

## Review Article

# Overview of Hyperspectral Image Classification

Wenjing Lv and Xiaofei Wang 

*School of Electronic Engineering, Heilongjiang University, Harbin 150080, China*

Correspondence should be addressed to Xiaofei Wang; [nk\\_wxf@hlju.edu.cn](mailto:nk_wxf@hlju.edu.cn)

Received 14 January 2020; Revised 20 May 2020; Accepted 19 June 2020; Published 8 July 2020

Academic Editor: Egidio De Benedetto

Copyright © 2020 Wenjing Lv and Xiaofei Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of remote sensing technology, the application of hyperspectral images is becoming more and more widespread. The accurate classification of ground features through hyperspectral images is an important research content and has attracted widespread attention. Many methods have achieved good classification results in the classification of hyperspectral images. This paper reviews the classification methods of hyperspectral images from three aspects: supervised classification, semisupervised classification, and unsupervised classification.

## 1. Introduction

In recent years, people have only begun to obtain hyperspectral remote sensing images with high spatial resolution and high spectral resolution relatively easily. Because hyperspectral images have strong resolving power for fine spectra, they have a wide range of applications in environmental [1], military [2], mining [3], and medical fields [4]. The acquisition of hyperspectral images depends on imaging spectrometers installed in different spaces. The imaging spectrum was established in the 1980s. It is used to image in the ultraviolet, visible, near-infrared, and mid-infrared regions of electromagnetic waves. The imaging spectrometer can image in many continuous and very narrow bands, so each pixel in the used wavelength range can get a fully reflected or emitted spectrum. Therefore, hyperspectral images have the characteristics of high spectral resolution, many bands, and abundant information. The processing methods of hyperspectral remote sensing images mainly include image correction [5], noise reduction [6], transformation [7], dimensionality reduction, and classification [8]. Unlike ordinary images, hyperspectral images are rich in spectral information, and this spectral information can reflect the physical structure and chemical composition of the object of interest, which is

helpful for image classification. Hyperspectral image classification is the most active part of the research in the hyperspectral field [9].

Computer classification of remote sensing images is to identify and classify the information of the earth's surface and its environment on the remote sensing images, so as to identify the feature information corresponding to the image information and extract the required feature information. Computer classification of remote sensing images is the specific application of automatic pattern recognition technology in the field of remote sensing. Moreover, with the further development of hyperspectral imaging systems, the information collected by hyperspectral images will be more detailed and richer. On the one hand, the spatial resolution will be higher; on the other hand, the spectral resolution will be significantly improved. With the further development of hyperspectral imagers, the amount of information contained in hyperspectral images will become even greater, and the scope of application of hyperspectral images will also be wider. The application of different occasions and the increasing amount of data also put forward more complicated requirements for the hyperspectral remote sensing ground observation technology. Compared with multispectral images, the number of imaging bands in hyperspectral images is greater, and

the ability to resolve objects is stronger, that is, the higher the spectral resolution. However, due to the high-dimensional characteristics of hyperspectral data, the similarity between the spectra and the mixed pixels, the hyperspectral image classification technology still faces a series of challenges, mainly including the following problems that need to be solved [10, 11].

- (1) The data of hyperspectral images have high dimensionality. Because hyperspectral images are obtained by using spectral reflectance values collected by airborne or space-borne imaging spectrometers in hundreds of bands, the corresponding spectral information dimension of hyperspectral images is also up to hundreds of dimensions.
- (2) Missing labeled samples. In practical applications, it is relatively easy to collect hyperspectral image data, but it is extremely difficult to obtain image-like label information. Therefore, the lack of labeled samples is often faced in the classification of hyperspectral images.
- (3) Spatial variability of spectral information. Affected by factors such as atmospheric conditions, sensors, composition and distribution of ground features, and surrounding environment, the spectral information of hyperspectral images changes in the spatial dimension, resulting in that the ground feature corresponding to each pixel is not single.
- (4) Image quality. During the acquisition of hyperspectral images, the interference of noise and background factors seriously affects the quality of the collected data. The image quality directly affects the classification accuracy of hyperspectral images.

The high dimensionality of hyperspectral image data and the lack of labeled samples can lead to the Hughes phenomenon [12]. Earlier in the research of hyperspectral image classification, people often focused on spectral information, using only spectral information to achieve image classification, and developed many classification methods, such as support vector machine (SVM) [13], random forest (RF), neural networks [14], and Polynomial logistic regression [15]. Dimension reduction methods such as feature extraction and feature selection have also been proposed, such as principal component analysis (PCA) [16, 17], independent component analysis (ICA) [18, 19], and linear discriminant analysis (LDA) [20]. The other is a nonlinear feature extraction method. For example, in 2000, the local linear embedding (LLE) [21] algorithm published by Science of Roweis and Saul in Science projects local high-dimensional data points into a low-dimensional coordinate system. The overall information is obtained by superimposing local neighborhoods, maintaining the same topological relationship, and retaining the overall geometric properties. At the same time, Tenenbaum et al. proposed (Isometric Feature Mapping, ISOMAP) [22] an algorithm based on the classic MDS [23]. It uses geodesic distance to embed high-dimensional data into low-dimensional coordinates. The neighborhood structure between high-dimensional spatial data points is still

retained in low-dimensional coordinate space. Belki and Niyogi proposed a similar pull to LLE in 2001. Laplacian Eigenmap (LE) [24, 25], also known as Spectral Clustering (SC); these nonlinear feature extraction methods are used in classification for practical applications.

Because the spatial context information is not considered, the above classification methods based on spatial information have not achieved good classification results. According to the research published in [26, 27], spatial context information plays an important role in the classification of hyperspectral images and can effectively avoid the phenomenon of homologous hyperspectral phenomena or homologous heterogeneous phenomena caused by using only spectral information. In a certain spectral region, two different features may show the same spectral line characteristics. This is the same spectrum of foreign objects, and may also be the same ground feature in different states, such as different relative angles to sunlight, different densities, and different water contents, showing different spectral line characteristics. The same-spectrum heterospectrum phenomenon or the same-object heterospectrum phenomenon has a certain effect on the classification effect, so spatial context information needs to be used in classification. The combination of spatial information and spectral information is another hot area in hyperspectral image classification. More and more scholars are beginning to explore this new direction. In [28, 29], the extended morphological profile (EMP) method was used to extract the spatial information of the image. In addition, the joint sparse representation models [30, 31] can also mine spatial information.

It is worth noting that deep learning [32] has excellent capabilities in image processing. Especially in recent years, image classification, target detection, and other fields have set off a wave of deep learning. Some deep learning network models have been used in remote sensing image processing, such as the Convolutional neural network (CNN), deep belief network (DBN) [33], and recurrent neural network (RNN). Moreover, in order to solve the problem of poor classification results due to the lack of training samples, a new tensor-based classification model [34–36] was proposed. Experiments confirmed that this method is superior to support vector machines and deep learning when the number of training samples is small.

Hyperspectral image classification methods are classified into supervised classification [37–39], unsupervised classification [40, 41], and semisupervised classification [42, 43] according to whether the classification information of training samples is used in the classification process.

## 2. Supervision Classification

The supervised classification method is a commonly used hyperspectral image classification method. The basic process is first, determine the discriminant criteria based on the known sample category and prior knowledge and calculate the discriminant function; Commonly used supervised classification methods include support vector machine method, artificial neural network classification method, decision tree classification method, and maximum likelihood classification method.

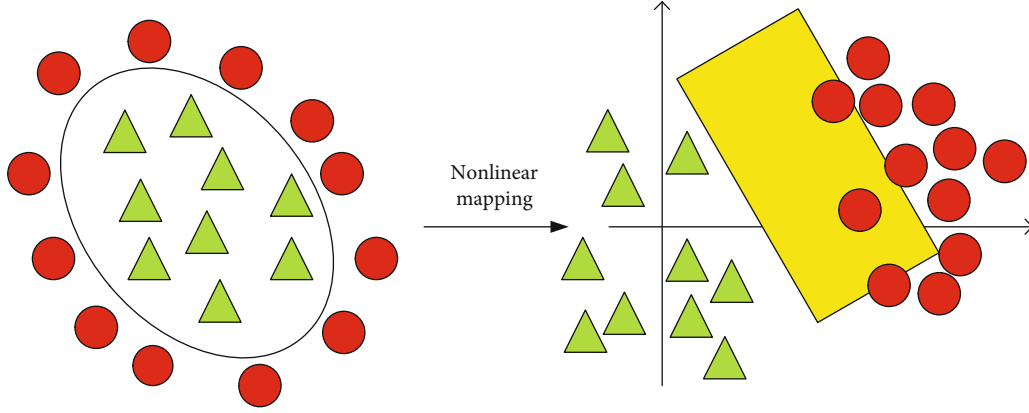


FIGURE 1: Kernel function support vector machine diagram.

**2.1. Support Vector Machines.** Support Vector Machine (SVM) [44] is a supervised classification method proposed by Boser et al. Based on statistical theory and based on the principle of minimizing structural risk, it solves a quadratic constraint with inequality constraints. As a machine learning method, the support vector machine method plays a huge role in image and signal processing and recognition. SVM applies the structural risk minimization principle to a linear classifier to find the optimal classification surface. It requires that the classification surface can not only separate the two types of sample points without error but also maximize the classification gap between the two types. Suppose a hyperspectral image  $X = \{x_1, x_2, \dots, x_n\}$ ,

$x_i = \{x_{i1}, x_{i2}, \dots, x_{iD}\}^T$  represents the spectral vector of the pixel  $i$  in the image,  $D$  represents the total number of bands, and  $n$  is the total number of pixels. In addition, we define  $y = (y_1, y_2, \dots, y_n)$  the classification mark image, in the formula  $y_i \in (-1, 1)$ . The mathematical process of a classic SVM classifier is

$$y_i = \text{sgn} \left( \sum_{j=1}^{l_n} y_j \alpha_j (x_j^T \cdot x_i) + b \right). \quad (1)$$

Among them,  $l_n$  is recorded as the number of a priori marks, and  $\alpha_i$  is a soft interval parameter. By setting  $b = 0$ , the optimal classification plane can pass the origin of the coordinate system, thereby simplifying the calculation. In practical operations, the situations we encounter are not all linearly separable, so we introduce slack variables, the mathematical expression of the support vector machine after introducing the slack variable is

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (X_i^T X_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \quad (2)$$

Among them,  $C$  is a constant, called the penalty factor or regularization parameter.

For nonlinear cases, that is, cases where the data is nonlinearly separable due to the hyperspectral data itself or the external environment. At this time, the classic support vector machine classification method can no longer meet the classification requirements, and the kernel function [45] came into being. After introducing the concept of a kernel function, the basic idea of SVM can be summarized as simple: firstly transform the input space to a high-dimensional space by nonlinear transformation, and then find the optimal linear classification surface in this new space, and this nonlinear transformation is achieved by defining the appropriate inner product function. The more commonly used kernel functions include linear kernel functions, polynomial kernel functions, and Gaussian kernel functions. Figure 1 is a schematic diagram of a kernel function support vector machine.

There is  $x_i^T x_j$  in the above formula, Order kernel function  $k(x_i, x_j) = x_i^T x_j$

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (X_i^T X_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0. \end{aligned} \quad (3)$$

**2.2. Minimum Distance Classification.** Minimum distance classifier (MDC) is a supervised classification based on the distance of pixels in the feature space as a classification basis. It is generally considered that in the feature space; feature points belonging to the same class are clustered in space. The mean vector determined by these feature points is used as the center of the category, and the covariance matrix is used to describe the dispersion of surrounding points. Points are similarly measured with each category. The basic assumption of the similarity measure is if the feature differences between the two modes are below a set threshold, the two modes are said to be similar. It uses the area formed by the collection of various training sample points to represent

various decision-making regions and uses distance as the main basis for measuring the similarity of samples. There are many forms of distance calculation, including Ming's distance, Mahalanobis distance, absolute value distance, Euclidean distance, Che's distance, and Barth's distance. Among them, the Mahalanobis distance and the Barth-Parametric distance not only consider the class means vector, but also consider the distribution of each feature point around the center of the class, so it has a more effective classification result than other distance criteria, but the calculation amount, it is larger than several other criteria.

The minimum distance method matching method: ① select a feature type from the spectral library; ② calculate the distance between the feature in the spectral library and the feature to be matched; ③ set a threshold for classification. He is one of the earliest applied methods for image classification research. Due to its advantages such as intuitiveness and simple calculation, it is still widely used today. For some classifications with insufficient training sample points, it can obtain better classification results than other complex classifiers. Figure 2 is a flowchart of the minimum distance classification method.

**2.3. Maximum Likelihood Classification.** Maximum Likelihood Classifier (MLC) is a classification method based on the Bayesian criterion. The maximum likelihood of discrimination classification is a nonlinear classification method. The statistical feature values of each type of training samples are calculated during classification. Establish a classification discriminant function, use the discriminant function to find the probability that each pixel in the hyperspectral remote sensing image belongs to various types, and classify the test sample into the category with the highest probability. Each category in the remote sensing image has a projection in either direction of the feature space, but when the projections of these different directions are difficult to distinguish, it means that the method of linear discrimination is also not ideal. This requires the establishment of a nonlinear. It is possible to obtain better results with classification boundaries. Due to a large amount of data in the hyperspectral remote sensing image, the covariance matrix generated will be very large, and it will be more difficult to calculate when using this covariance matrix. Therefore, the maximum likelihood discrimination classification method can generally obtain better results. When the training samples are normally distributed, the classification method obtained by the maximum likelihood classification method is better. A remote sensing ground feature image can use its spectral feature vector  $X$  as a measure to find a corresponding feature point in the spectral feature space: and each feature point from a similar feature will form a cluster of a certain probability in the feature space. The conditional probability  $P(\omega_i | X)$  of a feature point ( $X$ ) falling into a certain cluster ( $\omega_i$ ) can be used as a component category decision function, which is called a likelihood decision function. Assuming that  $g_i(x)$  is a discriminant function, the probability  $p(\omega_i | x)$  that a pixel  $x$  belongs to class  $\omega_i$  can be expressed as

$$g_i(x) = p(\omega_i | x). \quad (4)$$

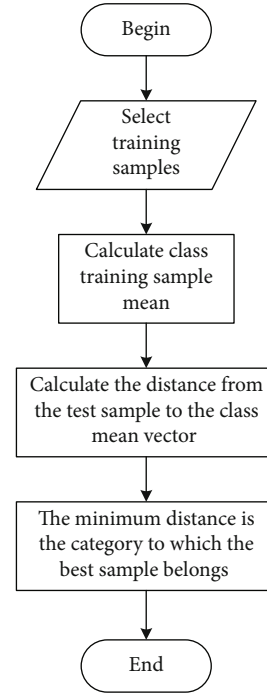


FIGURE 2: Schematic diagram of minimum distance classification.

According to the Bayesian formula, there is

$$g_i(x) = p(\omega_i | x) = p(x | \omega_i)p(\omega_i)/p(x), \quad (5)$$

where  $p(x | \omega_i)$  is the conditional probability that  $x$  belongs to  $\omega_i$ ,  $p(\omega_i)$  is the prior probability, and  $p(x)$  is the probability when  $x$  has nothing to do with the category. Maximum likelihood classification assumes normal distribution of hyperspectral data, and the discriminant formula is

$$g_i(x) = p(\omega_i | x) = p(\omega_i) = \frac{p(\omega_i)}{(2\pi)^{K/2} |\Sigma_i|^{1/2}} \exp \left[ -\frac{1}{2} (x - u_i)^T \Sigma_i^{-1} (x - u_i) \right], \quad (6)$$

where  $i$  is the number of classes,  $K$  is the number of features,  $\Sigma_i$  is the covariance matrix of the  $i$ -th class,  $|\Sigma_i|$  is the determinant of the matrix  $\Sigma_i$ , and  $u_i$  is the mean vector.

**2.4. Neural Network Classification.** Artificial Neural Networks (Artificial Neural Networks, ANN) is the most popular artificial intelligence classification method at present. It is characterized by simulating the processing and processing of information by human neurons. It has been widely used in intelligent control, information processing, and combinatorial optimization. However, artificial neural networks also have their own weaknesses, such as the need for a large amount of training data, slower operation speed, and difficulty in obtaining decision surfaces in the feature space. Neuron classification methods are commonly used in BP neural networks [46], radial basis neural networks [47], and wavelet neural networks [48]. Among them, the BP neural network model (feedforward network model) is currently the most



widely used neural network model. It consists of an input layer, a hidden layer, and an output layer. When an input mode is given, the input signal is from the input layer to the transmission of the output layer is a forward propagation process. If there is an error between the output signal and the desired signal, the error is transferred to the backward propagation process, and the weight of each layer is adjusted according to the magnitude of the error of each layer.

The implementation process of BP neural network classification mainly includes two stages: the first stage is the network self-learning of the sample data to obtain an optimized connection weight matrix. The main steps include the determination of the network system, the input of sample data and control parameters, initialize the weights, and adjust the connection weights of each layer. The second stage is to use the learning results to classify the entire image data. After inputting multispectral images, the network uses the connection weight matrix of the network obtained during the learning process to the image data is calculated. According to the comparison between the output result and the expected value of each type, each pixel is classified as the one with the smallest error.

SVM is a novel small sample learning method with a solid theoretical foundation, so it does not require a large number of training samples. However, it also has defects when dealing with large-scale samples, and it cannot solve the multi-classification problem well. The principle of the minimum distance classification method is simple; the classification accuracy is not high, but the calculation speed is fast; it can be used in a quick overview of the classification. In practical application, the maximum likelihood method, the minimum distance method, and the neural network method can be used, but strict and precise supervision by humans is required to ensure that the accuracy meets certain requirements.

### 3. Deep Learning

In recent years, hyperspectral image classification methods have introduced spatial information of hyperspectral images. This type of method is simply referred to as hyperspectral image classification methods based on spatial-spectral joint features. Deep learning originates from artificial neural networks. Compared with artificial neural networks, deep learning has a stronger pumping ability. Deep learning models have deeper layers, which also helps to extract feature information. This section mainly introduces convolutional neural networks (CNN) in deep learning [49, 50], deep belief network (DBN), and stacked autoencoder (SAE).

**3.1. CNN. Classification method based on spectral features:** Hyperspectral images have very rich spectral information and extremely high spectral resolution. Each pixel can extract one-dimensional spectral vectors. These vectors are composed of spectral information. Classification using only one-dimensional spectral vectors is called a classification method based on spectral information. In the classification method based on spectral information, generally, the pixel

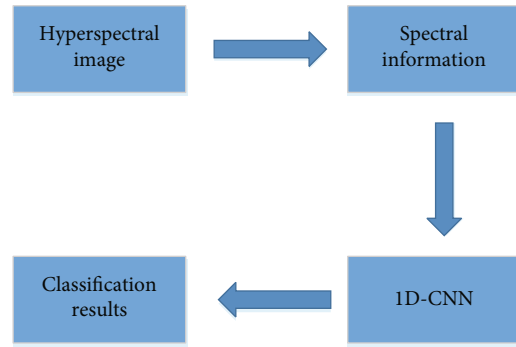


FIGURE 3: Schematic diagram of 1D-CNN.

is used to extract spectral information or to obtain certain specific features from spectral information through feature extraction to classify. Using CNN to classify spectral features of hyperspectral images is to use one-dimensional CNN (1D-CNN) to extract spectral features and classify them. The process is shown in Figure 3.

The specific process is: input labeled data in hyperspectral data into 1-DCNN, train 1-DCNN with class labels, and then iteratively update network weights through algorithms such as SGD, and finally use the trained 1-DCNN for each Pixel classification results in classification results. The one-dimensional convolution operation uses a one-dimensional convolution kernel to perform a convolution operation on a one-dimensional feature vector. Its expression is

$$v_{l,j}^x = f \left( \sum_m \sum_{h=0}^{H_l-1} k_{l,j,m}^h v_{(l-1),m}^{(x+h)} + b_{l,j} \right). \quad (7)$$

Among them,  $k_{l,j,m}^h$  represents the specific value of the  $l$ -th convolution kernel in the  $j$ -th layer at  $h$ , and the convolution kernel is connected to the  $m$ -th feature vector in the  $(l-1)$  layer network.  $H_l$  represents the length of the one-dimensional convolution kernel.  $b_{l,j}$  represents the offset of the  $j$ -th feature map of the  $l$ -th layer.  $v_{(l-1),m}^{(x+h)}$  represents the specific value of the  $m$ -th feature map at the  $(x+h, y+w)$  position in the  $l$ -th layer.

Classification method based on spatial features: spatial information, that is, context information. When classifying based on spatial information, instead of using the spectral information extracted from a certain pixel, the spatial information extracted from the neighborhood of the pixel is used instead. Due to the high latitude of hyperspectral data, the usual method for extracting spatial information is to first compress the data set, then use two-dimensional convolutional neural networks (2D-CNN) to extract deeper spatial information, and then use spatial information to classify. The specific process is shown in Figure 4.

The main difference between the two-dimensional convolution operation and the one-dimensional convolution operation is the dimensions of the convolution layer and the pooling layer. The two-dimensional convolution

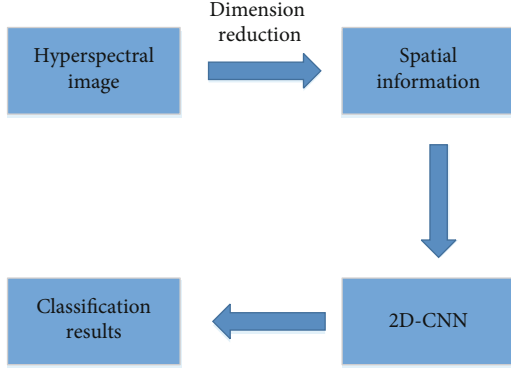


FIGURE 4: Schematic diagram of 2D-CNN.

operation uses a two-dimensional convolution kernel to perform a convolution operation on two-dimensional data:

$$\text{map}_{l,j}^{x,y} = f \left( \sum_m \sum_{h=0}^{H_l-1} \sum_{w=0}^{W_l-1} k_{l,j,m}^{h,w} \text{map}_{(l-1),m}^{(x+h),(y+w)} + b_{l,j} \right). \quad (8)$$

Among them,  $k_{l,j,m}^{h,w}$  represents the value of the  $j$ -th convolution kernel in the  $l$ -th layer at  $(h, w)$ , and this convolution kernel is connected to the  $m$ -th feature map in the  $l-1$  layer.  $H_l$  and  $W_l$ , respectively, represent the height and width of the convolution kernel, and  $b_{l,j}$  represents the offset of the  $j$ -th feature map of the  $l$ -th layer.  $\text{map}_{(l-1),m}^{(x+h),(y+w)}$  represents the specific value of the  $m$ -th feature map at the  $(x+h, y+w)$  position in the  $l-1$ th layer, and  $\text{map}_{l,j}^{x,y}$  represents the output data of the  $j$ -th feature map at the  $l$ -th layer at  $(x, y)$ .

Classification method based on spectral-spatial features: In traditional hyperspectral image classification, only spectral information is often used. However, the same ground features will show different spectral curves due to the influence of the external environment. Different ground objects may also exhibit the same spectral curve, which is the so-called same-object heterospectrum and foreign-object same-spectrum phenomenon. For example, some pixels connected in space are classified as parking lots, so the pixels whose spectral information appears very similar to the metal spectral information are likely to be cars. If many pixels around a pixel are grass, the pixels in the middle are likely to be grass too. Hyperspectral data presents a three-dimensional structure, containing both one-dimensional spectral information and two-dimensional spatial information. A three-dimensional convolutional neural network (3D-CNN) can simultaneously extract spectral information and spatial information. The specific process is shown in Figure 5.

**3.2. DBN.** The realization of a deep belief network (DBN) is based on restricted Boltzmann machine (RBM). DBN is a network model constructed by multiple RBMs layer by layer. A classic DBN network is composed of several RBMs and a layer of BP. The structure diagram is shown in Figure 6.

During training, a layer-by-layer unsupervised method is used to learn parameters. First, take the data and the first hid-

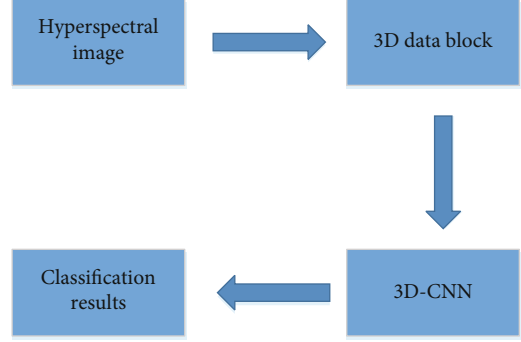


FIGURE 5: Schematic diagram of 3D-CNN.

den layer as an RBM, train the parameters of this RBM, then fix the parameters of this RBM, treat the first hidden layer as a visible vector, and the second hidden layer as a hidden vector, train the second RBM, get its parameters, then fix these parameters, and loop according to this method. The specific steps are

First, separately and unsupervisedly train each layer of RBM network to ensure that feature vectors retain as much feature information as possible when they are mapped to different feature spaces;

Second, set up the BP network in the last layer of the DBN, receive the RBM output feature vector as its input feature vector, and supervise the training of the entity relationship classifier. Moreover, each layer of the RBM network can only ensure that the weights in its own layer and the layer feature vector mapping is optimal, and it is not optimal for the entire DBN feature vector mapping, so the backpropagation network also propagates error information from top to bottom to each layer of RBM, and fine-tunes the DBN network and RBM network training model. The process can be seen as the initialization of the weight parameters of a deep BP network, so that the DBN overcomes the shortcomings of the BP network that is easy to fall into local optimization and long training time due to the random initialization of the weight parameters.

When using DBN to classify the spectral features of hyperspectral images, the main method is to use DBN to extract the deeper features of the spectral information collected from the positions of the pixels to be classified, and then use the deep features to complete the classification. Spatial features of hyperspectral images based on DBN, the classification method is very similar to the SAE-based hyperspectral image spatial feature classification method. In [51], Li et al. first used PCA to compress the original hyperspectral image, retain the first 3 principal components, and then extract the data in the  $7 \times 7$  neighborhood. As input, use the DBN network for feature extraction and classification. Literature [52] uses DBN to separately extract spectral features and spatial features, and then connect the spectral features and spatial features to form a spatial spectrum feature, and then complete the classification based on the spatial spectrum feature. The overall framework of the task and classification is basically the same as the framework in [53], and the method in [54] introduces sparse restrictions, Gaussian

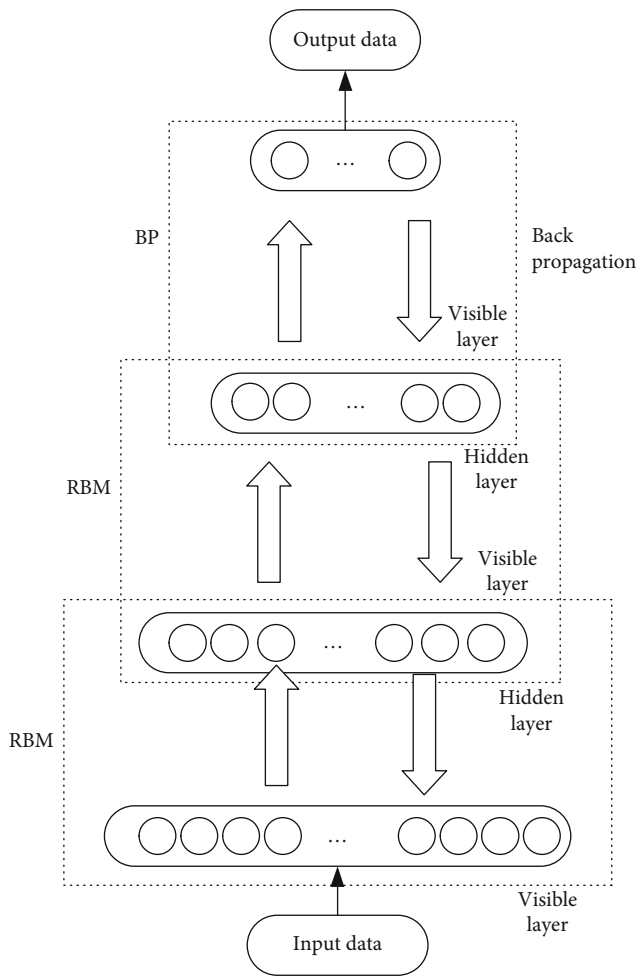


FIGURE 6: Classic DBN structure diagram.

noise, etc. in the training process of DBN. Figure 7 shows the flowchart of using DBN to classify hyperspectral images.

**3.3. SAE.** The SAE network structure is similar to the DBN, and it is also trained in an unsupervised form. Depending on the specific training method, the stacked SAE will also have different performance aspects. The hyperspectral image classification method based on the SAE network can also generally be used. It is divided into classification methods based on spectral features, classification methods based on spatial features, and classification methods based on spatial spectrum combined features.

When performing spectral feature classification of hyperspectral images based on SAE, the spectral vector extracted from the position of the pixel to be classified is usually used as the input data of the network, and SAE is used to extract deeper and more abstract depth spectral features, and then complete the classification task based on the depth spectral features.

When performing spatial feature classification of hyperspectral images based on SAE, the main method is to first reduce the dimensionality of the hyperspectral image and then extract all the information in the neighborhood with the pixels to be classified as the center, and pull the extracted

information into a vector as SAE input data. SAE cannot directly process data blocks with a two-dimensional structure in space. In order to meet SAE's input format requirements, the data needs to be stretched into a one-dimensional vector, which is different from the spatial feature classification of hyperspectral images based on CNN. The method differs greatly.

When SAE-based hyperspectral image spatial spectrum joint classification, the common practice is to separately extract spatial information and spectral information, and finally merge to obtain spatial spectrum joint information. The spectral vector extracted from the spatial position of the pixel to be classified is compressed in the hyperspectral image, extract spatial information from the neighborhood on the map and pull the spatial information into a vector, and merge the spectral vector and the vector drawn from the spatial information. Enter the merged information into the SAE, use the SAE to extract deep spatial-spectral features, and complete the classification.

It can be found through statistics that it can be found that the hyperspectral image classification methods based on SAE and DBN have developed earlier, and various hyperspectral image classification methods based on CNN have been proposed later. Spectral image classification methods have developed the fastest and have the largest number of papers. Overall, the classification results are better than the hyperspectral image classification based on SAE and DBN.

## 4. Unsupervised Classification

The unsupervised classification method refers to the classification based on the spectral similarity of the hyperspectral data, that is, the clustering method without any prior knowledge. Because no prior knowledge is used, unsupervised classification can only assume initial parameters, form clusters through preclassification processing, and then iterate to make the relevant parameters reach the allowable range.

**4.1. K-Means Classification.** The classification idea of the  $K$ -means clustering method is that the sum of the squares of the distances from all the pixels in each class to the center point of the class is the smallest. At the beginning of the clustering, the center point of the initial clustering is randomly selected, and other pixels to be classified are classified into one of the categories according to the prescribed principles to complete the initial clustering. Then recalculate the clustering center point of each class, modify the clustering center point, and classify again, so iterate until the position of the clustering center points no longer changes, find the best clustering center, and get the best Cluster results and stop the iteration. Figure 8 shows the algorithm flow of  $K$ -means clustering. During  $K$ -means clustering, the number of selected categories cannot be changed during the calculation process, and the position of the initial cluster center point initially selected will also affect the clustering result, so each time you may get a difference. Large experimental results, this is the disadvantage of  $K$ -means clustering. This kind of defects can help to find a better initial clustering center with some

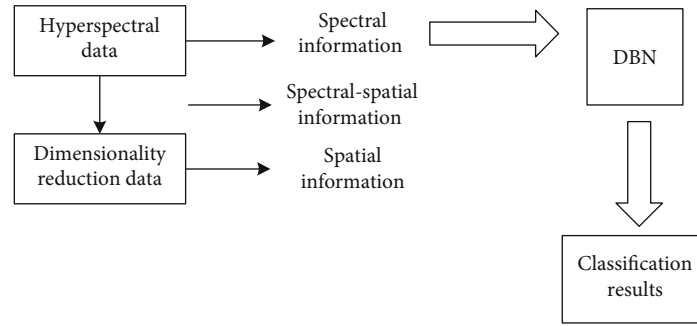


FIGURE 7: Flow chart of hyperspectral image classification based on DBN.

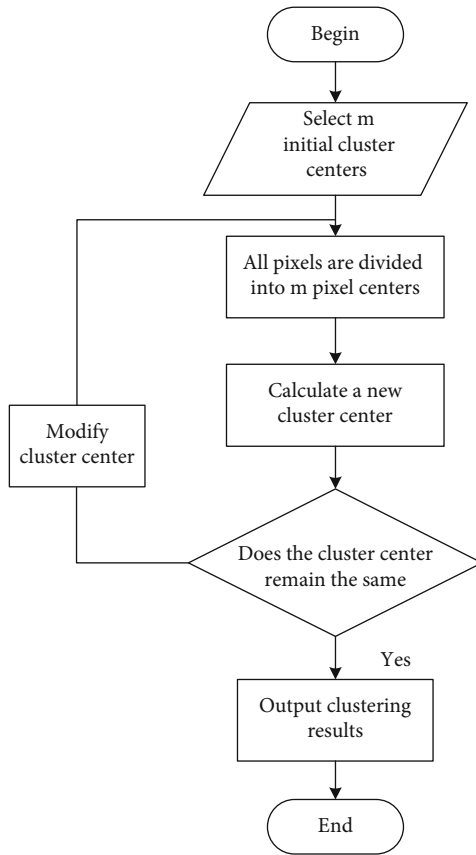


FIGURE 8: Schematic of  $K$ -means algorithm.

auxiliary methods, thereby improving the accuracy of classification.

**4.2. Iterative Self-Organizing Method.** ISODATA algorithm is also a commonly used clustering algorithm. The ISODATA algorithm is similar to the  $K$ -means classification algorithm. It is a method based on  $K$ -means classification improvement. Compared with  $K$ -means clustering, the ISODATA algorithm has more obvious advantages in classification. First, when clustering, there is no need to continuously adjust the cluster center during the calculation process, but all categories are calculated and then the samples are adjusted uniformly. Second, a big difference from  $K$ -means clustering is that when clustering the algorithm, the number of categories

can be automatically adjusted according to the actual situation, so that a reasonable clustering result can be obtained.

The main advantages of these two classification methods: there is no need to have a broad understanding of the classification area, only a certain amount of knowledge is required to explain the classified cluster groups; the chance of human error is reduced, and the initial parameters that need to be input are less; the clusters with small but unique spectral characteristics are more homogeneous than the supervised classification; the unique and small coverage categories can be identified. Main shortcomings: A lot of analysis and post-processing is required to obtain reliable classification results; the clusters and land categories that are classified may or may not correspond, plus the common “same spectrum” and “foreign material”. The phenomenon of “same spectrum” makes the matching of cluster groups and categories difficult. Because the spectral characteristics of each category change with time and terrain, the spectral cluster groups between different images cannot maintain their continuity and are difficult to compare.

## 5. Semisupervised Classification

The main disadvantage of the supervised method is that the classification model and classification accuracy mainly depend on the number of training data sets of label points, and obtaining a large number of hyperspectral image class labels is a time-consuming and cost-intensive task. Although unsupervised methods are not sensitive to labeled samples, due to the lack of prior knowledge, the relationship between clustering categories and real categories is uncertain [55]. Semisupervised classification uses both labeled and unlabeled data to train the classifier. It makes up for the lack of unsupervised and supervised learning. This classification method is based on the same type of labeled and unlabeled samples on the feature space. Closer assumptions, since a large number of unlabeled samples can more fully describe the overall characteristics of the data, the classifier trained using these two samples has better generalization.

This classification method is widely used in hyperspectral image classification. Typical semisupervised classification methods include model generation algorithms, semisupervised support vector machines, graph-based semisupervised algorithms, and self-training, collaborative training, and triple training.



Based on the above problems, this paper reviews a semi-supervised classification method. Semisupervised learning has attracted much attention from researchers in hyperspectral image classification because it only requires a small number of labeled samples [56]. Semisupervised learning combines labeled data with unlabeled data to improve classification accuracy.

**5.1. Laplace Support Vector Machine.** Laplacian support vector machine (Laplacian Support Vector Machine, LapSVM) is proposed on the basis of traditional SVM. By adding manifold regularization terms, the geometric information of unlabeled samples and labeled samples are fully utilized to construct a classifier. LapSVM can predict the labels of future test samples. And LapVM has the characteristics of strong adaptability and global optimization.

Given labeled samples and unlabeled samples  $\{x_i\}_{i=l+1}^{l+u}$ ,  $x_i \in R^m$ , and  $y_i \in \{-1, +1\}$ , the decision function is  $f$ . The

$$L = \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f) + \gamma_L \|f\|_H^2 + \gamma_M \|f\|_M^2. \quad (9)$$

Among them,  $V$  represents the mis-segmentation cost function of labeled samples,  $\gamma_L$  controls the complexity of the function  $f$  in Hilbert space, and  $\gamma_M$  controls the complexity of the geometric characteristics of the data distribution within the maximum distance of  $f$ . The structure of LapSVM is described in detail below. First, LapSVM uses the same loss function as traditional SVM:

$$V(x_i, y_i, f) = \max \{0, 1 - y_i f(x_i)\}. \quad (10)$$

Among them,  $f$  represents the classification decision function  $f(x) = \langle w, \varphi(x) \rangle + b$  of the selected classifier, where  $\varphi(\cdot)$  represents a nonlinear mapping function from a low-dimensional space to a high-dimensional Hilbert space, where

$$w = \sum_{i=1}^{l+u} \alpha_i \varphi(x_i) = \Phi \alpha, \Phi = [\varphi(x_1), \dots, \varphi(x_{l+u})]^T, \quad (11)$$

$\alpha = [\alpha_1, \dots, \alpha_{l+u}]$ , is a decision function after finishing:

$$f(x) = \sum_{i=1}^{l+u} \alpha_i K(x_i, x) + b. \quad (12)$$

Here,  $K$  represents the kernel function, and different learner functions are realized by selecting different kernel functions, so there are

$$\|f\|_H^2 = \|w\|^2 = (\Phi \alpha)^T (\Phi \alpha) = \alpha^T K \alpha. \quad (13)$$

LapSVM algorithm simulates the geometric distribution of data by constructing a graph using labeled samples and unlabeled samples. Using the smoothing assumption to normalize the graph, the fast-changing part of the penalty classification function is

$$\|f\|_H^2 = \frac{1}{(l+u)^2} \sum_{i,j=1}^{l+u} W_{ij} (f(x_i) - f(x_j))^2 = f^T L f. \quad (14)$$

Substituting the above formula into

$$\begin{aligned} \min_{\xi_i \in R^l, \alpha \in R^{l+u}} & \left\{ \frac{1}{l} \sum_{i=1}^l \xi_i + \gamma_L \alpha^T K \alpha + \frac{\gamma_M}{(l+u)^2} \alpha^T K L K \alpha \right\} \\ \text{s.t.} & \xi_i y_i \left( \sum_{i,j=1}^{l+u} \alpha_i K(x_i, x_j) + b \right) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, l, \end{aligned} \quad (15)$$

where  $\xi_i$  represents the relaxation factor of the labeled sample.

LapSVM algorithm fully reflects the role of unlabeled samples in the classification process through the geometric characteristics of the data, but often requires a large computational cost.

**5.2. Self-Training.** Self-training is a commonly used semisupervised classification algorithm. In the implementation of the algorithm, a classifier is first trained with labeled samples, and then a large number of unlabeled samples are labeled with this classifier. Data with high confidence are selected from the labeled samples, and these data are added to the initial training set together with their labels to retrain the classifier. Repeat this process until the end condition is met. The general process of self-training is as follows:

- (1) Use the initial labeled sample set to train the classifier
- (2) Use the classifier to label the data in the set of unlabeled samples, and select the samples with the highest confidence, and record as
- (3) Retrain the classifier using the new sample set
- (4) Repeat steps 2) and 3) until the end condition is met

Self-training algorithms have been widely used. This classification method is simple and convenient. However, because the number of initial training samples is generally limited, it is difficult to train a classifier with good generalization performance and high accuracy. When unlabeled samples are labeled, a large number of mislabeled samples will be generated. Such samples will always exist as noise samples when they are added to the original training set. With the iterative loop, errors accumulate, which will inevitably cause the classification performance of the classifier to decline.

## 6. Evaluation of Classification Results

After classification processing of hyperspectral remote sensing images, we need to judge the quality of the classification results, and then evaluate the performance of the classifier. Some evaluation criteria are used to evaluate the degree of agreement between the classification results and the real

features, that is, the classification accuracy is calculated to visually show the feasibility of the proposed algorithm is followed by a comparative analysis of other classification algorithms, and further improvements based on their shortcomings. Commonly used evaluation indicators are confusion matrix, producer accuracy, user accuracy, Kappa coefficient, overall classification accuracy, etc. [57], and its implementation principle will be briefly introduced below.

**6.1. Confusion Matrix.** The confusion matrix is also called the error matrix [58], which is mainly used to compare whether the classification result is the same as the actual ground cover. Various other evaluation criteria are based on this. Assume that the confusion matrix of order  $c \times c$  is

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1c} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ x_{c1} & x_{c2} & \cdots & x_{cc} \end{bmatrix}, \quad (16)$$

where  $c$  is the number of categories, and  $x_{ij}$ , ( $i, j = 1, 2, \dots, c$ ) is the number of samples of the  $i$ -th category divided into the  $j$ -th category, and the element  $x_{ii}$  on the diagonal represents the number of sample points that were correctly divided. Where  $n$  is the total number, and the total number of sample points is calculated as

$$n = \sum_{i=1}^c \sum_{j=1}^c x_{ij}. \quad (17)$$

**6.2. Overall Accuracy.** The overall accuracy refers to the ratio of the number of sample points that are correctly divided into the total number of samples. The calculation formula is

$$P_{OA} = \frac{\sum_{i=1}^c x_{ii}}{\sum_{i=1}^c \sum_{j=1}^c x_{ij}}. \quad (18)$$

**6.3. Kappa Coefficient.** The Kappa coefficient can comprehensively evaluate the division results of images. The Kappa coefficient comprehensively considers the number of sample points that are correctly divided and the number of incorrect divisions to evaluate the classification results, which is very convincing [59]. Based on the confusion matrix, the following formula can be used:

$$P_{KC} = \frac{n \sum_{i=1}^c x_{ii} - n \sum_{i=1}^c (x_{i+} x_{+i})}{n^2 - \sum_{i=1}^c (x_{i+} x_{+i})}, \quad (19)$$

where  $n$  is total and  $x_{i+}$  is the sum of the elements of each row, defined as follows:

$$x_{i+} = \sum_{j=1}^c x_{ij}. \quad (20)$$

TABLE 1: Comparison table of classification results.

Class	SVM	RF	BP	1D-CNN
OA	78.0	71.5	80.5	78.0
AA	86.8	82.0	82.3	81.0
KAPPA	73.3	65.0	76.6	77.9

$x_{+i}$  represents the sum of the elements in each column and is defined as follows:

$$x_{+i} = \sum_{j=1}^c x_{ij}. \quad (21)$$

Table 1 gives several classification methods of supervised classification; the data comes from Pavia University. It should be noted that percentages are added after the values of OA and AA.

For the above methods, the SVM training time is longer but the classification speed is fast; the RF training speed is fast, the implementation is simple but it is prone to overfitting; the BP algorithm has fault tolerance and strong self-learning ability, but the convergence speed is slow and the training time is long; 1D-CNN can handle high-dimensional data but requires a large number of samples and it is best to use a GPU for training.

Table 2 gives several classification methods based on deep learning. Similarly, the data also comes from Pavia University and gives a comparison table of the classification effects of each method using spectral information, spatial information, and spatial-spectral information.

Comparing Tables 1 and 2, it can be seen that the classification effect of the deep learning method is superior to the ordinary classification method, especially after combining spatial information and spectral information, the classification performance has been greatly improved. However, the deep learning method requires a large number of training samples for classification, and the training time is long. If there are enough training samples, the deep learning method is a good choice. If there are no training samples, then SVM, etc. can also be used for classification.

## 7. Conclusion

Classification and recognition of hyperspectral images are important content of hyperspectral image processing. This paper discusses several methods of hyperspectral image classification, including supervised and unsupervised classification and semisupervised classification. Although the supervised and unsupervised classification methods described in this article have their respective advantages to varying degrees, there are limitations in the application of various methods. For example, supervised classification requires a certain number of prior conditions, and human factors will affect the classification results have an impact. Therefore, based on different application requirements, combined with the acquisition of hyperspectral images with massive information, multiple methods need to be combined

TABLE 2: Comparison of the effects of deep learning classification methods.

Network model	OA	Spectral feature AA	K	OA	Spatial feature AA	K	OA	Spectral-spatial feature AA	K
CNN	92.3	92.6	90.4	94.0	97.6	92.4	99.5	99.8	99.6
DBN	95.1	94.0	93.7	98.1	97.3	97.6	98.5	97.9	98.0
SAE	96.5	95.1	95.3	98.7	98.0	98.1	99.0	98.4	98.8

with each other in order to achieve the desired classification effect. With the development of hyperspectral image technology, hyperspectral image classification has been widely used. Existing theories and methods still have certain limitations for more complicated hyperspectral image classification. Therefore, researching more targeted hyperspectral image classification methods will be an important research direction in the future.

### Data Availability

Data are available on request. Please contact Xiaofei Wang to request the data.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work is supported by the National Natural Science Foundations of China (61871150) and the National Key R&D Program of China (2016YFB0502502).

### References

- [1] M. Y. Teng, R. Mehrubeoglu, S. A. King, K. Cammarata, and J. Simons, "Investigation of epifauna coverage on seagrass blades using spatial and spectral analysis of hyperspectral images," in *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Gainesville, FL, USA, 2013.
- [2] C. Ke, "Military object detection using multiple information extracted from hyperspectral imagery," in *2017 International Conference on Progress in Informatics and Computing (PIC)*, pp. 124–128, Nanjing, China, 2017.
- [3] G. Notesco, E. Ben Dor, and A. Brook, "Mineral mapping of makhtesh ramon in israel using hyperspectral remote sensing day and night LWIR images," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Lausanne, Switzerland, 2014.
- [4] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 653–663, 2016.
- [5] X. Zhang, A. Zhang, and X. Meng, "Automatic fusion of hyperspectral images and laser scans using feature points," *Journal of Sensors*, vol. 2015, Article ID 415361, 9 pages, 2015.
- [6] H. Kwon, X. Hu, J. Theiler, A. Zare, and P. Gurrum, "Algorithms for multispectral and hyperspectral image analysis," *Journal of Electrical and Computer Engineering*, vol. 2013, Article ID 908906, 2 pages, 2013.
- [7] Y. Ma, R. Li, G. Yang, L. Sun, and J. Wang, "A research on the combination strategies of multiple features for hyperspectral remote sensing image classification," *Journal of Sensors*, vol. 2018, Article ID 7341973, 14 pages, 2018.
- [8] H. Binol, "Ensemble learning based multiple kernel principal component analysis for dimensionality reduction and classification of hyperspectral imagery," *Mathematical Problems in Engineering*, vol. 2018, Article ID 9632569, 14 pages, 2018.
- [9] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: an overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [10] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
- [11] G. Camps-Valls, T. V. Bandos Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044–3054, 2007.
- [12] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.
- [13] X. Wang and Y. Feng, "New method based on support vector machine in classification for hyperspectral data," in *2008 International Symposium on Computational Intelligence and Design*, pp. 76–80, Wuhan, China, 2008.
- [14] A. Fu, X. Ma, and H. Wang, "Classification of hyperspectral image based on hybrid neural networks," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2643–2646, Valencia, Spain, 2018.
- [15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.
- [16] P. Deepa and K. Thilagavathi, "Feature extraction of hyperspectral image using principal component analysis and folded-principal component analysis," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, pp. 656–660, Coimbatore, India, 2015.
- [17] M. Imani and H. Ghassemian, "Principal component discriminant analysis for feature extraction and classification of hyperspectral images," in *2014 Iranian Conference on Intelligent Systems (ICIS)*, pp. 1–5, Bam, Iran, 2014.
- [18] C. Jayaprakash, B. B. Damodaran, V. Sowmya, and K. P. Soman, "Dimensionality reduction of hyperspectral images for classification using randomized independent component analysis," in *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 492–496, Noida, India, 2018.
- [19] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in

- hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.
- [20] Q. Du, "Modified Fisher's linear discriminant analysis for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 503–507, 2007.
- [21] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [22] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [23] J. D. Carroll, P. Arabie, and L. J. Hubert, "Multidimensional scaling (MDS)," *Encyclopedia of Social Measurement*, vol. 45, no. 2, pp. 779–784, 2005.
- [24] J. Jost, *Riemannian Geometry and Geometric Analysis*, Universitext, Springer, 5th edition, 2008.
- [25] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [26] Y. Yan, Y. Zhao, H.-f. Xue, X.-d. Kou, and Y. Liu, "Integration of spatial-spectral information for hyperspectral image classification," in *2010 Second IITA International Conference on Geoscience and Remote Sensing*, pp. 242–245, Qingdao, 2010.
- [27] X. Xu, J. Li, Y. Zhang, and S. Li, "A subpixel spatial-spectral feature mining for hyperspectral image classification," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 8476–8479, Valencia, Spain, 2018.
- [28] J. Wang, G. Zhang, M. Cao, and N. Jiang, "Semi-supervised classification of hyperspectral image based on spectral and extended morphological profiles," in *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–4, Los Angeles, CA, USA, 2016.
- [29] X. Zhang, Y. Sun, and W. Qi, "Hyperspectral image classification based on extended morphological attribute profiles and abundance information," in *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5, Amsterdam, Netherlands, 2018.
- [30] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.
- [31] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7738–7749, 2014.
- [32] A. Voulodimos, "Deep learning for computer vision: a brief review," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 7068349, 13 pages, 2018.
- [33] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: a review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [34] K. Makantasis, A. D. Doulamis, N. D. Doulamis, and A. Nikitakis, "Tensor-based classification models for hyperspectral data analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 12, pp. 6884–6898, 2018.
- [35] K. Makantasis, A. Doulamis, N. Doulamis, A. Nikitakis, and A. Voulodimos, "Tensor-based nonlinear classifier for high-order data analysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, 2018.
- [36] K. Makantasis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Common mode patterns for supervised tensor subspace learning," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019.
- [37] H. Sima, P. Liu, L. Liu, A. Mi, and J. Wang, "Sparse representation classification based on flexible patches sampling of superpixels for hyperspectral images," *Mathematical Problems in Engineering*, vol. 2018, Article ID 8264961, 10 pages, 2018.
- [38] J. Jiang, L. Huang, H. Li, and L. Xiao, "Hyperspectral image supervised classification via multi-view nuclear norm based 2D PCA feature extraction and kernel ELM," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1496–1499, Beijing, July 2016.
- [39] L. Zhang, G.-S. Xia, T. Wu, L. Liang, and X. C. Tai, "Deep Learning for remote sensing image understanding," *Journal of Sensors*, vol. 2016, Article ID 7954154, 2 pages, 2016.
- [40] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 47, no. 12, pp. 4175–4188, 2009.
- [41] W. Jian, "Unsupervised Intrusion Feature Selection Based on Genetic Algorithm and FCM," in *Information Engineering and Applications*, R. Zhu and Y. Ma, Eds., vol. 154 of Lecture Notes in Electrical Engineering, pp. 1005–1012, Springer, London, 2012.
- [42] C. Wang, Z. Guo, S. Wang, L. Wang, and C. Ma, "Improving hyperspectral image classification method for fine land use assessment application using semisupervised machine learning," *Journal of Spectroscopy*, vol. 2015, Article ID 969185, 8 pages, 2015.
- [43] Y. Han, J. Ren, Z. Hong et al., "Active learning algorithms for the classification of hyperspectral sea ice images," *Mathematical Problems in Engineering*, vol. 2015, Article ID 124601, 10 pages, 2015.
- [44] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152, ACM Press, 1992.
- [45] C. Li, P. Hsieh, and B. Kuo, "Multiple SVMs based on random subspaces from kernel feature importance for hyperspectral image classification," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 574–577, Fort Worth, TX, USA, 2017.
- [46] Y. Lei, "Fusion method of PCA and BP neural network for face recognition," in *2011 International Conference on Computer Science and Service System (CSSS)*, pp. 3256–3259, Nanjing, China, 2011.
- [47] H. Sarnel, Y. Senol, and D. Sagirlibas, "Accurate and robust image registration based on radial basis neural networks," in *2008 23rd International Symposium on Computer and Information Sciences*, pp. 1–5, Istanbul, Turkey, 2008.
- [48] X. Gao, "A comparative research on wavelet neural networks," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02*, vol. 4, pp. 1699–1703, Singapore, 2002.
- [49] K. Makantasis, K. Karantzas, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *2015 IEEE*



- International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4959–4962, Milan, Italy, 2015.
- [50] T. Guo, J. Dong, H. Li, and Y. Gao, “Simple convolutional neural network on image classification,” in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 721–724, Beijing, China, 2017.
- [51] T. Li, J. P. Zhang, and Y. Zhang, “Classification of hyperspectral image based on deep belief networks,” in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5132–5136, Paris, France, 2014.
- [52] J. H. Le, A. P. Yazdanpanah, E. Regentova, and V. Muthukumar, “A deep belief network for classifying remotely-sensed hyperspectral data,” in *Advances in Visual Computing. ISVC 2015*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, I. Pavlidis, R. Feris, T. McGraw, M. Elendt, R. Kopper, E. Ragan, Z. Ye, and G. Weber, Eds., vol. 9474 of Lecture Notes in Computer Science, pp. 682–692, Cham, Springer, 2015.
- [53] Y. S. Chen, Z. H. Lin, X. Zhao, G. Wang, and Y. F. Gu, “Deep learning-based classification of hyperspectral data,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
- [54] Y. S. Chen, X. Zhao, and X. P. Jia, “Spectral-spatial classification of hyperspectral data based on deep belief network,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2381–2392, 2015.
- [55] Z. Cheng and F. Xie, “Semi-supervised classification of hyperspectral images based on spatial features and texture information,” *Bulletin of Surveying and Mapping*, vol. 51, no. 12, pp. 56–59, 2016.
- [56] T. Yanping, T. Chao, Z. Jun, Z. Yang, and X. He, “Hyperspectral image classification combined with active learning and semi-supervised graphs,” *Journal of Surveying and Mapping*, vol. 44, no. 8, pp. 919–926, 2015.
- [57] L. Zhimin, H. Panchao, H. Hong, and H. Wen, “Semi-supervised graph clustering with composite kernel and its application in hyperspectral image,” *Opto-electronic Engineering*, vol. 43, no. 4, pp. 33–39, 2016.
- [58] L. Xiangjun, *Discussion on land use change detection method by remote sensing*, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, 2006.
- [59] R. G. Congalton and R. A. Mead, “A quantitative method to test for consistency and correctness in photointerpretation,” *Photogrammetric Engineering and Remote Sensing*, vol. 49, no. 1, pp. 69–74, 1983.