

An Ensemble-Based Stacked Sequential Learning Algorithm for Remote Sensing Imagery Classification

Danillo R. Pereira, Rodrigo J. Pisani, André N. de Souza, and João P. Papa, *Member, IEEE*

Abstract—Contextual-based image classification attempts at considering spatial/temporal information during the learning process in order to make the classification process smarter. Sequential learning techniques are one of the most used ones to perform contextual classification, being based on a two-step classification process, in which the traditional noncontextual learning process is followed by one more step of classification based on an extended feature vector. In this paper, we propose two ensemble-based approaches to make sequential learning techniques less prone to errors, since their effectiveness is strongly dependent on the feature extension process, which ends up adding the wrong predicted label of the neighborhood samples as new features. The proposed approaches are validated in the context of land-cover classification, being their results considerably better than some state-of-the-art techniques in the literature.

Index Terms—Land-cover classification, optimum-path forest (OPF), sequential learning.

I. INTRODUCTION

MACHINE-LEARNING techniques have been considered a game-changing in the way one organizes and analyzes data in different areas. Although traditional pattern recognition techniques have been widely used in several research areas [1], such approaches usually assume that the dataset samples are independent and identically distributed. Such an assumption means that the samples' spatial/temporal dependence are not considered during the learning process, which may be a serious shortcoming when dealing with applications that require such knowledge. Time-series prediction of finance-related problems and meteorological observations are some examples that may not fit well in models that do not employ contextual information. In the context of image classification, a way to introduce prior knowledge in the problem formulation is to employ smoothness

constraints considering the spatial context of the data. When looking at a picture or a video, for instance, one can realize that the pixels vary smoothly with the exception of the high-frequency regions (borders and discontinuities).

Pixel-based classification has been the foremost approach for satellite image classification over the past decades, since the assumption of independent and identically distributed pixels is surmised and, hence, employed by a considerable number of works [2]–[6]. However, the problem of modeling each pixel as a sole entity without considering its neighborhood may lead us to a poorly labeled image, since pixels within homogeneous regions are likely to describe the similar content. A forthright approach to deal with this problem is the so-called contextual-based classification, in which a given pixel and its neighborhood are then used to enhance the labeling process [7], [8]. Such approaches, also referred to semantic-based classification [9], make use of additional data provided by spatial/temporal information to provide more accurate results. Among several approaches for such purpose, one may highlight the stacked sequential learning (SSL) proposed by Cohen and Carvalho [10], which aims at modeling contextual information by means of a two-stage classification process: 1) in the first one, a naïve classification step is performed, i.e., a training procedure followed by the classification of test samples are carried out; and 2) soon after, the original feature vectors of training and test samples are extended with the labels of their neighborhood, followed by an ordinary training and classification processes. The SSL uses a unique classifier in the first step of the two-stage classification.

Later on, Gatta *et al.* [11] proposed a multiscale sequential learning approach, in which the contextual information is obtained not only from the sample's neighborhood, but also from pixels far apart. The idea of multiple scales is driven by several Gaussian-convolved labeled images, which are former obtained by means of a traditional classification process. Afterward, Puertas *et al.* [12] addressed the aforementioned work in the context of multiclass-based classification problems. Sampedro *et al.* [13] proposed a similar approach to the one introduced by Puertas *et al.* [12], but now in a 3-D space and using error-correcting output codes in the context of medical image classification, and González *et al.* [14] employed the SSL paradigm for pedestrian detection. In 2015, Puertas *et al.* [15] presented an improvement of the former multiscale sequential learning approach proposed by Gatta *et al.* [11]. This new version now supports multiclass problems, as well as the authors proposed to compress the final extended feature vector with minimal effect in the recognition rates.

Manuscript received September 29, 2016; revised November 16, 2016 and December 8, 2016; accepted December 14, 2016. Date of publication January 15, 2017; date of current version March 22, 2017. This work was supported by FAPESP under Grant #2013/20387-7, Grant #2014/16250-9, and Grant #2014/12236-1, and by CNPq under Grant #303182/2011-3, Grant #470571/2013-6, Grant #487032/2012-8, and Grant #306166/2014-3. (*Corresponding author: João P. Papa.*)

D. R. Pereira is with the University of the West São Paulo, Presidente Prudente 19050-920, Brazil (e-mail: dpereira@ic.unicamp.br).

R. J. Pisani is with the Nature Sciences Institute, Federal University of Alfenas, Alfenas 37130-000, Brazil (e-mail: pisanigeo@gmail.com).

A. N. Souza is with the Department of Electrical Engineering, São Paulo State University, Bauru 17033-360, Brazil (e-mail: andrejau@feb.unesp.br).

J. P. Papa is with the Department of Computing, São Paulo State University, Bauru 17033-360, Brazil (e-mail: papa@fc.unesp.br).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2016.2645820

Roscher *et al.* [16] used an incremental learning algorithm based on support vector machines (SVMs) to classify hyperspectral data, being the main idea to improve the quality of the training set with new informative samples, but at the same time removing noninformative ones. Apart from remote sensing-based applications, but related to sequential learning, Subramanian and Suresh [17] presented a neuro-fuzzy-based sequential learning algorithm that aimed at controlling the learning process. Roughly speaking, the idea is to use the error over a misclassified sample in order to learn the best training strategy for that sample. Other works have employed online sequential learning techniques either, but using extreme learning machines and in the context of social networks [18] and time-series prediction [19].

Since sequential learning also concerns contextual classification, a number of related works can be referred in the literature. Susa *et al.* [20] have recently used the Gaussian process to classify remote sensing images by means of a Bayesian-theoretic framework, and Li *et al.* [21] highlighted the importance of employing spatiocontextual information in remote sensing image classification. SVMs have been considered in contextual-based learning either [22], [23], where Markov random fields (MRFs) play the role of considering neighboring information during the learning process. Recently, Osaku *et al.* [24] showed that one can improve the classification of satellite images using a contextual version of the optimum-path forest (OPF) classifier [25]–[27], which makes use of MRFs, the spatial dependence among nearby pixels either.

Although the aforementioned works obtained very promising results, it is possible to improve their approaches by means of ensemble of classifiers. The main drawback with respect to sequential learning is related to the misclassified samples during the first classification step, which can lead the second classifier to errors, since the feature vector of each sample is extended with the predicted labels of its neighbors. Albeit all the aforementioned approaches for sequential learning may employ different methodologies, all of them rely on the very same idea of extending the number of features for each sample. Therefore, this paper proposes two ensemble-based approaches to sequential learning, which are validated in the context of land-use classification. We have observed that a committee of learners can improve the classification results when dealing with sequential learning approaches, which are strongly dependent on the first (initial) classification results.

The proposed approaches are validated over OPF and a Naïve Bayes classifier, since they are parameterless and quite fast for training. However, it is important to shed light on that the proposed approaches can be used with any other supervised pattern recognition techniques. In regard to SSL-based implementations considering the OPF classifier, the reader can refer to only one very recent work conducted by Pereira *et al.* [28], which evaluated OPF in the context of land-use classification in satellite and radar images. Therefore, the main contributions of this paper are threefold: 1) to present a voting-based and a 2) concatenation-based approach to enhance sequential learning, and 3) to foster the research on the OPF classifier regarding remote sensing

images. The OPF classifier has gained attention and popularity in the past years [29], [30], since it has been consistently similar or even more accurate than SVMs, but faster for training. It has a number of advantages over some state-of-the-art pattern recognition techniques.

- 1) It is parameterless.
- 2) It does not make assumptions about separability of the samples in the feature space.
- 3) It can be easily adapted to different situations by just changing some of its main components.

Roughly speaking, OPF can be seen as a framework to the design of pattern classification techniques based on the theory about OPFs, which means one can design a new OPF classifier by just designing some of its modules (more details about it are given further).

The remainder of this paper is organized as follows. Section II presents a brief theoretical background about the OPF and Naïve-Bayes, and Section III revisits the techniques of SSL employed in this paper for comparison purposes. The novel proposed approaches are described in Section IV. The methodology and experimental results are discussed in Sections V and VI, respectively. Section VII states conclusions and future works.

II. THEORETICAL BACKGROUND

In this section, a brief theoretical background concerning the pattern recognition techniques used in this work is presented.

A. Optimum-Path Forest

The OPF framework is a recent highlight to the development of pattern recognition techniques based on graph partitions. The nodes are the data samples, which are represented by their corresponding feature vectors, and are connected according to some predefined adjacency relation (i.e., a complete or a k -NN graph). Given some key nodes (prototypes), they will compete among themselves aiming to conquer the remaining nodes. Thus, the algorithm outputs an OPF, which is a collection of optimum-path trees (OPTs) rooted at each prototype. This work employs the OPF classifier proposed by Papa *et al.* [25], [26], which uses a complete graph and a path-cost function that computes the maximum arc-weight along a path. Additionally, the key nodes are the nearest samples from different classes, which are obtained by means of a minimum spanning tree (MST) computation over the training set. Follow, below, a more detailed explanation about the OPF mechanism.

Let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ be a labeled dataset, such that \mathcal{D}_1 and \mathcal{D}_2 stand for the training and test sets, respectively. Let $\mathcal{P} \subset \mathcal{D}_1$ be a set of prototypes of all classes (i.e., key samples that best represent the classes). Let (\mathcal{D}_1, A) be a complete graph, whose nodes are the samples in \mathcal{D}_1 and any pair of samples defines an arc in $A = \mathcal{D}_1 \times \mathcal{D}_1$, as displayed in Fig. 1(a). Additionally, let π_s be a path in (\mathcal{D}_1, A) with terminus at sample $s \in \mathcal{D}_1$.

The OPF algorithm proposed by Papa *et al.* [25], [26] employs the path-cost function f_{\max} due to its theoretical properties for estimating prototypes (Section II-A1 gives further details about

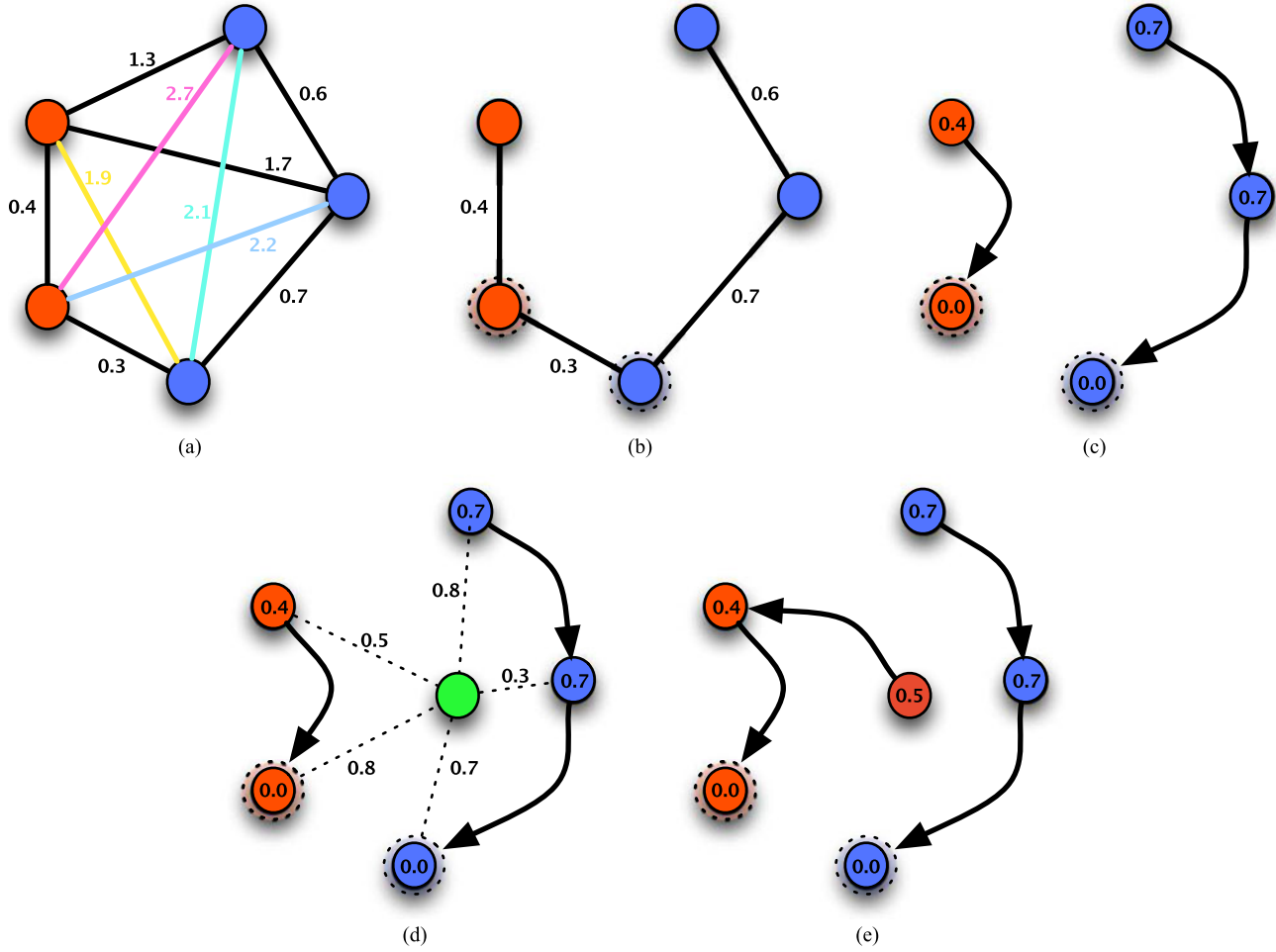


Fig. 1. (a) Training set modeled as a complete graph. (b) MST computation over the training set (prototypes are highlighted). (c) OPF over the training set. (d) Classification process of a “green” sample. (e) Test sample is finally classified.

this procedure):

$$f_{\max}(\langle s \rangle) = \begin{cases} 0, & \text{if } s \in P \\ +\infty, & \text{otherwise} \end{cases}$$

$$f_{\max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{\max}(\pi_s), d(s, t)\} \quad (1)$$

where $d(s, t)$ stands for a distance (e.g., Euclidean distance) among nodes s and t , such that $s, t \in \mathcal{D}_1$. Therefore, $f_{\max}(\pi_s)$ computes the maximum distance between adjacent samples in π_s , when π_s is not a trivial path. Roughly speaking, the main idea of OPF is to minimize $f_{\max}(\pi_t), \forall t \in \mathcal{D}_1$. By minimizing f_{\max} , one can create an OPF rooted at S , which is essentially a collection of OPTs that are composed of their most strongly connected nodes.

1) *Training*: Let \mathcal{P}^* be an optimum set of prototypes when the OPF algorithm minimizes the classification errors for every $s \in \mathcal{D}_1$. Notice that \mathcal{P}^* can be found by exploiting the theoretical relation between the MST and the OPT for f_{\max} [31]. The training essentially consists of finding \mathcal{P}^* and an OPF classifier rooted at \mathcal{P}^* .

By computing an MST in the complete graph (\mathcal{D}_1, A) [see Fig. 1(b)], one can obtain a connected acyclic graph whose nodes are all samples of \mathcal{D}_1 and the arcs are undirected and weighted by

the distances d between adjacent samples. The spanning tree is optimum in the sense that the sum of its arc weights is minimum as compared to any other spanning tree in the complete graph. In the MST, every pair of samples is connected by a single path, which is optimum according to f_{\max} . Hence, the MST contains one OPT for any selected root node.

The optimum prototypes are the closest elements of the MST with different labels in \mathcal{D}_1 [i.e., elements that fall in the frontier of the classes, as displayed in Fig. 1(c)]. By removing the arcs between different classes, their adjacent samples become prototypes in \mathcal{P}^* , and the OPF algorithm can define an OPF with minimum classification errors in \mathcal{D}_1 [see Fig. 1(d)].

2) *Classification*: For any sample $t \in \mathcal{D}_2$, one shall consider all arcs connecting t with samples $s \in \mathcal{D}_1$, as though t were part of the training graph [see Fig. 1(d)]. Considering all possible paths from \mathcal{P}^* to t , one finds the optimum path $F^*(t)$ from \mathcal{P}^* and label t with the class $\lambda(R(t))$ of its most strongly connected prototype $R(t) \in \mathcal{P}^*$. This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as

$$C(t) = \min\{\max\{C(s), d(s, t)\}\}, \forall s \in \mathcal{D}_1. \quad (2)$$

Let the node $s^* \in \mathcal{D}_1$ be the one that satisfies (2) (i.e., the predecessor $F(t)$ in the optimum path $F^*(t)$). Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ as the class of t [see Fig. 1(e)]. An error occurs when $L(s^*) \neq \lambda(t)$.

B. Bayesian Classifier

Although Naïve-Bayes is a well-known pattern recognition technique in the literature, we decided to leave some main background related to its formulation here. In this context, suppose the elements of the feature vector of each sample are conditionally independent of each other given the classification process. Roughly speaking, this means that given a certain class y_i , $i = 1, 2, \dots, c$, the values of different features do not affect others.

Let us consider sample x once again. Since Naïve-Bayes is a conditional-dependent model, one needs to compute the posterior distributions $P(y_i|\mathbf{x})$, $\forall i$, as follows:

$$P(y_i|\mathbf{x}) = \frac{P(y_i)P(\mathbf{x}|y_i)}{p(\mathbf{x})} \quad (3)$$

where $P(\omega_i)$ is the so-called prior probability. Since the denominator does not depend on ω_i , one can leave it behind to obtain a more compact representation of the posterior distribution as follows:

$$P(y_i|x) = P(y_i)P(\mathbf{x}|y_i). \quad (4)$$

By assuming the conditional independence of the features, one has a new formulation for the posterior distribution:

$$P(y_i|x^1, x^2, \dots, x^n) = P(y_i) \prod_{i=1}^n P(x^i|y_i) \quad (5)$$

where x^i stands for the i th feature of sample \mathbf{x} . Finally, Naïve-Bayes assigns the label y_j that maximizes the above equation. The term “naïve” comes from the fact we are assuming the features are independent of each other.

III. SEQUENTIAL LEARNING

As aforementioned in Section I, the main idea of sequential learning-based techniques is to model the contextual information by means of a two-step classification process, in which a traditional learning is conducted in the first phase, followed by the extension of the feature vectors by the predicted labels of a given sample’s neighborhood. After that, a new classification process is performed to refine the prior results. Roughly speaking, the methods that rely on such a context differ primarily with respect to the function that models each sample’s neighborhood.

Basically, the SSL [10] and the sliding window [32] are very similar to each other: while the former extends the feature vector of a given sample with the labels of its neighborhood, the latter approach employs the feature vectors of the neighboring samples for such purpose. The multiscale sequential learning proposed by Gatta *et al.* [11] considers two different roads to compute a given sample’s neighborhood: multiresolution- and pyramid-based decompositions. The first approach attempts at convolving the first-step-labeled image into a series of Gaussian kernels with different variances to simulate distinct scales;

thus, the feature vector is extended by concatenating the sample’s neighborhood at each scale. The pyramid decomposition also takes the Gaussian-convoluted images, but now the image is scaled down for further feature vector extension. The next sections describe in more details such procedures.

A. Standard SSL

The SSL is a metalearning procedure that has two basic steps: 1) a base classifier and 2) the classification of the extend features. Given a dataset $\mathcal{S} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \dots, (\mathbf{x}_m, y_m)\}$ composed of m pairs of samples (\mathbf{x}_i, y_i) , where $\mathbf{x}_i \in \mathbb{R}^n$ represents a feature vector and $y_i \in \{1, 2, \dots, c\}$ denotes the class of \mathbf{x}_i , the first step employs a base classifier f trained over a subset $\mathcal{S}_{Tr} \subset \mathcal{S}$, in which c stands for the number of classes. The next step extends the feature vector \mathbf{x}_i to a new vector $\tilde{\mathbf{x}}_i$ containing more features. The new feature vector $\tilde{\mathbf{x}}_i$ contains all features of \mathbf{x}_i , the output of the classifier $f(\mathbf{x}_i)$, and the output of the base classifier of the neighborhood samples \mathbf{x}_j , $\forall \mathbf{x}_j \in \Omega_i$, where Ω_i represents the neighborhood samples of \mathbf{x}_i , $j \neq i$. Thus, $\tilde{\mathbf{x}}_i = \mathbf{x}_i \cup \hat{y}_i \cup \hat{y}_{j_1} \cup \hat{y}_{j_2} \cup \dots \cup \hat{y}_{j_u}$, where u denotes the neighborhood size of \mathbf{x}_i , \hat{y}_i is the output of the classifier considering sample \mathbf{x}_i , i.e., $\hat{y}_i = f(\mathbf{x}_i)$, and \hat{y}_{j_u} is the classification output considering \mathbf{x}_{j_u} , i.e., $\hat{y}_{j_u} = f(\mathbf{x}_{j_u})$. The second step aims at training a new classifier $\tilde{f}(\cdot)$ using the extended feature vectors $\tilde{\mathbf{x}}_i$. The final output for the classification of \mathbf{x}_i is given by $\tilde{y}_i = \tilde{f}(\tilde{\mathbf{x}}_i)$. The first and second steps of the standard SSL paradigm are presented in Fig. 2.

B. Multiscale SSL

The multiresolution stacked sequential learning (MR-SSL) [11] is based on the multiresolution theory widely used in image-processing-oriented applications. The basic difference between SSL and MR-SSL is the neighborhood sampling process: MR-SSL samples feature vectors using different scales on the dataset \mathcal{S} .

Let $\rho^i(\mathbf{x}_j)$ be the probability of the sampled position \mathbf{x}_j belong to the class $i = \{1, 2, \dots, c\}$. Therefore, a multiresolution decomposition M_R^i is defined as follows:

$$M_R^i(\mathbf{x}_j, s) = \rho^i(\mathbf{x}_j) * G(0, \delta^{s-1}) \quad (6)$$

where $s = \{1, 2, \dots, S\}$ is the scale of the multiresolution decomposition, $*$ denotes the convolution operator, G is a multi-dimensional Gaussian distribution with zero mean and variance $\sigma = \delta^{s-1}$, and δ is the step of the multiscale decomposition.¹ After that, it is necessary to define the neighborhood sampling methodology, i.e., Ω . First, one should set the displacement vectors $\vec{d} = \{\vec{d}_1, \vec{d}_2, \dots, \vec{d}_9\}$ as follows:

$$\begin{aligned} \vec{d}_1 &= (-1, -1) & \vec{d}_2 &= (0, -1) & \vec{d}_3 &= (1, -1) \\ \vec{d}_4 &= (-1, 0) & \vec{d}_5 &= (0, 0) & \vec{d}_6 &= (1, 0) \\ \vec{d}_7 &= (1, -1) & \vec{d}_8 &= (0, 1) & \vec{d}_9 &= (1, 1). \end{aligned}$$

¹In hard classification techniques, i.e., the ones that output the label only, $\rho^i(\mathbf{x}_j) \in \{0, 1\}$.

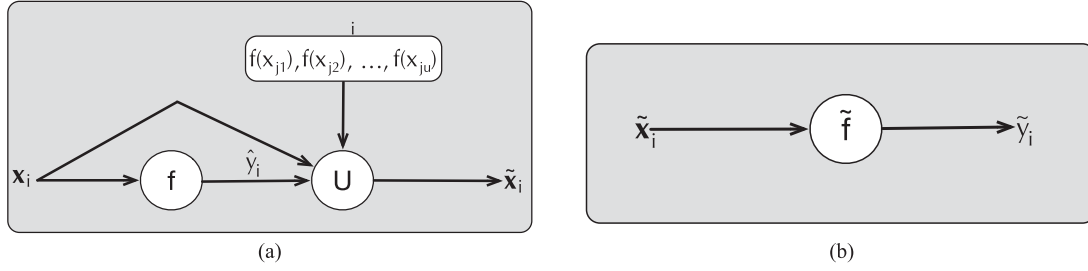


Fig. 2. Schematic diagram for the (a) first and (b) second steps of the aforementioned SSL paradigm.

In regard to MR-SSL, $\tilde{\mathbf{x}}_i$ is obtained as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_i = & \{ \mathbf{x}_i, M_R(\mathbf{x}_i + \vec{d}_1, 1), \dots, M_R(\mathbf{x}_i + \vec{d}_9, 1) \\ & \times M_R(\mathbf{x}_i + \delta \vec{d}_1, 2), \dots, M_R(\mathbf{x}_i + \delta \vec{d}_9, 2) \\ & \times \dots \times M_R(\mathbf{x}_i + \delta^{S-1} \vec{d}_1, S), \dots, \\ & \times M_R(\mathbf{x}_i + \delta^{S-1} \vec{d}_9, S) \}. \end{aligned} \quad (7)$$

Finally, $M_R(\mathbf{x}_i, s)$ is defined as follows:

$$M_R(\mathbf{x}_i, s) = \arg \max_j \{ M_R^j(\mathbf{x}_i, s) \}. \quad (8)$$

C. Pyramidal Stacked Sequential Learning

The pyramidal stacked sequential learning (PY-SSL) is a simple modification of the MR-SSL, and it differs in the sampling process only. The pyramidal decomposition can be obtained by

$$M_{PYR}^i(\mathbf{x}_j, s) = M_R^i(\lfloor k_s s \mathbf{x}_j \rfloor, s) \quad (9)$$

where k_s is the sampling step defined as $k_s = \delta^s / 2$.

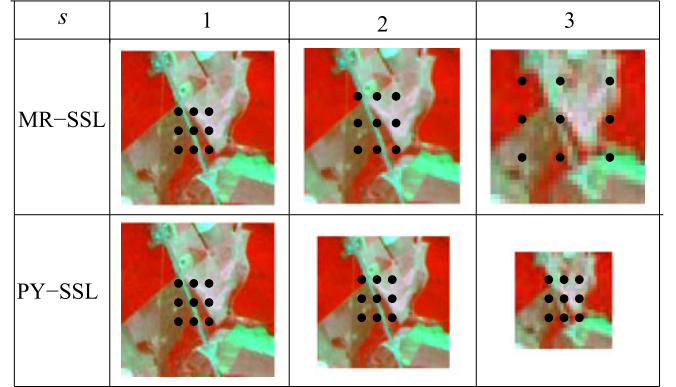
The PY-SSL approach to obtain the extended feature vector $\tilde{\mathbf{x}}_i$ from \mathbf{x}_i is given by

$$\begin{aligned} \tilde{\mathbf{x}}_i = & \{ \mathbf{x}_i, M_{PYR}(\mathbf{x}_i + \vec{d}_1, 1), \dots, M_{PYR}(\mathbf{x}_i + \vec{d}_9, 1) \\ & \times M_{PYR}(\lfloor \mathbf{x}_i / \delta \rfloor + \vec{d}_1, 2), \dots, M_{PYR}(\lfloor \mathbf{x}_i / \delta \rfloor + \vec{d}_9, 2) \\ & \times \dots \times M_{PYR}(\lfloor \mathbf{x}_i / \delta^{S-1} \rfloor + \vec{d}_1, S), \dots, \\ & \times M_{PYR}(\lfloor \mathbf{x}_i / \delta^{S-1} \rfloor + \vec{d}_9, S) \}. \end{aligned} \quad (10)$$

In order to clarify the main differences between MR-SSL and PY-SSL representations, Fig. 3 depicts the sampling process used when creating the neighborhood of a given sample for feature vector extension purposes. That pictorial representation was based on the work by Gatta *et al.* [11].

IV. PROPOSED ENSEMBLE-BASED SSL

In this section, two modifications of the standard SSL paradigm that can considerably improve the accuracy results are presented. The main difference between our approach and the classical SSL methods concerns the number of base classifiers. While the classical SSL uses a unique classifier in the first step, the proposed approach employs a variable number k of classifiers, each one trained on a partition \mathcal{S}_{Tr}^i of the training set \mathcal{S}_{Tr} , such that $\mathcal{S}_{Tr} = \mathcal{S}_{Tr}^1 \cup \mathcal{S}_{Tr}^2 \cup \dots \cup \mathcal{S}_{Tr}^k$. Based on such an assumption, two different approaches to extend the feature


 Fig. 3. Sampling the neighborhood by means of MR-SSL and PY-SSL approaches at different scales $s = \{1, 2, 3\}$.

vector \mathbf{x}_i are proposed: 1) concatenated and 2) voting. The motivation for such ensemble-based approaches comes from the following assumption: if the base classifier is prone to errors, it will degrade the learning process of the second one, since it propagates the wrong labels to the nearby samples. Consequently, such misclassified samples will corrupt the extended feature vector of a given sample, leading to even worse classification results. The next sections present more details about the proposed approaches.

A. Concatenated SSL

The concatenated stacked sequential learning (CN-SSL) extends the feature vector \mathbf{x}_i to a new vector $\tilde{\mathbf{x}}_i$ using the output of each classifier f_z for both \mathbf{x}_i and also for each sample that falls in its neighborhood Ω_i , $z \in \{1, 2, \dots, k\}$. Therefore, the idea is to use the output of each classifier f_z trained over the partition \mathcal{S}_{Tr}^z to augment the original feature vector \mathbf{x}_i from n features to $n + k + ku$ features, where n corresponds to the number of features extracted from \mathbf{x}_i , and u stands for its neighborhood size Ω_i . In short, we have that $\tilde{\mathbf{x}}_i = \mathbf{x}_i \cup \hat{y}_i^1 \cup \hat{y}_i^2 \cup \dots \cup \hat{y}_i^k \cup \hat{y}_{j_1}^1 \cup \hat{y}_{j_1}^2 \cup \dots \cup \hat{y}_{j_1}^k \cup \hat{y}_{j_2}^1 \cup \hat{y}_{j_2}^2 \cup \dots \cup \hat{y}_{j_2}^k \cup \dots \cup \hat{y}_{j_u}^1 \cup \hat{y}_{j_u}^2 \cup \dots \cup \hat{y}_{j_u}^k$, where $\hat{y}_i^z = f_z(\mathbf{x}_i)$, i.e., the output of classifier $f_z(\cdot)$ with respect to the sample \mathbf{x}_i . The next step (second classification) is performed as usual. Fig. 4 depicts the pipeline of CN-SSL.

The main idea of CN-SSL is to alleviate possible misclassifications by increasing the dimensionality of the feature vector \mathbf{x}_i . Suppose a situation in which a classifier f_z assigns the wrong

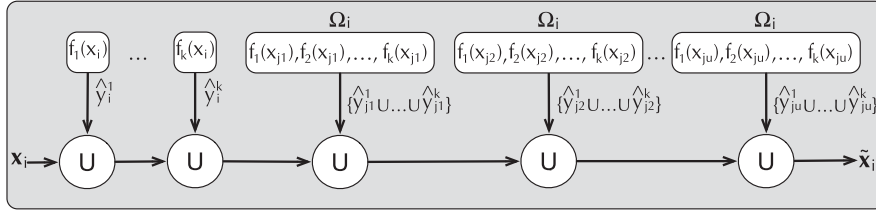


Fig. 4. Schematic diagram of CN-SSL.

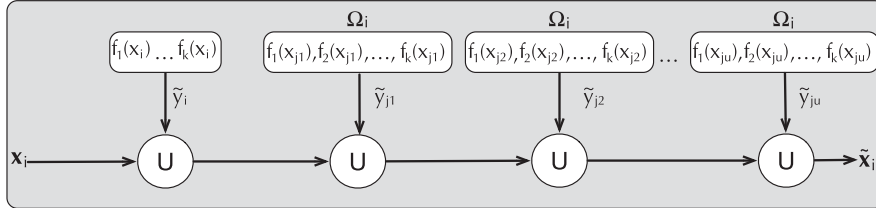


Fig. 5. Schematic diagram of VO-SSL.

label to some or even all neighbors of a sample \mathbf{x}_i , and the remaining classifiers f_v perform a correct classification, $\forall v \neq z$. Since the misclassified neighbors will take part of u dimensions only ($\tilde{\mathbf{x}}_i \in \mathbb{R}^{n+k+ku}$), they may not contribute a lot to corrupt $\tilde{\mathbf{x}}_i$.

B. Voting SSL

The voting stacked sequential learning (VO-SSL) extends the feature vector \mathbf{x}_i to a new vector $\tilde{\mathbf{x}}_i$ using the majority of the output considering the k base classifiers. In the case of a draw, the algorithm always returns the class with lower index. In this variant, the extended feature vector $\tilde{\mathbf{x}}_i$ has $n + 1 + u$ dimensions (features), as follows: $\tilde{\mathbf{x}}_i = \mathbf{x}_i \cup \tilde{y}_i \cup \tilde{y}_{j1} \cup \tilde{y}_{j2} \cup \dots \cup \tilde{y}_{ju}$, where \tilde{y}_i stands for the majority voting considering the classifiers in $\{\hat{y}_i^1, \hat{y}_i^2, \dots, \hat{y}_i^k\}$. Fig. 5 depicts the working mechanism of VO-SSL.

C. Variations

In order to evaluate the robustness of the proposed approaches, they were combined with standard SSL, MR-SSL, and PY-SSL, thus resulting in six new methods, as detailed below:

- 1) CN-SSL: Concatenated SSL.
- 2) VO-SSL: Voting SSL.
- 3) CN-MR-SSL: CN-SSL applied to MR-SSL.
- 4) VO-MR-SSL: VO-SSL applied to MR-SSL.
- 5) CN-PY-SSL: CN-SSL applied to PY-SSL.
- 6) VO-PY-SSL: VO-SSL applied to PY-SSL.

The rationale in combining the proposed approaches with others is to show that one can enhance them by simply adding the ensemble-based tool into their working mechanism.

V. METHODOLOGY

In this section, the methodology employed to validate the proposed approaches in the context of land-cover classification is presented. In regard to the experiments, images obtained from CBERS-2B and Landsat 5 TM covering the area of Itatinga, SP,

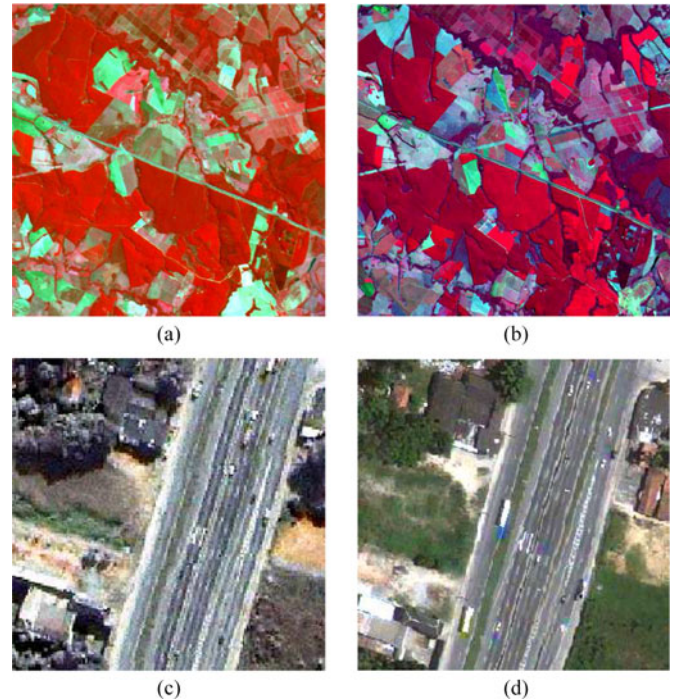


Fig. 6. Satellite images used in the experiments: covering the area of Itatinga, SP, Brazil by (a) CBERS-2B CCD (20 m) sensor (R2G3B4) and (b) Landsat 5 TM (30 m) sensor (R4G3B5), and covering the area of Duque de Caxias, RJ, Brazil by (c) Ikonos-2 MS sensor (R4G3B2) and (d) Geosyde sensor (R5G4B3). The CBERS-2B and Landsat 5 TM images have 526×492 pixels, and Ikonos-2 MS and Geosyde images have 258×250 and 268×250 pixels, respectively. Notice that Ikonos-2 MS and Geosyde images were obtained through a fusion process between the corresponding images from MS (4 m) and PAN (1 m) sensors using the pan-sharpening method. The final image has a spatial resolution of 1 m.

Brazil, and other images are obtained from Ikonos-2 MS and Geosyde covering the area of Duque de Caxias, RJ, Brazil [29]. Fig. 6 displays these images, being their respective ground truth versions illustrated in Fig. 7. Additionally, Table I presents the description of the land-cover classes for each image, and Tables II and III present the number of samples per

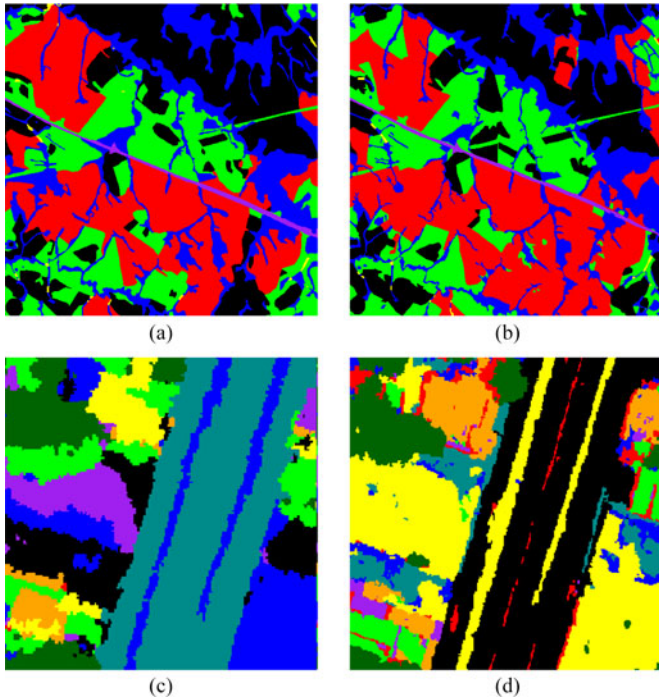


Fig. 7. Labeled images used in the experiments: (a) and (b) refer to the images displayed in Fig. 6(a) and (b), respectively, and (c) and (d) stand for images displayed in Fig. 6(c) and (d), respectively.

TABLE I
LAND-COVER DESCRIPTION FOR EACH COVERING AREA

Covering area	Sensor	# land-cover classes	Land use classes
Itatinga	CBERS-2B	6	grass lands, reforestation, cultures, roads, dams and bushes
Itatinga	Landsat 5 TM	6	grass lands, reforestation, cultures, roads, dams and bushes
Duque de Caxias	Ikonos-2 MS	8	roads, grass lands, bare soil moist, covering of tree, covering of clear tonality, covering of dark tonality, bare soil clear and shadows
Duque de Caxias	Geoeye	9	roads, grass lands, bare soil moist, covering of tree, covering of clear tonality, covering of average tonality, covering of dark tonality, bare soil clear and shadows

TABLE II
NUMBER OF SAMPLES FOR THE COVERING AREA OF ITATINGA

Label	CBERS-2B	Landsat 5 TM
culture	18 407	13 356
bushes	13 779	12 387
dams	142	123
grass lands	14 794	17 567
reforestation	18 825	22 581
roads	803	736

TABLE III
NUMBER OF SAMPLES FOR THE COVERING AREA OF DUQUE DE CAXIAS

Label	Ikonos-2 MS	Geoeye
covering of trees	5914	6132
shadows	6481	2822
grass lands	12 054	19 370
covering of dark tonality	3578	5073
roads	22 871	22 924
bare soil moist	4417	2380
bare soil clear	7400	4490
covering of clear tonality	1785	1026
covering of average tonality	-	2783

TABLE IV
COLORS ASSOCIATED WITH EACH LAND-COVER CLASS

Sensor	Color	Land cover classes
CBERS-2B and Landsat 5	Green	Grass lands
	Blue	Bushes
	Black	Cultures
	Red	Reforestation
	Yellow	Dams
	Purple	Roads
Ikonos-2 MS	Green	Shadows
	Blue	Grass lands
	Black	Bare soil clear
	Red	Covering of average tonality
	Yellow	Covering of dark tonality
	Purple	Bare soil moist
	Green	Covering of tree
	Orange	Covering of clear tonality
Geoeye	Green	Covering of average tonality
	Blue	Bare soil moist
	Black	Roads
	Red	Shadows
	Yellow	Grass lands
	Purple	Covering of clear tonality
	Green	Covering of tree
	Orange	Covering of dark tonality
	Teal	Bare soil clear

land-cover class.² Finally, Table IV presents the color map used to obtain the ground-truth images displayed in Fig. 7.

The standard OPF³ classifier was compared against nine different sequential learning approaches: OPF with standard stacked sequential learning [10] (OPF-SSL), OPF with multi-scale sequential learning and multiresolution-based decomposition (OPF-MR-SSL) [11], OPF with concatenated SSL (OPF-CN-SSL) and voting SSL (OPF-VO-SSL), OPF with multi-scale sequential learning and pyramid-based decomposition (OPF-PY-SSL) [11], OPF-MR-SSL combined with concatenated SSL (OPF-CN-MR-SSL) and voting SSL (OPF-VO-MR-SSL), and OPF-PY-SSL combined with concatenated SSL (OPF-CN-PY-

²Images are available at http://www.fc.unesp.br/~papa/reconga/remote_sensing.html

³We employed the LibOPF [33], which is an open-source library to the design of OPF-based classifiers.

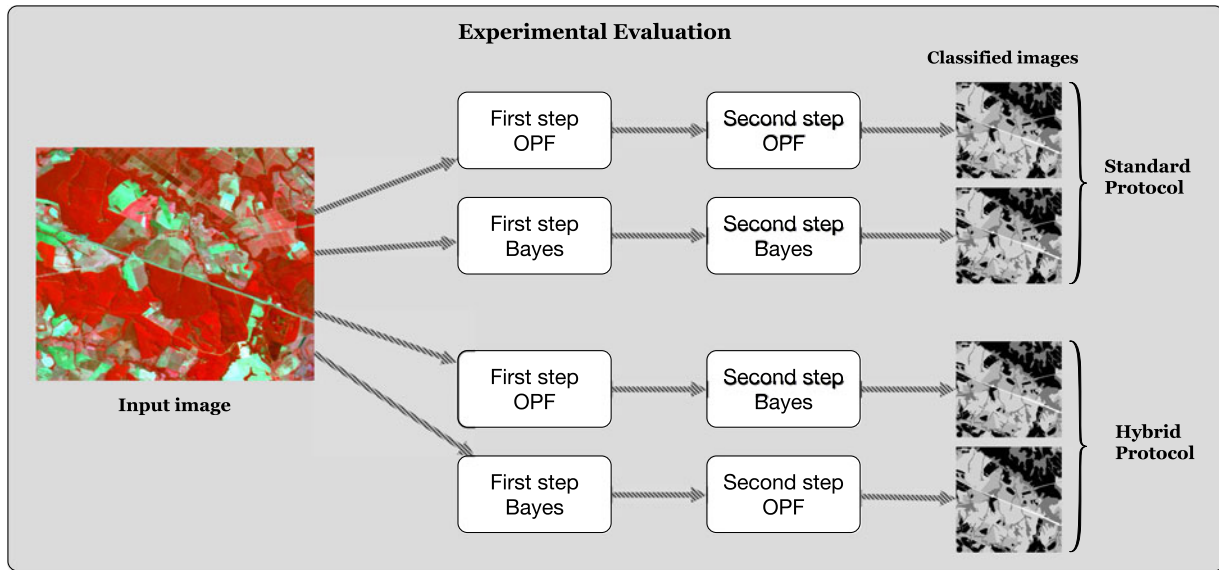


Fig. 8. Experimental methodology adopted in the work.

TABLE V
EXPERIMENTAL RESULTS REGARDING CBERS-2B IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF		66.8 ± 2.1			65.8 ± 2.2			65.2 ± 2.8	
OPF-SSL		71.9 ± 0.6			71.6 ± 0.4			72.2 ± 0.4	
OPF-VO-SSL	68.3 ± 0.1	68.7 ± 0.0	68.8 ± 0.0	67.9 ± 0.0	68.0 ± 0.0	67.9 ± 0.0	67.3 ± 0.1	66.9 ± 0.0	66.9 ± 0.0
OPF-CN-SSL	73.5 ± 0.0	74.3 ± 0.0	74.6 ± 0.0	74.0 ± 0.0	74.7 ± 0.0	74.9 ± 0.0	74.3 ± 0.0	75.0 ± 0.0	75.5 ± 0.0
OPF-MR-SSL		70.6 ± 0.5			70.2 ± 0.4			70.5 ± 0.6	
OPF-VO-MR-SSL	73.0 ± 0.0	71.2 ± 0.0	71.4 ± 0.0	71.1 ± 0.1	70.6 ± 0.0	70.7 ± 0.0	72.4 ± 0.0	70.8 ± 0.0	70.8 ± 0.0
OPF-CN-MR-SSL	72.6 ± 0.0	72.8 ± 0.0	73.0 ± 0.0	73.2 ± 0.0	73.3 ± 0.0	73.7 ± 0.0	73.5 ± 0.0	74.0 ± 0.0	74.2 ± 0.1
OPF-PY-SSL		68.7 ± 0.5			68.9 ± 0.4			69.5 ± 0.4	
OPF-VO-PY-SSL	69.1 ± 0.3	68.6 ± 0.3	68.8 ± 0.5	68.8 ± 0.5	67.6 ± 0.2	67.7 ± 0.5	66.6 ± 0.0	66.8 ± 0.2	67.4 ± 0.0
OPF-CN-PY-SSL	72.8 ± 0.1	73.6 ± 0.1	74.2 ± 0.2	73.6 ± 0.1	74.1 ± 0.2	74.6 ± 0.3	74.1 ± 0.1	74.6 ± 0.2	75.1 ± 0.1
Bayes		72.0 ± 0.0			72.2 ± 0.1			71.0 ± 0.0	
Bayes-SSL		66.5 ± 0.0			66.9 ± 0.0			68.0 ± 0.0	
Bayes-VO-SSL	72.5 ± 0.0	72.4 ± 0.0	72.2 ± 0.0	72.5 ± 0.0	72.5 ± 0.0	72.5 ± 0.0	72.5 ± 0.0	72.7 ± 0.0	72.7 ± 0.0
Bayes-CN-SSL	74.8 ± 0.1	75.2 ± 0.0	75.3 ± 0.1	75.4 ± 0.1	75.6 ± 0.1	75.8 ± 0.1	75.9 ± 0.1	76.2 ± 0.1	76.4 ± 0.1
Bayes-MR-SSL		64.7 ± 0.0			66.9 ± 0.1			68.0 ± 0.0	
Bayes-VO-MR-SSL	73.0 ± 0.0	72.9 ± 0.0	72.9 ± 0.0	73.1 ± 0.0	73.1 ± 0.0	73.0 ± 0.0	73.4 ± 0.0	73.6 ± 0.0	73.6 ± 0.0
Bayes-CN-MR-SSL	73.5 ± 0.0	73.7 ± 0.0	73.5 ± 0.0	74.1 ± 0.0	74.2 ± 0.0	74.4 ± 0.0	75.0 ± 0.0	75.1 ± 0.0	75.3 ± 0.0
Bayes-PY-SSL		64.9 ± 0.1			66.8 ± 0.1			68.3 ± 0.2	
Bayes-VO-PY-SSL	72.8 ± 0.1	72.6 ± 0.2	72.7 ± 0.0	72.6 ± 0.1	72.7 ± 0.0	72.7 ± 0.0	72.9 ± 0.2	73.0 ± 0.0	72.7 ± 0.0
Bayes-CN-PY-SSL	74.3 ± 0.1	74.7 ± 0.1	74.9 ± 0.2	75.5 ± 0.0	75.2 ± 0.2	75.4 ± 0.0	75.5 ± 0.0	75.9 ± 0.2	76.6 ± 0.1

SSL) and voting SSL (OPF-VO-PY-SSL). In addition, we implemented the very same proposed approaches considering naïve Bayes classifier in order to show that the proposed approaches can also obtain very good results when applied to other techniques. Following the above nomenclature, one has Bayes (standard Bayesian classifier), Bayes-SSL (Bayesian classifier with standard SSL), Bayes-MR-SSL (Bayes with multi-scale sequential learning and multi-resolution-based decomposition), Bayes with concatenated SSL (Bayes-CN-SSL) and voting SSL (Bayes-VO-SSL), Bayes with multi-scale sequential learning and pyramid-based decomposition (Bayes-PY-SSL) [11], Bayes-MR-SSL combined with concatenated SSL (Bayes-CN-

MR-SSL) and voting SSL (Bayes-VO-MR-SSL), and Bayes-PY-SSL combined with concatenated SSL (Bayes-CN-PY-SSL) and voting SSL (Bayes-VO-PY-SSL). Although one can use any pattern recognition technique, the proposed approach was validated over OPF and naïve Bayes classifier, since both techniques are parameterless and do not require a considerable computational load.⁴ Additionally, the main motivation for ensemble-based SSL is related to OPF classifier, since such graph-based approach has been consistently more accurate than SVM in several applications, but being faster for training patterns. An

⁴In regard to Bayesian classifier, our own implementation has been employed.

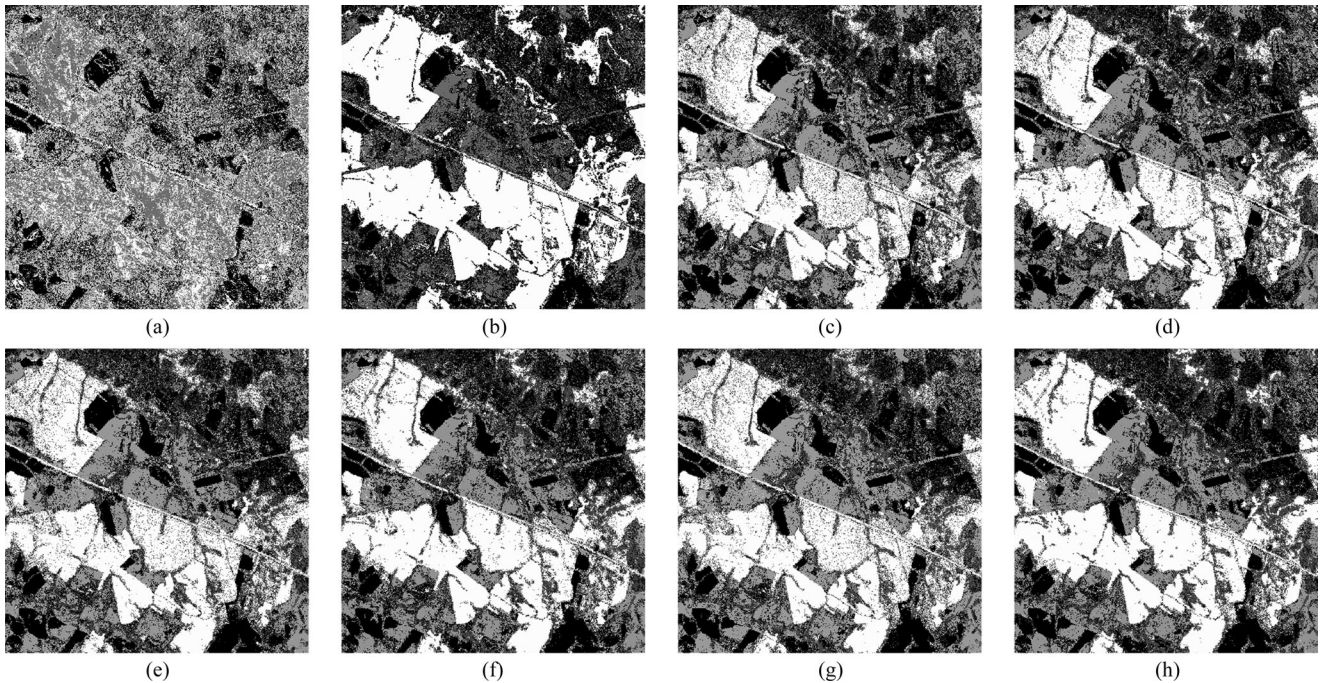


Fig. 9. CBERS-2B images classified using: (a) OPF (10%), (b) Bayes (10%), (c) OPF-CN-SSL (5% and seven base classifiers), (d) Bayes-CN-SSL (5% and seven base classifiers), (e) OPF-CN-SSL (10% and seven base classifiers), (f) Bayes-CN-SSL (10% and seven base classifiers), (g) OPF-CN-SSL (20% and seven base classifiers), and (h) Bayes-CN-SSL (20% and seven base classifiers).

extra round of experiment was conducted in order to check the robustness of the proposed approaches when one uses different classifiers for each step (this last experiment is called “Hybrid”). Fig. 8 displays the experimental evaluation procedure adopted in this paper.

The influence of different training set sizes with 5%, 10%, and 20% of the entire image was evaluated, being the remaining pixels used to compose the test set. Additionally, the influence in using a different number of base classifiers was also considered. For such purpose, we use $k \in \{3, 5, 7\}$ classifiers for each the training set sizes. In order to allow a robust statistical evaluation, a cross-validation with ten runnings for the further computation of the Wilcoxon signed-rank test [34] over the accuracy rates was performed.⁵ Additionally, each pixel has been described by its RGB values to compose the dataset samples.⁶ Finally, the experiments were conducted on a personal computer equipped with an Intel Xeon CPU E5-2603 1.60-GHz processor, 16 GB of RAM and Ubuntu 14.04 LTS as the operational system.

VI. EXPERIMENTS

In this section, the experimental results regarding sequential learning-based OPF classification using nine different approaches were presented, including the six new variations considered in this paper.

⁵We employed an accuracy measure proposed by Papa *et al.* [25] that considers unbalanced datasets, which is often faced in land-cover classification.

⁶In this work, we used an eight-neighborhood system for OPF-SSL, and an 11- and a three-neighborhood systems for OPF-MSSL-MR and OPF-MSSL-PY, respectively. We also employed seven scales of decomposition for OPF-MSSL-MR, and five scales of decomposition for OPF-MSSL-PY. Such values have been empirically set.

A. CBERS-2B Image

Table V presents the mean accuracy results with respect to the CBERS-2B image displayed in Fig. 6(a). The most accurate techniques considering the Wilcoxon signed-rank test are in bold.⁷ Considering such results, one can draw some interesting conclusions.

- 1) Standard OPF accuracy does not improve when one increases the training set size, which means there is no guarantee we are always adding good samples for training purposes (Bayes performance gets better slightly with 10% for training, but it drops again when one uses 20%).
- 2) Both OPF and Bayes techniques do not have substantially more accurate results using SSL when one increases the training set size, which corroborates our assumption that misclassified samples do not help sequential learning techniques, thus making them even worse.
- 3) The configuration with seven base classifiers worked better for almost all pairs of classifiers and training set sizes, which makes sense since we use more specialists to compose the ensemble, and thus, more specialists are considered into the decision-making process (obviously, there is a tradeoff between the number of classifiers to compose the ensemble and the training set size assigned to each of them).
- 4) Both the proposed approaches, i.e., CN-SSL and VO-SSL, improved the base techniques where they have been

⁷The statistical test is performed for each classifier and amount of training set (e.g., OPF-CN-SSL with seven base classifiers obtained the best results considering OPF and using 5% of the data for training purposes).

TABLE VI
EXPERIMENTAL RESULTS REGARDING THE “HYBRID” EXPERIMENT OVER CBERS-2B IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
First step classification using OPF and the second step using Bayes									
SSL		70.1 ± 0.6			70.9 ± 0.9			70.1 ± 0.3	
VO-SSL	69.7 ± 0.6	71.1 ± 0.0	69.8 ± 0.0	69.0 ± 1.1	68.2 ± 0.0	69.0 ± 0.2	67.7 ± 1.0	66.9 ± 0.4	66.7 ± 0.0
CN-SSL	70.2 ± 0.9	70.0 ± 0.0	68.7 ± 0.0	69.1 ± 1.1	68.2 ± 0.0	70.1 ± 1.0	69.1 ± 0.7	68.5 ± 0.6	66.8 ± 0.0
MR		72.6 ± 1.0			74.6 ± 1.1			74.3 ± 1.2	
VO-MR	72.1 ± 0.5	72.7 ± 0.0	70.8 ± 0.0	73.3 ± 1.2	68.2 ± 0.0	68.0 ± 0.0	74.1 ± 1.0	72.8 ± 1.0	69.4 ± 0.0
CN-MR	69.3 ± 0.7	72.4 ± 0.0	71.4 ± 0.0	71.2 ± 1.1	75.2 ± 0.0	75.2 ± 0.0	72.4 ± 0.9	76.1 ± 0.5	76.5 ± 0.2
PYR		70.2 ± 1.0			71.3 ± 0.6			71.2 ± 0.4	
VO-PYR	70.2 ± 0.9	70.0 ± 0.0	69.8 ± 0.0	69.1 ± 1.1	68.2 ± 0.0	68.3 ± 0.0	69.1 ± 0.7	68.6 ± 0.6	67.9 ± 0.0
CN-PYR	73.5 ± 0.8	75.4 ± 0.0	69.7 ± 0.0	74.7 ± 1.0	68.2 ± 0.0	67.0 ± 0.0	75.8 ± 0.6	76.7 ± 0.6	71.4 ± 0.2
First step classification using Bayes and the second step using OPF									
SSL		68.9 ± 0.3			69.5 ± 0.1			64.0 ± 0.4	
VO-SSL	63.8 ± 0.7	63.4 ± 0.0	64.1 ± 0.2	61.8 ± 0.4	62.5 ± 0.0	62.4 ± 0.1	60.9 ± 0.7	60.8 ± 0.6	60.9 ± 0.0
CN-SSL	65.4 ± 1.0	66.4 ± 0.0	65.1 ± 0.2	64.4 ± 0.5	62.5 ± 0.0	66.1 ± 0.0	65.0 ± 0.3	65.3 ± 0.4	63.9 ± 0.2
MR		68.1 ± 0.3			68.9 ± 0.1			68.4 ± 0.2	
VO-MR	65.3 ± 1.0	65.1 ± 0.0	65.0 ± 0.0	64.7 ± 0.4	62.5 ± 0.0	62.0 ± 0.3	63.5 ± 0.8	64.3 ± 0.6	62.1 ± 0.1
CN-MR	67.7 ± 0.3	67.9 ± 0.0	66.0 ± 0.0	68.3 ± 0.1	68.5 ± 0.0	68.7 ± 0.0	69.0 ± 0.2	69.1 ± 0.2	67.7 ± 0.0
PYR		68.2 ± 0.3			69.3 ± 0.2			64.5 ± 0.4	
VO-PYR	64.3 ± 0.8	63.7 ± 0.0	64.9 ± 0.0	62.0 ± 1.1	62.5 ± 0.0	62.9 ± 0.1	61.4 ± 0.6	61.5 ± 0.4	60.7 ± 0.0
CN-PYR	66.0 ± 0.6	66.2 ± 0.0	65.0 ± 0.0	64.8 ± 0.4	62.5 ± 0.0	66.1 ± 1.1	65.0 ± 0.2	65.2 ± 0.2	63.8 ± 0.1

TABLE VII
MEAN COMPUTATIONAL LOAD IN SECONDS REGARDING CBERS-2B IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF-SSL		5.3			11.3			29.3	
OPF-VO-SSL	3.4	3.5	3.4	6.4	6.2	7.2	11.0	11.7	9.6
OPF-CN-SSL	11.9	12.0	11.4	28.6	26.3	26.2	52.1	52.5	49.7
OPF-MR		40.7			76.8			170.0	
OPF-VO-MR	34.0	34.4	31.4	71.9	66.3	78.2	129.5	132.1	119.7
OPF-CN-MR	128.0	201.0	206.2	236.9	368.2	401.1	450.7	646.0	649.3
OPF-PYR		20.7			47.1			104.8	
OPF-VO-PYR	11.8	12.4	13.4	26.7	26.3	26.3	51.8	52.6	49.7
OPF-CN-PYR	47.1	48.5	43.4	100.1	106.3	101.4	219.8	208.5	219.6
Bayes-SSL		6.6			11.1			18.0	
Bayes-VO-SSL	3.4	3.4	3.3	5.6	5.8	6.8	8.8	9.2	9.0
Bayes-CN-SSL	5.9	6.2	6.3	10.7	12.8	11.5	19.1	23.3	19.0
Bayes-MR		10.3			20.9			47.1	
Bayes-VO-MR	10.8	11.0	8.3	20.2	25.8	22.0	43.3	45.8	49.0
Bayes-CN-MR	12.5	13.2	13.3	38.2	38.8	32.3	95.3	96.2	99.0
Bayes-PYR		11.2			43.3			101.5	
Bayes-VO-PYR	9.3	9.3	7.3	17.1	15.8	16.4	33.5	35.9	39.0
Bayes-CN-PYR	22.0	21.2	27.3	56.2	58.1	64.3	101.4	108.2	108.1

applied (only one exception was noticed with respect to OPF-VO-SSL and OPF-SSL for all training set sizes, in which the former did not outperform the base approach using SSL).

- 5) The number of base classifiers seemed to have no good influence over VO-SSL, since it has obtained the best results using less classifiers (i.e., 3), being an intuitive idea about that the fact of having a committee composed of not so high-skilled specialists;
- 6) CN-SSL worked better than VO-SSL, which also corroborates our assumption stated in Section IV-A that says one

can somehow mask the effect of misclassified samples increasing the dimensionality of the feature space.

Fig. 9 displays some of the best results obtained over CBERS-2B image, in which the percentage in parenthesis stands for the amount of training set used. Usually, the results using MR-SSL and PY-SSL are more accurate than standard SSL, as observed by Pereira *et al.* [28]. However, due to the computational load, we used less scales and smaller neighborhoods, which might have affected the results of both MR and PY with respect to SSL. Albeit, the main goal of this work is not related to the fact that MR-SSL and PY-SSL are better than naïve SSL, since

TABLE VIII
EXPERIMENTAL RESULTS CONCERNING IKONOS-2 IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN
BASE CLASSIFIERS

Classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF		69.2 ± 0.1			71.3 ± 0.2			74.3 ± 0.1	
OPF-SSL		68.5 ± 0.4			69.9 ± 0.3			72.8 ± 0.3	
OPF-VO-SSL	69.0 ± 0.1	68.4 ± 0.2	68.0 ± 0.4	70.0 ± 0.2	70.0 ± 0.4	70.2 ± 0.4	73.8 ± 0.3	73.3 ± 0.3	73.7 ± 0.4
OPF-CN-SSL	68.1 ± 0.2	68.0 ± 0.2	68.4 ± 0.0	69.3 ± 0.1	69.1 ± 0.1	69.8 ± 0.2	72.0 ± 0.1	71.4 ± 0.2	71.4 ± 0.2
OPF-MR-SSL		62.4 ± 0.3			64.2 ± 0.2			68.5 ± 0.1	
OPF-VO-MR-SSL	69.8 ± 0.2	69.5 ± 0.3	69.3 ± 0.2	71.6 ± 0.2	71.7 ± 0.2	72.5 ± 0.2	77.0 ± 0.1	77.2 ± 0.1	77.1 ± 0.1
OPF-CN-MR-SSL	69.7 ± 0.1	69.6 ± 0.1	69.6 ± 0.1	72.0 ± 0.2	72.3 ± 0.2	72.2 ± 0.2	76.9 ± 0.2	77.1 ± 0.2	77.0 ± 0.2
OPF-PY-SSL		62.4 ± 0.3			63.0 ± 0.0			67.2 ± 0.4	
OPF-VO-PY-SSL	69.1 ± 0.2	68.2 ± 0.5	69.5 ± 0.3	71.0 ± 0.2	70.9 ± 0.4	70.5 ± 0.2	75.0 ± 0.5	74.0 ± 0.2	73.8 ± 0.4
OPF-CN-PY-SSL	68.5 ± 0.2	68.5 ± 0.2	69.0 ± 0.3	70.3 ± 0.2	70.5 ± 0.3	70.4 ± 0.2	73.1 ± 0.2	73.1 ± 0.2	73.7 ± 0.2
Bayes		69.0 ± 0.9			70.1 ± 0.1			73.8 ± 0.1	
Bayes-SSL		67.2 ± 0.2			65.2 ± 0.1			70.1 ± 0.2	
Bayes-VO-SSL	67.8 ± 0.1	67.7 ± 0.2	67.9 ± 0.2	69.8 ± 0.1	69.8 ± 0.1	69.7 ± 0.2	73.5 ± 0.2	73.3 ± 0.2	73.2 ± 0.1
Bayes-CN-SSL	68.0 ± 0.2	67.9 ± 0.1	68.2 ± 0.1	69.5 ± 0.1	69.3 ± 0.2	69.1 ± 0.1	71.6 ± 0.1	71.2 ± 0.1	71.2 ± 0.1
Bayes-MR-SSL		60.1 ± 0.9			64.2 ± 0.2			68.0 ± 0.0	
Bayes-VO-MR-SSL	69.6 ± 0.3	69.5 ± 0.2	69.7 ± 0.2	72.5 ± 0.2	72.3 ± 0.1	72.8 ± 0.1	77.3 ± 0.2	77.3 ± 0.3	77.4 ± 0.3
Bayes-CN-MR-SSL	69.6 ± 0.2	69.8 ± 0.1	69.7 ± 0.2	72.6 ± 0.1	72.7 ± 0.0	72.7 ± 0.1	76.9 ± 0.0	77.6 ± 0.1	77.1 ± 0.4
Bayes-PY-SSL		61.0 ± 0.1			63.9 ± 0.0			69.1 ± 0.0	
Bayes-VO-PY-SSL	68.5 ± 0.2	68.0 ± 0.1	68.3 ± 0.1	70.3 ± 0.1	70.0 ± 0.1	70.4 ± 0.1	74.0 ± 0.2	74.6 ± 0.2	74.0 ± 0.2
Bayes-CN-PY-SSL	68.3 ± 0.0	69.0 ± 0.3	69.6 ± 0.2	70.3 ± 0.1	70.4 ± 0.1	70.4 ± 0.1	73.6 ± 0.1	73.6 ± 0.1	73.8 ± 0.1

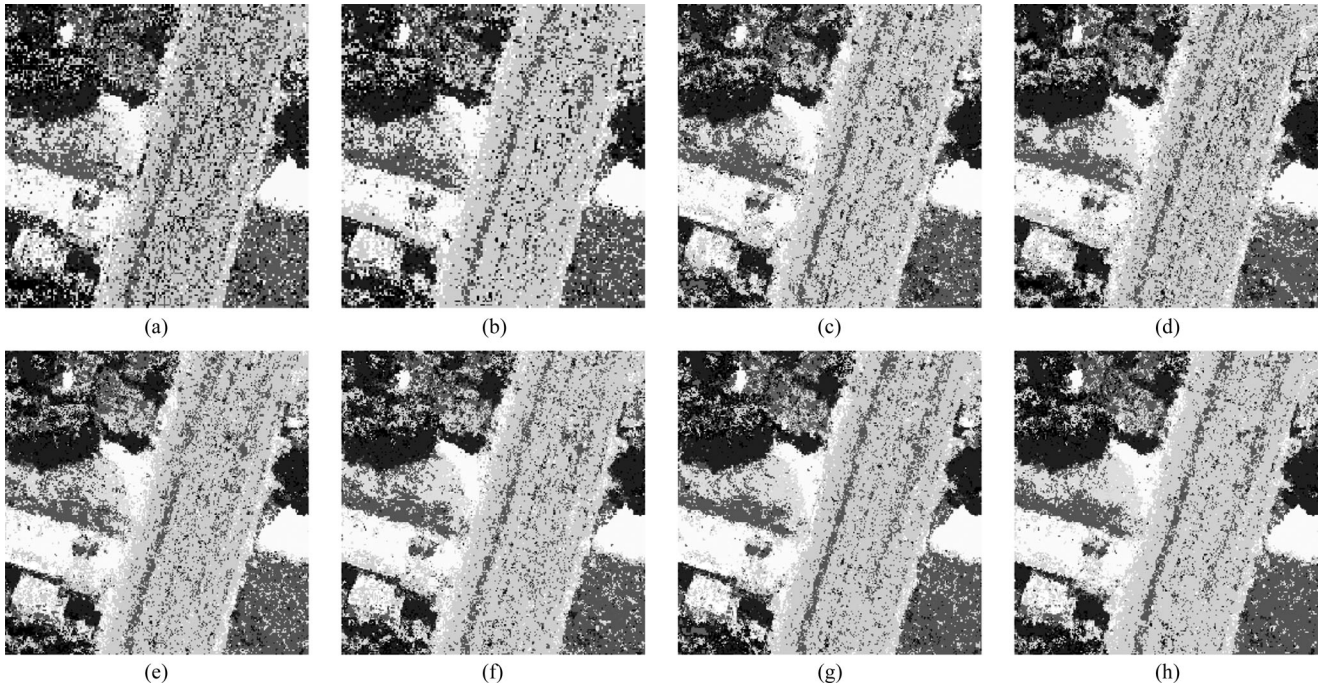


Fig. 10. Ikonos-2 MS images classified using: (a) OPF (10%), (b) Bayes (10%), (c) OPF-CN-SSL (5% and seven base classifiers), (d) Bayes-CN-SSL (5% and seven base classifiers), (e) OPF-CN-SSL (10% and seven base classifiers), (f) Bayes-CN-SSL (10% and seven base classifiers), (g) OPF-CN-SSL (20% and seven base classifiers), and (h) Bayes-CN-SSL (20% and seven base classifiers).

such points have been extensively discussed by Gatta *et al.* [11], but showing one can improve the aforementioned techniques using a committee of classifiers to alleviate the problem of misclassification.

Table VI presents the results with respect to the “Hybrid” experiment. From such data, one can draw some conclusions: 1) the results using OPF as the first classifier were considerably

better than the ones using Naïve-Bayes in the first step, which suggests that OPF achieves better initial results that are propagated to the next stacked classifier for feature extension; and 2) when using the OPF as the initial learner, the hybrid protocol obtained better results (for some situations) than using the standard protocol presented in Table V. Such results strengthen the validity of the proposed approaches.

TABLE IX
EXPERIMENTAL RESULTS REGARDING THE “HYBRID” EXPERIMENT OVER IKONOS-2 IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
First step classification using OPF and the second step using Bayes									
SSL		66.6 ± 0.5			71.1 ± 0.6			72.0 ± 1.1	
VO-SSL	67.0 ± 0.3	66.5 ± 0.0	67.2 ± 0.0	69.4 ± 0.7	69.4 ± 0.0	70.1 ± 0.0	73.7 ± 0.8	73.6 ± 0.6	74.4 ± 0.0
CN-SSL	66.8 ± 0.2	68.6 ± 0.0	68.1 ± 0.0	69.4 ± 0.1	69.4 ± 0.0	71.5 ± 0.1	73.3 ± 0.5	73.3 ± 0.2	74.4 ± 0.1
MR		69.4 ± 0.4			69.8 ± 0.3			70.1 ± 0.2	
VO-MR	62.2 ± 0.6	61.8 ± 0.0	64.2 ± 0.0	65.2 ± 0.3	69.4 ± 0.0	67.2 ± 0.0	68.8 ± 0.4	68.5 ± 0.2	66.4 ± 0.0
CN-MR	71.5 ± 0.1	72.1 ± 0.0	73.0 ± 0.0	76.4 ± 0.4	78.4 ± 0.0	78.0 ± 0.0	81.2 ± 0.4	81.2 ± 0.2	81.1 ± 0.0
PYR		69.8 ± 0.6			75.2 ± 0.8			73.2 ± 0.4	
VO-PYR	68.1 ± 0.5	68.6 ± 0.0	67.2 ± 0.0	70.4 ± 0.8	69.4 ± 0.0	71.2 ± 0.1	73.8 ± 0.4	74.2 ± 0.6	74.4 ± 0.0
CN-PYR	71.2 ± 0.6	73.1 ± 0.0	70.2 ± 0.0	74.8 ± 0.1	69.4 ± 0.0	72.5 ± 0.0	80.1 ± 0.4	81.6 ± 0.5	78.1 ± 0.3
First step classification using Bayes and the second step using OPF									
SSL		62.8 ± 0.3			66.6 ± 0.2			69.4 ± 0.5	
VO-SSL	62.5 ± 0.6	62.3 ± 0.0	62.4 ± 0.2	64.9 ± 0.8	65.5 ± 0.0	66.4 ± 0.0	68.9 ± 0.7	68.2 ± 0.2	68.0 ± 0.0
CN-SSL	62.9 ± 0.5	62.2 ± 0.0	61.8 ± 0.1	63.9 ± 0.2	65.5 ± 0.0	67.0 ± 0.0	66.5 ± 0.4	65.9 ± 0.1	66.1 ± 0.2
MR		63.9 ± 0.1			71.6 ± 0.5			71.2 ± 0.3	
VO-MR	64.0 ± 0.4	63.7 ± 0.0	62.2 ± 0.0	66.8 ± 0.4	65.47 ± 0.0	66.8 ± 0.0	71.6 ± 0.4	70.9 ± 0.6	69.1 ± 0.0
CN-MR	63.6 ± 0.2	63.6 ± 0.0	61.4 ± 0.0	66.7 ± 0.3	65.5 ± 0.0	69.2 ± 0.1	71.3 ± 0.2	70.9 ± 0.2	69.8 ± 0.5
PYR		63.3 ± 0.3			71.1 ± 0.3			71.3 ± 0.4	
VO-PYR	63.0 ± 0.6	62.6 ± 0.0	61.8 ± 0.2	65.4 ± 0.4	65.5 ± 0.0	67.0 ± 0.1	69.2 ± 0.8	69.2 ± 0.4	68.0 ± 0.0
CN-PYR	63.7 ± 0.1	64.0 ± 0.0	62.3 ± 0.1	66.5 ± 0.4	65.5 ± 0.0	66.7 ± 0.3	71.1 ± 0.2	71.2 ± 0.4	70.4 ± 0.2

TABLE X
MEAN COMPUTATIONAL LOAD IN SECONDS CONCERNING IKONOS-2 IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF-SSL		0.3			1.1			3.2	
OPF-VO-SSL	0.3	0.3	0.3	0.6	0.6	0.8	1.1	1.1	1.1
OPF-CN-SSL	0.8	0.5	0.4	1.5	1.2	0.1	4.9	4.5	4.7
OPF-MR		2.7			9.4			22.4	
OPF-VO-MR	2.7	2.9	2.3	5.3	5.6	6.1	9.5	10.3	9.0
OPF-CN-MR	7.9	12.8	12.6	14.7	22.0	18.7	24.4	36.8	37.8
OPF-PYR		1.1			4.8			11.9	
OPF-VO-PYR	1.0	1.2	1.3	2.1	2.6	2.6	6.8	6.0	6.1
OPF-CN-PYR	3.7	4.0	4.3	7.6	7.6	9.1	13.4	18.7	17.1
Bayes-SSL		0.8			1.6			3.2	
Bayes-VO-SSL	0.3	0.3	0.3	0.7	0.7	0.7	1.9	1.9	1.9
Bayes-CN-SSL	0.4	0.5	0.4	0.8	0.8	0.9	2.3	2.6	2.9
Bayes-MR		1.3			3.7			6.2	
Bayes-VO-MR	0.8	0.8	0.8	1.4	1.5	1.5	3.5	3.5	3.9
Bayes-CN-MR	1.9	1.0	1.3	3.8	3.5	3.1	9.3	10.9	11.9
Bayes-PYR		1.3			3.7			6.2	
Bayes-VO-PYR	0.8	0.7	0.7	1.4	1.5	1.4	2.4	2.5	2.9
Bayes-CN-PYR	0.8	0.8	1.0	1.8	1.9	2.0	3.9	3.8	3.8

Table VII presents the mean computational load in seconds for each approach considered in this work. One can observe the voting-based approaches are consistently more efficient than standard SSL for both OPF and Naïve-Bayes classifiers, since the original training set is now divided in smaller disjoint subsets to train each classifier in the ensemble. However, the complexity grows when one uses the concatenated-oriented approach, since one has larger feature vectors, which end up impacting in the computational burden when computing the Euclidean distance among feature vectors. OPF can handle that problem by using a

precomputed distance matrix, but at the price of requiring more memory space to store such distances.

B. Ikonos-2 MS Image

Table VIII states the results considering Ikonos-2 MS image, being their presentation the very same one used in Section VI-A, i.e., the best results according to Wilcoxon signed-rank test for each pair of classifier and training set size are in bold. Considering this image, we obtained some results that are slightly different from the previous ones (see Section VI-A). First, we observed the accuracy of both OPF and Bayes classifiers, as well as their SSL-based versions increased with larger training sets (only one exception with respect to Bayes-SSL with 10% of the data for training purposes). Second, the standard SSL versions did not improve the results of traditional OPF and Bayes techniques, probably due to large homogeneous regions, thus making no sense the application of standard sequential learning approaches, since when taking the neighborhood of a given sample, there is a high probability of most part of the pixels that fall in that neighborhood have the very same label of that sample. However, the multiresolution SSL-oriented techniques obtained the best results for both OPF and Bayes (i.e., MR-SSL). Actually, the proposed approaches enhanced even more the recognition rates of MR-SSL techniques, being 13.89% more accurate for some cases (e.g., Bayes-CN-MR-SSL using 5% of the dataset for the training set and five classifiers to compose the ensemble).

Fig. 10 displays some of the best results obtained over Ikonos-2 MS image, in which the percentage in parenthesis stands for the amount of training set used. Except for standard SSL, the proposed approaches obtained much more accurate results in all possible configurations, i.e., considering the classifier,

TABLE XI
EXPERIMENTAL RESULTS REGARDING LANDSAT 5 TM IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF		63.2 ± 0.4			63.1 ± 0.5			61.9 ± 0.2	
OPF-SSL		69.2 ± 0.1			70.0 ± 0.0			71.3 ± 0.1	
OPF-VO-SSL	67.5 ± 0.6	67.7 ± 0.6	66.5 ± 0.2	66.8 ± 0.6	66.5 ± 0.5	65.8 ± 0.7	64.0 ± 0.4	64.3 ± 0.3	63.9 ± 0.5
OPF-CN-SSL	71.6 ± 0.2	72.9 ± 0.5	73.3 ± 0.5	72.2 ± 0.1	73.0 ± 0.2	74.0 ± 0.4	72.6 ± 0.1	73.4 ± 0.1	73.5 ± 0.2
OPF-MR-SSL		69.2 ± 0.1			69.7 ± 0.0			71.1 ± 0.1	
OPF-VO-MR-SSL	66.5 ± 0.2	65.4 ± 0.3	65.8 ± 0.5	66.1 ± 0.1	65.4 ± 0.3	66.2 ± 0.2	65.2 ± 0.2	65.3 ± 0.1	65.3 ± 0.1
OPF-CN-MR-SSL	71.1 ± 0.1	71.3 ± 0.1	71.5 ± 0.1	71.4 ± 0.1	71.5 ± 0.1	71.9 ± 0.1	72.0 ± 0.1	72.1 ± 0.1	72.1 ± 0.1
OPF-PY-SSL		69.2 ± 0.1			69.0 ± 0.0			70.1 ± 0.2	
OPF-VO-PY-SSL	69.4 ± 0.5	68.0 ± 0.4	68.4 ± 0.1	68.0 ± 0.1	66.0 ± 0.5	67.5 ± 0.1	65.8 ± 0.4	65.1 ± 0.5	63.4 ± 0.5
OPF-CN-PY-SSL	72.4 ± 0.4	74.6 ± 0.4	75.7 ± 0.6	72.6 ± 0.3	74.9 ± 0.6	75.4 ± 0.4	73.4 ± 0.3	74.2 ± 0.5	74.1 ± 0.5
Bayes		69.6 ± 0.1			69.0 ± 0.2			69.0 ± 0.2	
Bayes-SSL		72.9 ± 0.0			73.4 ± 0.1			74.5 ± 0.3	
Bayes-VO-SSL	71.9 ± 0.7	72.0 ± 0.6	71.9 ± 0.4	71.5 ± 0.3	71.6 ± 0.4	71.6 ± 0.5	69.9 ± 0.4	70.9 ± 0.6	70.6 ± 0.5
Bayes-CN-SSL	72.7 ± 0.3	73.5 ± 0.4	73.5 ± 0.2	73.1 ± 0.1	73.4 ± 0.1	74.2 ± 0.3	73.2 ± 0.1	73.7 ± 0.1	73.9 ± 0.1
Bayes-MR-SSL		70.6 ± 0.1			71.4 ± 0.0			73.0 ± 0.0	
Bayes-VO-MR-SSL	70.9 ± 0.3	70.9 ± 0.1	70.9 ± 0.3	71.0 ± 0.0	70.8 ± 0.1	71.0 ± 0.1	71.0 ± 0.1	70.7 ± 0.2	70.6 ± 0.2
Bayes-CN-MR-SSL	71.6 ± 0.2	71.3 ± 0.1	71.7 ± 0.1	71.6 ± 0.2	71.6 ± 0.4	71.9 ± 0.1	72.2 ± 0.1	72.2 ± 0.1	72.3 ± 0.1
Bayes-PY-SSL		71.3 ± 0.1			72.5 ± 0.2			73.2 ± 0.0	
Bayes-VO-PY-SSL	72.9 ± 0.1	73.8 ± 0.3	72.6 ± 0.7	72.6 ± 0.2	71.4 ± 0.5	73.4 ± 0.7	71.1 ± 0.5	71.3 ± 0.5	71.1 ± 0.3
Bayes-CN-PY-SSL	74.1 ± 0.5	75.3 ± 0.6	75.7 ± 0.5	73.9 ± 0.6	75.3 ± 0.3	75.4 ± 0.5	75.0 ± 0.4	75.2 ± 0.6	75.8 ± 0.6

TABLE XII
EXPERIMENTAL RESULTS REGARDING THE “HYBRID” EXPERIMENT OVER LANDSAT 5 TM IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
First step classification using OPF and the second step using Bayes									
SSL		68.7 ± 0.9			69.2 ± 0.5			72.1 ± 0.3	
VO-SSL	65.4 ± 1.1	66.7 ± 0.0	68.3 ± 0.0	63.4 ± 0.6	64.4 ± 0.0	67.4 ± 0.5	61.8 ± 0.9	62.9 ± 1.1	61.6 ± 0.0
CN-SSL	65.0 ± 1.0	68.3 ± 0.0	68.4 ± 0.0	64.8 ± 1.5	64.4 ± 0.0	67.0 ± 0.2	63.4 ± 0.9	63.4 ± 1.0	62.3 ± 0.1
MR		65.1 ± 1.9			63.7 ± 2.1			64.8 ± 1.0	
VO-MR	64.4 ± 1.0	64.3 ± 0.0	66.2 ± 0.0	72.9 ± 3.5	71.4 ± 0.0	70.4 ± 0.0	73.1 ± 0.8	71.5 ± 2.7	69.6 ± 0.0
CN-MR	70.7 ± 1.9	71.1 ± 0.0	73.5 ± 0.0	72.7 ± 0.7	74.0 ± 0.0	75.4 ± 0.3	80.1 ± 1.1	73.6 ± 1.2	73.6 ± 0.0
PYR		69.1 ± 1.2			69.1 ± 0.7			69.8 ± 2.1	
VO-PYR	65.0 ± 1.0	68.3 ± 0.0	68.1 ± 0.0	64.8 ± 1.5	64.4 ± 0.0	64.8 ± 0.2	63.4 ± 0.9	63.4 ± 1.0	62.9 ± 0.0
CN-PYR	62.4 ± 0.8	63.3 ± 0.0	66.3 ± 0.0	63.3 ± 1.1	64.4 ± 0.0	70.1 ± 0.1	73.2 ± 1.1	74.4 ± 0.9	71.6 ± 0.5
First step classification using Bayes and the second step using OPF									
SSL		66.9 ± 0.2			67.1 ± 0.2			63.1 ± 0.6	
VO-SSL	60.9 ± 1.1	58.8 ± 0.0	60.4 ± 0.0	58.9 ± 0.8	58.8 ± 0.0	58.0 ± 0.0	57.1 ± 1.0	57.6 ± 0.9	56.8 ± 0.0
CN-SSL	65.2 ± 0.9	66.6 ± 0.0	65.4 ± 0.0	64.7 ± 0.6	58.8 ± 0.0	61.0 ± 0.0	65.0 ± 0.3	65.8 ± 0.8	63.8 ± 0.2
MR		65.9 ± 0.4			66.6 ± 0.3			66.6 ± 0.1	
VO-MR	60.6 ± 0.1	60.1 ± 0.0	64.6 ± 0.0	59.8 ± 0.6	58.8 ± 0.0	62.8 ± 0.7	58.7 ± 0.6	59.0 ± 0.7	58.1 ± 0.1
CN-MR	66.3 ± 0.2	65.8 ± 0.0	64.9 ± 0.0	66.4 ± 0.1	58.8 ± 1.0	69.5 ± 0.9	67.0 ± 0.1	66.9 ± 0.2	67.8 ± 0.1
PYR		66.1 ± 0.1			66.8 ± 0.2			64.9 ± 0.4	
VO-PYR	60.3 ± 0.9	60.9 ± 0.0	60.6 ± 0.0	58.2 ± 0.4	58.8 ± 0.0	60.0 ± 0.0	57.4 ± 0.6	57.5 ± 0.8	56.8 ± 0.0
CN-PYR	65.7 ± 0.3	66.0 ± 0.0	65.5 ± 0.0	65.6 ± 0.5	62.8 ± 0.0	67.0 ± 0.8	66.1 ± 0.5	65.9 ± 0.2	65.8 ± 0.0

percentage of training set, and the sequential learning base method. Table IX presents the results concerning the “Hybrid” protocol. Once again, better results than the “Standard” one were obtained, being the OPF classifier the best choice for the initial classification. For example, using 10% of the data for training purposes, the “Hybrid” protocol obtained 78.4% using five classifiers, while the “Standard” protocol achieved 72.5% using seven classifiers.

Table X presents the mean computational load in seconds concerning Ikonos-2 image. Once again, the voting-based

approaches are considerably faster than standard SSL; meanwhile, concatenation-driven ones require more computational burden. Since the pyramid decomposition reduces the image resolution, it requires less computational load when compared against the multiresolution approach.

C. Landsat 5 TM and Geoeye Images

In this section, a brief discussion the experiments obtained over Landsat 5 TM and Geoeye images is considered, since

TABLE XIII
MEAN COMPUTATIONAL LOAD IN SECONDS REGARDING LANDSAT 5 TM IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF-SSL		5.2			15.0			32.0	
OPF-VO-SSL	4.1	4.1	4.1	7.3	7.8	7.3	15.7	13.4	11.7
OPF-CN-SSL	13.0	12.9	14.1	31.5	27.9	31.1	55.1	53.5	51.6
OPF-MR		47.9			105.1			217.2	
OPF-VO-MR	31.2	31.3	34.1	63.3	67.8	65.1	109.2	110.3	111.6
OPF-CN-MR	118.5	184.6	188.2	324.7	357.1	322.1	631.3	628.7	634.4
OPF-PYR		22.1			55.9			101.3	
OPF-CN-PYR	44.0	49.3	44.1	95.7	89.1	91.0	181.1	218.0	211.7
OPF-VO-PYR	12.8	12.8	14.1	31.8	28.0	29.6	56.4	53.4	61.6
Bayes-SSL		9.8			19.2			21.0	
Bayes-VO-SSL	4.1	4.0	4.0	7.1	7.2	7.7	11.3	11.2	11.1
Bayes-CN-SSL	6.1	7.4	6.1	11.5	10.2	10.7	20.1	24.6	21.0
Bayes-MR		12.0			37.5			71.1	
Bayes-VO-MR	10.9	10.7	14.1	20.4	27.2	23.4	44.4	45.8	51.0
Bayes-CN-MR	11.4	10.5	14.1	37.1	37.2	34.2	94.2	91.9	101.1
Bayes-PYR		11.7			40.1			100.1	
Bayes-VO-PYR	10.3	10.4	11.1	19.1	17.1	19.1	40.5	40.9	41.1
Bayes-CN-PYR	21.2	20.9	25.0	57.1	57.1	54.3	104.3	101.9	101.1

TABLE XIV
EXPERIMENTAL RESULTS REGARDING GEOEYE IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF		69.9 ± 1.3			70.5 ± 1.4			71.6 ± 2.0	
OPF-SSL		71.9 ± 0.6			72.5 ± 0.4			73.4 ± 0.4	
OPF-VO-SSL	71.6 ± 0.6	71.0 ± 0.0	71.3 ± 0.7	71.8 ± 0.6	71.1 ± 0.6	71.6 ± 0.6	71.5 ± 0.4	72.1 ± 0.6	71.4 ± 0.5
OPF-CN-SSL	72.0 ± 0.5	72.4 ± 0.5	72.6 ± 0.7	72.9 ± 0.4	72.5 ± 0.4	72.7 ± 0.4	73.1 ± 0.2	73.6 ± 0.3	73.3 ± 0.3
OPF-MR-SSL		70.8 ± 1.0			71.5 ± 0.2			72.8 ± 0.1	
OPF-VO-MR-SSL	72.5 ± 0.1	72.6 ± 0.1	72.2 ± 0.1	71.1 ± 0.1	70.6 ± 0.0	70.7 ± 0.0	72.4 ± 0.0	70.8 ± 0.0	70.8 ± 0.0
OPF-CN-MR-SSL	72.9 ± 0.1	73.1 ± 0.1	73.1 ± 0.1	73.9 ± 0.1	74.1 ± 0.1	74.3 ± 0.1	75.2 ± 0.0	75.5 ± 0.1	75.5 ± 0.1
OPF-PY-SSL		70.7 ± 0.6			71.7 ± 0.5			73.0 ± 0.3	
OPF-VO-PY-SSL	70.4 ± 0.5	71.5 ± 0.5	72.2 ± 0.5	72.2 ± 0.6	72.0 ± 0.6	71.7 ± 0.6	72.3 ± 0.4	71.6 ± 0.2	73.6 ± 0.1
OPF-CN-PY-SSL	72.3 ± 0.5	71.9 ± 0.5	72.2 ± 0.5	72.4 ± 0.5	72.5 ± 0.4	72.7 ± 0.1	73.5 ± 0.1	73.8 ± 0.3	74.4 ± 0.2
Bayes		70.3 ± 0.0			70.8 ± 0.0			71.6 ± 0.1	
Bayes-SSL		69.0 ± 0.0			70.1 ± 0.0			72.4 ± 0.3	
Bayes-VO-SSL	70.2 ± 0.2	70.3 ± 0.2	70.5 ± 0.3	71.2 ± 0.1	71.3 ± 0.1	71.3 ± 0.2	72.1 ± 0.0	72.1 ± 0.1	72.0 ± 0.1
Bayes-CN-SSL	71.7 ± 0.2	71.9 ± 0.3	71.9 ± 0.4	72.5 ± 0.1	72.5 ± 0.2	72.3 ± 0.3	73.7 ± 0.2	73.6 ± 0.2	73.5 ± 0.2
Bayes-MR-SSL		65.6 ± 0.1			67.1 ± 0.0			68.9 ± 0.2	
Bayes-VO-MR-SSL	73.0 ± 0.1	72.9 ± 0.1	72.7 ± 0.2	73.8 ± 0.1	73.9 ± 0.1	73.7 ± 0.1	73.4 ± 0.0	75.1 ± 0.1	75.0 ± 0.1
Bayes-CN-MR-SSL	73.3 ± 0.2	73.4 ± 0.2	73.4 ± 0.2	74.2 ± 0.1	74.5 ± 0.1	74.7 ± 0.0	75.7 ± 0.1	75.9 ± 0.0	75.7 ± 0.1
Bayes-PY-SSL		66.6 ± 0.1			68.0 ± 0.0			69.1 ± 0.0	
Bayes-VO-PY-SSL	70.5 ± 0.1	70.7 ± 0.1	70.6 ± 0.1	71.8 ± 0.1	71.5 ± 0.1	71.6 ± 0.1	72.4 ± 0.1	72.7 ± 0.1	72.7 ± 0.1
Bayes-CN-PY-SSL	71.6 ± 0.1	71.8 ± 0.1	72.0 ± 0.1	72.5 ± 0.1	72.5 ± 0.1	72.8 ± 0.1	73.5 ± 0.1	73.5 ± 0.1	74.2 ± 0.1

their respective covering areas are the very same ones regarding CBERS-2B and Ikonos-2 MS images, respectively. Table XI presents the results with respect to Landsat 5 TM image. In this case, standard OPF did not get better when increasing the training set size; meanwhile, OPF-SSL obtained some results slightly more accurate with larger data for training. Once again, the proposed approaches (CN-SSL only) obtained the best results, but in this case with pyramidal-based SSL (see Section III-C). Considering OPF-CN-PY-SSL with 5% of the dataset for training, for instance, the proposed approach obtained results around 8.58% more accurate than OPF-PY-SSL. Additionally, Table XII presents the results using the “Hybrid” protocol. In

this case, both protocols (i.e., the hybrid and non-hybrid approaches) obtained similar results using 5% and 10% for training purposes. However, if one takes into account a larger training set (i.e., with 20%), the accuracy rate raised from 74.2% to 80.1% when using OPF as the first classifier, which is quite good. Finally, Table XIII presents the mean computational load of the compared techniques, which follows the very same behavior observed for the other images.

Table XIV presents the experimental evaluation with respect to Geoeye image, being the results somehow similar to the ones obtained for Ikonos-2 MS data. In this case, the most accurate techniques were the ones based on MR- and CN-SSL for both

TABLE XV
EXPERIMENTAL RESULTS REGARDING THE “HYBRID” EXPERIMENT OVER REGARDING GEOEYE IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
First step classification using OPF and the second step using Bayes									
SSL		70.6 ± 0.8			73.4 ± 0.6			74.1 ± 0.8	
VO-SSL	71.0 ± 1.6	68.3 ± 0.0	71.9 ± 0.3	70.7 ± 1.2	71.5 ± 0.0	73.6 ± 0.1	71.6 ± 1.3	70.9 ± 1.1	70.2 ± 0.2
CN-SSL	71.4 ± 1.1	69.7 ± 0.0	70.9 ± 0.2	72.1 ± 0.7	71.5 ± 0.0	70.1 ± 0.9	72.6 ± 1.1	71.6 ± 0.4	71.2 ± 0.0
MR		68.7 ± 0.5			75.2 ± 0.4			78.1 ± 0.8	
VO-MR	72.3 ± 0.4	70.6 ± 0.0	71.9 ± 0.1	74.4 ± 0.4	71.5 ± 0.0	72.4 ± 0.3	77.8 ± 0.7	77.1 ± 0.6	74.1 ± 0.5
CN-MR	80.7 ± 0.5	82.6 ± 0.0	81.2 ± 0.0	74.5 ± 0.4	74.8 ± 0.0	74.4 ± 0.4	77.8 ± 0.4	78.3 ± 0.3	78.7 ± 0.0
PYR		73.0 ± 1.2			74.4 ± 0.7			73.8 ± 1.4	
VO-PYR	71.4 ± 1.1	69.7 ± 0.0	70.7 ± 0.0	72.1 ± 0.7	71.5 ± 0.0	71.5 ± 0.0	72.6 ± 1.1	71.6 ± 0.4	70.3 ± 0.0
CN-PYR	74.6 ± 0.3	75.1 ± 0.0	73.6 ± 0.3	76.0 ± 0.6	71.5 ± 0.0	73.8 ± 0.0	78.2 ± 0.3	78.1 ± 0.3	75.2 ± 0.2
First step classification using Bayes and the second step using OPF									
SSL		66.3 ± 0.2			68.0 ± 0.1			69.8 ± 1.2	
VO-SSL	65.3 ± 1.3	67.4 ± 0.0	64.9 ± 0.0	67.1 ± 0.9	67.6 ± 0.0	71.6 ± 0.4	66.6 ± 1.5	67.2 ± 1.3	68.7 ± 0.1
CN-SSL	66.4 ± 0.7	68.8 ± 0.0	66.8 ± 0.0	67.1 ± 1.1	67.6 ± 0.10	69.1 ± 0.1	69.2 ± 0.9	68.2 ± 0.9	68.2 ± 0.3
MR		67.6 ± 0.2			67.6 ± 0.1			69.6 ± 0.1	
VO-MR	67.1 ± 0.3	67.2 ± 0.0	65.8 ± 0.2	68.0 ± 0.2	67.6 ± 0.0	69.6 ± 0.0	69.2 ± 0.2	69.1 ± 0.2	68.7 ± 0.0
CN-MR	67.8 ± 0.1	67.7 ± 0.0	66.9 ± 0.1	68.7 ± 0.1	67.6 ± 0.0	69.4 ± 0.4	69.8 ± 0.1	70.0 ± 0.1	72.7 ± 0.0
PYR		66.0 ± 0.2			68.4 ± 0.1			69.4 ± 0.2	
VO-PYR	65.7 ± 1.2	64.7 ± 0.0	64.9 ± 0.0	67.1 ± 0.9	67.6 ± 0.0	68.8 ± 0.2	67.5 ± 1.0	67.3 ± 1.0	66.4 ± 0.2
CN-PYR	67.2 ± 0.8	67.9 ± 0.0	66.8 ± 0.1	67.1 ± 0.9	67.6 ± 0.0	69.9 ± 0.1	68.7 ± 0.8	69.3 ± 0.6	68.9 ± 0.0

TABLE XVI
MEAN COMPUTATIONAL LOAD IN SECONDS REGARDING GEOEYE IMAGE USING 5%, 10%, AND 20% OF THE IMAGE FOR TRAINING PURPOSES WITH THREE, FIVE, AND SEVEN BASE CLASSIFIERS

Base classifiers	Accuracy (5%)			Accuracy (10%)			Accuracy (20%)		
	3	5	7	3	5	7	3	5	7
OPF-SSL		0.3			1.3			3.1	
OPF-VO-SSL	0.3	0.4	0.3	0.7	0.7	0.7	1.2	1.2	1.1
OPF-CN-SSL	1.2	1.2	1.3	2.7	2.7	2.6	6.2	5.8	5.1
OPF-MR		3.0			10.6			23.3	
OPF-VO-MR	3.1	3.1	3.3	6.4	6.7	6.1	11.6	11.5	11.1
OPF-CN-MR	8.9	9.9	9.4	16.1	24.9	18.7	37.6	41.6	43.1
OPF-PYR		1.2			5.5			15.1	
OPF-VO-PYR	1.3	1.3	1.3	2.7	2.7	2.9	6.0	6.0	6.1
OPF-CN-PYR	3.9	3.3	3.3	8.2	8.7	9.1	14.6	20.8	18.1
Bayes-SSL		0.8			2.8			5.1	
Bayes-VO-SSL	0.3	0.3	0.3	0.6	0.6	0.6	1.0	1.0	1.0
Bayes-CN-SSL	0.4	0.5	0.5	0.8	0.9	0.8	1.4	1.7	1.2
Bayes-MR		3.0			10.5			35.1	
Bayes-VO-MR	0.8	0.8	0.8	1.6	1.6	1.6	3.1	3.0	3.0
Bayes-CN-MR	2.0	3.1	1.4	4.0	3.6	4.9	8.2	11.9	9.0
Bayes-PYR		1.1			14.3			31.1	
Bayes-VO-PYR	0.8	0.8	0.9	1.5	1.6	1.6	2.8	2.9	1.0
Bayes-CN-PYR	0.9	0.9	1.0	1.9	2.0	2.1	3.8	3.9	4.0

classifiers. Another interesting point is the capability of CN-SSL to achieve better results with larger training sets. If one takes into account OPF-MR-SSL, for instance, it is possible to observe the accuracy raised from 70.8% using 5% for training to 72.8% with 20% to compose the training set, thus increasing the recognition rates in around 2.74%. However, considering OPF-CN-MR-SSL for the very same range of training set size percentages, the accuracy was increased in about 3.17%.

Table XV presents the results concerning the “Hybrid” protocol. Once again, this approach allowed better results than the

“Standard” protocol with OPF as the first classifier. If one takes into account a training set of 5%, the accuracy raised from 73.1% to 82.6% (i.e., the accuracy increased about 11.50%) by just considering different classifiers during the stacked approach. In fact, since each classifier works differently, we assume they might be complimentary to each other, thus strengthening the idea of using the ensemble. Finally, Table XVI presents the mean computational load in seconds concerning Geoeeye image, being the behavior pretty much similar to the aforementioned experiments, i.e., the voting-based approaches are considerably faster than naïve SSL, since one has classifiers learning over smaller training sets. The concatenated approaches pay the price of using larger feature vectors. An interesting observation concerns Naïve-Bayes, which is consistently more effective even using the concatenation-driven approach. Actually, OPF is penalized with larger feature vectors, since it needs to compute the distance among them several times during learning.

VII. CONCLUSION

In this paper, we coped with the problem of land-cover classification by means of SSL, which attempts at considering spatial information during the learning process. Such techniques perform an additional classification step with the original feature vector of each sample extended with the labels of the neighborhood samples (pixels). Therefore, the idea is to employ some sort of contextual learning to make the classification process smarter.

Although such techniques usually improve the recognition rates, they may also degrade the learning process when adding misclassified labels to extend the original feature vectors of the dataset samples. Therefore, in this paper, we propose two ensemble-based approaches to alleviate this problem, since we are now considering a committee of specialists to classify each

sample, thus obtaining a more reliable classification process for the further extension of the feature vectors. The first approach concatenates the outputs of each classifier in the ensemble, while the second takes the majority voting of them.

The proposed approaches were validated in nine different base SSL techniques, as well as using two different classifiers. Although the proposed approaches were based on the OPF classifier, we also showed they can be used with other machine learning techniques. In addition, four satellite images were used in this work: CBERS-2B, Landsat 5 TM, Ikonos-2 MS, and Geoeye. The proposed approaches were adapted in three distinct SSL-oriented learning algorithms: standard SSL, MR-SSL, and pyramidal-decomposition SSL (PY-SSL). In all situations, at least one of the proposed approaches obtained the best results according to the Wilcoxon signed-rank, usually the concatenated-based one, which can alleviate the problem of adding misclassified samples by increasing the dimension of the feature space.

An additional round of experiments (“Hybrid” protocol) showed to improve the results of the proposed approaches when we used different classifiers at each stage of the sequential learning process. As a matter of fact, the idea of using ensembles to promote the diversity plays a big role at different steps of the process, i.e., when we are in charge of deciding how to extend the feature vector, or even when we try to alleviate the shortcomings of a given classifier used in a previous state of the stacked paradigm. In regard to future works, we aim at comparing different schemas for the decision fusion step, such as weighted voting and metaheuristic-based ones.

REFERENCES

- [1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd. ed. New York, NY, USA: Wiley-Interscience, 2000.
- [2] M. A. Bencherif, Y. Bazi, A. Guessoum, N. Alajlan, F. Melgani, and H. AlHichri, “Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 527–531, Mar. 2015.
- [3] J. A. Santos, P. H. Gosselin, S. Philipp-Foliguet, R. S. Torres, and A. X. Falcão, “Interactive multiscale classification of high-resolution remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 2020–2034, Aug. 2013.
- [4] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, “Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, Jun. 2015.
- [5] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, “Convolutional neural networks for large-scale remote-sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.
- [6] M. K. S. Varma, N. K. K. Rao, K. K. Raju, and G. P. S. Varma, “Pixel-based classification using support vector machine classifier,” in *Proc. IEEE 6th Int. Conf. Adv. Comput.*, 2016, pp. 51–55.
- [7] B. Kumar and O. Dikshit, “Spectral contextual classification of hyperspectral imagery with probabilistic relaxation labeling,” *IEEE Trans. Cybern.*, 2016, doi: 10.1109/TCYB.2016.2609882.
- [8] A. Masjedi, M. J. Valadan Zoej, and Y. Maghsoudi, “Classification of polarimetric SAR images based on modeling contextual information and using texture features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 932–943, Feb. 2016.
- [9] P. Tokarczyk, J. D. Wegner, S. Walk, and K. Schindler, “Features, color spaces, and boosting: New insights on semantic classification of remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 280–295, Jan. 2015.
- [10] W. W. Cohen and V. R. Carvalho, “Stacked sequential learning,” in *Proc. 19th Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, 2005, pp. 671–676.
- [11] C. Gatta, E. Puertas, and O. Pujol, “Multi-scale stacked sequential learning,” *Pattern Recognit.*, vol. 44, no. 10–11, pp. 2414–2426, 2011.
- [12] E. Puertas, S. Escalera, and O. Pujol, “Multi-class multi-scale stacked sequential learning,” in *Multiple Classifier Systems (Lecture Notes in Computer Science)*, vol. 6713, C. Sansone, J. Kittler, and F. Roli, Eds. Berlin, Germany: Springer, 2011, pp. 197–206.
- [13] F. Sampedro, S. Escalera, and A. Puig, “Iterative multi-class multi-scale stacked sequential learning: Definition and application to medical volume segmentation,” *Pattern Recognit. Lett.*, vol. 46, pp. 1–10, 2014.
- [14] A. González, D. Vázquez, S. Ramos, A. M. López, and J. Amores, “Spatiotemporal stacked sequential learning for pedestrian detection,” in *Pattern Recognition and Image Analysis (Lecture Notes in Computer Science)*, vol. 9117, R. Paredes, J. S. Cardoso, and X. M. Pardo, Eds., Berlin, Germany: Springer, 2015, pp. 3–12.
- [15] E. Puertas, S. Escalera, and O. Pujol, “Generalized multi-scale stacked sequential learning for multi-class classification,” *Pattern Anal. Appl.*, vol. 18, no. 2, pp. 247–261, 2015.
- [16] R. Roscher, B. Waske, and W. Forstner, “Incremental import vector machines for classifying hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3463–3473, Sep. 2012.
- [17] K. Subramanian and S. Suresh, “A meta-cognitive sequential learning algorithm for neuro-fuzzy inference system,” *Appl. Soft Comput.*, vol. 12, no. 11, pp. 3603–3614, 2012.
- [18] Y. Sun, Y. Yuan, and G. Wang, “An on-line sequential learning method in social networks for node classification,” *Neurocomputing*, vol. 149, pp. 207–214, 2015.
- [19] X. Wang and M. Han, “Online sequential extreme learning machine with kernels for nonstationary time series prediction,” *Neurocomputing*, vol. 145, pp. 90–97, 2014.
- [20] S. Suna, P. Zhonga, H. Xiaoa, and R. Wanga, “Spatial contextual classification of remote sensing images using a gaussian process,” *Remote Sens. Lett.*, vol. 7, no. 2, pp. 131–140, 2016.
- [21] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, “A review of remote sensing image classification techniques: The role of spatio-contextual information,” *Eur. J. Remote Sens.*, vol. 47, pp. 389–411, 2014.
- [22] C. H. Li, B. C. Kuo, C. T. Lin, and C. S. Huang, “A spatial-contextual support vector machine for remotely sensed image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 784–799, Mar. 2012.
- [23] G. Moser and S. B. Serpico, “Combining support vector machines and Markov random fields in an integrated framework for contextual image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.
- [24] D. Osaku *et al.*, “Improving land cover classification through contextual-based optimum-path forest,” *Inf. Sci.*, vol. 324, pp. 60–87, 2015.
- [25] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, “Supervised pattern classification based on optimum-path forest,” *Int. J. Imag. Syst. Technol.*, vol. 19, no. 2, pp. 120–131, 2009.
- [26] J. P. Papa, A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares, “Efficient supervised optimum-path forest classification for large datasets,” *Pattern Recognit.*, vol. 45, no. 1, pp. 512–520, 2012.
- [27] J. P. Papa, S. E. N. Fernandes, and A. X. Falcão, “Optimum-path forest based on k-connectivity: Theory and applications,” *Pattern Recognit. Lett.*, 2016, doi: 10.1016/j.patrec.2016.07.026.
- [28] D. R. Pereira, R. Y. M. Nakamura, R. J. Pisani, and J. P. Papa, “Land-cover classification through sequential learning-based optimum-path forest,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 76–79.
- [29] R. J. Pisani, R. Y. N. Mizobe, P. S. Riedel, C. R. L. Zimback, A. X. Falcão, and J. P. Papa, “Toward satellite-based land cover classification through optimum-path forest,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6075–6085, Oct. 2014.
- [30] R. Y. M. Nakamura, L. M. G. Fonseca, J. A. dos Santos, R. S. Torres, X.-S. Yang, and J. P. Papa, “Nature-inspired framework for hyperspectral band selection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2126–2137, Apr. 2014.
- [31] C. Allène, J.-Y. Audibert, M. Couprie, and R. Keriven, “Some links between extremum spanning forests, watersheds and min-cuts,” *Image Vision Comput.*, vol. 28, no. 10, pp. 1460–1471, 2010.
- [32] T. G. Dietterich, “Machine learning for sequential data: A review,” in *Proc. Joint IAPR Int. Workshop Struct., Syntactic, Statist. Pattern Recognit.*, 2002, pp. 15–30.
- [33] J. P. Papa, C. T. N. Suzuki, and A. X. Falcão, *LibOPF: A Library for the Design of Optimum-path Forest Classifiers, Software version 2.1*. 2014. [Online]. Available: <http://www.ic.unicamp.br/afalcao/LibOPF>
- [34] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.



Danilo R. Pereira received the B.Sc. degree from the Faculty of Sciences and Technology, São Paulo State University, Presidente Prudente, Brazil, in 2006, and the M.Sc. and Ph.D. degrees from the University of Campinas, Campinas, Brazil, in 2009 and 2013, respectively, all in computer science.

He is a Professor with the University of the West of São Paulo, Presidente Prudente, and a Postdoctoral Student with the Department of Computer Science, Federal University of São Carlos, São Carlos, France. His research interests include machine learning and

computer graphics.



André N. de Souza received the B.Sc. degree in electrical engineering from Mackenzie Presbyterian University, São Paulo, Brazil, in 1991, and the M.Sc. and Ph.D. degrees in electrical engineering from the Polytechnic School, University of São Paulo, São Paulo, in 1995 and 1999, respectively.

He has been an Associate Professor with the Department of Electrical Engineering, São Paulo State University, São Paulo, since 2005. His interests include intelligent systems, transformers, fraud detection in electrical systems, atmospheric discharges,

and power quality.



Rodrigo J. Pisani received the graduate (B.Sc.) degree in geography from the Faculty of Sciences and Technology, São Paulo State University, Presidente Prudente, Brazil, in 2005, the M.Sc. degree in agronomy (energy in agriculture) from the Faculty of Agronomic Sciences, São Paulo State University, Botucatu, Brazil, in 2009, and the Ph.D. degree in geosciences and environment from São Paulo State University, in 2013.

He is an Adjunct Professor with the Nature Sciences Institute, Federal University of Alfenas, Alfenas, Brazil. His research interests include remote sensing, geographic information systems, and geoprocessing tools.



João P. Papa (M'09) received the B.Sc. degree in information systems from São Paulo State University, São Paulo, Brazil, in 2003, the M.Sc. degree in computer science from the Federal University of São Carlos, São Carlos, Brazil, in 2005, and the Ph.D. degree in computer science from the University of Campinas, Campinas, Brazil, in 2008.

During 2008–2009, he was a Postdoctorate Researcher at the University of Campinas, and during 2014–2015, he was a Visiting Professor at the Center for Brain Science, Harvard University. Since 2009, he

has been an Associate Professor with the Department of the Computer Science, São Paulo State University. His research interests include machine learning, pattern recognition, and image processing.