*Article*

# Hyperspectral Remote Sensing Image Classification Based on Partitioned Random Projection Algorithm

**Shuhan Jia, Quanhua Zhao * and Yu Li**

The Institute for Remote Sensing Science and Application, School of Geomatics, Liaoning Technical University, Fuxin 123000, China; 471810035@lntu.edu.cn (S.J.); liyu@lntu.edu.cn (Y.L.)
* Correspondence: zhaoquanhua@lntu.edu.cn; Tel.: +86-138-9854-5105

**Abstract:** Dimensionality reduction based on random projection (RP) includes two problems, namely, the dimensionality is limited by the data size and the class separability of the dimensionality reduction results is unstable due to the randomly generated projection matrix. These problems make the RP algorithm unsuitable for large-size hyperspectral image (HSI) classification. To solve these problems, this paper presents a new partitioned RP (PRP) algorithm and proves its rationality in theory. First, a large-size HSI is evenly divided into multiple small-size sub-HSIs. Afterwards, the projection matrix that maximizes the class separability is selected from multiple samplings in which the class dissimilarity measurement is defined as large inter-class distance and small intra-class variance. By using the same projection matrix, each small-size sub-HSI is projected to generate a low dimensional sub-HSI, thereby generating a low dimensional HSI. Next, the minimum distance (MD) classifier is utilized to classify the low dimensional HSI obtained by the PRP algorithm. Finally, four real HSIs are used for experiments, and three of the most popular classification algorithms based on RP are selected as comparison algorithms to validate the effectiveness of the proposed algorithm. The classification performance is evaluated with the kappa coefficient, overall accuracy (OA), average accuracy (AA), average precision rate (APR), and running time. Experimental results indicate that the proposed algorithm can obtain reliable classification results in a very short time.

**Keywords:** random projection; minimum distance classifier; hyperspectral remote sensing image classification; dimensionality reduction

## 1. Introduction

The wavelength range of hyperspectral images (HSIs) is continuous and dense from visible light to short-wave infrared [1]. Compared with panchromatic images and multi-spectral remote sensing images, HSIs contain richer spectral information and finer texture information of ground objects, which provide a solid data foundation for high-precision ground object classification [2,3]. Therefore, more and more researchers apply it to the fields of medicine [4,5] and agriculture [6,7]. However, for commonly configured hardware, the classification of large-size HSIs has some problems, such as long running time and a large amount of data. The most straightforward solution is to perform dimensionality reduction before HSI classification, thereby reducing the computational load through dimensionality reduction [8,9].

Popular HSI dimensionality reduction methods are roughly divided into two types: band selection [10,11] and feature extraction [12,13]. The band selection method directly discards most of the bands and only selects a part of the bands for subsequent HSI analysis. Currently, Zhang et al. [14] proposed a marginalized graph self-representation (MGSR) method for unsupervised hyperspectral band selection, which considers the differences between different homogenous regions. First, super-pixel segmentation is used to construct the structure map of HSI [15]. Additionally, considering the relationship between adjacent pixels of the same segmentation, the damage is marginalized by an alternating optimization

algorithm to search for the best band selection scheme. Computation of the graph matrix takes a lot of time, and the result of super-pixel segmentation affects the result of subsequent band selection. Mou et al. [16] studied deep reinforcement learning models for HSI analysis and first proposed an automatic learning strategy for band selection under the framework of the Markov decision. The method first used a parametric method to rank the bands and then utilized a deep reinforcement learning model to select the best subset of bands. Because it is a learning-based algorithm, it takes much time during the training phase to explore efficient band selection strategies. Feature extraction generally uses mapping to linearly transform all bands of HSIs. Tamilarasi et al. [17] proposed a new dimensionality reduction and classification technique, which extracted and classified road and building features from HSIs with high accuracy. This technique combined independent component analysis (ICA) [18], principal component analysis (PCA) [19], fully convolutional network (FCN) [20], and support vector machine (SVM) models [21]. The above dimensionality reduction method is cumbersome and cannot be calculated quickly.

The random projection (RP) algorithm is an emerging dimensionality reduction technique that satisfies the Johnson-Lindenstrauss (JL) lemma [22–24]. The algorithm does not require any prior knowledge of the original data [25]. In addition, a randomly generated projection matrix is used to achieve dimensionality reduction, where the column vector of the projection matrix is unit length. This makes the RP algorithm easier to implement and more convenient to calculate [26]. Additionally, the RP algorithm can guarantee that the distance between any pair of vectors before and after projection is kept within a small range [27,28]. Because many machine learning algorithms use the distance information between vectors to operate and execute [29,30], the dataset obtained after dimensionality reduction by the RP algorithm can be well applied to machine learning algorithms. Therefore, more and more researchers apply it to the dimensionality reduction of HSI data.

At present, many researchers interested in the field of the RP algorithm have proposed many classification algorithms based on the RP algorithm [31,32]. These methods can be grouped into three types: combined feature extraction method [33], separability-boosting method [34], and ensemble method [35]. Combined feature extraction methods combine RP and other feature extraction methods. Zhao and Mao [36] proposed a semi-random projection method, which uses linear discriminant analysis (LDA) to calculate each column vector in the projection matrix and calculates the projection matrix by repeating it multiple times. It strikes a good balance between computational complexity and classification accuracy. This method can effectively preserve the distance invariance of the data when dealing with high-dimensional data. Deegalla and Bostrom [37] compared the dimensionality reduction effect of PCA and the RP algorithm and combined PCA and RP to propose a nearest-neighbor classification algorithm. This method not only improves efficiency but also enables more efficient nearest-neighbor classification. However, the method cannot achieve dimensionality reduction when dealing with large-size HSIs. Separability-boosting methods preserve the class separability of the original data during projection. Zhao et al. [38] proposed a new tighter RP with minimal intra-class variance (TRP-MIV) algorithm for HSI classification, which supplied a promising way for HSI dimensionality reduction. Nevertheless, when the number of hyperspectral vectors is very large, the dimensionality reduction effectiveness of the TRP-MIV algorithm is still limited. The ensemble methods combine RP and the fuzzy clustering algorithm. Rathore et al. [39] studied the cumulative agreement fuzzy c-means (CAFCM) algorithm, which uses RP for generating multiple dimensionality reduction results. In addition, Anderlucci et al. [40] proposed a model-based high dimensional data clustering method based on the idea of ensemble methods. The method first generates a set of low dimensional independent projection results and performs model-based clustering on each of them. Then, the projection of the best grouping structure is displayed, and the final partition is obtained by aggregating the clusters found in the projection by consensus. However, the dimensionality reduction results produced by each projection are only guaranteed to be like the original data and cannot be applied to classification tasks. The RP algorithm has the following shortcom-

ings in large-size HSI classification. (1) The lowest projection dimensionality of the RP algorithm is independent of the high dimensionality, which is positively correlated with the number of hyperspectral vectors [41]. The larger number of hyperspectral vectors, the larger the lowest projection dimensionality. (2) Different dimensionality reduction results will be generated by different projection matrices, which lead to the class separability of dimensionality reduction results being very unstable. Thus, the RP algorithm cannot play a dimensionality reduction effect on the HSI with a large number of hyperspectral vectors, which is not conducive to large-size HSI classification.

Aiming at the above problems existing in the RP algorithm in large-scale dimensionality reduction and classification, an HSI classification algorithm based on partitioned RP (PRP) is proposed. Firstly, the PRP algorithm is proposed by means of image division, and the distance preservation of the hyperspectral vector pair after partitioning is still satisfied from two perspectives within each sub-HSI and between any two different sub-HSIs. In addition, a classifier based on the PRP algorithm is designed, in which the optimal projection matrix is determined according to the principle of large inter-class distance and small intra-class variance, and minimum distance (MD) [42,43] is used to classify the low dimensional HSI received by the PRP algorithm. The experimental results show that the PRP algorithm can project the large-size HISs into a subspace with a lower dimensionality than the RP algorithm, and the proposed classification algorithm can obtain good classification results. The structure of this paper is organized as below. Sections 2 and 3 give the materials and the proposed algorithm, respectively. The experimental results and discussion are provided in Sections 4 and 5. Finally, this paper is concluded in Section 6.

## 2. Materials

This paper uses four publicly available datasets with validation data, namely, real Pavia Centre, Salinas, Chikusei, and LongKou [44,45] images. The four images are respectively located in Pavia Centre in northern Italy, Salinas in Northwestern Indiana, Chikusei in Japan, and LongKou in China. Details of these images are listed in Table 1, including download address, year, sensor, spatial resolution, and number of bands. Spatial resolution represents the actual distance on the ground represented by a pixel.

**Table 1.** Details of four real HSIs.

|  | **Pavia Centre** | **Salinas** | **Chikusei** | **LongKou** |
|---|---|---|---|---|
| URL for data source | http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (accessed on 12 July 2021) | http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes (accessed on 12 July 2021) | https://naotoyokoya.com/Download.html (accessed on 11 April 2016) | http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm (accessed on 1 October 2020) |
| Year | 2003 | 1998 | 2014 | 2018 |
| Sensor | ROSIS | AVIRIS | Headwall Hyperspec-VNIR-C | Nano-Hyperspec |
| Spatial resolution | 1.3 m | 3.7 m | 2.5 m | 0.463 m |
| Number of bands | 102 | 204 | 128 | 270 |

It is worth noting that this paper uses HSIs of sizes $1096 \times 531$, $512 \times 127$, $1000 \times 800$, and $550 \times 400$, with homogeneous regions of 9, 13, 7, and 9, respectively, as shown in Figure 1. The classes included in the Pavia Centre image are water, trees, meadows, self-blocking bricks, bare soil, asphalt, bitumen, tiles, and shadows. The classes included in the Salinas image are broccoli green weeds2, fallow, fallow rough plow, fallow smooth, stubble, celery, grapes untrained, soil vineyard develop, corn senesced weeds, lettuce romaine 4 week, lettuce romaine 5 week, lettuce romaine 6 week, and lettuce romaine 7 week. The classes included in the Chikusei image are water, bare soil (school), bare soil (farmland), natural plants, glass, rice field (grown), and row crops. The classes included

in the LongKou image are corn, cotton, sesame, broad leaf soybean, narrow leaf soybean, rice, water, roads and houses, and mixed weed. Specifically, Figure 1 shows the false-color images, the standard images (i.e., validation data), and the legends. The band combinations used to replace the RGB bands in the false-color images of the four images are (49, 31, 15), (29, 20, 12), (54, 35, 22) and (130, 65, 18), respectively. Moreover, the specific classes are explained in the legends. In addition, because many vectors in these images do not contain any valid information, these vectors are discarded before applying. Therefore, the numbers of spectral vectors in these scenes are 109,794, 20,655, 9435, and 204,542, respectively.
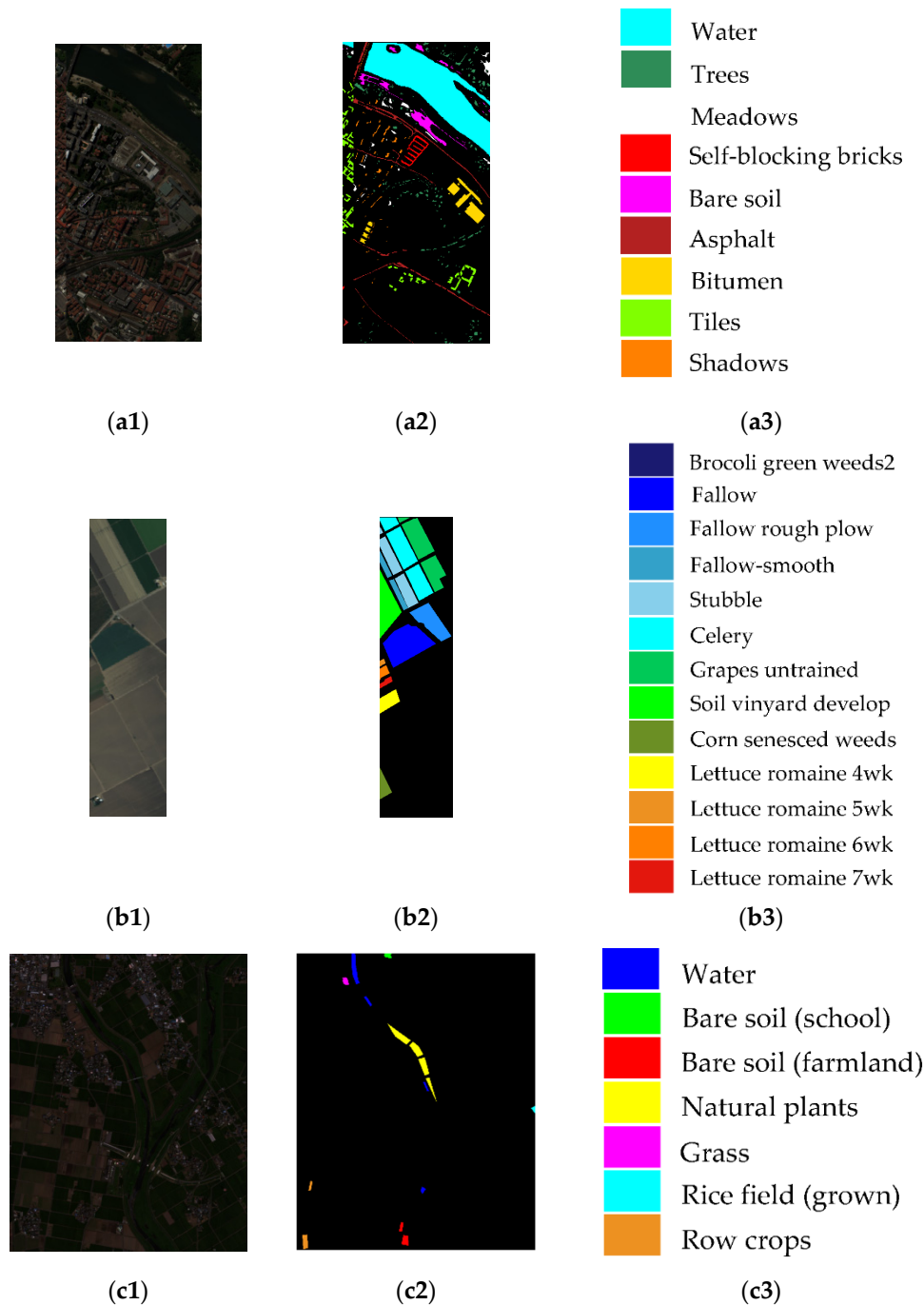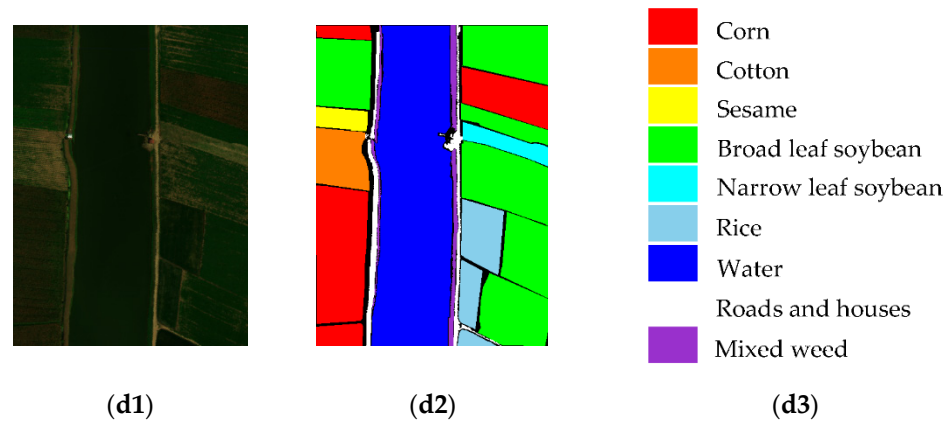


(**a1**)　　(**a2**)　　(**a3**)

Legend (a3):
- Water
- Trees
- Meadows
- Self-blocking bricks
- Bare soil
- Asphalt
- Bitumen
- Tiles
- Shadows

(**b1**)　　(**b2**)　　(**b3**)

Legend (b3):
- Brocoli green weeds2
- Fallow
- Fallow rough plow
- Fallow-smooth
- Stubble
- Celery
- Grapes untrained
- Soil vinyard develop
- Corn senesced weeds
- Lettuce romaine 4wk
- Lettuce romaine 5wk
- Lettuce romaine 6wk
- Lettuce romaine 7wk

(**c1**)　　(**c2**)　　(**c3**)

Legend (c3):
- Water
- Bare soil (school)
- Bare soil (farmland)
- Natural plants
- Grass
- Rice field (grown)
- Row crops

**Figure 1.** *Cont.*

**Figure 1.** Real HSIs. (**a1**–**d1**) represent the false-color images of the Pavia Centre, Salinas, Chikusei, and LongKou images, respectively. (**a2**–**d2**) represent the standard classified images (i.e., validation data) of four real HSIs, respectively. (**a3**–**d3**) represent the legends for all classes of four real HSIs, respectively.

## 3. The Proposed Algorithm

### 3.1. Random Projection

RP is an effective dimensionality reduction tool that can keep the distance between vector pairs before and after projection unchanged. The usual version of the RP algorithm is defined below [26].

**Theorem 1.** *Supposing that* $A$ *is any* $S \times D$ *matrix, where* $S$ *is the data size and* $D$ *is the feature dimensionality. Its row vector is* $a_s$ *in* $D$ *dimensional spaces, where* $s$ *is the index of row vectors and* $s = 1, \dots , S$. *For constants* $\varepsilon$ *and* $\beta > 0$, *an integer* $K_{RP}$ *is chosen as follows,*

$$K_{\text{RP}} \geq K_{\text{RP}}^0 = \frac{4 + 2\beta}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}} \log S \tag{1}$$

*where* $K^0{}_{RP}$ *is the lowest projection dimensionality of the RP algorithm.* $R_{RP}$ *is any* $D \times K_{RP}$ *matrix whose entry belongs to the standard normal Gaussian distribution. The* $S \times K_{RP}$ *matrix*

$$B = \frac{1}{\sqrt{K_{\text{RP}}}} A R_{\text{RP}} \tag{2}$$

*is the projection of* $A$ *in* $K_{RP}$ *dimensional subspaces. The row vector of matrix* $B$ *is* $b_s$. *Then, with the probability of no distortion at least*

$$P_{\text{RP}} = 1 - S^{-\beta} \tag{3}$$

*for any two row vectors* $a_s$ *and* $a_{s'}$ *of matrix* $A$, *the distance of the row vectors* $b_s$ *and* $b_{s'}$ *of matrix* $B$ *is preserved as follows,*

$$(1 - \varepsilon) \, ||a_s - a_{s'}||^2 \leq ||b_s - b_{s'}||^2 \leq (1 + \varepsilon) \, ||a_s - a_{s'}||^2 \tag{4}$$

*where* $s'$ *is the index of row vectors.*

It can be seen from the above theorem that RP is a computationally simple method. The order of forming a projection matrix $R_{\text{RP}}$ and projecting the high dimensional data $A$ into the $K_{\text{RP}}$ dimensional subspaces is $O(D \times K_{\text{RP}} \times S)$. If the number of vectors $S$ becomes larger, the lowest projection dimensionality $K^0{}_{\text{RP}}$ will become larger according to Equation (1) and the probability of no distortion $P_{\text{RP}}$ will also become larger according to Equation (3). Therefore, when the number of vectors is large, although the probability of no distortion is high, the low dimensionality $K_{\text{RP}}$ may be higher than the high dimensionality $D$, so the

dimensionality reduction effect cannot be achieved. Besides, the generation of the projection matrix is completely random, which will lead to instability in the class separability of the generated dimensionality reduction result. In simpler terms, it cannot guarantee that the class separability of the dimensionality reduction result is good for the classification task.

*3.2. Partitioned Random Projection*

The lowest projection dimensionality $K^0_{RP}$ of the RP algorithm is strictly limited by the parameters $\varepsilon$, $\beta$, and the number of hyperspectral vectors $S$. The Pavia Centre image with 102 bands is taken as an example, of which the number of hyperspectral vectors is 109,794. When the fixed parameters $\varepsilon = 1$ and $\beta = 0.5$, according to Equation (1), the lowest projection dimensionality of the RP algorithm is 349, which is much higher than the number of bands, which is 102. Therefore, when processing large-size HSIs, if the number of bands is very low, the RP algorithm will not be able to achieve the dimensionality reduction of large-size HSIs. Considering this problem, a new PRP algorithm is introduced in this section. PRP, a dimensionality reduction algorithm, can project a large-size HSI into a low dimensional subspace without serious distortion on pairwise distance.

The HSI can be expressed as a matrix $U = [u_1; \ldots ; u_s; \ldots ; u_S]$, where $S$ is the number of hyperspectral vectors, $s$ is the index of hyperspectral vectors, and $u_s$ is the $s$th hyperspectral vector. The implementation of the PRP algorithm involves partitioning an HSI into multiple sub-HSIs evenly according to the horizontal direction and then projecting each sub-HSI into a low dimensional subspace. Specifically, the matrix of the HSI can also be defined as $U = [U^1; \ldots ; U^m; \ldots ; U^M]$, where $M$ is the number of sub-matrices, and $m$ is the index of sub-matrices. Moreover, $U^m = [u^m_1; \ldots ; u^m_n; \ldots ; u^m_N]$ is the $m$th $N \times D$ sub-matrix, where $n$ is the index of hyperspectral vectors in the sub-matrix, and $N$ is the number of hyperspectral vectors in the sub-matrix that is the same for all sub-matrices. Thus, $S = M \times N$. Note that the dimensionality reduction method in this paper requires that an HSI must be divided equally. Otherwise, the first few hyperspectral vectors can be regarded as edges and discarded so that the HSI can be divided equally.

Next, the rationality of the distance preservation property guaranteed by this partition method is proven from two perspectives: within each sub-matrix and between any two different sub-matrices.

The distance changes before and after the projection of any two hyperspectral vectors within each sub-matrix are explained as follows. The conclusion of this theorem comes from the RP algorithm (Theorem 1).

**Theorem 2.** *For constants $\varepsilon$ and $\beta > 0$, an integer $K_{PRP}$ such that*

$$K_{\text{PRP}} \geq K^0_{\text{PRP}} = \frac{4 + 2\beta}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}} \log N = \frac{4 + 2\beta}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}} \log \frac{S}{M} \tag{5}$$

*where $K^0_{PRP}$ is the lowest projection dimensionality of the proposed algorithm. $R_{PRP}$ is a $D \times K_{PRP}$ matrix whose entry belongs to the standard normal Gaussian distribution. The $N \times K_{PRP}$ sub-matrix*

$$V^m = \frac{1}{\sqrt{K_{\text{PRP}}}} U^m R_{\text{PRP}} \tag{6}$$

*is the projection of sub-matrix $U^m$ in $K_{PRP}$ dimensional subspaces. The low dimensional hyperspectral vector of sub-matrix $V^m$ is $v^m_n$. Namely, $V^m = \{v^m_1; \ldots ; v^m_n; \ldots ; v^m_N\}$. Then, according to Equation (3), with the probability of no distortion at least*

$$P_{\text{PRP1}} = 1 - N^{-\beta} = 1 - (S/M)^{-\beta} \tag{7}$$

*for any two hyperspectral vectors $u^m_n$ and $u^m_{n'}$ of sub-matrix $U^m$, the distance of the two low dimensional hyperspectral vectors $v^m_n$ and $v^m_{n'}$ of sub-matrix $V^m$ is preserved based on Equation (4),*

$$(1 - \varepsilon) \, ||u^m_n - u^m_{n'}||^2 \leq ||v^m_n - v^m_{n'}||^2 \leq (1 + \varepsilon) \, ||u^m_n - u^m_{n'}||^2 \tag{8}$$

*where n′ is the index of hyperspectral vectors in the sub-matrix.*

According to Equation (5), the lowest projection dimensionality $K^0_{PRP}$ decreases with the increase in the number of sub-matrices $M$ and the decrease of parameter $\beta$. Moreover, with the increase of parameter $\varepsilon$, the change trend of lowest projection dimensionality $K^0_{PRP}$ is to first decrease and then increase. Comparing Equation (1) and Equation (5), the lowest projection dimensionality of the PRP algorithm is smaller than that of the RP algorithm under the same parameters. Note that when $M$ is equal to one, it is the projection without partitioning, that is, the RP algorithm. When $M$ is equal to $S$, there is only one hyperspectral vector in each sub-matrix. Meanwhile, the lowest projection dimensionality is zero, according to Equation (5), so it can be projected into any space with a dimensionality greater than zero.

A projection matrix is randomly generated, and each sub-matrix is projected into the same low dimensional space with this projection matrix. On this basis, the distance changes before and after the projection of any two hyperspectral vectors in any two different sub-matrices are introduced. It is described in detail as Theorem 3. (Proof of Theorem 3 can be found in Appendix A).

**Theorem 3.** *For constants ε and β > 0, an integer $K_{PRP}$ is chosen according to Equation (5). $\boldsymbol{R}_{PRP}$ is a $D \times K_{PRP}$ matrix whose entries belong to the standard normal Gaussian distribution. The sub-matrix $\boldsymbol{V}^m$ is obtained based on Equation (6). Then, for the same ε, β, $K_{PRP}$, and $\boldsymbol{R}_{PRP}$, with the probability of no distortion at least*

$$P_{\mathrm{PRP2}} = 1 - 2N^{-\beta} = 1 - 2\left(S/M\right)^{-\beta} \tag{9}$$

*for any hyperspectral vector $\boldsymbol{u}^m_n$ in sub-matrix $\boldsymbol{U}^m$ and hyperspectral vector $\boldsymbol{u}^{m'}_{n'}$ in another sub-matrix $\boldsymbol{U}^{m'}$, the distance of the low dimensional hyperspectral vector $\boldsymbol{v}^m_n$ in sub-matrix $\boldsymbol{V}^m$ and low dimensional hyperspectral vector $\boldsymbol{v}^{m'}_{n'}$ in another sub-matrix $\boldsymbol{V}^{m'}$ is preserved,*

$$(1-\varepsilon)\,||\boldsymbol{u}^m_n - \boldsymbol{u}^{m'}_{n'}||^2 \leq ||\boldsymbol{v}^m_n - \boldsymbol{v}^{m'}_{n'}||^2 \leq (1+\varepsilon)\,||\boldsymbol{u}^m_n - \boldsymbol{u}^{m'}_{n'}||^2 \tag{10}$$

*where m′ is the index of sub-matrices. Significantly, m′ is not equal to m ($m' \neq m$).*

Combining Theorem 2 and Theorem 3, the distance before and after the projection of any two hyperspectral vectors in HSIs can also remain unchanged with an acceptable probability of no distortion. Specifically, the minimum value of probabilities of no distortion within each sub-matrix and between two different sub-matrices is taken as the probability of no distortion of the PRP algorithm. The projection of the matrix $\boldsymbol{U} = [\boldsymbol{u}_1; \dots ; \boldsymbol{u}_s; \dots ; \boldsymbol{u}_S]$ can be expressed as a matrix $\boldsymbol{V} = [\boldsymbol{v}_1; \dots ; \boldsymbol{v}_s; \dots ; \boldsymbol{v}_S]$, where the low dimensional hyperspectral vector $\boldsymbol{v}_s$ is the projection of $\boldsymbol{u}_s$. The proposed algorithm is interpreted as Theorem 4.

**Theorem 4.** *For constants ε and β > 0, an integer $K_{PRP}$ is chosen according to Equation (5). $\boldsymbol{R}_{PRP}$ is a $D \times K_{PRP}$ matrix whose entries belong to the standard normal Gaussian distribution. The $N \times K_{PRP}$ sub-matrix $\boldsymbol{V}^m$ is defined based on Equation (6) to present the projection of $\boldsymbol{U}^m$ in $K_{PRP}$ dimensional subspaces. Then, for the same ε, β, $K_{PRP}$ and $\boldsymbol{R}_{PRP}$, with the probability of no distortion at least*

$$P_{\mathrm{PRP}} = \min\,[P_{\mathrm{PRP1}}, P_{\mathrm{PRP2}}] = 1 - 2\left(S/M\right)^{-\beta} \tag{11}$$

*for any two hyperspectral vectors $\boldsymbol{u}_s$ and $\boldsymbol{u}_{s'}$ of matrix $\boldsymbol{U}$, the distance of the hyperspectral vectors $\boldsymbol{v}_s$ and $\boldsymbol{v}_{s'}$ of matrix $\boldsymbol{V}$ remains as Equation (4).*

The PRP algorithm can greatly reduce the dimensionality of HSIs, thereby reducing the computational load of subsequent HSI classification. The probability of no distortion $P_{\mathrm{PRP}}$ increases as the parameter $\beta$ increases and the number of sub-matrices $M$ decreases based on Equation (11). Theorem 4 demonstrates that the PRP algorithm does not introduce a significant distortion in HSIs. Meanwhile, the proposed algorithm can also gain a high

probability of no distortion in the case of dividing a small number of sub-matrices. In this case, the proposed algorithm can approximate the distance between each hyperspectral vector pair while projecting HSI into a space with a lower dimensionality than the RP algorithm. Thus, considering Theorem 1 and Theorem 4 comprehensively, the PRP algorithm is feasible in calculation and dimensionality reduction for large-size HSI. Moreover, when the number of sub-matrices is too small, the proposed algorithm cannot ensure the achievement of the purpose of dimensionality reduction. Therefore, this paper gives the minimum number of sub-matrices that guarantees high dimensionality can be reduced. Additionally, it is defined as Lemma 1. (Proof of Lemma 1 can be found in Appendix B).

**Lemma 1.** *Suppose that 1.5 > ε > 0 and β > 0, then*

$$M > S \times e^{-\frac{D}{24}} \Leftrightarrow K_0 \leq D \tag{12}$$

Figure 2 presents an example of dividing an HSI with five million hyperspectral vectors into one million sub-HSIs. The number of hyperspectral vectors $N$ in each sub-HSI is five in Figure 2, which is calculated by dividing the number of hyperspectral vectors $S$ in the HSI by the number of sub-HSIs $M$, that is, $N = S/M = 5,000,000/1,000,000 = 5$.
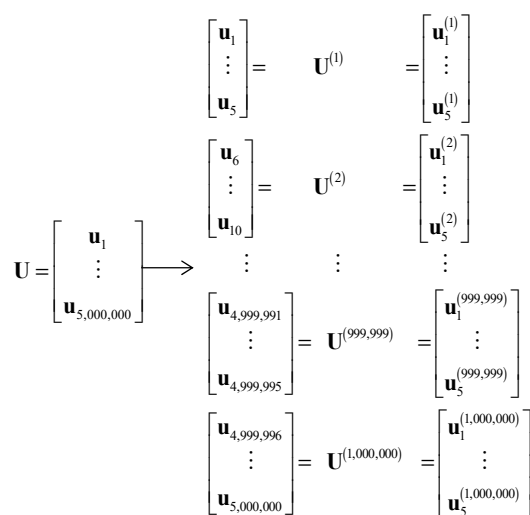


**Figure 2.** A partition of an HSI with five million hyperspectral vectors into one million sub-HSIs.

### 3.3. HSI Classification Based on the PRP Algorithm

The classification based on the PRP algorithm is introduced from two aspects. One is the optimization strategy of the projection matrix to increase the class separability of the dimensionality reduction result with the assistance of samples, and then the optimization strategy is applied to the PRP algorithm. The other is the HSI classification algorithm integrates the PRP algorithm for dimensionality reduction and MD classifier for classification.

### 3.3.1. Optimization Strategy of the Projection Matrix

Because the projection matrix is generated randomly, there is a case where the dimensionality reduction result with low class separability is not suitable for subsequent HSI classification. Therefore, it is necessary to exploit a projection matrix that is applicable for the classification task. The optimization strategy of the projection matrix is to increase the class separability of the low dimensional HSI. The class dissimilarity measurement is defined as large inter-class distance and small intra-class variance. The calculation method is the sum of the distance between classes divided by the variance within the class. Specifically, with the aid of samples, the projection matrix that maximizes the class dissimilarity is selected among multiple samplings.

Samples matrix for all classes of an HSI is expressed as $X$. Specifically, $X = [X^1; \dots; X^l; \dots; X^L]$, where $l$ is the index of classes and $L$ is the number of classes known as *a prior*. And $X^l$ is the $l$th samples matrix,

$$
X^l = \begin{bmatrix}
x^l_{11} & x^l_{12} & \cdots & x^l_{1D} \\
x^l_{21} & x^l_{22} & \cdots & x^l_{2D} \\
\vdots & \vdots & \ddots & \vdots \\
x^l_{H1} & x^l_{H2} & \cdots & x^l_{HD}
\end{bmatrix}
\tag{13}
$$

where $H$ is the number of samples in a single class, which is the same for each class. The selection of the projection matrix $R_{\text{PRP}}$ is very important for the HSI classification. To be specific, many matrices are generated based on the standard normal Gaussian distribution. These matrices constitute a sampling matrix of the projection matrix, $Q = [Q^1, \dots, Q^t, \dots, Q^T]$, where $t$ is the index of sampling, $T$ is the number of samplings, and $Q^t$ is the projection matrix of the $t$th sampling. Then, each projection matrix $Q^t$ is used to project the samples matrix $X$. The low dimensional samples matrix $W = [W^1, \dots, W^t, \dots, W^T]$ is the projection of samples matrix $X$ by the sampling matrices $Q$. Moreover, its calculation is derived from Equation (6), that is, $W^t = XQ^t / \sqrt{K_{\text{PRP}}}$. Specifically, $W^t = [W^{t1}, \dots, W^{tl}, \dots, W^{tL}]$, where $W^{tl}$ is the low dimensional samples matrix of class $l$ of the $t$th sampling. That is,

$$
W^{tl} = \begin{bmatrix}
w^{tl}_{11} & w^{tl}_{12} & \cdots & w^{tl}_{1K_{\text{PRP}}} \\
w^{tl}_{21} & w^{tl}_{22} & \cdots & w^{tl}_{2K_{\text{PRP}}} \\
\vdots & \vdots & \ddots & \vdots \\
w^{tl}_{H1} & w^{tl}_{H2} & \cdots & w^{tl}_{HK_{\text{PRP}}}
\end{bmatrix}
\tag{14}
$$

The feature mean hyperspectral vector of low dimensional samples matrix $W^{tl}$ of class $l$ of the $t$th sampling is described as

$$
W^{tl}_{\text{mean}} = \frac{1}{H} W^{tl} = \frac{1}{H} \begin{bmatrix} w^{tl}_{11} + \cdots + w^{tl}_{H1} & w^{tl}_{12} + \cdots + w^{tl}_{H2} & \cdots & w^{tl}_{1K_{\text{PRP}}} + \cdots + w^{tl}_{HK_{\text{PRP}}} \end{bmatrix}
\tag{15}
$$

Next, distance between $l$th class and $l'$th class divided by the variance of the $l$th class is defined as follows,

$$
I^t_{ll'} = \frac{\left\| W^{tl}_{\text{mean}} - W^{tl'}_{\text{mean}} \right\|^2}{var\left( W^{tl}_{\text{mean}} \right)}
\tag{16}
$$

where $l'$ is the index of classes. Then, the class dissimilarity is calculated as follows,

$$
J^t = \sum_{l=1}^{L} \sum_{l'=1}^{L} I^t_{ll'}
\tag{17}
$$

The better projection matrix is the $t^*$th sampled projection matrix with the large class dissimilarity. Then,

$$
t^* = \arg \max_{t=1,\dots,T} J^t
\tag{18}
$$

Finally, the projection matrix $R_{\text{PRP}}$ takes to the $t^*$th sampled projection matrix $Q^{t^*}$, where $t^*$ is obtained by Equation (18),

$$
R_{\text{PRP}} = Q^{t^*}
\tag{19}
$$

The projection matrix selection strategy can increase the separability of dimensionality reduction results, thereby increasing the accuracy of HSI classification.

### 3.3.2. HSI Classification Based on Optimization Strategy of the Projection Matrix

HSI classification algorithm is described in this section. First, the low dimensional samples matrix $Y = [Y^1; \ldots ; Y^l; \ldots ; Y^L]$ is the projection of samples matrix $X$ by the projection matrix obtained according to Equation (19) with $K_{\mathrm{PRP}}$ dimensionalities. To be special,

$$Y^l = \begin{bmatrix} y^l_{11} & y^l_{12} & \cdots & y^l_{1K_{\mathrm{PRP}}} \\ y^l_{21} & y^l_{22} & \cdots & y^l_{2K_{\mathrm{PRP}}} \\ \vdots & \vdots & \ddots & \vdots \\ y^l_{H1} & y^l_{H2} & \cdots & y^l_{HK_{\mathrm{PRP}}} \end{bmatrix} \tag{20}$$

is the $l$th low dimensional samples matrix. The feature mean hyperspectral vector of the $l$th low dimensional samples is expressed as

$$Y^l_{\mathrm{mean}} = \frac{1}{H}Y^l = \frac{1}{H}\left[ \ y^l_{11} + \cdots + y^l_{H1} \quad y^l_{12} + \cdots + y^l_{H2} \quad \cdots \quad y^l_{1K_{\mathrm{PRP}}} + \cdots + y^l_{HK_{\mathrm{PRP}}} \ \right] \tag{21}$$

Consequently, an HSI is denoted by $U = [u_1, \ldots , u_s, \ldots , u_S]$, where $s$ is the index of hyperspectral vectors, $S$ is the number of hyperspectral vectors, and $u_s$ is a hyperspectral vector with $D$ dimensionalities. The low dimensional HSI $V = [v_1, \ldots , v_s, \ldots , v_S]$ is obtained by the PRP algorithm, where $v_s$ is a low dimensional hyperspectral vector with $K_{\mathrm{PRP}}$ dimensionalities. Then, the final classes to which the low dimensional HSIs belong are determined by executing the MD classifier. That is to calculate the distance between the low dimensional hyperspectral vector and the samples feature mean hyperspectral vector of each class.

The distance matrix is characterized as $Z = [z_1, \ldots , z_s, \ldots , z_S]$, where $z_s$ is the $s$th distance vector. Especially, $z_s = [z_{s1}, \ldots , z_{sl}, \ldots , z_{sL}]$, where $z_{sl}$ is the distance between a low dimensional hyperspectral vector $v_s$ and feature mean hyperspectral vector $Y^l_{\mathrm{mean}}$ according to Equation (21), and the calculation method is as follows,

$$z_{sl} = ||Y^l_{\mathrm{mean}} - v_s|| \tag{22}$$

The determined classification result can be defined as $f = [f_1, \ldots , f_s, \ldots , f_S]$ in the HSI classification process. That is, the class with the smallest distance is the class to which the low dimensional hyperspectral vector belongs,

$$f_s = \mathrm{argmin} z_s = \arg \min_{l=1,\ldots,L} \{z_{sl}\} \tag{23}$$

### 3.4. The Complexity of the Proposed Algorithm

The space and time complexity of the proposed algorithm are analyzed here. This section studies the complexity in two parts: the optimization strategy of the projection matrix and the HSI classification algorithm based on the optimization strategy of the projection matrix.

The main contribution of the complexity of the optimization strategy of the projection matrix is to calculate the projection matrix $Q$. To update $Q$, it takes $O(HLD)$ space and $O(HLDK_{\mathrm{PRP}})$ time for the calculation of the low dimensional sample matrix $W$. Furthermore, $O(HK_{\mathrm{PRP}})$ space and $O(THL^2K_{\mathrm{PRP}})$ time are required to calculate the class dissimilarity $J^t$. The main contribution of the complexity of the HSI classification algorithm based on the optimization strategy of the projection matrix is to calculate the distance matrix $Z$. To update $Z$, it takes $O(MND)$ space and $O(MNDK_{\mathrm{PRP}})$ time for the calculation of the low dimensional image $V$. Furthermore, $O(MNK_{\mathrm{PRP}})$ space and $O(MNLK_{\mathrm{PRP}})$ time are required to calculate the distance $z_{sl}$. To sum up, the overall space complexity of the proposed algorithm is $O(MND)$, and the overall time complexity of the proposed algorithm is $O(HLDK_{\mathrm{PRP}} + THL^2K_{\mathrm{PRP}} + MNDK_{\mathrm{PRP}} + MNLK_{\mathrm{PRP}})$.

### 3.5. The Flow Chart of the Proposed Algorithm

The detailed process of the proposed Algorithm 1 is described as follows.

---

**Algorithm 1** The detailed process of the proposed classification algorithm

---

**Input:** HSI $U$ and samples $X$.
**Output:** Classification result $f$.
Step 1. Initializing parameters $\varepsilon$, $\beta$, $L$, $T$, $H$, $M$, $S$.
Step 2. Dividing an HSI evenly $U = [U^1; \ldots ; U^M]$.
Step 3. Calculating the lowest projection dimensionality $K^0_{PRP}$ by using Equation (5).
Step 4. Forming the projection matrix $R_{PRP}$ by using Equation (19).
Step 5. Generating a low dimensional sub-HSI $V^m$ by using Equation (6).
Step 6. Getting the low dimensional HSI $V = [V^1; \ldots ; V^M]$.
Step 7. Generating the feature mean hyperspectral vector of the low dimensional samples of $l$th class $Y^l_{mean}$ by using Equation (21).
Step 8. Calculating the distance $z_{sl}$ by using Equation (22).

---

In addition, Figure 3 illustrates a flow chart of the proposed classification algorithm in this paper.
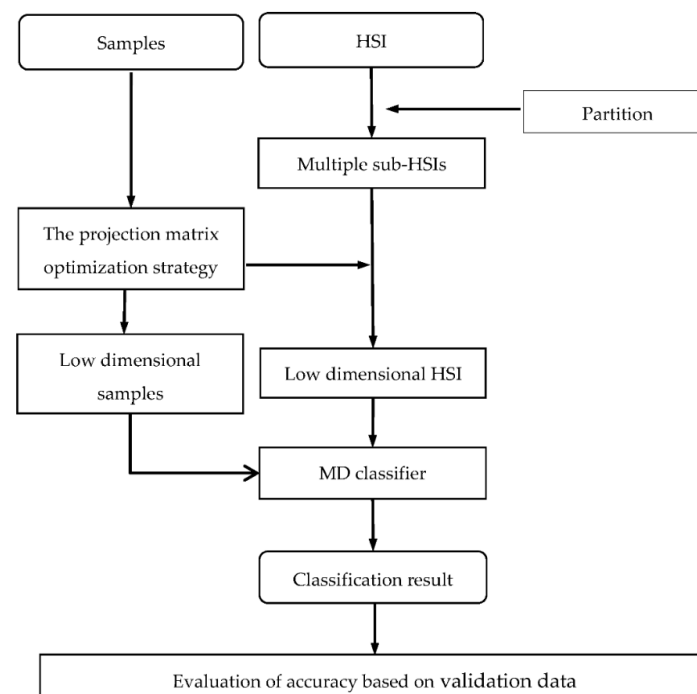


**Figure 3.** Flow chart of the proposed classification algorithm in this paper.

## 4. Experiments and Results

To validate the effectiveness of the proposed algorithm, classification experiments in real HSIs were performed on a PC with Intel (R) Xeon (R) CPU E7-8880 v3, 2.30 GHz and 96.0 GB memory using MATLAB R2020a. The experimental parameters are set as shown in Table 2. The table contains the number of samplings $T$, the number of samples in each class $H$, the number of sub-HSI $M$, and low dimensionalities. It can be seen from Table 2 that the $K_{PRP}$ is less than three-quarters of that of the $K_{TRP}$ and one-eighth of that of the $K_{RP}$, which greatly reduces the amount of computation. In addition, the $K_{RP}$ is larger than the high dimensionality $D$, which means that the RP algorithm cannot achieve the purpose of dimensionality reduction.

To verify the effectiveness of the projection matrix selection strategy, taking the LongKou image as an example, the randomly generated projection matrix (projection matrix without selection) is used as a comparison algorithm. Figure 4 shows the change of the spectral curve of the mean vector of samples of 9 classes of the LongKou image to evaluate the distance change between classes in the dimensionality reduction results. The horizontal coordinates represent the number of bands, and the vertical coordinates represent the intensity value. Figure 4a–c represent spectral curves of all classes without

dimensionality reduction, dimensionality reduction without matrix selection, and dimensionality reduction based on matrix selection, with legends for all classes, respectively. Comparing the three images in Figure 4, the low dimensional spectral curves obtained by the projection matrix selection strategy have good separability for each class in each dimensionality, which fully reflect the superiority of the projection matrix optimization strategy.

**Table 2.** Projection parameter settings in the experiment.

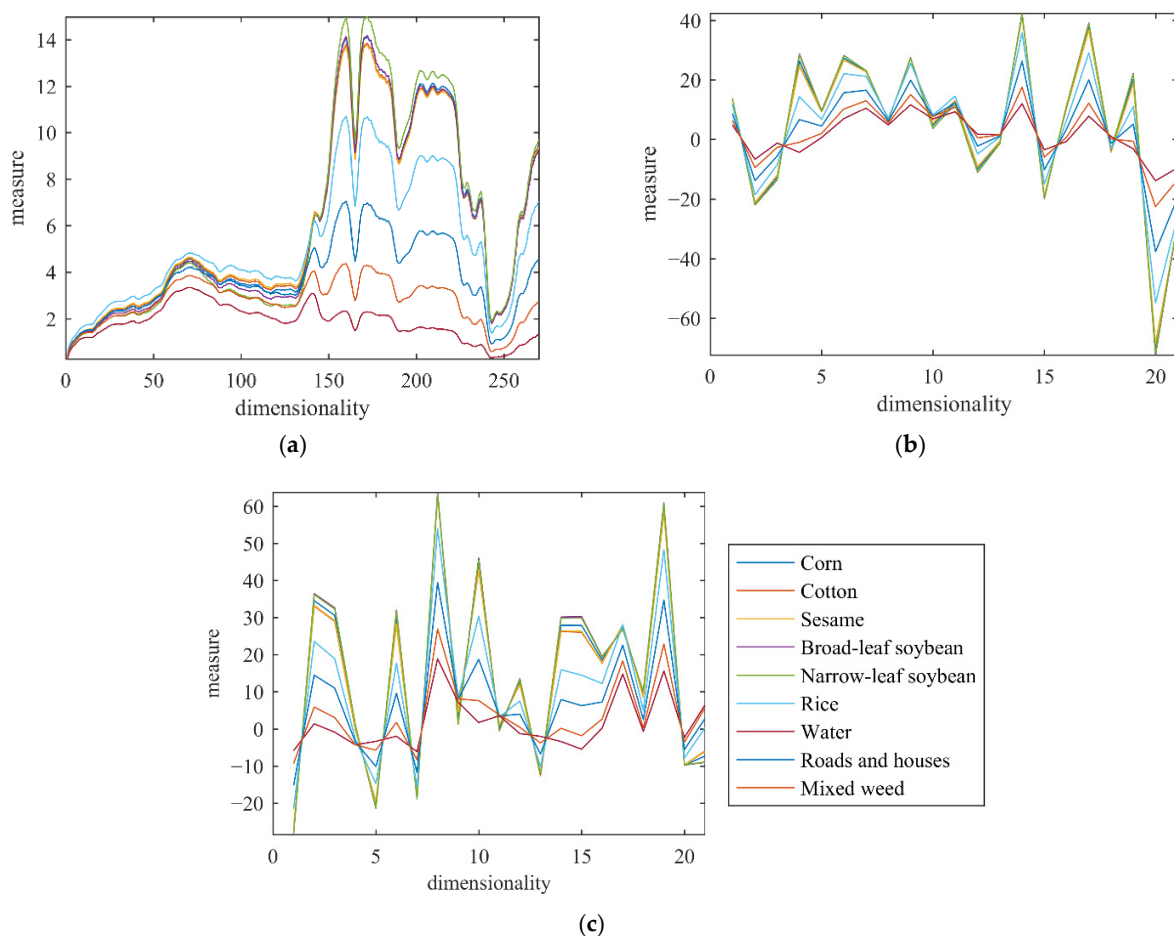|  | *T* | *H* | *M* | $K_{PRP}$ | $K_{TRP}$ | $K_{RP}$ |
|---|---|---|---|---|---|---|
| Pavia Centre image | 10 | 10 | 36,598 | 33 | 100 | 349 |
| Salinas image | 10 | 10 | 2295 | 66 | 86 | 299 |
| Chikusei image | 10 | 10 | 3145 | 33 | 79 | 275 |
| LongKou image | 10 | 10 | 102,271 | 21 | 106 | 367 |



**Figure 4.** Spectral curve analysis of sample mean vectors of all classes in the LongKou image. (**a**) High dimensional sample mean vectors of all classes. (**b**) Low dimensional sample mean vectors of all classes after dimensionality reduction without projection matrix selection. (**c**) Low dimensional sample mean vectors of all classes after dimensionality reduction based on projection matrix selection, with legends for all classes.

For the sake of validating the proposed classification algorithm, the TRP-MIV algorithm, the algorithm for combining PRP and MGSR (PRP-MGSR), and the CAFCM algorithm are used as comparison algorithms. It is worth noting that the classification experiments of all experimental images have been tested 100 times. For all algorithms, the classification results with the highest accuracy in 100 trials are exhibited in Figure 5. Figure 5(a1–a4,b1–b4,c1–c4,d1–d4) show the classification results for four experimental

images by the proposed algorithm and the TRP-MIV, PRP-MGSR, and CAFCM algorithms, respectively. Additionally, Figure 5(a5–d5) are legends for all classes of the four experimental images. Meanwhile, to better show the classification effect of the proposed algorithm, Figure 6 superimposes the final classification results of HSIs on the respective false-color images (see Figure 1(a1–d1) for details). Figure 6 shows the outlines of the superposition results of the four experimental images of the four algorithms.
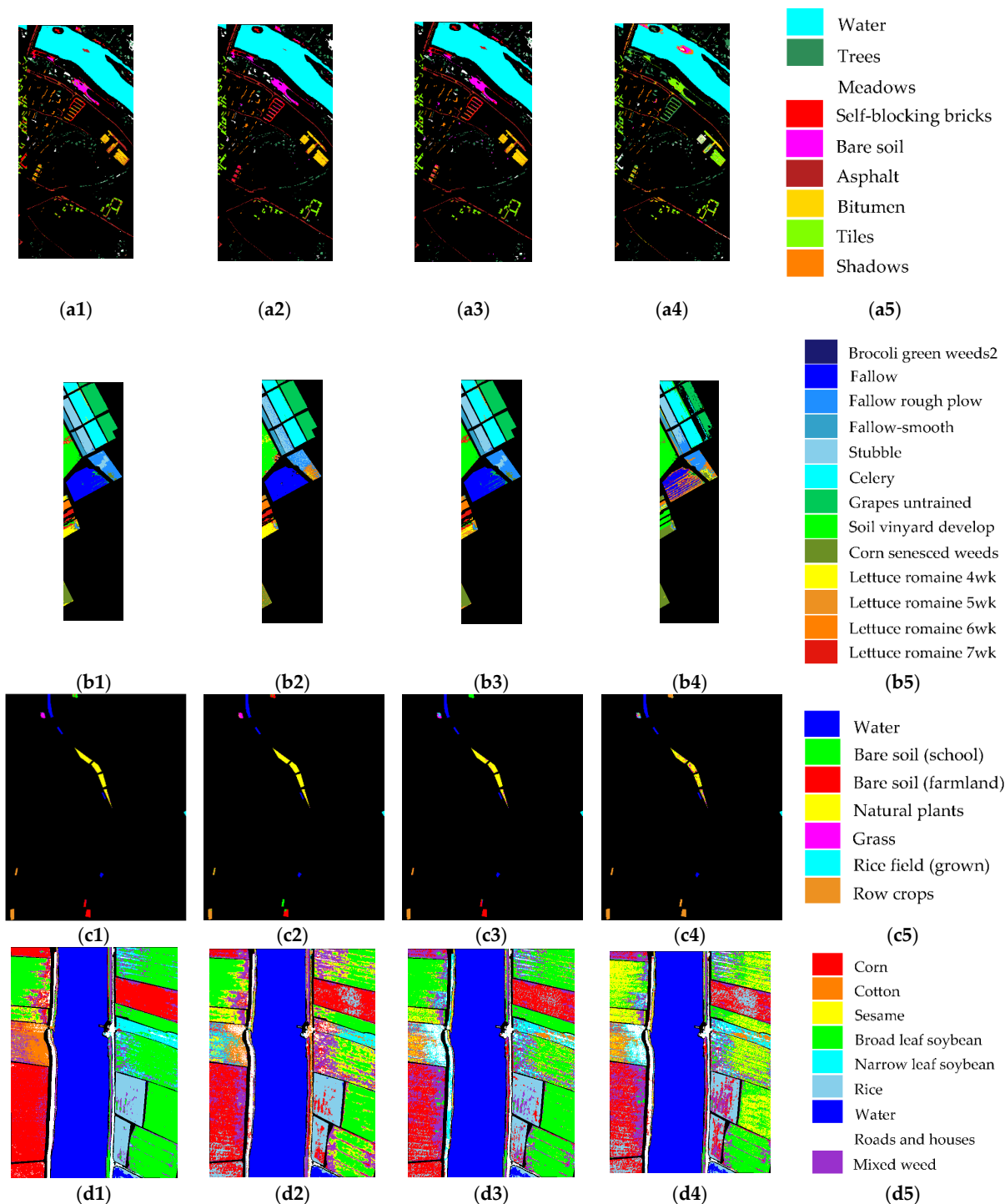


**Figure 5.** Comparison of the classification results of four images. (**a1**–**d1**) The proposed algorithm. (**a2**–**d2**) TRP-MIV algorithm. (**a3**–**d3**) PRP-MGSR algorithm. (**a4**–**d4**) CAFCM algorithm. (**a5**–**d5**) The legends for all classes.
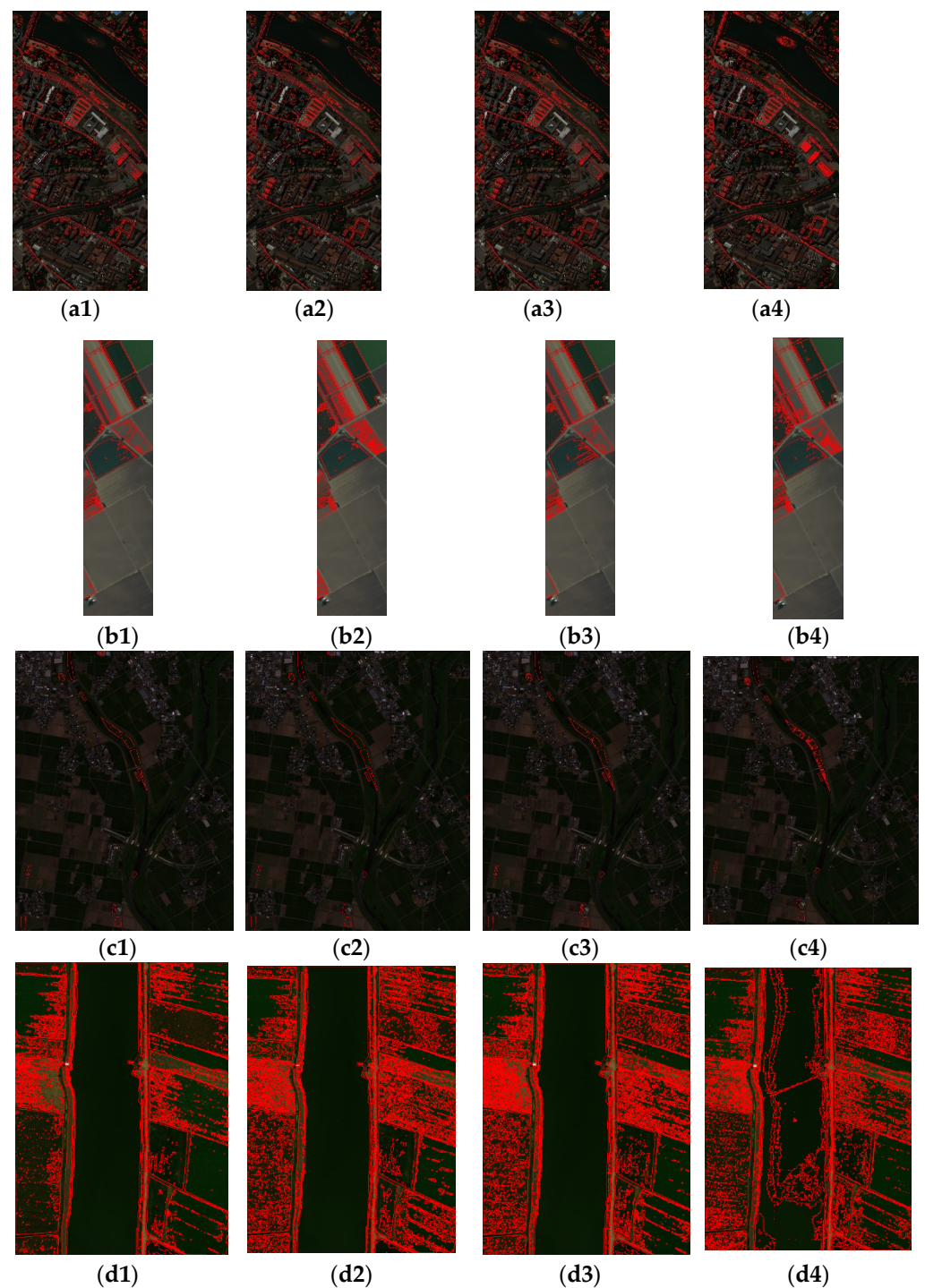
**Figure 6.** Comparison of outlines superposition results of four images. (**a1**–**d1**) The proposed algorithm. (**a2**–**d2**) TRP-MIV algorithm. (**a3**–**d3**) PRP-MGSR algorithm. (**a4**–**d4**) CAFCM algorithm. (The red lines are the outline of the classification results).

From a qualitative point of view, compared with the other three comparison algorithms, the classification performance of the proposed algorithm is superior, which proves the reliability of the proposed classification algorithm for large-size HSIs. According to qualitative experimental results, the TRP-MIV algorithm has misclassification for complex HSIs with large hyperspectral vectors. In addition, comparing the PRP-MGSR algorithm, the proposed algorithm simultaneously considers the inter-class distance and intra-class variance and can achieve better classification results. The proposed algorithm preserves the

class separability so that each class is classified exactly. In theory, the more dimensionalities are reduced, the more the loss of spectral information will be, which results in an unsatisfactory classification effect. However, the proposed algorithm attains a classification effect better than the CAFCM algorithm for four experimental images. Moreover, the CAFCM algorithm needs to be projected into a space with a higher dimensionality than the high dimensionality so that it cannot achieve the effect of dimensionality reduction.

The kappa coefficient, overall accuracy (OA), average accuracy (AA), average precision rate (APR), and running time are calculated to effectively evaluate the classification performance. Tables 3–6 reveal the quantitative evaluation results of four experimental images, respectively. The values in the tables are the mean and variance of the kappa coefficient, OA, AA, APR, and running time calculated by different algorithms for 100 trials, in which the values in brackets are the variance and those not in are the mean.

**Table 3.** Accuracy evaluation of classification results for the Pavia Centre image. (The values in brackets are the variance of 100 trials, and those not in are the mean).

|  | The Proposed Algorithm | TRP-MIV | PRP-MGSR | CAFCM |
|---|---|---|---|---|
| Kappa coefficient | 0.83 (0.02) | 0.81 (0.05) | 0.79 (0.02) | 0.38 (0.11) |
| OA/% | 89.65 (1.03) | 88.02 (2.94) | 87.34 (1.25) | 53.72 (11.68) |
| AA/% | 78.99 (1.94) | 75.43 (5.88) | 73.22 (3.02) | 38.03 (5.65) |
| APR/% | 79.80 (1.70) | 74.49 (5.81) | 72.72 (2.60) | 34.11 (6.01) |
| Running time/s | 1.16 (0.09) | 2.93 (0.29) | 10,209.72 (1538.24) | 5798.18 (855.03) |

**Table 4.** Accuracy evaluation of classification results for the Salinas image. (The values in brackets are the variance of 100 trials, and those not in are the mean).

|  | The Proposed Algorithm | TRP-MIV | PRP-MGSR | CAFCM |
|---|---|---|---|---|
| Kappa coefficient | 0.91 (0.00) | 0.87 (0.01) | 0.88 (0.00) | 0.59 (0.14) |
| OA/% | 91.98 (0.23) | 89.00 (0.71) | 90.20 (0.38) | 63.77 (11.47) |
| AA/% | 85.97 (0.92) | 84.24 (1.25) | 80.14 (1.70) | 48.95 (11.68) |
| APR/% | 79.89 (0.66) | 75.57 (1.20) | 75.28 (1.33) | 45.32 (17.93) |
| Running time/s | 0.37 (0.10) | 3.93 (0.11) | 100.08 (8.65) | 1623.16 (226.78) |

From a quantitative perspective, the classification results of the proposed algorithm surpass other algorithms. The running time of the proposed method for four image classification experiments is about 1.16, 0.37, 0.21, and 2.42 s, and the mean values of their kappa coefficient are larger than 0.82. All evaluation parameters intuitively state that the proposed algorithm accurately classifies the real HSIs. According to quantitative experimental results, although the mean values of AA and APR are less than 80% for the Pavia Centre image, the mean value of OA is more than 89%. For the Salinas image, the variance value of the kappa coefficient of the proposed algorithm in 100 trials is approximately equal to zero, which indicates that the classification results of the proposed algorithm are relatively stable. In addition, for the Chikusei image, the mean values of all evaluation parameters of the proposed algorithm are at least 0.05, 3.59%, 4.73%, 7.28%,

and 7.57 times larger than those of the comparison algorithms, which fully reflects the superiority of the proposed classification algorithm. Although the kappa coefficient of the proposed algorithm is not very high for the LongKou image, the accuracy of the proposed algorithm is still higher than that of all comparison algorithms. Additionally, the values of OA, AA, and APR outperform the comparison algorithms by at least 16.51%, 21.35%, and 35.19%. In summary, the proposed algorithm can obtain reliable classification results in a very short time.

**Table 5.** Accuracy evaluation of classification results for the Chikusei image. (The values in brackets are the variance of 100 trials, and those not in are the mean).

|  | The Proposed Algorithm | TRP-MIV | PRP-MGSR | CAFCM |
|---|---|---|---|---|
| Kappa coefficient | 0.97 (0.01) | 0.92 (0.05) | 0.90 (0.01) | 0.35 (0.10) |
| OA/% | 97.74 (0.38) | 94.15 (3.36) | 92.93 (0.74) | 48.52 (7.77) |
| AA/% | 96.66 (1.06) | 87.89 (8.85) | 91.63 (0.81) | 36.58 (9.75) |
| APR/% | 94.57 (0.56) | 87.29 (7.67) | 85.00 (1.93) | 33.62 (6.54) |
| Running time/s | 0.21 (0.76) | 1.59 (0.16) | 52.20 (11.34) | 289.21 (34.70) |

**Table 6.** Accuracy evaluation of classification results for the LongKou image. (The values in brackets are the variance of 100 trials, and those not in are the mean).

|  | The Proposed Algorithm | TRP-MIV | PRP-MGSR | CAFCM |
|---|---|---|---|---|
| Kappa coefficient | 0.82 (0.01) | 0.62 (0.01) | 0.62 (0.03) | 0.47 (0.01) |
| OA/% | 86.24 (0.51) | 68.90 (0.81) | 69.73 (2.93) | 56.84 (1.03) |
| AA/% | 77.94 (1.57) | 56.59 (2.12) | 55.51 (3.51) | 33.69 (2.56) |
| APR/% | 74.23 (1.50) | 52.05 (1.65) | 52.78 (2.36) | 39.04 (3.21) |
| Running time/s | 2.42 (0.24) | 7.76 (0.40) | 7215.75 (835.60) | 16,698.02 (1512.06) |

## 5. Discussion

In order to verify the superiority of the proposed classification algorithm, all classification experiments have been tested 100 times. The TRP-MIV algorithm [38] has better classification results for HSIs with a large inter-class distance. The PRP-MGSR algorithm is relatively stable in classification results in multiple tests. This is because the MGSR algorithm [14] has strong applicability to various data, but the running time is relatively long. The CAFCM algorithm [39] is more likely to be disturbed by noise, and the classification results of the algorithm are better for images with less noise. The proposed classification algorithm considers the class dissimilarity, which can reach the ideal classification result in a shorter time.

In order to use the proposed algorithm more widely, this section discusses the influence of the projection parameters on dimensionality reduction and the influence of the number of samples and the number of samplings on the classification accuracy. First, the focus is to analyze the effects of parameters $\varepsilon$, $\beta$, and the number of sub-HSIs $M$ on the lowest projection dimensionality $K^0_{PRP}$. According to Equation (5), taking the Salinas image as

an example, Figure 7 represents the change curves of the lowest projection dimensionality with different parameters, where the horizontal coordinates are different parameters, and the vertical coordinates are $K^0_{PRP}$. Figure 7a is the change curve of $\varepsilon$ and $K^0_{PRP}$ when $\beta = 0.5$ and $M = 2295$ (according to Table 2), where the position of the black dotted line is the inflection point. When $\varepsilon$ is equal to 1.5, the denominator term of Equation (5) is 0, so the curve in Figure 7a is discontinuous at $\varepsilon = 1.5$. When $\varepsilon < 1.5$, $K^0_{PRP}$ first decreases and then increases with the increase of $\varepsilon$ and takes the minimum value when $\varepsilon = 1$. When $\varepsilon > 1.5$, $K^0_{PRP}$ increases with the increase of $\varepsilon$ and is always negative, which is meaningless for dimensionality reduction. Figure 7b is the change curve of $\beta$ and $K^0_{PRP}$ when $\varepsilon = 1$ and $M = 2295$. As seen in Figure 7b, $K^0_{PRP}$ increases with the increase of $\beta$. Figure 7c is the change curve of $M$ and $K^0_{PRP}$ when $\varepsilon = 1$ and $\beta = 0.5$. The position of the black dotted line is the number of bands, and the values on the upper side and the left side of the boundary line cannot achieve dimensionality reduction. As seen from Figure 7c, $K^0_{PRP}$ decreases with the increase of $M$. Therefore, to reduce computation and memory, it is necessary to project the HSI to a smaller dimensional space. Then, it is best to set $\varepsilon$ to 1 within (0, 1.5); $\beta$ should be set as small as possible, and $M$ should be set as large as possible. Next, in order to reflect the impact of the number of samples $H$ and the number of samplings $T$ on classification performance, the change curves in the values of OA with different $T$ and $H$ are shown in Figure 8. The number of samples and samplings are taken from 10 to 650, respectively. Figure 8 shows that as $T$ and $H$ increase, the OA value varies only in a small range. The larger the number of samplings and samples, the smaller the range of variation. Among them, Figure 8a shows that with the increase of $T$, the error of the OA value is 1%, and Figure 8b shows that with the increase of $H$, the error of the OA value is 2.5%. Therefore, in practical applications, only a small number of samples and few projection matrix samplings can be used to obtain better classification results. In addition, in order to ensure that the PRP algorithm can still maintain the distance after the HSI is divided, the number of samples in each class in this paper is set to be the same. The focus of future research will be on dividing an HSI into sub-HSIs with different numbers of hyperspectral vectors to improve utility.
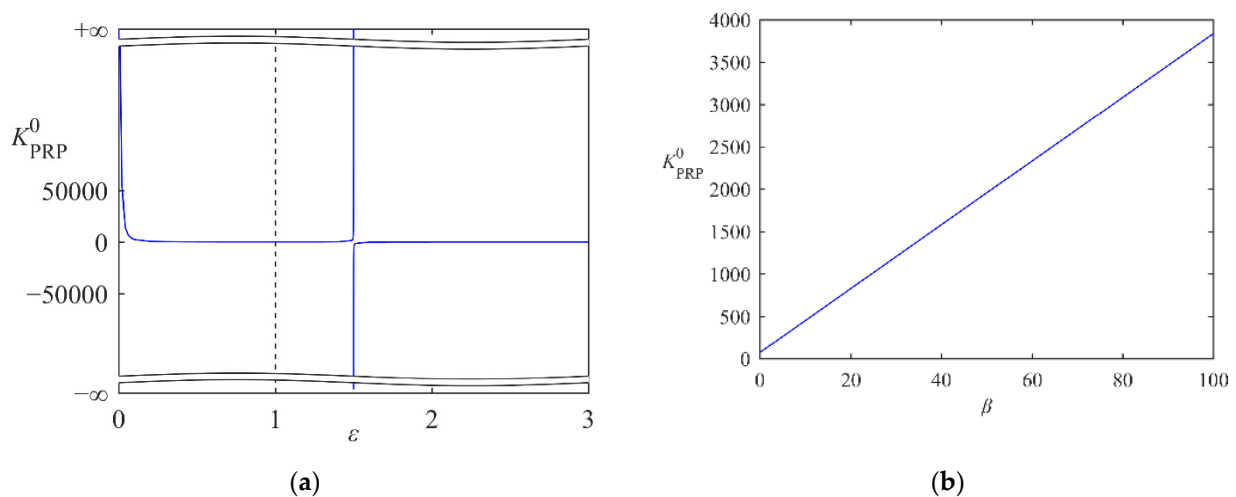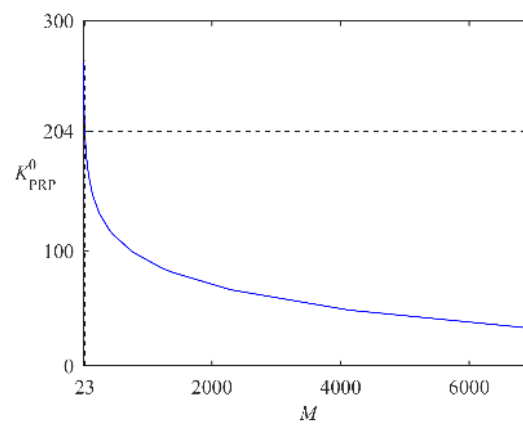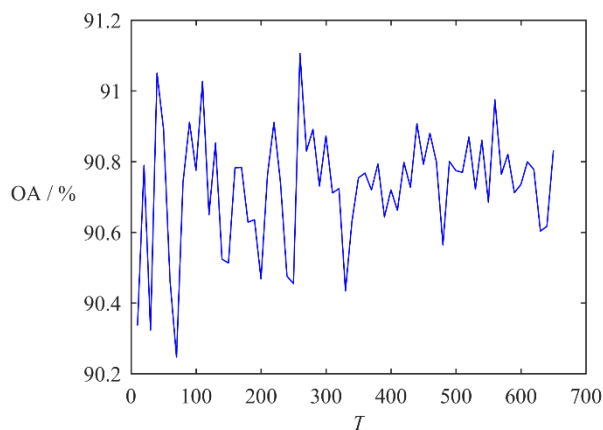


(a)

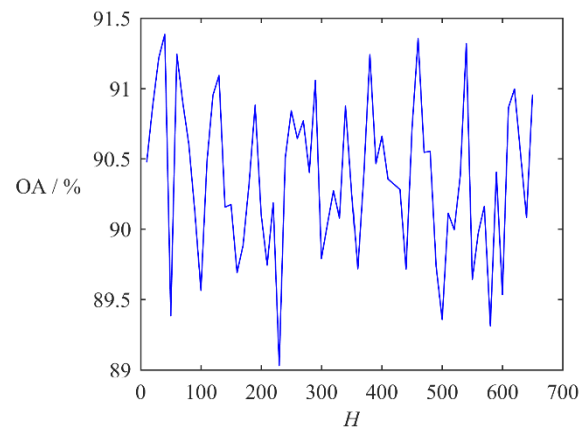

(b)

**Figure 7.** *Cont.*

(**c**)

**Figure 7.** Analysis of the change of the lowest projection dimensionality $K^0_{\mathrm{PRP}}$ with different parameters. (**a**) The change curve of $\varepsilon$ and $K^0_{\mathrm{PRP}}$. (The position of the black dotted line is the inflection point.) (**b**) The change curve of $\beta$ and $K^0_{\mathrm{PRP}}$. (**c**) The change curve of $M$ and $K^0_{\mathrm{PRP}}$. (The position of the black dotted line is the number of bands).



(**a**)



(**b**)

**Figure 8.** Analysis of the change of OA with the number of samplings $T$ and the number of samples in each class $H$. (**a**) The change curve of $T$ and OA. (**b**) The change curve of $H$ and OA.

## 6. Conclusions

A new dimensionality reduction method, the PRP algorithm, is presented for large-size HSIs in this paper. It is worth mentioning that this paper also theoretically proves the distance preservation property of the PRP algorithm in detail. The PRP algorithm projects the large-size HSI into a space with a lower dimensionality than the RP algorithm. Therefore, when applied to the large-size HSI, the PRP algorithm no longer has the problem of the inability to reduce the dimensionality. In addition, the proposed algorithm utilizes the large inter-class distance and small intra-class variance as the class dissimilarity measurement for selecting the projection matrix, which can preserve the class separability of the low dimensional HSI well. Compared with TRP-MIV, PRP-MGSR, and CAFCM algorithms, the proposed algorithm has the shortest running time and the highest classification accuracy.

**Author Contributions:** Conceptualization, S.J., Q.Z. and Y.L.; methodology, S.J., Q.Z. and Y.L.; software, S.J.; validation, S.J.; formal analysis, S.J.; writing—original draft preparation, S.J.; writing—review and editing, S.J., Q.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** The data presented in this study are openly available at http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm, and https://naotoyokoya.com/Download.html.

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

## Appendix A

**Proof of Theorem 3.** To facilitate understanding, some basic statistical concepts and properties that need to be used in the proof are introduced.

Markov inequality. The Markov inequality defines an upper bound on the probability that a function of a random variable is greater than or equal to a positive number. Let X be a non-negative random variable $\alpha > 0$, and then,

$$Pr(X \geq \alpha) \leq \frac{E(X)}{\alpha} \tag{A1}$$

Standard normal distribution. If $\alpha$ obeys the standard normal distribution, then

$$p(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right) \tag{A2}$$

Function integral. When $1/2 > a > 0$,

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2a^2}\right) dx = \sqrt{2\pi}a \tag{A3}$$

Taylor expansion. A formula that describes the value around a point of an exponential function is as follows.

$$\exp(-\varepsilon) = 1 - \varepsilon + \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{6} + \cdots \tag{A4}$$

The specific proof is given below. Provided that the probability that the distance between the vectors before and after projection exceeds the range of $(1 \pm \varepsilon)$ is $P_{\text{distortion}}$

$$P_{\text{distortion}} = Pr\left(\left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2 \vee \left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2\right) \tag{A5}$$

Either of these two cases signifies that the distance exceeds a factor of $\varepsilon$. Without loss of generality, thinking of the situation $Pr\left(||v_n^m - v_{n'}^{m'}||^2 \geq (1+\varepsilon)||u_n^m - u_{n'}^{m'}||^2\right)$, suppose that

$$C = \frac{K_{\text{PRP}}\left\|v_n^m - v_{n'}^{m'}\right\|^2}{\left\|u_n^m - u_{n'}^{m'}\right\|^2} \tag{A6}$$

Then,

$$Pr\left(\left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2\right) = Pr(C \geq (1+\varepsilon)K_{\text{PRP}}) \tag{A7}$$

For any constant $\lambda > 0$, continue to construct the above formula, then,

$$Pr\left(\left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2\right) = Pr(\exp(C\lambda) \geq \exp((1+\varepsilon)K_{\text{PRP}}\lambda)) \quad \text{(A8)}$$

The probability that the projected distance will remain the same can be linked to the mathematical expectation of the projected distance according to Equation (A1). Then, the first scaling can be done for $Pr\left(||v^m{}_n - v^{m'}{}_{n'}||^2 \geq (1+\varepsilon)||u^m{}_n - u^{m'}{}_{n'}||^2\right)$,

$$Pr\left(\left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2\right) \leq \frac{\text{E}(\exp(C\lambda))}{\exp((1+\varepsilon)K_{\text{PRP}}\lambda)} \quad \text{(A9)}$$

Meanwhile, according to Equation (6), the distance between vector $v^m{}_n$ and vector $v^{m'}{}_{n'}$ is

$$\left\|v_n^m - v_{n'}^{m'}\right\|^2 = \left\|\frac{1}{\sqrt{K_{\text{PRP}}}}R(u_n^m - u_{n'}^{m'})\right\| \quad \text{(A10)}$$

Then,

$$C = \frac{K_{\text{PRP}}\left\|v_n^m - v_{n'}^{m'}\right\|^2}{\left\|u_n^m - u_{n'}^{m'}\right\|^2} = \frac{\left\|R(u_n^m - u_{n'}^{m'})\right\|^2}{\left\|u_n^m - u_{n'}^{m'}\right\|^2} \quad \text{(A11)}$$

Supposing that $C_k = R_k (u^m{}_n - u^{m'}{}_{n'})$. Because it is the weighted average of the independent standard normal variables, it belongs to the standard normal distribution. Hence,

$$C = \sum_{k=1}^{K_{\text{PRP}}} \frac{\left\|R_k(u_n^m - u_{n'}^{m'})\right\|^2}{\left\|u_n^m - u_{n'}^{m'}\right\|^2} = \frac{1}{\left\|u_n^m - u_{n'}^{m'}\right\|^2} \sum_{k=1}^{K_{\text{PRP}}} C_k^2 \quad \text{(A12)}$$

Since $C_k$ is independent and identically distributed, $Pr\left(||v^m{}_n - v^{m'}{}_{n'}||^2 \geq (1+\varepsilon)||u^m{}_n - u^{m'}{}_{n'}||^2\right)$ defined in Equation (6) can be calculated as follows,

$$
\begin{aligned}
Pr\left(\left\|v_n^m - v_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|u_n^m - u_{n'}^{m'}\right\|^2\right) \quad &\leq \frac{\text{E}(\exp(C\lambda))}{\exp((1+\varepsilon)K_{\text{PRP}}\lambda)} \\
&= \frac{\prod_{k=1}^{K_{\text{PRP}}} \text{E}\left(\exp(C_k^2\lambda)\right)}{\exp((1+\varepsilon)K_{\text{PRP}}\lambda)} \\
&= \left(\frac{\text{E}\left(\exp(C_1^2\lambda)\right)}{\exp((1+\varepsilon)\lambda)}\right)^{K_{\text{PRP}}}
\end{aligned}
\quad \text{(A13)}
$$

Intriguingly, because $C_1$ obeys the standard normal distribution, then according to Equation (A2)

$$
\begin{aligned}
\text{E}\left(\exp(C_1^2\lambda)\right) \quad &= \int_{-\infty}^{+\infty} \exp(C_1^2\lambda)p(C_1)dC_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(C_1^2\lambda) \exp\left(-\frac{C_1^2}{2}\right)dC_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-(\frac{1}{2} - \lambda)C_1^2\right)dC_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(1 - 2\lambda)C_1^2\right)dC_1
\end{aligned}
\quad \text{(A14)}
$$

Based on Equation (A3), combining the above equation and Equation (A13),

$$\text{E}\left(\exp(C_1^2\lambda)\right) = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi}\frac{1}{\sqrt{1-2\lambda}} = \frac{1}{\sqrt{1-2\lambda}} \quad \text{(A15)}$$

Therefore,

$$
Pr\left(\left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2\right) \leq \left(\frac{\mathrm{E}\left(\exp(C_1^2\lambda)\right)}{\exp((1+\varepsilon)\lambda)}\right)^{K_{\mathrm{PRP}}}
$$
$$
= \left(\frac{1}{\sqrt{1-2\lambda}\exp((1+\varepsilon)\lambda)}\right)^{K_{\mathrm{PRP}}} \tag{A16}
$$
$$
= \left(\frac{\exp(-2(1+\varepsilon)\lambda)}{1-2\lambda}\right)^{\frac{K_{\mathrm{PRP}}}{2}}
$$

In order to perform the next scaling on the base term of the above equation, we provide that

$$
g(\lambda) = \frac{\exp(-2(1+\varepsilon)\lambda)}{1-2\lambda} \tag{A17}
$$

At this point, its minimum value can be found at $\lambda_0 = \varepsilon/2\,(1+\varepsilon)$, and $g\,(\lambda_0) = (1+\varepsilon)$ $\exp(-\varepsilon)$ here. And then,

$$
Pr\left(\left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2\right) \leq ((1+\varepsilon)\exp(-\varepsilon))^{\frac{K_{\mathrm{PRP}}}{2}} \tag{A18}
$$

Therewith, according to Equation (A4), the probability that a single hyperspectral vector pair satisfies the distance preservation property is

$$
Pr\left(\left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2\right) \leq \exp\left(\frac{K_{\mathrm{PRP}}}{2}\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)\right) \tag{A19}
$$

Similarly,

$$
Pr\left(\left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \leq (1-\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2\right) \leq \exp\left(\frac{K_{\mathrm{PRP}}}{2}\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)\right) \tag{A20}
$$

According to the above derivation, the probability of distortion is as follows,

$$
Pr\left(\left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \geq (1+\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2 \vee \left\|\boldsymbol{v}_n^m - \boldsymbol{v}_{n'}^{m'}\right\|^2 \leq (1-\varepsilon)\left\|\boldsymbol{u}_n^m - \boldsymbol{u}_{n'}^{m'}\right\|^2\right)
$$
$$
\leq 2\exp\left(\frac{K_{\mathrm{PRP}}}{2}\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)\right) \tag{A21}
$$

Since the number of hyperspectral vector pairs is $N^2$, the probability of making these hyperspectral vector pairs undistorted based on Equation (5) can be calculated as below,

$$
1 - N^2 \cdot 2\exp\left(\frac{K_{\mathrm{PRP}}}{2}\left(-\frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}\right)\right) \geq 1 - 2N^2 \cdot \exp(-(2+\beta)\log N)
$$
$$
= 1 - 2\exp(-(2+\beta)\log N + 2\log N)
$$
$$
= 1 - 2\exp(-\beta\log N) \tag{A22}
$$
$$
= 1 - 2N^{-\beta}
$$

□

## Appendix B

**Proof of Lemma 1.** For Theorem 4, if the lowest projection dimensionality $K^0{}_{\mathrm{PRP}}$ is smaller than the high dimensionality $D$ based on Equation (5), then

$$
\frac{4+2\beta}{\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}}\log\frac{S}{M} \leq D \ (\forall \varepsilon, \beta > 0) \tag{A23}
$$

For $\beta > 0$, the item $4 + 2\beta$ is larger than 4. Moreover, $S/M$ is the number of hyperspectral vectors in the sub-HSI $N$, which is larger than 1. Therefore, the logarithmic item $\log S/M$ is larger than zero. When the range of the parameter $\varepsilon$ is larger than 1.5, the item $\varepsilon^2/2 - \varepsilon^3/3$

is smaller than 0. Therefore, the value on the left side of the above equation is always negative. At this time, the above equation is always established, so this paper will not discuss it. In addition, if the range of the parameter $\varepsilon$ is 0 to 1.5, the item $\varepsilon^2/2 - \varepsilon^3/3$ is larger than 0. Thus, the value on the left side of the above equation is always positive here. Then,

$$\frac{4+2\beta}{D} \log \frac{S}{M} \leq \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} \ (\forall 1.5 > \varepsilon > 0, \beta > 0) \tag{A24}$$

The right side can be seen as a function $g(\varepsilon) = \varepsilon^2/2 - \varepsilon^3/3$ of $\varepsilon$. Within this range, the function obviously increases first and then decreases. $g(\varepsilon)$ takes the maximum value $g(\varepsilon_0) = 1/6$ at $\varepsilon_0 = 1$. Thus, in case the above inequality is always established in the range of $1.5 \geq \varepsilon \geq 0$, then

$$\frac{4+2\beta}{D} \log \frac{S}{M} \leq \frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3} \leq \frac{1}{6} \ (\forall 1.5 > \varepsilon > 0, \beta > 0) \tag{A25}$$

Moreover, when the parameter $\beta$ is equal to zero, the left side takes the value $(4 \log S/M)/D$. Similarly, for $\beta > 0$, if the above equation always holds, then

$$\frac{4}{D} \log \frac{S}{M} < \frac{4+2\beta}{D} \log \frac{S}{M} \leq \frac{1}{6} \tag{A26}$$

In this way, the minimum number of sub-matrices to ensure dimensionality reduction can be calculated as follows,

$$\begin{aligned} \frac{4}{D} \log \frac{S}{M} &< \frac{1}{6} \\ \log \frac{S}{M} &< \frac{D}{24} \\ \frac{S}{M} &< e^{\frac{D}{24}} \\ M &> S e^{-\frac{D}{24}} \end{aligned} \tag{A27}$$

$\square$

**References**

1. Pan, H.; Chen, Z. Application of UVA hyperspectral remote sensing in winter wheat leaf area index inversion. *Chin. J. Agric. Resour. Reg. Plan.* **2018**, *39*, 32–37. [CrossRef]
2. Xu, Q.; Ma, Y.; Jiang, Q.; Tong, C.; Zhao, Z. Estimation of rice leaf water content based on hyperspectral remote sensing. *Remote Sens. Infom.* **2018**, *33*, 136–143. [CrossRef]
3. Chutia, D.; Bhattacharyya, D.K.; Sarma, K.K.; Kalita, R.; Sudhakar, S. Hyperspectral remote sensing classifications: A perspective survey. *Trans. GIS* **2016**, *20*, 463–490. [CrossRef]
4. Wang, Y.; Reardon, C.P.; Read, N.; Thorpe, S.; Evans, A.; Todd, N.; Krauss Thomas, F. Attachment and antibiotic response of early-stage biofilms studied using resonant hyperspectral imaging. *NPJ Biofilms Microbiomes* **2020**, *6*, 57. [CrossRef]
5. Koprowski, R. Hyperspectral imaging in medicine: Image pre-processing problems and solutions in Matlab. *J. Biophotonics* **2015**, *8*, 935–943. [CrossRef] [PubMed]
6. Gu, X.; Wang, Y.; Sun, Q.; Yang, G.; Zhang, C. Hyperspectral inversion of soil organic matter content in cultivated land based on wavelet transform. *Comput. Electron. Agric.* **2019**, *167*, 105053–105059. [CrossRef]
7. Mathieu, M.; Roy, R.; Launeau, P.; Cathelineau, M.; Quirt, D. Alteration mapping on drill cores using a hyspex swir-320m hyperspectral camera: Application to the exploration of an unconformity related uranium deposit (Saskatchewan, Canada). *J. Geochem. Explor.* **2017**, *172*, 71–88. [CrossRef]
8. Su, H.; Sheng, Y.; Du, P.; Chen, C.; Liu, K. Hyperspectral image classification based on volumetric texture and dimensionality reduction. *Front. Earth Sci.* **2015**, *9*, 225–236. [CrossRef]
9. Dong, Y.; Du, B.; Zhang, L.; Zhang, L. Exploring locally adaptive dimensionality reduction for hyperspectral image classification: A maximum margin metric learning aspect. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 1136–1150. [CrossRef]
10. Yao, D.; Zhao, P.; Pham, T.N.; Cong, G. High-Dimensional Similarity Learning Via Dual-Sparse Random Projection. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), Stockholm, Sweden, 13–19 July 2018; pp. 3005–3011. [CrossRef]
11. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Han, L. Band selection of hyperspectral images using attention-based autoencoders. *IEEE Geosci. Remote. Sens. Lett.* **2020**, *18*, 147–151. [CrossRef]

12. Fan, J.; Liao, Y.; Wang, W. Projected principal component analysis in factor models. *Ann. Stat.* **2016**, *44*, 219–254. [CrossRef] [PubMed]

13. Hua, W.; Lin, Y.; Huang, H.; Ding, C. From protein sequence to protein function via multi-label linear discriminate analysis. *IEEE-ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 503–513. [CrossRef]

14. Zhang, Y.; Wang, X.; Jiang, X.; Zhou, Y. Marginalized graph self-representation for unsupervised hyperspectral band selection. *IEEE Trans. Geosci. Remote. Sens.* **2021**, *60*, 1–12. [CrossRef]

15. Saranathan, A.M.; Parente, M. Uniformity-based superpixel segmentation of hyperspectral images. *IEEE Trans. Geosci. Remote. Sens.* **2016**, *54*, 1419–1430. [CrossRef]

16. Mou, L.; Saha, S.; Hua, Y.; Bovolo, F.; Zhu, X.X. Deep reinforcement learning for band selection in hyperspectral image classification. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–14. [CrossRef]

17. Tamilarasi, R.; Prabu, S. Automated building and road classifications from hyperspectral imagery through a fully convolutional network and support vector machine. *J. Supercomput.* **2021**, *77*, 13243–13261. [CrossRef]

18. Filipović, M.; Kopriva, I. A comparison of dictionary based approaches to inpainting and denoising with an emphasis to independent component analysis learned dictionaries. *Inverse Probl. Imag.* **2017**, *5*, 815–841. [CrossRef]

19. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [CrossRef]

20. Li, H.; Jiang, G.; Wang, R.; Zhang, J.; Wang, Z.; Zheng, W.S.; Menze, B. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage* **2018**, *183*, 650–665. [CrossRef]

21. Meier, T.B.; Desphande, A.S.; Vergun, S.; Nair, V.A.; Prabhakaran, V. Support vector machine classification and characterization of age-related reorganization of functional brain networks. *NeuroImage* **2012**, *60*, 601–613. [CrossRef]

22. Frankl, P.; Maehara, H. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Comb. Theory* **1988**, *44*, 355–362. [CrossRef]

23. Dasgupta, S. Experiments with Random Projection. In Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, Francisco, CA, USA, 30 June–3 July 2000; pp. 143–151. [CrossRef]

24. Dasgupta, S.; Gupta, A. An elementary proof of the Johnson-Lindenstrauss lemma. *Random Struct. Algor.* **2003**, *22*, 60–65. [CrossRef]

25. Zhu, Y.; Chen, S. Growing neural gas with random projection method for high-dimensional data stream clustering. *Soft Comput.* **2020**, *24*, 9789–9807. [CrossRef]

26. Vempala, S.S. *The Random Projection Method*; American Mathematical Society: Providence, RI, USA, 2004. [CrossRef]

27. Achlioptas, D. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *J. Comput. Syst. Sci.* **2003**, *66*, 671–687. [CrossRef]

28. Huang, W.; Wong, P.K.; Wong, K.I.; Vong, C.M.; Zhao, J. Adaptive neural control of vehicle yaw stability with active front steering using an improved random projection neural network. *Veh. Syst. Dyn.* **2021**, *59*, 396–414. [CrossRef]

29. Zhe, X.; Chen, S.; Yan, H. Directional statistics-based deep metric learning for image classification and retrieval. *Pattern Recogn.* **2018**, *93*, 113–123. [CrossRef]

30. Yellamraju, T.; Boutin, M. Clusterability and clustering of images and other "real" high-dimensional data. *IEEE Trans. Image Process.* **2018**, *27*, 1927–1938. [CrossRef]

31. Fan, Y.; Xuan, L.; Li, Q.; Tao, L. Exploring the diversity in cluster ensemble generation: Random sampling and random projection. *Expert Syst. Appl.* **2014**, *41*, 4844–4866. [CrossRef]

32. Hou, B.; Na, L.; Shuang, W.; Zhang, X. SAR image segmentation based on random projection and signature frame. *Geosci. Remote Sens. Symp.* **2014**, *1*, 3726–3729. [CrossRef]

33. Fowler, J.E.; Qian, D.; Wei, Z.; Younan, N.H. Classification performance of random-projection-based dimensionality reduction of hyperspectral imagery. *Geosci. Remote Sens. Symp.* **2009**, *5*, 76–79. [CrossRef]

34. Alshamiri, A.K.; Singh, A.; Surampudi, B.R. Combining ELM with Random Projections for Low and High Dimensional Data Classification and Clustering. In Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015), Copenhagen, Denmark, 17–19 December 2015; Springer: Cham, Switzerland, 2015; pp. 89–107.

35. Fern, X.Z.; Brodley, C.E. Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach. In Proceedings of the 20th International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 186–193.

36. Zhao, R.; Mao, K. Semi-random projection for dimensionality reduction and extreme learning machine in high-dimensional space. *IEEE Comput. Intell. Mag.* **2015**, *10*, 30–41. [CrossRef]

37. Deegalla, S.; Bostrom, H. Reducing High-Dimensional Data by Principal Component Analysis vs. Random Projection for Nearest Neighbor Classification. In Proceedings of the 5th International Conference on Machine Learning & Applications (ICMLA'06), Orlando, FL, USA, 14–16 December 2006; pp. 245–250. [CrossRef]

38. Zhao, Q.H.; Jia, S.H.; Li, Y. Hyperspectral remote sensing image classification based on tighter random projection with minimal intra-class variance algorithm. *Pattern Recogn.* **2021**, *111*, 107635. [CrossRef]

39. Rathore, P.; Bezdek, J.C.; Erfani, S.M.; Rajasegarar, S.; Palaniswami, M. Ensemble fuzzy clustering using cumulative aggregation on random projections. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 1510–1524. [CrossRef]

40. Anderlucci, L.; Fortunato, F.; Montanari, A. High-dimensional clustering via Random Projections. *J. Classif.* **2021**, *38*, 191–216. [CrossRef]

41. Menon, A.K. Random Projections and Applications to Dimensionality Reduction. Bechelor's Thesis, The University of Sydney, Darlington, Australia, March 2007.

42. Sarvia, F.; Xausa, E.; Petris, S.D.; Cantamessa, G.; Borgogno-Mondino, E. A possible role of copernicus sentinel-2 data to support common agricultural policy controls in agriculture. *Agronomy* **2021**, *11*, 110. [CrossRef]

43. Zhan, Y.; Muhammad, S.; Hao, P.; Niu, Z. The effect of EVI time series density on crop classification accuracy. *Optik* **2018**, *157*, 1065–1072. [CrossRef]

44. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* **2020**, *250*, 112012. [CrossRef]

45. Zhong, Y.; Wang, X.; Xu, Y.; Wang, S.; Jia, T.; Hu, X.; Zhao, J.; Wei, L.; Zhang, L. Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 46–62. [CrossRef]