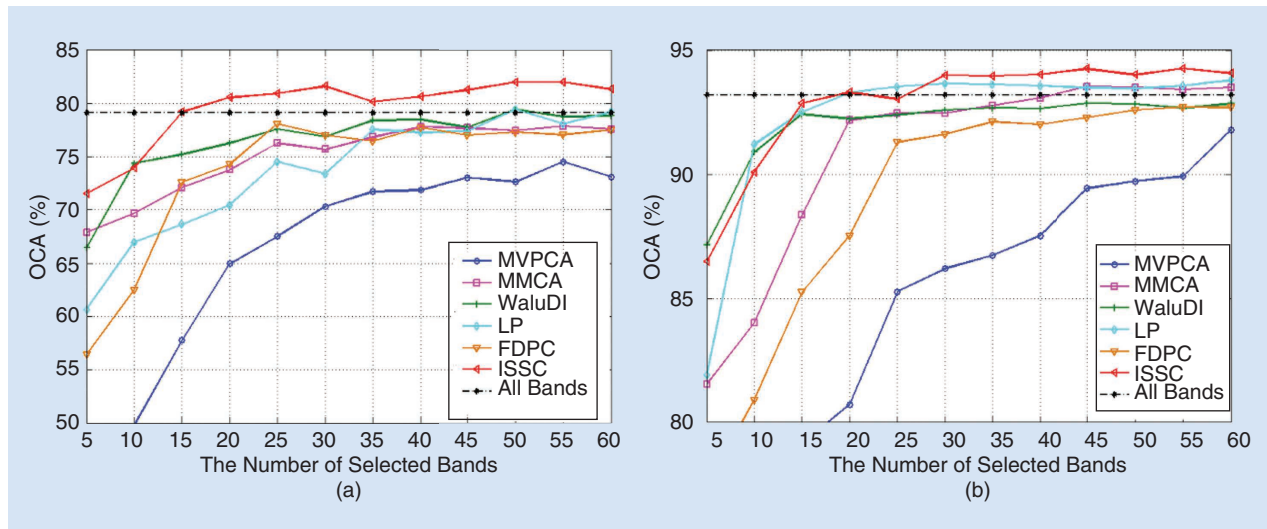


FIGURE 3. Overall classification accuracy curves of different band selection methods for the (a) Indian Pines and (b) PaviaU data sets.

TABLE 4. THE CLASSIFICATION ACCURACIES (%) ACHIEVED USING 30 SELECTED BANDS FROM THE INDIAN PINES DATA SET.

	MVPCA	MMCA	WaluDI	LP	FDPC	ISSC	ALL BANDS
Alfalfa	65.85	63.14	51.22	29.27	36.59	34.15	36.59
No-till corn	58.52	63.66	69.11	61.25	69.03	75.8	75.41
Minimal-till corn	41.5	62.92	57.7	46.72	66.27	65.19	66.8
Corn	47.89	59.15	55.4	48.83	56.81	53.05	59.14
Grass/pasture	81.38	89.2	86.44	87.36	82.99	82.07	82.53
Grass/trees	92.24	96.04	93.61	91.32	93.3	96.65	96.04
Mowed grass/pasture	44	76	44	68	32	92	56
Windrowed hay	98.6	89.53	96.28	98.14	94.88	96.74	98.61
Oats	27.78	50	22.22	33.33	22.22	61.11	38.89
No-till soybeans	59.09	68.69	75.89	63.2	66.29	79.43	66.4
Minimal-till soybean	76.14	75.55	78.27	79.95	81.62	82.53	80.76
Clean soybean	41.76	69.48	67.6	56.93	70.22	83.33	69.85
Wheat	92.93	95.65	96.74	96.2	96.74	96.2	98.91
Woods	93.59	95.52	94.38	93.59	93.5	96.84	94.38
Buildings/grass/trees/drives	41.79	51.01	49.28	51.87	48.41	53.6	52.74
Stone/steel towers	92.86	42.86	71.43	66.67	78.57	80.95	89.29
ACA	65.99	71.79	69.35	67.04	68.09	76.85	72.65
OCA	70.18	75.79	77.15	73.5	77.49	81.61	79.12
KC	65.83	72.27	73.86	69.56	74.22	78.98	76.05

NOTE: ACA: average classification accuracy; KC: Kappa coefficient.

bands increases. MVPCA takes the least amount of time; MMCA takes the second-least amount of time. ISSC is faster than FDPC, LP, and WaluDI, although it is slower than MVPCA and MMCA. The WaluDI and LP take longer than the four other methods, and the time consumed by the WaluDI is the longest.

SUMMARY AND DISCUSSION

Numerous narrow bands of hyperspectral data provide the potential for discerning the subtle spectral difference of ground objects but bring about high computational costs and information redundancy. Therefore, in recent last decades, a variety of band selection methods has been proposed for selecting proper bands and reducing the dimensionality of HSI data. In this article, we have symmetrically reviewed state-of-the-art band selection methods grouped into six categories, i.e., ranking based, searching based, clustering based, sparsity based, embedding-learning based, and hybrid-scheme based. From our research, we found that the recent development of band selection methods is unbalanced. The achievements of ranking-based, searching-based, and clustering-based methods are relatively more remarkable, particularly the former two. Sparsity-based methods have achieved more attention due to the popularity of sparsity theory.

Furthermore, we compared the classification performance of using bands chosen by several popular band selection approaches on two widely used hyperspectral data sets. Although a band selection method might behave divergently on different HSI data depending on the level of scene complexity and classification difficulty, using selected bands can yield better accuracy than using all the original bands.

The ranking-based methods have low computational complexity and are suitable for larger hyperspectral data

sets, but the classification accuracies from their selected bands are unsatisfactory. The searching-based, clustering-based, and deep-learning-based methods usually require higher computational costs and, accordingly, are more suitable for smaller hyperspectral data sets. Sparse representation-based band selection methods, e.g., ISSC [113], FRSSR [115], and SSR [117], may bring about high classification accuracy while taking lower computational costs in practice. Specifically, the ISSC and SSR methods have clear physical meanings for selected bands. Additionally, when dealing with noisy hyperspectral bands, the MSNRPCA [23], OSP [23], and minimum-noise band selection [84] methods can be considered for real-world applications.

Band selection is an important technique for hyperspectral dimensionality reduction and has room for further improvement. Based on our current research status, we recommend the future directions of hyperspectral band selection as follows.

DEVELOPING SEMISUPERVISED METHODS

Hyperspectral data typically come with a small number of labeled samples, with a majority of samples unlabeled. Unlabeled samples may assist with the mathematical modeling of band selection; for instance, they may help accurately quantify spectral redundancy. However, most existing supervised band selection methods do not utilize the information from unlabeled samples; therefore, more semisupervised methods should be developed to promote band selection performance.

DEVELOPING ACTIVE LEARNING-BASED AND DEEP-LEARNING-BASED METHODS

Current applications of active learning in hyperspectral data alleviate the act of labeling for classification. Active learning can absorb the most informative and representative training samples in low-density areas and ensure

TABLE 5. THE CLASSIFICATION ACCURACIES (%) ACHIEVED USING 30 SELECTED BANDS FROM THE PAVIAU DATA SET.

	MVPCA	MMCA	WaluDI	LP	FDPC	ISSC	ALL BANDS
Asphalt	86.54	90	90.45	93	90.06	93.4	91.62
Meadows	96.22	97.74	97.8	97.74	96.93	98.28	98.16
Gravel	72.53	70.99	71.2	79.62	74.17	79.3	77.27
Trees	83.79	91.81	93.33	93.76	91.52	92.97	89.75
Painted metal sheets	99.09	98.6	99.17	99.59	97.77	98.26	98.95
Bare soil	52.19	85.95	84.49	89.48	84.38	89.66	90.14
Bitumen	79.78	82.37	86.8	85.63	77.78	86.3	85.38
Self-blocking bricks	83.43	90.86	91.85	88.05	84.82	89.86	90.2
Shadows	99.65	100	99.53	99.88	99.06	100	99.89
ACA	83.69	89.81	90.51	91.86	88.50	92	91.26
OCA	86.04	92.42	92.70	93.76	91.32	94.14	93.56
KC	81.11	89.91	90.27	91.70	88.45	92.20	91.42

NOTE: ACA: average classification accuracy; KC: Kappa coefficient.

TABLE 6. THE COMPUTING TIMES OF DIFFERENT BAND SELECTION METHODS USED IN THE INDIAN PINES EXPERIMENT.

NUMBER OF SELECTED BANDS	COMPUTING TIME (S)					
	MVPCA	MMCA	WALUDI	LP	FDPC	ISSC
$m = 10$	0.081	0.111	596.068	23.631	2.67	0.828
$m = 20$	0.081	0.114	596.073	29.657	2.689	1.182
$m = 30$	0.086	0.112	596.076	40.021	2.702	1.663
$m = 40$	0.085	0.117	596.077	53.859	2.703	2.054
$m = 50$	0.087	0.118	596.08	71.173	2.711	2.608
$m = 60$	0.09	0.122	596.083	96.607	2.714	3.162

TABLE 7. THE COMPUTING TIMES OF DIFFERENT BAND SELECTION METHODS USED IN THE PAVIAU EXPERIMENT.

NUMBER OF SELECTED BANDS	COMPUTATIONAL TIME (S)					
	MVPCA	MMCA	WALUDI	LP	FDPC	ISSC
$m = 10$	0.285	0.467	2.001E + 3	180.107	9.205	0.599
$m = 20$	0.304	0.486	2.002E + 3	209.417	9.199	0.876
$m = 30$	0.365	0.492	2.002E + 3	262.243	9.34	1.195
$m = 40$	0.387	0.506	2.002E + 3	321.597	9.548	1.553
$m = 50$	0.407	0.514	2.002E + 3	384.186	9.468	1.832
$m = 60$	0.467	0.562	2.003E + 3	448.645	9.524	2.623

a good classification result for hyperspectral data. The idea of active learning in selecting informative samples can be transferred to select representative bands. On the other hand, deep learning can investigate various abstract features from hyperspectral data and demonstrate their powerful advantages. However, all current trials involving deep learning for band selection were made on simple CNN models, and more advanced deep networks models have not been investigated on the issue. Therefore, active learning-based and more advanced deep-learning-based band selection methods should be developed.

ESTIMATING THE NUMBER OF BANDS TO BE SELECTED

A too-small number of bands cannot preserve specific spectral information for applications, while a too-large number of bands may still contain redundancy. A few approaches have been attempted to estimate the band subset size, e.g., the concept of virtual dimensionality [21] and the Neyman–Pearson detection theory-based eigen-thresholding approach [44]; however, the estimate may be larger than actually needed. How best to estimate an appropriate size of selected bands is an open question for hyperspectral band selection.

EXPLAINING THE BAND SELECTION

Band selection preserves the spectral information of ground objects with selected bands. Current research may overemphasize the mathematical procedure of band selection but overlooks the spectral explanation for

selected bands. Taking band selection for classification as an example, each selected band should correspond to certain spectral diagnostic characteristics for separating specific classes. However, such correspondence is often unclear, particularly when band selection is modeled as a sophisticated mathematical problem. Associating selected bands with their contributions in practical applications may also improve the generation capability of a proposed approach.

EVALUATING BAND SELECTION PERFORMANCE

Band selection performance is often evaluated by classification or detection accuracy. Aside from the selected bands themselves, many potential factors such as the nature of the adopted classifier and its parameter configurations would impact classification accuracy and affect band selection performance evaluation. The objective of band selection is to reduce BC, preserve the spectral information of main ground objects, and lower the computational costs. Thus, how best to comprehensively evaluate the performance of band selection, including spectral information, intracorrelations, and computational cost, rather than only classification or detection accuracy, is an interesting topic.

ACKNOWLEDGMENTS

This work was funded by the National Science Foundation of China (NSFC) under grants 41671342 and U1609203, by the Zhejiang Provincial NSF of China under grant LR19D010001, by the NSF of Ningbo under grant

2017A610294, and by the K.C. Wong Magna Fund at Ningbo University, China.

AUTHOR INFORMATION

Weiwei Sun (nbsww@outlook.com) received his B.S. degree in surveying and mapping and his Ph.D. degree in cartography and geographic information engineering, both from Tongji University, Shanghai, China, in 2007 and 2013, respectively. He is currently an associate professor at Ningbo University, Zhejiang, China. From 2011 to 2012, he studied in the Department of Applied Mathematics at the University of Maryland, College Park, working as a visiting scholar with Prof. John Benedetto to study the dimensionality reduction of hyperspectral images. From 2014 to 2016, he studied in the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China, working as a postdoctoral researcher studying intelligent processing in hyperspectral imagery. He has published more than 70 journal papers, and his current research interests include hyperspectral image processing and manifold learning. He is a Member of the IEEE.

Qian Du (du@ece.msstate.edu) received her Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2000. She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville. She is also an adjunct professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. She received the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She has served as an associate editor of *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *IEEE Signal Processing Letters*, and *Journal of Applied Remote Sensing*. Since 2016, she has been editor-in-chief of *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. She is a fellow of the Society of Photo-Optical Instrumentation Engineers and a Fellow of the IEEE.

REFERENCES

- [1] A. F. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environment*, vol. 113, pp. S5–S16, Sept. 2009.
- [2] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, 2014.
- [3] K. S. He, D. Rocchini, M. Neteler, and H. Nagendra, "Benefits of hyperspectral remote sensing for tracking plant invasions," *Diversity Distributions*, vol. 17, pp. 381–392, Mar. 2011.
- [4] F. D. Van der Meer et al., "Multi-and hyperspectral geologic remote sensing: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 14, no. 1, pp. 112–128, 2012.
- [5] P. S. Thenkabail, J. G. Lyon, and A. Huete, *Hyperspectral Remote Sensing of Vegetation*. Boca Raton, FL: CRC, 2016.
- [6] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, 2014.
- [7] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, 2014.
- [8] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.
- [9] S. J. Hook, "NASA 2014 The hyperspectral infrared imager (HypIRI)—science impact of deploying instruments on separate platforms," JPL CA Inst. Technol., Pasadena, White Paper, 2014.
- [10] P. K. Varshney and M. K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Berlin: Springer-Verlag, 2004.
- [11] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [12] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, 2002.
- [13] M. Marshall and P. Thenkabail, "Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 205–218, Oct. 2015.
- [14] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogramm. Eng. Remote Sens.*, vol. 70, pp. 793–802, July 2004.
- [15] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, 1994.
- [16] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, 2014.
- [17] X. Miao, P. Gong, S. Swope, R. Pu, R. Carruthers, and G. L. Anderson, "Detection of yellow star thistle through band selection and feature extraction from hyperspectral imagery," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 9, pp. 1005–1015, 2007.
- [18] W. Sun et al., "UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 89, pp. 25–36, Mar. 2014.
- [19] I. Dópido, A. Villa, A. Plaza, and P. Gamba, "A quantitative and comparative assessment of unmixing-based feature extraction techniques for hyperspectral image classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 421–435, 2012.
- [20] A. Martínez-UsÓMartínez-Uso, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, 2007.
- [21] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land

- cover classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 544–554, 2012.
- [22] C.-I. Chang and K.-H. Liu, "Progressive band selection of spectral unmixing for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2002–2017, 2014.
 - [23] 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.
 - [24] J.-H. Kim, J. Kim, Y. Yang, S. Kim, and H. S. Kim, "Covariance-based band selection and its application to near-real-time hyperspectral target detection," *Opt. Eng.*, vol. 56, May 2017. doi: 10.1117/1.OE.56.5.053101.
 - [25] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, 2006.
 - [26] H.-C. Li, C.-I. Chang, L. Wang, and Y. Li, "Constrained multiple band selection for hyperspectral imagery," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2016, pp. 6149–6152.
 - [27] H. Su, Y. Sheng, and P. Du, "A new band selection algorithm for hyperspectral data based on fractal dimension," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 37, pp. 279–283, July 2008.
 - [28] M. K. Pal and A. Porwal, "Dimensionality reduction of hyperspectral data: Band selection using curve fitting," in *Proc. SPIE Asia-Pacific Remote Sensing*, 2016, pp. 98801W–98801W-7.
 - [29] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, pp. 1492–1496, June 2014.
 - [30] 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, 2016.
 - [31] S. Le Moan, A. Mansouri, Y. Voisin, and J. Y. Hardeberg, "A constrained band selection method based on information measures for spectral image color visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5104–5115, 2011.
 - [32] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, 2006.
 - [33] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 156–159, 2005.
 - [34] B. Demir and S. Ertürk, "Phase correlation based redundancy removal in feature weighting band selection for hyperspectral images," *Int. J. Remote Sens.*, vol. 29, pp. 1801–1807, Feb. 2008.
 - [35] S. Feng, Y. Itoh, M. Parente, and M. F. Duarte, "Hyperspectral band selection from statistical wavelet models," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2111–2123, 2017.
 - [36] S. Wang and C.-I. Chang, "Variable-number variable-band selection for feature characterization in hyperspectral signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 9, pp. 2979–2992, 2007.
 - [37] H. Du, H. Qi, X. Wang, R. Ramanath, and W. E. Snyder, "Band selection using independent component analysis for hyperspectral image processing," in *Proc. Applied Imagery Pattern Recognition Workshop*, 2003, pp. 93–98.
 - [38] B. Somers and G. P. Asner, "Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests," *Remote Sens. Environment*, vol. 136, pp. 14–27, Sept. 2013.
 - [39] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 2, pp. 101–106, 2004.
 - [40] Y. He, D. Liu, and S. Yi, "Recursive spectral similarity measure-based band selection for anomaly detection in hyperspectral imagery," *J. Opt.*, vol. 13, p. 015401, 2010. doi: DOI:10.1088/2040-8978/13/1/015401.
 - [41] X. Cao, X. Li, Z. Li, and L. Jiao, "Hyperspectral band selection with objective image quality assessment," *Int. J. Remote Sens.*, vol. 38, pp. 3656–3668, Mar. 2017.
 - [42] M. Kamandar and H. Ghassemian, "Maximum relevance, minimum redundancy band selection for hyperspectral images," in *Proc. 19th Iranian Conf. Electrical Engineering (ICEE)*, pp. 1–5, 2011.
 - [43] B. Wang, X. Wang, and Z. Chen, "Spatial entropy based mutual information in hyperspectral band selection for supervised classification," *Int. J. Numer. Anal. Model.*, vol. 9, no. 2, pp. 181–192, 2012.
 - [44] Q. Du, "Band selection and its impact on target detection and classification in hyperspectral image analysis," in *Proc. IEEE Workshop Advances in Techniques Analysis of Remotely Sensed Data*, 2003, pp. 374–377.
 - [45] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, 2008.
 - [46] L. C. B. dos Santos, S. J. F. Guimarães, and J. A. dos Santos, "Efficient unsupervised band selection through spectral rhythms," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1016–1025, 2015.
 - [47] L. Wang, X. Jia, and Y. Zhang, "A novel geometry-based feature-selection technique for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 1, pp. 171–175, 2007.
 - [48] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, 2004.
 - [49] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, 2011.
 - [50] M. Fauvel, C. Dechesne, A. Zullo, and F. Ferraty, "Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2824–2831, 2015.
 - [51] B. Wu, X. Wang, H. Shen, and X. Zhou, "Feature selection based on max-min-associated indices for classification of remotely sensed imagery," *Int. J. Remote Sens.*, vol. 33, pp. 5492–5512, Mar. 2012.
 - [52] B. Wu, L. Zhang, and Y. Zhao, "Feature selection via Cramer's V-test discretization for remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2593–2606, 2014.
 - [53] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, 2015.

- [54] Y. Liu, H. Xie, L. Wang, and K. Tan, "Hyperspectral band selection based on a variable precision neighborhood rough set," *Appl. Opt.*, vol. 55, no. 3, pp. 462–472, 2016.
- [55] L. Wang et al., "Band subset selection for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 4887–4898, 2017.
- [56] Y. Xu, Q. Du, and N. H. Younan, "Particle swarm optimization-based band selection for hyperspectral target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 554–558, 2017.
- [57] S. Li, Y. Zhu, D. Wan, and J. Feng, "Spectral similarity-preserving hyperspectral band selection," *Remote Sens. Lett.*, vol. 4, pp. 969–978, July 2013.
- [58] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, 2001.
- [59] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, 2007.
- [60] S. De Backer, P. Kempeneers, W. Debruyne, and P. Scheunders, "A band selection technique for spectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 319–323, 2005.
- [61] J.-P. Fang, Y.-L. Chang, H. Ren, C.-C. Lin, W.-Y. Liang, and J.-F. Fang, "A simulated annealing band selection approach for hyperspectral imagery," *Proc. SPIE*, pp. 63781G-1, Oct. 2006.
- [62] R. Huang and X. Li, "Band selection based on evolution algorithm and sequential search for hyperspectral classification," in *Proc. Int. Conf. Audio, Language and Image Processing (ICALIP)*, 2008, pp. 1270–1273.
- [63] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, 2016.
- [64] A. Ghosh, A. Datta, and S. Ghosh, "Self-adaptive differential evolution for feature selection in hyperspectral image data," *Appl. Soft Comput.*, vol. 13, no. 4, pp. 1969–1977, 2013.
- [65] J. Gao, Q. Du, L. Gao, X. Sun, and B. Zhang, "Ant colony optimization-based supervised and unsupervised band selections for hyperspectral urban data classification," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 085094–085094, 2014.
- [66] S. Zhou, J.-p. Zhang, and S. B.-k., "Feature selection and classification based on ant colony algorithm for hyperspectral remote sensing images," in *Proc. 2nd Int. Congr. Image and Signal Processing (CISP)*, Oct. 2009, pp. 1–4.
- [67] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images," *Pattern Recogn.*, vol. 51, pp. 295–309, Mar. 2016.
- [68] 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, 2016.
- [69] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, 2007.
- [70] J. Yin, Y. Wang, and J. Hu, "A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 935–943, 2012.
- [71] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, 2014.
- [72] A. Shi, H. Gao, Z. He, M. Li, and L. Xu, "A hyperspectral band selection on game theory and differential evolution algorithm," *Int. J. Smart Sens. Intell. Syst.*, vol. 9, no. 4, pp. 1971–1990, 2016.
- [73] T. Imbiriba, J. C. M. Bermudez, and C. Richard, "Band selection for nonlinear unmixing of hyperspectral images as a maximal clique problem," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2179–2191, 2017.
- [74] J. Feng, L. Jiao, X. Zhang, and T. Sun, "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4092–4105, 2014.
- [75] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, 2015.
- [76] R. Y. Nakamura, L. M. G. Fonseca, J. A. Dos Santos, R. d. S. Torres, X.-S. Yang, and J. P. Papa, "Nature-inspired framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2126–2137, 2014.
- [77] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson, "A novel feature selection approach based on FODPSO and SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2935–2947, 2015.
- [78] L. Feng, A.-H. Tan, M.-H. Lim, and S. W. Jiang, "Band selection for hyperspectral images using probabilistic memetic algorithm," *Soft Comput.*, vol. 20, no. 12, pp. 4685–4693, 2016.
- [79] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Gray Wolf optimizer for hyperspectral band selection," *Appl. Soft Computing*, vol. 40, pp. 178–186, Mar. 2016.
- [80] M. Wang, Y. Wan, Z. Ye, X. Gao, and X. Lai, "A band selection method for airborne hyperspectral image based on chaotic binary coded gravitational search algorithm," *Neurocomputing*, vol. 273, pp. 57–67, 2018.
- [81] K. Sun, X. Geng, and L. Ji, "A band selection approach for small target detection based on CEM," *Int. J. Remote Sens.*, vol. 35, pp. 4589–4600, June 2014.
- [82] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7111–7119, 2014.
- [83] K. Sun, X. Geng, and L. Ji, "An efficient unsupervised band selection method based on an autocorrelation matrix for a hyperspectral image," *Int. J. Remote Sens.*, vol. 35, pp. 7458–7476, Oct. 2014.
- [84] K. Sun, X. Geng, L. Ji, and Y. Lu, "A new band selection method for hyperspectral image based on data quality," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2697–2703, 2014.

- [85] X. Geng, K. Sun, L. Ji, H. Tang, and Y. Zhao, "Joint skewness and its application in unsupervised band selection for small target detection," *Sci. Rep.*, vol. 5, p. 9915, Apr. 2015.
- [86] A. Datta, S. Ghosh, and A. Ghosh, "Band elimination of hyperspectral imagery using partitioned band image correlation and capacity discrimination," *Int. J. Remote Sens.*, vol. 35, pp. 554–577, June 2014.
- [87] C. Cariou, K. Chehdi, and S. Le Moan, "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 565–569, 2011.
- [88] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, 2017.
- [89] X. Cao, B. Wu, D. Tao, and L. Jiao, "Automatic band selection using spatial-structure information and classifier-based clustering," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4352–4360, 2016.
- [90] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, 2011.
- [91] R. Yang, L. Su, X. Zhao, H. Wan, and J. Sun, "Representative band selection for hyperspectral image classification," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 396–403, Oct. 2017.
- [92] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret, "Band selection in RKHS for fast nonlinear unmixing of hyperspectral images," in *Proc. 23rd European Signal Processing Conf. (EUSIPCO)*, 2015, pp. 1651–1655.
- [93] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, 2016.
- [94] B. Mojaradi, H. Emami, M. Varshosaz, and S. Jamali, "A novel band selection method for hyperspectral data analysis," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 37, pp. 447–454, July 2008.
- [95] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, 2011.
- [96] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vision*, vol. 3, no. 4, pp. 213–222, 2009.
- [97] C. Yang, S. Liu, L. Bruzzone, R. Guan, and P. Du, "A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1152–1156, 2013.
- [98] C. Yang, Y. Tan, L. Bruzzone, L. Lu, and R. Guan, "Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images," *Remote Sens.*, vol. 9, no. 8, p. 782, 2017.
- [99] J. Feng, L. Jiao, T. Sun, H. Liu, and X. Zhang, "Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6516–6530, 2016.
- [100] H. Su, Y. Sheng, P. Du, and K. Liu, "Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection," *Appl. Opt.*, vol. 51, no. 14, pp. 2656–2663, 2012.
- [101] L. Jiao, J. Feng, F. Liu, T. Sun, and X. Zhang, "Semisupervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2760–2773, 2015.
- [102] S. Li, J. Qiu, X. Yang, H. Liu, D. Wan, and Y. Zhu, "A novel approach to hyperspectral band selection based on spectral shape similarity analysis and fast branch and bound search," *Eng. Appl. Artif. Intell.*, vol. 27, pp. 241–250, Jan. 2014.
- [103] V. Kumar, J. Hahn, and A. M. Zoubir, "Band selection for hyperspectral images based on self-tuning spectral clustering," in *Proc. 21st European Signal Processing Conf. (EUSIPCO)*, 2013, pp. 1–5.
- [104] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 51–64, 2017.
- [105] R. Hedjam and M. Cheriet, "Hyperspectral band selection based on graph clustering," in *Proc. 11th Int. Conf. Information Science, Signal Processing and Their Applications (ISSPA)*, 2012, pp. 813–817.
- [106] G. Zhu, Y. Huang, J. Lei, Z. Bi, and F. Xu, "Unsupervised hyperspectral band selection by dominant set extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239, 2016.
- [107] W. Xia, B. Wang, and L. Zhang, "Band selection for hyperspectral imagery: A new approach based on complex networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1229–1233, 2013.
- [108] J.-m Li and Y.-t Qian, "Clustering-based hyperspectral band selection using sparse nonnegative matrix factorization," *J. Zhejiang Univ.-Sci. C*, vol. 12, no. 7, pp. 542–549, 2011.
- [109] W. Sun, W. Li, J. Li, and Y. M. Lai, "Band selection using sparse nonnegative matrix factorization with the thresholded earth's mover distance for hyperspectral imagery classification," *Earth Sci. Informat.*, vol. 8, no. 4, pp. 907–918, 2015.
- [110] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. 18th IEEE Int. Conf. Image Processing (ICIP)*, 2011, pp. 2693–2696.
- [111] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, 2015.
- [112] Q. Du, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral band selection using a collaborative sparse model," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2012, pp. 3054–3057.
- [113] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, 2015.
- [114] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Squaring weighted low-rank subspace clustering for hyperspectral image band selection," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2016, pp. 2434–2437.

- [115] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5087–5098, 2017.
- [116] W. Sun, L. Zhang, L. Zhang, and Y. M. Lai, "A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4374–4388, 2016.
- [117] W. Sun, M. Jiang, W. Li, and Y. Liu, "A symmetric sparse representation based band selection method for hyperspectral imagery classification," *Remote Sens.*, vol. 8, no. 3, p. 238, 2016.
- [118] Z. Guo, H. Yang, X. Bai, Z. Zhang, and J. Zhou, "Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, 2013, pp. 1474–1477.
- [119] B. B. Damodaran, N. Courty, and S. Lefèvre, "Sparse Hilbert Schmidt independence criterion and surrogate-kernel-based feature selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2385–2398, 2017.
- [120] K. Sun, X. Geng, and L. Ji, "A new sparsity-based band selection method for target detection of hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 329–333, 2015.
- [121] H. Li, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Group sparsity based semi-supervised band selection for hyperspectral images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2013, pp. 3225–3229.
- [122] Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, "Semisupervised hyperspectral band selection via spectral-spatial hypergraph model," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, 2015.
- [123] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, 2007.
- [124] R. Zhang and J. Ma, "Feature selection for hyperspectral data based on recursive support vector machines," *Int. J. Remote Sens.*, vol. 30, pp. 3669–3677, July 2009.
- [125] B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 317–326, 2014.
- [126] V. Sharma, A. Diba, T. Tuytelaars, and L. Van Gool, "Hyperspectral CNN for image classification & band selection, with application to face recognition," KU Leuven, Belgium, Rep. KUL/ESAT/PSI/1604, 2016. [Online]. Available: <http://www.esat.kuleuven.be/psi>
- [127] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, 1992.
- [128] P. Zhong, P. Zhang, and R. Wang, "Dynamic learning of SMLR for feature selection and classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 280–284, 2008.
- [129] P. Pant, V. Heikkinen, I. Korpela, M. Hauta-Kasari, and T. To-kola, "Logistic regression-based spectral band selection for tree species classification: Effects of spatial scale and balance in training samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 9, pp. 1604–1608, 2014.
- [130] A. Zare and P. Gader, "Hyperspectral band selection and end-member detection using sparsity promoting priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 256–260, 2008.
- [131] X. Geng, K. Sun, and L. Ji, "Band selection for target detection in hyperspectral imagery using sparse CEM," *Remote Sens. Lett.*, vol. 5, pp. 1022–1031, Dec. 2014.
- [132] Y. Zhan, D. Hu, H. Xing, and X. Yu, "Hyperspectral band selection based on deep convolutional neural network and distance density," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2365–2369, 2017.
- [133] J. Tschannerl, J. Ren, J. Zabalza, and S. Marshall, "Segmented autoencoders for unsupervised embedded hyperspectral band selection," in *Proc. 7th European Workshop Visual Information Processing*, 2018, pp. 1–6.
- [134] P. R. Lorenzo, L. Tulczyjew, M. Marcinkiewicz, and J. Nalepa, "Band selection from hyperspectral images using attention-based convolutional neural networks. 2018. [Online]. Available: <https://arxiv:1811.02667>
- [135] R. Cai, Y. Yuan, and X. Lu, "Hyperspectral band selection with convolutional neural network," in *Proc. Chinese Conf. Pattern Recognition and Computer Vision (PRCV)*, 2018, pp. 396–408.
- [136] J. Yin, Y. Wang, and Z. Zhao, "Optimal band selection for hyperspectral image classification based on inter-class separability," in *Proc. Symp. Photonics and Optoelectronic (SOPO)*, 2010, pp. 1–4.
- [137] A. Datta, S. Ghosh, and A. Ghosh, "Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 12, pp. 2814–2823, 2015.
- [138] Y.-L. Chang, J.-N. Liu, Y.-L. Chen, W.-Y. Chang, T.-J. Hsieh, and B. Huang, "Hyperspectral band selection based on parallel particle swarm optimization and impurity function band prioritization schemes," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014. doi: 10.1117/1.JRS.8.084798.
- [139] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, 2016.
- [140] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Syst.*, vol. 24, no. 1, pp. 40–48, 2011.
- [141] A. Paul, S. Bhattacharya, D. Dutta, J. R. Sharma, and V. K. Dadhwal, "Band selection in hyperspectral imagery using spatial cluster mean and genetic algorithms," *GIScience & Remote Sens.*, vol. 52, no. 6, pp. 643–659, 2015.
- [142] Y. Jiang and C. Li, "mRMR-based feature selection for classification of cotton foreign matter using hyperspectral imaging," *Comput. Electron. Agriculture*, vol. 119, pp. 191–200, Nov. 2015.
- [143] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, 1998.