

Journal of  
**Applied Remote Sensing**

RemoteSensing.SPIEDigitalLibrary.org

**Band clustering using expectation–  
maximization algorithm and  
weighted average fusion-based  
feature extraction for hyperspectral  
image classification**

Manoharan Prabukumar  
Sawant Shrutika

**SPIE.**

Manoharan Prabukumar, Sawant Shrutika, “Band clustering using expectation–maximization algorithm and weighted average fusion-based feature extraction for hyperspectral image classification,” *J. Appl. Remote Sens.* **12**(4), 046015 (2018), doi: 10.1117/1.JRS.12.046015.

# Band clustering using expectation–maximization algorithm and weighted average fusion-based feature extraction for hyperspectral image classification

Manoharan Prabukumar\* and Sawant Shrutika

Vellore Institute of Technology, School of Information Technology and Engineering,  
Vellore, India

**Abstract.** The presence of a significant amount of information in the hyperspectral image makes it suitable for numerous applications. However, extraction of the suitable and informative features from the high-dimensional data is a tedious task. A feature extraction technique using expectation–maximization (EM) clustering and weighted average fusion technique is proposed. Bhattacharya distance measure is used for computing the distance among all the spectral bands. With this distance information, the spectral bands are grouped into the clusters by employing the EM clustering method. The EM algorithm automatically converges to an optimum number of clusters, thereby specifying the absence of need for the required number of clusters. The bands in each cluster are fused together applying the weighted average fusion method. The weight of each band is calculated on the basis of the criteria of minimizing the distance inside the cluster and maximizing the distance among the different clusters. The fused bands from each cluster are then considered as the extracted features. These features are used to train the support vector machine for classification of the hyperspectral image. The performance of the proposed technique has been validated against three small-size standard bench-mark datasets, Indian Pines, Pavia University, Salinas, and one large-size dataset, Botswana. The proposed method achieves an overall accuracy (OA) of 92.19%, 94.10%, 93.96%, and 84.92% for Indian Pines, Pavia University, Salinas, and Botswana datasets, respectively. The experimental results prove that the proposed technique attains significant classification performance in terms of the OA, average accuracy, and Cohen’s kappa coefficient ( $k$ ) when compared to the other competing methods. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JRS.12.046015](https://doi.org/10.1117/1.JRS.12.046015)]

**Keywords:** hyperspectral image; band clustering; expectation–maximization clustering; weighted average fusion; classification.

Paper 180488 received Jun. 9, 2018; accepted for publication Oct. 11, 2018; published online Nov. 2, 2018.

## 1 Introduction

Hyperspectral sensors [e.g., Airborne Visible Infrared Imaging Spectrometer (AVIRIS), hyperspectral digital imagery collection experiment, HyMap, and EO-1 Hyperion] record a scene over the wide wavelength ranging from the visible region to the infrared spectrum, which provides detailed spectral information about the objects in numerous and continuous spectral bands (from tens to several hundreds) as well as a high spatial resolution.<sup>1</sup> Due to the high spectral resolution, hyperspectral images offer very high-discrimination capabilities among similar ground cover objects.<sup>2</sup> However, the huge numbers of bands always bring the curse of dimensionality, reducing the discriminating ability of the data as the dimensionality increases with fewer numbers of labeled training samples.<sup>3,4</sup> This behavior is also referred to as “Hughes phenomenon.”<sup>5</sup> Moreover, the high dimensionality of the hyperspectral image also consists of redundant and noisy information, which increases the computational burden of the data processing. So dimensionality reduction becomes an essential task in the hyperspectral image processing.

Dimensionality reduction is the process of reducing redundant data and extracting meaningful features. In other words, dimensionality reduction is a convenient way of reducing the

---

\*Address all correspondence to: Prabukumar Manoharan, E-mail: [mprabukumar@vit.ac.in](mailto:mprabukumar@vit.ac.in)

number of spectral bands and transforming the data from a high-dimensional space to a lower dimensional space, where the most significant information is conserved.<sup>6,7</sup> Dimensionality reduction can be done through the feature selection or feature extraction method. In the feature selection method, a few informative bands are selected on the basis of the adopted selection criteria, namely, the distance measures (Euclidean distance, spectral angle mapping, Bhattacharyya distance, Hausdorff distance, and Jeffreys–Matusita distance), information theoretic approaches (divergence, transformed divergence, and mutual information), and Eigen analysis [principal component analysis (PCA)], where the original physical significant properties of the bands can be preserved.<sup>8–15</sup> One of the popular band selection methods is the constrained band selection (CBS) method.<sup>9</sup> It minimizes the correlation and dependency in the selection of the bands. Based on correlation and dependence, CBS method offers four different approaches, which arise from two different approaches: (1) constrained energy minimization (CEM) and (2) linearly constrained minimum variance (LCMV). There are four specific criteria for band selection such as, band correlation minimization (BCM), band correlation constraint (BCC), band dependence minimization (BDM), and band dependence constraint (BDC). These four criteria divide the CEM and LCMV approach into to four parts: CEM-BCC/BDC, CEM-BCM/BDM, LCMV-BCC/BDC, and LCMV-BCM/BDM. Feature selection provides suitable features for classification but is computationally expensive and often not robust in complex scenes (variation in spectral signatures across scenes). On the other hand, feature extraction methods transform the higher dimensional data into the lower dimensional space. They are computationally superior and more robust to the complex scenes. However, extraction of efficient and suitable features in the classification of large hyperspectral data is a highly crucial task.

Feature extraction methods transform the original high-dimensional feature space into a low-dimensional feature space, which faces loss of the physical meaning of the bands but preserves the significant discriminative information needed for further analysis.<sup>16–25,26</sup> PCA is one of the most widely used approaches for feature extraction.<sup>16</sup> This is due to the fact of PCA being an invertible transformation, which facilitates the interpretation of the extracted features. PCA offers high-computational load and operates on the global features but loses local information.<sup>27</sup> The extension of PCA, segmented PCA method,<sup>17</sup> is presented for addressing this issue. Here, for using the local information, PCA is applied to the groups of bands formed using the correlation between bands. Another most useful feature extraction method is independent component analysis (ICA),<sup>19</sup> which is used for the extraction of class discriminant features from the hyperspectral images. But the complexity of ICA method increases the computational load. In general, the hyperspectral data are nonlinear in nature. Hence, the linear classifier usually provides unsatisfied classification performance. In recent times, some nonlinear methods such as maximum noise fraction<sup>20</sup> and kernel PCA,<sup>21</sup> and probabilistic PCA (PPCA) are proposed as an extension to the conventional PCA. PPCA is a constraint Gaussian generative latent variable model. PPCA extracts features using the maximum likelihood estimates for the parameters associated with the covariance matrix that can be efficiently calculated from the data principal component.<sup>22</sup> In most of the situations, the labeled samples are limited and obtaining the labeled samples is a very expensive and time-consuming task. On the other hand, unlabeled samples are available in large quantities at low cost. Hence, semisupervised PPCA is proposed as an extension of PPCA, which uses both the labeled as well as unlabeled information into the projection for overcoming the problem of the scarcity of the labeled samples.<sup>18</sup> Apart from the PCA, there are two other best known feature extraction approaches, discriminant analysis feature extraction<sup>28</sup> and linear discriminant analysis (LDA).<sup>23</sup> In recent times, many other extensions to the above-mentioned two methods have been proposed, namely, regularized LDA,<sup>23</sup> nonparametric weighted feature extraction (NWFE),<sup>24</sup> and kernel NWFE.<sup>25</sup> Another most popular feature extraction approach is the clustering-based feature extraction (CBFE). Clustering makes partitions of the hyperspectral image into several uncorrelated subband groups, each of which contains contiguous bands. Clustering has received increasing attention in the hyperspectral remote sensing community due to its better performance toward the curse of dimensionality problem.<sup>29–35</sup> Clustering technique removes redundancies and the correlated data from the high-dimensional data and provides uncorrelated low-dimensional data. In Ref. 30, CBFE is proposed. It works well in a small sample size scenario using the most popular *k*-means clustering algorithm. A semisupervised *k*-means clustering method is proposed for utilizing the easily

available unlabeled samples.<sup>36</sup> It uses the multiple classifiers for each cluster of band and the final output is the fused result of the multiple classifiers. Clustering methods do not require *a priori* knowledge in advance to the band grouping process, but make the cluster of the bands as per the distribution of the spectral features of hyperspectral image. Moreover, clustering methods are too sensitive to the randomly initialized cluster center and selected subset of bands may be unstable. Hence, in Ref. 37, an automatic clustering method [fast density peak-based clustering (FDPC)] is proposed, which selects the cluster centers using the fast search method. But it is not a fully automatic cluster center selection method and loses the data points. Hence, improvements in FDPC are proposed, namely, enhanced fast density peak-based clustering (E-FDPC),<sup>38</sup> and *k*-means fast density peak-based clustering.<sup>39</sup> Dual clustering-based band selection by context analysis (DCCA)<sup>33</sup> does the clustering by considering the context information in the bands of the hyperspectral image. Recently, along with the algorithm development for the hyperspectral image classification, fusion methods such as decision level and feature level fusion methods have gained great interest,<sup>40–43</sup> and these methods demonstrated the ability of the combination of the selected features to improve the classification performance. Considering the above study of the feature extraction techniques, the authors of this work found the following challenges:

1. Though the existing clustering-based feature extraction approaches show a significant performance, the emphasis of these conventional clustering strategies is on raw spectral features rather than exploiting more complementary information from the bands of the hyperspectral cube.
2. The existing clustering-based feature extraction approaches fail to find an optimal number of clusters and are very sensitive to the number of clusters.
3. The existing feature extraction methods work well in small-size data, but fail to show the effectiveness in the case of the large-size data.

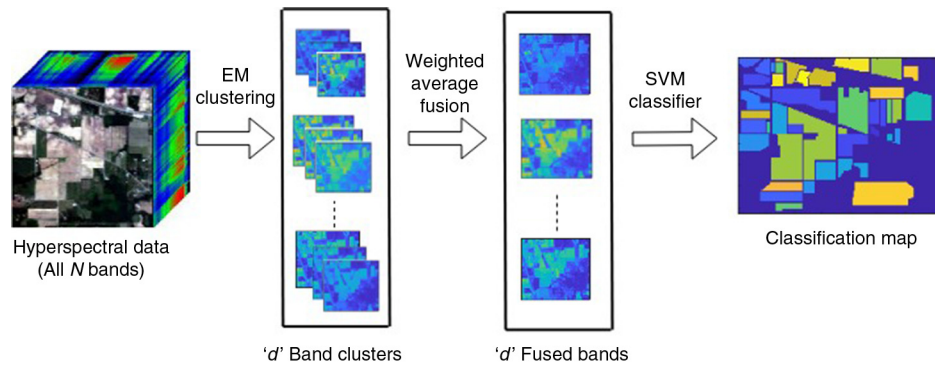
The main contributions of the proposed method are summarized as follows.

1. An effective expectation–maximization clustering and weighted average fusion (EM-WAF)-based feature extraction method is proposed for the hyperspectral image classification.
2. The EM algorithm automatically converges to an optimal number of clusters. Therefore, the proposed technique circumvents the necessity to specify the number of clusters by making the use of the EM clustering algorithm.
3. The bands from each cluster are combined by adopting the weighted average fusion method. This process usually improves the classification performance by giving more weight to the particular band, thereby providing more discriminative and complementary information. Calculation of the weight is done on the basis of the criteria of minimizing the intracluster distance and maximizing the intercluster distance. The fused bands obtained from each cluster are then considered as extracted features, which are further used for the hyperspectral image classification.
4. Finally, the experimentation is done on both small and large-size datasets to prove the effectiveness of the proposed method.

The remainder of this paper is arranged as follows: in Sec. 2, the proposed architecture of EM clustering and weighted average fusion-based hyperspectral image classification is explained in detail. Mathematical details of EM clustering and weighted average fusion are also discussed. Experimental analysis of four standard datasets is presented in Sec. 3. More precisely, the proposed method is compared with other clustering and fusion-based methods. Comparison is done for both quantitative accuracy and visual interpretation. Section 4 provides the concluding remarks.

## 2 Proposed Architecture

This section discusses the proposed architecture of the feature extraction for hyperspectral image classification in detail. The proposed feature extraction architecture is presented in Fig. 1, which



**Fig. 1** The architecture of the proposed EM-WAF method for hyperspectral image classification.

depicts the proposed approach as comprising three stages, namely, band clustering, the fusion of the bands of each cluster, and classification. The following section provides a detailed explanation of the various stages present in the proposed system.

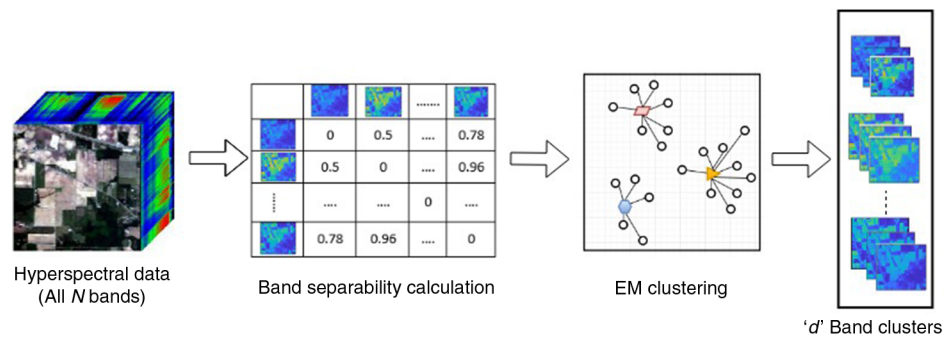
### 2.1 Band Clustering

Hyperspectral data consist of the hundreds of spectral bands, which are highly redundant due to similar sensor responses in two adjacent bands. The objective of the band clustering is to group the highly correlated bands and group them into distant clusters. Figure 2 shows the workflow of the band clustering procedure. Here the Bhattacharya distance<sup>28</sup> is used as band separability measure for computing the distance between each pair of spectral bands. The Bhattacharya distance between bands  $b_i$  and  $b_j$  is defined as

$$b_{i,j} = \frac{1}{8}(\mu_i - \mu_j)^T \left( \frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left[ \frac{|\Sigma_i + \Sigma_j|/2}{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}} \right]. \quad (1)$$

Here,  $\mu_i$  and  $\mu_j$  are band means,  $\Sigma_i$  and  $\Sigma_j$  are band covariance matrices.

Using the distance information, the bands are clustered using the EM clustering algorithm. The band clustering procedure using the EM clustering algorithm is explained in detail in the following section.



**Fig. 2** The band clustering procedure. In this procedure, the pairwise band separability information is calculated, and then EM clustering is conducted to generate “d” band clusters.

### 2.1.1 Band clustering using EM algorithm

Using the generated distances between each pair of spectral bands, all the original bands are grouped into “ $d$ ” clusters. Clustering is done using the EM algorithm.

The EM clustering algorithm features the partial allotment of points to different clusters instead of assigning them to the closest cluster center. This can be achieved by modeling each cluster using the probabilistic distribution. Finally, the algorithm is converged into the cluster with the highest probability. The  $K$ -means clustering algorithm is an incremental heuristic approach, whereas the EM algorithm is a statistical algorithm that assumes a statistical model that describes the data. The assumption of the EM algorithm to cluster analysis is that the patterns are drawn from one or several distributions. The goal here is to identify the parameters of each distribution. In this case, the parameters of a Gaussian mixture model have to be estimated. The EM algorithm<sup>44</sup> is a probabilistic model used for finding the maximum likelihood estimates of the parameters from the patterns. Assume that bands belonging to the same cluster are drawn from a multivariate Gaussian probability distribution for forming the cluster of bands. The EM clustering algorithm converges to an optimal value of the clusters. It is considered as converged when there is no further change in the assignment of the bands to cluster. The EM clustering algorithm is explained in Algorithm 1.

---

#### Algorithm 1 Band clustering using EM algorithm.

---

**Input:**  $b = \{b_1, b_2, b_3, \dots, b_n\}$  be the set of the bands and  $C = \{C_1, C_2, \dots, C_c\}$  be the set of centroid centers, max\_iteration  $k$ .

**Output:** An optimal number of “ $d$ ” band clusters.

##### Step 1: Initialization

- i) Initially select  $c$  bands randomly from the set  $b$  as cluster center. Let us consider,  $\mu_j$  is the mean,  $\Sigma_j$  is covariance matrix, and  $\alpha_j$  is the weight. Each cluster  $C_j$  is represented by a Gaussian distribution  $N(\mu_j, \Sigma_j)$  and  $\alpha_j$ .

##### Step 2: Iteration

- i) While (iteration  $< k$ )
- ii) Expectation step (E-step)

Assign each band to one of the clusters according to the maximum a posteriori probability criteria.

The probability of cluster  $C_j$  over  $b_i$ , for each distance point  $b_i$  and each cluster  $C_j$ :

$$p(C_j|b_i) = \frac{p(b_i|C_j)p(C_j)}{\sum_j p(b_i|C_j)p(C_j)}. \quad (2)$$

The probability density function  $p(b_i|C_j)$  for a bivariate Gaussian distribution is given by

$$p(b_i|C_j) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} e^{\left[ -\frac{1}{2}(b_i - \mu_j)^T \Sigma_j^{-1} (b_i - \mu_j) \right]}. \quad (3)$$

- iii) Maximization step (M-step):

Recompute the parameter values  $\mu_j$ ,  $\Sigma_j$ , and  $\alpha_j$  for the cluster  $C_j$  by using the probability  $p(C_j|b_i)$  obtained in expectation step.

The mean  $\mu_j$  is computed as

$$\mu_j = \frac{\sum_i p(C_j|b_i) b_i}{\sum_i p(C_j|b_i)}. \quad (4)$$


---



The covariance matrix  $\Sigma_j$  is computed as

$$\Sigma_j = \frac{\sum_i p(C_j|b_i)(b_i - \mu_j)(b_i - \mu_j)^T}{\sum_i p(C_j|b_i)}. \quad (5)$$

The weight  $\alpha_j$  is given as

$$\alpha_j = \frac{\sum_i p(C_j|b_i)}{N}. \quad (6)$$

where  $N$  is the total number of bands.

- iv) Eliminate the cluster  $C$  if  $p(C_j|b_i)$  is less. The bands that belonged to the deleted clusters will be reassigned to the other clusters in the next iteration.

### Step 3: Stopping criteria

- i) If the convergence criterion is not achieved, repeat the step 2.
- 

## 2.2 Weighted Average Fusion

Following the band clustering process, all the bands from each cluster are fused together using the weighted average fusion method. The fused bands should have the following characteristics:

1. *Decorrelation.* Correlation among the clusters should be greatly reduced.
2. *Separability.* Discrimination capability of fused bands should be increased.

The simple average fusion method proposed in Ref. 29 does not ensure any satisfactory way for removing redundant information. Hence, here, the weighted average fusion method is used for the preservation of the discriminative information of the original bands. Since the weight factor preserves the discriminative information of the original bands, it improves the classification results. Therefore,  $m$  bands in  $d$ 'th cluster are fused as shown in

$$F_d = \frac{\sum_{j \in m} w_d(j) * b_j}{m} \quad \forall d, \quad (7)$$

where  $b_j$  is the  $j$ 'th band in  $d$ 'th cluster, and  $w_d(j)$  is the weight factor for  $j$ 'th band in  $d$ 'th cluster. Here we provide each band a weight value of  $w$ . An optimal weight value of each band is determined by updating the weight value  $w$ .<sup>45</sup> Let the sum of band weight in each cluster be one, i.e.,

$$\sum_{j \in d} w_d(j) = 1. \quad (8)$$

The initial weight value of each band is evaluated by considering the variance of each band. The initial value of weight  $w_d^0(j)$  is calculated as:

$$w_d^0(j) = \frac{s_j}{\sum_{i \in N} s_i}, \quad (9)$$

where  $s_j$  represents variance of  $j$ 'th band image and  $N$  represents the total number of bands in the hyperspectral image data.

The weight updating procedure is iterated for  $t$  times for finding the optimal weight value of each band. The weight value  $w_d^t(j)$  is determined using the following equation:

$$w_d^t(j) = \alpha \left[ w_d^0(j) + \sum_{b_i \in d} x(b_i, b_j) w_d^{t-1}(i) \right] - \frac{1 - \alpha}{d - 1} \sum_{d=2,3,\dots,d} \sum_{b_i \in d} x(b_i, b_j) w_d^{t-1}(i), \quad (10)$$

where  $t$  is the number of iterations,  $\alpha$  is the balance factor between first and second term of Eq. (10), and  $x(b_i, b_j)$  is the distance between band  $b_i$  and  $b_j$  calculated by using Eq. (1).

In this propagation process, each time updating one band's weight is done using all other information relating to the bands based on the distance between them. This process continues until all bands in the cluster have been updated once. The weight updating procedure indicated in Eq. (10) ensures following two characteristics of fused bands. The first term measures the compactness within the same cluster, whereas the second term measures the scatteredness among the discriminative clusters. There exists a concise form for Eq. (10):

$$w_d^t(j) = \alpha w_d^0(j) + a(x_i, x_j)A_{ij}, \quad (11)$$

where

$$a(x_i, x_j) = \sum_{b_i \in d} x(b_i, b_j) w_d^{t-1}(i). \quad (12)$$

The coefficient matrix  $A$  is defined as

$$A = \begin{cases} \alpha, & \text{if } b_i, b_j \in d \\ \frac{1-\alpha}{d-1} \sum_{d=2,3,\dots,d}, & \text{if } b_j \in d \text{ and } b_i \notin c_d. \end{cases} \quad (13)$$

Following the  $t$  iterations, the weight value of band  $b_j$  is chosen by maximizing Eq. (10), i.e.,

$$w_d(j) = \arg \max_{w_d^t(j), j \in d} \in [w_d^t(j)] \quad \forall t. \quad (14)$$

Then weight value in each band cluster is normalized as follows:

$$w_d(j) = \frac{w_d(j)}{\sum_{b_j \in d} w_d(j)}. \quad (15)$$

Calculation of the weighted average of bands in each subgroup removes the noise from bands and also the redundant information for each subgroups. Weighted average fusion decorrelates the intercorrelated hyperspectral bands into a set of uncorrelated bands. The fused bands  $F_d$  from each cluster are then considered as set of extracted features. After fusion of bands using the weighted average fusion technique, the actual classification is performed with SVM classifier. The extracted features are used for training the SVM classifier. Its remarkable benefits in solving the complex problems such as nonlinear and high dimensionality of the data and limited training samples make the SVM classifier the most commonly used in the hyperspectral image classification.<sup>46</sup>

### 2.3 Computational Cost Analysis

In this section, the theoretical computational cost of the proposed EM-WAF method is discussed. Both the arithmetic operations and the big  $O$  notation are used for calculation of the computational cost. The theoretical computational cost of the proposed method depends on four steps, namely, the Bhattacharya distance-based band distance measure, the EM band clustering, the weighted average fusion, and SVM classifier. The computational cost of the Bhattacharya distance measure for all pairs of bands scales is  $O(n^2)$ , where  $n$  is the number of the spectral bands. The computational cost of EM clustering method is  $O(nkd)$ , where  $k$  is the number of iterations in EM clustering and  $d$  is the number of clusters formed. In the weighted average fusion, the computation cost comes mainly from Eq. (10), which scales as  $O(n^2td)$ , where  $t$  is the number of iteration in the process. For the SVM with RBF kernel, the computational cost is  $O(d^2)$ , where  $d$  is the number of input dimensions. Hence, the total computational cost of the proposed algorithm is the arithmetic sum of the computational costs of all stages, which is given as:



$$O(n^2) + O(nkd) + O(n^2dt) + O(d^2). \quad (16)$$

Although the proposed method shows a significant classification performance, its training phase requires the determination of an optimal weight value of the band in the fusion process, which is computationally expensive.

### 3 Results and Discussion

This section presents the experimental analysis of the proposed method using a standard benchmark hyperspectral datasets widely used in the literature.

#### 3.1 Dataset Description

A series of experiments were conducted on four standard bench-mark datasets, namely, Indian Pines, Pavia University, Salinas, and Botswana dataset, available in Ref. 47. Datasets such as Indian Pines, Pavia University, and Salinas are small-size datasets captured by airborne sensor, whereas Botswana dataset is a large-size hyperspectral dataset, which is captured by space borne or satellite sensors. The detailed description of each dataset is given below:

- a. *Indian Pines dataset.* It was acquired by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) over North-Western Indiana region in June 1992. This dataset consists of 16 different classes of agriculture as well as vegetation species, namely, “alfalfa,” “corn-notill,” “corn-mintill,” “corn,” “grass-pasture,” “grass-trees,” “grass-pasture-mowed,” “hay-windrowed,” “oats,” “soybean-notill,” “soybean-mintill,” “soybean-clean,” “wheat,” “woods,” “buildings-grass-trees-drives,” and “stone-steel-towers.” The size of the dataset is  $145 \times 145$  pixels with 20-m spatial resolution and 10-nm spectral resolution over the range of 400 to 2500 nm. It contains 224 spectral bands where only 200 bands remain for experimentation after the removal of 24 water absorption bands.
- b. *Pavia University dataset.* It was captured by Reflective Optical System Imaging Spectrometer over Pavia, Northern Italy, in July 2002. This dataset contains nine different classes such as “water,” “trees,” “asphalt,” “self-blocking bricks,” “bitumen,” “tiles,” “shadows,” “meadows,” and “bare soil.” The size of the dataset is  $610 \times 340$  pixels with 1.3-m spatial resolution over the range of 430 to 860 nm. It contains 103 spectral bands.
- c. *Salinas dataset.* It was captured by AVIRIS over Salinas Valley, California. This dataset contains 16 different classes, namely, “brocoli-green-weeds1,” “brocoli-green-weeds2,” “fallow,” “fallow-rough-plow,” “fallow-smooth,” “stubble,” “celery,” “grapes-untrained,” “soil-vinyard-develop,” “corn-senesced-green-weeds,” “lettuce-romaine-4wk,” “lettuce-romaine-5wk,” “lettuce-romaine-6wk,” “lettuce-romaine-7wk,” “vinyard-untrained,” and “vinyard-vertical-trellis.” The size of the dataset is  $512 \times 217$  pixels with 3.7-m spatial resolution over the range of 400 to 2500 nm. It contains 224 spectral bands.
- d. *Botswana dataset.* It was captured by NASA EO-1 satellite over the Okavango Delta, Botswana from 2001 to 2004. The hyperion sensor on EO-1 acquires data at 30-m pixel resolution over a 7.7-km strip in 242 bands covering the 400- to 2500-nm portion of the spectrum in 10-nm windows. Only 145 bands remain for experimentation after removal of noisy and water absorption bands. The size of dataset is  $1476 \times 256$  pixels with 30-m spatial resolution. The data contain 14 classes, namely, “water,” “hippo grass,” “floodplain grasses1,” “floodplain grasses2,” “reeds1,” “riparian,” “firescar2,” “island interior,” “Acacia woodlands,” “Acacia shrublands,” “Acacia grasslands,” “short mopane,” “mixed mopane,” and “exposed soils.”

#### 3.2 Evaluation Measures

The classification performance of the proposed EM-WAF technique is assessed using three commonly used quality metrics, i.e., overall accuracy (OA), average accuracy (AA), and  $k$ .

## a. OA

Percentage of the correctly classified pixels in the whole scene:

$$OA = \frac{\text{no. of correctly classified samples}}{\text{no. of test samples}}. \quad (17)$$

## b. AA

Mean of the percentage of the correctly labeled pixels for each class:

$$AA = \frac{1}{\text{no. of classes}(c)} \sum_{i=1}^c (OA)_i. \quad (18)$$

c. Kappa coefficient ( $k$ )

It is a robust measure of the degree of agreement, which integrates diagonal and off-diagonal entries of a confusion matrix.

### 3.3 Parameters Settings

For the EM clustering algorithm, the number of iteration  $k$  is set to 10. For an optimal weight finding procedure, the balance factor  $\alpha$  is set to 0.5 and the number of iterations  $t$  is set to 100. The SVM classifier with RBF kernels has two parameters: the penalty parameter  $C$  and the RBF parameter  $\gamma$  are tuned through fivefold cross validation ( $\gamma = 2^{-8}, 2^{-7}, \dots, 2^8$ ,  $C = 2^{-8}, 2^{-7}, \dots, 2^8$ ).

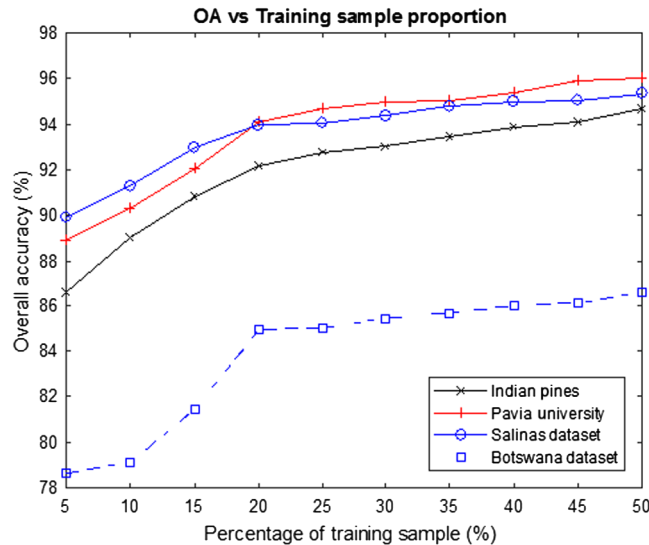
### 3.4 Experimental Results

In this section, the impact of different proportions of training samples on OA, the classification results obtained for Indian Pines, Pavia University, Salinas, and Botswana dataset, analysis of the features extracted by the proposed method, and remarkable findings are discussed. All the experiments are conducted using MATLAB 2018a on PC with 16 GB RAM and 2.70 GHz CPU. In the beginning, to evaluate the effectiveness of the proposed method with fewer amounts of labeled data, 20% of the samples for each class from the Indian Pines, Pavia University, Salinas, and Botswana dataset are randomly chosen as training samples, and the remaining samples in each class are used for testing purpose. Section 3.4.1 provides a detailed analysis of the different proportions of the training samples on OA. The experiment is conducted ten times to evaluate an average of OA, AA, and kappa coefficient. Four different categories of methods have been considered for comparison for verification of the superiority of the proposed method.

- In the first category, clustering-based feature extraction methods, namely, CBF<sup>30</sup> and DCCA<sup>33</sup> are considered.
- In the second category, CBS methods<sup>9</sup> considered are, CEM-BCC/BDC, CEM-BCM/BDM, LCMV-BCC/BDC, and LCMV-BCM/BDM.
- In the third category, clustering- and ranking-based band selection method considered is E-FDPC.<sup>38</sup>
- In the fourth category, a comparison of the proposed method is made with clustering and band fusion method for demonstrating the significance of the weights of the bands,<sup>29</sup> where a simple average fusion method is used for fusing the bands from a cluster.

#### 3.4.1 Influence of different proportion of training samples on OA obtained by the proposed method for all four hyperspectral datasets

The performance of the proposed method is validated against different proportions of training samples, namely, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50% of the labeled

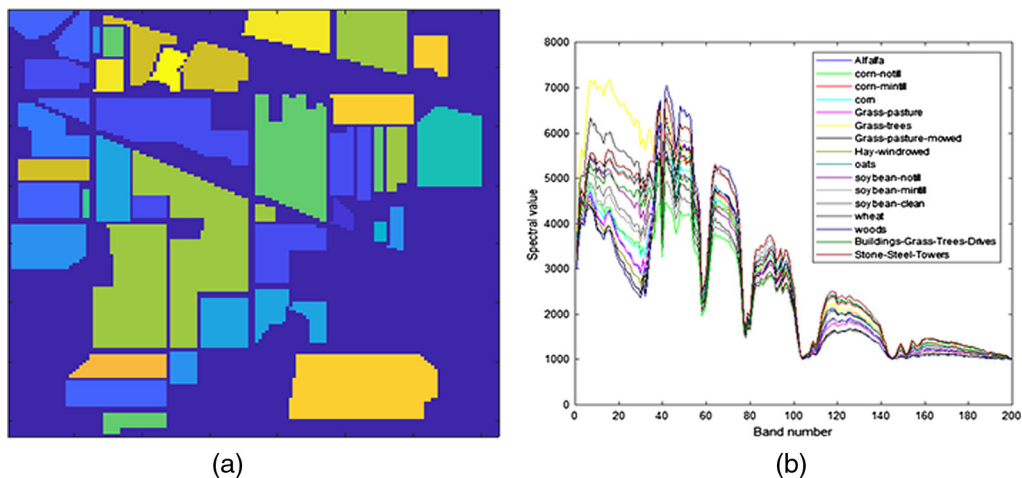


**Fig. 3** Influence of different proportions of training samples on OA for Indian Pines, Pavia University, Salinas, and Botswana dataset.

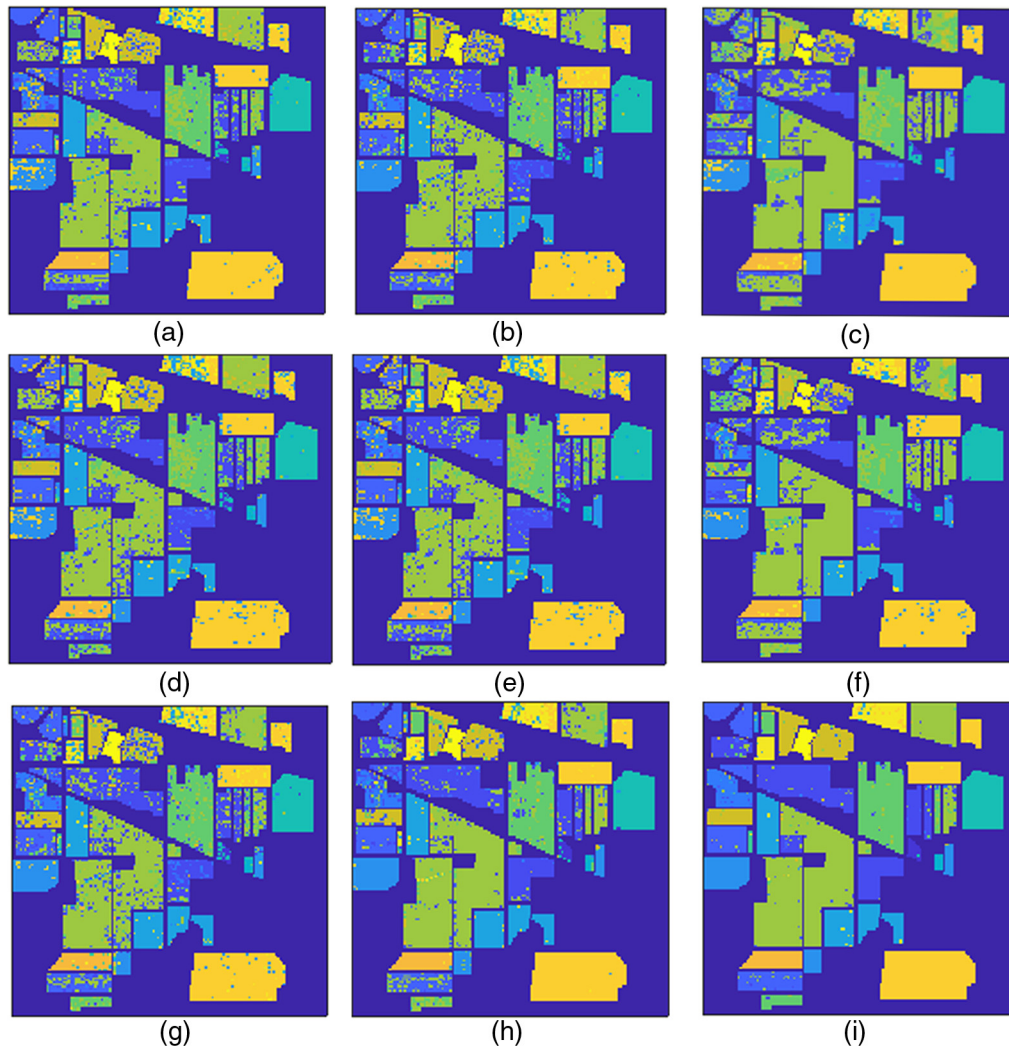
training samples per class. Figure 3 shows OA obtained using the proposed method for different proportions of the training samples. The proposed method has seen a good discriminative ability to deal even with a smaller size of the labeled samples, 5% of training sample size per class. With increase in the number of training sample, the classification performance of the proposed method increases gradually for all four datasets. The sample size of more than 20% does not have much impact on the OA. However, the increase in sample size increases the computational burden in the training phase. Hence, the proposed method is tested with 20% of the training samples.

### 3.4.2 Results analysis by comparing the proposed method with different classification methods on Indian Pines dataset

The ground truth data of Indian Pines dataset are shown in Fig. 4(a), where the different colors signify the various land cover categories. Figure 4(b) shows the spectral signature or the reflectance of each category. The classification maps obtained for all the competing methods on Indian Pines dataset as shown in Fig. 5 and the classification results (i.e., OA, class wise accuracy, AA, and  $k$ ) are reported in Table 1. Figure 5 and Table 1 show that the proposed method achieves the



**Fig. 4** Indian Pines dataset information: (a) ground truth data and (b) spectral response of each category



**Fig. 5** Classification map of Indian Pine dataset for all competing methods: (a) CBFE, (b) DCCA, (c) CEM-BCC/BDC, (d) CEM-BCM/BDM, (e) LCMV-BCC/BDC, (f) LCMV-BCM/BDM, (g) E-FDPC, (h) IF, and (i) EM-WAF.

**Table 1** Comparison of classification accuracies (%) obtained by the proposed method with other competing methods for Indian Pine dataset.

Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection method	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Alfalfa	85.13	84.13	69.53	54.63	69.3	67.23	57.53	89.43	<b>94.01</b>
Corn-no till	72.09	71.09	56.49	41.59	56.26	54.19	44.49	76.39	<b>81.09</b>
Corn-min till	59.04	58.04	62.64	47.74	62.41	60.34	50.64	63.34	<b>87.24</b>
Corn	59.71	58.71	66.55	51.65	66.32	64.25	54.55	64.01	<b>91.15</b>

**Table 1** (Continued).

Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection method	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Grass-pasture	99.01	98.01	76.1	61.2	75.87	73.8	64.1	<b>100</b>	99.87
Grass-tree	93.03	92.03	76.1	61.2	75.87	73.8	64.1	97.33	<b>99.1</b>
Grass-pasture-mowed	64.93	63.93	59.55	44.65	59.32	57.25	47.55	69.23	<b>84.15</b>
Hay-windrowed	90.08	89.08	74.48	59.58	74.25	72.18	62.48	94.38	<b>99.08</b>
Oat	68.89	67.89	58.44	43.54	58.21	56.14	46.44	73.19	<b>83.04</b>
Soybean-no till	59.93	58.93	65	50.1	64.77	62.7	53	64.23	<b>89.6</b>
Soybean-min till	88.9	87.9	73.3	58.4	73.07	71	61.3	93.2	<b>97.9</b>
Soybean-clean	57.93	56.93	63.77	48.87	63.54	61.47	51.77	62.23	<b>88.37</b>
Wheat	94.02	93.02	76.1	61.2	75.87	73.8	64.1	98.32	<b>100</b>
Woods	88.03	87.03	72.43	57.53	72.2	70.13	60.43	92.33	<b>97.03</b>
Buildings-grass-trees-drives	61.68	60.68	55.92	41.02	55.69	53.62	43.92	65.98	<b>80.52</b>
Stone-steel-towers	99.03	98.03	76.1	61.2	75.87	73.8	64.1	<b>100</b>	99.25
OA	79.88	78.67	69.94	53.56	69.33	67.94	57.68	83.56	<b>92.19</b>
AA	77.97	76.59	67.65	52.75	67.43	65.36	55.65	81.9	<b>91.96</b>
K	0.7751	0.7639	0.6602	0.5276	0.6701	0.6601	0.5701	0.817	<b>0.9085</b>

Note: Highest value across the method is represented in bold font.

better result when compared with the competing methods in terms of OA, AA, and  $k$ . It is due to the use of EM clustering algorithm for band partitioning and weighted average fusion for fusing the correlated band that leads to increase the interclass separation and decrease the intraclass separation.

Table 1 shows EM-WAF method achieving a good performance compared to clustering-based methods, namely, CBFE, DCCA, CEM-BCC/BDC, CEM-BCM/BDM, LCMV-BCC/BDC, LCMV-BCM/BDM, and E-FDPC. The proposed technique shows a noticeable performance due to the presence of larger discriminative information by the clustering and fusing of highly correlated bands. The classification accuracy of the proposed EM-WAF method is much better than that of the simple IF method and highlights the importance of weight factor in the fusion process. Clustering-based methods and IF method only consider the intracluster distance, which limits the discriminative ability, whereas the proposed method considers the intercluster distance as well as intracluster distance, which leads to a better discriminative ability. Hence, the proposed EM-WAF technique preserves the useful as well as the discriminative information of the original data. When compared to the other competing approaches, the proposed EM-WAF approach achieves a substantial improvement in terms of the class wise classification accuracy as shown in Table 1 (boldface). It is evident that the classification accuracy of

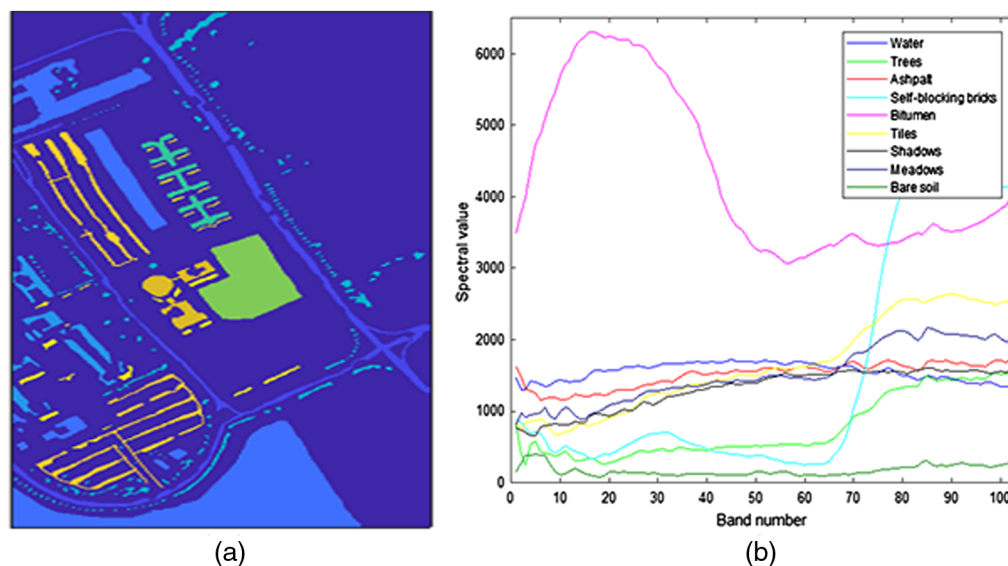


the classes “alfalfa,” “corn-no till,” “corn-min till,” “corn,” “grass-pasture-mowed,” “hay-wind-rowed,” “oat,” “soybean-no till,” “soybean-min till,” “soybean-clean,” and “woods” increases from 54.63% to 94.01%, 41.59% to 81.09%, 47.54% to 87.24%, 51.65% to 91.15%, 44.65% to 84.15%, 59.58% to 99.08%, 43.54% to 83.04%, 53% to 89.06%, 58.4% to 97.9%, 48.87% to 88.37%, and 57.53% to 97.03%, respectively. In particular, in the class “wheat” all the pixels are correctly classified through the use of the proposed method. However, it is observed that the proposed method achieves slightly lesser accuracy for the individual classes such as “grass-pasture” and “stone-steel-towers” when compared to the IF method (achieves 100% accuracy for both classes) as shown in Table 1.

### 3.4.3 Results analysis by comparing the proposed method with different classification methods on Pavia University dataset

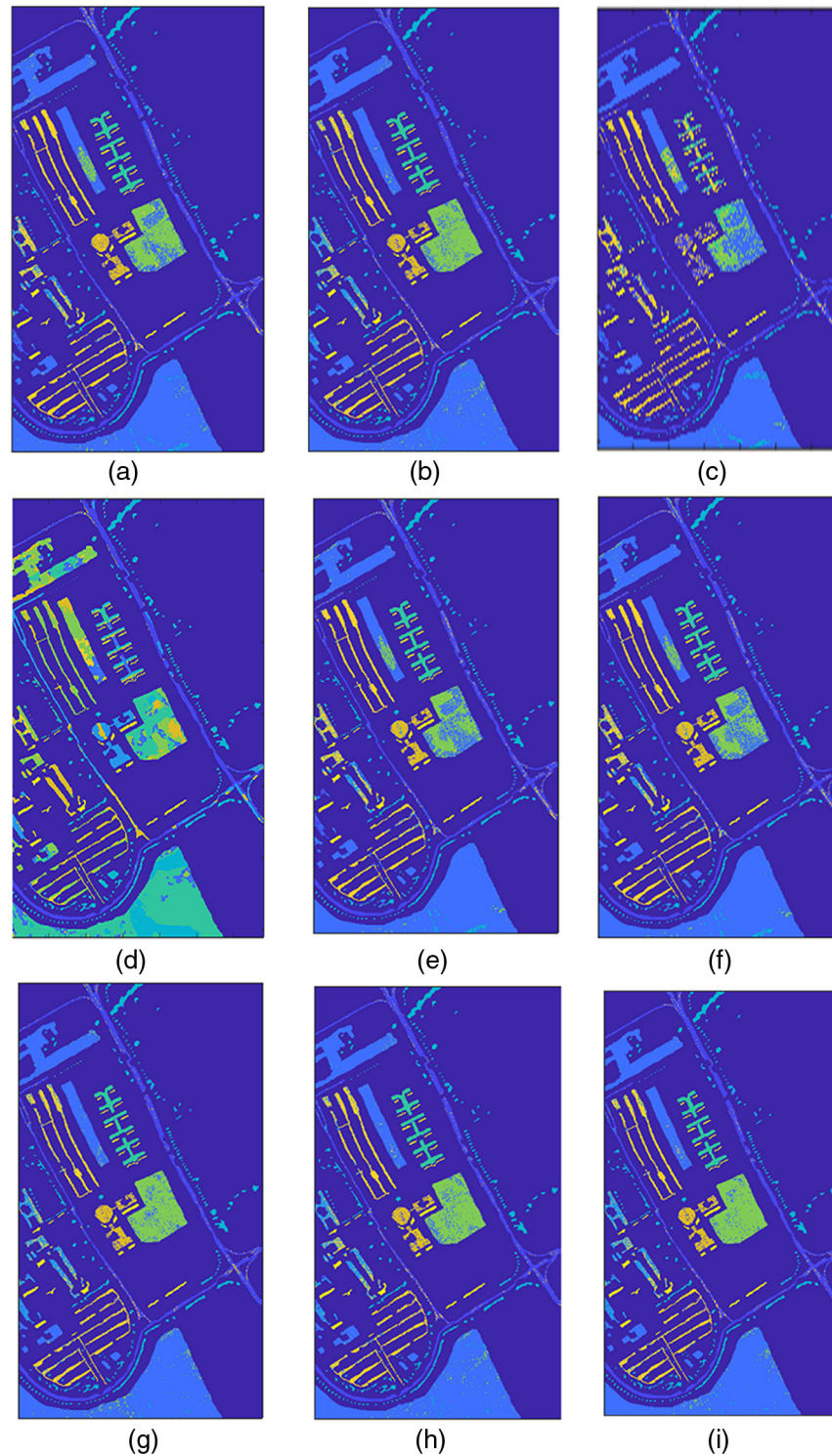
The ground truth data of Pavia University dataset are shown in Fig. 6(a), where the different colors denote the different categories. Figure 6(b) shows the spectral signature or the reflectance of each category. The classification maps obtained for all the competing techniques along with the proposed technique on Pavia University dataset are depicted in Fig. 7 and the classification results (i.e., OA, class wise accuracy, AA, and  $k$ ) are presented in Table 2. Figure 7 and Table 2 show that the proposed EM-WAF technique achieving the best result among all the competing methods in terms of OA, AA, and  $k$ . It is due to the fact of EM clustering extracts more useful information and increases the separation among the spectral classes. As shown in Table 2, the classification accuracy of the proposed EM-WAF method is much better than the IF method showing the importance of the weight factor in the fusion process. In other words, the proposed method preserves the complementary information of all bands well.

As shown in Fig. 7, the proposed approach helps in the elimination of most of the noisy pixels generated by the other methods, and the overall classification accuracy increases by more than 2%. For instance, the misclassified pixels are corrected in the green region at the center of Fig. 7, which is very close to the ground truth and the classification map becomes smoother. When compared to the other competing approaches, the proposed approach shows a significant improvement in the class wise classification accuracy as shown in Table 2 (boldface). For instance, the classification accuracy of class “Gravel” increases from 7.98% to 85.21%. Moreover, the proposed method correctly classified the class “painted metal sheets.” However, EM-WAF



**Fig. 6** Pavia University dataset information: (a) ground truth data and (b) spectral response of each category.





**Fig. 7** Classification map of Pavia University dataset for all competing methods: (a) CBFE, (b) DCCA, (c) CEM-BCC/BDC, (d) CEM-BCM/BDM, (e) LCMV-BCC/BDC, (f) LCMV-BCM/BDM, (g) E-FDPC, (h) IF, and (i) EM-WAF.

approach is seen producing lesser classification accuracy for individual class, namely, “self-blocking bricks” when compared to LCMV-BCM/BDM method as shown in Table 2. The reason is that fusion of the spectral bands eliminates the important spectral features of the respective land cover class.

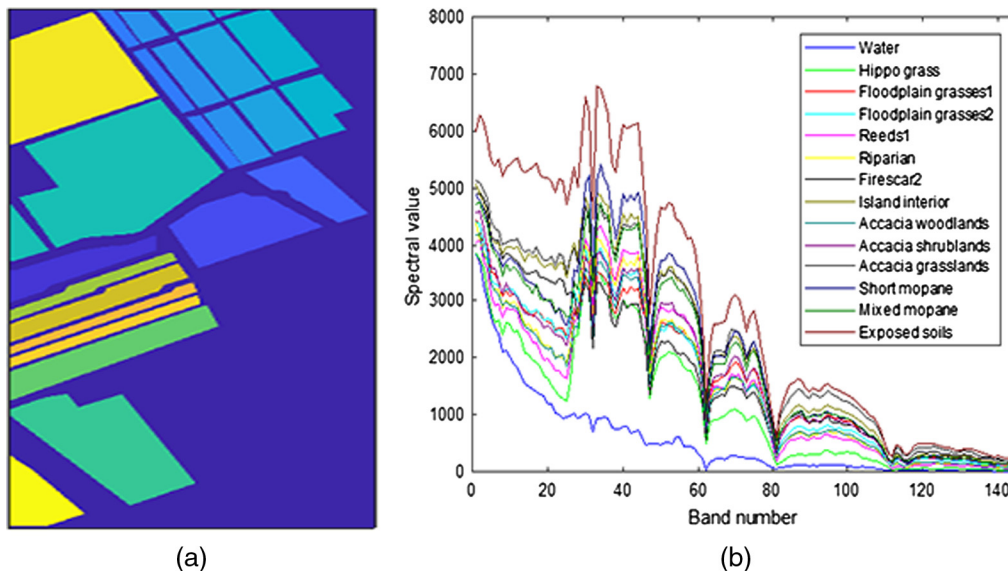
**Table 2** Comparison of classification accuracies (%) obtained by the proposed method with other competing methods for Pavia University dataset.

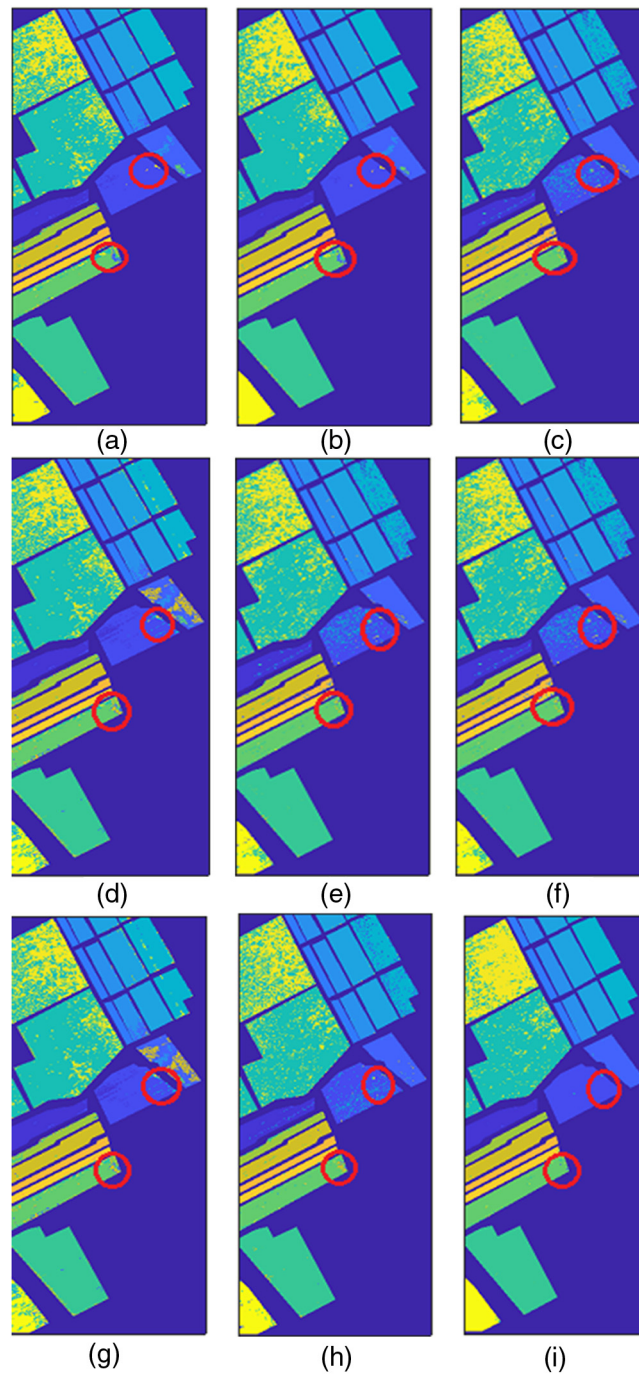
Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection method	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Asphalt	89.88	93.53	76.15	72.69	88.88	90.98	91.46	92.48	<b>95.87</b>
Meadows	94.47	95.61	82.55	79.09	93.47	95.57	96.62	97.04	<b>99.85</b>
Gravel	31.15	72.30	11.44	7.98	45.67	32.25	72.60	65.46	<b>85.21</b>
Trees	82.54	89.02	69	65.54	81.54	83.64	90.33	89.11	<b>91.36</b>
Painted metal sheets	98.70	98.70	86.05	82.59	97.7	99.8	98.88	98.51	<b>100</b>
Bare soil	62.59	89.81	34.72	31.26	61.59	63.69	83.62	81.18	<b>92.15</b>
Bitumen	78.95	82.89	70.05	66.59	77.95	80.05	79.14	78.01	<b>85.23</b>
Self-blocking bricks	87.71	84.04	75.18	71.72	86.71	<b>88.81</b>	83.33	83.02	86.38
Shadows	100	99.87	87.31	83.85	90.02	98.43	99.60	100	<b>100</b>
OA	85.50	89.92	67.23	63.21	84.52	84.52	91.11	90.67	<b>94.10</b>
AA	80.66	89.75	65.82	62.36	80.39	81.76	88.40	87.20	<b>92.89</b>
K	0.8012	0.8831	0.6690	0.6287	0.8123	0.8102	0.8816	87.53	<b>91.12</b>

Note: Highest value across the method is represented in bold font.

### 3.4.4 Results analysis by comparing the proposed method with different classification methods on Salinas dataset

The ground truth data of the Salinas dataset are shown in Fig. 8(a), where the different colors represent the different categories. Figure 8(b) shows the spectral signature or the reflectance of

**Fig. 8** Salinas dataset information: (a) ground truth data and (b) spectral response of each category.



**Fig. 9** Classification map of Salinas dataset for all competing methods: (a) CBFE, (b) DCCA, (c) CEM-BCC/BDC, (d) CEM-BCM/BDM, (e) LCMV-BCC/BDC, (f) LCMV-BCM/BDM, (g) E-FDPC, (h) IF, and (i) EM-WAF.

each category. The classification maps of all the competing techniques on Salinas dataset are shown in Fig. 9 and the classification results (i.e., OA, class wise accuracy, AA, and  $k$ ) are reported in Table 3. Table 3 and Fig. 9 show that the proposed method achieves the best performance in terms of the quantitative results and visual interpretation.

Though all the competing methods are quite useful for dimensionality reduction, CBFE and DCCA methods attain noticeable performance over E-FDPC and other CBS methods. However, the proposed method shows the significant performance over all the

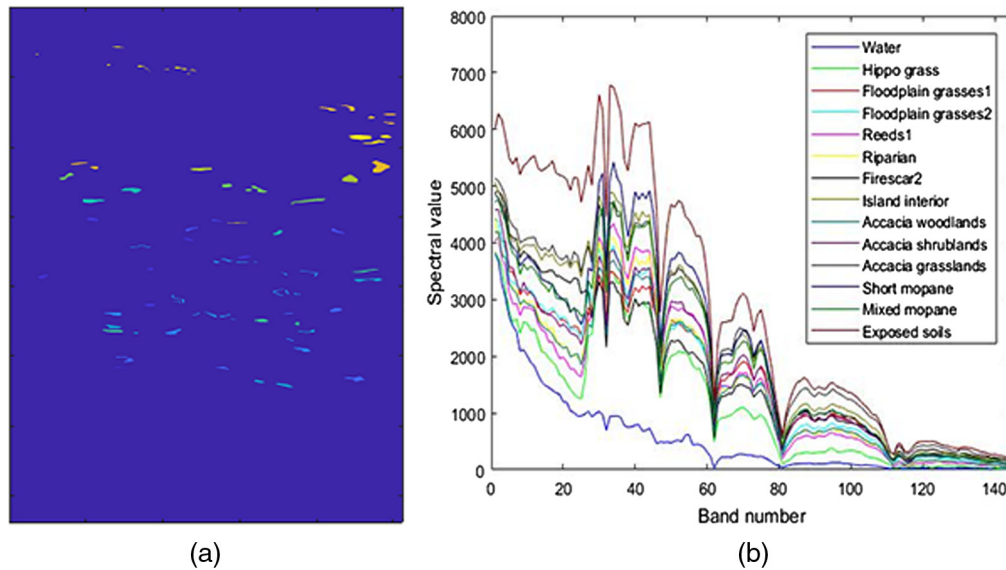
**Table 3** Comparison of classification accuracies (%) obtained by the proposed method with other competing methods for Salinas dataset.

Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection method	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Broccoli-green-weeds1	96.33	88.03	85.03	79.41	84.41	83.41	97.76	94.83	<b>97.83</b>
Broccoli-green-weeds2	98.15	74.99	71.99	66.37	71.37	70.37	88.22	85.63	<b>98.91</b>
Fallow	85.89	81.14	78.14	72.52	77.52	76.52	52.41	94.54	<b>97.94</b>
Fallow-rough-plow	98.39	85.05	82.05	76.43	81.43	80.43	<b>99.55</b>	97.13	99.16
Fallow-smooth	93.46	94.6	91.6	85.98	90.98	89.98	90.01	98.50	95.26
Stubble	99.02	94.6	91.6	85.98	90.98	89.98	97.82	98.64	<b>99.34</b>
Celery	98.81	78.05	75.05	69.43	74.43	73.43	96.23	86.65	<b>99.57</b>
Grapes-untrained	83.88	<b>92.98</b>	89.98	84.36	89.36	88.36	84.70	85.62	88.54
Soil-vinyard-develop	96.45	76.94	73.94	68.32	73.32	72.32	95.57	<b>98.51</b>	97.48
Corn-senesced-green-weeds	80.40	83.5	80.5	74.88	79.88	78.88	80.05	89.33	<b>90.46</b>
Lettuce-romaine-4 wk	80.80	<b>91.8</b>	88.8	83.18	88.18	87.18	78.57	89.57	87.7
Lettuce-romaine-5 wk	99.09	82.27	79.27	73.65	78.65	77.65	99.22	97.92	<b>99.56</b>
Lettuce-romaine-6 wk	98.36	94.6	91.6	85.98	90.98	89.98	99.04	96.85	<b>97.28</b>
Lettuce-romaine-7 wk	88.90	90.93	87.93	82.31	87.31	86.31	87.27	91.23	<b>92.64</b>
Vinyard-untrained	44.55	74.42	<b>71.42</b>	65.80	70.8	69.8	40.45	44.15	52.2
Vinyard-vertical-trellis	84.71	94.6	91.6	85.98	90.98	89.98	61.52	94.53	<b>98.36</b>
OA	85.14	89.94	84.01	78.86	83.18	83.14	81.55	85.85	<b>93.96</b>
AA	89.20	86.15	83.15	77.54	82.53	81.54	84.28	90.68	<b>92.45</b>
<i>K</i>	0.8339	0.8789	0.8289	0.7689	0.8145	0.8237	0.7937	0.8418	<b>0.9036</b>

Note: Highest value across the method is represented in bold font.

other competing methods. It is due to the fact that the clustering and weighted average fusion of the highly correlated bands provide more discriminative information. It shows the proposed EM-WAF technique extracting the significant features of the data. Consequently, the superiority of the EM-WAF approach can be explained by the use of weighted average of useful bands. When compared to the other competing methods, the performance of the proposed method is superior in terms of OA, AA, and *k*. In most of the classes, the class wise accuracy of the proposed method exceeds 90%. However, the proposed method fails to obtain a good performance for a few classes. For instance, the pixels of class “grapes-untrained” are misclassified with the pixels of “vinyard-untrained” class. This misclassification occurs as the spectral signatures of these two classes are almost the same. Figure 9 shows that the region uniformity of the classes “fallow” and “corn-senesced-green-weeds” (marked by red circles) as improved by the proposed method when compared to the other competing methods.



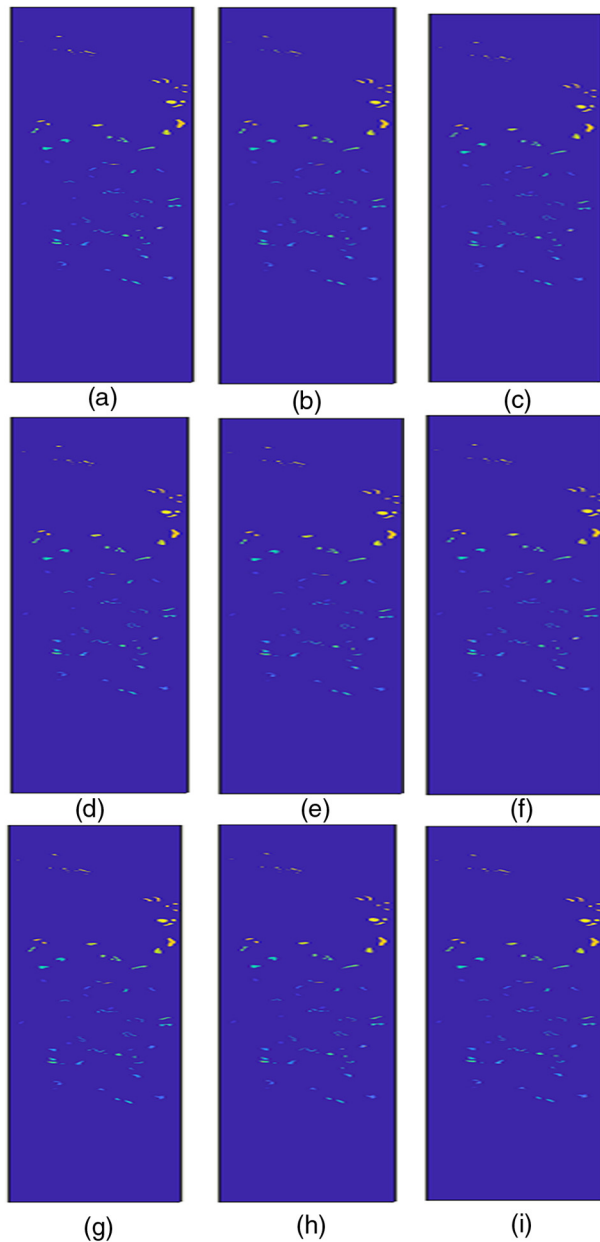


**Fig. 10** Botswana dataset information: (a) ground truth data and (b) spectral response of each category.

### 3.4.5 Results analysis by comparing the proposed method with different classification methods on Botswana Dataset

The ground truth information relating to Botswana dataset used for experimentation is shown in Fig. 10(a), where the different colors signify the different land cover categories. Figure 10(b) shows the spectral signature or the reflectance of each category. The classification maps of all the competing techniques on Botswana dataset are shown in Fig. 11 and the classification results (i.e., OA, class wise accuracy, AA, and  $k$ ) are summarized in Table 4.

The results reported in Table 4 lead to the observation of the proposed EM-WAF method delivering a better performance than the other competing methods. Table 4 shows the classification results obtained by the proposed clustering and fusion-based method are very promising, which indicates the possibility of classification of the large-size dataset using the proposed method. Table 4 shows the E-FDPC method obtains significant performance superior that of other clustering and constrained-based selection methods, this is mainly due to the band selection strategy of the ranking-based methods. However, the proposed method is better than the E-FDPC method, since the latter technique only considers the intracluster distance between the data points, whereas the former technique considers intercluster as well as intracluster distance between the data points, resulting in good discriminative capabilities for the classification. Table 4 shows that the proposed method achieves better class wise accuracies for most of the classes. It is observed that the proposed method classifies all pixels of the class “water” correctly. Compared to the other competing methods, classes such as “hippo grass,” “Acacia woodlands,” “short mopane,” and “mixed mopane” are better distinguished by the proposed method. The performance of the proposed method is better than that of the other competing methods for the classes, “reeds1” and “riparian,” though it is not satisfactory. The main reason is that the samples selected from such classes consist of more redundant information.



**Fig. 11** Classification map of Botswana dataset for all competing methods: (a) CBFE, (b) DCCA, (c) CEM-BCC/BDC, (d) CEM-BCM/ BDM, (e) LCMV-BCC/BDC, (f) LCMV-BCM/ BDM, (g) E-FDPC, (h) IF, and (i) EM-WAF.

**Table 4** Comparison of classification accuracies (%) obtained by the proposed method with other competing methods for Botswana dataset.

Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection methods	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Water	96.53	98.14	97.68	99.53	96.04	99.53	99.50	99.00	<b>100</b>



**Table 4 (Continued).**

Class name	Clustering-based methods		Constrained-based selection methods				Clustering and ranking-based selection methods	Clustering and fusion-based methods	
	CBFE <sup>30</sup>	DCCA <sup>33</sup>	CEM-BCC/BDC <sup>9</sup>	CEM-BCM/BDM <sup>9</sup>	LCMV-BCC/BDC <sup>9</sup>	LCMV-BCM/BDM <sup>9</sup>	E-FDPC <sup>38</sup>	IF <sup>29</sup>	EM-WAF (proposed method)
Hippo grass	81.25	85.00	78.75	86.25	83.75	78.75	82.66	77.33	<b>90.66</b>
Floodplain grasses1	77.50	86.00	82.00	92.00	89.50	85.00	91.45	90.95	<b>91.48</b>
Floodplain grasses2	80.81	<b>87.20</b>	61.62	75.00	74.41	72.67	84.47	78.88	78.26
Reeds1	58.60	60.46	39.06	65.11	64.65	59.53	61.69	60.17	<b>65.67</b>
Riparian	61.60	46.51	53.02	50.69	42.81	50.69	45.87	47.78	<b>61.69</b>
Firescar2	94.20	<b>98.55</b>	96.13	97.10	93.10	98.55	96.90	97.42	96.90
Island interior	90.12	88.08	86.41	86.41	86.74	90.74	<b>93.42</b>	80.94	88.15
Acacia woodlands	65.33	68.22	60.55	60.15	57.37	57.37	64.25	68.31	<b>73.19</b>
Acacia shrublands	58.08	83.85	62.62	61.11	60.06	58.58	59.67	59.91	<b>84.40</b>
Acacia grasslands	88.93	89.11	93.03	92.62	86.88	93.44	87.71	90.01	<b>94.29</b>
Short mopane	45.13	87.50	62.50	53.47	67.47	50.69	59.25	67.44	<b>94.81</b>
Mixed mopane	75.23	89.25	70.56	61.21	76.63	73.36	60.19	78.10	<b>91.04</b>
Exposed soils	82.89	88.15	85.59	80.26	82.05	82.89	<b>90.14</b>	87.32	78.87
OA	75.25	81.73	72.86	75.33	74.60	74.87	83.01	77.53	<b>84.92</b>
AA	75.25	82.70	73.54	75.78	75.89	75.13	84.9	77.37	<b>84.96</b>
K	0.7319	0.8020	0.7058	0.7326	0.7249	0.7276	0.8253	0.7563	<b>0.8336</b>

Note: Highest value across the method is represented in bold font.

### 3.4.6 Analysis of number of selected bands or features for all four hyperspectral datasets

Table 5 shows the number of selected bands or features and OA for four hyperspectral datasets. Table 5 shows the ability of the proposed approach to achieve a better classification accuracy through selection of features of an optimal number. In other words, the proposed approach selects the features that separate the land cover classes well.

As shown in Table 5, the features extracted by the proposed method for all datasets achieve the highest classification accuracy. For the Indian Pines dataset, the proposed method provides a maximum OA of 92.19% among all the competing methods for only seven features, which is found to be optimal. For the Pavia University dataset, the proposed method delivers the highest OA of 94.10% among all the competing methods for only 11 optimal features. For the Salinas dataset, CBFE method provides 85.14% OA for only 12 features, which are the minimum number of features extracted by CBFE among all other competing methods. However, the proposed method achieves a maximum OA of 93.96% among all the competing methods for an optimal number of 13 averaged bands. For Botswana dataset, the proposed method provides OA, which is slightly better than E-FDPC method. However, the proposed method achieves maximum OA (84.92%) among all the competing methods for only 20 features, which is found to be optimal one. Table 5 shows that the proposed approach extracts meaningful features from the hyperspectral data. These features are suitable and adequate for the hyperspectral image classification. These results indicate that: (a) the pairwise distance-based band separability is an important aspect for feature extraction; (b) consideration of intracluster and intercluster distance provides

**Table 5** Number of selected bands or features and OA (%) for all four hyperspectral datasets.

Dataset		Method								
		DCCA	CEM- BCC/BDC	CEM- BCM/BDM	LCMV- BCC/BDC	LCMV- BCM/BDM	CBFE	E- FDPC	IF	EM- WAF
Indian Pines	Number of bands or features	20	20	15	20	20	13	10	25	7
	OA (%)	78.67	69.94	53.56	69.33	67.94	79.88	57.68	83.56	92.19
Pavia University	Number of bands or features	20	20	20	20	20	15	14	25	11
	OA (%)	89.92	67.23	63.21	84.52	84.52	85.50	91.11	90.67	94.10
Salinas	Number of bands or features	15	15	20	20	20	12	14	25	13
	OA (%)	89.94	84.01	78.86	83.18	83.14	85.14	81.55	85.85	93.96
Botswana	Number of bands or features	30	30	30	30	30	30	30	25	20
	OA (%)	75.25	81.73	72.86	75.33	74.60	74.87	83.01	77.53	84.92

more discriminative information; and (c) an appropriate weighting mechanism for the weighted average fusion improves the performance of feature extraction significantly.

## 4 Conclusion

In this paper, EM clustering and weighted average fusion technique-based feature extraction for hyperspectral image classification has proposed. The proposed method explores the information among the clusters and removes redundancy among the bands. The EM algorithm converges to the best number of clusters, thereby providing an effective way to determine an optimal number of features. The weight factor of the bands is calculated on the basis of the criteria of minimizing the distance inside each cluster and maximizing the distance among the different clusters, which highlights the importance of the particular band in the fusion process. The significance of this technique lies in its highly discriminative ability, which leads to a better classification performance. Experimental results and comparison with the existing approaches prove the efficiency of the proposed method for hyperspectral image classification. When compared with the other competing methods on four standard datasets, the proposed method achieves higher classification accuracy and better visual results. For the Botswana dataset, the proposed method provides better OA among all other competing methods, which makes it evident that the proposed method can classify a large-size dataset effectively. Moreover, the proposed method performs equally well for all four hyperspectral datasets, showing the robustness of the proposed method in both small- and large-size datasets.

In our future work, we will focus on integrating the spatial features with the spectral features to improve the classification performance.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their comments and valuable suggestions, which greatly helped us to improve the technical quality and presentation of the manuscript. The authors thank VIT for providing a VIT seed grant for carrying out this research work

and the Council of Scientific & Industrial Research (CSIR), New Delhi, India for the award of CSIR-SRF.

## References

1. H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.* **39**(4), 1232–1249 (2003).
2. M. Khodadadzadeh et al., "A new framework for hyperspectral image classification using multiple spectral and spatial features," in *IEEE Geoscience and Remote Sensing Symp.*, Quebec City, Canada, pp. 4628–4631 (2014).
3. S. S. Sawant and M. Prabukumar, "Semi-supervised techniques based hyper-spectral image classification: a survey," in *Innovations in Power and Advanced Computing Technologies (i-PACT)* (2017).
4. J. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin (1999).
5. G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory* **14**(1), 55–63 (1968).
6. C. Burges, "Dimension reduction: a guided tour," *Found. Trends Mach. Learn.* **2**(4), 275–364 (2010).
7. R. Vaddi and M. Prabukumar, "Comparative study of feature extraction techniques for hyper spectral remote sensing image classification: a survey," in *Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*, pp. 543–548 (2017).
8. C. I. Chang, "A joint band prioritization and band decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **37**(6), 2631–2641 (1999).
9. C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1575–1585 (2006).
10. X. Bai et al., "Semisupervised hyperspectral band selection via spectral–spatial hypergraph model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(6), 2774–2783 (2015).
11. X. Cao et al., "Fast hyperspectral band selection based on spatial feature extraction," *J. Real-Time Image Process.* 1–10 (2018).
12. C. Yu, M. Song, and C. Chang, "Band subset selection for hyperspectral image classification," *Remote Sens.* **10**, 113 (2018).
13. Q. Chen, "Band selection algorithm based on information entropy for hyperspectral image classification," *J. Appl. Remote Sens.* **11**(2), 026018 (2017).
14. W. Zhang, X. Li, and L. Zhao, "Hyperspectral band selection based on triangular factorization," *J. Appl. Remote Sens.* **11**(2), 025007 (2017).
15. S. Samiappan, S. Prasad, and L. Bruce, "Non-uniform random feature selection and kernel density scoring with SVM based ensemble classification for hyperspectral image analysis," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**(2), 792–800 (2013).
16. J. C. Davis, "Introduction to statistical pattern recognition: 2nd edition, by Keinosuke Fukunaga, Academic Press, San Diego, 1990, 591 p., ISBN 0-12-269851-7, US\$69.95," *Comput. Geosci.* **22**(7), 833–834 (1990).
17. F. Tsai, E.-K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *Int. J. Remote Sens.* **28**(5), 1023–1039 (2007).
18. X. Junshi et al., "(Semi-) supervised probabilistic principal component analysis for hyperspectral remote sensing image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(6), 2224–2236 (2014).
19. A. Villa et al., "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.* **49**(12), 4865–4876 (2011).
20. X. Liu et al., "A maximum noise fraction transform with improved noise estimation for hyperspectral images," *Sci. China Ser. F* **52**(9), 1578–1587 (2009).
21. M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.* **2009**, 783194 (2009).

22. M. E. Tipping, “Probabilistic principal component analysis,” *J. R. Stat. Soc. Ser. B* **61**(3), 611–622 (1999).
23. T. V. Bandos, L. Bruzzone, and G. Camps-Valls, “Classification of hyperspectral images with regularized linear discriminant analysis,” *IEEE Trans. Geosci. Remote Sens.* **47**(3), 862–873 (2009).
24. B. C. Kuo and D. A. Landgrebe, “Nonparametric weighted feature extraction for classification,” *IEEE Trans. Geosci. Remote Sens.* **42**(5), 1096–1105 (2004).
25. B. C. Kuo, C. H. Li, and J. M. Yang, “Kernel nonparametric weighted feature extraction for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.* **47**(4), 1139–1155 (2009).
26. M. Prabukumar et al., “Threedimensional discrete cosine transform-based feature extraction for hyperspectral image classification,” *J. Appl. Remote Sens.* **12**(4), 046010 (2018).
27. I. Makki et al., “A survey of landmine detection using hyperspectral imaging,” *ISPRS J. Photogramm. Remote Sens.* **124**, 40–53 (2017).
28. A. R. Webb, *Statistical Pattern Recognition*, Vol. **71**, no. 8, John Wiley & Sons, Ltd., England (2011).
29. S. Sawant and M. Prabukumar, “Band fusion based hyper spectral image classification,” *Int. J. Pure Appl. Math.* **117**(17), 71–76 (2017).
30. M. Imani and H. Ghassemian, “Band clustering-based feature extraction for classification of hyperspectral images using limited training samples,” *IEEE Geosci. Remote Sens. Lett.* **11**(8), 1325–1329 (2014).
31. Q. Yan et al., “Class probability propagation of supervised information based on sparse subspace clustering for hyperspectral images,” *Remote Sens.* **9**, 1017 (2017).
32. X. Peng et al., “Constructing the L2-graph for subspace learning and subspace clustering,” *IEEE Trans. Cybern.* **6**, 1–14 (2016).
33. Y. Yuan, J. Lin, and Q. Wang, “Dual-clustering-based hyperspectral band selection by contextual analysis,” *IEEE Trans. Geosci. Remote Sens.* **54**(3), 1431–1445 (2016).
34. M. Khoder et al., “Multicriteria classification method for dimensionality reduction adapted to hyperspectral images,” *J. Appl. Remote Sens.* **11**(2), 025001 (2017).
35. X. Sun et al., “Hyperspectral image clustering method based on artificial bee colony algorithm,” in *Sixth Int. Conf. on Advanced Computational Intelligence (ICACI)*, pp. 106–109 (2013).
36. H. Su and P. Du, “Multiple classifier ensembles with band clustering for hyperspectral image classification,” *Eur. J. Remote Sens.* **47**(1), 217–227 (2014).
37. R. Liu, H. Wang, and X. Yu, “Shared-nearest-neighbor-based clustering by fast search and find of density peaks,” *Inf. Sci.* **450**, 200–226 (2018).
38. S. Jia et al., “A novel ranking-based clustering approach for hyperspectral band selection,” *IEEE Trans. Geosci. Remote Sens.* **54**(1), 88–102 (2016).
39. H. Xie et al., “Unsupervised hyperspectral remote sensing image clustering based on adaptive density,” *IEEE Geosci. Remote Sens. Lett.* **15**(4), 632–636 (2018).
40. B. Peng et al., “Weighted-fusion-based representation classifiers for hyperspectral imagery,” *Remote Sens.* **7**, 14806–14826 (2015).
41. S. Prasad and L. M. Bruce, “Decision fusion with confidence-based weight assignment for hyperspectral target recognition,” *IEEE Trans. Geosci. Remote Sens.* **46**(5), 1448–1456 (2008).
42. T. Lu et al., “From subpixel to superpixel: a novel fusion framework for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.* **55**(8), 4398–4411 (2017).
43. B. Kumar and O. Dikshit, “Hyperspectral image classification based on morphological profiles and decision fusion,” *Int. J. Remote Sens.* **38**(20), 5830–5854 (2017).
44. A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. R. Stat. Soc. Ser. B* **39**(1), 1–38 (1977).
45. R. Yang et al., “Representative band selection for hyperspectral image classification,” *J. Vision Commun. Image Represent.* **48**, 396–403 (2017).
46. F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing,” *IEEE Trans. Geosci. Remote Sens.* **42**(8), 1778–1790 (2004).

47. “Hyperspectral remote sensing scenes,” Computational Intelligence Group, 2007, [http://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

**Manoharan Prabukumar** received his BE degree in electronics and communication engineering from Periyar University, Tamilnadu, India, in 2002, his MTech degree in computer vision and image processing from Amrita School of Engineering, Coimbatore, India, in 2007, and his PhD in computer graphics from Vellore Institute of Technology (VIT), Tamilnadu, India, in 2014. Currently, he is working as an associate professor in the School of Information Technology and Engineering, VIT. His research interests include hyperspectral remote sensing, image processing, computer graphics, and machine learning.

**Sawant Shrutika** received her BE and ME degrees in electronics and telecommunication engineering from Shivaji University, Maharashtra, India, in 2009 and 2012, respectively. Currently, she is pursuing her PhD in hyperspectral image processing from VIT, Vellore, Tamilnadu, India. She has been awarded with the senior research fellowship from the Council of Scientific and Industrial Research, New Delhi, India. Her research interests include hyperspectral remote sensing, image processing, and machine learning.