To appear in the *International Journal of Remote Sensing*
Vol. 00, No. 00, Month 20XX, 1–20

RESEARCH ARTICLE

ELM-based Spectral-Spatial Classification of Hyperspectral Images using Extended Morphological Profiles and Composite Feature Mappings

Francisco Argüello[a*] and Dora B. Heras[b]

[a]Dept. Electronics and Computing, University of Santiago de Compostela, Spain
[b]Centro de Investigación en Tecnoloxías da Información, University of Santiago de Compostela, Spain

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Extreme Learning Machine (ELM) is a supervised learning technique for a class of feed forward neural networks with random weights that has recently been used with success for the classification of hyperspectral images. In this work we show that the morphological techniques can be integrated in this kind of classifiers using several composite feature mappings which are proposed for ELM. In particular, we present a spectral-spatial ELM-based classifier for hyperspectral remote sensing images that integrates the information provided by extended morphological profiles. The proposed spectral-spatial classifier allows different weights for both (spatial and spectral) features outperforming other ELM-based classifiers in terms of accuracy for land cover applications. The accuracy classification results are also better than those obtained by equivalent spectral-spatial SVM-based classifiers [1].

1. Introduction

The development of image sensor technology has made it possible to capture hyperspectral images composed of hundreds of bands covering a wide range in the electromagnetic spectrum (Landgrebe 2002). The high dimensionality of hyperspectral data offers new opportunities for recognizing and classifying materials. Applications range from agriculture, mineralogy, medicine, quality control, material inspection, and surveillance, among others. Hyperspectral sensors usually collect two dimensional spatial images comprising hundreds of spectral bands, so that specific techniques are required to exploit all the information available in the hyperspectral data (Faubel et al. 2013).

Supervised classification of hyperspectral images is a challenging problem due to the high dimensionality of the data and the limited availability of training samples. The most basic method for classifying the pixels of a hyperspectral image is based on the use of only the spectrum of each pixel. In particular, neural networks (Heras et al. 2012) and Support Vector Machines (SVM) (Melgani and Bruzzone 2004) are widely used owing to their performance. SVMs use a learning process

*Corresponding author. Email: francisco.arguello@usc.es

[1]This is an Accepted Manuscript of an article published by Taylor Francis in International Journal of Remote Sensing on 23 Jan 2015, available online: http://wwww.tandfonline.com/10.1080/01431161.2014.999882

which is known as structural risk minimization. The training algorithm searches for a hyperplane that separates the dataset into different classes. SVMs are intrinsically binary classifiers, but different strategies have been developed to allow them to simultaneously discriminate between multiple classes (Mountrakis, Im, and Ogole 2011).

The accuracy of the results provided by a pixelwise classification can be improved by integrating spatial and spectral information in the classifier (Plaza et al. 2009; Faubel et al. 2013; Mura et al. 2011; Palmason et al. 2005). This means that the decision to assign a pixel to a specific class is based both on the spectral feature, which is the pixel value, and on certain information derived from the pixel's neighborhood. In the bibliography various methods have been proposed for the spectral-spatial classification of remote sensing images, usually using neural networks or SVMs for processing the spectral information and different techniques for integrating the spatial information. These include the watershed transform (Zhang, Feng, and Le 2008; Tarabalka, Chanussot, and Benediktsson 2010a), partitional clustering techniques (Tarabalka, Benediktsson, and Chanussot 2009), Markov random fields (Tarabalka et al. 2010), minimum spanning forests (Tarabalka, Chanussot, and Benediktsson 2010b), hierarchical segmentation (Plaza et al. 2009), kernel methods (Marconcini, Camps–Valls, and Bruzzone 2009; Faubel, Chanussot, and Benediktsson 2012), and decision fusion (Benediktsson and I. Kanellopoulos 1999), among others.

The spatial information can be extracted from hyperspectral data using the tools of mathematical morphology. Two widely used morphological operators are opening and closing, which are based on the fundamental operations of erosion and dilation. From these basic operations the so-called morphological profile (MP) can be constructed (Pesaresi and Benediktsson 2001; Soille and Pesaresi 2002; Palmason et al. 2005). An MP contains information of the structures of the image at different resolution sizes. In remote sensing, MPs are usually computed from hyperspectral data using Principal Component Analysis (PCA) (Benediktsson, Pesaresi, and Arnason 2003; Faubel et al. 2013; Marpu et al. 2012). If several components are retained, the MPs can be used all together in one Extended Morphological Profile (EMP) (Benediktsson, Palmason, and Sveinsson. 2005; Licciardi et al. 2011). This can be further generalized in order to model the spatial information more accurately. For example, a morphological attribute profile (MAP) is created in the same way as the MP but considering attribute operators (Mura et al. 2010). In EMAP (Mura et al. 2011) an extended morphological attribute profile is created using morphological attribute filters. These attributes extract information related to size, geometry, and homogeneity of regions. On the other hand, the use of multi-channel morphological profiles for feature extraction is explored in (Plaza, Plaza, and Barra 2009).

In spectral-spatial classification methods, once the spatial information has been extracted, it must be taken into account in the classification. In the approaches based on neural-networks or SVM, spatial information can be integrated into a pixelwise classifier as a preprocessing stage of the hyperspectral data, or as a postprocessing stage on the output of the classifier. In particular, majority voting (Tarabalka, Chanussot, and Benediktsson 2010a,b) and regularization (Tarabalka et al. 2010) have been widely exploited as postprocessing stages of pixelwise SVM-based classifiers. For example, in (Faubel et al. 2013) some techniques based on SVM that use marker selection followed by MSF (minimum spanning forest) segmentation and majority voting are presented. Other approaches presented in the work are based on marker selection using the MSSC approach (multiple classifications based on SVM are performed in parallel applying majority voting) followed by MSF. Alternatively, kernel methods (Marconcini, Camps–Valls, and Bruzzone 2009; Faubel, Chanussot, and Benediktsson 2012; Camps–Valls and Bruzzone 2005) can jointly use the spatial and spectral information in the classification. Finally, decision fusion methods can combine information from several individual classifiers or pre-processing steps according to a certain criteria (Benediktsson and I. Kanellopoulos 1999).

In the field of neural networks, *Extreme Learning Machine* (ELM) is a term used to describe a class of Single-hidden Layer Feedforward Neural Networks (SLFN) with random weights (Wang and Wan 2008; Huang 2014). The idea of randomness was introduced in an early paper by Broomhead and Lowe (Broomhead and Lowe 1988), who showed that hidden radial basis function (RBF) neurons randomly selected are sufficient to allow functional interpolation, and later extended by Lowe (Lowe 1989). Later, Schmidt et al. (Schmidt, Kraaijveld and Duin 1992) showed that an SLFN with sigmoid activation function and fixed random values in the weights of the hidden layer can be trained providing good results. The learning and generalization characteristics of the Random Vector Funcional-Link (RVFL) net were explored by Pao et al. (Pao, Park, and Sobajic 1994), while a theoretical justification of RVFL as universal approximator was developed by Igelnit and Pao (Igelnik and Pao 1995). Finally, Huang et al. (Huang, Zhu, and Siew 2004, 2006) stated the universal aproximation of SLFNs for a wide type of activation functions with random weights and introduced the term ELM.

The use of ELM for hyperspectral images has only recently been explored. Two datasets, one multispectral and another hyperspectral, were used for classification in (Pal 2009). Besides, (Pal. Maxwell, Warner 2013) compares a pixelwise SVM classifier to a kernel-based ELM one using a RBF function. Results from the aforesaid work suggest that ELM results are slightly better in terms of accuracy with significantly lower execution times. ELM has also been used for temporal-change detection in urban environments (Chang 2011). Two ELM-based techniques incorporating spatial information of the image were proposed in (Heras, Argüello, and Quesada–Barriuso 2014). The first scheme uses a majority vote approach in order to combine the results of a pixelwise ELM classifier with the segmentation map obtained by a watershed algorithm. The second one uses the information of the neighboring pixels in order to modify the ELM classifier output. The results are compared to those obtained by similar SVM-based techniques and show improved classification accuracies and much lower execution times. More recently (Samat et al. 2014) proposes two new algorithms of ensemble extreme learning machines (bagging-based and AdaBoost-based ELMs) for the classification task, considering EMPs and principal components and providing improvements over previous works.

In this paper, we show that morphological techniques can be successfully applied to ELM classifiers. In particular, we present a spectral-spatial ELM-based classifier for hyperspectral images that integrates the spatial information provided by EMPs. Both features, spectral and spatial, are integrated in the ELM classifier using composite feature mappings. This scheme allows different weights for both features and provides significant improvements in the accuracy of the classification when the results are compared to the state of the art. The rest of this paper is organized as follows: Section 2 describes the ELM technique and its application to remote sensing; in Section 3 we present the spectral-spatial ELM-based classifier; the evaluation of the classifier is performed in Section 4; and finally in Section 5

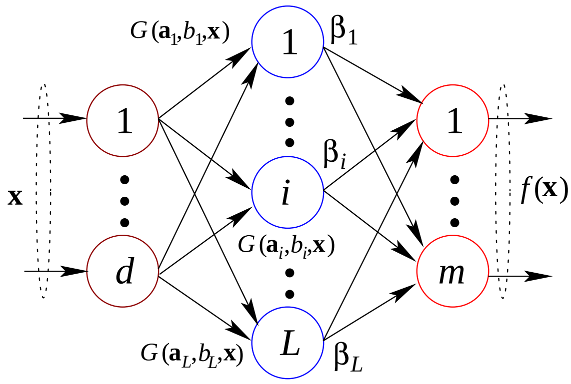


Figure 1. A single-hidden layer feedforward network such as that used by ELM.

we present the conclusions.

2. Extreme Learning Machines

Neural networks make it possible to change the network structure or its parameters by incorporating certain learning algorithms based on internal o external information. ELM was originally developed as a training technique for a class of SLFNs with random weights (Huang, Wang, and Lan 2011; Huang, Chen, and Siew 2006). The structure of an SLFN is shown in Fig. 1. The output function of an SLFN with L hidden nodes and input $\boldsymbol{x}$ can be written as

$$f(\boldsymbol{x}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i G(\boldsymbol{a}_i, b_i, \boldsymbol{x}), \quad \boldsymbol{x} \in \mathrm{R}^d, \;\; \boldsymbol{\beta}_i \in \mathrm{R}^m, \tag{1}$$

where $G(\boldsymbol{a}_i, b_i, \boldsymbol{x})$ denotes the output function of the ith hidden node and $\boldsymbol{\beta}_i$, $\boldsymbol{a}_i$, and b_i are the weights and biases which must be generated in the training phase. For the case of additive nodes with activation function g, it can be expressed as

$$G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) = g(\boldsymbol{a}_i \cdot \boldsymbol{x} + b_i), \quad \boldsymbol{a}_i \in \mathrm{R}^d, \;\; b_i \in \mathrm{R}, \tag{2}$$

while for RBF nodes,

$$G(\boldsymbol{a}_i, b_i, \boldsymbol{x}) = g(b_i ||\boldsymbol{x} - \boldsymbol{a}_i||), \quad \boldsymbol{a}_i \in \mathrm{R}^d, \;\; b_i \in \mathrm{R}^+. \tag{3}$$

An SLFN with L hidden nodes can approximate N arbitrary distinct samples and targets $(\boldsymbol{x}_i, \boldsymbol{t}_i) \in \mathrm{R}^d \times \mathrm{R}^m$, if the following equation system can be solved:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{4}$$

where

$$\mathbf{H} = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{h}(\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} G(\boldsymbol{a}_1, b_1, \boldsymbol{x}_1) & \dots & G(\boldsymbol{a}_L, b_L, \boldsymbol{x}_1) \\ \vdots & \dots & \vdots \\ G(\boldsymbol{a}_1, b_1, \boldsymbol{x}_N) & \dots & G(\boldsymbol{a}_L, b_L, \boldsymbol{x}_N) \end{bmatrix}_{N \times L}, \tag{5}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m}, \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}_{N \times m}. \tag{6}$$

$\mathbf{H}$ is called the hidden layer output matrix of the neural network. An SLFN with randomly generated additive or RBF nodes in the hidden layer can universally approximate any continuous target functions over any compact subset $\chi \subset \mathrm{R}^d$ (Igelnik and Pao 1995; Huang, Chen, and Siew 2006; Huang, Wang, and Lan 2011). For the case of additive nodes, the activation function g can be any infinitely differentiable function, including sigmoidal functions, and also the radial basis, sine, cosine and exponential functions among others. Once they are randomly generated, the hidden node parameters $(\boldsymbol{a}_i, b_i)$ remain fixed and training an SLFN is equivalent to finding a least-squares solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$, i.e.,

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T}, \tag{7}$$

where $\mathbf{H}^{\dagger}$ is the *Moore-Penrose generalized inverse* of matrix $\mathbf{H}$ (Courrieu 2005). Thus, ELM can be summarized as follows (Huang, Wang, and Lan 2011; Huang, Chen, and Siew 2006):

Algorithm ELM: Given a training set $\{(\boldsymbol{x}_i, \boldsymbol{t}_i) | \boldsymbol{x}_i \in \mathrm{R}^d, \boldsymbol{t}_i \in \mathrm{R}^m, i = 1, \dots, N\}$, hidden node output function $G(\boldsymbol{a}_i, b_i, \boldsymbol{x})$, and hidden node number L,

(1) Randomly generate hidden node parameters $(\boldsymbol{a}_i, b_i)$, $i = 1, \dots, L$.

(2) Calculate the hidden layer output matrix $\mathbf{H}$.

(3) Calculate the output weight vector, $\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{T}$.

Usually, a sigmoid activation function is used, although sine function, hard limit, triangular basis function, etc., are also possible. ELM has the advantage over SVM of being inherently a multiclass classifier, so that it does not require the additional stage of voting the results of multiple binary classifiers. The classification accuracy of ELM is, as in any other supervised learning method, dependent on the size and quality of the training set. Additionally, the design of the ELM classifier requires the setting of a single user-defined parameter: the number of nodes in the hidden layer. On the other hand, ELM methods can be extended to include kernel learning (Huang 2012), incremental learning (Huang and Chen 2008), effective ELM (Wang, Cao, and Yuan 2011; Cao, Liu, and Park 2013), modified ELM with sigmoidal activation functions (Chen, Zhu, and Wang 2013), and voting (Cao et al. 2012).

ELM presents important differences from other methods of classification of hyperspectral images, such as SVM. In general, supervised classification and regres-

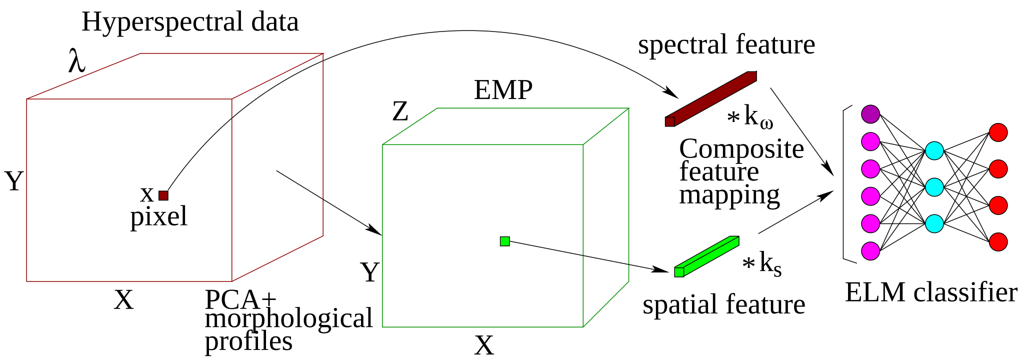


Figure 2. The proposed ELM-EMP spectral-spatial classifier.

sion methods can be categorized as "feature-based" or "geometry-based" methods (Baldi, Azencott, and Swamidass 2011). Feature-based methods, as ELM and neural networks, tend to derive a prediction from numerical features extracted from the input data. On the other hand, geometry-based methods, as exemplified by kernel methods, derive a prediction by looking at the similarities between the current input instance and all the other instances available in the training set, hence looking at the geometry of the neighborhood of the input instance.

In SVM, the training data are usually first mapped into a higher dimensional feature space through a nonlinear feature mapping function. A standard optimization method is then used to find the solution by maximizing the separating margin of two different classes. Suppose $\phi(\boldsymbol{x})$ is the mapping from the input vector $\boldsymbol{x}$ to a higher dimensional feature space. The optimization stage relies on inner products $\phi(\boldsymbol{x}_1)\phi(\boldsymbol{x}_2)$, and the "kernel trick" can be used to compute $\phi(\boldsymbol{x}_1)\phi(\boldsymbol{x}_2)$ using the input vectors directly, i.e., $\phi(\boldsymbol{x}_1)\phi(\boldsymbol{x}_2) = k(\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is referred to as a kernel. As the training of SVMs involves a quadratic programming problem, the computational complexity of SVM training algorithms is at least quadratic with respect to the number of training samples.

Different from SVM, in ELM a feature mapping (in this case, a hidden-layer output vector) is usually known to users. Almost all nonlinear piecewise continuous functions can be used as the hidden-node output functions, and thus, very different feature mappings can be used in ELM. For example, we can have a nonlinear piecewise continuous function $G(\boldsymbol{a}_i, b_i, \boldsymbol{x})$ with $\{\boldsymbol{a}_i, b_i\}_{i=1}^{L}$ randomly generated according to any continuous probability distribution. However, if the feature mapping is unknown to users, ELM can be generalized to use kernels (Huang 2012).

3. Spectral-spatial ELM-based classification

In this section we present an ELM-based classifier (ELM-EMP) that can jointly use the spectral and spatial information. Specifically, the input to the classifier has two components: the spectral feature, which is the value of each pixel, and a spatial feature, which is obtained from the hyperspectral data using morphological techniques. The process is outlined in Fig. 2.

Two widely used morphological operators are opening and closing, which are based on the fundamental operations of erosion and dilation (Crespo, Serra, and Schafer 1995; Soille 2005). The precise details of the effect of the operator on the image are determined by a structuring element. Moreover, it is possible to define morphological operators that satisfy the following property: if a spatial structure of the image contain the structuring element then it is fully preserved; otherwise it is fully removed. These operators are called opening and closing by reconstruction (Faubel et al. 2013). The MP is constructed based on the repeated use of openings and closings with a structuring element of increasing size (Pesaresi and Benediktsson 2001; Soille and Pesaresi 2002; Palmason et al. 2005). In this way, information about the contrast and the size of the structures present in the image can be extracted. The MP of order n from the image I can be expressed as,

$$\mathrm{MP}^{(n)}(I) = [\phi_r^{(n)}(I), \ldots, \phi_r^{(1)}(I), I, \gamma_r^{(1)}(I), \ldots, \gamma_r^{(n)}(I)], \tag{8}$$

being $\phi_r^{(i)}$ and $\gamma_r^{(i)}$ the closing and opening by reconstruction operators, respectively, with a structuring element i from 1 to n, whose size increases normally in steps of 1 or 2. For a single panchromatic image, the MP of order n results in $2n + 1$ components.

When the MP approach is applied to hyperspectral data, which can comprise hundreds of bands, the most significant principal components are usually used as base images. PCA is an linear transformation defined in such a way that the principal components are arranged according to their variance (variability in the data). The result is an EMP, i.e., a profile based on more than one original image (Benediktsson, Palmason, and Sveinsson. 2005). The EMP computed from the m principal components of the hyperspectral image using MPs of order n is written as,

$$\mathrm{EMP}_m^{(n)}(I) = [\mathrm{MP}_1^{(n)}(I), \ldots, \mathrm{MP}_m^{(n)}(I)], \tag{9}$$

and contains $m(2n + 1)$ components.

Once the spatial information has been extracted from the image, it must be integrated into the classifier. When a SVM-based classifier is used, this can be done by using the so-called kernel methods (Faubel, Chanussot, and Benediktsson 2012; Camps–Valls and Bruzzone 2005). SVM uses kernels to map the data with a nonlinear transformation to a higher dimensional space, in order to find a linear separating surface between the two classes. Several types of kernels, such as linear, polynomial, splines or RBF kernels can be used in SVM. In this case, the spectral and spatial features lead to two different kernel matrices which can be easily processed using a suitable kernel function (Scholkopf and Smola 2002; Faubel, Chanussot, and Benediktsson 2006; Mercier and Lennon 2003). In this approach the composite kernel can be built through four different approaches: concatenation, direct summation, weighted summation, and cross-information (Camps–Valls et al. 2006).

ELM, as commented in section 2, does not require such nonlinear transformation and therefore does not require kernels. Instead, ELM usually uses a feature mapping $G(\boldsymbol{a}_i, b_i, \boldsymbol{x})$ with $\{\boldsymbol{a}_i, b_i\}_{i=1}^{L}$ randomly generated according to any continuous probability distribution (if feature mappings are unknown to users, kernels can be applied in ELM as well) (Huang 2012). Almost all nonlinear piecewise continuous functions can be used as the hidden-node output functions, for example, sigmoid, hard-limit, Gaussian, and multiquadratic, among others. In particular, in what follows we will consider the case of a sigmoid function,

$$G(\boldsymbol{a}, b, \boldsymbol{x}) = \frac{1}{1 + \exp(-(\boldsymbol{a} \cdot \boldsymbol{x} + b))}. \tag{10}$$

In the following, we present several approaches for the joint consideration of spectral and spatial information in an ELM-based classifier. Let $\boldsymbol{x} \in \mathrm{R}^n$ be a pixel-vector of the hyperspectral image, which by means of certain transformations can be redefined simultaneously both in the spectral domain, $\boldsymbol{x}^\omega \in \mathrm{R}^{n_\omega}$, and in the spatial domain, $\boldsymbol{x}^s \in \mathrm{R}^{n_s}$. The possibilities to construct a composite feature mapping in ELM are the following:

(1) Weighted averaging. In this approach the spectral and spatial features are just added before being included in the feature mapping,

$$G(\boldsymbol{a}, b, \boldsymbol{x}) = G(\boldsymbol{a}, b, k_\omega \boldsymbol{x}^\omega + k_s \boldsymbol{x}^s), \tag{11}$$

where k_ω and k_s are the weights assigned to each of these two features. In this case both vectors should have the same size. This approach could be used, for example, when the pixel-vector is used as spectral feature and the spatial feature is computed as the vector median value of the neighbors of the original pixel.

(2) Weighted summation. In this approximation we have two feature mappings, one for the spectral feature and the other for the spatial one,

$$G(\boldsymbol{a}, b, \boldsymbol{x}) = k_\omega G(\boldsymbol{a}, b, \boldsymbol{x}^\omega) + k_s G(\boldsymbol{a}, b, \boldsymbol{x}^s). \tag{12}$$

(3) Weighted concatenation. The spectral and spatial features are concatenated into one stacked vector,

$$G(\boldsymbol{a}, b, \boldsymbol{x}) = G(\boldsymbol{a}, b, [k_\omega \boldsymbol{x}^\omega, k_s \boldsymbol{x}^s]^T). \tag{13}$$

The weights k_ω and k_s allow to adjust the influence of each feature in the classification. Since an ELM classifier discriminates classes according to the magnitude of the vector components, a smaller weight in one of the components implies it will have less impact in the classification. For example, consider a sigmoid type feature mapping and the concatenation approach. In Equation (10) we have the following operation of multiplication and accumulation (MAC),

$$\mathrm{MAC} = \boldsymbol{a} \cdot \boldsymbol{x} + b = \sum_{j=1}^{d_\omega} a_j k_\omega x_j^\omega + \sum_{j=1}^{d_s} a'_j k_s x_j^s + b. \tag{14}$$

Since coefficients of the hidden-layer output vector, $(\boldsymbol{a}, b)$, are randomly assigned once during the training phase, the terms $a_j k_\alpha x_j^\alpha$ with greater magnitude in Equation (14) contribute more to MAC and therefore to the feature mapping. In ELM the training phase is not an iterative process, and therefore there is no possibility of changes in the coefficients $(\boldsymbol{a}, b)$.

In order to speed up the classification, the spatial feature can be computed in a pre-processing step on the hyperspectral data, just before the application of the ELM classifier, as it is shown in Fig. 2. The same weights should be applied to both training and testing stages.

4. Evaluation

The ELM classifier was programmed in C using the OpenBLAS (OpenBLAS 2011) implementation of the LAPACK linear algebra library. This implementation takes advantage of multicore CPUs by running multiple threads. Additionally, the computation of the activation functions was parallelized using OpenMP. The SLFN was configured with additive nodes and sigmoid activation function (10) with $a_i \in [-1, 1]$ and $b_i \in [0, 1]$, and the hyperspectral data was scaled in the range [0,1]. The *drand48* generator from the standard C library was used for the selection of the random parameters, while the Moore-Penrose generalized inverse was computed using the method described in (Courrieu 2005). The number of training samples was selected to be similar to that of other works on which the comparison was performed, while the number of nodes in the hidden layer was selected to obtain the best results. As the results of the classification depend on the training set and on the random values generated by the ELM classifier, each measure shown is the average of 100 runs varying these parameters. For each dataset the samples are randomly distributed between the training and testing sets. During the testing stage all the pixels of the image are classified, but the samples used in the training stage are excluded for calculating the accuracy results.

In the case of the spectral-spatial classifier, the EMP was computed as a preprocessing step. The best results were obtained for an EMP with $m = 7$ principal components and $n = 7$ openings and closings by reconstruction, using disks with increasing radius 2 (105 components in total), and a composite feature mapping of type “weighted concatenation”. In order to maximize the discriminative capacity of the classifier, each dataset χ was shifted to the range $[0, a]$, where $a = \max(\chi) - \min(\chi)$. This process was performed independently for the hyperspectral data and for each one of the $m(2n+1)$ individual components of the EMP. Then, the weight of the spectral feature, k_ω, was set to 1, while the weight of the spatial feature, k_s, was adjusted for each case. Finally, after concatenating both features, the entire data set was scaled in the range [0,1].

As usual in remote sensing (Faubel et al. 2013), measures of classification accuracy are given in terms of *overall accuracy* (OA), that is the percentage of correctly classified pixels, *average accuracy* (AA), that is computed as the mean of the classes accuracies, and *kappa coefficient* (κ) (Pontius and Millones 2010), which is the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone. Besides, *quantity disagreement* (QD) and *allocation disagreement* (AD) are also given for the ELM classifiers. They measure the disagreement between classification map and reference data in terms of proportion (QD) and spatial allocation (AD) of the classes (Pontius and Millones 2010). The execution times were measured on an Intel Core i7 860 CPU at 2.80 GHz with 8 GB of RAM. The code was compiled in 64-bit Linux using the GNU compiler with -O3 flag.

Three data sets, commonly used in the field of hyperspectral image classification, were used to evaluate the method proposed in this work[2]. The following sections

[2]Details available in `http://wiki.citius.usc.es/software/ELM-EMP`

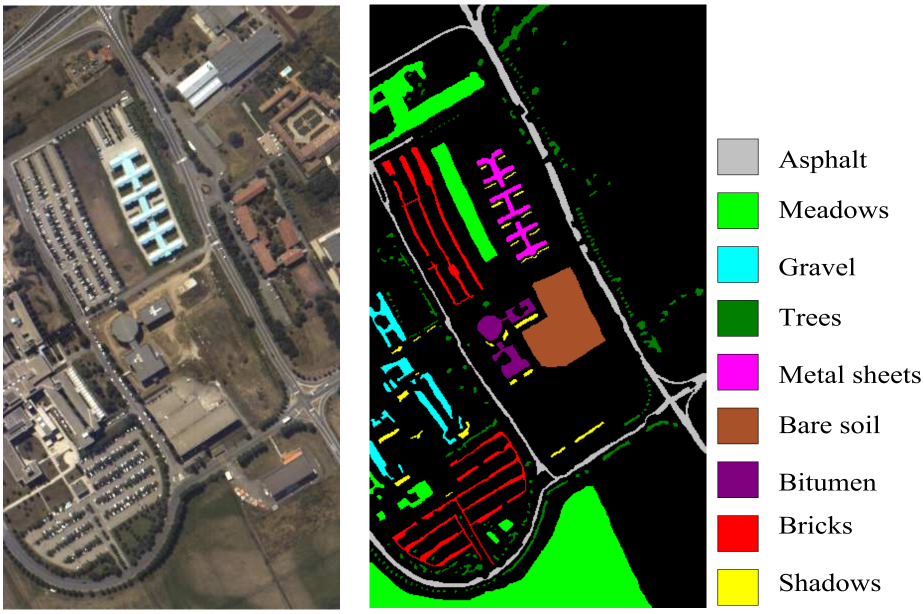


Figure 3. *University of Pavia* image: (a) RGB color space composite of the hyperspectral data; (b) reference data.

describe the characteristics of these images and the results obtained in the classification. In each case, the results of the proposed classifier are compared with equivalent spectral-spatial classifiers based on SVM.

4.1. *Classification of the University of Pavia image*

The first dataset is a 103-band ROSIS (Reflective Optics System Imaging Spectrometer) image of the *University of Pavia*, Italy (ROSIS University of Pavia dataset 2002). This image is an urban area surrounding the University of Pavia, with a spatial size of 610 × 340 pixels, and a spatial resolution of 1.3 m per pixel. The ROSIS-03 sensor captures 115 spectral bands (the twelve most noisy bands of this image were removed) in the spectral range from 0.43 to 0.86 μm, corresponding mainly to the visible spectrum. The number of classes in the reference data is nine: asphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, bricks, and shadows. Hyperspectral image and reference data (ground truth) are displayed in Fig. 3.

The results of classification of this image are shown in Fig. 4. Part (a) of the figure shows the classification map of the pixelwise ELM, i.e., considering only the spectral feature, while part (b) shows the classification map obtained with the proposed ELM-EMP classifier. The false color maps use the same palette as the ground truth shown in Fig. 3.(b).

The accuracy measures for the classification of the *University of Pavia* image are shown in Table 1. The first two columns of the table show the classes available in the reference data and the number of samples used in the training which is the same as in the reference works in the literature (the total number of pixels in the reference data is 42,776). The next column shows the accuracy results for the proposed spectral-spatial classifier.

These results of the proposed ELM-based classification scheme are compared with other methods incorporating spatial information. In this case we compare

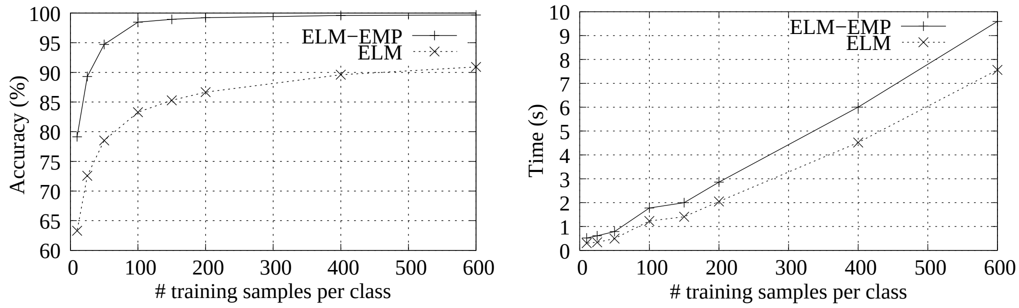
(a) (b)

Figure 4. *University of Pavia* image: (a) classification map for the pixelwise ELM; (b) classification map for the ELM-EMP.

with EMP-KPCA and MSSC-MSF which are the methods in (Faubel et al. 2013) that provide the best results for this image, and with EMAP (Mura et al. 2011) and WT-EMP (Quesada-Barriuso, Argüello, and Heras 2014), that are recent methods based on morphological profiles. EMP-KPCA uses the first 12 KPCs (an extension of PCA using kernel methods) and 4 openings and 4 closings for the construction of the profile (resulting in 108 components). In the case of MSSC-MSF different segmentation techniques are applied in order to extract spatial dependencies through markers. A later minimum spanning forest is used in order to perform thematic classification from these markers using a SAM dissimilarity measure. In EMAP an extended morphological attribute profile is created using morphological attribute filters. These attributes extract information related to size, geometry, and homogeneity of regions and the main features are not extracted by PCA but using ICA (independent component analysis). In WT-EMP the extended morphological profile is created from the features extracted by wavelets. In addition, the hyperspectral image is denoised, also using wavelets. All these methods use SVM for classification.

In (Samat et al. 2014), results of applying techniques based on the use of ensembles with ELM classifiers are also shown to the *University of Pavia* image. The methods, called Bagging-based and AdaBoost-based ELMs, have not been included in the table because the best result that they offer is an OA value of 94.3% obtained by the BoostELMs considering the first 10 PCs with EMPs, which is lower than the values obtained for the methods included in the table.

Table 1 shows that all the presented results are above 96% of OA values. The use of classification-derived markers for marker-controlled region growing (MSSC-MSF) improves the extraction of spatial features by EMP performed by EMP-KPCA. But all these results are improved if the EMP features are extracted from a EMP extracted from a previously wavelet-compressed dataset (WT-EMP). The



(a) (b)

Figure 5. Classification of the *University of Pavia* image depending on the number training samples: (a) overall accuracies; (b) SLFN execution times.

results are improved further if the results are obtained by the proposed ELM-EMP method, which considers a composite feature mapping of weighted concatenation with the same weight for the spatial and the spectral features, achieving the highest accuracy results per class and considering all the classes (OA value of 99.65%).

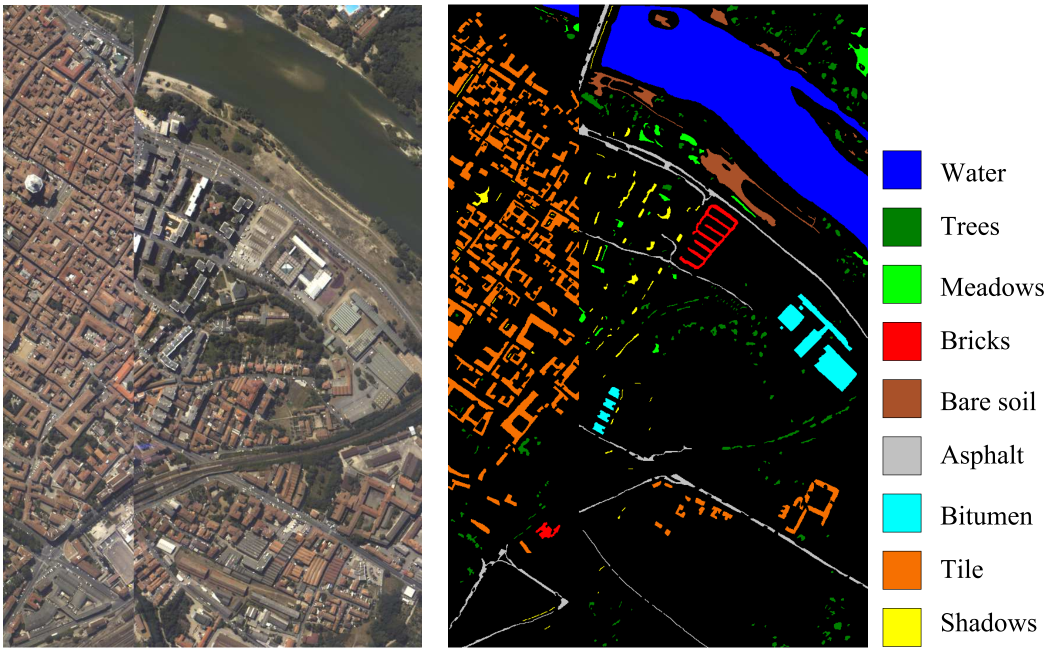
Class	#train	Proposed ELM-EMP	EMP-KPCA	MSSC-MSF	EMAP	WT-EMP
1-asphalt	548	**99.48**	96.0	98.0	-	98.8
2-meadows	540	**99.76**	97.5	96.7	-	98.8
3-gravel	392	**99.76**	81.1	97.8	-	98.8
4-trees	524	99.27	**99.3**	98.8	-	99.2
5-metal sheets	265	99.48	99.4	**99.9**	-	**99.9**
6-bare soil	532	99.78	92.2	**100**	-	98.5
7-bitumen	375	99.52	98.8	**99.9**	-	99.0
8-bricks	514	99.48	99.4	**99.8**	-	98.0
9-shadows	231	99.84	98.0	96.5	-	**99.9**
OA		**99.65** $\pm$ 0.08	96.3	97.9	94.5	98.8
AA		**99.60** $\pm$ 0.06	95.7	98.6	96.6	99.0
κ		**99.52** $\pm$ 0.11	95	97	93	98
QD		0.15 $\pm$ 0.07	-	-	-	-
AD		0.19 $\pm$ 0.04	-	-	-	-
Time (SLFN)		6.01 s	-	-	-	-

Table 1. Information and accuracies in percentage for the *University of Pavia* image. For ELM-EMP the parameters used are $L = 1000$ nodes in the hidden layer and spatial weight $k_s = 1$. The results shown are the average of 100 runs ($\pm$ standard deviations).

Fig. 5 shows the graphs of accuracy and computing times as a function of the size of the training set (number of samples per class) for the *University of Pavia* image. The execution times indicate the total time (training plus testing) for all the pixels of the image (610×340 pixels), considering only the SLFN. The number of nodes in the hidden layer was selected to obtain the best results in each case and ranges between 25 nodes (for 10 samples/class) and 1400 (for 600 samples/class). In all cases the spatial weight was set to 1. Moreover, these graphs compare the results of the pixelwise ELM classifier with those of the ELM-EMP one. As expected, in all cases the accuracy increases with the size of the training set, while at the same time the runtime increases. Moreover, the ELM-EMP classifier provides accuracies a 10% higher than the pixelwise one, and requires fewer training samples for achieving the same accuracy. For example, to achieve 90% accuracy, ELM-EMP requires only 25 training samples/class, while the pixelwise one requires 400.

4.2. *Classification of the Pavia Centre image*

The second ROSIS data set is the *Pavia Centre* (ROSIS Pavia Centre dataset 2002). The image was originally of a size of 1096×1096 pixels, but a 381-pixel-



(a) (b)

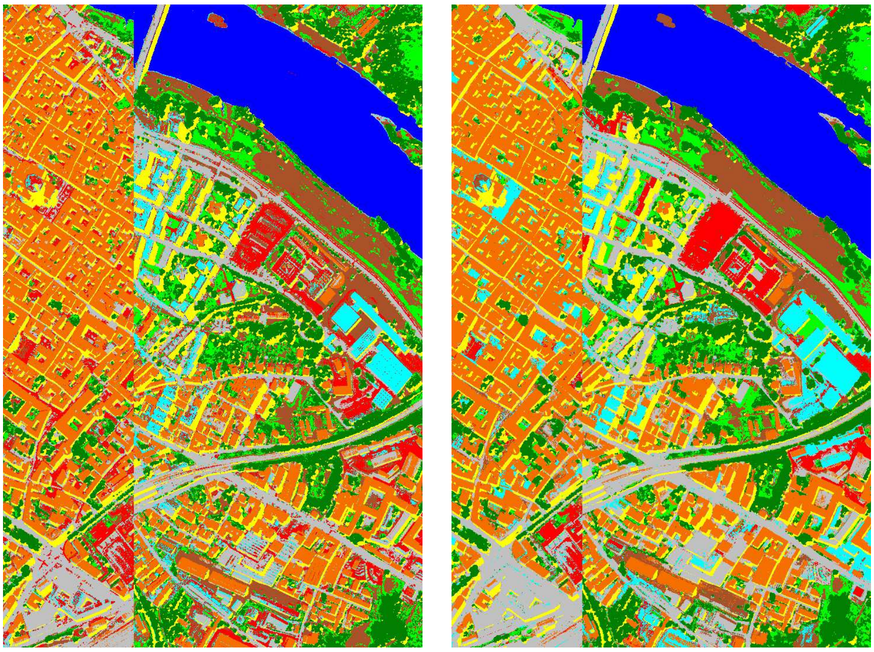
Figure 6. *Pavia Centre* image: (a) RGB color space composite; (b) reference data.

wide black vertical band in the middle of the image was removed, resulting in a two-part image of 1096×715 pixels. Thirteen channels have been removed due to noise, containing the resulting image 102 spectral channels. Nine classes of interest are considered: water, trees, meadows, bricks, soil, asphalt, bitumen, tiles, and shadows. The hyperspectral image and the reference data are displayed in Fig. 6.

The classification results for the *Pavia Centre* image are shown in Fig. 7 and the accuracy measures in Table 2. The total number of pixels available in the reference data is 414,815, which is high. Accordingly, a training set of around 800 samples per class has been selected (the same number per class as in the SVM classifiers in the literature shown in the table). The rest of the parameters chosen are: 1000 nodes in the hidden layer and a spatial weight of 1. It is interesting to note in the figure that the "salt-and-pepper" classification noise of the thematic map obtained with the spectral information alone (pixelwise ELM) is removed or reduced when the spatial information is added (proposed ELM-EMP).

Regarding Table 2 the accuracy measures of the ELM-based proposed technique are compared to the best results in (Faubel et al. 2013) for the image, EMP-KPCA and spec-EMP, and to ICA-EMP (Palmason et al. 2005) and WT-EMP (Quesada-Barriuso, Argüello, and Heras 2014). All these methods use SVM as classifier. EMP-KPCA uses the first 10 KPCs (an extension of PCA using kernel methods) and 4 openings and 4 closings for the construction of the profile (resulting in 108 components). Spec-EMP uses the first three principal components, 4 openings and 4 closings for the construction of the profile, and a stacked vector. ICA-EMP denotes the EMP created from the first three principal components extracted by independent component analysis. Finally for WT-EMP the EMP is created from the features extracted by wavelets and a stacked vector is built joining the original denoised image and the EMP. For this last method the WT-EMP is created with five levels of decomposition by a CDF97 wavelet and the EMP is calculated from the four approximation bands, thus obtaining a stacked vector of 138 features.

The accuracy results for this image are very good for all the classifiers. The



(a) (b)

Figure 7. *Pavia Centre* image: (a) classification map for the pixelwise ELM; (b) classification map for the proposed ELM-EMP.

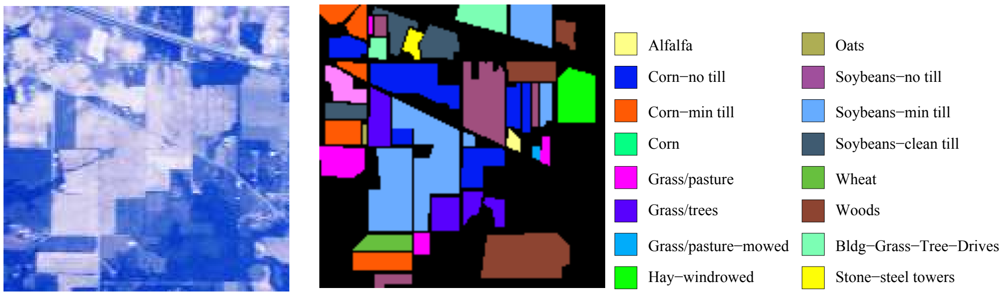
Class	#train	Proposed ELM-EMP	EMP-KPCA	spec-EMP	ICA-EMP	WT-EMP
1-water	824	**100**	98.9	98.7	99.5	99.9
2-trees	820	**99.16**	92.0	93.5	91.7	97.4
3-meadows	824	**99.67**	96.3	96.0	85.3	99.1
4-bricks	808	99.76	99.6	98.8	99.2	**99.8**
5-bare soil	820	**99.93**	99.8	99.4	98.4	99.9
6-asphalt	816	**99.79**	99.2	98.4	98.6	99.5
7-bitumen	808	**99.01**	98.6	98.2	98.1	98.5
8-tiles	1260	**99.91**	99.9	99.8	99.8	99.9
9-shadows	476	99.88	99.6	99.9	99.2	**100**
OA		**99.86** $\pm$ 0.02	98.9	99.7	98.8	99.7
AA		**99.68** $\pm$ 0.06	98.3	98.1	96.6	99.3
κ		**99.80** $\pm$ 0.03	98	98	-	99
QD		0.10 $\pm$ 0.02	-	-	-	-
AD		0.04 $\pm$ 0.01	-	-	-	-
Time (SLFN)		20.66 s	-	-	-	-

Table 2. Information and accuracies in percentage for the *Pavia Centre* image. For ELM-EMP the parameters used are $L = 1000$ nodes in the hidden layer and spatial weight $k_s = 1$. The results shown are the average of 100 runs ($\pm$ standard deviations).

methods ICA-EMP and WT-EMP, that are based on SVM and differ mainly in using ICA or wavelets for feature extraction, obtain OA values of 98.8 and 99.7 respectively. The proposed method based in ELM improves them achieving 99.86. The execution times are higher in this case due to the size of the image (1096×715 pixels).

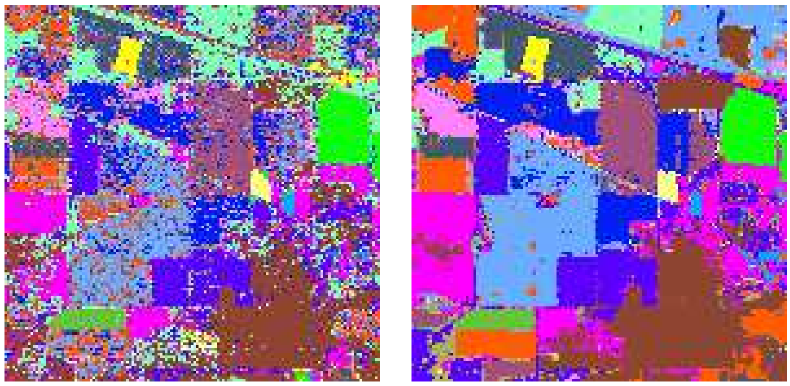
4.3. *Classification of the Indian Pine image*

The third dataset is a 220-band AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) image taken over the *Indian Pine* Test Site 3 (2×2 mile portion of NW Tippecanoe County, Indiana) (AVIRIS Indian Pine dataset 1992). This image is a rural area with a size of 145×145 pixels and a spatial resolution of 20 m per pixel. The AVIRIS sensor has a number of spectral bands of 220 and a spectral



(a) (b)

Figure 8. *Indian Pine* image: (a) RGB color space composite; (b) reference data.



(a) (b)

Figure 9. *Indian Pine* image: (a) classification map for the pixelwise ELM; (b) classification map for the ELM-EMP.

range from 0.4 of 2.5 μm, corresponding to the visible and infrared spectrum. The scene contains two-thirds agriculture and one-third forest or other natural perennial vegetation. Some of the crops present (corn, soybeans, etc) are in early stages of growth with less than 5% coverage. There are two major dual lane highways, smaller roads, a rail line, some low density housing and other built structures. Sixteen classes of interest are considered, which are not all mutually exclusive. This image is displayed in Fig. 8.

The results for the classification of the *Indian Pine* image are shown in Fig. 9 and information about the classes and accuracy measures in Table 3. The total number of pixels available in the reference data is 10,366, but some classes contain very few samples (only 20 samples in the *oats* class and 28 samples in the *grass-pasture-mowed* one). So, a number of samples of 50 per class (or 15 in the small classes) has been selected for training. Accordingly, the hidden layer of the ELM includes 300 nodes, a smaller number that in the previous images.

In the table, the results are compared with those provided by SVM-MSF+MV and MSSC-MSF, that are the best methods in (Faubel et al. 2013) for this image, and with other two very accurate methods in the literature, NW-NW (Marpu et al. 2012) and WT-EMP (Quesada-Barriuso, Argüello, and Heras 2014). SVM-MSF+MV uses marker selection using SVM followed by MSF (minimum spanning forest) segmentation and majority voting. For this image the MSF algorithm uses SAM as dissimilarity function. On the other hand, MSSC-MSF in based on marker selection using the MSSC approach (multiple classifications based on SVM are performed in parallel) followed by MSF without the majority voting step.

Class	#train	Proposed ELM-EMP	SVM-MSF +MV	MSSC-MSF	NW-NW	WT-EMP
1-corn-no till	50	86.86	**93.2**	89.7	90.24	91.5
2-corn-min till	50	92.65	96.6	87.0	**98.85**	83.7
3-corn	50	96.49	95.7	95.1	**97.28**	85.7
4-soybean-no till	50	90.15	93.9	91.8	**95.52**	94.6
5-soybean-min till	50	89.60	82.0	89.2	**99.56**	92.3
6-soybean-clean	50	92.95	97.2	97.4	**99.54**	97.2
7-alfalfa	15	92.56	94.9	94.9	86.27	**100**
8-grass-pasture	50	94.34	94.6	94.6	94.58	**99.1**
9-grass-trees	50	98.32	97.3	97.9	93.61	**100**
10-grass-past-mowed	15	95.64	**100**	**100**	99.38	84.4
11-hay-windrowed	50	99.81	**99.8**	**99.8**	92.76	74.9
12-oats	15	**100**	**100**	**100**	99.09	84.6
13-wheat	50	99.5	99.4	99.4	**100**	99.2
14-woods	50	98.89	**99.7**	99.4	94.87	87.7
15-bldg-grass-trees-dr	50	97.12	68.8	73.6	**100**	94.3
16-stone-steel-towers	50	95.49	95.6	97.8	**100**	96.9
OA		92.81 ± 0.76	91.8	92.3	**96.35**	86.6
AA		**95.02** ± 0.65	94.3	94.2	94.17	92.3
κ		91.77 ± 0.86	91	91	**93.33**	84.8
QD		2.63 ± 0.65	-	-	-	-
AD		4.55 ± 0.58	-	-	-	-
Time (SLFN)		0.25 s	-	-	-	-

Table 3. Information and accuracies in percentage for the *Indian Pine* image. For ELM-EMP the parameters used are $L = 300$ nodes in the hidden layer and spatial weight $k_s = 5$. The results shown are the average of 100 runs ($\pm$ standard deviations).

The best results among these two methods are achieved when the MSSC approach is used as it happens for the other two images considered. NW-NW consists in a supervised feature reduction using NWFE, then the EMAP considering 30 features is built and, finally, other NWFE feature reduction is performed on the EMAP before the final SVM classification. This technique obtains the best OA result (96.35) showing, as the authors of the method claim, that NWFE performs better in the case where there are a limited number of samples. Nevertheless, for this image the best AA value (95.02) is obtained for the proposed ELM-EMP technique and not for NW-NW. It is interesting to note that, although in (Marpu et al. 2012) the computational cost of NW-NW is not analyzed, using SVM and two phases of feature reduction, NW-NW should be computationally more costly than ELM-EMP. The WT-EMP method obtain worse results for this image.

5. Conclusions

In this paper we have proposed a spectral-spatial ELM-based classifier and evaluated its accuracy results on several hyperspectral images commonly used in remote sensing for land-cover applications. The classifier is based on the ELM method, which is an alternative to the more usual ones based on neural networks or SVM.

The proposed ELM classifier uses jointly the spatial and spectral information available in the hyperspectral image. In particular, the proposed classifier has as input the original spectral-vector (the spectral feature) and an extended morphological profile (the spatial feature). Additionally, we have studied the alternatives to construct a composite feature mapping that integrates both features in the classifier. Moreover, the feature mapping makes it possible to assign different weights to both features in order to give more dominance to the spectral or the spatial part of the classification.

We have performed several experiments on three hyperspectral images (*University of Pavia*, *Pavia Centre* and *Indian Pine*), commonly used for testing in remote sensing. These images correspond to rural and urban scenes and contain various types of crops, forests and other natural vegetation, asphalt, gravel, bricks, bitumen, etc. In these experiments, the proposed spectral-spatial ELM-based classifier substantially improved the accuracy results compared to equivalent SVM-based classifiers. The number of training samples was selected to be similar to that of similar classifiers used for comparison, while the number of nodes in the hidden layer and the weights of the spatial and the spectral features were selected to obtain the most accurate classification results. In our experiments, the proposed classifier provided in most of the cases better accuracy results than equivalent spectral-spatial SVM-based classifiers.

In general, ELM-based methods are a good choice for the classification of hyperspectral images over other methods based on SVM or neural networks, due to their simplicity and ease of use and because they provide very competitive results. Furthermore, the possibility of incorporating into them spectral-spatial techniques such as the one presented in this work, makes them highly suitable for hyperspectral imaging.

Funding

This work was supported in part by the Ministry of Science and Innovation, Government of Spain, cofounded by the FEDER funds of European Union [grant number TIN2013-41129-P] and by Xunta de Galicia, Programme for Consolidation of Competitive Research Groups [grant number 2014/008].

Supplemental data

The underlying research materials for this article can be accessed at `http://wiki.citius.usc.es/software/ELM-EMP`

References

AVIRIS Indian Pine Test Site 3 dataset (2x2 mile portion of Northwest Tippecanoe County, Indiana). 1992. Accessed April 3, 2014. `ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.tif.zip` (220–band hyperspectral image) and `ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/ThyFiles.zip` (ground truth).

Baldi, P., C. Azencott, and S. J. Swamidass. 2011. “Bridging the Gap Between Neural Network and Kernel Methods: Applications to Drug Discovery.” in *Frontiers in Artificial Intelligence and Applications*, vol. 226, B. Apolloni, S. Bassis, A. Esposito, and C.F. Morabito, Eds, p. 3–13, IOS Press.

Benediktsson, J. A., and I. Kanellopoulos. 1999. “Classification of multisource and hyperspectral data based on decision fusion.” *IEEE Trans. Geoscience Remote Sensing* 37: 1367–1377.

Benediktsson, J. A., M. Pesaresi, and K. Arnason. 2003. “Classification and feature extraction for remote sensing images from urban areas based on morphological transformations.” *IEEE Trans. Geoscience Remote Sensing* 41(9): 1940–1949.

Benediktsson, J. A., J. A. Palmason, and J. R. Sveinsson. 2005. “Classification of hyperspectral data from urban areas based on extended morphological profiles.” *IEEE Trans. Geoscience and Remote Sensing* 43(3): 480–491.

Broomhead, D. S. and D. Lowe. 1988. “Multivariable functional interpolation and adaptive networks.” *Complex Syst.* 6(2): 321–355. classification.” *IEEE Geoscience Remote Sensing Letters* 3(1): 93–97.

Camps–Valls, G., and L. Bruzzone. 2005. “Kernel–based methods for hyperspectral image classification.” *IEEE Trans. Geoscience and Remote Sensing* 43(6): 1351–1362.

Camps–Valls, G., L. Gomez–Chova, J. Munoz–Mari, J.Vila–Francés, and J. Calpe–Maravilla. 2006. "Composite kernels for hyperspectral image

Cao, J., Z. Lin, G.–B. Huang, and N. Liub. 2012. "Voting based extreme learning machine." *Information Sciences* 185(1): 66–77.

Cao, F., B. Liu, and D. S. Park. 2013. "Image classification based on effective extreme learning machine." *Neurocomputing* 102: 90–97.

Chang, N.–B. 2011. "Satellite–based multitemporal–change detection in urban environments." *SPIE Newsroom*, 23 February 2011, DOI: 10.1117/2.1201101.003502

Chen, Z. X., H. Y. Zhu, and Y. G. Wang. 2013. "A modified extreme learning machine with sigmoidal activation functions." *Neural Computing & Applications* 22(3–4): 541–550.

Courrieu, P. 2005. "Fast computation of Moore–Penrose inverse matrices." *Neural Information Processing* 8(2): 25–29.

Crespo, J., J. Serra, and R. W. Schafer. 1995. "Theoretical aspects of morphological filters by reconstruction." *Signal Processing* 47(2): 201–225.

Fauvel, M., J. Chanussot, and J. A. Benediktsson. 2006. "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data." in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, II813–II816.

Fauvel, M., J. Chanussot, and J. A. Benediktsson. 2012. "A spatial–spectral kernel–based approach for the classification of remote–sensing images." *Pattern Recogn.* 45(1): 381–392.

Fauvel, M., Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. 2013. "Advances in spectral–spatial classification of hyperspectral images." *Proceedings of the IEEE* 101(3): 652–675.

Heras, D. B., F. Argüello, J. L. Gomez, J. A. Becerra, and R. J. Duro. 2012. "Towards real–time hyperspectral image processing, a GP–GPU implementation of target identification." Chap. 9 in *Digital Image and Signal Processing for Measurement Systems*, R. J. Duro and F. Lopez Pea, Eds., 235–265, River Publishers.

Heras, D. B., F. Argüello, and P. Quesada–Barriuso. 2014. "Exploring ELM–based spatial–spectral classification of hyperspectral images." *International Journal of Remote Sensing* 35(2): 401–423.

Huang, G.–B., Q.–Y. Zhu, and C.–K. Siew. 2004. "Extreme learning machine: a new learning scheme of feedforward neural networks." in *Proc. Int. Joint Conf. on Neural Networks (IJCNN)*, 985–990.

Huang, G.–B, L. Chen, and C.–K. Siew. 2006. "Universal approximation using incremental constructive feedforward networks with random hidden nodes". *IEEE Trans. Neural Networks* 17(4): 879–892.

Huang, G.–B., Q.–Y. Zhu, and C.–K. Siew. 2006. "Extreme learning machine: theory and applications." *Neurocomputing* 70(1): 489–501.

Huang G.–B. and L. Chen. 2008. "Enhanced random search based incremental extreme learning machine." *Neurocomputing* 71(16): 3460–3468.

Huang, G.–B., D.–H. Wang, and Y. Lan. 2011. "Extreme learning machines: a survey." *International Journal of Machine Learning and Cybernetics* 2(2): 107–122.

Huang, G.–B. 2012. "Extreme learning machine for regression and multiclass classification." *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics* 42(2): 513–529.

Huang, G.–B. 2014. "An insight into extreme learning machines: random neurons, random features and kernels". *Cognitive Computation* 6(3): 376–390.

Igelnik B. and Y.–H. Pao. 1995. "Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans. Neural Networks* 6(6): 1320-1329.

Landgrebe, D. 2002. "Hyperspectral image data analysis." *IEEE Signal Processing Magazine* 19(1): 17–28.

Licciardi, G., P. R. Marpu, J. A. Benediktsson, and J. Chanussot. 2011. "Extended morphological profiles using auto–associative neural networks for hyperspectral data classification." in *Proc. IEEE. Workshop Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–4.

Lowe, D. 1989. "Adaptive radial basis function nonlinearities, and the problem of generalisation", in *Proc. 1st Inst. Electr. Eng. Int. Conf. Artif. Neural Netw.*, 171-175.

Marconcini, M., G. Camps–Valls, and L. Bruzzone. 2009. "A composite semisupervised SVM for classification of hyperspectral images." *IEEE Geoscience and Remote Sensing Letters* 6(2): 234–238.

Marpu, P.R., M. Pedergnana, M. D. Mura, S. Peeters, J.A. Benediktsson and L. Bruzzone. 2012. "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques." *International Journal of Image and Data Fusion* 3(3): 269–298.

Melgani, F., and L. Bruzzone. 2004. "Classification of hyperspectral remote sensing images with support vector machines." *IEEE Trans. Geoscience and Remote Sensing* 42(8): 1778–1790.

Mercier, G., and M. Lennon. 2003. "Support vector machines for hyperspectral image classification with spectral–based kernels." in *Proc. IEEE Int. Conf. Geoscience and Remote Sensing Symposium (IGARSS)*, 288–290.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support vector machines in remote sensing: A review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3): 247–259.

Mura, M. D., J.A. Benediktsson, B. Waske, and L. Bruzzone. 2010. "Morphological attribute profiles for the analysis of very high resolution images." *IEEE Trans. Geoscience and Remote Sensing* 48(10): 3747–3762.

Mura, D. M., A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone. 2011. "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis." *IEEE Geoscience and Remote Sensing Letters* 8(3): 542–546.

OpenBLAS. 2011. Accessed April 3, 2014. `http://xianyi.github.com/OpenBLAS`.

Pal, M. 2009."Extreme–learning–machine–based land cover classification." *International Journal of Remote Sensing* 30(14): 3835–3841.

Pal, M., A. E. Maxwell, and T. A. Warner. 2013. "Kernel–based extreme learning machine for remote–sensing image classification." *Remote Sensing Letters* 4(9): 853–862.

Palmason, J. A., J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot. 2005. "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis." in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*, 176–179.

Pao, Y.–H., G.–H. Park, and D. J. Sobajic. 1994. "Learning and generalization characteristics of the random vector Functional-link net", *Neurocomputing.* 6(2): 163–180.

Pesaresi, M., and J. A. Benediktsson. 2001. "A new approach for the morphological segmentation of high–resolution satellite imagery." *IEEE Trans. Geoscience and Remote Sensing* 39(2): 309–320.

Plaza, J., A. J. Plaza, and C. Barra. 2009. "Multi–channel morphological profiles for classification of hyperspectral images using support vector machines." *Sensors* 9(1): 196–218.

Plaza, A., J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps–Valls, J. Chanussot, et al. 2009. "Recent advances in techniques for hyperspectral image processing." *Remote Sens. Environ.* 113: S110–S122.

Pontius Jr, R. G., and M. Millones. 2010. "Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment." *International Journal of Remote Sensing* 32(15): 4407–4429.

Quesada-Barriuso, P., F. Argüello, and D. B. Heras. 2014. "Spectral-Spatial Classification of Hyperspectral Images Using Wavelets and Extended Morphological Profiles", *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4): 1177-1185.

ROSIS Pavia Centre dataset. 2002. Accessed April 3, 2014 `http://www.ehu.es/ccwintco/uploads/e/e3/Pavia.mat` (102–band hyperspectral image) and `http://www.ehu.es/ccwintco/uploads/5/53/Pavia_gt.mat` (ground truth).

ROSIS University of Pavia dataset. 2002. Accessed April 3, 2014. `http://www.ehu.es/ccwintco/uploads/e/ee/PaviaU.mat` (103–band hyperspectral image) and `http://www.ehu.es/ccwintco/uploads/5/50/PaviaU_gt.mat` (ground truth).

Samat, A., P. Du, S. Liu, J. Li and L. Cheng. 2014. "E2LMs: Ensemble Extreme Learning Machines for Hyperspectral Image Classification." *IEEE Journal of Selected Topics in*

Applied Earth Observations and Remote Sensing 7(4): 1060–1069.

Scholkopf, B., and A. J. Smola. 2002. *Learning with Kernels. Support Vector Machines Regularization Optimization and Beyond*, MIT Press.

Schmidt, W.F., M.A. Kraaijveld, and R.P.W. Duin. 1992 "Feed forward neural networks with random weights." In *Proc. 11th IAPR Int. Conf. on Pattern Recognition*, Vol.II. Conference B: Pattern Recognition Methodology and Systems: 1-4.

Soille, P., and M. Pesaresi. 2002. "Advances in mathematical morphology applied to geoscience and remote sensing." *IEEE Trans. Geoscience and Remote Sensing* 40(9): 2042–2055.

Soille, P. 2005. "Beyond self–duality in morphological image analysis." *Image and Vision Computing* 23(2): 249–257.

Tarabalka, Y., J. A. Benediktsson, and J. Chanussot. 2009. "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques." *IEEE Trans. Geoscience and Remote Sensing* 47(8): 2973–2987.

Tarabalka, Y., J. Chanussot, and J. A. Benediktsson. 2010 "Segmentation and classification of hyperspectral images using watershed transformation." *Pattern Recognit.* 43(7): 2367–2379.

Tarabalka, Y., J. Chanussot, and J. A. Benediktsson. 2010. "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers." *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics* 40(5): 1267–1279.

Tarabalka, Y., M. Fauvel, J. Chanussot, and J.A. Benediktsson. 2010. "SVM and MRF–based method for accurate classification of hyperspectral images." *IEEE Geoscience Remote Sensing Letters* 7(4): 736–740.

Wang, L. and C. Wan. 2008. "Comments on "the extreme learning machine"" *IEEE Trans. Neural Networks* 19(8): 1494–1495.

Wang, Y., F. Cao, and Y. Yuan. 2011. "A study on effectiveness of extreme learning machine." *Neurocomputing* 74(16): 2483–2490.

Zhang, Y., X. Feng, and X. Le. 2008. "Segmentation on multispectral remote sensing image using watershed transformation." in *Proc. Int. Congress on Image and Signal Processing (CISP)*, 773–777.