Uludağ University Journal of The Faculty of Engineering, Vol. 23, No. 3, 2018 RESEARCH

DOI:10.17482/uumfd.435723

A COMPARATIVE STUDY FOR HYPERSPECTRAL DATA CLASSIFICATION WITH DEEP LEARNING AND DIMENSIONALITY REDUCTION TECHNIQUES

Gizem ORTAÇ[*]
Gıyasettin ÖZCAN[**]

Received: 25.06.2018; revised: 27.09.2018; accepted: 16.10.2018

Abstract: In recent years, hyperspectral imaging has been a popular subject in the remote sensing community by providing a rich amount of information for each pixel about fields. In general, dimensionality reduction techniques are utilized before classification in statistical pattern-classification to handle high-dimensional and highly correlated feature spaces. However, traditional classifiers and dimensionality reduction methods are difficult tasks in the spectral domain and cannot extract discriminative features. Recently, deep convolutional neural networks are proposed to classify hyperspectral images directly in the spectral domain. In this paper, we present comparative study among traditional data reduction techniques and convolutional neural network. The obtained results on hyperspectral image data sets show that our proposed CNN architecture improves the accuracy rates for classification performance, when compared to traditional methods by increasing the classification accuracy rate by 3% and 6%.

Keywords: Hyperspectral Imaging, Deep Learning, Dimensionality Reduction, Classification, Convolutional Neural Networks,

Hiperspektral Verilerin Sınıflandırmasında Derin Öğrenme ve Boyut İndirgeme Tekniklerinin Karşılaştırılması

Öz: Son yıllarda, hiperspektral görüntüleme yüzey pikselleri ile ilgili zengin miktarda bilgi sağlamasıyla uzaktan algılama alanında popüler bir konu olmuştur. Genel olarak, elde edilen yüksek boyutlu ve ilişkisel veriyi işlemek için, sınıflandırmadan önce boyut indirgeme teknikleri uygulanmaktadır. Bununla birlikte geleneksel sınıflandırıcılar ve boyut azaltma yöntemleri, spektral alanda hala zorlu bir işlemdir ve ayırt edici öznitelikler çıkarmaz. Son zamanlarda ise derin konvolüsyonel sinir ağları, hiperspektral görüntüleri doğrudan spektral alanda sınıflandırmak için geliştirilmiştir. Önerilen çalışmada, geleneksel sınıflandırma ve konvolüsyonel sinir ağları arasında karşılaştırmalı bir çalışma ve analiz yapılmıştır. Çeşitli hiperspektral görüntü verilerine dayanarak elde edilen sonuçlar, önerilen konvolüsyonel sinir ağının, geleneksel yöntemlerden %3 ve %6 oranında daha iyi bir sınıflandırma oranı sağladığını göstermiştir.

Anahtar Kelimeler: Hiperspektral Görüntüleme, Derin Öğrenme, Boyut Azaltma, Sınıflandırma, Konvolüsyonel Sinir Ağları

1. INTRODUCTION

Hyperspectral remote sense imaging technology, HSI, is widely used for monitoring Earth's surface (Chang, 2003, P. F. Hsieh,1998). In contrast to traditional multispectral sensors with

[*] Bursa Technical University, Faculty of Engineering and Natural Sciences, Department Of Computer Engineering, 16330 Bursa, Turkey
[**] Uludağ University, Faculty of Engineering, Department Of Computer Engineering, 16059 Bursa, Turkey
Correspondence Author: Gıyasettin ÖZCAN (gozcan@uludag.edu.tr)

low spectral resolution, hyperspectral remote sensing imaging is advanced from the developments of hyperspectral sensors and provides better discrimination among ground cover classes (Scott, 2015). The sensors provide a vast amount of spectral and spatial information, comprise highly correlated and very narrow spectral bands under a specific spectral frequency. The information is exploited in HSI classification such as agriculture, environmental management, urban planning, mineral detection and urban mapping (Liang et.al., 2016).

Hyperspectral image comprises of two-dimensional images at a series of wavelengths. Spectral information is provided by the grey information of the same pixel point at each wavelength (Wang et.al., 2018). The traditional HSI classification based on pixel-wise approach (Landgrebe, 2005) that classifies each pixel by its digital numbers and reflectance values from different spectral bands. In particular, the classification introduces good performance due to the high spatial and spectral resolution, although some pitfalls can affect classification results negatively. For instance, training samples and spectral information (i.e., hundreds of correlated spectral bands) collection is complex and causes Hughes phenomenon (Hughes, 1968).

As a consequence, classification accuracy may be insufficient. The Hughes phenomenon, also known as the curse of dimensionality, emerges when the number of features and available training samples are unbalanced and causes complete failure of the traditional classifiers (Bazi et.al., 2006). On the other hand, the classification process can suffer from high-resolution images since the process can increase the intra-class variation or decrease the interclass variation in both spectral and spatial domains (Chen et.al., 2011).

In the literature, various studies have been carried out to overcome the issues. The studies are based on the following approaches (Bazi et.al., 2006):

1) The using of the sample covariance matrix (Hoffbeck et.al.,1996a, Tadjudin et.al., 1999);

2) The exploitation of the classified samples (Shahshahani, 1994, Jackson, 2001);

3) Reducing/transforming the original feature space into lower dimensionality with feature selection/extraction techniques (Lee et.al., 1993, Jimenez et.al., 1999);

4) Modeling the class spectral signatures with shape description techniques; and (Hoffbeck et.al., 1996b, Tsai et.al., 2002)

5) Support vector machine (SVM) classifiers (Gualtieri et.al,2000, Huang et.al., 2002, Melgani et.al.,2004, Camps-Valls et.al., 2004, Foody et.al., 2004, Camps-Valls et.al, 2006, Pal et.al., 2005)

Regarding classification, the transformation of a hyperspectral image into a meaningful domain without losing the relevant object information has become an important research topic, recently. Ideally, the reduced image should correspond to a minimum number of variables for efficient image modeling.

Instead of using the full spectral bands, dimensionality reduction techniques are effective methods for data processing and for finding the class-specific subspace. However, determination of the most effective dimensionality reduction technique is difficult in practice. In the early stage, spectral-based methods, including principal component analysis (PCA) (Licciardi et.al., 2012), independent component analysis (ICA) (Villa et.al., 2011), linear discriminant analysis (Villa et.al., 2011), etc. can be thought as linear transformations to extract better features of the input image in the lower dimensions (Bruce et.al., 2002, Jimenez et.al., 1999). Nonetheless, the linear transformation-based methods are not suitable for neither analyzing inherently nonlinear hyperspectral data (Chen et.al., 2016) nor in the existence of interference sources such as striping (Chang et.al., 1999).

In recent years, deep learning based methods also provide promising results to explore the higher level and more effective spatial features (Fang et.al., 2014). In the computer vision field, deep learning methods are designed as automatic multi-layer feature learning and exploration tools by using non-linear activation functions and provide more robust features compared to lower level ones (Fang et.al., 2014).

In deep learning, convolutional neural networks, CNNs, play a dominant role in the implementation on GPUs and have recently outperformed other conventional method (Hinton et.al., 2006). However, CNNs have been mostly used for visual-related problems, a relatively newer method for hyperspectral image classification.

A convolutional neural network is used to extract spectral and spatial feature maps by linear convolution filters followed by nonlinear activation functions. The classical CNNs were proposed by Lecun and has recently become popular in image processing applications including object detection (Bruna et.al., 2015), face recognition (Sun et.al., 2014), image denoising (Li, 2014).

In recent works, the convolutional neural networks have been used to learn the discriminating features to classify hyperspectral images adaptively. For instance, Hu et.al. (2015) developed a deep convolutional neural network and compared the experimental results for some traditional methods. The experimental results on different hyperspectral datasets showed that the proposed neural network architecture which was contained five layers with weights achieved better classification performance. Also, Chen et.al. (2016) presented a CNN-based deep feature extraction method for HSI classification. The proposed method performed on three public hyperspectral datasets with some state-of-the-art way and provided competitive results. Yu et.al. (2017) introduced an efficient CNN architecture that overcomes some limitations such as over-fitting. The designed architecture included different principles such as data augmentation, more substantial drop rates and discarding max-pooling layers. The experimental results for different hyperspectral datasets showed that the well-designed deep learning model CNNs can achieve better classification performance.

In summary, reduction of the spectral information is a necessary pre-processing step to hyperspectral analysis. Although, these methods can be affected by the small number of training samples and they usually need a large number of samples. They also suffer from unbalanced structure between curse of dimensionality of the data and the limited availability of training samples.

In this work, we develop a 2-D deep CNN model for classifying hyperspectral data after building appropriate architecture. The model presents a powerful tool to extract the spatial feature representation. We also produce a comparative study with traditional classifiers.

This paper is organized as follows: In Section 2, a brief introduction to CNN and dimensionality reduction is presented. In Section 3, the CNN architecture and training process is presented. In Section 4, we experimentally compare the performance of the CNN with the classification of lower-dimensional hyperspectral datasets generated by different dimensionality reduction techniques. Finally, we summarize our experimental results in Section 5.

2. DEFINITIONS AND RELATED WORK

In this section, some general aspects of CNN and dimensionality reduction in hyperspectral image classification are presented.

2.1. Convolutional Neural Networks

CNN is a special type of feed-forward neural network that is composed of one or more pairs of convolution layers and pooling layers. A CNN architecture can be designed according to different tasks such as image classification (Agarwal et.al., 2007), speech recognition (Xu et.al., 2015) and text recognition (Tuia et.al., 2014). However, there is relatively less CNN technique for HSI classification in the literature. In general, CNN is composed of the convolutional layers, pooling layers, and fully connected layers. Convolutional layer extracts the previous layer feature maps by using linear convolution filters. At least one layer of the nonlinear activation functions (e.g., rectifier, sigmoid, tanh, etc.) is applied to obtain the output feature map. Let $X \in R^{N\ x\ M}$ be a training input image or the layer and $n\ x\ n$ is a square region

extracted from the image and w be a weighted filter of kernel size with the size of $(m\ x\ m)$. The output layer is computed as:

$$h_{ij}^{l} = f\ (\sum_{k=0}^{m-1}\sum_{l=0}^{m-1} w_{ab} x_{(i+k)(j+l)}^{l-1} +\ b_{ij}^{l}\ ) \tag{1}$$

(Fotiadou et.al., 2014) where b is the bias term and $f(.)$ is an activation unit of the neuron. Every neuron is presented with a spatial location (i,j) concerning the input image in the convolutional layer.

The pooling layer provides a group of the local features from adjacent pixels to correct deformations of objects. The input is partitioned into a set of patches and returns the max or mean value for each partition. By pooling, down-sampled input maps are created to reduce computational complexity for the upper layers. The pooling operation is formulated as:

$$h_{ij}^{l} = f\ (\beta_j^{l} down\left(h_{ij}^{l-1} + b_{ij}^{l}\right)) \tag{2}$$

(Fotiadou et.al., 2014) where $down(.)$ is the sub-sampling function that sums over each distinct patch in the input feature and β is the multiplicative bias of the output feature maps.

The last layer is generally a fully-connected layer with a softmax function that generates the probability of class membership for each unit. The amount of neurons is equal to the number of classes to be categorized in a softmax layer. The last layer can be defined as (Liang et.al., 2016).

$$v_{lj}^{xy} = f\ (\sum_{m}\sum_{h=0}^{H_j-1}\sum_{w=0}^{W_j-1} k_{ljm}^{(x+h)(y+w)} +\ b_{lj}\ ) \tag{3}$$

where l is the layer that is processed, j is the number of feature maps in layer l. v_{lj}^{xy} is the output at position (x,y) in that feature map and layer. m indexes in the $(l-1)th$ layer connected to the current (jth) feature map and k_{ljm}^{hw} is the value at position (h,w) of the kernel connected to the jth feature map. H_j refer to the height and width of the spatial convolution kernel, respectively (Chen et.al., 2016).

In the proposed network, a hyperspectral image is considered as a 3D tensor of dimensions $h\ x\ w\ x\ c$ where h and w refers the height and width of the image and c is the spectral bands (channels). The images are decomposed into square patches to align with the specific nature of CNNs. Each square patch contains spectral and spatial information for a specific pixel p_{xy} to classification. l_{xy} is the class label of the pixel at location (x,y) and w_{xy} the patch centered at pixel p_{xy}. In final, the dataset is formed $D = \{(w_{xy}, l_{xy})\}$ for $X = 1,2,\cdots,w$ and $y = 1,2,\cdots,h$. Patch w_{xy} is also a 3D tensor with dimension $s \times s \times c$. It contains spectral and spatial information for the pixel located at (x,y). Parameter c corresponds to the number of spectral bands (Makantasiset. et. al., 2015).

2.2. Dimensionality Reduction Technique

The hyperspectral images are composed of several hundred images obtained with different frequencies. In general, the ability of classification increases with detailed information about the land cover. However, some reasons make the classification of pixels challenging such as high spectral resolution, insufficient training samples and a large number of bands. The computational time is significantly increased because of these reasons.

Dimensionality reduction transforms the data into a lower dimensional space. It is an effective method to eliminate irrelevant variance in the data and extract low-dimensional features which include some desired information. Instead of using the all spectral bands, the lower-dimensional representation with better specific subspace could effectively improve classification performance.

In the study, we consider Principal component analysis, PCA, linear discriminative analysis, LDA, and independent component analysis, ICA, Factor Analysis, FA and Truncated Singular Value Decomposition, SVD, has been applied as classical dimensionality reduction methods. PCA (Fukunaga, 2013) is the most widely used unsupervised dimensionality reduction method and removes the dependencies among the spectral bands by eigenvector decomposition. Therefore, it is often used in hyperspectral image processing (Rodarmel et.al., 2002). It generates a lower dimensional representation of data that describe as much of the large variance. It keeps the most significant singular vectors for the projection of the data to decrease dimensionality (Lee et.al., 1993). In the study, PCA utilizes Singular Value Decomposition, SVD. SVD is a method for performing PCA by diagonalization of the covariance matrix and principal components of data are calculated more efficient and robust way for transformation (Wall et.al., 2003).

LDA seeks the best projection that maximizes the between-class scatter while minimizing the within-class scatter. It optimizes the Fisher score and does not require the tuning of free parameters. Due to these reasons, LDA is extensively used in remote sensing and hyperspectral imaging for feature reduction (Bandos et.al., 2009). In the study, another linear dimensionality reduction method, Truncated SVD, is applied. This method does not center the data before computing the singular value decomposition contrary to PCA (Halko, et.al., 2011).

FA is a linear statistical method that is developed for potential factors from observed variables to replace the original data (Bartholomew et.al., 2008). It is a very useful method for high-dimensional data generation model since it allows different regions in the input space to build a model of local factor data (Wang et.al., 2015).

In this study, the effectiveness of CNN based model is tested by comparison of different dimensionality reduction and different classification methods with the low-dimensional data.

3. MATERIAL AND METHODS

3.1. Hyperspectral Datasets

For the experimentation, we exploit Indian Pines and Pavia University hyperspectral datasets which are prominent and publicly available.

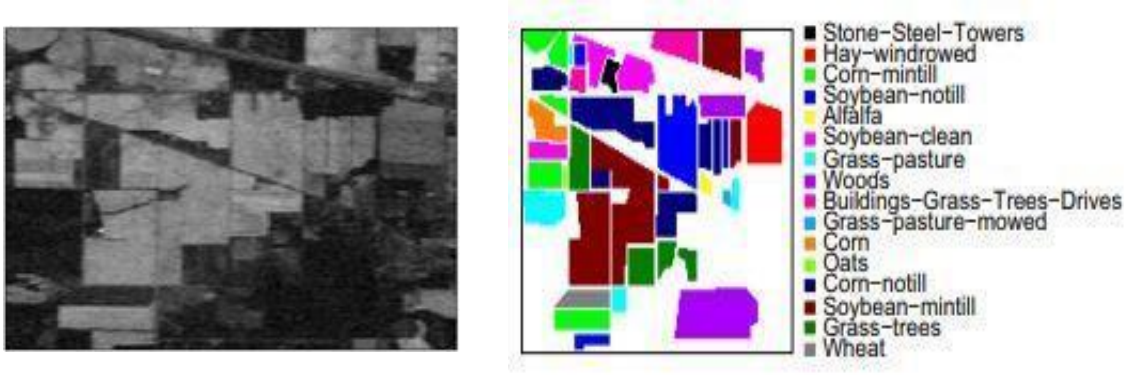
The Indian Pines dataset is collected by Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor from a test site in the northeast of Indian Pine state, the USA in 1992. The dataset contains $145\,x\,145$ pixels with $20\,m$ spatial resolution and 224 spectral bands in the wavelength range of $0.4-2.5\,\mu m$. 20 water absorption bands are $([104-108],[150-163],220)$. The dataset contains 10.249 labeled samples and a 16-classes ground-truth map (Gamba, 2004).

The Pavia University dataset (Engineering School at the University of Pavia, Pavia, Italy) is obtained by the reflective optics system imaging spectrometer (ROSIS-03) airborne optical sensor.

The dataset has *610 x 340* pixels with a spatial resolution of *1.3 m* and *103* spectral bands in the wavelength range 0.43–0.86 μm. Pavia University dataset has ground truth maps of *9* classes and *42.776* labeled samples (or pixels) (Huang et.al., 2009).

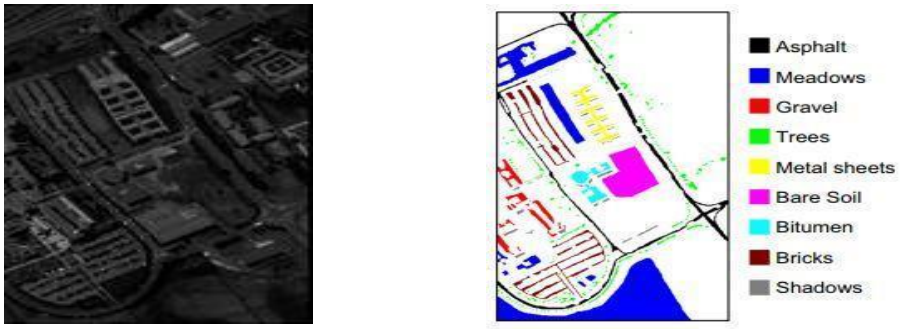
3.2. Experiment Setup

Different experiments are performed to evaluate the performance of classification and convolutional neural network approaches in Python environment (version 3, 64-bit) language and Tensorflow library (Abadi, et.al., 2016). The results are generated on a PC equipped with Intel(R) Core(TM) i7-7700HQ CPU @ 2.8 GHz Processor and 16.00 GB memory (RAM).



a. *b.*

Figure 1:
The Indian Pines hyperspectral data;
<http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes>
a. *a sample band and* ***b.*** *Ground-truth map of the Indian Pines dataset (sixteen land cover classes*



a. *b.*

Figure 2:
The Pavia University hyperspectral data;
<http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes>
a. *a sample band and* ***b.*** *Ground-truth map of the Pavia University dataset (nine land cover classes)*

3.3. The Architecture of the Proposed CNN

We present the architecture of our CNN in Figure 3. In the architecture, there exist 2 convolutional layers in the network. The convolutional kernel size, pixels of the first

convolutional layer, is *5x5* and the number of maps in this layer is *200*. The number of feature maps of the second layers is 100 and the size of each feature map is *3x3*. After each convolution step, a *2x2* max-pooling is operated on each channel. After these processes, we "flatten" the data in the third layer, i.e., stretch it to a 1-D vector, and feed it into two fully connected layers with *150* and *50* nodes. The output-layer size is set to be the same as the total number of classes. The ReLU non-linearity function is selected as the activation function to the output of every convolutional layer.

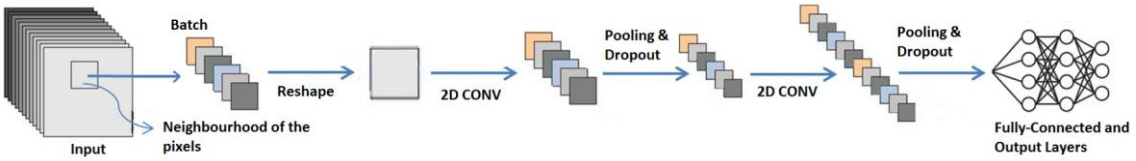


Figure 3:
The architecture of proposed CNN for HSI classification

In Table I, we present scheme of the proposed architecture in more detail. First, the hyperspectral images are split into 3-D patches. The size of the neighboring regions (patch size) in pixels is 5*x5x200* for Indian Pines and *5x5x103* for Pavia University. The created data with the patches divided into the number of parts (batches) that is the number of instances used in one iteration. Then, the batches are reshaped two-dimensional images and sent as input volume to the first convolutional layer, Conv1. After applying the RELU function, the generated feature maps by the first convolutional layers are sent to the first max pool layer (Pool1) with a *5x5* kernel. The resulting output volume is sent to the last convolutional layer (Conv2) with a *3x3* filter size. Again, after applying RELU function, the generated feature maps by Conv2 are sent to the second max pool layer (Pool2) with a *2x2* kernel. Since there is no third max pool layer, the output volume is reshaped to send it to fully-connected layers. Three fully-connected layers are implemented to the networks. The first two fully-connected layers (F1 and F2) compute the outputs according to their weights, their biases, the output of the previous layer and the activation function RELU. Finally, the last fully connected-layer (F3) computes the outputs of the network with a softmax function.

To minimize the loss function in a network, a backward propagation algorithm can be useful in a general way. Mostly, variations of the stochastic gradient descent algorithm (SGD) is applied to optimize the parameters (Liang, et.al., 2016) The optimizers require careful initialization and adjustment of the model hyper-parameters such as the learning rate used in optimization. The learning rate hyper-parameter controls the tuning the weights of out network respect the loss gradient. In this work, the Xavier initializer (Glorot et.al., 2010) is used to initialize of all weights and bias of the network.

The Adam optimizer is also implemented for optimizing the parameters *k* and *b*, trainable parameters (Kingma et.al., 2014). The Adam optimizer has various advantages such as working sparse gradients, naturally performing a form of step size annealing and invariant parameter updates to a rescaling of the gradient (Kingma et.al., 2014). In the study, the cross-entropy is used to determine the loss of the CNN and measure the deviation from the target and predicted labels. The network is trained by minimizing the cross-entropy loss function by the Adam optimizer (Kingma et.al., 2014).

Table 1. The Configuration of the 2-D Convolution Neural Network

Datasets	**Patch Size**	**Conv 1 Pool1**	**REL U**	**Conv 2 Pool2**	**REL U**	**F1**	**F2**	**F3**
Indian Pines	5x5x200	5x5 2x2	Yes	3x3 2x2	Yes	Fully Connected	Fully Connected	1x16
Pavia University	5x5103	5x5 2x2	Yes	3x3 2x2	Yes	Fully Connected	Fully Connected	1x9
Feature Map		200		100			150	50

The parameters can be updated according to the derivatives. *k* and *b* are determined by applying the backpropagation firstly. Then, new error derivatives are generated with a feed-forward step. These derivatives could be used for parameter updating for another round. The feed-forward and back-propagation steps are repeated until obtaining optimal *k* and *b* or a predefined number of iterations is reached (Liang et.al., 2016). In our study, the number of training iteration set in *2000*.

3.4. Application of Different FE Methods and Classifiers

Hyperspectral images are high-dimensional data with a limited number of training samples. Since training supervised classifiers are time-consuming and costly in classification, a small part of the data is used for training classifiers. In this set of experiments, CNN was compared with the effectiveness of different dimensionality reduction techniques performances through classification results.

In the dimensionality reduction step, we utilized Python's scikit-learn machine learning package (Pedregosa et.al.,2011). For a detailed comparison, we tested various unsupervised and supervised dimensionality reduction techniques which have been described in Section 2. The number of reduced dimensions is iteratively increased to find an appropriate dimension for each technique.

After dimensionality reduction is applied and new data is obtained, this data is dividing 10 groups called folds. In the process, the reduced data divided into *k* mutually subsets of equal size and each subset are used for training while the rest subsets are used for the test. After *k* times of classification, the average accuracy is calculated. Various classifiers in scikit-learn are performed to evaluate different dimensionality reduction techniques through classification results.

4. EXPERIMENTAL RESULTS AND VALIDATIONS

In the CNN training process, the training samples are divided into *100* batches with the equal number of samples, randomly. Approximately *60%* of the available samples were used as the training dataset, whereas remaining of them served as the test dataset in the experiment. The number of train and test samples of each class is presented in Table 2 and Table 3. The total number of training and test samples are *6153* and *4096* for Indian Pines, *25670* and *17106* for

Pavia University dataset. One batch is sent into the network for each iteration. The training process continues until it reaches the maximal number of iterations. In the test process, the test sample is sent into the trained network.

Table 2. The Configuration of the 2-D Convolution Neural Network

Classes	Train	Test
1	28	18
2	857	571
3	498	332
4	143	94
5	290	193
6	438	292
7	17	11
8	287	191
9	12	8
10	584	388
11	1473	982
12	356	237
13	123	82
14	759	506
15	232	154
16	56	37

Table 3. The Indian Pines dataset and per class training sets and corresponding test sets

Classes	Train	Test
1	3979	2652
2	11190	7459
3	1260	839
4	1839	1225
5	807	538
6	3018	2011

7	798	532
8	2210	1472
9	569	378

To verify that the proposed CNN is suitable for classifying hyperspectral data sets with limited training samples, we compare the CNN with different traditional classification techniques. The dimensionality reduction methods are also performed before the classification to improve the classification performance. The number of dimensions was found from *2* to *50* for two hyperspectral data sets, iteratively. Then, *k*-fold cross-validation is used to the reduced data in the current dimension for classification. The average classification results for all dimensionality of the data sets for the classifiers with the dimensionality reduction techniques are reported in Table 4 and Table 5. As seen from the tables, the maximum average accuracies of *87.23%* and *92.47%* are obtained with FA by Random Forest classifier for Indian Pines and Pavia University data sets. The experimental results also show that the FA algorithm outperforms than the other dimensionality reduction methods. FA assumes that variables within a particular group are highly correlated among themselves, but they have relatively small correlations with variables in a different group. While PCA is widely used in hyperspectral data analysis, it is not a useful dimensionality reduction method when the components of maximum variation do not coincide with a large intra-class variation.

Table 4. Average classification accuracies of dimensions from 2 to 50 for the Indian Pines dataset

Classifier DR Technique	Random Forest	Decision Tree	Gaussian Naive Bayes	Quadratic Discriminant Analysis	Logistic Regression
Factor Analysis (FA)	**87.23**	81.32	87.23	67.78	76.64
Independent Component Analysis (ICA)	74.03	65.00	74.03	53.56	63.292
Linear Discriminant Analysis (LDA)	81.62	75.81	81.65	79.51	82.44
Truncated SVD	77.17	70.77	77.17	59.29	65.13
Principal Component Analysis (PCA)	77.47	71.29	77.542	59.44	63.33

The classification results for the CNN is presented in Figure 4 and Figure 5 for the datasets. Compared with the conventional classification methods, the proposed CNN achieves higher accuracy using all spectral bands even with a small number of training samples. As seen in Figure 4 and Figure 5, the best accuracy of *95.24%* is obtained with *2000* iterations for Pavia

University using CNN. Moreover, the best accuracy result *(93.87%)* is obtained for Indian Pines with CNN. In Figure 6, we can observe the evaluation of the error regarding training iteration. The value of the loss function is decreased with an increasing number of iterations. The results demonstrate that the test accuracy is relatively increasing while the cost value is reducing for both datasets. Early stopping can be considered for the training process to reduce computational cost since the proposed CNN converge in almost 900 iterations. Concerning the conventional classification method, the suggested CNN architecture provide averagely 6% classification improvements and 3% Indian Pines and Pavia University, respectively. Obviously, the proposed CNN increased the classification accuracy significantly under insufficient training data.

Table 5. Average classification accuracies of dimensions from 2 to 50 for the Pavia University dataset

Classifier DR Technique	Random Forest	Decision Tree	Gaussian Naive Bayes	Quadratic Discriminant Analysis	Logistic Regression
Factor Analysis (FA)	**92.47**	89.97	92.45	83.12	92.39
Independent Component Analysis (ICA)	89.67	85.93	89.68	84.44	92.08
Linear Discriminant Analysis (LDA)	90.74	87.32	90.74	86.84	89.20
Truncated SVD	89.94	86.61	89.92	80.89	92.10
Principal Component Analysis (PCA)	89.65	86.59	89.65	81.46	92.08

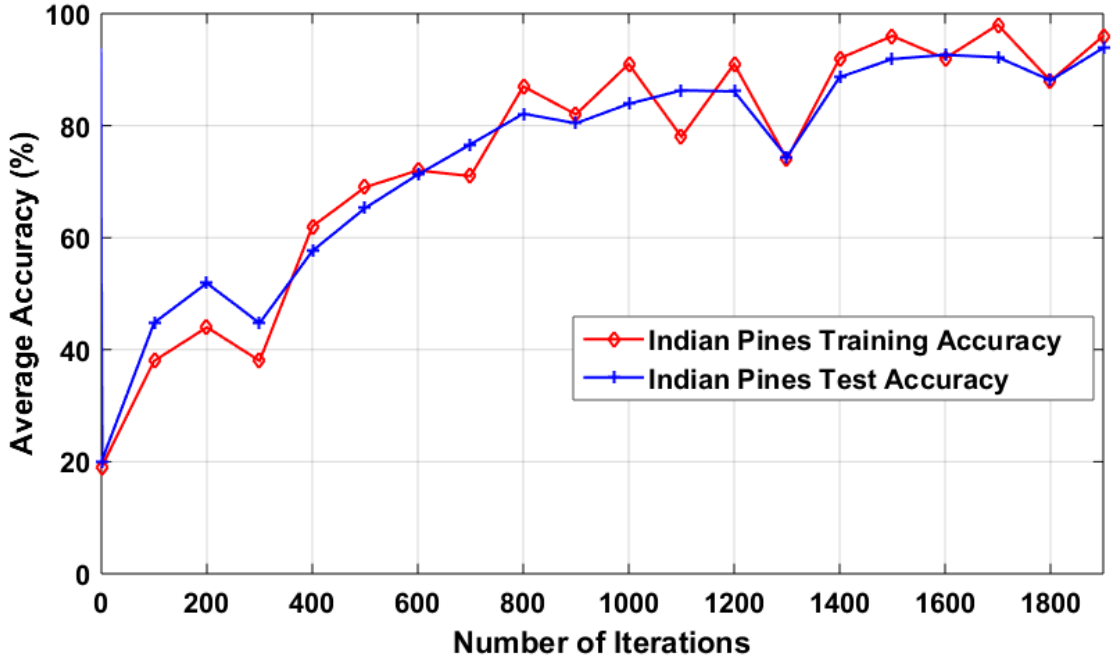


Figure 4:
Classification accuracies of CNN for the Indian Pines dataset

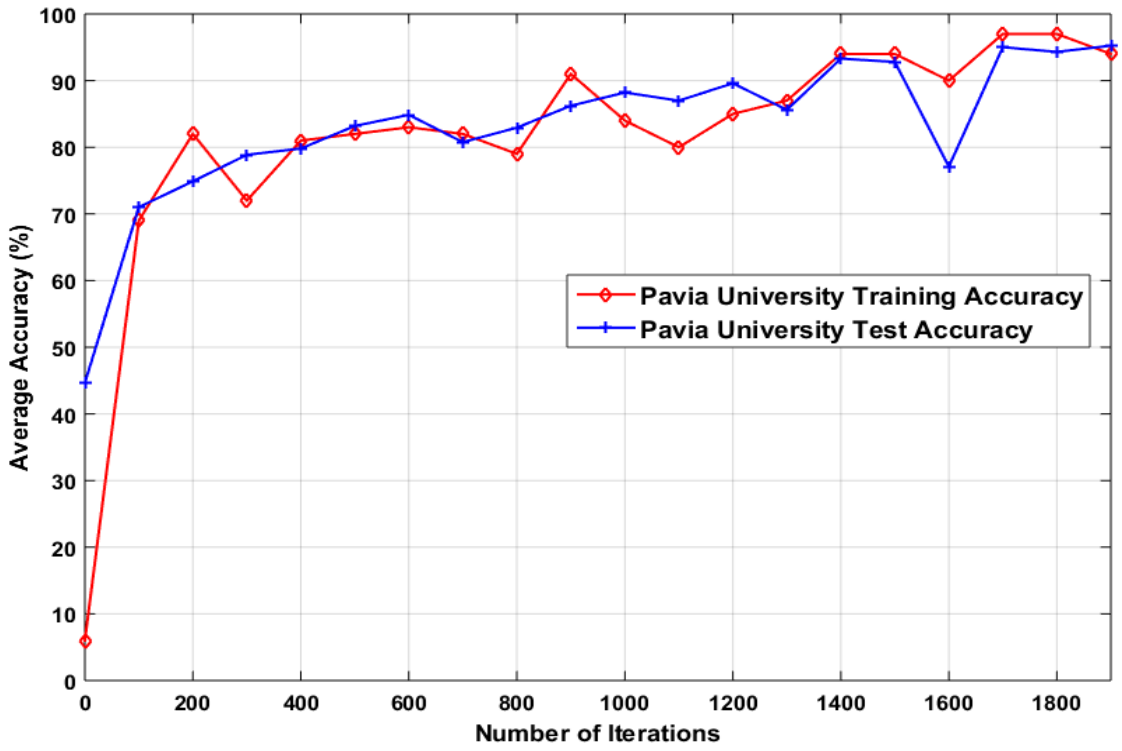


Figure 5:
Classification accuracies of CNN for the Pavia University dataset

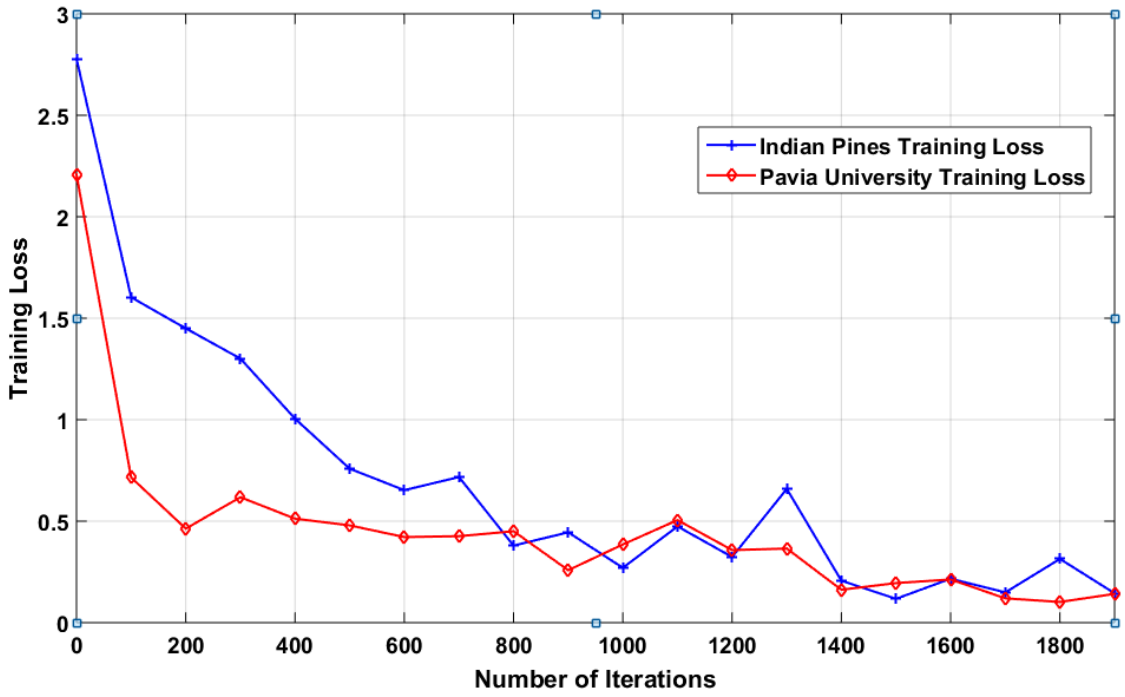


Figure 6:
Cost value versus the training iteration for the hyperspectral data sets

5. CONCLUSION

This study considered data classification problem on hyperspectral imagery, where the size of the data set is very large. To reduce the computational burden and improve classification accuracy, we utilized dimensionality reduction and deep learning techniques. We evaluated the most efficient the dimensionality reduction techniques and the proposed convolutional neural network using accuracy performance.

In hyperspectral imagery, dimensionality reduction without loss of critical information is one of the fundamental goals for efficient classification. However, finding the suitable dimensionality reduction technique is highly relying on domain knowledge.

Unlike conventional hyperspectral classification approaches, we propose a 2D CNN architecture for efficient classification. In the study, we compared our design to traditional dimensionality reduction and classification techniques on two publicly available hyperspectral datasets. Experimental results demonstrate that our CNN features can yield superior accurate results with using all spectral bands.

In the proposed CNN architecture, two convolutional and fully connected layers are used because of the limited number of training samples. We intend to improve multiple layers of CNN frameworks to improve our classification results, in the future works.

KAYNAKLAR

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016, November). TensorFlow: A System for Large-Scale Machine Learning. In OSDI (Vol. 16, pp. 265-283).

2. Agarwal, A., El-Ghazawi, T., El-Askary, H., & Le-Moigne, J. (2007, December). Efficient hierarchical-PCA dimension reduction for hyperspectral imagery. In Signal Processing and

Information Technology, 2007 IEEE International Symposium on (pp. 353-356). IEEE. DOI: 10.1109/ISSPIT.2007.4458191

3. Bandos, T. V., Bruzzone, L., & Camps-Valls, G. (2009). Classification of hyperspectral images with regularized linear discriminant analysis. IEEE Transactions on Geoscience and Remote Sensing, 47(3), 862-873. DOI: 10.1109/TGRS.2008.2005729

4. Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). Analysis of multivariate social science data. Chapman and Hall/CRC.

5. Bazi, Y., & Melgani, F. (2006). Toward an optimal SVM classification system for hyperspectral remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 44(11), 3374-3385. DOI: 10.1109/TGRS.2006.880628.

6. Bruce, L. M., Koger, C. H., & Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. IEEE Transactions on geoscience and remote sensing, 40(10), 2331-2338. DOI: 10.1109/TGRS.2002.804721.

7. Bruna, J., Sprechmann, P., & LeCun, Y. (2015). Super-resolution with deep convolutional sufficient statistics. arXiv preprint arXiv:1511.05666.

8. Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Martín-Guerrero, J. D., Soria-Olivas, E., Alonso-Chordá, L., & Moreno, J. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery. IEEE Transactions on Geoscience and Remote sensing, 42(7), 1530-1542. DOI: 10.1109/TGRS.2004.827262.

9. Camps-Valls, G., Gomez-Chova, L., Muñoz-Marí, J., Vila-Francés, J., & Calpe-Maravilla, J. (2006). Composite kernels for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 3(1), 93-97. DOI: 10.1109/LGRS.2005.857031.

10. Chang, C. I. (2003). Hyperspectral imaging: techniques for spectral detection and classification (Vol. 1). Springer Science & Business Media.

11. Chang, C. I., & Du, Q. (1999). Interference and noise-adjusted principal components analysis. IEEE transactions on geoscience and remote sensing, 37(5), 2387-2396. DOI: 10.1109/36.789637. DOI: 10.1109/36.789637.

12. Chen, S., & Zhang, D. (2011). Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 8(2), 369-373. DOI: 10.1109/LGRS.2010.2076407.

13. Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 6232-6251. DOI: 10.1109/TGRS.2016.2584107. DOI: 10.1109/TGRS.2016.2584107

14. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction L. O. Jimenez, D. A. Landgrebe (Nov. 1999) Hyperspectral data analysis and supervised feature reduction via projection pursuit", *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2653-2667. DOI: 10.1109/TGRS.2002.804721.

15. Fang, L., Li, S., Kang, X., & Benediktsson, J. A. (2014). Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 52(12), 7738-7749. DOI: 10.1109/TGRS.2014.2318058.

16. Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on geoscience and remote sensing, 42(6), 1335-1343. DOI: 10.1109/TGRS.2004.827257.

17. Fotiadou, K., Tsagkatakis, G., & Tsakalides, P. (2017). Deep Convolutional Neural Networks for the Classification of Snapshot Mosaic Hyperspectral Imagery. Electronic Imaging, 2017(17), 185-190. DOI: https://doi.org/10.2352/ISSN.2470-1173.2017.17.COIMG-445.

18. Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Academic press.

19. Gamba, P. (2004, September). A collection of data for urban area characterization. In Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International (Vol. 1). IEEE. DOI: 10.1109/IGARSS.2004.1368947.

20. Girshick, R. (2015) Fast R-CNN. In Proceedings of the International Conference on Computer Vision, Santiago, Chile,; pp. 1440–1448.

21. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

22. Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256).

23. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

24. Gualtieri, J. A., & Chettri, S. (2000). Support vector machines for classification of hyperspectral data. In Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International (Vol. 2, pp. 813-815). IEEE. DOI: 10.1109/IGARSS.2000.861712

25. Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review, 53(2), 217-288. DOI: https://doi.org/10.1137/090771806.

26. He, K., Zhang, X., Ren, S., & Sun, J. (2014, September). Spatial pyramid pooling in deep convolutional networks for visual recognition. In european conference on computer vision (pp. 346-361). Springer, Cham. DOI: 10.1109/TPAMI.2015.2389824.

27. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507. DOI: 10.1126/science.1127647.

28. Hoffbeck, J. P., & Landgrebe, D. A. (1996). Classification of remote sensing images having high spectral resolution. Remote Sensing of Environment, 57(3), 119-126.

29. Hoffbeck, J. P., & Landgrebe, D. A. (1996). Covariance matrix estimation and classification with limited training data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(7), 763-767. DOI: 10.1109/34.506799.

30. http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, Date of Access: 01.06.2018, Topic: Hyperspectral Remote Sensing Scenes

31. Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors, 2015. DOI: http://dx.doi.org/10.1155/2015/258619

32. Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. International Journal of remote sensing, 23(4), 725-749. DOI: https://doi.org/10.1080/01431160110040323

33. Huang, X., & Zhang, L. (2009). A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy. International Journal of Remote Sensing, 30(12), 3205-3221. DOI: https://doi.org/10.1080/01431160802559046

34. Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. IEEE transactions on information theory, 14(1), 55-63. DOI: 10.1109/TIT.1968.1054102.

35. Jackson, Q., & Landgrebe, D. A. (2001). An adaptive classifier design for high-dimensional data analysis with a limited training data set. IEEE Transactions on Geoscience and Remote Sensing, 39(12), 2664-2679. DOI: 10.1109/36.975001.

36. Jimenez, L. O., & Landgrebe, D. A. (1999). Hyperspectral data analysis and supervised feature reduction via projection pursuit. IEEE Transactions on Geoscience and Remote Sensing, 37(6), 2653-2667. DOI: 10.1109/36.803413.

37. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

38. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

39. Landgrebe, D. A. (2005). Signal theory methods in multispectral remote sensing (Vol. 29). John Wiley & Sons.

40. Lee, C., & Landgrebe, D. A. (1993). Feature extraction based on decision boundaries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(4), 388-400. DOI: 10.1109/34.206958.

41. Li, H. (2014). Deep learning for image denoising. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(3), 171-180. DOI: http://dx.doi.org/10.14257/ijsip.2014.7.3.14

42. Liang, H., & Li, Q. (2016). Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sensing, 8(2), 99. DOI:10.3390/rs8020099.

43. Licciardi, G., Marpu, P. R., Chanussot, J., & Benediktsson, J. A. (2012). Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. IEEE Geoscience and Remote Sensing Letters, 9(3), 447-451. DOI: 10.1109/LGRS.2011.2172185.

44. Liu, F., Shen, C., & Lin, G. (2015). Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5162-5170).

45. Makantasis, K., Karantzalos, K., Doulamis, A., & Doulamis, N. (2015, July). Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International (pp. 4959-4962). IEEE. DOI: 10.1109/IGARSS.2015.7326945.

46. Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on geoscience and remote sensing, 42(8), 1778-1790. DOI: 10.1109/TGRS.2004.831865.

47. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).

48. P. F. Hsieh (1998) D. Landgrebe, Classification of high dimensional data.

49. Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. International Journal of Remote Sensing, 26(5), 1007-1011. DOI: https://doi.org/10.1080/01431160512331314083

50. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

51. Rodarmel, C., & Shan, J. (2002). Principal component analysis for hyperspectral image classification. Surveying and Land Information Science, 62(2), 115.

52. Scott, D. W. (2015). Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons.

53. Shahshahani, B. M., & Landgrebe, D. A. (1994). The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Transactions on Geoscience and remote sensing, 32(5), 1087-1095. DOI: 10.1109/36.312897.

54. Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. In Advances in neural information processing systems (pp. 1988-1996).

55. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. Cvpr.

56. Tadjudin, S., & Landgrebe, D. A. (1999). Covariance estimation with limited training samples. IEEE Transactions on Geoscience and Remote Sensing, 37(4), 2113-2118.

57. Tsai, F., & Philpot, W. D. (2002). A derivative-aided hyperspectral image analysis system for land-cover classification. IEEE Transactions on Geoscience and Remote Sensing, 40(2), 416-425. DOI: 10.1109/36.774728.

58. Tuia, D., Volpi, M., Dalla Mura, M., Rakotomamonjy, A., & Flamary, R. (2014). Automatic feature learning for spatio-spectral image classification with sparse SVM. IEEE Transactions on Geoscience and Remote Sensing, 52(10), 6062-6074. DOI: 10.1109/TGRS.2013.2294724.

59. Villa, A., Benediktsson, J. A., Chanussot, J., & Jutten, C. (2011). Hyperspectral image classification with independent component discriminant analysis. IEEE transactions on Geoscience and remote sensing, 49(12), 4865-4876. DOI: 10.1109/TGRS.2011.2153861.

60. Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In A practical approach to microarray data analysis (pp. 91-109). Springer, Boston, MA.

61. Wang, S., & Wang, C. (2015). Research on dimension reduction method for hyperspectral remote sensing image based on global mixture coordination factor analysis. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(7), 159. DOI:10.5194/isprsarchives-XL-7-W4-159-2015.

62. Wang, Y., Lv, Y., Liu, H., Wei, Y., Zhang, J., An, D., & Wu, J. (2018). Identification of maize haploid kernels based on hyperspectral imaging technology. Computers and Electronics in Agriculture, 153, 188-195. DOI: https://doi.org/10.1016/j.compag.2018.08.012.

63. Xu, C., Lu, C., Gao, J., Zheng, W., Wang, T., & Yan, S. (2015). Discriminative analysis for symmetric positive definite matrices on lie groups. IEEE Transactions on Circuits and Systems for Video Technology, 25(10), 1576-1585. DOI: 10.1109/TCSVT.2015.2392472.

64. Yu, S., Jia, S., & Xu, C. (2017). Convolutional neural networks for hyperspectral image classification. Neurocomputing, 219, 88-98. DOI: https://doi.org/10.1016/j.neucom.2016.09.010