processes

MDPI

Article

A Comparison of Three Different Group Intelligence Algorithms for Hyperspectral Imagery Classification

Yong Wang and Weibo Zeng *

Geographic Information and Tourism College, Chuzhou University, Chuzhou 239099, China
* Correspondence: njzwb@chzu.edu.cn

Abstract: The classification effect of hyperspectral remote sensing images is greatly affected by the problem of dimensionality. Feature extraction, as a common dimension reduction method, can make up for the deficiency of the classification of hyperspectral remote sensing images. However, different feature extraction methods and classification methods adapt to different conditions and lack comprehensive comparative analysis. Therefore, principal component analysis (PCA), linear discriminant analysis (LDA), and locality preserving projections (LPP) were selected to reduce the dimensionality of hyperspectral remote sensing images, and subsequently, support vector machine (SVM), random forest (RF), and the k-nearest neighbor (KNN) were used to classify the output images, respectively. In the experiment, two hyperspectral remote sensing data groups were used to evaluate the nine combination methods. The experimental results show that the classification effect of the combination method when applying principal component analysis and support vector machine is better than the other eight combination methods.

Keywords: hyperspectral remote sensing; image; classification; feature extraction

check for updates

Citation: Wang, Y.; Zeng, W. A Comparison of Three Different Group Intelligence Algorithms for Hyperspectral Imagery Classification. *Processes* **2022**, *10*, 1672. https://doi.org/10.3390/pr10091672

Academic Editors: Amir H. Gandomi and Laith Abualigah

Received: 7 July 2022
Accepted: 20 August 2022
Published: 23 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Hyperspectral remote sensing has high spectral resolution, and can obtain the spectral characteristics and differences of ground objects comprehensively and carefully, thus, greatly improving the accuracy of ground object classification [1]. Hyperspectral Images are highly innovative remote sensing imageries, and contain hundreds of continuous narrow spectral bands [2]. However, classifying HSIs efficiently is a major challenge for scientists and researchers [3]. Some of these challenges have been indicated, such as the increased presence of redundant spectral information and high dimensionality in observed data [4]. Several machine learning classifiers have been used for classifying HSIs. In recent years, deep-learning-based classifiers have been extensively studied in hyperspectral image classification [5–7]; they can achieve better classification effects, but the parameters involved are complicated. Unlike deep-learning-based classifiers, the traditional unsupervised machine learning classifiers include fuzzy C-means (FCM) and K-means (KM). Alternatively, the supervised classifiers, (e.g., k-nearest neighbor (KNN), Gaussian mixture model (GMM), support vector machine (SVM), random forest (RF), and artificial neural network (ANN)) have been widely used in the classification of HSIs [8–10]. The single classifier is simple to implement and suitable for classifying data with small samples and high dimensional features. However, both theory and practice show that no classifier is superior to others in nature due to the characteristics of hyperspectral remote sensing data, training samples, and the classifier itself [11,12].

Because hyperspectral remote sensing has many bands, a strong correlation between adjacent bands, and a large amount of data, it is easy to cause problems such as "dimension disaster" [13], which has a great influence on ground object classification; therefore, before classification, hyperspectral remote sensing images are often reduced to retain the original image information to the maximum extent and facilitate better understanding, analysis,

and processing of hyperspectral data [12]. One method of dimensionality reduction is band selection, which selects a band subset of the original image according to certain metric criteria or methods. Although this method can select specific bands with key functions, it is easy to ignore important information about other bands. The other method is feature extraction, which makes the original image data achieve the optimal feature in a certain sense. Feature extraction can be divided into linear and nonlinear feature extraction methods [14]. Common linear feature extraction methods include: principal component analysis (PCA) [15], linear discriminant analysis (LDA) [16], and locality preserving projections (LPP) [17], etc. These methods can successfully preserve the spectral characteristic information of local objects, and are simple to implement and fast to calculate. Nonlinear feature extraction methods include: kernel principal component analysis (KPCA) [18], kernel independent component analysis (KICA) [19], local linear embedding (LLE) [20], Laplacian eigenmaps (LE) [21], etc. However, which can better represent the structure of hyperspectral data is uncertain [22], and although homotopy disruption strategy (HPM) can be utilized to examine the logical surmised answer for the nonlinear control issue [23], their implementation is relatively complex. Selecting appropriate feature extraction methods can improve the processing speed and reduce the time of extracting valuable information. Therefore, many researchers have studied the comparison of different feature extraction methods [24–26], which provide references for the classification of different types of hyperspectral data.

In the application of hyperspectral remote sensing, the simplest classification method is usually adopted to ensure the accuracy of the classification and improve the efficiency of operation. In addition, the combination of different feature extraction methods and classifiers has different effects on hyperspectral image classification. However, the current research mainly compares the classification effects of different feature extraction methods or classifiers alone, which is not conducive to the further application of classification methods in the hyperspectral field. Therefore, the present comprehensive analysis of the current research on feature extraction adopted three of the most typical feature extraction methods and three different classifiers, respectively, selected two study areas of hyperspectral image datasets to design the classification experiment, and finally, compared the classification effects of different combination methods.

The research has two main advantages: (1) it discusses the advantages and disadvantages of different combination methods, and thus can provide a reference for the classification of hyperspectral remote sensing images; (2) on the basis of easy acquisition and simple operation, the applicability of different combination methods to different types of hyperspectral data is explored, which provides a reference for the practical application of hyperspectral remote sensing image classification and saves time in method selection.

2. Theoretical Methods

In this paper, the unsupervised feature extraction method PCA, and the supervised feature extraction methods LDA and LPP are selected from among common feature extraction methods to reduce the dimensionality of original hyperspectral images, and the common single classifiers SVM, RF, and KNN are selected to classify the images after dimensionality reduction.

2.1. Feature Extraction Method

2.1.1. Principal Component Analysis (PCA)

PCA is an unsupervised feature extraction method. Its main function is to reduce dimension by mapping the sample data from high-dimensional space to low-dimensional space through an orthogonal matrix. In the original space, the largest variance in the first set of axes represent the original data direction, the second set of axes represent the largest variance in the first set of the orthogonal plane coordinate system, the third set of axes represent the largest variance in the first and second group axis orthogonal plane, and so on. Therefore, most of the variance is retained in the first K coordinate axes, that is, the

K-dimensional space is reconstructed based on the original feature space, and this space contains most of the important information regarding the original space.

Let the original sample matrix be $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_n] \in \mathbf{R}^{m \times n}$, where m and n represent the characteristic dimension and sample number, respectively, and the sample mean is 0. Therefore, when $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = 0$, the low-dimensional matrix is obtained after mapping. The PCA algorithm tries to find a set of optimal orthogonal basis vectors to minimize the reconstruction error function:

$$\delta = \sum_{i=1}^{n} \|\boldsymbol{x}_i - \sum_{a=1}^{k} (\boldsymbol{\beta}_a^T \cdot \boldsymbol{x}_i)\boldsymbol{\beta}_a\|^2 \tag{1}$$

$\{\boldsymbol{\beta}_a \| a = 1, \ldots, k\}$ is the set of partial orthogonal basis vectors.

If the mapping matrix is $\boldsymbol{U} = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k\}$, $\boldsymbol{U} \in \mathbf{R}^{d \times k}$, then $\boldsymbol{y}_i = \boldsymbol{U}^T \boldsymbol{x}_i$, and thus $\boldsymbol{Y} = \boldsymbol{U}^T \boldsymbol{X}$, and if it satisfies the constraint conditions $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}$, then $\boldsymbol{I}$ is the identity matrix. Subsequently, the objective function can be expressed as:

$$\underset{\boldsymbol{U}}{\operatorname{argmin}} \sum_{i} \left\| \boldsymbol{x}_i - \boldsymbol{U}\left(\boldsymbol{U}^T \boldsymbol{x}_i\right) \right\|^2 \tag{2}$$

To solve the objective function is to solve the eigenvalue problem:

$$\boldsymbol{X}\boldsymbol{X}^T \boldsymbol{\beta}_i = \lambda_i \boldsymbol{\beta}_i \tag{3}$$

where λ_i is the eigenvalue, and the optimal orthogonal basis vector is composed of the k maximum eigenvalue pairs of the eigenvectors solved by the above formula, which can finally form the mapping matrix $\boldsymbol{U}$.

The PCA algorithm is the most commonly used dimension reduction method, and is suitable for the condition of the global linear low-dimensional structure, and has a good effect on linear data.

2.1.2. Linear Discriminant Analysis (LDA)

LDA is a feature extraction method of supervised learning. Its basic function is to project the sample data of high-dimensional space into low-dimensional space to meet the requirement of "the minimum variance within the class and the maximum variance between classes"; that is, the projection points of the data of the same category are as close as possible, and those of different classes are as far as possible.

In the LDA algorithm, the mapping matrix is set as $\boldsymbol{U}$ and satisfies Fisher's criterion function:

$$\arg\max_{\boldsymbol{U}} \frac{\operatorname{Tr}\left(\boldsymbol{U}^T \boldsymbol{S}_b \boldsymbol{U}\right)}{\operatorname{Tr}\left(\boldsymbol{U}^T \boldsymbol{S}_w \boldsymbol{U}\right)} \tag{4}$$

where $\boldsymbol{S}_b$ is the interclass dispersion matrix of samples, $\boldsymbol{S}_w$ is the in-class dispersion matrix of the sample. Assuming that the number of categories is C, n_i is the class i sample, $\overline{\boldsymbol{x}}_i$ and $\overline{\boldsymbol{x}}$ are the mean of the class i sample and the population mean of the sample, respectively. Considering that $\boldsymbol{S}$ is the sample population discrete matrix, then $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$ can be defined as:

$$\boldsymbol{S}_b = \frac{1}{n}\sum_{i=1}^{C} n_i (\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})^T \tag{5}$$

$$\boldsymbol{S} = \frac{1}{n}\sum_{j=1}^{n} \left(\boldsymbol{x}_j - \overline{\boldsymbol{x}}\right)\left(\boldsymbol{x}_j - \overline{\boldsymbol{x}}\right)^T \tag{6}$$

$$\begin{aligned} \boldsymbol{S}_w &= \boldsymbol{S} - \boldsymbol{S}_b \\ &= \tfrac{1}{n}\sum_{j=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_j{}^T - \tfrac{1}{n}\sum_{i=1}^{C} n_i \overline{\boldsymbol{x}}_i \overline{\boldsymbol{x}}_i{}^T \end{aligned} \tag{7}$$

Solving the optimal mapping matrix $\boldsymbol{U}$ is equivalent to solving the generalized eigenvalue problem: $\boldsymbol{S_b U}_i = \lambda_i \boldsymbol{S_w U}_i$, as the rank of $\boldsymbol{S_b}$ is, at most, $C - 1$; therefore, the maximum dimension of the space after LDA mapping is also $C - 1$.

The LDA algorithm assumes that data conform to Gaussian distribution, so it is not suitable for dimensionality reduction in non-Gaussian distribution samples. As LDA information is measured by mean size, the effect of dimension reduction is poor when sample classification information depends on the variance rather than the mean.

2.1.3. Locality Preserving Projections (LPP)

LPP mainly constructs a graph containing neighborhood information of a dataset comprising high-dimensional data, and then calculates a transformation matrix using the concept of the Laplacian operator [15] to map data points onto a low-dimensional subspace. This linear transformation preserves local neighborhood information well.

The LPP algorithm assumes that the two sample points, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, which are very close to each other in the original space, are also very close to the corresponding points, $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$, after being projected into the low-dimensional space. Its objective function is:

$$\min \sum_{ij} \left(\boldsymbol{y}_i - \boldsymbol{y}_j \right)^2 \boldsymbol{W}_{ij} \tag{8}$$

where $\boldsymbol{W}_{ij}$ represents weights. There are two ways to construct these weights.

The first method is a thermonuclear method that utilizes the Euclidean distance between samples to determine the corresponding weight, i.e., the closer the distance, the greater the weight, and vice versa. By introducing thermonuclear parameter t, the weight can be expressed as:

$$\boldsymbol{W}_{ij} = e^{-\frac{\|\boldsymbol{x}_i \boldsymbol{x}_j\|^2}{t}} \tag{9}$$

The second method is relatively simple. As long as the two points are adjacent, the weight between samples is set as 1; however, this method cannot distinguish the affinity between sample points effectively. The objective function can be further derived as:

$$\begin{aligned} &= \min \sum_{ij} \left(\boldsymbol{U}^T \boldsymbol{x}_i - \boldsymbol{U}^T \boldsymbol{x}_j \right)^2 \boldsymbol{W}_{ij} \\ &= \min \sum_{ij} \boldsymbol{U}^T \boldsymbol{x}_i \boldsymbol{D}_{ii} \boldsymbol{x}_i{}^T \boldsymbol{U} - \sum_{ij} \boldsymbol{U}^T \boldsymbol{x}_i \boldsymbol{D}_{ij} \boldsymbol{x}_j{}^T \boldsymbol{U} \\ &= \min \boldsymbol{u}^T \boldsymbol{X} (\boldsymbol{D} - \boldsymbol{W}) \boldsymbol{X}^T \boldsymbol{u} \\ &= \min \boldsymbol{u}^T \boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \boldsymbol{u} \end{aligned} \tag{10}$$

where $\boldsymbol{L}$ is the Laplace matrix, $\boldsymbol{D}_{ii} = \sum_j \boldsymbol{W}_{ij}$. Since the larger $\boldsymbol{D}_{ii}$ is, the more important $\boldsymbol{y}_i$ is, the formula $\boldsymbol{U}^T \boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \boldsymbol{U} = 1$ is introduced. Finally, the objective function can be transformed into an eigenvalue to solve the problem:

$$\boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \boldsymbol{U} = \lambda \boldsymbol{X} \boldsymbol{D} \boldsymbol{X}^T \boldsymbol{U} \tag{11}$$

The LPP algorithm is suitable for processing nonlinear sample data because LPP can maintain the nonlinear relationship after dimensionality reduction.

2.2. Classification Methods

The most representative machine learning classification methods in hyperspectral remote sensing image classification mainly include SVM, RF, and KNN.

2.2.1. Support Vector Machine (SVM)

The SVM is a supervised learning algorithm widely used in two or more linear and nonlinear classifications. The purpose of the SVM algorithm is to try to find an optimal hyperplane and ensure the maximum distance between various sample points and this plane.

For linear problems, any hyperplane can be expressed by the linear equation shown below:

$$\boldsymbol{w}^T\boldsymbol{x} + b = 0 \tag{12}$$

where $\boldsymbol{w}$ is the weight vector and b is bias. In higher dimensional space, the distance from the sample point to the hyperplane is:

$$\frac{\boldsymbol{w}^T\boldsymbol{x} + b}{\|\boldsymbol{w}\|} \tag{13}$$

To maximize the distance between the sample point closest to the hyperplane and the hyperplane, this distance can be transformed into a minimization function $\boldsymbol{L}(\boldsymbol{w})$ under additional constraints, which implies that the hyperplane correctly classifies all training samples, $\boldsymbol{x}_i$, as:

$$\min\frac{1}{2}\|\boldsymbol{w}\|^2, \text{ s.t. } \boldsymbol{y}_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1 \tag{14}$$

This is a Lagrangian optimization problem. The weight vector $\boldsymbol{w}$ and bias b of the optimal hyperplane can be obtained by Lagrangian multiplication.

For nonlinear problems, linearly separable support vector machines cannot solve them well, so nonlinear transformation is used to transform nonlinear problems into linear problems. In addition, $\Phi(\boldsymbol{x})$ represents the feature vector after the original data is mapped, and the hyperplane can be expressed as:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\Phi(\boldsymbol{x}) + b \tag{15}$$

therefore, we have the minimization function:

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2, \text{ s.t. } \boldsymbol{y}_i\left(\boldsymbol{w}^T\Phi(\boldsymbol{x}_i) + b\right) \geq 1 \quad (i = 1, 2, \ldots, m) \tag{16}$$

The SVM can model linear and nonlinear problems based on the kernel, but it is not suitable for large and/or noisy datasets.

2.2.2. Random Forest (RF)

RF is a classifier model composed of multiple decision trees, and the final output of the model is jointly determined by every moment in the decision tree in the forest. RF can deal with regression problems and classification problems. When dealing with classification problems, each decision tree will randomly select training samples and distinguish categories. Finally, the output categories of each decision tree will be considered comprehensively to determine the category of test samples by voting. The main steps to build an RF are:

(1) Extract k training subsets from the original training set, corresponding to k decision trees, respectively.
(2) The growth of each decision tree includes two processes. First, random feature variables are selected, and n features ($n \leq N$) are randomly selected at each node of each tree. The other is node splitting. The information contained in each feature is calculated, and the feature with the best classification ability is selected among n features for node splitting.
(3) Generate a random forest, do not prune each tree to maximize its growth, and finally, all decision trees constitute a random forest.

(4) After the random forest is constructed, the samples are input into the classifier. Each decision tree predicts the corresponding category for each sample, and records it by voting. The category with the most votes becomes the determined category of the sample.

2.2.3. K-Nearest Neighbor (KNN)

KNN is a classification algorithm. It sets the number of samples of each sample and its nearest neighbor. If most of these nearest neighbor samples belong to a certain category, it identifies that the sample also belongs to the same category. According to the distance between different characteristic values, data separation is generally carried out using Euclidean distance:

$$\boldsymbol{L} = \left(\sum_{l=1}^{N} \left| \boldsymbol{x}_i^{(l)} - \boldsymbol{x}_j^{(l)} \right|^p \right)^{\frac{1}{p}} \tag{17}$$

where p is a variable parameter. When p is 1, $\boldsymbol{L}$ is the Manhattan distance (corresponding to L1 norm), when p is 2, $\boldsymbol{L}$ is the Euclidean distance (corresponding to L2 norm), and when p tends to infinity, $\boldsymbol{L}$ is the Chebyshev distance, namely, the maximum distance of each coordinate axis. Additionally, l represents the vector dimension of the sample, and i and j represent the training sample vector of the ith and jth input, respectively.

As for the selection of the nearest neighbor sample number k value, when the k value is small, the existing training set can be well predicted, but the overall model becomes complex and prone to the overfitting phenomenon. When the k value is very large, the test error of the test set can be reduced, but the overall model becomes simpler, and the approximate error will increase. Therefore, in application, the value of k is generally selected as a small value, and the optimal value is usually found via a suboptimal verification method.

3. Data and Implementation

In this paper, hyperspectral datasets of two regions are selected for experimentation. The first hyperspectral dataset is for the Yellow River Estuary Experimental Zone, Dongying City, Shandong Province, China. The hyperspectral remote sensing images were acquired by the AHSI sensor on China's Gaofen-5 satellite in 2018, covering 330 bands from visible to shortwave infrared region (0.39–2.51 μm) with a spatial resolution of 30 m. After eliminating the substandard bands, image data of 285 bands were used for the experiment. The size of the experiment area was 721 pixels × 676 pixels, including 17 types of ground objects, such as the *Suaeda salsa*, pond, and floodplain. To obtain sufficient training samples, eight types of ground objects were removed, and the remaining nine types were used for experimental analysis. During the experiment, 10 samples in each category were selected for training. The distribution of the false-color image map and ground sample data for this area is shown in Figure 1, and the sample distribution is shown in Table 1. The second hyperspectral dataset comprises the Pavia University data (PaviaU for short) from the University of Pavia, Italy, acquired by the airborne reflective optical spectral imager ROSIS-03 in Germany in 2003. The spectral imager acquired 115 continuous band images in the wavelength range of 0.43 to 0.86 μm with a spatial resolution of 1.3 m. The bands affected by noise were removed, and the remaining 103 spectral bands were retained. The size of the area was 610 pixels × 340 pixels, including nine types of ground objects, such as trees, asphalt roads, bricks, meadows, etc. During the experiment, 5% of all samples were selected as training samples, and the rest as test samples. The distribution of the false-color image map and ground sample data in this area is shown in Figure 1, and the sample distribution is shown in Table 1.

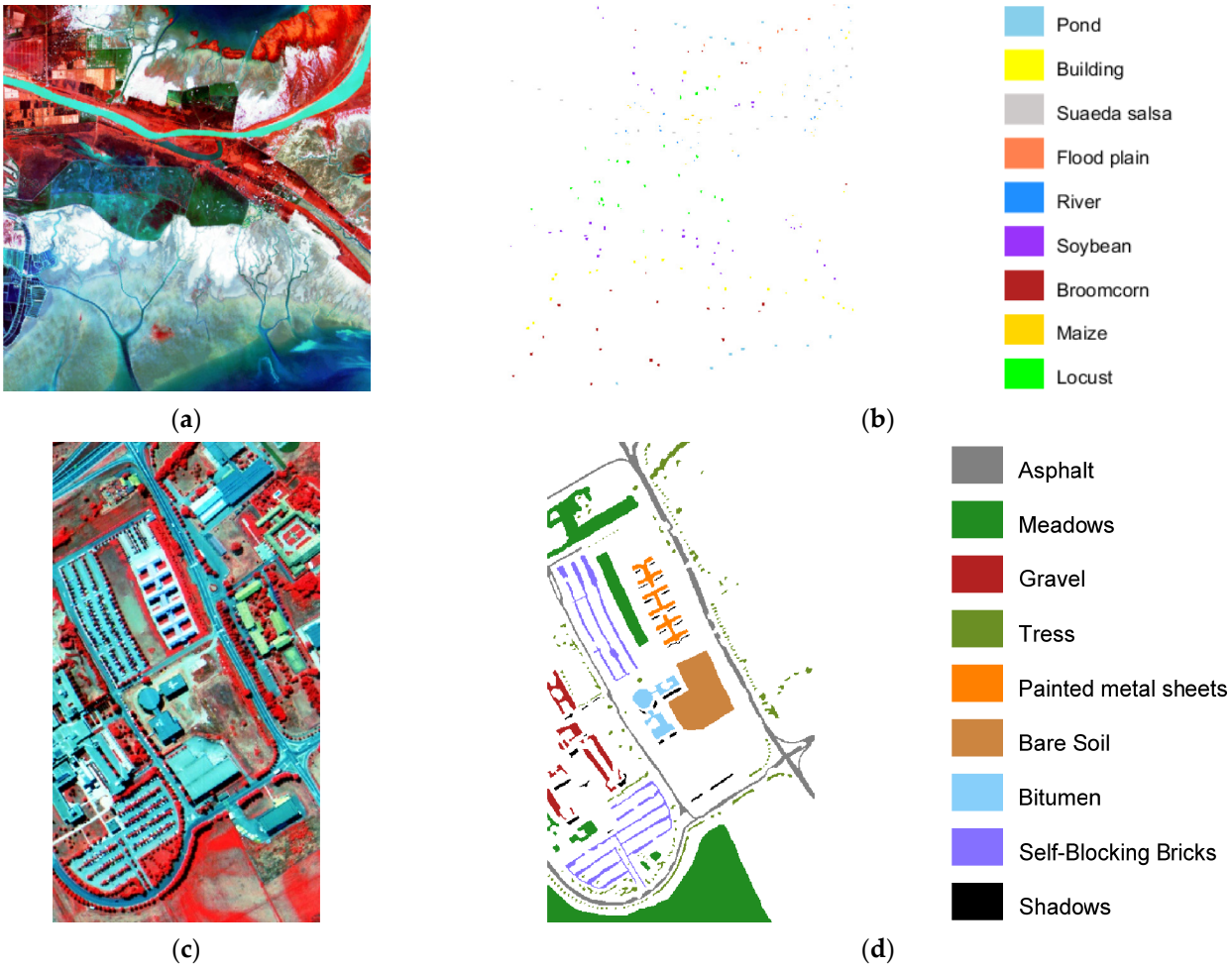

Figure 1. The distribution of false-color image maps and ground sample data: (**a**) false-color image for Yellow River Estuary data; (**b**) ground truth distribution map for Yellow River Estuary data; (**c**) false-color image map for Pavia University data; (**d**) ground truth distribution map for Pavia University data.

Table 1. Training sample of datasets.

No.	Yellow River Estuary Data			PaviaU Data		
	Classes	Training Sample	Test Sample	Class	Training Sample	Test Sample
1	Pond	10	300	Asphalt	332	6299
2	Building	10	406	Meadows	932	17,717
3	*Suaeda salsa*	10	255	Gravel	105	1994
4	Flood plain	10	95	Tress	153	2911
5	River	10	162	Painted metal sheets	67	1278
6	Soybean	10	538	Bare Soil	251	4778
7	Broomcorn	10	369	Bitumen	67	1263
8	Maize	10	123	Self-Blocking Bricks	184	3498
9	Locust	10	367	Shadows	47	900

In this experiment, principal component analysis (PCA), linear discriminant analysis (LDA), and local reserved projection (LPP) algorithms were used to extract features from hyperspectral remote sensing images, and then, support vector machine (SVM), random forest (RF), and k-nearest neighbor (KNN) classifiers were used to classify the feature images after dimensionality reduction. It is worth noting that before the experiment, the research data were preprocessed, and the preprocessing method involved data normalization (min–max

normalized data). The technical route is shown in Figure 2. In the experiment, when the PCA method was used for feature extraction, the classification accuracy was stable after dimensionality reduction to 30 dimensions [27]; thus, the data with dimensionality reductions to 30 dimensions were selected for accuracy evaluation. When LDA and LPP methods were used for feature extraction, the maximum dimension was reduced to $C - 1$ (C is the number of categories). Related parameters in the classifier, such as the penalty parameter c and kernel function parameter g in SVM, the number of decision trees in RF, and the nearest neighbor value in KNN, were determined by grid searches and tenfold cross-validation. The sample set selected by the decision number of the random forest to discriminate classification is random. Therefore, the mean value of classification accuracy is taken as the accuracy evaluation index after 10 runs. The classification performance evaluation indexes included overall accuracy (OA), average accuracy (AA), and the Kappa coefficient. At the same time, the average running time of the feature extraction and classification algorithm were recorded five times, and the operation efficiencies of the different algorithm combinations were calculated.

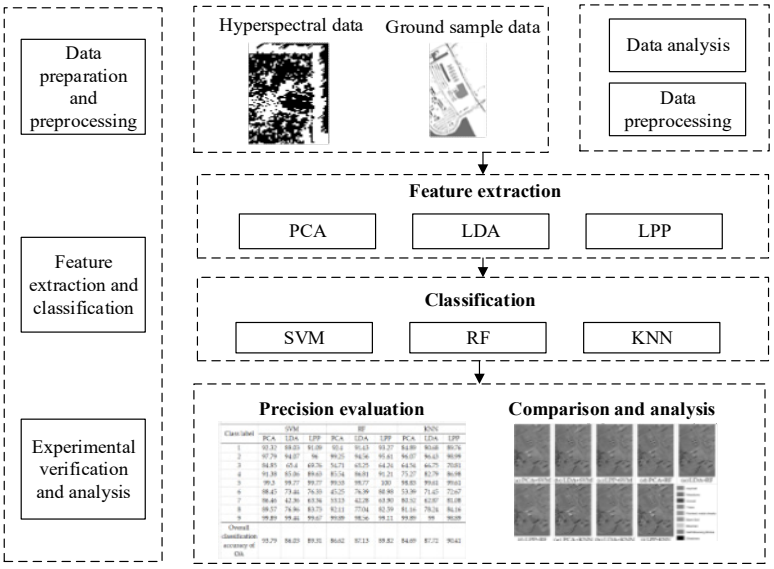

Figure 2. Technical route.

4. Results and Discussion

Table 2 shows the experimental results of the combination of three feature extraction methods and three classification methods for the Yellow River Estuary dataset. The classification accuracy from high to low in terms of method combination is PCA+SVM, PCA+RF, LDA+SVM/KNN, LPP+SVM/KNN, PCA+KNN, LDA+RF, and LPP+RF. The highest classification accuracy was PCA+SVM, for which OA was 94.68%, and the Kappa coefficient was 0.9385, followed by PCA+RF. The classification accuracy of LDA+RF and LPP+RF was poor. Compared with the classification results of PCA+SVM, OA differed by 4.5% and 4.56%, and the Kappa coefficient differed by 5.2% and 5.24%, respectively. The PCA feature extraction method not only retains the initial sample information to the maximum extent, and contains the most important information, it also removes image noise. Therefore, after the PCA algorithm was used to reduce the dimension, three kinds of classifiers were used for classification to test for high classification accuracy. Compared with PCA, the dimensionality reduction effect of LPP and LDA for Yellow River Estuary data was average. The possible reason for this is that Yellow River Estuary data meet the condition of global linearity, while the LPP algorithm is suitable for processing nonlinear sample data, thus impacting the dimensionality reduction effect. LDA considers the influence of categories in the dimensionality reduction process, and ensures that the sample sets of different classes have a large interval after dimensionality reduction. However, if

there is no significant difference between the mean values of the two types of sample sets, and the covariances are greatly different, the dimensionality reduction effect will have no obvious advantage. Because of the small number of training samples in the Yellow River Estuary dataset, SVM can achieve a better classification effect than other methods on the small sample training set. In addition, SVM represents a convex optimization problem, which can find the global minimum of the objective function rather than obtain the local optimal solution, thus, the classification accuracy is high. However, KNN is suitable for the classification of large samples, and it is easy to misclassify the Yellow River Estuary dataset with a small sample size; therefore, the classification effect had no obvious advantage. In terms of feature extraction and classification method combination for classification accuracy, we found PCA and SVM to be superior in feature extraction and classification, thus achieving the optimal classification accuracy. Furthermore, RF noise due to larger data classification easily produced the phenomenon of fitting, and after using the PCA process to obtain a better classification effect, the effects of LDA and LPP treatment were poor.

Table 2. Classification accuracy of Yellow River Estuary dataset (%).

Class Label	SVM			RF			KNN		
	PCA	LDA	LPP	PCA	LDA	LPP	PCA	LDA	LPP
1	81.67	99.67	100	95.67	100.00	98	74.33	99.67	100
2	100	100	100	100	97.04	93.6	99.75	100	100
3	100	100	100	100	98.82	99.61	100	100.	100
4	100	94.74	93.68	100	85.26	85.26	100	94.74	93.68
5	95.68	83.33	81.48	89.51	66.05	77.16	92.59	83.33	81.48
6	100	98.7	97.4	99.44	93.68	88.66	100	98.7	97.4
7	99.73	99.19	98.64	99.46	96.48	98.37	100	99.19	98.64
8	83.74	95.12	94.31	91.87	80.49	95.93	90.24	95.12	94.31
9	84.74	74.11	75.2	76.02	74.66	71.93	81.2	74.11	75.2
Overall classification accuracy of OA	94.68	94.49	94.15	94.66	90.18	90.12	93.46	94.49	94.15
Average classification accuracy of AA	94.12	93.66	93.03	94.18	88.02	87.57	92.99	93.66	93.03
Kappa coefficient	93.85	93.63	93.24	93.83	88.65	88.61	92.43	93.63	93.24

To evaluate the classification effect more directly, Figure 3 shows the land type classification diagram produced by the different methods. The experimental results show that PCA+SVM, PCA+RF, and PCA+KNN achieved better classification results, and the classification results of buildings, *Suaeda salsa*, and floodplain were clear and accurate. The possible reason for this is that PCA extracted the main information of ground objects, including buildings, *Suaeda salsa*, and floodplain. Therefore, the classification effect after PCA treatment was generally good. However, after feature extraction by LDA and LPP and classification by the three classifiers, the ground features were not smooth enough, especially in rivers and buildings covered with parts of corn ground features. The possible reason is that there is no significant difference between the mean values of rivers and maize, or buildings and maize, which affects the effect of dimension reduction and finally leads to the misclassification of maize as buildings and rivers. However, due to the obvious difference between the mean values of ponds and *Acacia*, the final differentiation degree was high. In addition, after the same feature extraction method was used, the overall classification effect of RF was poor, which may be because the RF classifier was greatly affected by noise, which affects the classification effect.

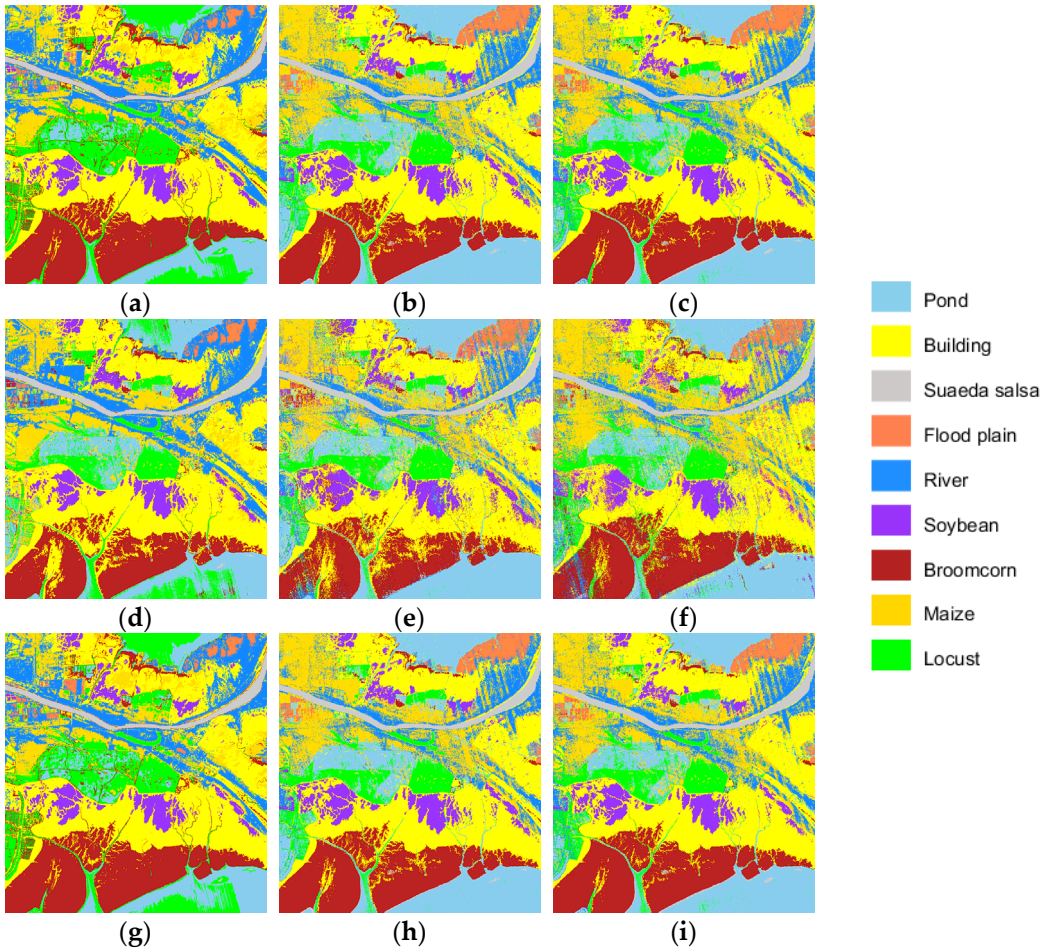

Figure 3. Classification maps of land types for different combination algorithms of Yellow River Estuary: (**a**) PCA+SVM; (**b**) LDA+SVM; (**c**) LPP+SVM; (**d**) PCA+RF; (**e**) LDA+RF; (**f**) LPP+RF; (**g**) PCA+KNN; (**h**) LDA+KNN; (**i**) LPP+KNN.

The classification accuracy of The University of Pavia is shown in Table 3, where the classification accuracy from high to low in terms of method combination is PCA+SVM, LPP+KNN, LPP+RF, LPP+SVM, LDA+KNN, LDA+RF, PCA+RF, LDA+SVM, PCA+KNN. Among them, PCA+SVM had the highest classification accuracy, while PCA+KNN had the worst classification accuracy, with a 9.1% difference in OA, 7.9% difference in AA, and 12.52% difference in the Kappa coefficient. For the University of Pavia dataset, the classification accuracy of PCA+SVM was still the best, which is because PCA effectively reduces the image noise, and SVM grasps the key samples in classification, thus the perception of the outliers is unknown, and the classification effect is superior. In addition, after processing by the LPP method, the classification results are all good, which may be due to a large number of training samples in the University of Pavia dataset. This method avoids the divergence of the sample set, and effectively retains the local neighborhood structure of the University of Pavia dataset. Because the LDA method considers the type of samples, the dimension projected into low-dimensional space is limited, and there may be an overfitting phenomenon when LDA is used to process the dataset of the University of Pavia, resulting in a generally poor final classification effect. PCA handling with KNN classification was the worst; this is because the PCA has to extract the principal component information, and sample dimension reduction causes imbalance, meanwhile, in KNN calculation, samples of nearest neighbor are only achieved if of the sample size is large. Thus, this type of sample may either be too far from the target sample, or too close to the target sample, leading to poor classification results.

Table 3. Classification accuracy of Pavia University data (%).

Class Label	SVM			RF			KNN		
	PCA	**LDA**	**LPP**	**PCA**	**LDA**	**LPP**	**PCA**	**LDA**	**LPP**
1	92.32	89.03	91.09	92.4	91.43	93.27	84.89	90.68	89.76
2	97.79	94.07	96	99.25	94.56	95.61	96.07	96.43	98.99
3	84.95	65.4	69.76	54.71	65.25	64.24	64.54	66.75	70.81
4	91.38	85.06	89.63	85.54	86.81	91.21	75.27	82.79	86.98
5	99.3	99.77	99.77	99.53	99.77	100	98.83	99.61	99.61
6	88.45	73.44	76.33	45.25	76.39	80.98	53.39	71.45	72.67
7	86.46	42.36	63.34	53.13	42.28	63.90	80.52	62.87	81.08
8	89.57	76.96	83.73	92.11	77.04	82.59	81.16	78.24	84.16
9	99.89	99.44	99.67	99.89	98.56	99.11	99.89	99	98.89
Overall classification accuracy of OA	93.79	86.03	89.31	86.62	87.13	89.82	84.69	87.72	90.41
Average classification accuracy of AA	92.84	83.94	88.18	91.18	86.13	89.24	84.94	87.02	89.88
Kappa coefficient	91.74	81.34	85.72	81.64	82.82	86.46	79.22	83.49	87.1

According to the land type classification map of The University of Pavia in Figure 4, it can be seen that the combination of PCA and SVM had the best classification effect, followed by LPP+KNN and LPP+RF, which can accurately classify grassland, metal plate, and shadow. This is because the University of Pavia is primarily urban terrain, with a distinct terrain shape, (e.g., rectangles, and arcs are easy to identify), and has obvious differences between plot properties. This is advantageous to the LPP in terms of dimension reduction, and better preserves the local field information of the samples. Moreover, the classification results of the three classifiers were also good. KNN classification relies mainly on the surrounding adjacent samples. Therefore, compared with other classifiers, the classification effect of KNN is better for data with local structures. In addition, the classification effects of LDA+RF, PCA+RF, LDA+SVM, and PCA+KNN were not good, and the classification error rate of gravel and asphalt roof was high. The possible reason for this is that there was a strong correlation between any two decision trees of the RF, which led to the occurrence of repeated information in the classification, especially for sand and asphalt. Although LDA can retain local information of samples, the extracted edge information does not have good consistency with the boundary of ground object distribution in some categories, and the classification accuracy of SVM is quite different [28] (for example, the classification accuracy of asphalt is only 42.36%, whereas sand is 65.4%, and a metal plate is 99.77%). The result of the final classification effect is not good. PCA is not able to distinguish samples of different classes. When the KNN classifier was used for classification after dimensionality reduction, due to the large value of the selected nearest neighbor sample number k, the model was not fitted enough, and its ability to recognize differentiated ground objects was reduced. On the whole, PCA+SVM had the best classification effect, while PCA+RF, LDA+SVM, and PCA+KNN had poor classification effects.

To further compare the time complexity of different algorithms, Table 4 presents the operation times of the experimental data under different combination algorithms (the average of five run times). It can be seen that when the amount of experimental data was small, the computation speeds of SVM and KNN were faster, and the computation time of RF was the longest. When the amount of experimental data was large, the computation time of KNN was the shortest, and the computation speed of RF was the slowest. The computation speed of RF classification after PCA processing is 61 times that of KNN. This is because the larger the number of RF decision trees, the longer each decision tree will participate in classification, and the slower the operation speed will be.

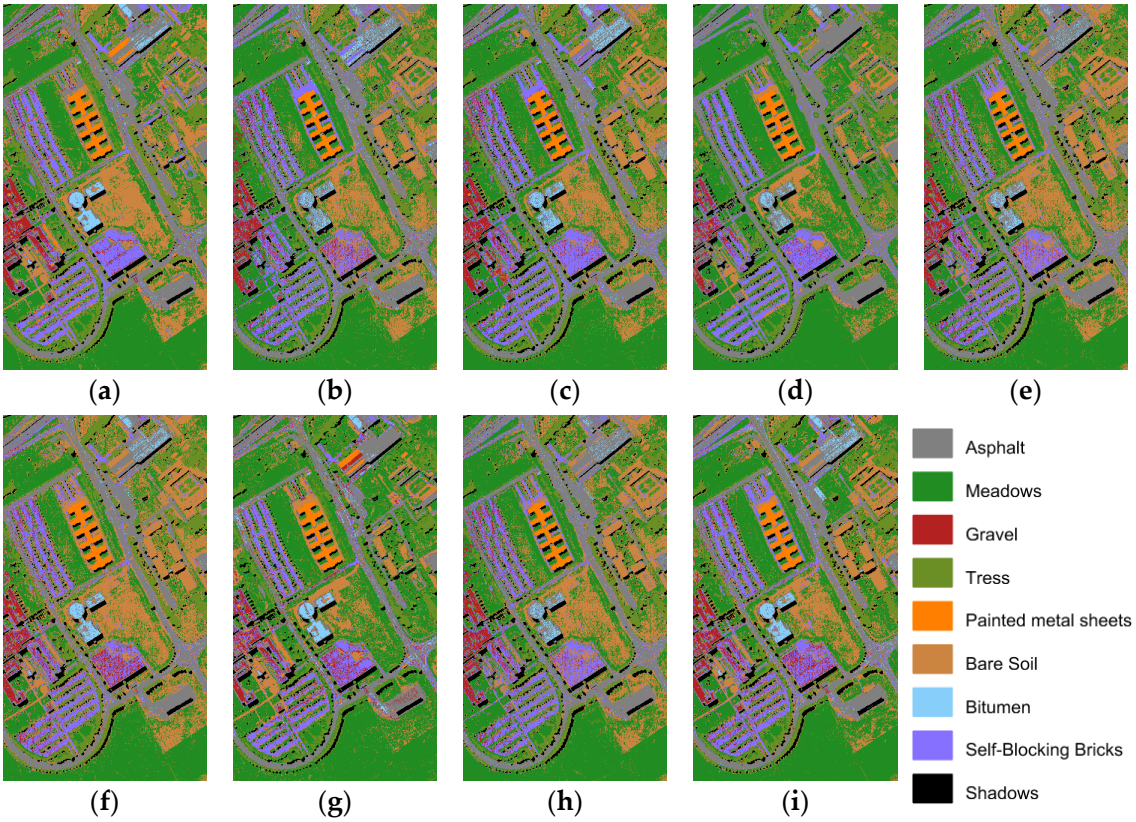

Figure 4. Classification maps of land types for different combination algorithms of Pavia University data. (**a**) PCA+SVM; (**b**) LDA+SVM; (**c**) LPP+SVM; (**d**) PCA+RF; (**e**) LDA+RF; (**f**) LPP+RF; (**g**) PCA+KNN; (**h**) LDA+KNN; (**i**) LPP+KNN.

Table 4. Computational time of different methods (s).

Dataset	SVM			RF			KNN		
	PCA	LDA	LPP	PCA	LDA	LPP	PCA	LDA	LPP
Yellow River Estuary	0.08	0.03	0.01	3.13	1.65	2.73	0.06	0.01	0.01
University of Pavia	2.15	0.85	1.26	29.23	11.36	18.41	0.48	0.03	0.3

5. Conclusions

In this paper, several feature extraction and classification methods are combined and applied for hyperspectral remote sensing image classification. Through comparative analysis of experimental results, the following conclusions are drawn: (1) For the feature extraction method, PCA can extract most of the important information from the original data, and the visual effect and classification accuracy are good after classification via the SVM classifier, and the calculation speed is fast. The combination of PCA and SVM is an effective method for hyperspectral remote sensing image classification. (2) For datasets with a large number of training samples, LPP can achieve a better effect for dimensionality reduction, and there is little difference in effect of classification after using different classifiers. (3) For datasets with a small amount of data, the classification effect of PCA+RF is better. For large datasets, LPP+KNN and LPP+RF can achieve better classification.

In this paper, several common hyperspectral remote sensing image feature extraction methods and classifiers are preliminarily compared. Future research work will focus on proposing the best method for processing the images. At the same time, we will compare more feature extraction and classification methods, and apply them to hyperspectral images with a large number of samples. In this process, the applicability of different combination

methods, optimization of dimensionality reduction methods, and classification results will also be discussed.

Author Contributions: Funding acquisition, project administration, writing—original draft: Y.W.; project administration, data curation, writing—original draft, methodology, formal analysis: W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grant from Chuzhou University (The research of funding name is 2022/2024 and funding number is No. 2022qd008).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Pavia University data from the experiment can be downloaded from the website (https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes). The Yellow River Estuary data in the experiment is not convenient to be disclosed because it involves confidentiality.

Acknowledgments: We acknowledge Shuying Zang's supervision and discussion. We would like to thank Huiqiao Sui (Nanjing Normal University) for the English and grammar corrections. We would also like to thank Mengyu Gu (Hohai University) for help with the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tong, Q.X.; Zhang, B.; Zhang, L.F. Advances in hyperspectral remote sensing in China. *J. Remote Sens.* **2016**, *20*, 19.
2. Zhang, M.; Li, W.; Du, Q. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 2623–2634. [CrossRef] [PubMed]
3. Li, P.; Wang, D.; Wang, L.; Lu, H. Deep visual tracking: Review and experimental comparison. *Pattern Recognit.* **2018**, *76*, 323–338. [CrossRef]
4. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A. Advanced supervised spectral classifiers for hyperspectral images: A review. *J. Latex Cl. Files* **2007**, *6*, 1–23.
5. Yang, X.; Ye, Y.; Li, X.; Lau RY, K.; Zhang, X.; Huang, X. Hyperspec-tral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5408–5423. [CrossRef]
6. Yin, X.; Wang, R.; Liu, X.; Cai, Y. Deep forest-based classification of hyperspectral images. *Proc. Chin. Control Conf.* **2018**, *2018*, 10367–10372.
7. Yu, D.; Ma, Z.; Wang, R. Efficient smart grid load balancing via fog and cloud computing. *Math. Probl. Eng.* **2022**, *22*, 3151249. [CrossRef]
8. Wang, X.; Feng, Y. New Method Based on Support Vector Machine in Classification for Hyperspectral Data. In Proceedings of the International Symposium on Computational Intelligence and Design, Wuhan, China, 17–18 October 2008; pp. 76–80.
9. Joelsson, S.R.; Benediktsson, J.A.; Sveinsson, J.R. Random Forest Classifiers for Hyperspectral Data. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005 IGARSS '05, Seoul, Korea, 29–29 July 2005.
10. Ma, L.; Crawford, M.M.; Tian, J. Local manifold learning-based K-nearest-neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [CrossRef]
11. Du, P.J.; Xia, J.S.; Xue, Z.H.; Tan, K.; Su, H.J.; Bao, R. Advances in classification of hyperspectral remote sensing images. *J. Remote Sens.* **2016**, *20*, 21.
12. Du, P.J.; Xia, J.S.; Zhang, W.; Tan, K.; Liu, Y.; Liu, S.C. Multiple classifier system for remote sensing image classification: A review. *Sensors* **2012**, *12*, 4764–4792. [CrossRef]
13. Hughes, G.F.; Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
14. Zhang, B. Frontier of hyperspectral image processing and information extraction. *J. Remote Sens.* **2016**, *20*, 1062–1089. [CrossRef]
15. Farrell, M.D.; Mersereau, R.M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 192–195. [CrossRef]
16. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear discriminant analysis: A detailed tutorial. *AI Commun.* **2017**, *30*, 169–190. [CrossRef]
17. He, X.; Niyogi, P. Locality Preserving Projections. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 153–160.
18. Schölkopf, B.; Smola, A.; Müller, K.R. Kernel Principal Component Analysis. In Proceedings of the 7th International Conference on Artificial Neural Networks—ICANN 1997, Lausanne, Switzerland, 8–10 October 1997; Springer-Verlag GmbH: Cham, Switzerland; pp. 583–588.
19. Bach, F.R.; Jordan, M.I. Kernel independent component analysis. *J. Mach. Learn. Res.* **2003**, *3*, 1–48.
20. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]

21. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [CrossRef]
22. Bachmann, C.M.; Ainsworth, T.L.; Fusina, R.A. Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 441–454. [CrossRef]
23. Gepreel, K.A.; Higazy, M.; Mahdy AM, S. Optimal control, signal flow graph, and system electronic circuit realization for nonlinear Anopheles mosquito model. *Int. J. Mod. Phys. C* **2020**, *31*, 2050130. [CrossRef]
24. Uddin, M.P.; Mamun, M.A.; Hossain, M.A. PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Technol. Rev.* **2021**, *38*, 377–396. [CrossRef]
25. Fabiyi, S.D.; Murray, P.; Zabalza, J.; Ren, J. Folded LDA: Extending the linear discriminant analysis algorithm for feature extraction and data reduction in hyperspectral remote sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 12312–12331. [CrossRef]
26. Ayesha, S.; Hanif, M.K.; Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf. Fusion* **2020**, *59*, 44–58. [CrossRef]
27. Su, H.J.; Gu, M.Y. Extraction of local alignment feature from hyperspectral remote sensing image based on optimization and discriminant. *J. Remote Sens.* **2021**, *25*, 16.
28. Shao, W.J.; Sun, W.W.; Yang, G. Comparative analysis of texture feature extraction from hyperspectral remote sensing images. *Remote Sens. Technol. Appl.* **2021**, *36*, 10.