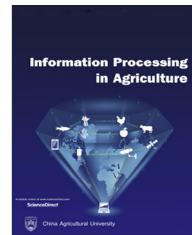


Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE xxx (xxxx) xxx

journal homepage: www.elsevier.com/locate/inpa

Dimensionality reduction of hyperspectral images of vegetation and crops based on self-organized maps

David Ruiz Hidalgo*, Bladimir Bacca Cortés, Eduardo Caicedo Bravo

Perception and Intelligent Systems Research Group, School of Electrical and Electronic Engineering, Universidad del Valle, Santiago de Cali 76001, Colombia

ARTICLE INFO

Article history:

Received 14 August 2019

Received in revised form

11 July 2020

Accepted 19 July 2020

Available online xxxx

Keywords:

Dimensionality reduction

Hyperspectral images

Principal component analysis

Radial basis function

Remote sensing

Self-organized maps

Wavelet transform

ABSTRACT

Hyperspectral images are multidimensional massive sets of information that have shown a great potential for different kind of applications as urban mapping, environmental management, vegetation and crops supervision and mineral detection. However, due to its high dimensional nature and the high variability of the spectral information, the dimensionality reduction process is one of the main challenges in processing hyperspectral images. The aim of dimensionality reduction is to eliminate redundant information and simplify the subsequent processes of classification and the search of information. In this context, several dimensionality reduction methods have been proposed, but most of them are not flexible enough to deal with the particular features of the hyperspectral images. In this way, the use of intelligent methods as neural networks and specially an unsupervised approach as self-organized maps, may improve the dimensionality reduction stage and the final classification process. This paper proposes an unsupervised method for the dimensionality reduction of hyperspectral images based on Kohonen self-organized maps, which, compared with other traditional methods such as principal component analysis (PCA) and wavelet decomposition, provides better classification results. The results provided in this paper use an RBF (radial basis function) classifier. On average, the proposed method provides a 64% dimensionality reduction and an 88.5% classification accuracy. These results suggest that the dimensionality reduction algorithm based on self-organized maps is an efficient approach compared with other popular algorithms. This is due to the ability of self-organized maps to automatically detect (self-organizing) relationships within the set of input patterns, which provides flexibility to deal with the special features of the hyperspectral images.

© 2020 China Agricultural University. Production and hosting by Elsevier B.V. on behalf of KeAi. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Recent advances in remote sensing have shown great potential for applications as urban mapping, environmental management, vegetation and crops supervision and mineral

* Corresponding author.

E-mail addresses: ruiz.david@correounivalle.edu.co (D. Ruiz Hidalgo), bladimir.bacca@correounivalle.edu.co (B. Bacca Cortés), eduardo.caicedo@correounivalle.edu.co (E. Caicedo Bravo).

Peer review under responsibility of China Agricultural University.

<https://doi.org/10.1016/j.inpa.2020.07.002>

2214-3173 © 2020 China Agricultural University. Production and hosting by Elsevier B.V. on behalf of KeAi.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

detection [1,2]. Talking about vegetation and crops supervision, currently it is possible to make diagnostics and discriminate species of crops through hyperspectral image processing [3]. This is possible because a hyperspectral image is multidimensional massive set of information that can be used to identify special features in a scene.

Although, the high spectral dimensionality of the hyperspectral images benefits the precision of pattern recognition, it challenges the computational capacity and the performance of conventional signal processing techniques [4]. A hyperspectral image of 145×145 pixels and 220 spectral bands has approximately 4.6 million data points. This volume of information generates several challenges in data processing and analysis [5]. Additionally, in hyperspectral image analysis it is common to find that many spectral bands are correlated, which implies processing of redundant information [6]. Therefore, one of the main steps in hyperspectral data processing is dimensionality reduction. Dimensionality reduction allows the identification and removal of redundant information that could drastically reduce the classifier performance [7]. In this way, the aim of dimensionality reduction in hyperspectral image processing is to reduce computational costs and the unnecessary use of resources without losing information quality [7]. A correct choice for a subset of spectral information generated from the original data set directly affects speed and system efficiency [6–8].

Dimensionality reduction methods used to create a subset of data can use supervised or unsupervised approaches [8,9]. Supervised methods use a priori knowledge about the classes of data or the phenomenon of interest. It means that these methods are based on the existence of datasets with labeled samples to infer class separability. Some of the most used dimensionality reduction methods for hyperspectral image processing are linear discriminant analysis (LDA) [10], non-parametric weighted feature extraction (NWFE) [11], sequential forward band extraction (SFBE) [12] or genetic algorithms [13]. Recent supervised dimensionality reduction methods for hyperspectral images are oriented to reduce the dimensionality of the data while processing the image in a spatial level. Proposals as Fisher's LDA [14], regularized local discriminant embedding [15], supervised band selection [16], discriminative local-metric-learning [17] or sparse-graph-based discriminant analysis [7] are able to perform a dimensionality reduction of data using local neighborhood information to add spatial information to data generated. However, the main disadvantage of these methods is the necessity of use labeled samples to perform the dimensionality reduction process. Many times, this kind of information is not available, which why unsupervised dimensionality reduction methods has gain attention for remote sensing applications.

Unsupervised dimensionality reduction methods deal with the problem of absence of labeled samples and seek to find another representation of the data in a lower dimensional space [4]. In this way, several methods have been proposed. Table 1 presents some unsupervised methods used for the dimensionality reduction of hyperspectral data [6].

The different methods presented in this table are divided into four groups of unsupervised methods based on their operation principle.

Table 1 shows that some of the methods based on information content are theoretical knowledge [5,6], band variance [6] and entropy [6]. The theoretical knowledge method is based on understanding how the target feature interacts with different spectral reflectance bands. Many vegetation indices are generated in this way [6]. Band variance and spectral entropy are measures of information content. High values of variance or entropy can represent high spatial crop variability or a high percentage of information content. These methods are simple and easy to implement but do not allow for the rejection of redundant bands [6].

A second kind of dimensional reduction method presented in Table 1 is based on projections such as ICA or PCA, which turn highly dimensional information into small dimensional spaces through projections or linear combinations [19]. PCA is one of the most commonly used methods in hyperspectral image analysis for dimensional reduction [22].

A third kind of approach for dimensionality reduction presented in Table 1 is based on similarity measures such as band correlation, graphics lambda (λ_1) vs lambda (λ_2) or spectral derivatives. These methods look for similarity indices between each spectral band. These kinds of methods allow the rejection of redundant data, but they increase computational costs [6].

Finally, methods based on frequency analysis, such as wavelet decomposition, are a good choice for feature and relevant information extraction [23–25]. An object may be easily identifiable based on the shape of its spectral signature. Wavelet-based approaches decompose a signal into approximation and detail coefficients, which represent the low- and high-frequency components, respectively. Then, the fewest wavelet coefficients required to perform dimensionality reduction are selected [25]. According to [25], wavelet-based dimensional reduction methods show better classification performance compared to PCA because they allow the local energy variation of hyperspectral signals in different spectral bands to be detected.

These proposals are some of the dimensionality reduction methods used in hyperspectral image processing. Nevertheless, several of these methods have shown not to be flexible enough to address the variability of the hyperspectral images [26]. The curse of the dimensionality, the interclass similarity and the intraclass variability make it extremely difficult to extract representative data of hyperspectral images and achieve a robust classification algorithm [27,4]. In this way, the use of intelligent methods as neural networks and specially an unsupervised approach as self-organized maps, may improve the dimensionality reduction stage and the final classification process.

This paper proposes a dimensional reduction method based on self-organized maps to be used as an unsupervised dimensional reduction approach. Self-organized maps (SOMs) are inspired by the capability of the human brain to extract relevant features from the world [28]. SOM networks

Table 1 – Unsupervised Methods of Dimensionality Reduction.

Unsupervised Methods	Algorithm
Based on Information Content	<ul style="list-style-type: none"> Theoretical Knowledge [5,6]. Band Variance [6]. Entropy [6].
Projection Based	<ul style="list-style-type: none"> PCA, Principal Component Analysis [18]. ICA, Independent Component Analysis [19].
Similarity Measures	<ul style="list-style-type: none"> Bands Correlation [5]. Graphs lambda (λ_1) vs lambda (λ_2) [6]. Spectral Derivative Analysis [5].
Frequency Analysis	<ul style="list-style-type: none"> Wavelet Decomposition [20,21].

automatically detect (self-organizing) relationships within the set of input patterns. This property can be applied to the problem of image mapping from higher dimensions to a two-dimensional feature set space to perform image classification [29]. This new data set will preserve the main features of the original data [28,30].

The paper is structured as follows. In Sections 2.1 and 2.2, some dimensional reduction methods are shown. Section 2.3 presents the dimensional reduction method based on self-organized maps. Section 3 presents the performance results of the self-organized map reduction method, which is compared with PCA and wavelet decomposition, using a radial base function (RBF) neural network classifier [28]. Conclusions and future works are given in Section 4.

2. Materials and methods

This section presents the three algorithms used in this paper to compare and develop the proposed dimensionality reduction method. Sections 2.1 and 2.2 describe two algorithms which are some of the most used and relevant for hyperspectral data dimensional reduction: PCA and wavelet decomposition [6,18–20,24,25]. The performance obtained through these algorithms will be compared with the dimensional reduction method based on self-organized maps, which is presented in the Section 2.3.

2.1. Principal component analysis (PCA)

PCA is a multivariable method used for dimensionality reduction [19]. The principal components are linear combinations based on the original band group. Each principal component is orthogonal to each other, and in this way, they eliminate redundant information [18].

Principal components are organized according to variance levels from highest to lowest. In this way, most information included in the hyperspectral image is captured in a few com-

ponents. Then, PCA can achieve dimensional reduction of the information [19].

$$X_{(n \times p)} = \begin{bmatrix} X_{11} & \cdots & X_{1j} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{i1} & \cdots & X_{ij} & \cdots & X_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nj} & \cdots & X_{np} \end{bmatrix} \quad (1)$$

$$C_{(p \times p)} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1p}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{p1}^2 & \cdots & \sigma_{pp}^2 \end{bmatrix} \quad (2)$$

In the context of hyperspectral data analysis, to estimate the principal components associated with a hyperspectral image with a number of pixels $n = k \times l$, where k is the image width and l is the image height, the $X_{(n \times p)}$ matrix is built, where p are all spectral bands, including all pixels and band information from the original data set [18]. This is shown in Eqs. (1) and (2), where $C_{(p \times p)}$ represents the covariance matrix computed from the $X_{(n \times p)}$ matrix. The principal components are obtained through linear combinations of $X_{(n \times p)}$, as shown in Eq. (3).

$$Z_{f(n \times p)} = X_{(n \times p)} U_{f(p \times 1)} \quad (3)$$

where $Z_{f(n \times p)}$ represents each of the principal components and f takes values between one and the total number of p bands; U_f contains the eigenvectors of $C_{(p \times p)}$ matrix [18], and its relationship with the covariance matrix is depicted in Eq. (4).

$$C_{(p \times p)} U_f = \lambda_f U_f \quad (4)$$

where λ_f are the eigenvalues of the covariance matrix described in Eq. (2). The total variance proportion associated with each component Z_f is calculated through the expression represented in Eq. (5).

$$\frac{\lambda_i}{\lambda_1 + \cdots + \lambda_p} \quad (5)$$

2.2. Wavelet decomposition

Wavelet decomposition is a method based on kernel functions that works with frequency components of the hyperspectral signals. The method allows for the analysis of the signals at different scales and resolutions. This makes it possible to identify the localization of variations in the spectrum, noise reduction, and redundant data reduction [6,19,21].

Wavelet decomposition is a mathematical tool that converts a signal into a different form. This conversion has the goal of revealing features hidden within the original signal and representing a more condensed form of the original signal [31].

The fundamental operator of wavelet decomposition is called the wavelet mother function (ψ_{jk}), which, through dilatation j and translation k operations, enables the decomposition of the original signal into two components: approximation and detail [21].

For a hyperspectral image $X_{(n \times p)}$ with n number of pixels and p spectral bands, the approximation components W_ϕ and the detail components W_ψ for each spectral vector x_i are calculated through Eqs. (6)–(8) [32].

$$W_\phi(0, k) = \frac{1}{\sqrt{p}} \sum_{l=1}^p x_i(l) \psi_{0k}(l) \quad (6)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{p}} \sum_{l=1}^p x_i(l) \psi_{jk}(l) \quad (7)$$

$$p = 2^j, j = [0, 1, \dots, J - 1], \\ k = [0, 1, \dots, 2^j - 1] \quad (8)$$

For data compression purposes, the best way to describe the wavelet decomposition method is through a series of cascaded filters. The input spectral vectors x_i from the image $X_{(n \times p)}$ feed low- and high-pass filters represented by approximation W_ϕ and detail W_ψ components, respectively [33]. The approximation components include most of the information content of the original signal. For this reason, in the processes of compression or dimensional reduction of information, it is possible to neglect the detailed components and only work with the approximation components [23,24]. Fig. 1 shows the structure of cascaded filters used by wavelet decomposition to develop the compression process for each one of the spectral vectors x_i [22,33].

2.3. Dimensional reduction of hyperspectral images based on self-organized maps

In this section, a dimensional reduction process based on self-organized maps is proposed, which aims to improve the performance of a crop classification process. This process is depicted in Fig. 2. First, the hyperspectral image is projected into a lower dimensional space, which is performed through the dimensional reduction process proposed in this work and based on self-organized maps. After using the new subset of data obtained by the dimensional reduction process, a classification process is implemented that generates a classification map labeling different types of vegetation and crops contained in the input image.

Self-organized maps (SOMs) are inspired by the fact that the human brain can extract relevant features from the world [28]. In this work, self-organized maps are used as a grouping tool of high-dimensional data in hyperspectral images. This kind of network contains no hidden layers but is made up of one input and one output layer. The SOM network automatically detects (self-organizing) relationships within the set of input patterns. This property can be applied to the problem of image mapping from higher dimensions to a two-dimensional feature space to perform image classification [29].

The number of neurons in the input and output layers defines a SOM network. The number of input neurons is equal to the number of input features. However, there are no clear rules regarding the specification of the number of output neurons. Generally, the output layer of an SOM is a two-dimensional layer composed of $M \times M$ neurons [28].

The neurons in the input layer and the output layer are linked by synaptic weights w_{ij} , where i and j are the identifiers of the input and output neurons, respectively. The weights w_{ij} are initialized randomly and are then continually adjusted during training to organize the relationships among the input patterns. Once the training is complete, the final weights w_{ij} represent the characterization of the input features [28]. An example of a SOM structure is shown in Fig. 3. Basically, it consists of making a projection of a multidimensional input space into an output space with lower dimensions. [28,30].

The SOM training strategy is based on the concept of competitive learning. The neurons in the output layer must compete with each other to “win” the opportunity to interact with the input pattern. As a result, the weights connecting

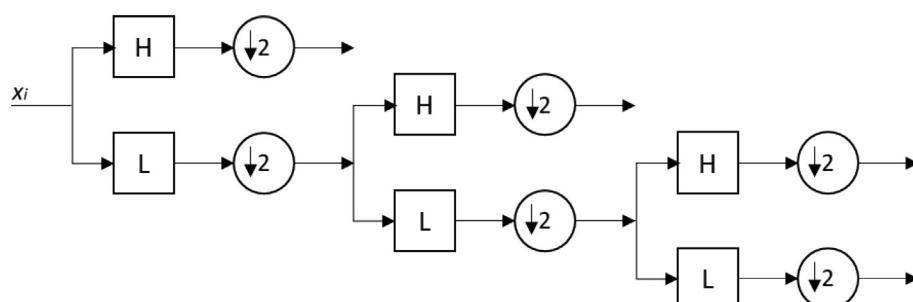


Fig. 1 – Cascaded Filters Used by Wavelet Decomposition.

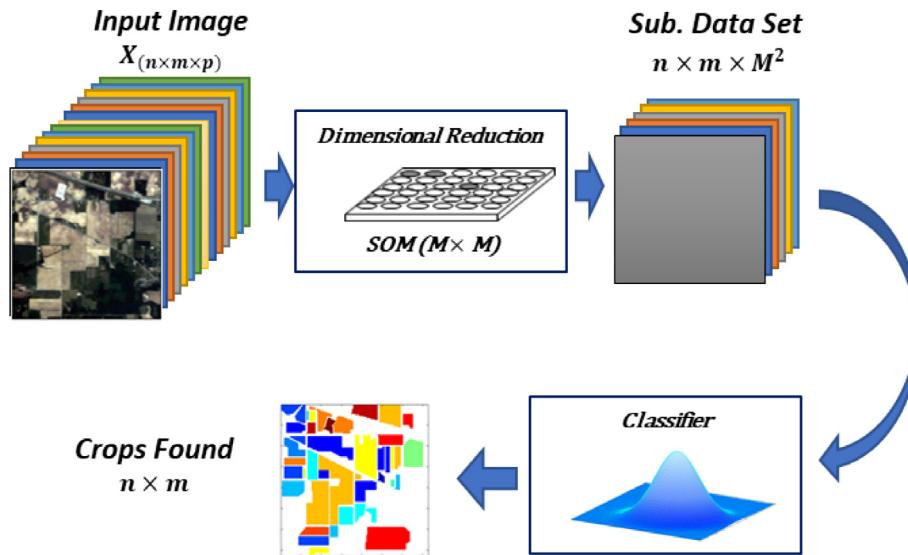


Fig. 2 – Diagram of Reduction Process and Crops Classification.

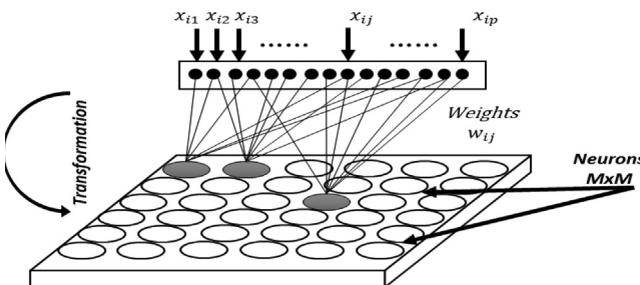


Fig. 3 – Self-organized Map Structure.

the input layer to the winning neuron and its neighbors are adjusted simultaneously, while other weights remain unchanged.

The dimensional data reduction process proposed in this paper uses Kohonen self-organized maps to project the p spectral bands of each pixel vector x_i of the original data set into a reduced dimensional space of $M \times M$ neurons.

Then, taking a hyperspectral image $X_{(N \times p)}$ where $N = n \times m$ is the total number of pixels and p is the number of spectral bands, the input vector to the SOM is defined as $x_{i(p \times 1)}$, which contains each spectral feature of the i^{th} pixel; then, though an iterative process, the Kohonen self-organized map generates a synaptic weights matrix $w_{ij(M^2 \times p)}$, which multiplies each spectral input vector $x_{i(p \times 1)}$ to generate output vectors with dimension M^2 . In this way, the dimensionality reduction is proportional to the self-organized map size. Choosing a self-organized map where $M \times M < p$, a dimensional reduction ratio of $1 - M^2/p$ can be obtained, as shown in Eq. (9).

$$\hat{x}_{i(M^2 \times 1)} = w_{ij(M^2 \times p)} x_{i(p \times 1)} \quad (9)$$

$$\%Reduction = 1 - M^2/p$$

Algorithm 1 shows the pseudo-code that describes each stage of the dimensional reduction process proposed in this work and based on self-organized maps.

Algorithm 1 (Dimensional Reduction Based on SOM).

```

1. Load input image  $X_{(N \times p)} = [x_{1(p \times 1)}, \dots, x_{i(p \times 1)}, \dots, x_{N(p \times 1)}]$ 
2. Define SOM  $M \times M < p$ 
3. Define weights matrix  $w_{ij(M^2 \times p)} = [w_1, \dots, w_r, \dots, w_{M^2}]$ 
4. Define control variables  $\sigma_{ini}, \sigma_{end}, \alpha_{ini}, \alpha_{end}, t_{max}$ 
5. While  $t < t_{max}$ 
  6. Select a random input vector  $x_{i(p \times 1)}$ 
  7. Find the winner neuron index  $s = argmin \|x_i - w_{ij}\|$ 
  8. Update each weight  $w_{ij}$ 
  9. For  $r = 1$  to  $M^2$ 
    10.  $\Delta w_{ij} = \alpha(t) h_{js}(t) (x_i - w_{ij})$ 
    11.  $w_{ij} = w_{ij} + \Delta w_{ij}$ 
  12. End For
  13.  $t = t + 1$ 
14. End While
15.  $\hat{X}_{(N \times M^2)} = w_{ij(M^2 \times p)} * X_{(N \times p)}$ 

```

Algorithm 1 starts by loading $X_{(N \times p)}$ and defining a self-organized map with $M \times M$ neurons and the synaptic weights matrix $w_{ij(M^2 \times p)}$, which is initialized randomly as shown in lines 1 to 3. The iterative process selects a random input vector $x_{i(p \times 1)}$ (line 6) through which the winner neuron location s is calculated based on the lower distance between the input vector x_i and each weight vector w_{ij} as described in line 7 through Eq. (10).

$$s = argmin \|x_i - w_{ij}\| \quad (10)$$

The synaptic weight variation Δw_{ij} (line 10) of the self-organized map is defined by Eq. (11), which is a function of the input vector x_i , the learning rate $\alpha(t)$, the neighborhood function $h_{js}(t)$, each synaptic weight w_{ij} and the iteration index t .

$$\Delta w_{ij} = \alpha(t) h_{js}(t) (x_i - w_{ij}) \quad (11)$$

The learning rate $\alpha(t)$ ranges between α_{ini} and α_{end} , the initial and final learning rate, and t_{max} is the highest iteration number allowed for the training process. The aim of Eq. (12) is that, at the beginning of the training process, the neurons fit quickly to the input data. When the process advances, the learning rate is lower to obtain a better fit in training.

$$\alpha(t) = \alpha_{ini} (\alpha_{end} \alpha_{ini})^{t/t_{max}} \quad (12)$$

The neighborhood function $h_{js}(t)$ is calculated using Eq. (13), which is a Gaussian function. In this function, the farther the neuron is from the winner neuron, the less the neuron is being affected.

$$h_{js}(t) = e^{\left(\frac{-d_{js}^2}{2\sigma_{ini}^2 (\sigma_{end}/\sigma_{ini})^{2t/t_{max}}} \right)} \quad (13)$$

where d is the Euclidean distance between the winner neuron s and neuron r ; here, the synaptic weights of neuron j are modified. σ is the neighborhood variable rank that guarantees the stability of the learning algorithm convergence [28].

At the end of the iterative process previously described, the synaptic weights matrix $w_{ij(M^2 \times p)}$ is obtained. It is multiplied by the input image $X_{(n \times m \times p)}$ to generate an output matrix \hat{X} with dimension $n \times m \times M^2$ (line 15). Fig. 4 shows the block diagram corresponding to the dimensional reduction process described in this paper. The appropriate value of M is obtained in each application heuristically. This is shown in the tests and results section.

After the previous dimensional reduction process, a data subset with $n \times m$ pixels and M^2 dimensions are obtained. In this work, this data subset is processed with a classifier that allows identification of different kinds of crops in the original data set, as shown in Fig. 2.

2.4. Radial basis function (RBF) neural networks

The classifier used to process the data generated by the dimensional reduction stage $\hat{X}_{(N \times M^2)}$ is a radial basis function (RBF) neural network. Fig. 5 presents the structure of the net, where the inputs are each one of the vectors $\hat{x}_{i(M^2 \times 1)}$ of the

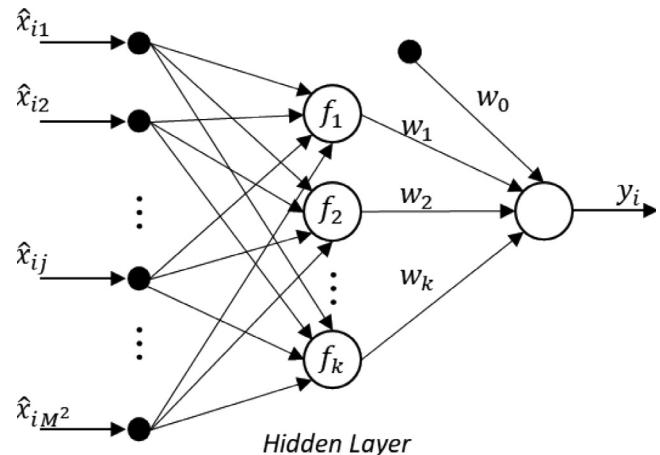


Fig. 5 – Structure of the Radial Basis Function Network.

matrix $\hat{X}_{(N \times M^2)}$. The output of the network is the label of the different types of vegetation and crops contained in the input image.

An RBF classifier is a neural network that allows solving nonlinear classification problems. Neural networks are non-parametric methods that have high dimension decision boundaries adjusted by an interactive process. RBF classifiers use a radial basis activation function as described in Eq. (14) [28].

$$f(x) = w_0 + \sum w_i e^{-\left(\frac{\|x_i - c_i\|^2}{2\sigma_i^2}\right)} \quad (14)$$

where x_i is a vector of features, c_i is the center of a class, σ_i is the variance, and w_i are the synaptic weights used to adjust the decision boundaries. The values of c_i , and w_i are calculated through a supervised training method. This training method is presented in detail in [22]. The trained network generates the final classification map, as shown in Fig. 2.

3. Results and discussion

To validate the dimensional reduction algorithm based on self-organized maps, three publicly available hyperspectral image data sets taken by the Airborne Visible-Infrared Imaging Spectrometer sensor (AVIRIS) were used [34]. Field aerial pictures of datasets selected are shown in Fig. 6. These hyperspectral images were selected because they have ground truth

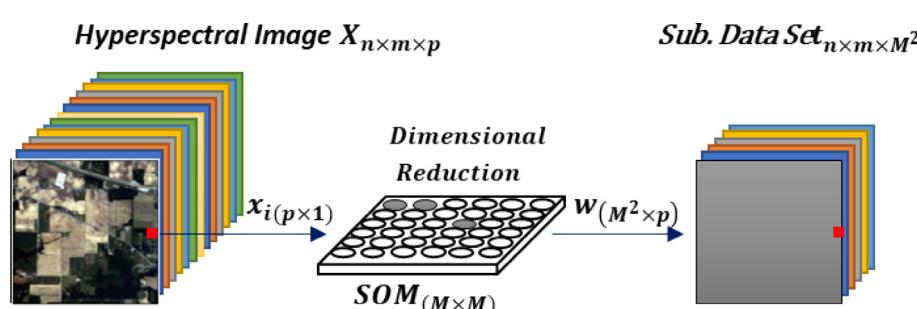


Fig. 4 – Dimensional Reduction Based on Self Organized Maps.

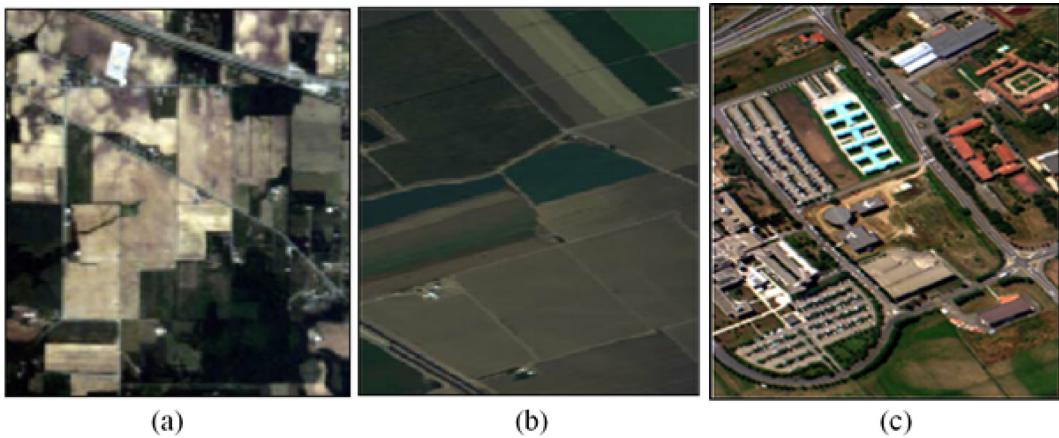


Fig. 6 – Aerial pictures of datasets selected: (a) Indian Pines (b) Salinas Valley (c) Pavia University.

information captured through field observations and pixel-by-pixel labeling.

The first data set was taken over the test place of Indian Pines in northwest Indiana. Fig. 7 shows the RGB image and

the ground truth with different kinds of crops in the Indian Pines data set. This data set has images of 145x145 pixels and 220 spectral reflectance bands in a wavelength range between 400 and 2500 nm (Fig. 8).

The second data set was taken over the test place of Salinas Valley in California. Fig. 9 shows the RGB image and the



Fig. 7 – Indian Pines Data Set. Right: RGB Image. Left: Ground Truth.

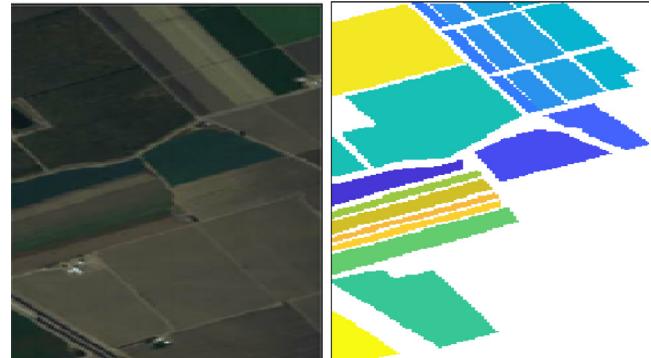


Fig. 9 – Salinas Valley Data Set. Right: RGB Image. Left: Ground Truth.

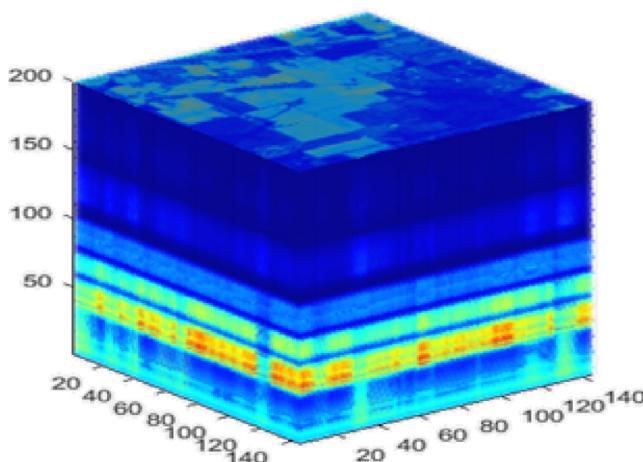


Fig. 8 – Stacked Images of 220 spectral reflectance bands for the Indian Pines Data Set.

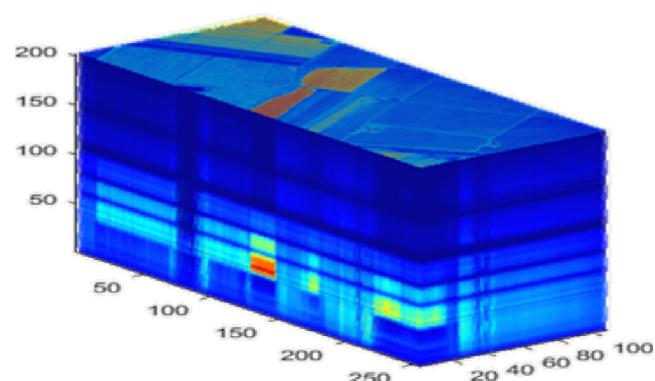


Fig. 10 – Stacked Images of 204 spectral reflectance bands for Salinas Valley Data Set.

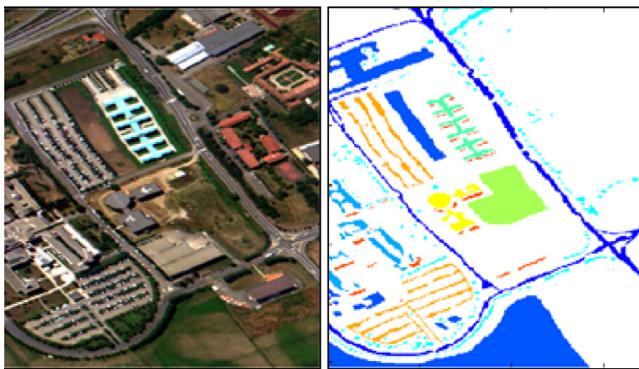


Fig. 11 – Pavia University Data Set. Right: RGB Image. Left: Ground Truth.

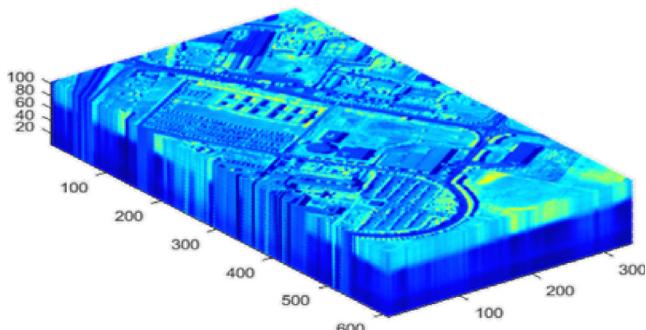


Fig. 12 – Stacked Images of 103 spectral reflectance bands for Pavia University Data Set.

ground truth with different crops in the Salinas Valley data. This data set has images of 512×217 pixels, 204 spectral reflectance bands in a wavelength range between 400 and 2500 nm and 3.7 m of spatial resolution by pixel (Fig. 10).

The third data set was taken over the test place of Pavia University, in northern Italy. Fig. 11 shows the RGB image and the ground truth with different kinds of elements, such as vegetation and buildings, in the Pavia University data set. This data set has images of 610×340 pixels and 103 spectral reflectance bands (Fig. 12).

The aim of these tests is to classify the different types of vegetation, crops or buildings in each hyperspectral data set. Initially, the dimensional reduction algorithm proposed in this work was applied to the data set. Then, the resulting spectral bands were used to perform a classification process through an RBF neural network. The process used a mean square error and an experimentally tuned variance of 0.05 and 0.5, respectively. The RBF classification was followed by a morphological operation of dilation and erosion to improve the accuracy of the results [35].

The performance of the dimensional reduction algorithm based on self-organized maps was compared with PCA and the wavelet decomposition method described in Sections 2.1 and 2.2. PCA and the wavelet decomposition method are some of the most common methods used in the dimensional reduction of data with high dimensionality.

3.1. Test Data: Indian Pines

Table 2 shows the types of vegetation and crops to be classified, as well as the total samples by class present in the data set. The RBF classifier used for the crop classification process uses 30 percent of the data set for training.

3.1.1. Self-organized map size

Considering the Indian Pines data set, it is important to define a proper self-organized map size. The size of the self-organized map defines the percentage of dimensional reduction of data. As shown in Section 2.3, the dimensional reduction method proposed in this paper makes a projection of an input hyperspectral image $X_{(N \times p)}$ where $N = n \times m$ is the total number of pixels and p is the number of spectral bands into a

Table 2 – Vegetation and Crop Types in Test Place 1.

Class	No. Samples
Alfalfa	46
Corn-no-till	1428
Corn-min-till	830
Corn	237
Grass-pasture	483
Grass-trees	730
Grass pasture mowed	28
Hay	478
Oats	20
Soybean-no-till	972
Soybean-min-till	2455
Soybean-clean	593
Wheat	205
Woods	1265
Buildings-Grass-Trees	386
Stone-Steel	93

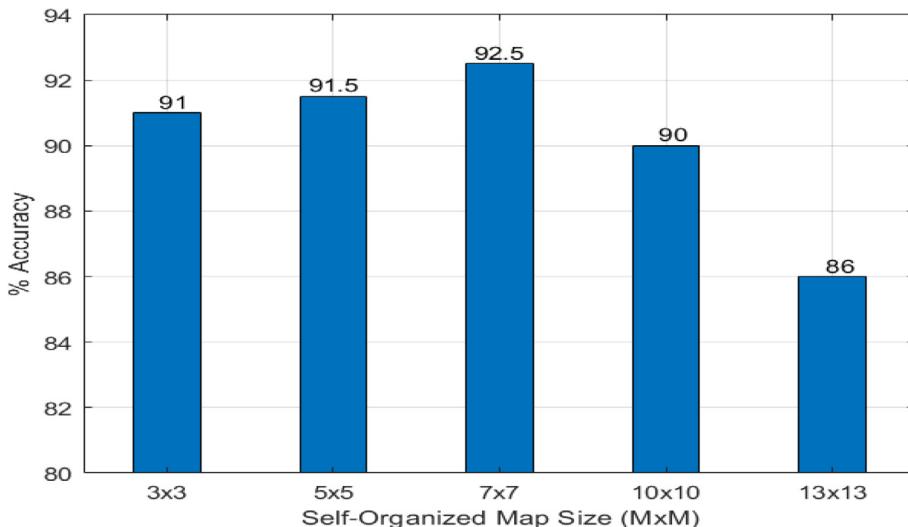


Fig. 13 – Success Percentage of Classification vs SOM Size.

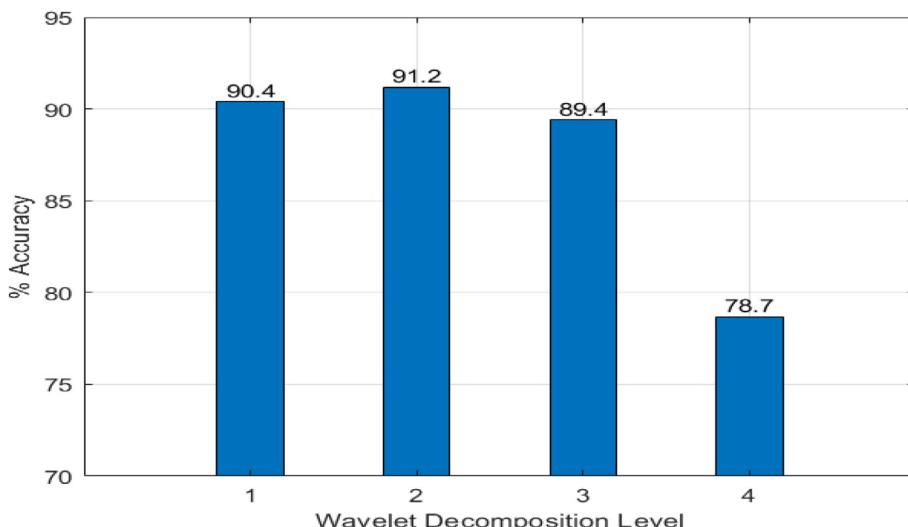


Fig. 14 – Success Percentage of Classification vs Decomposition Level.

smaller matrix $\hat{X}_{(N \times M^2)}$, where M is defined by the size of the self-organized map. Selecting an appropriate map size allows for better data reduction.

To do so, this test aims to define the relationship between the self-organized map size and the success percentage of the vegetation and crop classification process. For this reason, it is necessary to perform several tests with different dimensions of self-organized maps. The results are presented in Fig. 13.

Fig. 13 shows that the best classification result was 92.5%, which was obtained by a self-organized map with 7×7 elements. This implies a 77.7% of dimensional reduction regarding the original set of 220 spectral bands ($1 - M^2/p$).

3.1.2. Dimensional reduction for vegetation and crop classification using PCA, wavelet, and SOM

The goal of this test is to determine the performance of the algorithm proposed based on self-organized maps. To do so,

it is necessary to develop a comparison with the classification results obtained using different dimensional reduction methods and the RBF classifier. Then, the classifier is fed by the PCA, the wavelet method or the algorithm proposed in this work as dimensional reduction methods.

As mentioned before, the PCA variance proportion is used to select a suitable number of principal components. In this test, six principal components were taken that possess 98% of the total proportion of data variance.

In the same way as the self-organized map size test, the wavelet decomposition method needs to know the number of decomposition levels needed. This value may be different for each one of the datasets tested. Then, the number of decomposition levels is found experimentally for this data set using the RBF classifier performance. Fig. 14 shows the result of this test, where the best classification results are achieved with two decomposition levels.

Now, the PCA, wavelet and SOM algorithms are used in the vegetation and crop classification process depicted in Fig. 2,

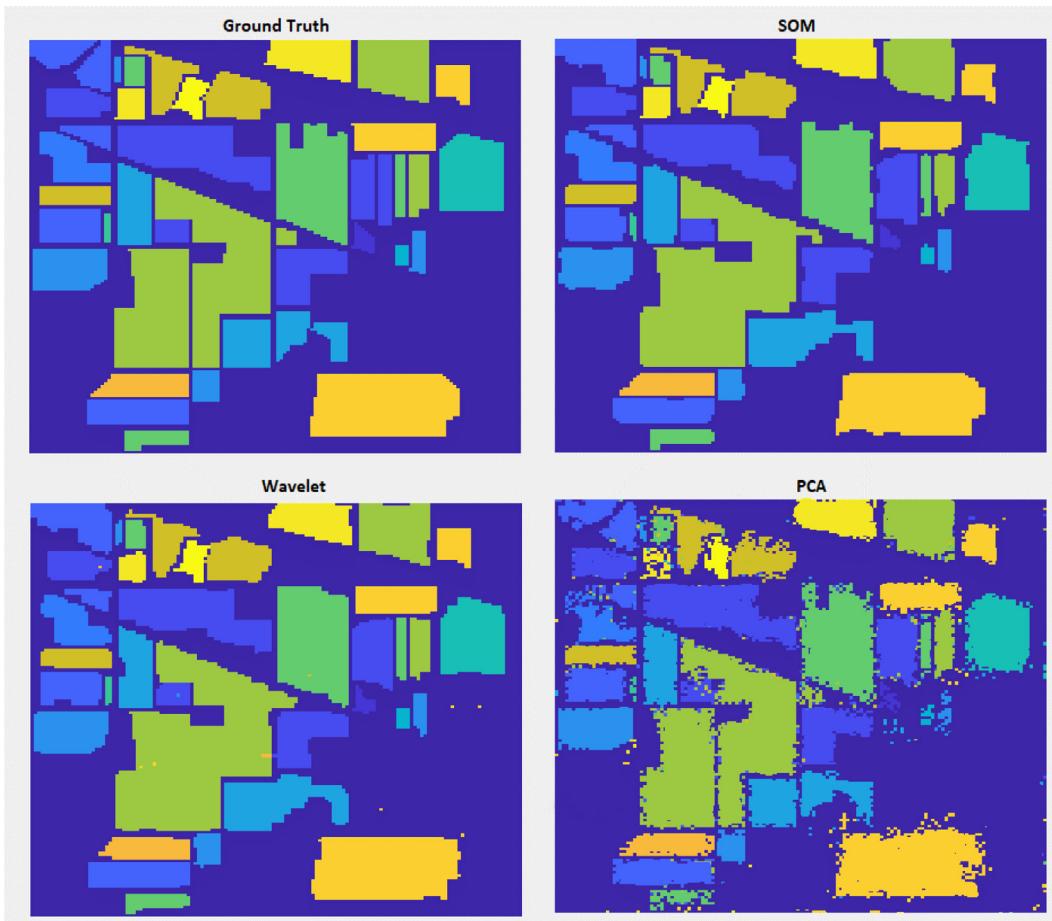


Fig. 15 – Ground Truth Indian Pines. Classification Results using SOM Dimensional Reduction, Wavelet and PCA.

Table 3 – Performance per Class for each Dimensional Reduction Method.

Class	Reduction Method		
	PCA	Wavelet	SOM
			% Percentage of Accuracy
Alfalfa	77.9	84.8	71.7
Corn-no-till	79.5	88.4	95.4
Corn-min-till	80.1	89.1	94.8
Corn	69.3	87.2	95.8
Grass-pasture	92.6	98.5	91.7
Grass-trees	87.2	88.7	95.1
Grass pasture mowed	75	86.4	89.3
Hay	89.7	93.8	94.4
Oats	67.9	85.5	85
Soybean-no-till	91.7	98.3	93.7
Soybean-min-till	91.2	95.4	96.7
Soybean-clean	73.8	89.5	94.4
Wheat	96.4	95.6	98.5
Woods	84.7	92.4	96.2
Buildings-Grass-Trees	63.4	89.7	95.6
Stone-Steel	67.8	94.9	91.4
Total Accuracy	80.5	91.2	92.5
% Reduction	97.2	75	77.5



Fig. 16 – Confusion Matrix using SOM Dimensionality Reduction for Indian Pines Scene.

and the resulting plots can be observed in Fig. 15. Here, the dimensional reduction of the self-organized map was performed using a 7×7 map size, the wavelet decomposition used 2 levels, and the number of PCA principal components was six.

Table 3 shows the performance per class obtained by each dimensional reduction method using an RBF classifier. The

results show that the dimensional reduction algorithm based on self-organized maps provides 77.5% data reduction and 92.5% classification accuracy. These are improvements considering the results obtained with PCA and wavelet decomposition. On the other hand, the largest mistakes in the classification process using the SOM reduction algorithm were generated in the classes with a reduced number of samples.

Table 4 – Vegetation and Crop’s Types in Test Place 2.

Class	No. Samples
Broccoli 1	2009
Broccoli 2	3726
Fallow	1976
Fallow rough plow	1394
Fallow smooth	2678
Stubble	3959
Celery	3579
Grapes	11,271
Soil vineyard develop	6203
Corn senesced	3278
Lettuce 4 weeks	1068
Lettuce 5 weeks	1927
Lettuce 6 weeks	916
Lettuce 7 weeks	1070
Vineyard untrained	7268
Vineyard vertical trellis	1807

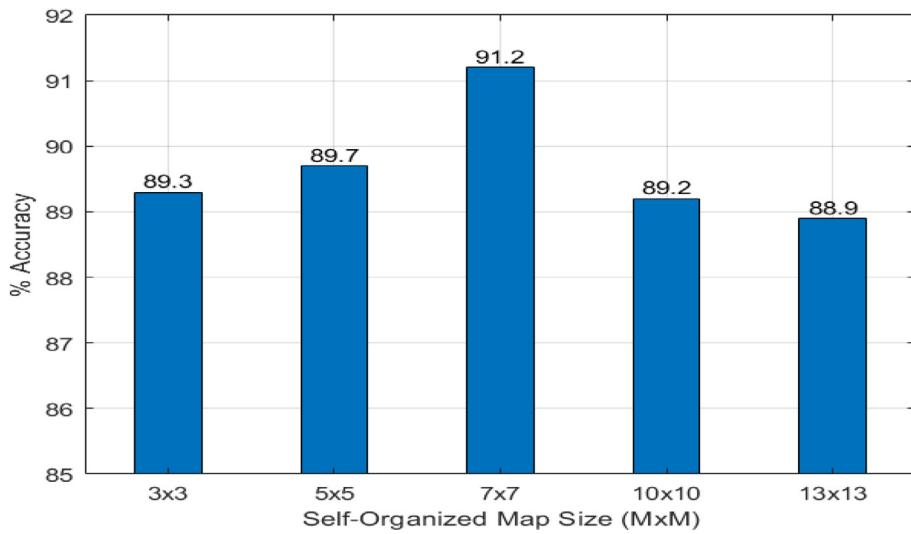


Fig. 17 – Success Percentage of Classification vs SOM Size.

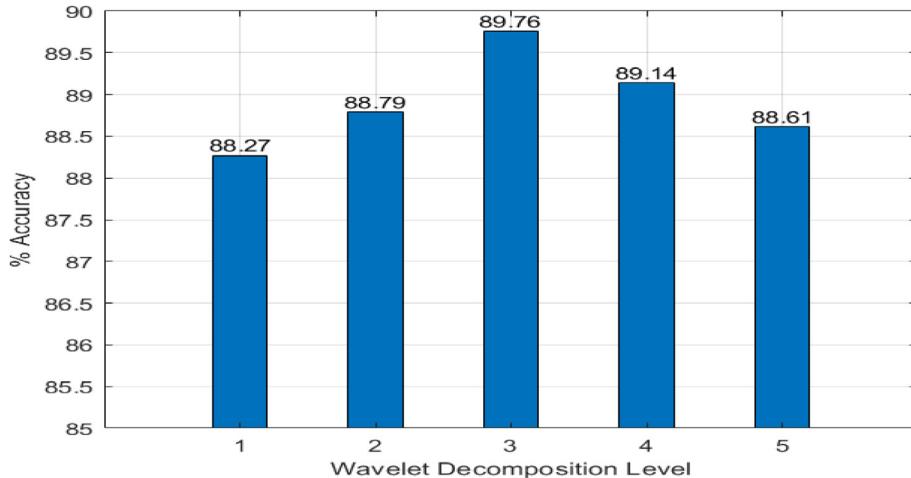


Fig. 18 – Success Percentage of Classification vs Decomposition Level.

Fig. 16 presents the confusion matrix of the classification system using the dimensional reduction algorithm based on self-organized maps. The first class corresponds to the background of the scene. The background may be soil or uninteresting elements such as native vegetation or roads. The next sixteen classes correspond to the classes presented in **Table 2**. The results of the confusion matrix show that most of the mistakes occur due to the background. This may happen due to the presence of unclassified elements that are related to interest classes.

3.2. Test data: Salinas Valley

Table 4 shows the vegetation and crop types to be classified using the Salinas Valley data set. The RBF classifier uses 30% of the total data for training. In the same way as before, the dimensional reduction algorithms were compared using this data set, and their results are described below.

3.2.1. Self-organized map size

Fig. 17 presents the classification performance with respect to the map size of the self-organized map. Observing **Fig. 17**, the best result of classification is 91.2% for this data set, and it is obtained by a self-organized map size of 7×7 elements. This implies a 75.9% of dimensional reduction regarding the original set of 204 spectral bands ($1 - M^2/p$).

3.2.2. Dimensional reduction for vegetation and crop classification using PCA, wavelet, and SOM

In PCA, to select a suitable number of principal components, the variance proportion data were used. In this experiment, five principal components were taken for this data set, which possess 98% of the total proportion of data variance.

For wavelet decomposition, it is necessary to test with different decomposition levels to determine the most suitable number of levels to be used on this dataset. This is done through an experimental process which uses different

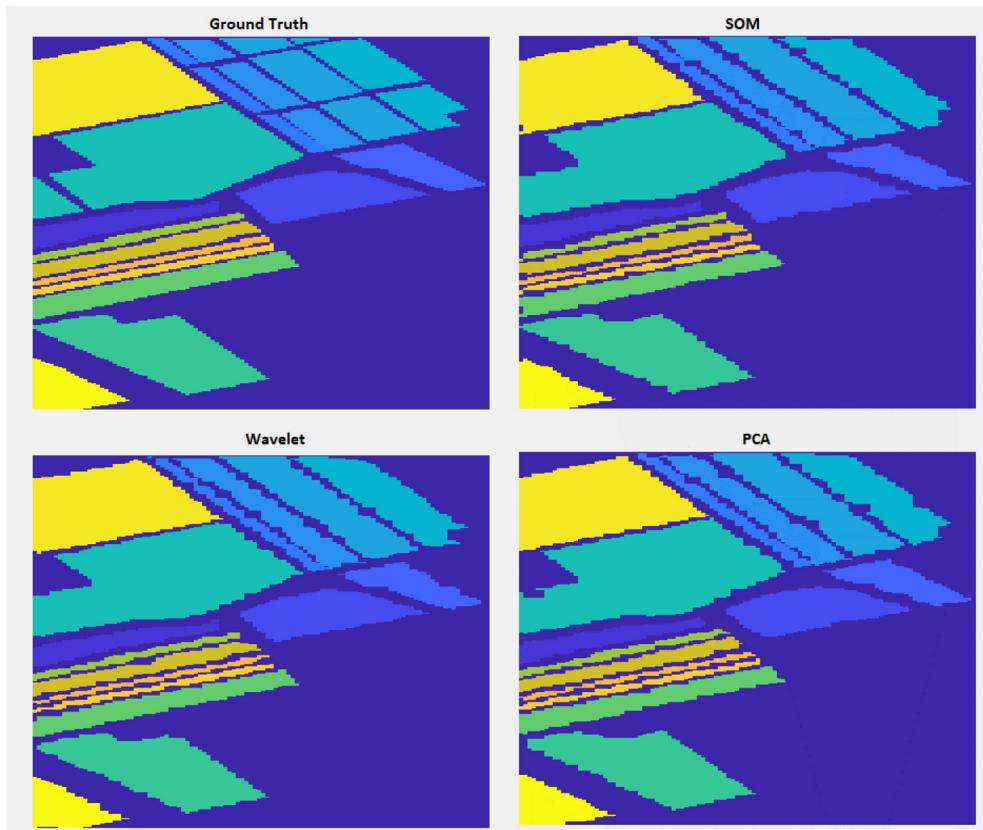


Fig. 19 – Ground Truth Salinas Valley. Classification Results using SOM Dimensional Reduction, Wavelet and PCA.

decomposition levels to determine the best classification score. Fig. 18 shows that the best classification results are achieved with three decomposition levels for this data set.

Once the SOM map size, the number of principal components, and the number of wavelet levels are defined, the orig-

inal data of the Salinas Valley data set are fed into these algorithms. Afterwards, their results are fed to the RBF classifier. Finally, Fig. 19 shows the ground truth and the crop classification plots. Here, the dimensional reduction of the self-organized map was performed using a 7×7 map size,

Table 5 – Performance per Class for each Dimensional Reduction Method.

Class	Reduction Method		
	PCA	Wavelet	SOM
			% Percentage of Accuracy
Broccoli 1	90.9	91.3	92.4
Broccoli 2	93.7	94.5	94.6
Fallow	91.2	89.9	91.5
Fallow rough plow	83.4	78.5	84.9
Fallow smooth	88.9	88.3	92.5
Stubble	91.4	92.5	94.2
Celery	88.9	92.2	93.4
Grapes	95.1	95.8	96.9
Soil vineyard develop	93.6	95.3	96.1
Corn senesced	90.4	90.3	91.6
Lettuce 4 weeks	82.5	83.6	85.0
Lettuce 5 weeks	89.0	91.0	90.5
Lettuce 6 weeks	75.39	74.4	81.1
Lettuce 7 weeks	78.2	78.2	82.5
Vineyard untrained	95.7	96.0	97.0
Vineyard vertical trellis	93.7	93.5	95.7
Total Accuracy	88.9	89.7	91.2
% Reduction	97.5	87.2	75.9

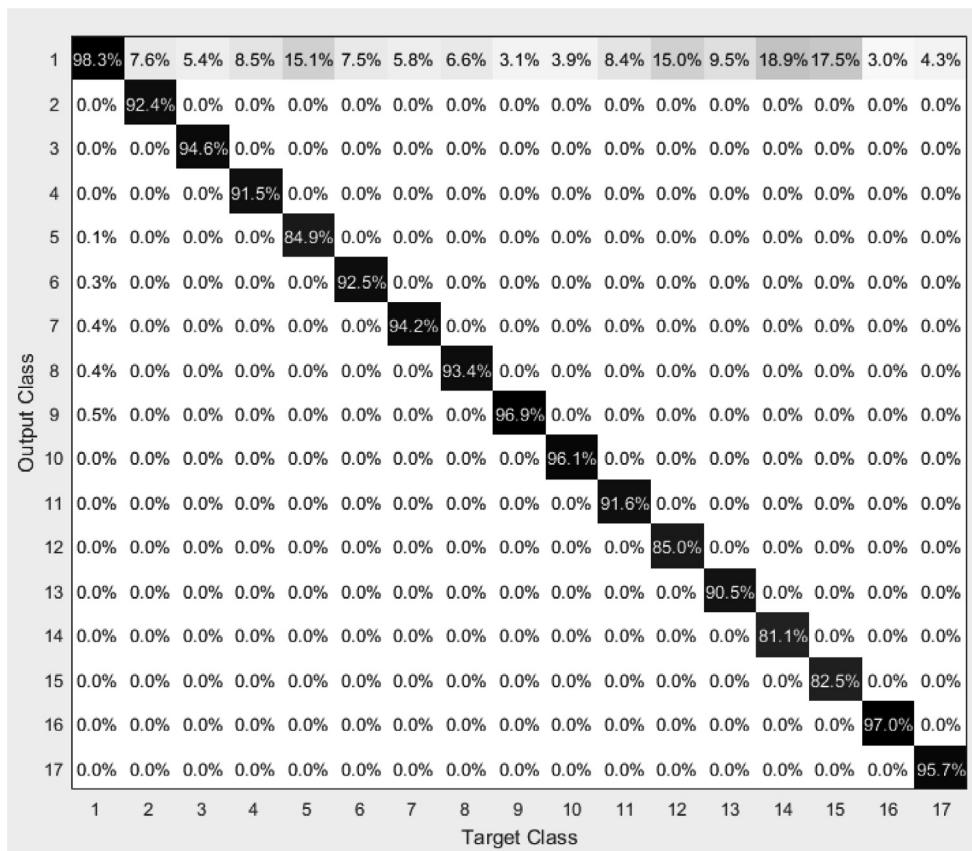


Fig. 20 – Confusion Matrix using SOM Dimensionality Reduction for Salinas Valley Scene.

the wavelet decomposition used 3 levels, and the number of PCA principal components was five.

Table 5 resumes the dimensional reduction and classification performance between PCA, wavelet decomposition and self-organized maps using the Salinas Valley data set. The average results show that the dimensional reduction algorithm based on self-organized maps provides a 75.9% data reduction and a 91.2% classification accuracy.

Fig. 20 presents the confusion matrix of the classification system using the dimensional reduction algorithm based on self-organized maps. The first class corresponds to the background of the scene. The next sixteen classes correspond to the class presented in **Table 4**. The background may be soil

or uninteresting elements, such as unclassified vegetation or soil. The results of the confusion matrix show that the similarity of some classes with the background generates most of the mistakes in the classification process. This may occur due to the presence of unclassified elements that are related to interest classes.

3.3. Test Data: Pavia University

Table 6 shows the different classes of elements to be classified and their total samples by class present in the data set. The RBF classifier used for the crop classification process uses 30 percent of the data set for training.

Table 6 – Vegetation and Crop Types in Test Place 3.

Pavia University	No. Samples
Class	
Asphalt	6631
Meadows	18,649
Trees	3064
Painted metal sheets	1345
Bare soil	5029
Self-Blocking Bricks	3682
Shadows	947

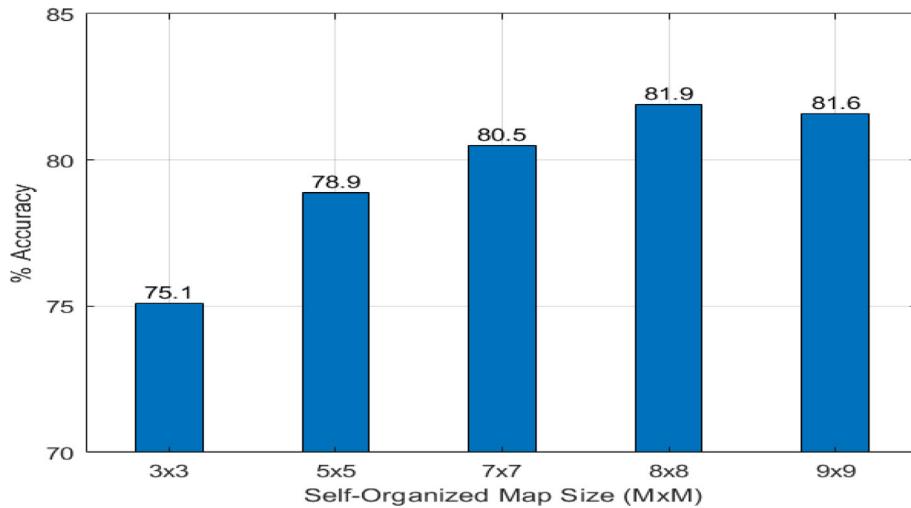


Fig. 21 – Success Percentage of Classification vs SOM Size.

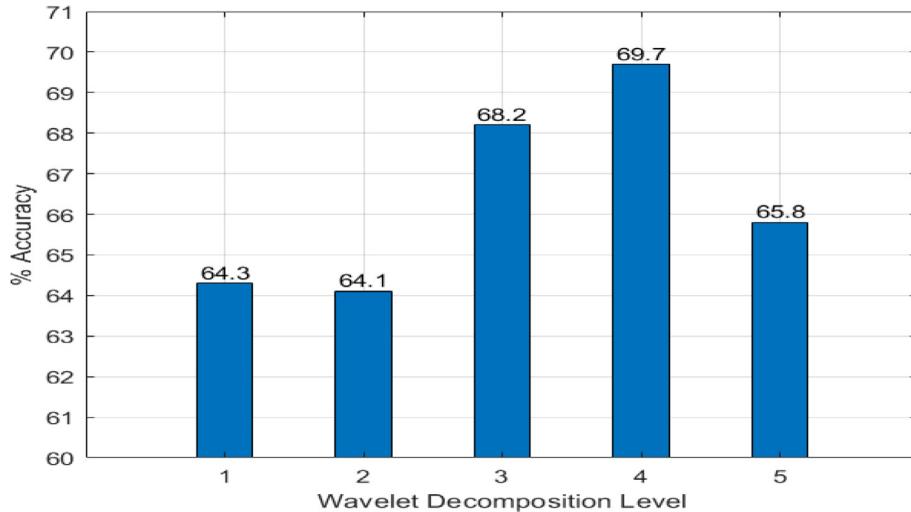


Fig. 22 – Success Percentage of Classification vs Decomposition Level.

3.3.1. Self-organized map size

Fig. 21 presents the classification performance with respect to the map size of the self-organized map. An 81.9% classification performance is obtained using a self-organized map size of 8×8 elements for this data set. This implies a 37.8% of dimensional reduction regarding the original set of 103 spectral bands

3.3.2. Dimensional reduction for vegetation and crop classification using PCA, wavelet, and SOM

To select a suitable number of principal components, the variance proportion data were used. In this experiment, four principal components were taken for this data set, which possess 98% of the total proportion of data variance.

For wavelet decomposition, an experimental test was performed to determine the most suitable number of decomposition levels for this dataset. Fig. 22 shows that the best classification results are achieved with four decomposition levels for this data set.

Once the dimensional reduction processes have been performed through the three algorithms presented, the reduced data can be fed to the RBF classifier. Fig. 23 shows the ground truth and the classification plots. Here, the dimensional reduction of the self-organized map was performed using an 8×8 map size, the wavelet decomposition used 4 levels, and four principal components were used.

Table 7 resumes the dimensional reduction and classification performance between PCA, wavelet decomposition and self-organized maps using the Pavia University data set. The results show that the dimensional reduction algorithm based on self-organized maps provides a 37.8% data reduction and an 81.9% classification accuracy.

Fig. 24 presents the confusion matrix of the classification system using the dimensional reduction algorithm based on self-organized maps. The first class corresponds to the background of the scene, which may be composed of unclassified elements. The next seven classes correspond to the class presented in Table 6. The results of the confusion matrix show

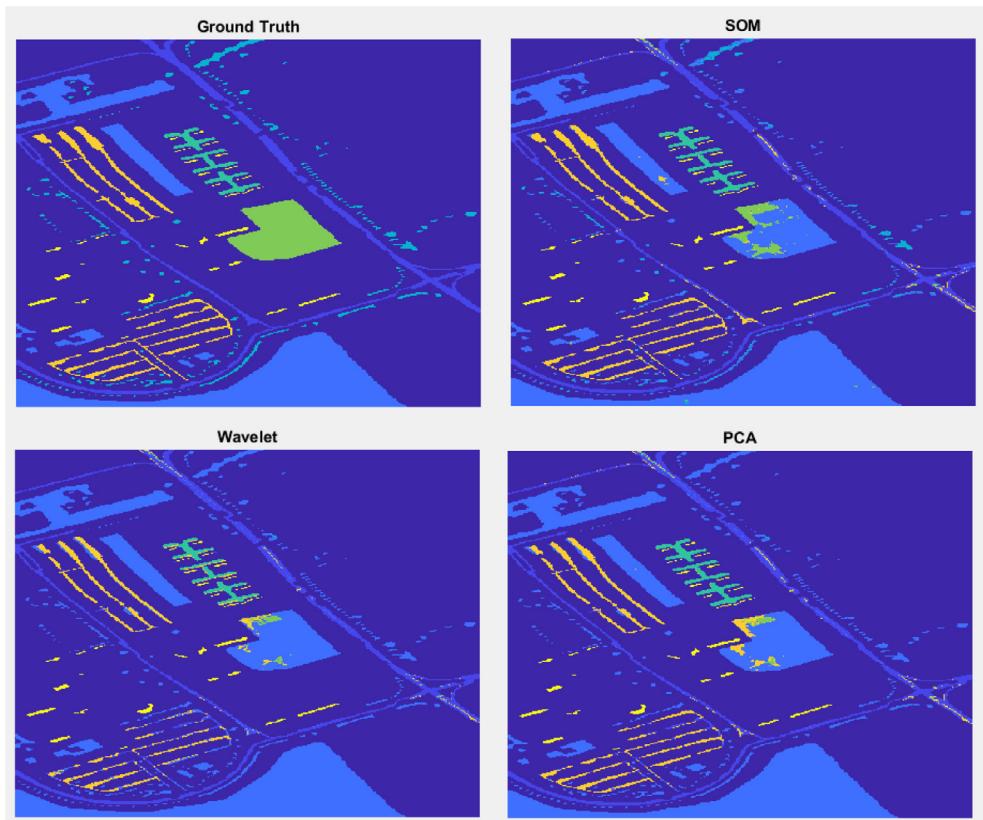


Fig. 23 – Ground Truth Pavia University. Classification Results using SOM Dimensional Reduction, Wavelet and PCA.

Table 7 – Performance per Class for each Dimensional Reduction Method.

Class	Reduction Method		
	PCA	Wavelet	SOM
% Percentage of Accuracy			
Asphalt	90.2	92.5	94.7
Meadows	99.9	100	99.6
Trees	5.33	9.4	46.8
Painted metal sheets	92.7	86.5	98.6
Bare soil	7.1	9.4	37.3
Self-Blocking Bricks	82.8	81.5	96.4
Shadows	96.9	98.1	99.9
Total Accuracy	67.8	68.2	81.9
% Reduction	96.1	93.2	37.8

that class four (painted metal sheets) and class six (self-blocking bricks) have high levels of confusion regarding class three (trees). It is possible that these classes have similar spectral signatures that may not be represented totally for the classification method.

4. Conclusions

One of the most important steps in hyperspectral data analysis is dimensional data reduction, which has the goal of identifying and removing redundant information that may obstruct subsequent classification processes. Dimensionality

reduction looks to reduce computational costs and the unnecessary use of resources, without sacrificing in a significant way the quality of information.

In this paper, an algorithm for the dimensional reduction of data based on self-organized maps is proposed. This algorithm uses a self-organized map to make a projection of the original data set with p spectral bands into a new dimensional space of $M \times M$ neurons. This process allows achieving a dimensional reduction proportion of $1 - M^2/p$.

To validate the results of the proposed dimensional reduction algorithm, the hyperspectral imagery data set taken by the AVIRIS sensor over the test sites Indian Pines and Salinas

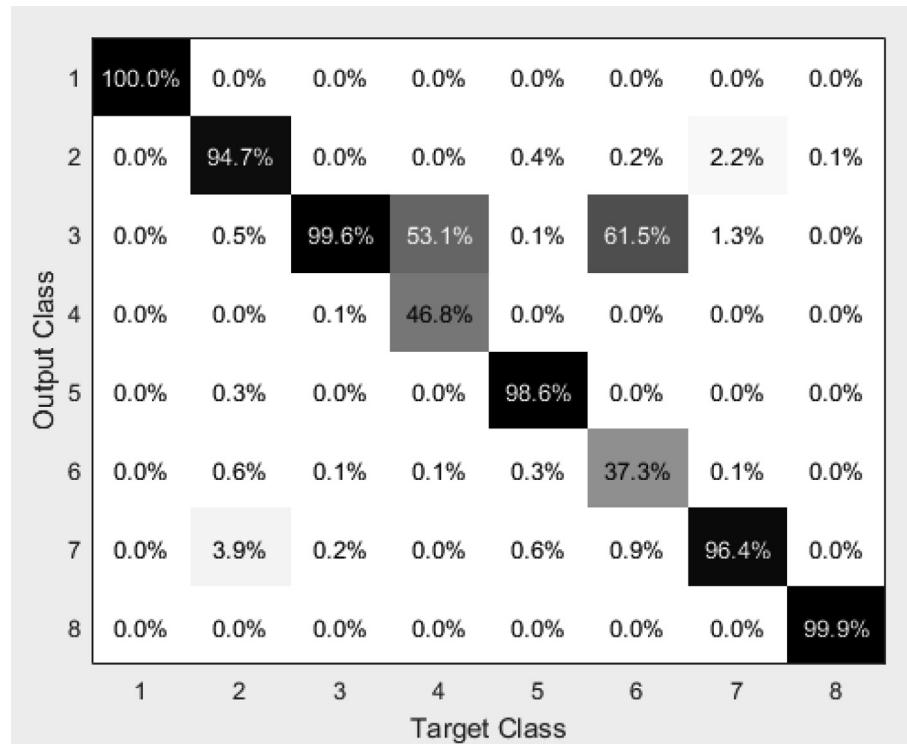


Fig. 24 – Confusion Matrix using SOM Dimensionality Reduction for Pavia University Scene.

Valley were used. The goal of the tests performed was to compare different dimensional reduction algorithms, such as PCA and wavelet decomposition. Then, these results were used to classify different vegetation types through an RBF (radial basis function) classifier.

Considering the three data sets used, the tests performed show that the proposed algorithm allows a 64% of dimensional reduction on average and an 88.5% classification performance. These results suggest that the dimensional reduction algorithm based on self-organized maps is an efficient approach compared with other popular algorithms, such as PCA or wavelet decomposition. This is due to the ability of self-organized maps to automatically detect (self-organizing) relationships within the set of input patterns. This property is applied to the problem of image mapping from higher dimensions to a two-dimensional feature space to perform image classification [28].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- [1] He L, Li J, Liu C, Li S. Recent Advances on Spectral Spatial Hyperspectral Image Classification: An Overview and New Guidelines. *IEEE Trans Geosci Remote Sens* 2018;56(3):1–19.
- [2] Shang X, Song M, Yu C. Hyperspectral Image Classification With Background. In: IGARSS - IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan. IEEE; 2019. p. 2714–7.
- [3] Loncan L, Almeida LB, Bioucas-Díaz JM, Briottet X, Chanussot J, Dobigeon N, et al. Hyperspectral pansharpening : a review. *IEEE Geosci Remote Sens Mag* 2015;3(9):27–46.
- [4] Ghamisi P, Yokoya N, Li J, Member S, Liao W, Member S, et al. Advances in Hyperspectral Image and Signal Processing: A comprehensive overview of the state of the art. *IEEE Geosci Remote Sens Mag* 2017;5(4):37–78.
- [5] Lira J. Tratamiento Digital de Imágenes Multiespectrales. 2nd ed. México D.F: Universidad Nacional Autónoma de México; 2010.
- [6] Thenkabail PS, Lyon JG, Huete A. Hyperspectral remote sensing of vegetation. 1st ed. New York: CRC Press Taylor & Francis Group; 2012.
- [7] Feng F, Li W, Du Q, Zhang B. Dimensionality Reduction of Hyperspectral Image with Graph-Based Discriminant Analysis Considering Spectral Similarity. *Remote Sens Environ* 2017;9(4):323–36.
- [8] Landgrebe D. Hyperspectral Image Data Analysis. *IEEE Signal Process Mag* 2002;19(1):17–28.
- [9] Yang X, Xu W, Liu H, Zhu L. Research on Dimensionality Reduction of Hyperspectral Image Under Close Range. In: IEEE International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China. p. 171–4.
- [10] Li W, Prasad S, Ye Z, Fowler JE, Cui M. Locality-preserving discriminant analysis for hyperspectral image classification using local spatial information. In: Geoscience and Remote Sensing Symposium (IGARSS). Munich, Germany: IEEE International; 2012. p. 4134–7.
- [11] Kuo BC, Landgrebe DA. Nonparametric weighted feature extraction for classification. *IEEE Trans Geosci Remote Sens* 2004;42(5):1096–105.

- [12] Serpico SB, Moser G. Extraction of spectral channels from hyperspectral images for classification purposes. *IEEE Trans Geosci Remote Sens* 2007;45(2):484–95.
- [13] Gao H, Zhou H, Xu L, Shi A. Classification of hyperspectral remote sensing images based on simulated annealing genetic algorithm and multiple instance learning. *J Central South Univ* 2014;21(1):262–71.
- [14] Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *J Mach Learn Res* 2007;8(5):1027–61.
- [15] Zhou Y, Peng J, Chen CLP. Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification. *IEEE Geosci Remote Sens* 2015;53(3):1082–95.
- [16] Cao X, Xiong T, Jiao L, Member S. Supervised Band Selection Using Local Spatial Information for Hyperspectral Image. *IEEE Geosci Remote Sens Lett* 2016;13(3):329–33.
- [17] Dong Y, Member S, Du B, Member S, Zhang L, Member S. Dimensionality Reduction and Classification of Hyperspectral Images Using Ensemble Discriminative Local Metric Learning. *IEEE Geosci Remote Sens* 2017;55(5):2509–24.
- [18] Ángel RL. Análisis de componentes principales, aplicaciones en teledetección. Teledetección Aplicada, Universidad Autónoma de México 2003;1:15.
- [19] Palsson F, Sveinsson JR, Ulfarsson MO, Benediktsson JA. Model-Based Fusion of Multi- and Hyperspectral Images Using PCA and Wavelets. *IEEE Trans Geosci Remote Sens* 2015;53(5):2652–63.
- [20] Hsu P-H, Tseng Y-H. Wavelet-Based Analysis of Hyperspectral Data for Detecting Spectral Features. *Int Arch Photogramm Remote Sens* 2000;33(7):61–8.
- [21] Karami A, Yazdi M, Mercier G. Compression of Hyperspectral Images Using Discrete Wavelet Transform and Tucker. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2012;5(2):444–50.
- [22] Theodoridis S, Koutroumbas K. Pattern Recognition. 2nd ed. Greece: Elsevier Academic Press; 2003.
- [23] Stephani H, Herrmann M, Bauer F, Heise B. Wavelet-based dimensionality reduction for hyperspectral THz imaging. *Int J Terahertz Sci Technol* 2010;3(3):117–29.
- [24] Agarwal A, LeMoigne J, Joiner J, El-Ghazawi T, Cantonnet F. Wavelet dimension reduction of AIRS infrared (IR) hyperspectral data. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS, Anchorage, USA. p. 3249–52.
- [25] Gormus ET, Canagarajah N, Achim A. Dimensionality reduction of hyperspectral images using empirical mode decompositions and wavelets. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2012;5(6):1821–30.
- [26] Zhang L, Xu F, Fraundorfer F. Deep Learning in Remote Sensing: A comprehensive review and list of resources. *IEEE Geosci Remote Sens Mag* 2017;5(4):8–36.
- [27] Chen Y, Jiang H, Li C, Jia X, Ghamisi P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans Geosci Remote Sens* 2016;54(10):6232–51.
- [28] Aggarwal C. Neural Networks and Deep Learning. 1st ed. Yorktown Heights, NY, USA: Springer, Cham; 2018.
- [29] Tso B, Mather PM. Classification methods for remotely sensed data. 2nd ed. New York: CRC Press Taylor & Francis Group; 2009.
- [30] Salazar F, Osorio C, Caicedo M, Alfonso W, Caicedo E. Identificación de la Delimitación Administrativa de la Malaria usando Redes Neuronales Artificiales. *Revista Orinoquia* 2017;21(1):11–9.
- [31] Gao RX, Yan R. Wavelets. 1st ed. Boston, MA: Springer US; 2011.
- [32] Proakis JG, Manolakis DG. Digital Signal Processing. 4th ed. Prentice Hall, USA: Pearson Education; 2007.
- [33] Alshareefi N. Image compression using wavelet transform. *J Babylon Univ Pure Appl Sci* 2013;21(4):1181–90.
- [34] Graña M. Hyperspectral Remote Sensing Scenes. Grupo de Inteligencia Computacional de La Universidad Del País Vasco; 2019. http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes [accessed November 27, 2019].
- [35] Fauvel M, Tarabalka Y, Benediktsson JA, Chanussot J, Tilton JC. Advances in spectral spatial classification of hyperspectral images. *Proc IEEE* 2013;101(3):652–75.