


## Article

# Joint Sparse Local Linear Discriminant Analysis for Feature Dimensionality Reduction of Hyperspectral Images

Cong-Yin Cao <sup>1</sup>, Meng-Ting Li <sup>1</sup>, Yang-Jun Deng <sup>1,\*</sup> , Longfei Ren <sup>2</sup>, Yi Liu <sup>1</sup> and Xing-Hui Zhu <sup>1</sup>

<sup>1</sup> College of Information and Intelligence, Hunan Agricultural University, Changsha 410128, China; 19913809590@stu.hunau.edu.cn (C.-Y.C.); lmt@stu.hunau.edu.cn (M.-T.L.); yiliu@hunau.edu.cn (Y.L.); zhuxh@hunau.net (X.-H.Z.)

<sup>2</sup> Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; renlf@aircas.ac.cn

\* Correspondence: dyj2012@yeah.net

**Abstract:** Although linear discriminant analysis (LDA)-based subspace learning has been widely applied to hyperspectral image (HSI) classification, the existing LDA-based subspace learning methods exhibit several limitations: (1) They are often sensitive to noise and demonstrate weak robustness; (2) these methods ignore the local information inherent in data; and (3) the number of extracted features is restricted by the number of classes. To address these drawbacks, this paper proposes a novel joint sparse local linear discriminant analysis (JSL LDA) method by integrating embedding regression and locality-preserving regularization into the LDA model for feature dimensionality reduction of HSIs. In JSL LDA, a row-sparse projection matrix can be learned, to uncover the joint sparse structure information of data by imposing a  $L_{2,1}$ -norm constraint. The  $L_{2,1}$ -norm is also employed to measure the embedding regression reconstruction error, thereby mitigating the effects of noise and occlusions. A locality preservation term is incorporated to fully leverage the local geometric structural information of the data, enhancing the discriminability of the learned projection. Furthermore, an orthogonal matrix is introduced to alleviate the limitation on the number of acquired features. Finally, extensive experiments conducted on three hyperspectral image (HSI) datasets demonstrated that the performance of JSL LDA surpassed that of some related state-of-the-art dimensionality reduction methods.



**Citation:** Cao, C.-Y.; Li, M.-T.; Deng, Y.-J.; Ren, L.; Liu, Y.; Zhu, X.-H. Joint Sparse Local Linear Discriminant Analysis for Feature Dimensionality Reduction of Hyperspectral Images. *Remote Sens.* **2024**, *16*, 4287. <https://doi.org/10.3390/rs16224287>

Academic Editor: Prashant K Srivastava

Received: 20 October 2024

Accepted: 15 November 2024

Published: 17 November 2024



**Copyright:** © 2024 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/>).

**Keywords:** hyperspectral image (HSI); dimensionality reduction; linear discriminant analysis; embedding regression regularization

## 1. Introduction

With the rapid advancement of hyperspectral imaging technology, its applications have proliferated across various fields, including agricultural [1–3] and forestry management [4,5], urban planning [6–8], international relations [9], and resource exploration [10,11]. Compared to traditional remote sensing images, hyperspectral images (HSIs) offer superior spectral resolution and a wealth of spectral information, which facilitates the recognition of ground objects [12]. Nevertheless, the high spectral dimensionality of HSIs presents significant challenges for their processing and analysis.

Dimensionality reduction (DR) is an effective approach for addressing the challenges of high data redundancy and huge data volumes [13]. DR techniques can eliminate redundant information by transforming high-dimensional data into more discriminative low-dimensional subspaces, which has become an important step for the analysis of HSIs. Over the past decades, researchers have proposed a variety of DR methods, which can mainly be categorized into linear and nonlinear DR approaches. The most well-known linear DR methods are principal component analysis (PCA) [14] and linear discriminant analysis (LDA) [15], whereas Laplacian eigenmaps (LE) [16], isometric mapping (ISOMAP) [17], and locality-preserving projection (LPP) [18] represent the foremost methods for tackling nonlinear data. These methodologies

primarily focus on learning projections from distinct local geometric structures within the raw data. In contrast to PCA and LDA, they are capable of preserving the spatial local geometric structure inherent in the original data, yet they are not inherently suited for classification tasks, as the features they derive do not inherently possess discriminability [19].

LDA stands as one of the most frequently utilized methods for obtaining discriminative features in image classification tasks [20]. It projects the data into a low-dimensional space where samples from the same class are as close together as possible, while samples from different classes are positioned as far apart as possible. Various extended models based on LDA have also been proposed, such as orthogonal LDA (OLDA) [21] and discriminative local alignment (DLA) [22]. Acknowledging LDA's approximation error, Wang et al. [23] introduced trace ratio LDA (TRLDA) as an optimal LDA solution. They further proposed optimal dimensionality LDA (ODLDA) to identify the optimal subspace dimension. To overcome the limitation that LDA cannot correctly describe internal structure, Zhu et al. [24] defined a scatter matrix based on a neighborhood composed of reverse nearest neighbors. However, these LDA extension methods were all developed by utilizing  $L_2$ -norm as a metric criterion, which can make models sensitive to noise and outliers. To mitigate the impact of noise, robust feature extraction techniques have garnered significant attention across various fields. A commonly adopted approach to enhancing robustness involves integrating robust metrics to refine existing feature extraction models. For instance,  $L_1$ -norm was applied to mitigate the negative effects of noise, and  $L_1$ -norm PCA and  $L_1$ -norm-2DPCA were proposed for robust feature extraction [25,26]. Lu et al. [27] enhanced the robustness of two-dimensional locality-preserving projections (2DLPP) by introducing the nuclear norm. Zhang et al. [28] minimized reconstruction errors based on the nuclear norm and  $L_{2,1}$ -norm, and proposed two 2DNPP methods. Gu et al. [29] addressed the issue of the classical CSP based on the squared Frobenius norm being sensitive to noise by utilizing the  $L_{2,1}$ -norm, and proposed a regularized version of the common spatial pattern based on the  $L_{2,1}$ -norm (RCSP- $L_{2,1}$ ). Nie et al. [30] proposed a robust PCA method based on  $L_{2,1}$ -norm maximization, which addressed the processing challenges of high-dimensional data through non-greedy optimization algorithms, overcoming the sensitivity to outliers and the high computational complexity of traditional PCA. Zhang et al. [31] proposed a novel low-rank-preserving embedding regression (LRPER) approach, which employs the robust  $L_{2,1}$ -norm to measure the reconstruction error and regression loss, thereby effectively modeling noise and occlusions. Deng et al. [32] utilized the adaptive  $L_{2,p}$ -norm to model noise and residuals in a low-dimensional space, enhancing the robustness and generalization.

As the  $L_{2,1}$ -norm is an effective approach for obtaining joint sparse projections for discriminative feature selection or extraction, a plethora of joint sparse projection learning methods based on the  $L_{2,1}$ -norm have been developed. In [33], Nie et al. employed  $L_{2,1}$ -norm regularization to select features across all data points through joint sparsity. In an unsupervised learning environment, Yang et al. [34] introduced a discriminative  $L_{2,1}$ -norm minimization algorithm (UDFS) for effective feature selection without labels. Robust sparse linear discriminant analysis (RSLDA) [35] strategically incorporates the  $L_{2,1}$ -norm to dynamically select the most discriminative features for enhanced classification, thus optimizing the discriminative analysis process. Tang et al. [36] devised an unsupervised linear feature selective projection (FSP) method for feature extraction, by integrating low-rank embedding, dual Laplacian regularization, and  $L_{2,1}$ -norm minimization. Li et al. [37] constructed a new linear discriminant analysis method characterized by robustness and sparsity using the  $L_1$ -norm and  $L_{2,1}$ -norm, which can obtain all discriminant directions simultaneously. Long et al. [38] combined LPP with the  $L_{2,1}$ -norm to achieve simultaneous feature extraction and feature selection.

Although employing a sparse constraint allows selecting the most significant features for the purpose of feature extraction, it also has several inherent limitations. Firstly, these methods tend to focus primarily on the global structural information of data samples, often neglecting the intrinsic local geometric structure of the data. This local information is crucial for capturing subtle differences between data points and their local neighborhood

relationships. The neglect of local information may result in the reduced feature representations failing to fully preserve the complex structure of the original data, thereby affecting the accuracy of subsequent analysis tasks. Secondly, most of these methods lack robustness to noise. Additionally, while the low-dimensional feature spaces generated by current methods achieve effective data compression, they struggle to directly identify which features are more important or discriminative for specific target tasks. Furthermore, many existing methods emphasize enhancing the discriminability of features by designing regression loss functions that maximize the feature differences between different categories or samples. However, this process often considers projection learning and regression optimization in isolation, overlooking the potential interactions and synergies between these two stages.

In response to the aforementioned issues, we propose a novel robust dimensionality reduction method based on LDA, termed joint sparse local linear discriminant analysis (JSLLDA). The LPP-regularization term is introduced to maintain the neighborhood relationship of the original samples by constructing a nearest neighbor weighted graph, so as to learn the projection by simultaneously using the global information and local information of the data. To enhance the robustness of the model, the robust measure  $L_{2,1}$ -norm is employed to evaluate the regression loss. Since the  $L_{2,1}$ -norm does not involve squaring operations, it reduces the impact of outliers on the distance measurements. Unlike the  $L_1$ -norm, the  $L_{2,1}$ -norm also enforces row sparsity, which can make the error terms sparse and thereby fit the noise during the projection learning process. This enhances the robustness against outliers. In addition, JSLLDA also uses  $L_{2,1}$ -norm regularization to constrain the projection matrix. Unlike the  $L_1$ -norm, the  $L_{2,1}$ -norm enforces row sparsity; that is, more discriminative features are given higher weights, while redundant information and noise are given lower weights when obtaining low-dimensional features. This not only improves the robustness of the model, but also enhances the interpretability of the obtained low-dimensional features. Most importantly, the proposed method relates the projected data to the original label information through an orthogonal matrix; that is, the interaction between projection learning and regression is considered through a linear regression variant. The main contributions of this work are summarized as follows:

(1) To address the issues such as insufficient robustness, limitations in feature dimensions, and the absence of locality manifold structures that currently affect LDA-based subspace learning methods, a new robust dimensionality reduction method named JSLLDA is proposed by integrating embedding regression and locality-preserving regularization into the LDA model.

(2) In JSLLDA, the locality-preserving regularization can capture the locality manifold geometric structural information, while the  $L_{2,1}$ -norm imposed on the projection matrix can reveal the joint sparse information of the data, significantly enhancing the robustness against noise. Moreover, the embedding regression term can overcome the limitations in the feature dimensions of LDA and improve the discriminability of low-dimensional features by fully using the prior information.

(3) An alternating iteration algorithm was developed to improve the JSLLDA model, and its computational complexity was analyzed both theoretically and numerically. Experiments were conducted on three public HSI datasets, and the associated results verified that the proposed method achieved a better classification performance than several state-of-the-art methods. Specifically, the robustness analysis and ablation study of the proposed JSLLDA method are discussed in detail.

## 2. Methodology

In this section, some notations commonly used in this study are first introduced. Then, we review related foundational works, which include linear regression (LR) and linear discriminant analysis (LDA). Finally, the formulation and optimization of the proposed method are presented.

### 2.1. Notations and Definitions

$X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$  represents the original high-dimensional data with  $n$  samples, and  $Y \in \mathbb{R}^{c \times n}$  is the label information, where  $c$  denotes the number of classes. For a square matrix  $M$ , its  $i$ th row data are denoted  $m_{i,:}$ , and its  $i$ th row and  $j$ th column data are denoted  $m_{ij}$ . The trace of matrix  $M$  is written as  $Tr(M)$ . The transposition of matrix  $M$  is denoted by  $M^T$ . The  $L_{2,1}$ -norm is defined as follows:

$$\|M\|_{2,1} = \sum_{i=1}^n \|m_{i,:}\| = \sum_{j=1}^m \sqrt{\sum_{i=1}^n m_{ij}^2}.$$

### 2.2. Linear Regression

Traditional linear regression (LR) mines some kind of mapping relationship hidden in the given sample and label information [39]. It is one of the most important machine techniques and has a wide range of applications in image classification and feature extraction. Generally, the formula for LR is as follows:

$$W^* = \underset{W}{\operatorname{argmin}} \|W^T X - Y\|_F^2. \quad (1)$$

In Equation (1),  $W$  is the regression matrix, the linear regression model will choose the best regression matrix as much as possible, so it can make the output value  $W^T X$  as close as possible to the real label  $Y$ . Generally, this method of solving a model based on minimizing the mean square error is called the “least squares method” and is solved as follows:

$$W^* = (X^T X)^{-1} X^T Y. \quad (2)$$

Unfortunately, the classical LR model measures the regression error with the  $F$ -norm, which is sensitive to noise. Thereby, the generalization and robustness of LR is generally unsatisfactory.

### 2.3. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a widely used supervised dimensionality reduction algorithm [40]. Its main purpose is to make data compact within the same class and as dispersed as possible between classes after transforming the data into a low-dimensional space. Thus, the objective function of LDA is defined as follows:

$$P = \arg \min_{P^T P = 1} P^T (S_w - \lambda S_b) P \quad (3)$$

where  $\lambda$  is a small positive constant. LDA finds a subspace to distinguish different classes by minimizing the rank of the intra-class dispersion matrix  $S_w$  and maximizing the rank of the inter-class dispersion matrix  $S_b$ . Specifically,  $S_w$  and  $S_b$  are defined as follows:

$$S_w = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_j^i - m_i) (x_j^i - m_i)^T \quad (4)$$

$$S_b = \frac{1}{n} \sum_{i=1}^c n_i (m_i - m) (m_i - m)^T \quad (5)$$

where  $c$  is the number of classes in the sample, the number of samples in the  $i$ th class is denoted as  $n_i$ , and the  $j$ th sample of the  $i$ th class is denoted as  $x_j^i$ . The mean vector of class  $i$  samples is  $m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i$ , and the mean feature of all samples is  $m = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} x_j^i$ . Nevertheless, when LDA is applied to HSI dimensionality reduction, the number of extractable features is inevitably constrained by the number of classes. Although the projections obtained by LDA can reduce the distance between samples of the same class and increase the

separability between different classes, they overlook the locality structural information of HSI data. However, the locality information often plays an important role in the feature dimensionality reduction of HSIs. What is worse, the number of features extracted by LDA is limited by the number of classes, with a maximum of  $c - 1$  features.

#### 2.4. Formulation of JSLLDA

As previously mentioned, most existing LDA-based dimensionality reduction methods overlook the locality structural information within data, which includes critical information such as the spatial neighborhood relationships between data samples. Ignoring this information may prevent the model from capturing subtle differences in the data. To address this issue, we introduce a locality-preserving regularization term into the LDA model to reveal the locality structural information. Mathematically, this can be expressed as follows:

$$\min_{P,W,E} \text{Tr}\left(P^T(S_w - \mu S_b)P\right) + \lambda_1 \text{Tr}\left(P^T X L X^T P\right) \quad (6)$$

where  $P \in \mathbb{R}^{m \times c}$  is the projection matrix, and  $\mu$  is a balance factor.  $S_b$  is the between-class scatter matrix, and  $S_w$  is the within-class matrix.  $S$  is an undirected domain graph, in which the points represent sample points and the edges represent the nearest neighbor relationships between data samples. The Laplace matrix is  $L = D - W$ , where  $D$  is a diagonal matrix where each element on the diagonal is the sum of the row or column elements of  $S$ .

Further considering linking the projection with regression, a linear regression variant is introduced into the model. In order to obtain more feature information, the regression matrix is decomposed into a regression matrix  $W$  and a projection matrix  $P$ . In detail,  $W$  and  $P$  are the orthogonal matrix and projection matrix of  $c \times k$ ,  $m \times k$ , respectively, and  $k$  denotes the number of features to be extracted. The projection matrix  $P$  can only obtain  $c - 1$  projections, while the orthogonal matrix  $W$  has the size  $c \times k$ , implying that  $k$  projections can be obtained.  $k$  can be set to any positive integer by the user, as required. At this point, the number of features acquired by LDA is no longer limited by the number of classes. More importantly, a regression matrix  $W$  is introduced to consider the reconstruction relationship between the projected samples and the label information, i.e., the constraint  $Y = W P^T X + E$  can make the samples projected onto the low-dimensional space retain the label energy of the original data as much as possible. Equation (6) is rewritten as follows:

$$\begin{aligned} \min_{P,W,E} \text{Tr}\left(P^T(S_w - \mu S_b)P\right) + \lambda_1 \text{Tr}\left(P^T X L X^T P\right) \\ \text{s.t. } Y = W P^T X + E, W^T W = I. \end{aligned} \quad (7)$$

$E$  denotes the reconstruction error, which is utilized to fit noise. Considering the sparseness and robustness of the  $L_{2,1}$ -norm, which mitigates the influence of outliers, we utilize this norm to measure the reconstruction error  $E$ .

$$\begin{aligned} \min_{P,W,E} \text{Tr}\left(P^T(S_w - \mu S_b)P\right) + \lambda_1 \text{Tr}\left(P^T X L X^T P\right) + \lambda_2 \|E\|_{2,1} \\ \text{s.t. } Y = W P^T X + E, W^T W = I. \end{aligned} \quad (8)$$

In addition, the  $L_{2,1}$ -norm can be further applied to constrain the projection matrix to further mine the joint sparse information and to distinguish high-low-dimensional features. Specifically, imposing an  $L_{2,1}$ -constraint on the projection matrix can render the matrix row-sparse, thereby assigning larger weights to significant features, while assigning weights close to zero to redundant features or noise. Evidently, these features endowed with larger weights are precisely what we need. Then, the objective function of JSLLDA is formulated as follows:

$$\begin{aligned} \min_{P,W,E} \text{Tr}(P^T(S_w - \mu S_b)P) + \lambda_1 \text{Tr}(P^T X L X^T P) + \lambda_2 \|E\|_{2,1} + \lambda_3 \|P\|_{2,1} \\ \text{s.t. } Y = W P^T X + E, W^T W = I. \end{aligned} \quad (9)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are regularization parameters.  $P$  is the projection matrix, and  $\mu$  is balance factor. In the above model, the first term is a linear discriminant analysis term, which promotes the low-dimensional subspaces to be more cohesive in the same class. The second term is the LPP term, which makes up for the disadvantage of ignoring local information in the linear discriminant analysis and makes full use of local information to improve the representation ability of the extracted features. The third term utilizes the  $L_{2,1}$ -norm to measure the reconstruction error. Compared with the Frobenius norm, it lacks a square operation, which is beneficial for improving the robustness of the model. In addition, the fourth term imposes a  $L_{2,1}$ -constraint on the projection matrix  $P$ , which gives more weight to important features and less weight to redundant features or noise close to 0, showing that those features are needed for the actual task.

### 2.5. Optimization to JSLLDA

In this section, an iterative method is designed to solve the optimization problem of JSLLDA by using the alternating direction method of multipliers (ADMM) [41]. The corresponding generalized Lagrangian function for problem (9) is written as

$$\begin{aligned} L(W, P, E, Y) = \text{Tr}(P^T(S_w - \mu S_b)P) + \lambda_3 \|P\|_{2,1} + \lambda_2 \|E\|_{2,1} + \\ \lambda_1 \text{Tr}(P^T X L X^T P) - \frac{1}{2\beta} \|\eta\|_F^2 + \frac{\beta}{2} \left\| Y - W P^T X - E + \frac{\eta}{\beta} \right\|_F^2, \end{aligned} \quad (10)$$

where  $\beta$  is the penalty factor and  $Y$  is the Lagrange multiplier. The specific solution scheme is as follows:

- Step 1: Fix  $W$  and  $E$  to update  $P$ . The problem Equation (10) is transformed into the following optimization problem:

$$\begin{aligned} \min_P \text{Tr}(P^T(S_w - \mu S_b)P) + \lambda_3 \|P\|_{2,1} + \lambda_1 \text{Tr}(P^T X L X^T P) + \\ \frac{\beta}{2} \left\| Y - W P^T X - E + \frac{\eta}{\beta} \right\|_F^2, \end{aligned} \quad (11)$$

where  $H$  is defined as  $H_{ii} = \frac{1}{2\|p^i\|_2}$ ,  $Y - E + \frac{\eta}{\beta}$  is defined as  $M$ .

By evaluating the derivative of Equation (11) with respect to  $P$  and setting it to 0, the following can be obtained:

$$2(S_w - \mu S_b)P + \lambda_3 H P + \beta (X X^T P - X M^T W) + \lambda_1 X L X^T P = 0. \quad (12)$$

then,  $P$  is calculated as follows:

$$P = \left[ 2(S_w - \mu S_b) + \lambda_3 H + \beta X X^T + \lambda_1 X L X^T \right]^{-1} \beta X M^T W. \quad (13)$$

- Step 2: Fix  $P$  and  $E$  to update  $W$ . The solution to  $W$  can be obtained by minimizing the equivalence problem (7)

$$\min_{W^T W = I} \left\| Y - W P^T X - E + \frac{\eta}{\beta} \right\|_F^2. \quad (14)$$

Let  $Y - E + \frac{\eta}{\beta} = M$ , and problem (14) is rewritten as

$$\min_{W^T W = I} \|M - WP^T X\|_F^2 = \max_{W^T W = I} \text{Tr}(W^T M X^T P). \quad (15)$$

Suppose the SVD of  $M X^T P$  is

$$\text{SVD}(M X^T P) = USV^T. \quad (16)$$

Then, the solution of  $W$  can be obtained by

$$W = UV^T. \quad (17)$$

- Step 3: Fix  $W$  and  $P$  to update  $E$ . Let us discuss a situation where  $W$  and  $P$  are provided. At this point,  $E$  can be solved using the following function:

$$\min_E \lambda_2 \|E\|_{2,1} + \frac{\beta}{2} \left\| Y - WP^T X + \frac{\eta}{\beta} - E \right\|_F^2. \quad (18)$$

According to [42],  $E$  can be expressed as the following closed solution:

$$E = \Omega_{\frac{\lambda_2}{\beta}} \left( Y - WP^T X + \frac{\eta}{\beta} \right), \quad (19)$$

where  $\Omega$  is the shrinkage operator.

- Step 4: update  $\eta, \beta$

$$\begin{aligned} \eta &= \eta + \beta (Y - WP^T X - E), \\ \beta &= \min(\rho\beta, \beta_{\max}). \end{aligned} \quad (20)$$

where  $\rho$  and  $\beta_{\max}$  are constant. The iterative solving process of JSLLDA is summarized in Algorithm 1.

---

#### Algorithm 1: The Iterative Algorithm for Solving JSLLDA

---

Input: Sample data  $X$ , label matrix  $Y$ , class compactness graph weight matrix  $S$ , reduced dimension  $m$ , parameters  $\lambda_1, \lambda_2, \lambda_3$ , and maximum number of iteration steps  $T$ .

Initialize:  $\beta = 0.1, \rho = 1.01, \beta_{\max} = 10^5, \mu = 10^{-5}$ .  $Q = \mathbf{0}, E = \mathbf{0}, \eta = \mathbf{0}$ , where  $\mathbf{0}$  is zero matrix, initialize  $P$  as an orthogonal matrix.

while not converge do

1. Update  $P$  by using (13),
2. Update  $W$  by using (17),
3. Update  $E$  by using (19),
4. Update  $\eta, \beta$  by using (20).
5. Check the convergence conditions

end while

Output:  $W, P, E$ .

---

### 3. Experiments

In this section, the effectiveness of JSLLDA was validated on HSIs using the Salinas dataset, the University of Pavia dataset, and the Heihe dataset. To provide a comprehensive comparison, results from several state-of-the-art algorithms, including LDA [15], LPP [18], RSLDA [35], TRLD [23], LRP [31], and  $L_{2,p}$ -RER [32], are presented. SVM classifiers were utilized to classify the dimensionality-reduced data, with classification accuracy serving as the quantitative evaluation metric for each method. The objective metrics

employed were the overall accuracy (OA), the average accuracy (AA), the kappa coefficient ( $\kappa$ ), and standard deviation.

### 3.1. Experimental Datasets

(1) The Salinas hyperspectral dataset was captured by the AVIRIS sensor over Salinas Valley, California, with a spatial resolution of 3.7 m per pixel. Originally consisting of 224 bands, the dataset was reduced to 204 usable bands by removing bands [108–112], [154–167], and the 224th band, due to water absorption. The image dimensions are  $512 \times 217$  pixels, yielding a total of 111,104 pixels, of which 54,129 pixels are of practical use. These pixels are categorized into 16 classes, including fallow, grapes\_untrained, and soil\_vineyard\_develop.

(2) The University of Pavia hyperspectral dataset was captured by the Reflective Optics Spectrographic Imaging System in 2003, covering part of the city of Pavia, Italy, with a spatial resolution of 1.3 meters. The dataset originally contained 115 bands, but 12 bands were removed due to damage, leaving 103 spectral bands in use. The image dimensions are  $610 \times 340$  pixels, totaling 2,207,400 pixels. Excluding background elements, 42,776 pixels are of practical use. These pixels are classified into nine different categories, including bitumen, trees, and gravel.

(3) The Heihe dataset was collected by CASI/SASI sensors. There are eight types of land cover in the dataset, and the dataset size is  $684 \times 453$  pixels, with a spatial resolution of 2.4 m. After removing 14 bands that are heavily affected by noise, the remaining 135 bands were utilized.

### 3.2. Evaluation Index

This subsection describes how OA, AA, and  $\kappa$  are calculated. To compute OA, AA, and  $\kappa$ , it is first necessary to compute the following confusion matrix:

$$M = \begin{pmatrix} m_{11} & \dots & m_{1C} \\ \vdots & \ddots & \vdots \\ m_{C1} & \dots & m_{CC} \end{pmatrix} \quad (21)$$

In Equation (21),  $m_{ij}$  denotes the number of samples that originally belonged to class  $i$  but were predicted to be class  $j$ , and  $C$  denotes the total number of classes. OA is the ratio of the number of data samples accurately classified to the total number of samples tested and is calculated as follows:

$$OA = \sum_{i=1}^C \frac{m_{ii}}{N_{test}} \quad (22)$$

In Equation (22),  $N_{test}$  represents the total number of test set samples and  $m_{ii}$  is the diagonal element in the confusion matrix, representing the number of samples of class  $i$  correctly classified. AA is the average of the accuracy of each category and is calculated as follows:

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{m_{ii}}{N_i} \quad (23)$$

where  $N_i$  is the total number of samples of class  $i$ .  $\kappa$  is a statistical measure used to assess the agreement between the classification results and the true classification.  $\kappa$  ranges from  $-1$  to  $1$ , where higher values indicate a stronger alignment between the method's results and the actual classifications. The coefficient is calculated as follows:

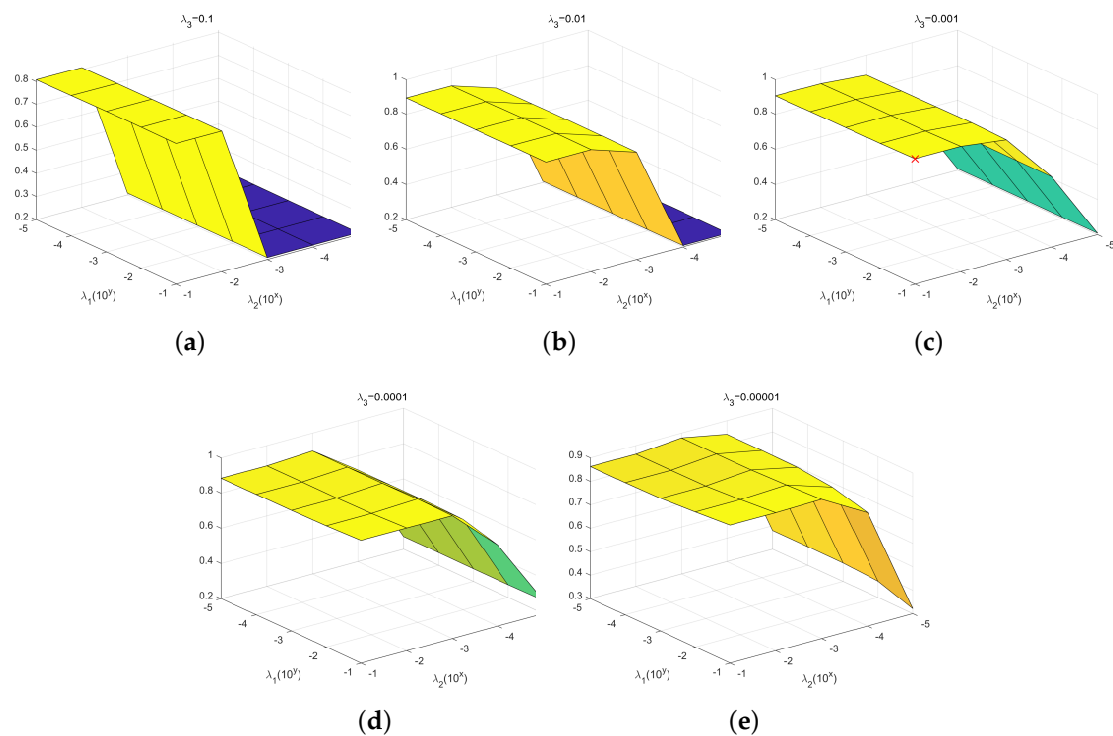
$$\kappa = \frac{OA - p_e}{1 - p_e} \quad (24)$$



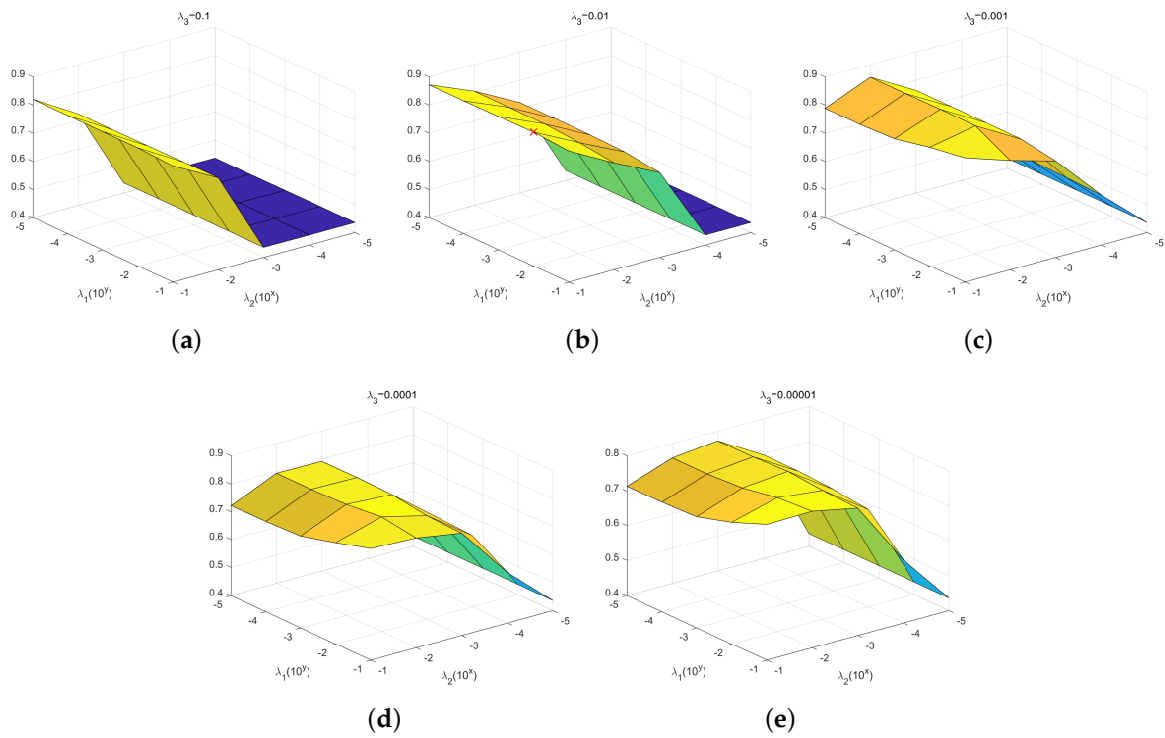
where  $p_e$  is the expected random classification accuracy calculated as  $p_e = \frac{1}{N^2} \sum_{i=1}^C (M_{i.} \times M_{.i})$ .  $M_{i.}$  is the sum of row  $i$  in the confusion matrix, and  $M_{.i}$  is the sum of column  $i$  in the confusion matrix. For the metrics OA, AA, and  $\kappa$ , higher values indicate better model performance. Conversely, for the standard deviation, a lower value signifies greater stability in the model's performance.

### 3.3. Experiment Setup

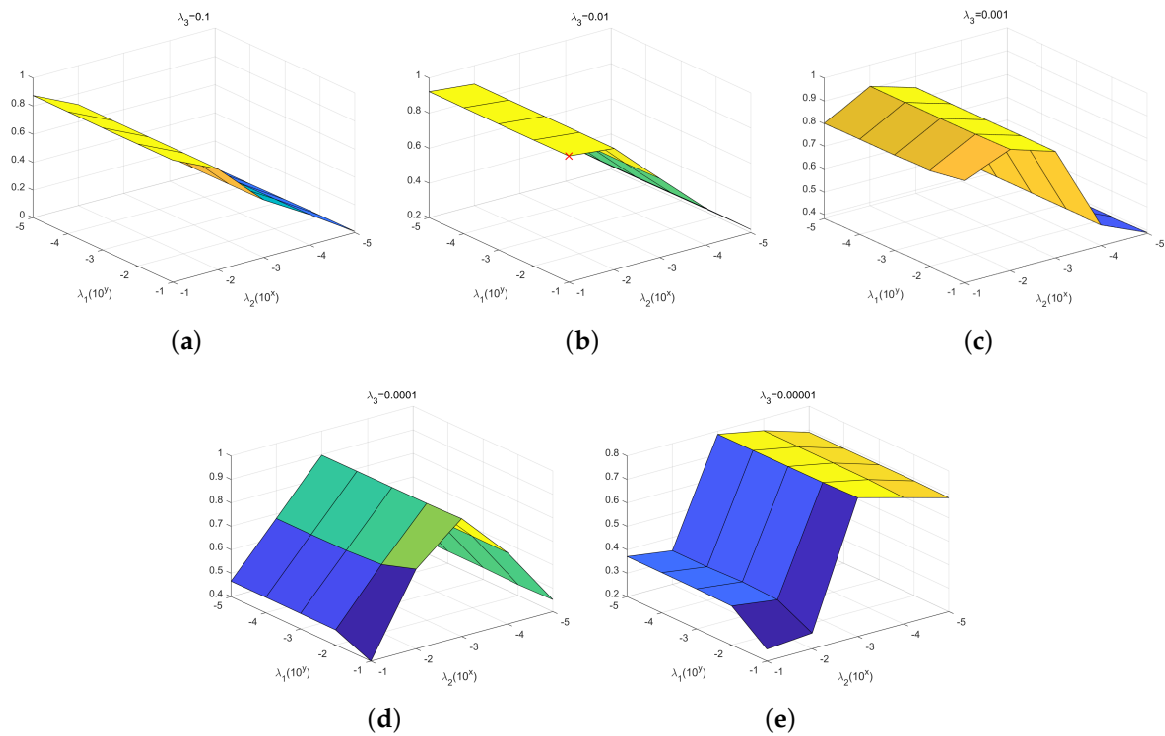
For the proposed JSLLDA, there are three crucial parameters: the regularization parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ . These parameters control the effect of the locality-preserving term, the reconstruction error term, and the joint sparse term in the objective function, respectively. To achieve the best performance for JSLLDA, we conducted cross-validation experiments on the three datasets separately. The parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  were tested with values set to 0.1, 0.01, 0.001, 0.0001, 0.00001. Figures 1–3 present the results of tuning  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  for the Salinas, University of Pavia, and Heihe datasets, respectively. The red crosses in these three figures are the best parameter positions we found. For ease of presentation,  $\lambda_3$  was fixed at 0.1, 0.01, 0.001, 0.0001, 0.00001, and the variations in OA with respect to  $\lambda_1$  and  $\lambda_2$  were then analyzed. As illustrated in Figure 1, the third panel shows that when  $\lambda_3$  was set to 0.001, there existed a combination of  $\lambda_1$  and  $\lambda_2$  where the overall accuracy (OA) reached its peak. Specifically, for the Salinas dataset, the optimal parameters were  $(\lambda_1, \lambda_2, \lambda_3) = (0.1, 0.1, 0.001)$ . Similarly, for the University of Pavia dataset, the second panel of Figure 2 indicates that the model performed best when  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  were set to (0.01, 0.1, 0.01). For the Heihe dataset, the optimal parameters, as shown by the red markers in the second panel of Figure 3, were  $(\lambda_1, \lambda_2, \lambda_3) = (0.1, 0.1, 0.01)$ .



**Figure 1.** For the Salinas dataset, the OA varied with  $\lambda_1$  and  $\lambda_2$  when  $\lambda_3$  was set to (a) 0.1, (b) 0.01, (c) 0.001, (d) 0.0001, and (e) 0.00001.



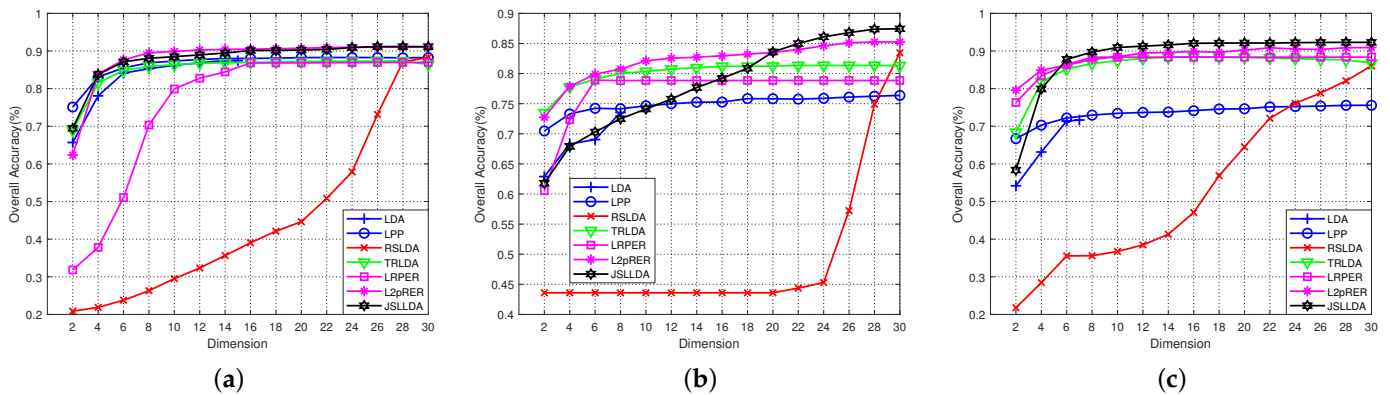
**Figure 2.** For the University of Pavia dataset, the OA varied with  $\lambda_1$  and  $\lambda_2$  when  $\lambda_3$  was set to (a) 0.1, (b) 0.01, (c) 0.001, (d) 0.0001, and (e) 0.00001.



**Figure 3.** For the Heihe dataset, the OA varied with  $\lambda_1$  and  $\lambda_2$  when  $\lambda_3$  was set to (a) 0.1, (b) 0.01, (c) 0.001, (d) 0.0001, and (e) 0.00001.

According to Figure 4, for the three hyperspectral datasets, the OA values of all methods in the experiment, except for LDA, tended to stabilize after reaching a certain number of retained features. Specifically, OA tended to stabilize after retaining 30 features

for all methods except LDA. However, for LDA, since it can only obtain at most  $c - 1$  features, in Figure 4, the low dimensional feature dimensions obtained by LDA for the Salinas dataset, University of Pavia dataset, and Heihe dataset are set to 15, 8, and 7, respectively.



**Figure 4.** The average recognition rates for the different datasets versus the different dimensions of the features extracted by different methods. (a) The Salinas dataset. (b) The University of Pavia dataset. (c) The Heihe dataset.

### 3.4. Experimental Results and Analysis

The AA, OA,  $\kappa$ , and their standard deviation for the different dimensionality reduction methods on the three hyperspectral datasets are presented in Tables 1–3. And the corresponding highest classification accuracy for each class is bolded. Figures 5–7 illustrate the visualization results obtained by applying the LDA, LPP, RSLDA, TRLDA, LRPER,  $L_{2,p}$ -RER, and JSLDDA methods for dimensionality reduction on the three hyperspectral datasets, followed by classification using an SVM classifier.

For the Salinas dataset, 1% of the samples in each category were randomly selected for training. Table 1 and Figure 5 present the numerical results and visualization of SVM classification after dimensionality reduction using various methods on the Salinas dataset. Table 1 provides the detailed classification accuracy for each method, as well as the OA, AA, and the  $\kappa$  coefficient. Figure 5 illustrates a corresponding visual representation of the classification results. The classification results presented in Table 1 demonstrate that the proposed JSLDDA achieved superior performance in terms of average AA, OA, and the  $\kappa$ . Specifically, JSLDDA surpassed LDA, LPP, RSLDA, TRLDA, LRPER, and  $L_{2,p}$ -RER by 0.52% to 8.84%, 0.2% to 6.54%, and 0.22% to 7.4% for AA, OA, and  $\kappa$  coefficient, respectively. It is evident that the standard deviations of the metrics for our proposed method were the smallest. This indicates that the performance of JSLDDA exhibited minimal fluctuations, which typically means that results are relatively stable across multiple experiments. Notably, JSLDDA also achieved the highest single-class accuracies for Class 2 (Broccoli-green-weeds-2), Class 3 (Fallow), Class 4 (Fallow-rough-plow), Class 5 (Fallow-smooth), Class 6 (Stubble), Class 9 (Soil-vineyard-develop), Class 10 (Corn-senesced-green-weeds), and Class 16 (Vineyard-vertical-trellis). These findings are further corroborated by Figure 5, which shows that the regions corresponding to these classes were more homogeneous compared to those obtained using other methods. In Figure 5, Classes 2 and 3 are circled with light cyan dashed lines. By comparing Class 2 in Figure 5f,h, it can be observed that the region classified by JSLDDA, which had the highest recognition rate, is smoother than that classified by TRLDA, which had the lowest recognition rate. This smoother appearance was due to fewer misclassifications in JSLDDA. The same result was found for the similar Class 3.

**Table 1.** Classification accuracies obtained by applying SVM on the features extracted from the Salinas hyperspectral dataset.

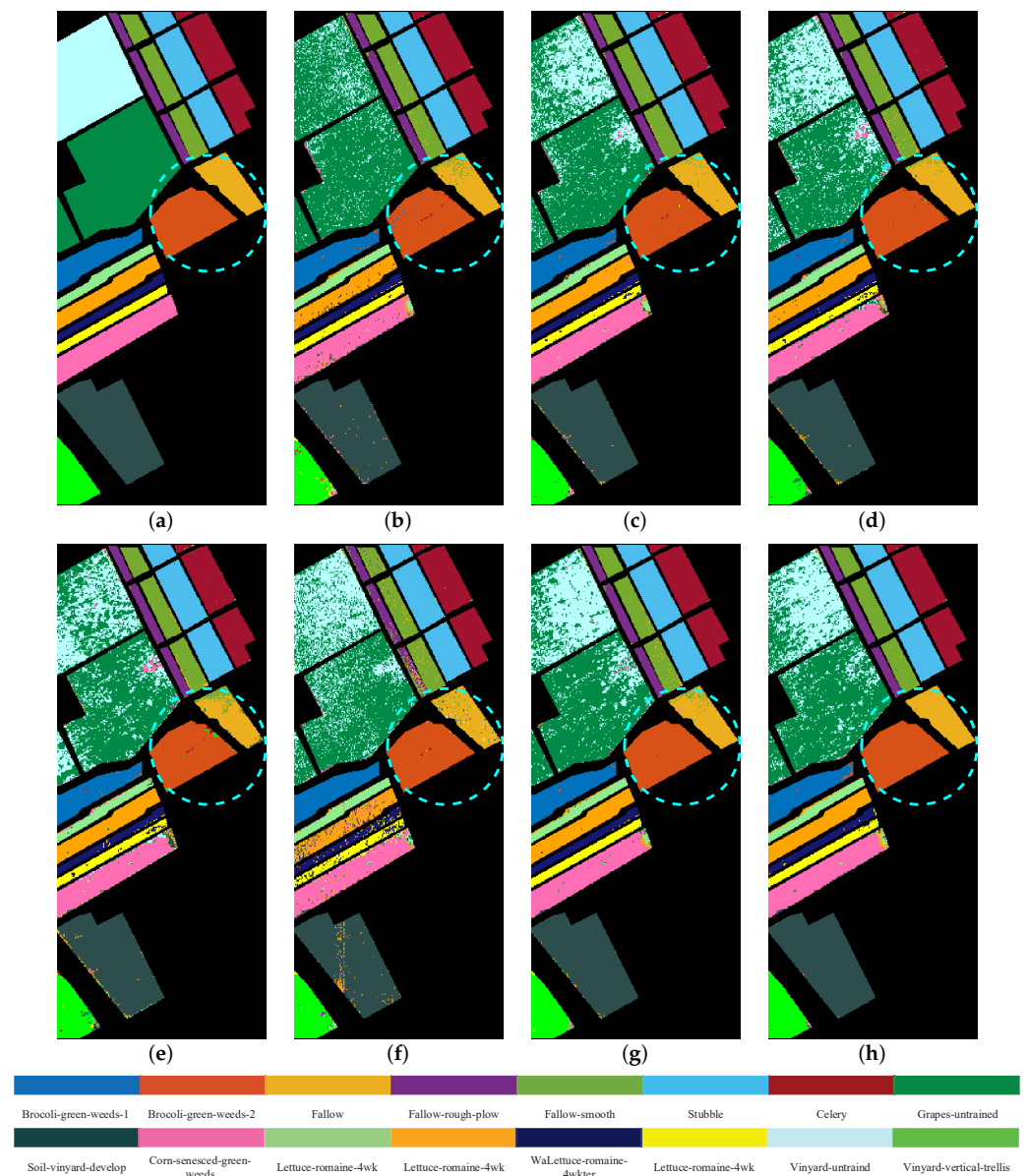
Class #	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLDA
1	93.42 ± 6.41	98.48 ± 0.36	98.24 ± 1.81	96.74 ± 1.02	100 ± 0.00	<b>98.85 ± 0.86</b>	98.24 ± 1.50
2	98.77 ± 0.42	99.14 ± 0.20	98.53 ± 0.46	98.45 ± 0.33	99.45 ± 0.15	99.49 ± 0.22	<b>99.91 ± 0.05</b>
3	92.57 ± 2.96	94.65 ± 4.30	92.72 ± 4.58	86.93 ± 6.05	77.58 ± 6.60	92.79 ± 3.74	<b>99.12 ± 0.75</b>
4	96.52 ± 2.54	97.68 ± 0.99	97.48 ± 1.95	98.73 ± 0.92	97.54 ± 0.07	99.23 ± 0.46	<b>99.41 ± 0.53</b>
5	94.73 ± 2.51	96.66 ± 1.49	95.53 ± 1.31	93.87 ± 2.45	95.07 ± 3.21	91.83 ± 4.98	<b>98.15 ± 0.75</b>
6	99.65 ± 0.16	99.55 ± 0.12	98.83 ± 0.36	98.54 ± 1.02	98.92 ± 0.43	99.69 ± 0.10	<b>99.71 ± 0.21</b>
7	99.59 ± 0.19	99.46 ± 0.11	98.57 ± 0.46	80.92 ± 5.19	99.50 ± 0.13	<b>99.77 ± 0.09</b>	99.72 ± 0.12
8	<b>92.20 ± 5.60</b>	87.18 ± 1.97	85.23 ± 2.99	80.92 ± 5.19	82.27 ± 1.81	85.37 ± 2.51	86.25 ± 1.82
9	97.55 ± 1.24	98.94 ± 0.79	97.79 ± 1.08	96.99 ± 0.85	96.36 ± 1.41	99.27 ± 0.63	<b>99.60 ± 0.29</b>
10	90.57 ± 2.25	84.20 ± 6.80	86.61 ± 4.28	83.39 ± 5.38	84.33 ± 2.05	93.24 ± 3.08	<b>93.87 ± 2.07</b>
11	91.41 ± 1.61	<b>91.92 ± 1.54</b>	85.87 ± 6.61	90.79 ± 3.96	33.52 ± 4.22	90.02 ± 4.02	90.84 ± 3.27
12	94.87 ± 2.27	96.99 ± 4.01	94.92 ± 3.51	98.24 ± 1.79	78.83 ± 7.24	<b>99.44 ± 0.57</b>	99.39 ± 0.94
13	97.92 ± 0.43	97.62 ± 0.51	96.66 ± 1.31	96.16 ± 3.05	94.08 ± 1.99	<b>98.51 ± 0.68</b>	97.66 ± 1.20
14	95.26 ± 1.89	93.42 ± 1.29	90.49 ± 4.70	90.09 ± 4.93	84.87 ± 3.35	<b>96.96 ± 1.13</b>	94.58 ± 1.72
15	16.05 ± 12.54	48.85 ± 4.67	60.33 ± 3.93	53.67 ± 7.27	64.38 ± 4.87	<b>67.75 ± 3.78</b>	63.44 ± 3.68
16	95.28 ± 3.45	96.77 ± 1.78	91.29 ± 2.46	90.82 ± 0.63	90.09 ± 4.09	97.70 ± 1.05	<b>98.30 ± 0.79</b>
AA	90.40 ± 0.73	92.59 ± 0.58	91.82 ± 0.70	86.05 ± 0.77	90.52 ± 0.86	94.37 ± 0.30	<b>94.89 ± 0.22</b>
OA	84.61 ± 0.63	88.23 ± 0.64	88.64 ± 0.35	86.41 ± 0.55	86.09 ± 0.37	90.95 ± 0.50	<b>91.15 ± 0.26</b>
$\kappa$	82.73 ± 0.74	86.84 ± 0.72	87.32 ± 0.39	84.83 ± 0.61	84.47 ± 0.42	89.91 ± 0.55	<b>90.13 ± 0.29</b>

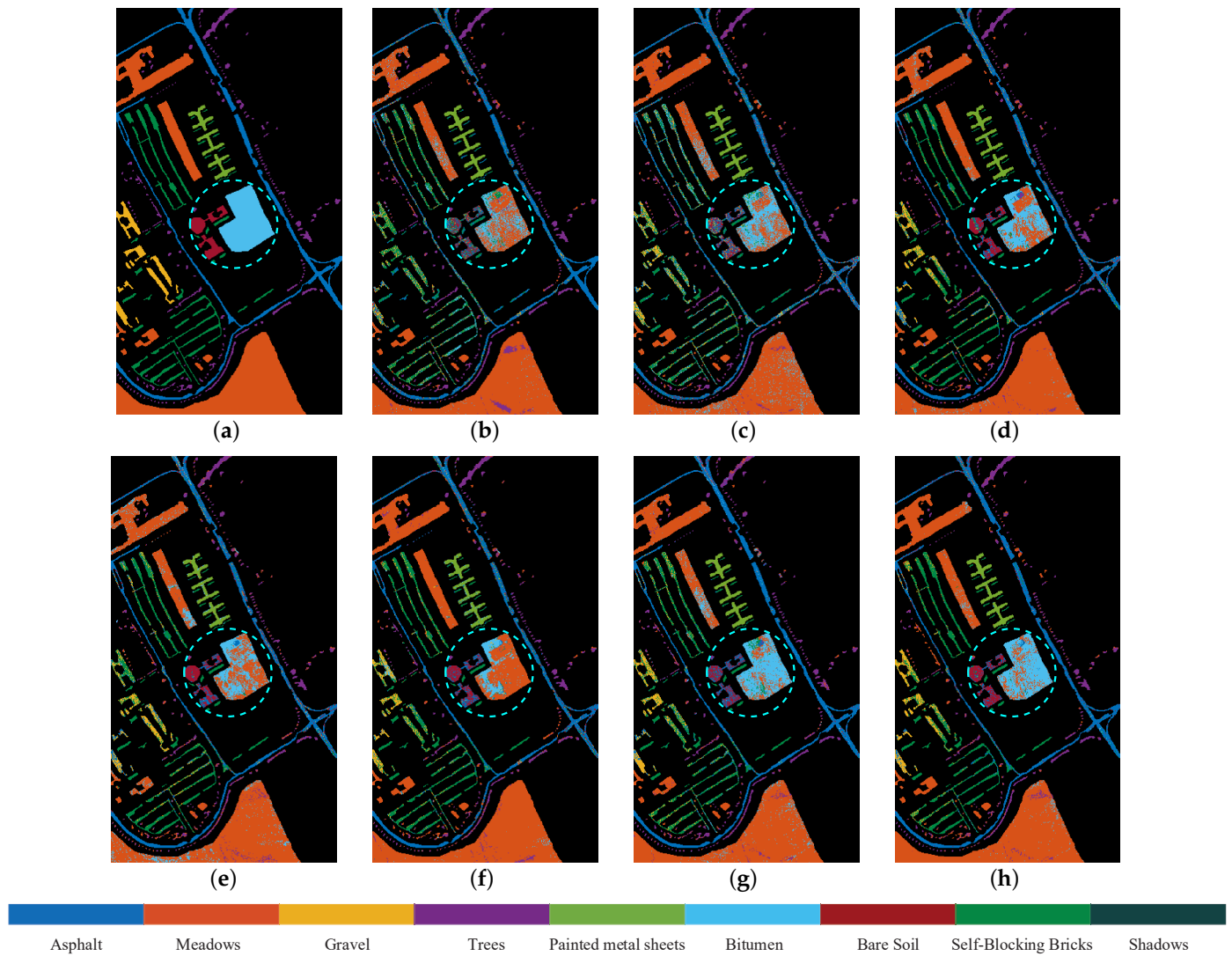
For the University of Pavia dataset, 1% of the samples from each class were randomly selected for training, while the remaining samples were used for testing. As shown in Table 2, JSLDA achieved an OA that was 14.24%, 11.66%, 4.58%, 6.73%, 9.58%, and 2.57% higher than LDA, LPP, RSLDA, TRLDA, LRPER, and  $L_{2,p}$ -RER, respectively. The standard deviation for OA was slightly higher than that of LPP and  $L_{2,p}$ -RER, but it was still lower than that of the other methods. Additionally, the AA and  $\kappa$  coefficient were also 4.29% to 18.75% and 3.44% to 19.79% higher, respectively. Similarly, in terms of the stability of the AA and Kappa coefficients, JSLDA was only slightly less stable than LPP and  $L_{2,p}$ -RER. Figure 6 visually confirms that JSLDA produced smoother classification regions compared to the other dimensionality reduction methods, particularly for Class 3 (Gravel), Class 4 (Trees), Class 5 (Painted metal sheets), Class 6 (Bitumen), Class 7 (Bare soil), and Class 9 (Shadows). Furthermore, despite the relatively small number of samples in Class 7 (Bitumen), the regions classified as such were notably larger and more homogeneous. Classes 6 and 7 have been highlighted in our analysis. As shown in Table 2, LDA's recognition rates for Classes 6 and 7 were below 50%. Consequently, in Figure 6b, more than half of the areas corresponding to Classes 6 and 7 are displayed in incorrect colors due to the high number of misclassified samples. In contrast, in Figure 6h, Classes 6 and 7 exhibit a relatively higher classification accuracy, leading to a greater display of homogeneous regions.

For the Heihe dataset, 20 samples from each class were selected for the training set, with the remaining samples used for testing. As shown in Table 3, the proposed method enhanced the OA by 1.31% to 20.81%, and the Kappa coefficient was 0.22% to 7.4% higher compared to the other methods. Observing the standard deviation of OA, it is evident that our proposed method performed comparably to LRPER, with only a slight difference. For the AA and  $\kappa$  coefficient, the performance of our method was slightly inferior only to that of  $L_{2,p}$ -RER. In Figure 7, Class 3 has been specifically highlighted. It is observed that the higher the recognition rate, the closer the corresponding classification result map aligned with the actual land cover distribution of Heihe, as shown in Figure 7a.

**Table 2.** Classification accuracies obtained by applying SVM on the features extracted from the university of Pavia hyperspectral dataset.

Class #	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLDA
1	66.89 ± 3.04	79.82 ± 4.30	84.86 ± 2.01	83.79 ± 2.07	83.06 ± 1.70	<b>90.57 ± 1.97</b>	87.01 ± 2.47
2	93.83 ± 2.06	90.81 ± 1.05	94.52 ± 1.83	92.17 ± 2.55	<b>96.53 ± 1.18</b>	93.98 ± 1.24	95.56 ± 1.22
3	46.97 ± 11.03	45.91 ± 3.49	49.66 ± 3.54	49.15 ± 4.88	50.19 ± 1.12	57.22 ± 6.11	<b>63.16 ± 3.97</b>
4	78.95 ± 5.22	82.23 ± 1.82	78.71 ± 5.21	79.27 ± 4.96	69.25 ± 5.31	81.87 ± 2.87	<b>82.53 ± 3.81</b>
5	99.78 ± 0.09	98.69 ± 0.43	98.43 ± 0.84	92.35 ± 4.13	95.61 ± 5.04	98.36 ± 1.28	<b>98.47 ± 0.93</b>
6	31.94 ± 11.53	48.88 ± 4.58	53.83 ± 7.57	48.76 ± 7.22	21.28 ± 3.25	70.43 ± 3.66	<b>75.77 ± 2.97</b>
7	40.62 ± 11.58	39.82 ± 8.08	68.63 ± 9.15	66.37 ± 9.75	53.76 ± 5.38	57.66 ± 10.74	<b>73.02 ± 5.87</b>
8	53.951 ± 9.65	49.23 ± 6.13	84.22 ± 2.58	83.10 ± 2.49	77.96 ± 2.50	74.97 ± 3.90	85.87 ± 2.01
9	79.30 ± 6.24	87.75 ± 2.85	99.57 ± 0.27	98.98 ± 0.46	98.40 ± 1.21	97.28 ± 2.68	<b>99.54 ± 0.31</b>
AA	65.80 ± 3.20	69.24 ± 0.56	79.16 ± 1.24	77.10 ± 1.73	71.78 ± 0.97	80.26 ± 1.20	<b>84.55 ± 0.82</b>
OA	73.79 ± 1.79	76.37 ± 0.42	83.45 ± 1.00	81.30 ± 1.25	78.45 ± 1.16	85.46 ± 0.51	<b>88.03 ± 0.70</b>
$\kappa$	64.20 ± 2.57	68.20 ± 0.52	77.63 ± 1.37	74.78 ± 1.67	70.15 ± 1.64	80.55 ± 0.66	<b>83.99 ± 0.94</b>

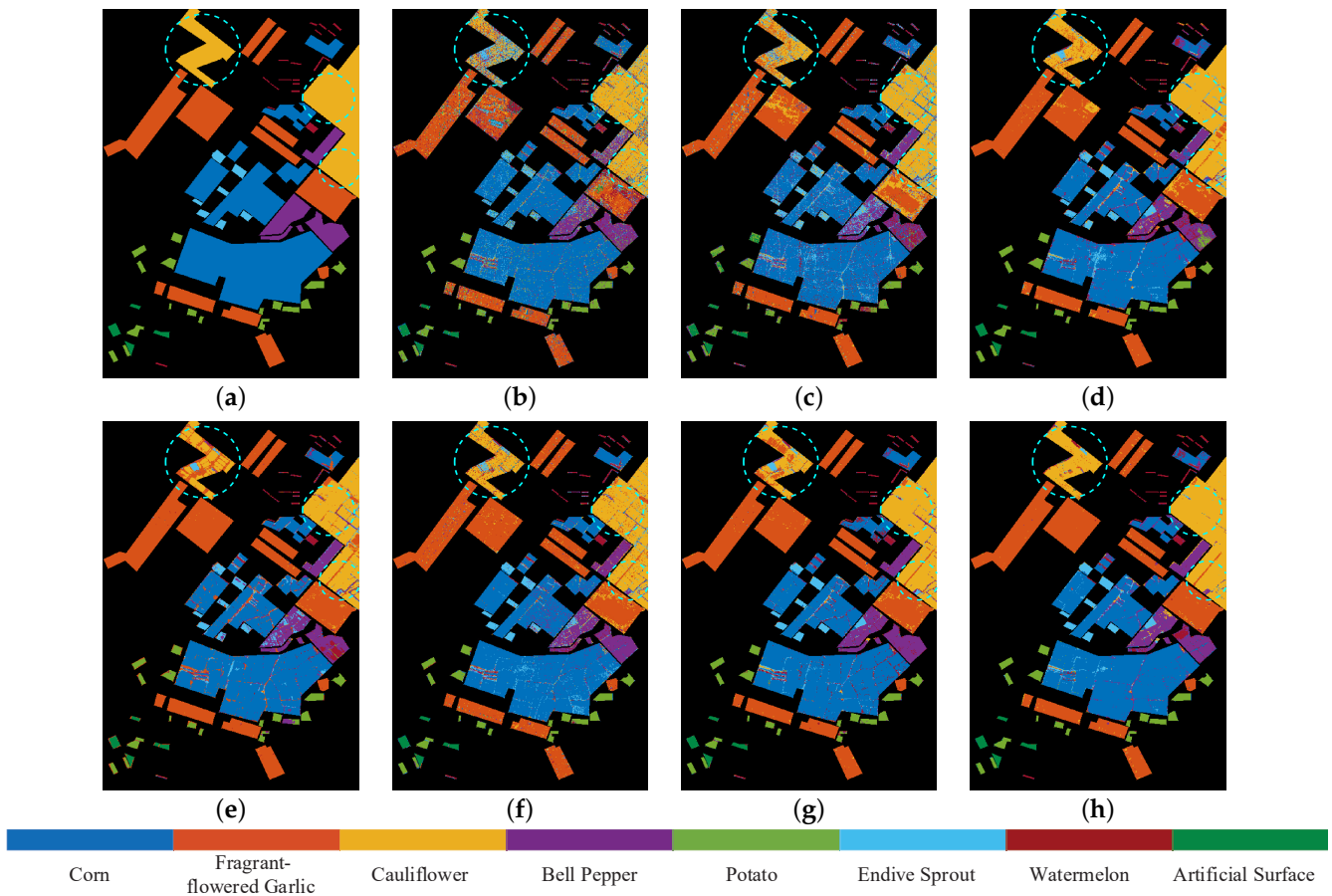
**Figure 5.** Classification maps of the different methods for the Salinas dataset. (a) Ground truth, (b) LDA, (c) LPP, (d) RSLDA, (e) TRLDA, (f) LRPER, (g)  $L_{2,p}$ -RER, (h) JSLDA.



**Figure 6.** Classification maps of different methods for the University of Pavia dataset. (a) Ground truth, (b) LDA, (c) LPP, (d) RSLDA, (e) TR LDA, (f) LRPER, (g)  $L_{2,p}$ -RER, (h) JSLLDA.

**Table 3.** Classification accuracies obtained by applying SVM on the features extracted from the Heihe hyperspectral dataset.

Class #	LDA	LPP	RSLDA	TR LDA	LRPER	$L_{2,p}$ -RER	JSLLDA
1	73.83 ± 11.04	78.56 ± 4.76	87.61 ± 2.49	85.12 ± 2.45	90.86 ± 1.84	90.79 ± 2.29	<b>91.37 ± 1.52</b>
2	73.93 ± 6.91	77.56 ± 3.79	89.06 ± 6.26	98.78 ± 0.31	92.41 ± 1.04	96.63 ± 1.21	<b>96.96 ± 0.88</b>
3	64.60 ± 6.46	76.05 ± 2.44	84.83 ± 2.59	78.52 ± 3.52	81.84 ± 1.49	84.71 ± 0.12	<b>89.54 ± 3.40</b>
4	65.26 ± 8.21	49.41 ± 12.51	71.18 ± 5.70	77.08 ± 4.18	75.75 ± 9.39	<b>86.45 ± 1.75</b>	85.27 ± 4.90
5	83.17 ± 5.10	79.69 ± 6.05	88.68 ± 4.19	84.78 ± 3.56	94.25 ± 1.39	<b>94.05 ± 1.42</b>	96.04 ± 1.01
6	55.77 ± 4.31	78.82 ± 4.84	81.52 ± 5.65	83.82 ± 4.00	83.88 ± 4.39	89.64 ± 1.50	<b>90.67 ± 3.50</b>
7	61.68 ± 9.87	58.19 ± 11.89	69.51 ± 5.88	75.70 ± 4.98	80.49 ± 2.60	81.61 ± 9.01	<b>91.64 ± 2.50</b>
8	86.20 ± 5.68	82.11 ± 5.98	90.30 ± 3.42	78.50 ± 5.48	92.22 ± 1.90	90.09 ± 1.66	<b>92.23 ± 2.85</b>
AA	70.56 ± 3.75	72.55 ± 1.73	82.84 ± 2.34	86.41 ± 1.55	86.17 ± 0.46	89.51 ± 1.44	<b>92.09 ± 0.98</b>
OA	71.48 ± 5.69	75.57 ± 2.58	86.07 ± 2.07	86.80 ± 1.25	88.36 ± 0.99	90.91 ± 0.39	<b>92.29 ± 0.68</b>
$\kappa$	63.40 ± 6.46	68.22 ± 3.06	81.39 ± 2.63	82.15 ± 1.65	84.30 ± 1.27	87.74 ± 0.47	<b>89.58 ± 0.89</b>



**Figure 7.** Classification maps of the different methods for the Heihe dataset. (a) Ground truth, (b) LDA, (c) LPP, (d) RSLDA, (e) TRLDA, (f) LRPER, (g)  $L_{2,p}$ -RER, (h) JSL LDA.

#### 4. Model Analysis

In this section, the computational complexity of the proposed method is first considered. Subsequently, the robustness of the model was validated through a series of experiments, and the specific roles of each component of the model were explored. Based on these comprehensive experimental results and analyses, a thorough discussion and evaluation of JSL LDA is then conducted, highlighting its advantages, while also candidly acknowledging its limitations. Finally, future research directions for JSL LDA are outlined, and its potential value in diverse application scenarios is explored.

##### 4.1. Computational Complexity Analysis

In this subsection, the computational complexity of JSL LDA is analyzed. Suppose that  $m$  is the dimension of the original data, projecting  $P$  reduces the original dimension to a low-dimensional space with  $d$  features. The major time-consuming computation is divided into two parts, originating from step 1 and step 2 in Algorithm 1. In step 1, the main computational cost originates from the inverse operation of the matrix, which has a time complexity of approximately  $O(m^3)$ . While in step 2, the singular value decomposition of the matrix contributes the second part of the complexity, which is approximately  $O(md^2)$ . Furthermore, considering that JSL LDA requires  $T$  iterations to achieve convergence, the total time complexity of JSL LDA can be estimated as  $O(T(m^3 + md^2))$ . In general,  $m > d$ , when the total computational complexity is about  $O(Tm^3)$ .

To provide a straightforward comparison of the computational efficiency of various methods, each method was executed 10 times, and the average runtime in seconds was recorded as the experimental result. These results are presented in Table 4. Notably, all the programs used in the aforementioned experiments were run on a computer equipped

with a 3.40 GHz Intel(R) Core(TM) i7-13700KF CPU and 32 GB of RAM, utilizing MATLAB 2021a as the programming platform. As shown in Table 4, the LRPER method exhibited the longest runtime across the three datasets, while LDA had the shortest runtime. The single execution time of JSLLDA was consistently under one second, indicating a significant advantage over comparative algorithms and demonstrating its negligible computational burden in practical applications.

**Table 4.** Average calculation time (seconds) for different dimensionality reduction methods.

Dataset	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLLDA
Salinas	0.018	0.054	0.2247	0.0943	20.7665	5.0521	2.6303
University of Pavia	0.0056	0.0026	0.1156	0.0455	25.6457	2.5703	1.1550
HeiHe	0.0113	0.0036	0.1159	0.0644	1.441	0.3218	0.2203

#### 4.2. Robustness Analysis

To comprehensively assess the robustness of the proposed model, zero-mean Gaussian noise with a local variance of 0.5 was introduced into the datasets: the first 100 spectral bands of the Salinas dataset, the first 50 bands of the University of Pavia dataset, and the first 40 bands of the Heihe dataset. The performance of the various dimensionality reduction methods was then evaluated on these noise-corrupted datasets, with the results presented in Tables 5–7. And, we bolded the highest value for each metric. This thorough evaluation aimed to demonstrate the stability and effectiveness of the model under challenging noisy conditions.

Upon a detailed examination and analysis of the experimental results presented in Tables 5–7, it is clear that the JSLLDA method proposed in this study significantly outperformed the other comparative approaches in terms of OA, AA, and  $\kappa$  across three datasets with varying levels of data corruption: Salinas, University of Pavia, and Heihe. Specifically, for the Salinas dataset, JSLLDA demonstrated a substantial improvement in OA, ranging from 3.14% to 42.29%, highlighting its superior performance. Similarly, on the University of Pavia dataset, JSLLDA achieved notable results, with an OA enhancement between 0.11% and 35.22%, markedly surpassing the other benchmark algorithms. Furthermore, the evaluation on the Heihe dataset revealed that JSLLDA delivered an OA improvement within the range of 1.56% to 46.26%, further corroborating its stability and superiority across diverse corrupted scenarios. In summary, the JSLLDA method not only demonstrated significant efficiency in hyperspectral data classification tasks but also highlighted its robust capability to manage complex data corruption scenarios.

**Table 5.** Comparison of the OA (%) classification results of the SVM classifiers after processing the Salinas dataset corrupted with Gaussian noise, using the different dimensionality reduction methods.

Metrics	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLLDA
OA	46.53 ± 1.39	76.82 ± 0.93	85.63 ± 0.62	79.54 ± 3.05	82.85 ± 0.67	77.69 ± 4.33	<b>88.77 ± 0.72</b>
AA	45.62 ± 1.23	77.34 ± 1.35	89.79 ± 0.96	80.78 ± 4.50	84.56 ± 0.74	75.37 ± 6.35	<b>91.73 ± 0.57</b>
$\kappa$	40.80 ± 1.50	74.11 ± 1.03	83.95 ± 0.69	77.09 ± 3.41	80.87 ± 0.74	75.15 ± 4.83	<b>87.48 ± 0.80</b>

**Table 6.** Comparison of the OA (%) classification results of the SVM classifiers after processing the Pavia dataset corrupted with Gaussian noise, using the different dimensionality reduction methods.

Metrics	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLLDA
OA	44.91 ± 2.58	74.41 ± 0.22	80.02 ± 0.82	78.75 ± 1.03	61.56 ± 1.50	77.19 ± 2.12	<b>80.13 ± 0.58</b>
AA	12.11 ± 1.85	68.08 ± 1.25	68.87 ± 2.11	66.53 ± 2.66	43.22 ± 2.85	67.05 ± 5.68	<b>72.44 ± 2.15</b>
$\kappa$	3.30 ± 6.21	65.76 ± 0.27	72.59 ± 1.21	70.77 ± 1.50	46.25 ± 3.06	69.39 ± 2.85	<b>73.35 ± 0.76</b>



**Table 7.** Comparison of the OA (%) classification results of the SVM classifiers after processing the Heihe dataset corrupted with Gaussian noise, using the different dimensionality reduction methods.

Metrics	LDA	LPP	RSLDA	TRLDA	LRPER	$L_{2,p}$ -RER	JSLDA
OA	40.23 ± 6.16	80.37 ± 2.31	84.93 ± 1.68	78.15 ± 1.76	65.13 ± 1.64	76.68 ± 2.64	<b>86.49 ± 1.28</b>
AA	41.82 ± 4.56	79.34 ± 1.64	86.25 ± 1.15	76.04 ± 1.45	65.49 ± 0.39	71.58 ± 1.86	<b>83.85 ± 0.82</b>
$\kappa$	28.49 ± 7.09	74.25 ± 2.82	80.09 ± 2.11	71.31 ± 2.13	56.08 ± 1.74	69.22 ± 3.23	<b>82.01 ± 1.63</b>

### 4.3. Ablation Study

In this subsection, the impact of the regression regularization (RR), joint sparsity (JS), and locality-preserving regularization (LP) terms on the performance of the proposed JSLDA model was analyzed via ablation experiments. Notably, when JSLDA was stripped of these regularization terms (RR, JS, and LP), it was reduced to the classical LDA model, which served as the baseline for comparison. We evaluated the dimensionality reduction models using OA, AA, and the  $\kappa$  coefficient. The results of the ablation study are presented in Tables 8–10, where ‘✓’ indicates that the model contains the corresponding component and ‘✗’ indicates that it does not.

As is evident from Table 8, the introduction of the RR term enhanced the baseline model in terms of OA, AA, and  $\kappa$  by 3.45%, 3.48%, and 3.68%, respectively. This demonstrates that the introduction of regression allowed the orthogonal matrix to better preserve the label information in the low-dimensional data. Additionally, the orthogonal components generated more discriminative information, which enhanced the model’s resistance to external noise.

Furthermore, upon integrating the JS and LP terms into the baseline + RR model (i.e., baseline + RR + JS and baseline + RR + LP), we observed OA improvements of 4.69% and 6.16%, respectively, compared to the baseline+RR model. The JS term accentuates discriminative features in high-dimensional data by assigning larger weights to salient features, while suppressing less important ones, resulting in a more discriminative information extraction. Meanwhile, the LPP-R term preserves the neighborhood relationships in the original data, ensuring that the projection matrix retains the primary energy of the raw data, thereby enhancing the discriminability of the low-dimensional space.

Crucially, the model that incorporated all three regularization terms—RR, JS, and LPP-R (our proposed JSLDA)—achieved the optimal classification performance, with the OA, AA, and kappa coefficient reaching 91.24%, 95.07%, and 90.22%, respectively. These values marked improvements of 8.73%, 9.55%, and 9.65% over the baseline model. Similar trends were observed on the University of Pavia dataset, where the integration of RR, JS, and LPP into the baseline model led to significant enhancements in OA, AA, and  $\kappa$  by 13.85%, 16.49%, and 19.29%, respectively. Analogous conclusions held true for the Heihe dataset.

In summary, we can confidently conclude that the inclusion of RR, JS, and LP terms played a pivotal role in enhancing the discriminative feature extraction capabilities of JSLDA. Each term contributes uniquely and synergistically to achieving a superior classification performance.

**Table 8.** Ablation experiments on the Salinas dataset.

Baseline	RR	JS	LPP-R	OA	AA	$\kappa$
✓	✗	✗	✗	82.51 ± 0.97	85.52 ± 1.12	80.57 ± 1.07
✓	✓	✗	✗	85.96 ± 1.05	89.00 ± 1.57	84.25 ± 1.21
✓	✓	✓	✗	90.65 ± 0.42	94.75 ± 0.13	89.58 ± 0.47
✓	✓	✗	✓	88.67 ± 0.60	93.13 ± 0.47	87.35 ± 0.68
✓	✓	✓	✓	91.24 ± 0.35	95.07 ± 0.15	90.22 ± 0.39

**Table 9.** Ablation experiments on the University of Pavia dataset.

Baseline	RR	JS	LPP-R	OA	AA	$\kappa$
✓	✗	✗	✗	73.79 ± 1.79	65.80 ± 3.20	64.20 ± 2.57
✓	✓	✗	✗	76.23 ± 0.75	66.28 ± 1.64	68.15 ± 1.01
✓	✓	✓	✗	87.19 ± 0.64	82.41 ± 1.55	82.86 ± 0.86
✓	✓	✗	✓	77.51 ± 0.83	68.45 ± 1.39	69.79 ± 1.09
✓	✓	✓	✓	87.64 ± 0.94	82.29 ± 2.21	83.49 ± 1.25

**Table 10.** Ablation experiments on the Heihe dataset.

Baseline	RR	JS	LPP-R	OA	AA	$\kappa$
✓	✗	✗	✗	71.48 ± 5.69	70.56 ± 3.75	63.40 ± 6.46
✓	✓	✗	✗	72.58 ± 3.86	67.33 ± 4.20	64.21 ± 4.82
✓	✓	✓	✗	90.45 ± 2.33	90.91 ± 0.97	87.16 ± 2.98
✓	✓	✗	✓	71.57 ± 4.47	66.67 ± 2.60	63.06 ± 5.21
✓	✓	✓	✓	92.32 ± 0.67	92.16 ± 0.96	89.63 ± 0.89

#### 4.4. Discussion

The classification results across the three hyperspectral datasets demonstrated that our proposed method consistently achieved superior performance. Notably, the representative methods RSLDA and JSLDA, which extend LDA by imposing an  $L_{2,1}$  constraint on the projection matrix, generally outperformed both LDA and TR LDA. This improvement was due to the  $L_{2,1}$  constraint, which enforces row sparsity in the projection matrix, thereby giving greater weight to important features and enhancing the discriminative power of the acquired features. Furthermore, regression-based methods such as LRPER,  $L_{2,p}$ -RER, and JSLDA tend to surpass traditional models like LDA and LPP. This advantage arises because regression-based models account for the interplay between projection and regression, with the orthogonal matrix providing the classifier with more comprehensive feature information, significantly improving performance. Among the algorithms compared, only LPP focuses exclusively on local information, while LDA, RSLDA, TR LDA, LRPER, and  $L_{2,p}$ -RER emphasize global information. In contrast, our proposed method integrates both local and global information, which was a crucial factor in JSLDA's superior classification performance.

Furthermore, the complexity analysis demonstrated the feasibility of our approach in practical scenarios. However, our proposed method still has some shortcomings. The ablation experiments also indirectly explained the advantages of the model. Specifically, the RR term establishes a link between projection and regression, aiding the projection in extracting more discriminative low-dimensional features. The LP term ensures that neighborhood information in the high-dimensional space is retained in the low-dimensional space. Additionally, by imposing an  $L_{2,1}$  constraint on the projection matrix, the model captures joint sparse information, allowing the low-dimensional space to preserve more of the essential information from the original data. The sparse error term further enhances the robustness of the model. In the pre-processing process, three-dimensional hyperspectral image data are converted into two-dimensional data and input into JSLDA for processing, which destroys the original spatial structure of the hyperspectral image data and loses certain spatial information. In future work, we hope to continue to investigate how to directly use high-dimensional data as an input to preserve more original information.

Although hyperspectral images are highly effective in capturing the spectral characteristics of objects, they are also prone to interference from external noise. Our proposed robust feature extraction method addresses this issue by mitigating the impact of varying environmental conditions, such as changes in lighting and shadows, thereby enhancing the accuracy of crop classification in agricultural production. Similarly, in mineral exploration,

where the presence of irrelevant spectral bands and complex surface cover can introduce noise, our method improves the detection and classification accuracy of mineral types.

## 5. Conclusions

In this study, a robust LDA-based dimensionality reduction method called JSLLDA was proposed. In JSLLDA, a locality-preserving regularization term was introduced to capture the locality structure of data, while a regression variant was utilized to fully exploit the interaction between regression and projection. Moreover, the model's robustness was further enhanced using a  $L_{2,1}$ -norm constraint, which is imposed on the regression loss. Additionally, the interpretability of the projections was improved through a  $L_{2,1}$ -norm constrained projection matrix. Furthermore, an alternating iterative algorithm was designed to optimize the proposed JSLLDA model, and its time complexity was analyzed theoretically and experimentally. To evaluate the performance of JSLLDA, experiments were conducted on three public hyperspectral datasets. To further assess the robustness of JSLLDA, Gaussian noise of varying concentrations was randomly added to the datasets for testing. Compared with several related dimensionality reduction methods, the results demonstrated that JSLLDA was more effective in extracting distinctive low-dimensional features for hyperspectral image classification. Finally, to validate the contribution of each component of the model, a series of ablation experiments were performed, confirming that each part was beneficial for JSLLDA.

**Author Contributions:** Methodology, C.-Y.C.; Software, M.-T.L.; Validation, Y.-J.D. and L.R.; Investigation, M.-T.L. and Y.L.; Writing—original draft, C.-Y.C. and Y.-J.D.; Writing—review and editing, L.R., Y.L. and X.-H.Z.; Visualization, M.-T.L. and Y.L.; Funding acquisition, X.-H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 62401203 and Grant 62401546, in part by the Hunan Provincial Key Research and Development Program under Grant 2023NK2011 and Grant 2023NK2002, and in part by the Hunan Provincial Natural Science Foundation of China under Grant 2022JJ40189.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Li, M.; Luo, Q.; Liu, S. Application of hyperspectral imaging technology in quality inspection of agricultural products. In Proceedings of the 2022 International Conference on Computers, Information Processing and Advanced Education (CIPAE), Ottawa, ON, Canada, 26–28 August 2022; pp. 369–372.
- Manohar Kumar, C.V.S.; Nidamanuri, R.R.; Dadhwal, V.K. Sub-pixel discrimination of soil and crop in drone-based hyperspectral imagery. In Proceedings of the 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Athens, Greece, 31 October–2 November 2023; pp. 1–4.
- Tang, X.J.; Liu, X.; Yan, P.F.; Li, B.X.; Qi, H.Y.; Huang, F. An MLP network based on residual learning for rice hyperspectral data classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6007405. [[CrossRef](#)]
- Liang, J.; Li, P.; Zhao, H.; Han, L.; Qu, M. Forest species classification of UAV hyperspectral image using deep learning. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 7126–7130.
- Singh, L.; Mutanga, O.; Mafongoya, P.; Peerbhay, K.; Crous, J. Hyperspectral remote sensing for foliar nutrient detection in forestry: A near-infrared perspective. *Remote Sens. Appl. Soc. Environ.* **2022**, *25*, 100676. [[CrossRef](#)]
- Nisha, A.; Anitha, A. Current advances in hyperspectral remote sensing in urban planning. In Proceedings of the 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT), Kannur, India, 11–12 August 2022; pp. 94–98.
- Ferreira, M.P.; Martins, G.B.; de Almeida, T.M.H.; da Silva Ribeiro, R.; da Veiga Júnior, V.F.; da Silva Rocha Paz, I.; de Siqueira, M.F.; Kurtz, B.C. Estimating aboveground biomass of tropical urban forests with UAV-borne hyperspectral and LiDAR data. *Urban For. Urban Green.* **2024**, *96*, 128362. [[CrossRef](#)]
- Li, X.; Wang, L.; Guan, H.; Chen, K.; Zang, Y.; Yu, Y. Urban tree species classification using UAV-based multispectral images and LiDAR point clouds. *J. Geovis. Spat. Anal.* **2024**, *8*, 5. [[CrossRef](#)]

9. Zhi, Z.; Liu, J.; Liu, J.; Li, A. Geospatial structure and evolution analysis of national terrestrial adjacency network based on complex network. *J. Geovis. Spat. Anal.* **2024**, *8*, 12. [[CrossRef](#)]
10. Li, Z.; Yang, X.; Meng, D.; Cao, X. An adaptive noisy label-correction method based on selective loss for hyperspectral image-classification problem. *Remote Sens.* **2024**, *16*, 2499. [[CrossRef](#)]
11. Durojaiye, A.I.; Olorunsogo, S.T.; Adejumo, B.A.; Babawuya, A.; Muhamad, I.I. Deep learning techniques for the exploration of hyperspectral imagery potentials in food and agricultural products. *Food Humanit.* **2024**, *3*, 100365. [[CrossRef](#)]
12. Ren, L.; Hong, D.; Gao, L.; Sun, X.; Huang, M.; Chanussot, J. Orthogonal subspace unmixing to address spectral variability for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5501713. [[CrossRef](#)]
13. Su, Y.; Jiang, M.; Gao, L.; You, X.; Sun, X.; Li, P. Graph-cut-based node embedding for dimensionality reduction and classification of hyperspectral remote sensing images. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 1720–1723.
14. Uddin, M.P.; Mamun, M.A.; Hossain, M.A. PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech. Rev.* **2021**, *38*, 377–396. [[CrossRef](#)]
15. 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](#)]
16. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
17. Wang, C.; Fu, W.; Huang, H.; Chen, J. Isomap-based three-dimensional operational modal analysis. *Sci. Program.* **2020**, *2020*. [[CrossRef](#)]
18. Wang, A.; Zhao, S.; Liu, J.; Yang, J.; Liu, L.; Chen, G. Locality adaptive preserving projections for linear dimensionality reduction. *Expert Syst. Appl.* **2020**, *151*, 113352. [[CrossRef](#)]
19. Qiao, Z.; Zhou, L.; Huang, J.Z. Sparse linear discriminant analysis with applications to high dimensional low sample size data. *IAENG Int. J. Appl. Math.* **2009**, *39*, 48–60.
20. Cui, H.; Deng, Y.; Zhong, R.; Li, W.; Yu, C.; Danyushevsky, L.V.; Belousov, I.; Li, Z.; Wang, H. Determining the ore-forming processes of Dongshengmiao Zn-Pb-Cu deposit: Evidence from the linear discriminant analysis of pyrite geochemistry. *Ore Geol. Rev.* **2023**, *163*, 105782. [[CrossRef](#)]
21. Ye, J.; Xiong, T. Null space versus orthogonal linear discriminant analysis. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 1073–1080.
22. Zhang, T.; Tao, D.; Yang, J. Discriminative locality alignment. In Proceedings of the Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008; Proceedings, Part I 10; Springer: Berlin/Heidelberg, Germany, 2008; pp. 725–738.
23. Wang, J.; Wang, L.; Nie, F.; Li, X. A novel formulation of trace ratio linear discriminant analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 5568–5578. [[CrossRef](#)]
24. Zhu, F.; Gao, J.; Yang, J.; Ye, N. Neighborhood linear discriminant analysis. *Pattern Recognit.* **2022**, *123*, 108422. [[CrossRef](#)]
25. Kwak, N. Principal Component Analysis Based on L1-Norm Maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1672–1680. [[CrossRef](#)]
26. Wang, R.; Nie, F.; Yang, X.; Gao, F.; Yao, M. Robust 2DPCA With Non-greedy  $\ell_1$ -Norm Maximization for Image Analysis. *IEEE Trans. Cybern.* **2015**, *45*, 1108–1112. [[CrossRef](#)]
27. Lu, Y.; Yuan, C.; Lai, Z.; Li, X.; Wong, W.K.; Zhang, D. Nuclear norm-based 2DLPP for image classification. *IEEE Trans. Multimed.* **2017**, *19*, 2391–2403. [[CrossRef](#)]
28. Zhang, Z.; Li, F.; Zhao, M.; Zhang, L.; Yan, S. Robust neighborhood preserving projection by nuclear/ $L_{2,1}$ -norm regularization for image feature extraction. *IEEE Trans. Image Process.* **2017**, *26*, 1607–1622. [[CrossRef](#)]
29. Gu, J.; Cai, Q.; Gong, W.; Wang, H. L21-Norm-Based Common Spatial Pattern with Regularized Filters. In Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 18–20 June 2021; Volume 4, pp. 1746–1751.
30. Nie, F.; Tian, L.; Huang, H.; Ding, C. Non-greedy L21-norm maximization for principal component analysis. *IEEE Trans. Image Process.* **2021**, *30*, 5277–5286. [[CrossRef](#)]
31. Zhang, T.; Long, C.F.; Deng, Y.J.; Wang, W.Y.; Tan, S.Q.; Li, H.C. Low-rank preserving embedding regression for robust image feature extraction. *IET Comput. Vis.* **2024**, *18*, 124–140. [[CrossRef](#)]
32. Deng, Y.J.; Yang, M.L.; Li, H.C.; Long, C.F.; Fang, K.; Du, Q. Feature dimensionality reduction with  $L_{2,p}$ -norm-based robust embedding regression for classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5509314. [[CrossRef](#)]
33. Nie, F.; Huang, H.; Cai, X.; Ding, C. Efficient and robust feature selection via joint  $L_{2,1}$ -norms minimization. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 1813–1821.
34. Yang, Y.; Shen, H.T.; Ma, Z.; Huang, Z.; Zhou, X. L2,1-norm regularized discriminative feature selection for unsupervised learning. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence—Volume Volume Two, Catalonia, Spain, 16–22 July 2011; IJCAI'11; AAAI Press: Barcelona, Spain, 2011; pp. 1589–1594.
35. Wen, J.; Fang, X.; Cui, J.; Fei, L.; Yan, K.; Chen, Y.; Xu, Y. Robust sparse linear discriminant analysis. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 390–403. [[CrossRef](#)]

36. Tang, C.; Liu, X.; Zhu, X.; Xiong, J.; Li, M.; Xia, J.; Wang, X.; Wang, L. Feature selective projection with low-rank embedding and dual Laplacian regularization. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1747–1760. [[CrossRef](#)]
37. Li, C.N.; Li, Y.; Meng, Y.H.; Ren, P.W.; Shao, Y.H.  $L_{2,1}$ -norm regularized robust and sparse linear discriminant analysis via an alternating direction method of multipliers. *IEEE Access* **2023**, *11*, 34250–34259. [[CrossRef](#)]
38. Long, C.F.; Wen, Z.D.; Deng, Y.J.; Hu, T.; Liu, J.L.; Zhu, X.H. Locality preserved selective projection learning for rice variety identification based on leaf hyperspectral characteristics. *Agronomy* **2023**, *13*, 2401. [[CrossRef](#)]
39. Liu, Q.; Huang, X.; Bai, R. Bayesian modal regression based on mixture distributions. *Comput. Stat. Data Anal.* **2024**, *199*, 108012. [[CrossRef](#)]
40. Lam, B.S.; Choy, S.; Yu, C.K. Linear discriminant analysis with trimmed and difference distribution modeling. *Knowl.-Based Syst.* **2024**, *299*, 112093. [[CrossRef](#)]
41. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
42. Yang, J.; Yin, W.; Zhang, Y.; Wang, Y. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sci.* **2009**, *2*, 569–592. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.