

Pruning optimum-path forest ensembles using metaheuristic optimization for land-cover classification

Silas Evandro Nachif Fernandes, André Nunes de Souza, Danilo Sinkiti Gastaldello, Danillo Roberto Pereira & João Paulo Papa

To cite this article: Silas Evandro Nachif Fernandes, André Nunes de Souza, Danilo Sinkiti Gastaldello, Danillo Roberto Pereira & João Paulo Papa (2017) Pruning optimum-path forest ensembles using metaheuristic optimization for land-cover classification, International Journal of Remote Sensing, 38:20, 5736-5762, DOI: [10.1080/01431161.2017.1346402](https://doi.org/10.1080/01431161.2017.1346402)

To link to this article: <https://doi.org/10.1080/01431161.2017.1346402>



Published online: 05 Jul 2017.



Submit your article to this journal [↗](#)



Article views: 62



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



Pruning optimum-path forest ensembles using metaheuristic optimization for land-cover classification

Silas Evandro Nachif Fernandes^a, André Nunes de Souza^b, Danilo Sinkiti Gastaldello^b, Danilo Roberto Pereira^c and João Paulo Papa^c

^aDepartment of Computing, Federal University of São Carlos, São Carlos, SP, Brazil; ^bDepartment of Electrical Engineering, São Paulo State University, Bauru, SP, Brazil; ^cDepartment of Computing, São Paulo State University, Bauru, SP, Brazil

ABSTRACT

Machine learning techniques have been actively pursued in the last years, mainly due to the increasing number of applications that make use of some sort of intelligent mechanism for decision-making processes. In this context, we shall highlight ensemble pruning strategies, which provide heuristics to select from a collection of classifiers the ones that can really improve recognition rates and provide efficiency by reducing the ensemble size prior to combining the model. In this article, we present and validate an ensemble pruning approach for Optimum-Path Forest (OPF) classifiers based on metaheuristic optimization over general-purpose data sets to validate the effectiveness and efficiency of the proposed approach using distinct configurations in real and synthetic benchmark data sets, and thereafter, we apply the proposed approach in remote-sensing images to investigate the behaviour of the OPF classifier using pruning strategies. The image data sets were obtained from CBERS-2B, LANDSAT-5 TM, IKONOS-2 MS, and GEOEYE sensors, covering some areas of Brazil. The well-known Indian Pines data set was also used. In this work, we evaluate five different optimization algorithms for ensemble pruning, including that Particle Swarm Optimization, Harmony Search, Cuckoo Search, and Firefly Algorithm. In addition, we performed an empirical comparison between Support Vector Machine and OPF using the strategy of ensemble pruning. Experimental results showed the effectiveness and efficiency of ensemble pruning using OPF-based classification, especially concerning ensemble pruning using Harmony Search, which shows to be effective without degrading the performance when applied to large data sets, as well as a good data generalization.

ARTICLE HISTORY

Received 28 September 2016
Accepted 14 June 2017

1. Introduction

The use of multiple classifiers has become an area of great interest in pattern recognition, being highly motivated due to the likeliness in including other classifiers during the decision process as a good approach to improve the generalization ability of the ensemble. Additionally, the strategy of polling the decisions of classifiers that are complementary to each other can improve the classification accuracy of an ensemble with respect to the

individual learners. Generally speaking, the literature proposed different methods for combining classifiers, such as Bagging (Breiman 1996), Boosting (Breiman 1998), and Random Subspaces (Ho 1998), just to name a few. The Random Subspaces technique, for example, creates multiple classifiers using different spaces of features, while Bagging generates different learners by randomly selecting subsets of samples to train base classifiers. Although Boosting uses part of the data to train the classifiers either, the most difficult samples to be classified have a higher probability of being selected to compose the final training set.

Roughly speaking, the task of polling decisions can be defined according to the outputs of the classifiers, which are often divided into three levels (Al-Ani and Deriche 2002): abstract, ranking, and confidence. In the first level, the classifiers associate a single label to each data set sample, while in ranking-based approaches, the possible labels for a sample are stored in a priority queue according to some criterion. In confidence-oriented techniques, the classifier computes some metric that will reflect the probability of each label being assigned to a particular sample.

Among the different methods to combine classifiers, we can cite voting, weighted voting, and methods based on Dempster–Shafer theory of evidence (Kuncheva 2004). Some authors model the combination of classifiers as an optimization problem (Jia, Wang, and Fan 2014), where the weights of the confidence levels for each classifier are determined by evolutionary optimization techniques (Joly, Verstraete, and Paniagua 2014). Nabavi-Kerizi, Abadi, and Kabir (2010) proposed a linear combination of Artificial Neural Networks using Particle Swarm optimization, while Sheen et al. (2012) used the Harmony Search to optimize the composition of a multiple classifier system. Chandrashekar and Sahin (2014) investigated feature selection methods and the effect of the curse of dimensionality in the predictor performance, for example, and Diao and Shen (2011) provided a classifier ensemble selection approach based on fuzzy-rough feature selection (Mendoza, Vellasco, and Figueiredo 2014) and Harmony Search that uses the latter technique to select a minimal subset of features to maximize the fuzzy-rough dependency measure. Lee et al. (2013) proposed an ensemble learning method that employs a set of regional classifiers by weighted voting. Later on, Coletta et al. (2015) employed metaheuristics to fine-tune parameters of clustering ensembles. Also, one can refer to a considerable number of methods for combining classifiers (Lee and Srihari 1995; Kuncheva, Skurichina, and Duin 2002; Kuncheva 2002; Langdon, Barrett, and Buxton 2002; Kuncheva 2004; Macas et al. 2007; Mao et al. 2011).

The Boosting approach has been considered one of the most effective methods for constructing ensembles (Martínez-Muñoz, Hernández-Lobato, and Suárez 2009; Quinlan 1996; Dietterich 2000; Bauer and Kohavi 1999). In some classification problems, such as noisy domains, Boosting methods have poor generalization (Martínez-Muñoz, Hernández-Lobato, and Suárez 2009; Quinlan 1996; Dietterich 2000; Bauer and Kohavi 1999; Caruana and Niculescu-Mizil 2006), and overfitting can also be observed when a considerable number of classifiers are combined (Rätsch, Onoda, and Müller 2001; Martínez-Muñoz, Hernández-Lobato, and Suárez 2009). Thus, the bagging-based methods can be considered safer and more robust than Boosting in some cases (Martínez-Muñoz, Hernández-Lobato, and Suárez 2009; Caruana and Niculescu-Mizil 2006). However, combining a large number of classifiers requires a large memory at the price of a slower classification phase. One way to speed up such process is the selection of a subset of the classifiers from the original ensemble. Such approach, also known as *ensemble pruning*, has provided several benefits in different

approaches (Zhou, Wu, and Tang 2002; Li et al. 2009). Usually, large ensembles of classifiers may comprise both high and low predictive performance models (Markatopoulou, Tsoumakas, and Vlahavas 2010). Pruning these low predictive models for a given classification problem while maintaining the highly predictive ones may contribute to improve the overall predictive performance of the whole ensemble. In general, ensemble pruning selects a subset of the models, thus providing efficiency by reducing the ensemble size prior to combining the model. Some methods to select the subset of classifiers are usually divided as follows: (i) static methods that select a fixed subset of the original ensemble for all test instances and (ii) dynamic methods that select different subsets of the original ensemble for each different test instance (Markatopoulou, Tsoumakas, and Vlahavas 2010). It is worth noting to highlight that finding the optimal sub-ensemble is a hard problem, whose solution is computationally expensive (Martínez-Muñoz, Hernández-Lobato, and Suárez 2009). Therefore, an interesting way of dealing with such problem is to model ensemble pruning as a metaheuristic-driven optimization task (Zhou, Wu, and Tang 2002). In general, it is expected that such methods may provide a near-optimum ensemble selection with a significant reduction of the number of classifiers. A number of works also employed metaheuristics to prune ensembles of classifiers as well (Krawczyk 2015; Jodavi, Abadi, and Parhizkar 2015), with very suitable results.

Furthermore, it is worth noting that land-cover classification is one of the most important applications of remote-sensing images. Some studies indicate significant improvements in the field of combination of classifiers, as highlighted by Abe, Gidudu, and Marwal (2010), with use feature selection techniques to create diversity in ensemble classification. Tinoco et al. (2013), for example, showed the efficiency in using an ensemble of classifiers concerning hyperspectral land-cover analysis. Li and Yin (2013) proposed a variational Bayesian-independent component analysis together with Support Vector Machine (VBICA-SVM) for high-spatial resolution remote-sensing images in order to improve their automatic classification.

Some years ago, Papa et al. (Papa, Falcão, and Suzuki 2009; Papa et al. 2012; Papa, Fernandes, and Falcão 2016) introduced the Optimum-Path Forest (OPF) classifier, which is a graph-based pattern recognition technique that uses a generalization of Dijkstra algorithm for multiple sources and path-cost functions. The OPF classifier has demonstrated interesting results in terms of efficiency and effectiveness, being some of them comparable to the ones obtained by SVM, but usually faster for training, since OPF is parameterless and its training step has a quadratic complexity (Papa, Falcão, and Suzuki 2009; Papa et al. 2012). However, the reader can refer to a very few works that deal with the problem of combining ensembles of OPF classifiers (Ponti and Papa 2011; Ponti, Papa, and Levada 2011). It is worth noting some works that use OPF classifier for improving land-cover classification, such as Osaku et al. (2015, 2016), which employed Markov Random Fields and OPF to classify satellite imagery, and Pisani et al. (2014) that showed OPF can outperform some of the well-known approaches in the context of land-cover classification. Moreover, to the best of our knowledge, we have not observed any work that attempted at pruning ensembles of OPF-based classifiers. Therefore, the main contributions of this work are twofold: (i) to investigate the use of ensemble pruning methods based on metaheuristic techniques considering the OPF classifier and (ii) to evaluate the effectiveness and efficiency of ensemble pruning when applied to remote-sensing image classification. In this work, we evaluated five different optimization algorithms for ensemble pruning, including Particle Swarm Optimization (PSO) (Kennedy and Eberhart 2001), Harmony Search (HS) (Geem 2009), Cuckoo Search (CS) (Yang and Deb 2010), and Firey Algorithm (FFA) (Yang 2010).

2 Optimum-Path Forest

2.1. Theoretical background

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. The OPF classifier framework comprises both supervised and unsupervised versions; however, this work focused only on the supervised model, which is used to develop the proposed approach. Currently, there are two distinct versions of the OPF for supervised learning: (i) one that makes use of a complete graph (Papa, Falcão, and Suzuki 2009; Papa et al. 2012) and (ii) another version that uses a graph k -neighbourhood (Papa and Falcão 2008; Papa, Falcão, and Suzuki 2009; Papa, Fernandes, and Falcão 2016).

The idea of the OPF classifier is to rule a competition process among some key samples (prototypes). Thus, the algorithm outputs an optimum path forest, which is a collection of optimum-path trees (OPTs) rooted at each prototype. Each OPT defines a discrete optimal partition (cluster) of samples that belong to the very same class, thus labelling the whole data set with the label of each prototype sample. This work employs the OPF classifier proposed by Papa et al. (Papa, Falcão, and Suzuki 2009; Papa et al. 2012), which is parameterless for training phase and is explained in more details below.

Let $D = D_1 \cup D_2$ be a λ -labelled data set (i.e. λ stands for a function that can correctly label all data set samples) such that D_1 and D_2 stand for the training and test sets, respectively. Let $S \subset D_1$ be a set of prototypes of all classes (i.e. key samples that best represent the classes). Let (D_1, A) be a complete graph whose nodes are the samples in D_1 , and any pair of samples defines an arc in $A = D_1 \times D_1$. Additionally, let p_s be a path (i.e. a sequence of adjacent and distinct samples) in (D_1, A) with terminus at sample $s \in D_1$. The OPF algorithm employs the path-cost function f_{\max} due to its theoretical properties for estimating prototypes (Section 2.2 gives further details about this procedure):

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

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

where $d(s, t)$ stands for the distance between nodes s and t . Additionally, $\langle s \rangle$ stands for a trivial path, and the operator \cdot denotes a concatenation among paths. Therefore, $f_{\max}(p_s)$ computes the maximum distance among adjacent samples in p_s , when p_s is not a trivial path. In short, the OPF algorithm tries to minimize $f_{\max}(p_t)$, $\forall t \in D_1$.

2.2. Training phase

In regard to the training phase, the optimum set of prototypes, defined as S^* , is obtained when Algorithm 1 minimizes the classification errors for every $s \in D_1$. The set S^* can be found by exploiting the theoretical relation between the Minimum Spanning Tree (MST) and the OPT for f_{\max} (Alléne et al. 2010). The training essentially consists in finding S^* and an OPF classifier rooted at S^* . By computing a MST in the complete graph (D_1, A) , one obtain a connected acyclic graph whose nodes are all samples of D_1 and the arcs are undirected and weighted by the distances d among adjacent samples. In the MST, every

pair of samples is connected by a single path, which is optimum according to f_{\max} . Hence, the minimum-spanning tree contains one OPT for any selected root node. After that, a competition process among samples in S takes place, which try to conquer (i.e. to offer the lowest cost) samples in $D_1 \setminus S$ using f_{\max} (Equation 1).

The optimum prototypes are the closest elements of the MST with different labels in D_1 (i.e. elements that fall in the frontier of the classes). By removing the arcs among different classes, their adjacent samples become prototypes in S^* , and Algorithm 1 can define an OPF with minimum classification errors in D_1 . Algorithm 1 implements the above procedure for the OPF training phase.

Algorithm 1: Optimum-Path Forest Training Algorithm

Input: A labelled training set D_1 .

Output: Optimum-path forest P , cost map C , label map L , and ordered set D'_1 .

Auxilliary: Priority queue Q , set S of prototypes, and variable cst .

```

1 Set  $D'_1 \leftarrow \emptyset$  and compute the set of prototypes  $S \subset D_1$  by MST
2 For each  $s \in D_1 \setminus S$ , set  $C(s) \leftarrow +\infty$ 
3 for each  $s \in S$  do
4    $C(s) \leftarrow 0$ ,  $P(s) \leftarrow nil$ ,  $L(s) \leftarrow \lambda(s)$ , and insert  $s$  in  $Q$ 
5 while  $Q$  is not empty do
6   Remove from  $Q$  a sample  $s$  such that  $C(s)$  is minimum
7   Insert  $s$  in  $D'_1$ 
8   for each  $t \in D_1$  such that  $C(t) > C(s)$  do
9     Compute  $cst \leftarrow \max\{C(s), d(s, t)\}$ 
10    if  $cst < C(t)$  then
11      if  $C(t) \neq +\infty$  then
12        Remove  $t$  from  $Q$ 
13       $P(t) \leftarrow s$ ,  $L(t) \leftarrow L(s)$ ,  $C(t) \leftarrow cst$ ;
14    Insert  $t$  in  $Q$ 
15 return Classifier  $[P, C, L, D'_1]$ 

```

Lines 1–4 initialize maps and insert prototypes in Q (the function $\lambda(\cdot)$ in Line 4 assigns the true label to each training sample). Notice the cost map C stores the optimum cost of each training sample. The main loop computes an optimum path from S to every sample s in a non-decreasing order of cost (Lines 5–14). At each iteration, a path of minimum cost $C(s)$ is obtained in P when we remove its last node s from Q (Line 6). Ties are broken in Q using first-in-first-out policy. That is, when two optimum paths reach an ambiguous sample s with the same minimum cost, s is assigned to the first path that reached it. Note that $C(t) > C(s)$ in Line 8 is false when t has been removed from Q and, therefore, $C(t) \neq +\infty$ in Line 11 is true only when $t \in Q$. Lines 10–14 evaluate whether the path that reaches an adjacent node t through s is cheaper than the current path with terminus t , and update the position of t in Q , $C(t)$, $L(t)$ and $P(t)$ accordingly. The complexity for training OPF is given by $\theta(|D_1|^2)$, due to the main (Lines 5–14) and inner loops (Lines 8–14) in Algorithm 1, which run $\theta(|D_1|)$ times each.

2.3. Classification phase

For any sample $t \in D_2$, we consider all arcs connecting t with samples $s \in D_1$, as though t were part of the training graph. Considering all possible paths from S^* to t , we wish to find the optimum path $P^*(t)$ from S^* , and label t with the class $\lambda(R(t))$ of its most strongly connected prototype $R(t) \in S^*$. This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as follows:

$$C(t) = \min_{\forall s \in D_1} \{ \max \{ C(s), d(s, t) \} \}. \quad (2)$$

Let the node $s^* \in D_1$ be the one that satisfies Equation 2 (i.e. the predecessor $P(t)$ in the optimum path $P^*(t)$). Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ as the class of t . An error occurs when $L(s^*) \neq \lambda(t)$.

Algorithm 2 implements the OPF classification procedure. The main loop (Lines 1–9) performs the classification of all nodes in D_2 . The inner loop (Lines 4–9) visits each node $k_{i+1} \in D'_1, i = 1, 2, \dots, |D'_1| - 1$ until an optimum path $p_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle$ is found. In the worst case, the algorithm visits all nodes in D'_1 . Line 5 evaluates $f_{\max}(p_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle)$ and Lines 7–8 update the cost, label and predecessor of t whenever $p_{k_{i+1}} \cdot \langle k_{i+1}, t \rangle$ is better than the current path pt (Line 6). Although the reader can note the complexity of the OPF classification phase is given by $\theta(|D_1||D_2|)$ (for each classification node we have to visit all training samples), Papa et al. (2012) showed that, in practice, the complexity is given by $O(p|D_2|)$, in which $p \in O(|D_1|)$, i.e. $0 < p \leq |D_1|$.

Algorithm 2: Optimum-Path Forest Classification Algorithm.

Input: Classifier $[P, C, L, D'_1]$ and test set D_2 .

Output: Label L' , cost C' and predecessor P' maps defined for D_2 .

Auxiliary: Cost variables tmp and mincost.

```

1 for each  $t \in D_2$  do
2    $i \leftarrow 0$ , mincost  $\leftarrow \max \{ C(k_i), d(k_i, t) \}$ 
3    $L'(t) \leftarrow L(k_i)$ ,  $P'(t) \leftarrow k_i$  and  $C'(t) \leftarrow \text{mincost}$ 
4   while  $i < |D'_1|$  and mincost  $> C(k_{i+1})$  do
5     Compute tmp  $\leftarrow \max \{ C(k_{i+1}), d(k_{i+1}, t) \}$ 
6     if tmp  $<$  mincost then
7       mincost  $\leftarrow$  tmp and  $C'(t) \leftarrow \text{mincost}$ 
8        $L'(t) \leftarrow L(k_{i+1})$  and  $P'(t) \leftarrow k_{i+1}$ ;
9    $i \leftarrow i + 1$ 
10 return  $[L', C', P']$ 
```

3. Ensemble pruning as a metaheuristic-based optimization

In this section, we describe the strategy used to prune ensembles based on metaheuristic techniques. The naïve OPF classifier uses the abstract output method when classifying samples, i.e. the output of the classifier is a single label, as mentioned in Section 2.3. Xu, Krzyzak, and Suen (1992) defined an interesting approach to combine the outputs of

L classifiers in an ensemble depending on the information obtained from the individual members. Such approach considers that each classifier assigns a class label to every sample in the data set. Therefore, the ensemble of classifiers generates a collection of L possible outputs to each sample to be classified.

Consider an ensemble of L OPFs in a bagging approach, i.e. the classifiers are aggregated by using different bootstrapped samples of the original training data. Let $M = \{M_1, M_2, \dots, M_L\}$ be a set of L classifiers, and $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ be a set of K class labels. Each classifier takes an n -dimensional input vector and associates it to a class label, i.e. $M_i : \mathcal{X}^n \rightarrow \Omega, i = 1, 2, \dots, L$. Therefore, for any sample z to be classified, the ensemble of classifiers generates a collection $\Psi_z = [\psi_z(M_1), \dots, \psi_z(M_L)]$ of possible outputs, where $\psi_z(M_i)$ stands for the output of classifier M_i considering sample z .

However, instead of considering the outputs of all classifiers, the idea of ensemble pruning is to select a subset $M' \subset M$ such that the recognition rate over a validating set is maximized. Let $D_v \subset D_1$ be a validating set, and $\hat{D}_1 = D_1 \setminus D_v$ be an L -folded training set such that $\hat{D}_1 = \hat{D}_1^1 \cup \hat{D}_1^2 \cup \dots \cup \hat{D}_1^L$. The main idea is to train each classifier M'_i over \hat{D}_1^i , and then classify D_v using the majority voting. In short, the idea is to turn 'on' or 'off' each classifier in M to build M' , and then classify D_v . The task to consider or not a given classifier is performed by the metaheuristic optimization technique, which essentially aims at learning what classifier will be turned 'on' or 'off'. Therefore, such choice will be guided by the maximum accuracy over D_v .

In order to decide whether we are going to select a given classifier or not, we associate a weight $w_i \in [0, 1]$ to each classifier, which is further selected to compose M' if $w_i > \tau$, being τ an adaptive threshold (Larkins and Mayo 2008) updated as follows:

$$\tau = \rho - \sigma, \quad (3)$$

where ρ is the mean weight, and σ is calculated by:

$$\sigma = \sqrt{\frac{1}{m} \sum_{i=1}^m (w_i - \rho)^2} \quad \forall w_i < \rho, \quad (4)$$

where m is the number of classifiers whose weight is smaller than ρ . Notice the above equation considers a classifier M_i whose weight w_i is smaller than ρ only.

In short, the main idea is to model each possible solution of the search space as a set of weights, and the optimization technique aims at finding the best values for such weights that maximize the OPF accuracy over D_v . Figure 1 depicts the proposed approach.

4. Data and experimental design

In this section, we present the methodology and experiments employed to validate the effectiveness and efficiency of the proposed approach. Firstly, in Section 4.1, we validate the effectiveness of OPF ensemble pruning using meta-heuristic techniques in some benchmark problems. Further, in Section 4.2 we validate the proposed approach in remote-sensing image classification.

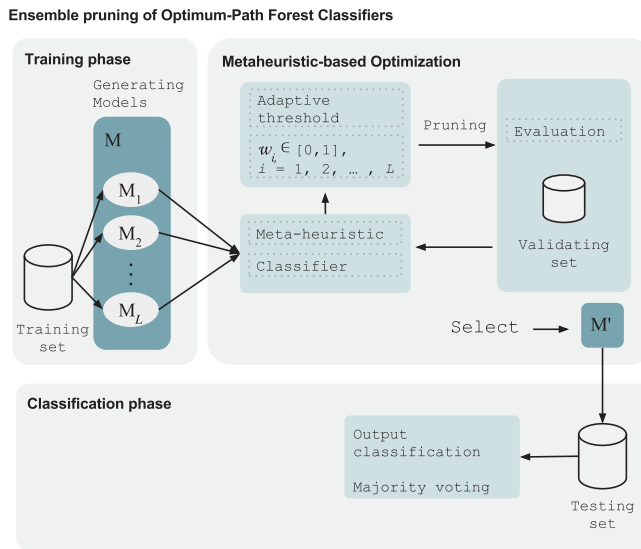


Figure 1. Proposed approach based on ensemble pruning as a metaheuristic-based optimization.

4.1. Experiments over general-purpose data sets

The proposed meta-heuristic-based ensemble pruning approach concerning OPF classifiers is compared against standard OPF over 6 real and synthetic benchmark classification data sets (Sonnenburg 2016; Chang and Lin 2016; Lichman 2013), whose main characteristics are presented in Table 1. Notice that synthetic data sets were generated through an isotropic Gaussian function.

In order to obtain statistically meaningful results, three different ranges of training, validating, and testing sets were used: (i) in a first stage, each data set was partitioned into three subsets, the training (50%), validating (10%), and testing sets (40%), hereinafter denoted as 50:10:40; (ii) in a second stage, the data sets were partitioned as 45:20:35; and (iii) in the last stage, the data sets were partitioned as 40:30:30. For each range, training, validating, and testing sets were selected randomly, and the process was repeated 20 times (cross-validation). Notice the percentages have been empirically chosen, being more intuitive to provide a larger validating set for ensemble pruning learning.

In order to allow a fair comparison, we trained standard OPF over $D_1 \cup D_v$ considering the aforementioned three stages. In addition, the mean accuracy and computational load are also computed for each technique. Notice the idea of using different percentages for D_1 , D_v , and D_2 is motivated to assess the effectiveness of the proposed approach under distinct scenarios. The final results were evaluated through the Wilcoxon signed-rank test with significance of 0.05 (Wilcoxon 1945). For the sake of implementation purposes, we used the open-source libraries LibOPF (Papa, Suzuki, and Falcão 2009) and LibOPT (Papa et al. 2017).

Table 1. Description of the data sets.

Data set	No. of samples	No. of features	No. of classes
Pima-Indians-Diabetes	768	8	2
Statlog-Australian	690	14	2
Statlog-Heart	270	13	2
Synthetic1	500	2	2
Synthetic2	1000	2	2
UCI-Breast-Cancer	683	10	2

In order to evaluate the influence of different ensembles, we performed a comparison with ensembles comprising 3, 5, 7, and 9 classifiers. Furthermore, with respect to the meta-heuristic optimization, we opted to employ the following techniques:

- Harmony Search (HS): 5 harmonies with 20 iterations, HMCR = 0.7, PAR = 0.7 and $\beta = 10$. Variables HMCR and PAR, which stand for ‘Harmony Memory Considering Rate’ and ‘Pitch Adjusting Rate’, are used to guide HS onto the search space, as well as to avoid traps from local optima. Variable β denotes the ‘bandwidth’ (step size) used with PAR.
- Particle Swarm Optimization (PSO): 5 particles with 20 iterations, $c_1 = 1.4$, $c_2 = 0.6$ and $\xi = 0.7$. While c_1 and c_2 are ad-hoc parameters, ξ stands for the well-known ‘inertia weight’, which is used as a step size towards better solutions.
- Firefly Algorithm (FFA): population size of 5 with 20 iterations, $\gamma = 0.3$ and $\mu = 1.0$. The variables γ and μ are used to control the randomness and the attractiveness, respectively.
- Cuckoo Search (CS): 5 particles with 20 iterations, and $p_a = 0.25$. Variable p_a is used to control the elitism and the balance of the randomization and local search.

Table 2 presents the mean accuracies and standard deviation over all data sets under standard OPF and OPF ensemble pruning considering HS, PSO, FFA, and CS optimization techniques, hereinafter called OPF-pruning_{HS}, OPF-pruning_{PSO}, OPF-pruning_{FFA}, and OPF-pruning_{CS}, respectively. The recognition rates are computed according to Papa, Falcão, and Suzuki (2009). The values in bold stand for the most accurate techniques according to the Wilcoxon signed-rank test. One can observe the proposed OPF ensemble pruning has obtained the best results in almost all data sets. If we take a look at Figure 2(a,b) (‘Synthetic1’ and ‘Synthetic2’ data sets, respectively), a considerable amount of overlapping among samples from different classes can be observed. In these cases, an ensemble with more classifiers seemed to be more effective (Table 2), since the feature space ends up being partitioned into different sub-regions. Moreover, the improvement concerning ‘Synthetic1’ and ‘Synthetic2’ data sets with respect to the baseline were around 1.79% and 6.07% considering the set of nine classifiers, respectively.

In order to provide a robust statistical analysis, we performed the non-parametric Friedman test, which is used to rank the algorithms for each data set separately. Friedman test provides meaningful results to reject the null-hypothesis (h_0 : all techniques are equivalent), we can perform a post-hoc test further. For this purpose, we conducted the Nemenyi test, proposed by Nemenyi (Nemenyi 1963), which allows us to verify whether there is a critical difference (CD) among techniques or not. The results of the Nemenyi test can be represented in a simple diagram, in which the average ranks

Table 2. Mean accuracy results (%) and standard deviation over all data sets for standard OPF and OPF ensemble pruning under different optimization techniques over 3, 5, 7, and 9 ensemble configuration.

Data set	Approach	Ensemble of 3 OPFs	Ensemble of 5 OPFs	Ensemble of 7 OPFs	Ensemble of 9 OPFs
Pima-Indians-Diabetes	OPF	65.50 ± 2.08	65.50 ± 2.08	65.50 ± 2.08	65.50 ± 2.08
	OPF-pruning _{CS}	66.87 ± 2.78	68.81 ± 2.46	68.70 ± 3.30	67.65 ± 2.84
	OPF-pruning _{FFA}	63.50 ± 6.01	68.14 ± 5.23	66.33 ± 5.96	65.95 ± 6.00
	OPF-pruning _{HS}	66.90 ± 2.61	68.73 ± 2.44	68.39 ± 3.23	68.26 ± 2.52
	OPF-pruning _{PSO}	67.33 ± 3.39	69.30 ± 2.72	68.09 ± 3.08	67.72 ± 3.41
Statlog-Australian	OPF	77.96 ± 2.04	77.96 ± 2.04	77.96 ± 2.04	77.96 ± 2.04
	OPF-pruning _{CS}	77.02 ± 3.21	80.65 ± 2.52	82.02 ± 1.83	83.57 ± 1.79
	OPF-pruning _{FFA}	75.59 ± 6.19	78.38 ± 7.38	81.66 ± 4.67	80.38 ± 7.57
	OPF-pruning _{HS}	78.20 ± 3.00	81.71 ± 1.93	82.82 ± 2.30	83.52 ± 2.65
	OPF-pruning _{PSO}	76.64 ± 3.06	81.18 ± 2.28	82.40 ± 2.20	83.86 ± 2.16
Statlog-Heart	OPF	74.94 ± 1.91	74.94 ± 1.91	74.94 ± 1.91	74.94 ± 1.91
	OPF-pruning _{CS}	78.78 ± 2.85	77.66 ± 1.63	80.12 ± 3.55	81.10 ± 3.50
	OPF-pruning _{FFA}	70.89 ± 6.71	79.48 ± 4.57	80.30 ± 1.59	80.64 ± 3.32
	OPF-pruning _{HS}	76.18 ± 2.39	80.91 ± 2.40	80.32 ± 2.81	80.91 ± 2.18
	OPF-pruning _{PSO}	73.85 ± 2.99	79.76 ± 2.78	80.05 ± 1.43	81.14 ± 2.28
Synthetic1	OPF	53.86 ± 3.62	55.84 ± 2.61	55.80 ± 3.40	56.31 ± 3.02
	OPF-pruning _{CS}	54.79 ± 2.55	55.76 ± 3.76	56.04 ± 3.56	54.71 ± 2.93
	OPF-pruning _{FFA}	52.97 ± 4.37	55.83 ± 3.21	56.33 ± 3.90	55.78 ± 3.22
	OPF-pruning _{HS}	52.97 ± 4.37	55.83 ± 3.21	56.33 ± 3.90	55.78 ± 3.22
	OPF-pruning _{PSO}	53.38 ± 4.47	54.79 ± 2.60	55.78 ± 4.26	56.29 ± 2.71
Synthetic2	OPF	71.65 ± 2.05	71.65 ± 2.05	71.65 ± 2.05	71.65 ± 2.05
	OPF-pruning _{CS}	71.89 ± 2.89	75.89 ± 2.03	77.25 ± 2.19	77.47 ± 2.60
	OPF-pruning _{FFA}	71.04 ± 5.90	75.69 ± 3.61	75.31 ± 5.31	77.39 ± 3.34
	OPF-pruning _{HS}	73.29 ± 2.43	76.31 ± 2.37	77.03 ± 1.91	77.87 ± 1.81
	OPF-pruning _{PSO}	73.35 ± 2.45	76.33 ± 2.42	77.04 ± 2.09	78.15 ± 1.68
UCI-Breast-Cancer	OPF	94.42 ± 1.05	94.42 ± 1.05	94.42 ± 1.05	94.42 ± 1.05
	OPF-pruning _{CS}	92.33 ± 2.95	94.85 ± 1.63	95.50 ± 1.49	95.57 ± 1.31
	OPF-pruning _{FFA}	88.11 ± 14.48	92.18 ± 6.60	93.66 ± 6.29	92.73 ± 6.71
	OPF-pruning _{HS}	94.10 ± 2.68	95.42 ± 1.20	96.17 ± 1.19	96.09 ± 0.97
	OPF-pruning _{PSO}	91.76 ± 3.36	94.94 ± 1.22	96.23 ± 1.07	95.70 ± 0.87

The most accurate techniques for the Wilcoxon test are highlighted in bold.

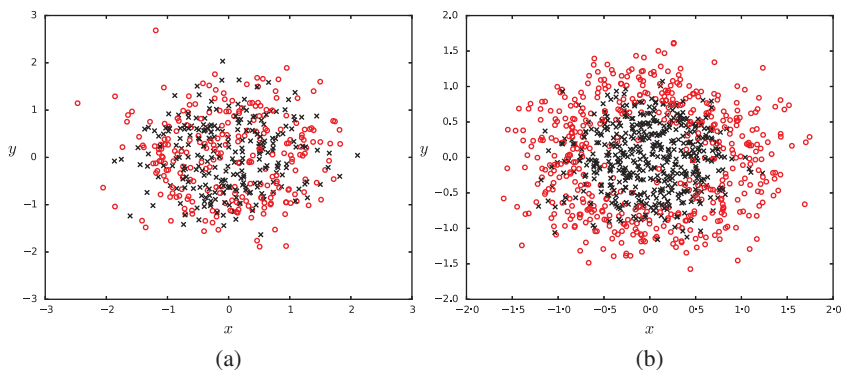


Figure 2. Scatter plot from: (a) 'Synthetic1' and (b) 'Synthetic2' data sets. Notice 'X' and 'O' stand for samples from two distinct classes.

of the methods are plotted on the horizontal axis, where the lower the average rank is, the better the technique is. Moreover, the groups with no significant difference are then connected with a horizontal line.

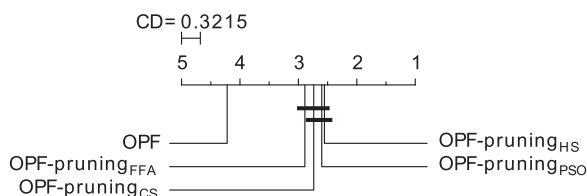


Figure 3. Results of the Nemenyi test for a comparison of standard OPF against OPF ensemble pruning using 9 classifiers and its variations under different optimization techniques based on accuracy results. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

Figure 3 depicts the statistical analysis considering the accuracy results for ensemble pruning using 9 classifiers. As one can observe, the pruning-based approaches can be considered as the most accurate ones. Such point reflects the OPF ensemble pruning using meta-heuristic techniques achieved the best accuracy rates in the majority of data sets. Furthermore, the HS approach can be considered as the most accurate technique. However, the statistical test did not point out a CD between HS, PSO, and CS meta-heuristics in the first group, which means they performed similarly.

Figure 4(a) depicts the computational load considering the training time (training and validating) with the Nemenyi test for ensemble pruning using 9 classifiers. The group considered the fastest one in the training and validating phase consists of standard OPF and OPF pruning using HS, where with there is no CD between them. On average, the OPF pruning using HS has been about 2.3559 times slower than standard OPF in training and validating phase. In regard to the testing phase (Figure 4(b)), HS, FFA, and PSO have been the fastest ones, since such techniques pruned more classifiers with respect to the original ensemble, i.e. fewer classifiers for testing phase are now considered. In short, we can draw some conclusions:

- the proposed approach has demonstrated to improve standard OPF classification results by pruning ensemble-based OPF using meta-heuristic optimization and
- the proposed approach provides a faster testing phase without significantly affecting the training time, since the HS approach and standard OPF have been considered equivalent.

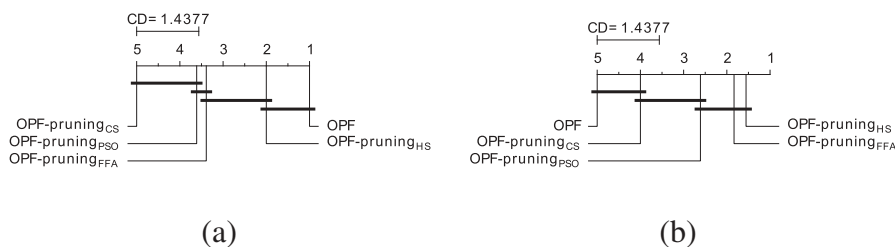


Figure 4. Results of the Nemenyi test for a comparison of standard OPF against OPF ensemble pruning using 9 classifiers and its variations under different optimization techniques based on the computational load for (a) training (training and validating) and (b) testing phases. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

4.2. Experiments over remote-sensing data sets

In this section, we employed the proposed approach in the context of remote-sensing image classification over five image data sets (Pisani et al. 2014; Pisani 2016; Romay 2016). The images were obtained from CBERS-2B CCD (20m), sensor (R2G3B4), and LANDSAT-5 Thematic Mapper (TM) (30m), sensor (R4G3B5), covering the area of Itatinga, São Paulo – Brazil, and images obtained from IKONOS-2 Multispectral (MS), sensor (R4G3B2), and GEOEYE, sensor (R5G4B3), covering the area of Duque de Caxias, Rio de Janeiro – Brazil. In addition, we considered well-known Indian Pines, which was obtained from AVIRIS sensor over the Indian Pines test site in North-western Indiana.

Figure 5 displays these images, being their respective ground truth versions illustrated in Figure 6. The main characteristics are presented in Table 3 and the ground truth classes for the images and their respective samples number are presented in Tables 4–8. Notice that IKONOS-2MS and GEOEYE 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. Roughly speaking, the idea is to use meta-heuristics to guide OPF ensemble pruning optimization in order to assess the robustness of the proposed approach under different performance scenarios.

In regard to the ranges of training, validating, and testing sets, each data set was partitioned into three subsets, the training (20%), validating (10%), and testing sets (70%), hereinafter denoted as 20:10:70. For each range, training, validating, and testing sets were selected randomly and the process was repeated 15 times (cross-validation). Notice the percentages have been empirically chosen, being more intuitive to provide a larger validating set for ensemble pruning learning, and concerning the ensemble strategy, we used nine classifiers. Furthermore, in regard to the meta-heuristic optimization, we opted to employ the same parameters and techniques used in the previous section.

Moreover, we also compared the standard OPF and OPF ensemble without pruning strategy. It is worth noting that these last two approaches were trained over $D_1 \cup D_v$, considering the aforementioned three stages. Table 9 presents the mean accuracies and standard deviation over all data sets, being the recognition rates computed according to Papa, Falcão, and Suzuki (2009), and Table 10 presents the mean *F*-measure values concerning the very same group of data sets. The values in bold stand for the most accurate techniques according of the Wilcoxon signed-rank test.

One can observe the proposed OPF ensemble pruning has obtained the best results in some data sets, except IKONOS-2 MS and Indian Pines data sets, which are better generalized by standard OPF, as it can be observed in Table 9. However, in regard to the *F*-measure statistical analysis, the proposed OPF ensemble pruning has obtained the best results in almost all data sets, except IKONOS-2MS data set, as it can be observed in Table 10. It is worth noting that all data sets used in this section are unbalanced, and is well-known that the *F*-measure statistical analysis is favoured over accuracy when we have an unbalanced data sets. Furthermore, the OPF pruning strategy of multiple classifiers may be more effective when applied to more complex problems, otherwise the results may show a low generalization of the data, as it can be observed in IKONOS-2MS data set.

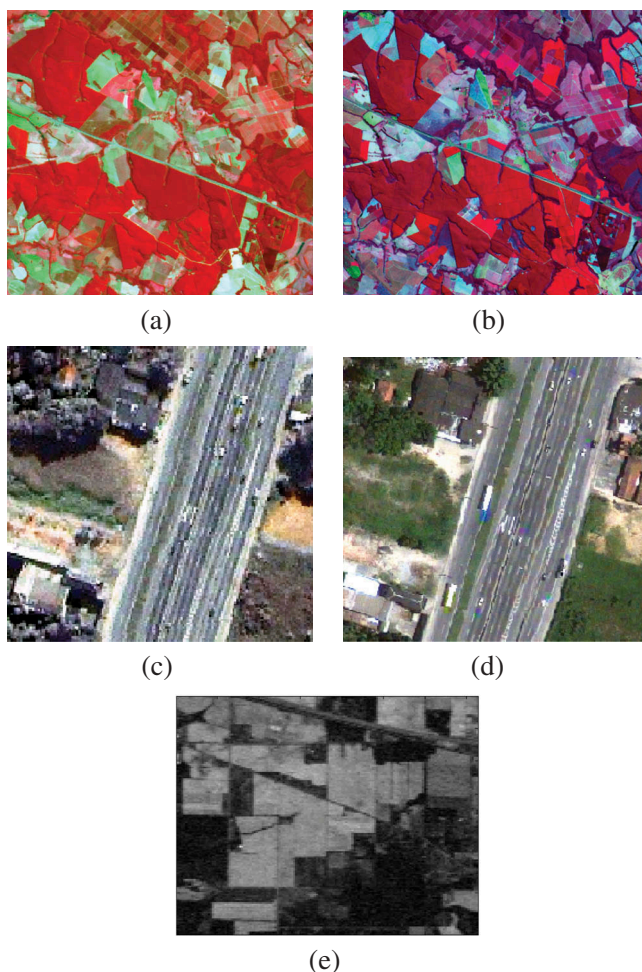


Figure 5. Satellite images used in the experiments: covering the area of Itatinga, São Paulo – Brazil, acquired by (a) CBERS-2B sensor and (b) LANDSAT-5 TM sensor, covering the area of Duque de Caxias, Rio de Janeiro – Brazil, acquired by (c) IKONOS-2 MS sensor, and (d) GEOEYE sensor, and (e) covering the area of Indian Pines test site in North-western Indiana.

Figure 7 depicts the statistical analysis considering the accuracy and F -measure values, respectively. As one can observe in Figure 7(a), the pruning-based approaches can be considered as the most accurate ones by Nemenyi test. Such point reflects that OPF ensemble pruning using meta-heuristic techniques achieved the best accuracy rates in the majority of the data sets. Furthermore, the HS approach can be considered the most accurate technique.

However, the statistical test did not point out a CD between meta-heuristics approaches and the ensemble without pruning and standard OPF, which means they performed similarly. In regard to the F -measure values (Figure 7(b)), the PSO and HS can be considered the most accurate techniques in the first group. In second group, we have the HS, CS, and FFA pruning strategies and ensemble without pruning. Lastly, the standard OPF as the least accurate compared to the others approaches.

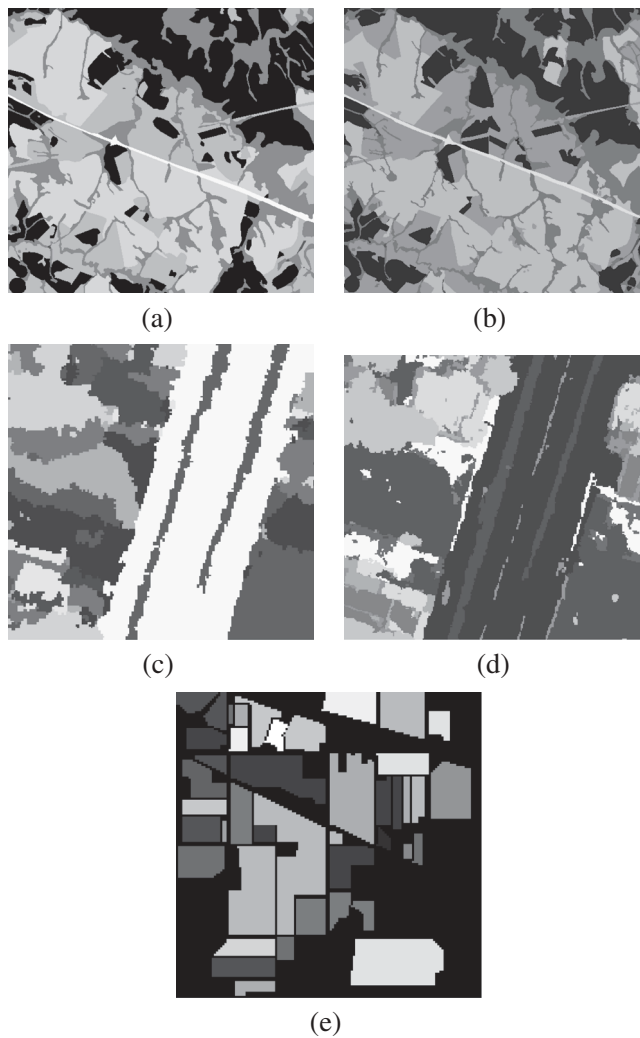


Figure 6. Labelled images used in the experiments: (a) and (b) correspond to the images displayed in Figure 5(a, and 5b), respectively, (c) and (d) correspond to images displayed in Figure 5(c, and 5d), respectively, and (e) corresponds to the images displayed in Figure 5(e).

Table 3. Description of the image data sets.

Data set	No. of samples (pixels)	No. of features	No. of classes
CBERS-2B	526×492	3	6
LANDSAT-5 TM	526×492	3	6
IKONOS-2 MS	258×250	3	9
GEOEYE	268×250	3	9
Indian Pines	145×145	200	17

Figure 8(a) depicts the computational load considering the training time (training and validating) with the Nemenyi test. Both PSO and FFA approaches showed similar results, wherewith there is no CD between them, and the CS approach is considered the slowest

Table 4. Ground truth classes for the CBERS-2B satellite.

No.	Class	No. of Samples
1	Plantations	71,434
2	Bushes	53,499
3	Dams	605
4	Grass lands	57,268
5	Reforested	72,866
6	Roads	3120

Table 5. Ground truth classes for the LANDSAT-5 TM satellite.

No.	Class	No. of samples
1	Plantations	62,327
2	Bushes	47,985
3	Grass lands	59,890
4	Reforested	85,189
5	Dams	464
6	Roads	2937

Table 6. Ground truth classes for the IKONOS-2MS satellite.

No.	Class	No. of samples
1	Tree cover	5914
2	Shadows	6481
3	Grass lands	12,054
4	Covering of dark tonality	3578
5	Roads	22,871
6	Bare soil – moist	4417
7	Bare soil – clear	7400
8	Covering of clear tonality	1738
9	Covering of average tonality	47

Table 7. Ground truth classes for the GEOEYE satellite.

No.	Class	No. of Samples
1	Bare soil – moist	2380
2	Tree cover	6132
3	Grass lands	19,370
4	Bare soil – clear	4490
5	Shadows	2822
6	Covering of dark tonality	5073
7	Roads	22,924
8	Covering of clear tonality	1026
9	Covering of average tonality	2783

one. On average, the OPF pruning using HS has been about 3.5327 times slower than standard OPF in training and validating phase.

In regard to the testing phase ([Figure 8\(b\)](#)), the fastest group consists of all approaches that use meta-heuristics, being PSO the fastest one, since it has pruned more classifiers compared to the other approaches. The last group (slowest techniques) concerns the standard OPF and OPF-ensemble. On average, the HS has been about 2.6457 times faster than OPF, i.e. fewer classifiers were used for the testing phase compared to the original ensemble. It is worth noting that HS approach has demonstrated interesting results in the statistical test ([Figure 7](#)) according to the recognition accuracies, and showed a good efficiency in the training and testing phases.

Table 8. Ground truth classes for the Indian Pines satellite.

No.	Class	No. of samples
1	Background	10,776
2	Alfalfa	46
3	Corn-notill	1428
4	Corn-mintill	830
5	Corn	237
6	Grass-pasture	483
7	Grass-trees	730
8	Grass-pasture-mowed	28
9	Hay-windrowed	478
10	Oats	20
11	Soybean-notill	972
12	Soybean-mintill	2455
13	Soybean-clean	593
14	Wheat	205
15	Woods	1265
16	Buildings-Grass-Trees-Drives	386
17	Stone-Steel-Towers	93

Table 9. Mean accuracy results (%) and standard deviation over all data sets for standard OPF and OPF ensemble pruning under different optimization techniques.

Approach	Data set				
	CBERS-2B	GEOEYE	IKONOS-2MS	LANDSAT-5 TM	Indian Pines
OPF	64.28 \pm 0.76	71.40 \pm 1.68	77.85 \pm 1.07	60.82 \pm 0.78	60.22 \pm 0.38
OPF-ensemble	73.28 \pm 1.02	73.89 \pm 1.30	70.01 \pm 0.62	71.35 \pm 0.55	56.14 \pm 0.12
OPF-pruning _{CS}	73.72 \pm 0.95	74.95 \pm 1.59	68.87 \pm 0.62	72.17 \pm 0.48	56.02 \pm 0.33
OPF-pruning _{FFA}	73.26 \pm 1.93	75.17 \pm 1.57	69.02 \pm 0.71	72.02 \pm 0.39	56.37 \pm 0.84
OPF-pruning _{HS}	73.93 \pm 0.40	75.24 \pm 1.50	69.22 \pm 0.55	72.16 \pm 0.63	56.10 \pm 0.10
OPF-pruning _{PSO}	73.73 \pm 0.59	75.06 \pm 1.58	68.97 \pm 0.73	72.15 \pm 0.47	56.21 \pm 0.20

The most accurate techniques for the Wilcoxon test are highlighted in bold.

Table 10. Mean *F*-measure values and standard deviation over all data sets for standard OPF and OPF ensemble pruning under different optimization techniques.

Approach	Data set				
	CBERS-2B	GEOEYE	IKONOS-2MS	LANDSAT-5 TM	Indian Pines
OPF	0.4186 \pm 0.02	0.5947 \pm 0.01	0.6766 \pm 0.00	0.4139 \pm 0.03	0.4114 \pm 0.06
OPF-ensemble	0.6210 \pm 0.02	0.7053 \pm 0.00	0.6206 \pm 0.00	0.6809 \pm 0.02	0.4571 \pm 0.03
OPF-pruning _{CS}	0.6365 \pm 0.02	0.7027 \pm 0.01	0.6021 \pm 0.00	0.7112 \pm 0.02	0.4599 \pm 0.00
OPF-pruning _{FFA}	0.6311 \pm 0.04	0.7035 \pm 0.01	0.6020 \pm 0.01	0.7045 \pm 0.01	0.4613 \pm 0.01
OPF-pruning _{HS}	0.6430 \pm 0.01	0.7040 \pm 0.01	0.6046 \pm 0.00	0.7087 \pm 0.02	0.4416 \pm 0.03
OPF-pruning _{PSO}	0.6405 \pm 0.01	0.7042 \pm 0.01	0.6016 \pm 0.00	0.7122 \pm 0.01	0.4619 \pm 0.00

The most accurate techniques for the Wilcoxon test are highlighted in bold.

Figures 9–13 depict some images classified with standard OPF, OPF-ensemble, and OPF pruning using HS (approach considered the most accurate in the majority of data sets in accuracy results), considering CBERS-2B, LANDSAT-5 TM, IKONOS-2 MS, GEOEYE, and Indian Pines, respectively. In addition, the aforementioned figures also show the confusion matrices concerning these data sets, which clarify that both ensemble of OPF classifiers and their pruned versions can improve the recognition results. Clearly, the results using meta-heuristic techniques and OPF-ensemble were much more accurate than standard OPF when taking into account the ground-truth images (Figure 6), except for Figures 11 and 13 (IKONOS-2 MS and Indian Pines data sets), where the standard OPF was more effective.

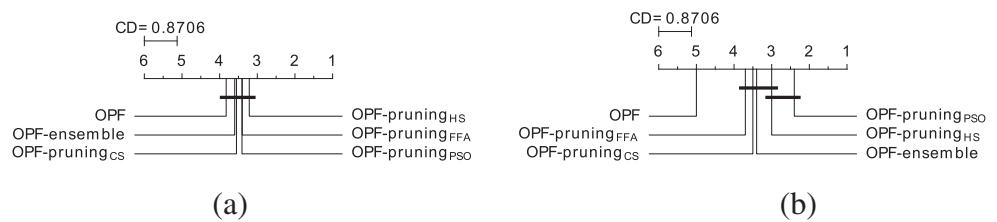


Figure 7. Results of the Nemenyi test for a comparison of standard OPF against OPF ensemble pruning using 9 classifiers and its variations under different optimization techniques based on the (a) accuracy results and (b) *F*-measure values over all image data sets. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

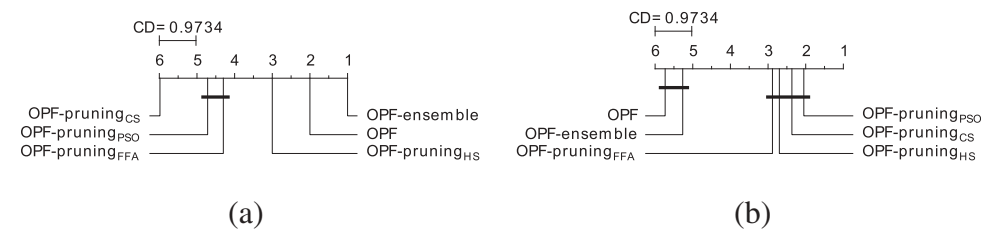


Figure 8. Results of the Nemenyi test for a comparison of standard OPF against OPF ensemble pruning using 9 classifiers and its variations under different optimization techniques based on the computational load for: (a) training (training and validating) and (b) testing phases. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

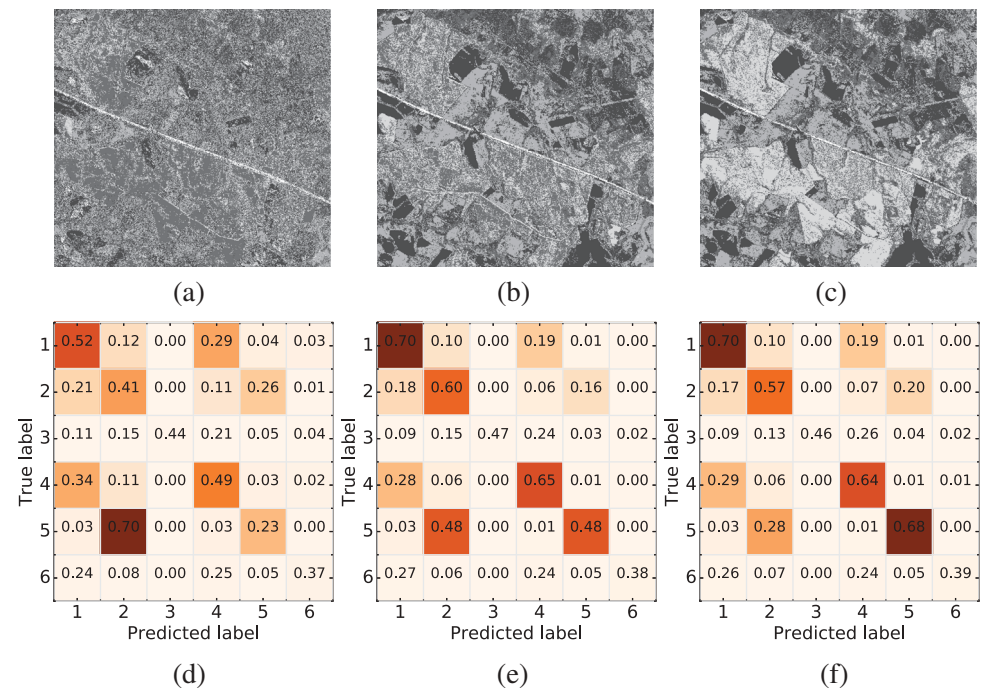


Figure 9. Classified CBERS-2B satellite images obtained using (a) standard OPF, (b) OPF-ensemble and (c) OPF pruning using HS, and parts (d), (e), and (f) are the corresponding confusion matrices.

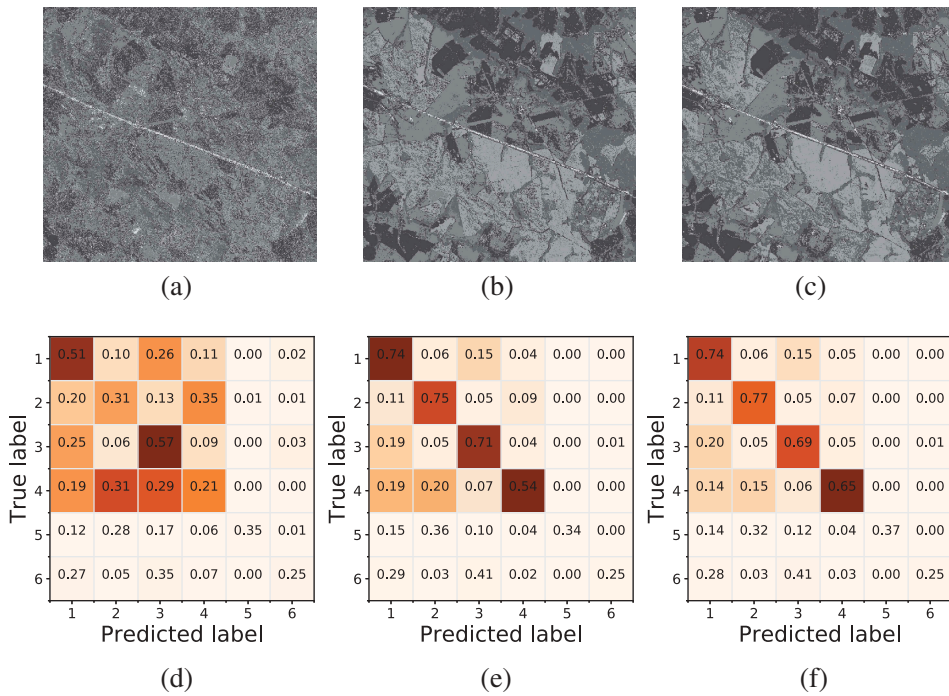


Figure 10. Classified LANDSAT-5 TM satellite images obtained using (a) standard OPF, (b) OPF-ensemble and (c) OPF pruning using HS, and parts (d), (e), and (f) are the corresponding confusion matrices.

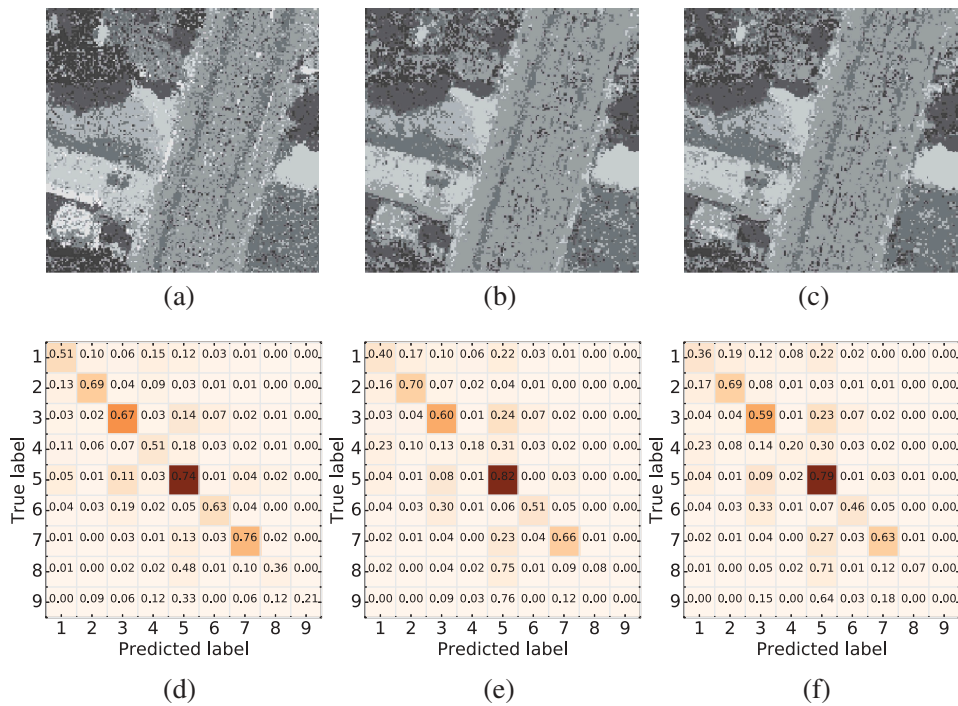


Figure 11. Classified IKONOS-2 MS satellite images obtained using (a) standard OPF, (b) OPF-ensemble and (c) OPF pruning using HS, and parts (d), (e), and (f) are the corresponding confusion matrices.

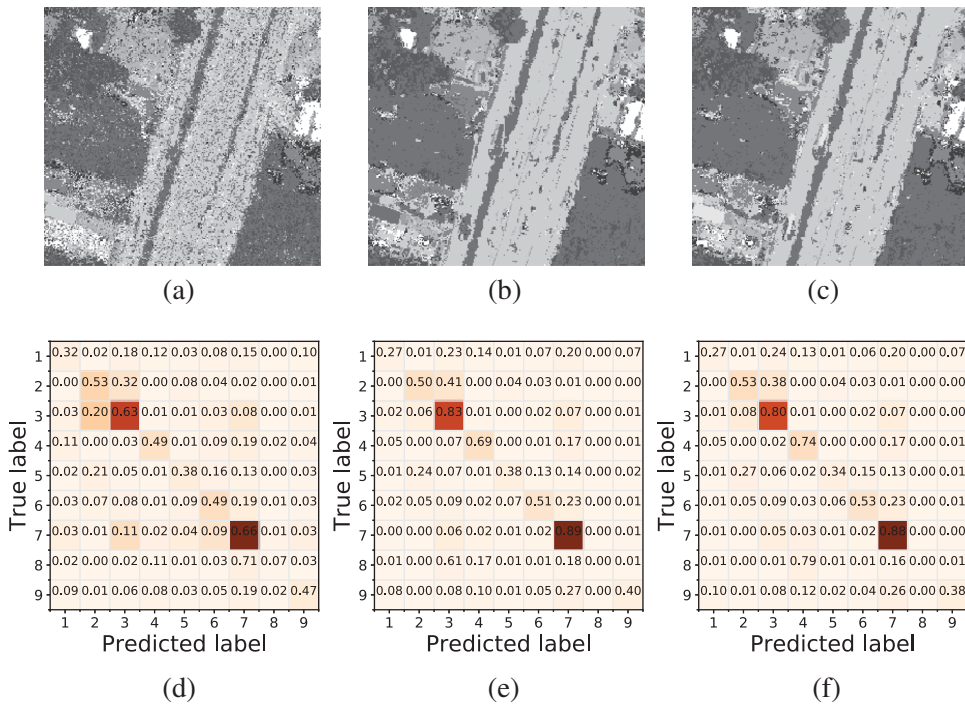


Figure 12. Classified GEOEYE satellite images obtained using (a) standard OPF, (b) OPF-ensemble and (c) OPF pruning using HS, and parts (d), (e), and (f) are the corresponding confusion matrices.

In order to obtain some insight about the representation of the data sets, we employed the Andrews curve method (Andrews 1972), which represents feature spaces by means of finite Fourier series. The transformation has to maintain some inherent properties of the data, thus making possible to identify some behaviours (Kozioł and Hacke 1991). Each line in this plot stands for a sample, and the variation in colours corresponds to a given class. Figure 14 shows the structure of the data for the IKONOS-2 MS and CBERS-2B data sets.

As we can observe in Figure 14(a), patterns from different classes are somehow separated. Note that IKONOS-2MS is the data set with the best results concerning the standard OPF approach. In addition, Figure 14(b) shows a more complex pattern, being better classified by OPF pruning. Considering Figure 14, we can establish that more complex data sets and with considerable amount of overlapping among samples from different classes may be better generalized by OPF pruning using meta-heuristics due to the near optimal sub-ensemble provided by the optimization process.

4.2.1. Empirical comparison between SVM and OPF ensemble pruning

Additionally, we compared the proposed approach against naïve SVM and SVM using the very same ensemble pruning strategy proposed in this article. The idea is to perform the very same procedure using metaheuristic optimization process for SVM with radial basis function. In order to fulfil this purpose, we apply the same procedure, i.e. 20:10:70 range for training, validating, and testing sets, respectively, repeated 15 times (cross-validation). The SVM parameters were optimized through a cross-validation procedure

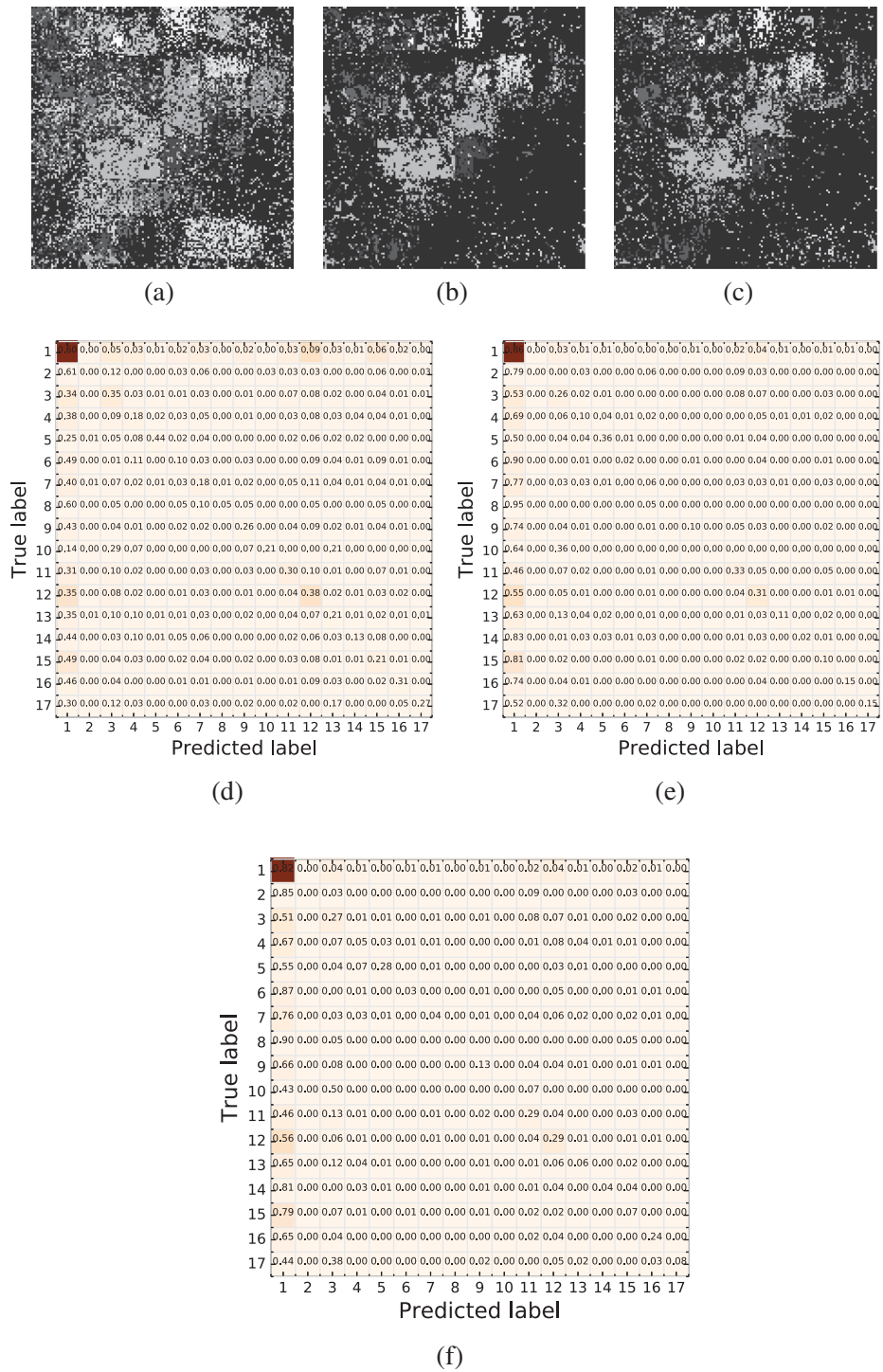


Figure 13. Classified Indian Pines satellite images obtained using (a) standard OPF, (b) OPF-ensemble and (c) OPF pruning using HS, and parts (d), (e), and (f) are the corresponding confusion matrices.

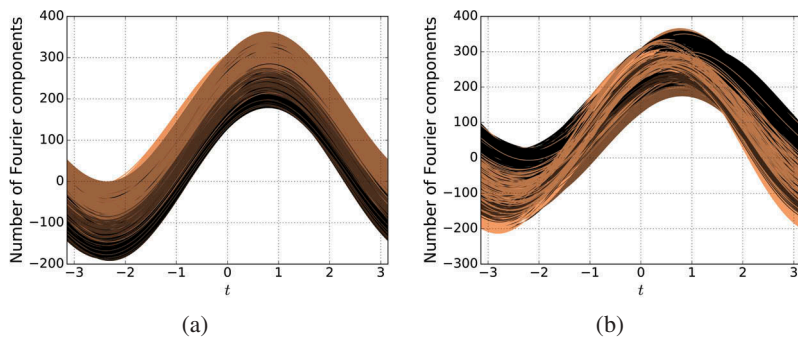


Figure 14. Andrews plot for (a) IKONOS-2 MS and (b) CBERS-2B images in the range of $-3.14 < t < 3.14$.

over a validation set using a grid-search ($\gamma \in \{0.001, 0.01, 0.1, 1\}$ and $C \in \{1, 10, 100, 1000\}$). In regard to the metaheuristic techniques, we considered SVM pruning using HS only, since the HS approach was considered the most accurate in the majority of data sets. The HS parameters for SVM ensemble pruning were the same used in Section 4.1. Tables 11 and 12 present the mean accuracy and F -measure results, respectively. The values in bold stand for the most accurate techniques according to the Wilcoxon signed-rank test.

We can observe the proposed OPF ensemble pruning strategy applied in the SVM approach did not reflect significant results. On the other hand, the SVM without ensemble pruning was the most accurate one considering the mean accuracies, as one can observe in Figure 15(a), which depicts the Nemenyi test. However, the F -measure values (Table 12) showed that the proposed OPF ensemble pruning using HS has obtained the best results in all data sets against standard SVM and SVM with ensemble pruning, as it can be observed by the Nemenyi test in Figure 15(b). The poor

Table 11. Mean accuracy results (%) and standard deviation over all data sets for both OPF and SVM ensemble pruning using 9 classifiers and standard SVM approach.

Approach	Data set				
	CBERS-2B	GEOEYE	IKONOS-2 MS	LANDSAT-5 TM	Indian Pines
OPF-pruning _{HS}	73.93 \pm 0.40	75.24 \pm 1.50	69.22 \pm 0.55	72.16 \pm 0.63	56.10 \pm 0.10
SVM-pruning _{HS}	76.15 \pm 0.48	76.38 \pm 0.24	68.28 \pm 0.33	73.92 \pm 0.72	50.00 \pm 0.00
SVM	76.77 \pm 0.30	76.51 \pm 0.13	68.64 \pm 0.48	74.58 \pm 0.37	50.00 \pm 0.00

The most accurate techniques for the Wilcoxon test are highlighted in bold.

Table 12. Mean F -measure values for both OPF and SVM ensemble pruning using 9 classifiers and standard SVM approach.

Approach	Data set				
	CBERS-2B	GEOEYE	IKONOS-2 MS	LANDSAT-5 TM	Indian Pines
OPF-pruning _{HS}	0.6430 \pm 0.01	0.7040 \pm 0.01	0.6046 \pm 0.00	0.7087 \pm 0.02	0.4416 \pm 0.03
SVM-pruning _{HS}	0.6311 \pm 0.01	0.5884 \pm 0.00	0.4307 \pm 0.01	0.5449 \pm 0.01	0.0399 \pm 0.00
SVM	0.6421 \pm 0.00	0.5911 \pm 0.00	0.4384 \pm 0.01	0.5580 \pm 0.01	0.0399 \pm 0.00

The most accurate techniques for the Wilcoxon test are highlighted in bold.

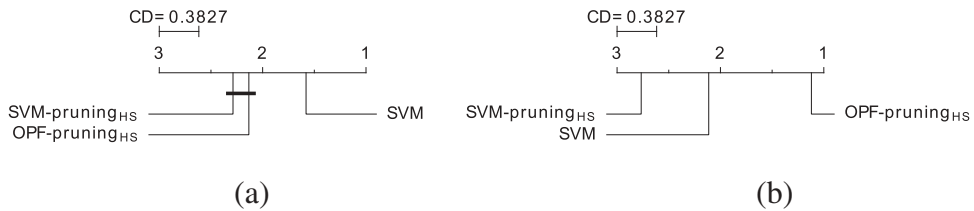


Figure 15. Results of the Nemenyi test for a comparison of both OPF and SVM ensemble pruning using 9 classifiers and standard SVM approach base on the (a) accuracy results and (b) *F*-measure values over all image data sets. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

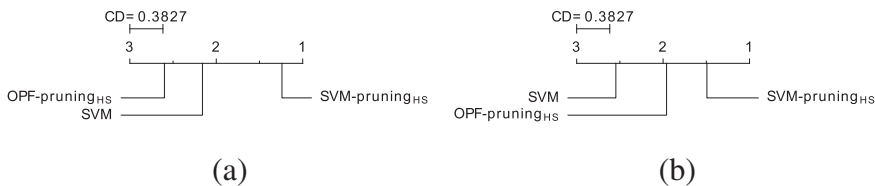


Figure 16. Confusion matrices for SVM without ensemble pruning for (a) CBERS-2B, (b) LANDSAT-5 TM, (c) IKONOS-2 MS, (d) GEOEYE, and (e) Indian Pines data sets, respectively.

generalization in IKONOS-2MS and Indian Pines data sets may be due to the low complexity concerning these data sets, with a low degree of overlapping among samples from different classes, as well as due to the few number of samples. Figure 16 shows the confusion matrix concerning all images data sets for the SVM without ensemble pruning.

In regard to the computational load, Figure 17(a,b) shows the training time (training and validating) and testing time with Nemenyi test, respectively. The training phase using disjoint subsets, which is required by the pruning strategy, has shown to be more efficient when compared with a training procedure over all sub-ensembles for a single classifier (standard SVM training), as stated in Figure 17(a) concerning SVM with ensemble pruning using HS and SVM without pruning. Figure 17(b) shows the testing phase with the proposed approach placed in second, and SVM without pruning as the slowest technique.

5. Conclusions

Remote-sensing image classification has been widely studied in the last years, mainly due to the complexity of the data. In this work, we presented a methodology to build an ensemble pruning strategy in OPF-based classifiers using metaheuristic optimization, which is useful for finding a near optimal sub-ensemble concerning a given application. Roughly speaking, the idea is to build a reduced ensemble of OPF classifiers using metaheuristic-based optimization algorithms, such as HS, PSO, CS, and FFA. The applications were represented by unbalanced data sets, distinct number of classes, and feature space dimensions.

Two main experiments were performed. Firstly, the proposed methodology exploits the learning and generalization capability of ensemble pruning in data sets with

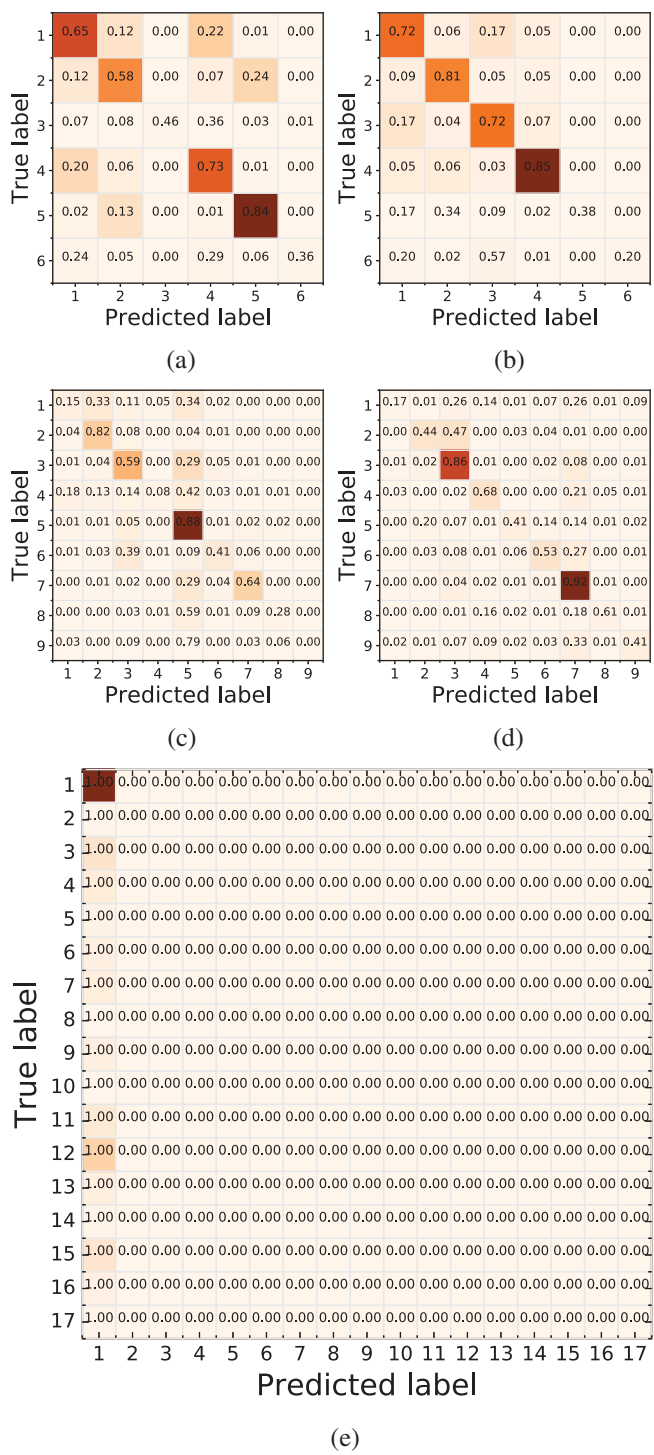


Figure 17. Results of the Nemenyi test for a comparison of both OPF and SVM ensemble pruning using 9 classifiers and standard SVM approach based on the computational load for: (a) training (training and validating) and (b) testing phases over all image data sets. Groups that are not significantly different (at $p = 0.05$) are connected to each other.

reduced samples and low complexity. For this purpose, we build a set of ensembles with 3, 5, 7, and 9 classifiers to investigate the impact of different layout of classifiers. Statistical tests showed that even with a larger number of classifiers (i.e. nine techniques), the time training would not be affected deeply. In addition, the proposed approach achieved good efficiency in the testing time, since it pruned more classifiers with respect to the original ensemble. It is worth noting that even though all metaheuristic techniques have been considered similarly with respect to the effectiveness, the HS approach can be considered the most accurate one.

In the second experiment, we investigate the proposed approach in the context of remote sensing image classification in order to show it can obtain reasonable effectiveness and efficiency when applied to more complex problems, which require a longer time to find the best sub-ensemble. For this purpose, we build a set of nine classifiers, since it achieved the best results in the majority voting strategy, and the statistical tests showed a good trade-off between efficiency and effectiveness. Lastly, we performed an empirical comparison between SVM and OPF using the same strategy applied in the second experiment. Note that the advantage of the proposed approach can be better observed when applied to complex problems with highly overlapped samples.

Therefore, the experiments over real and synthetic data sets showed the robustness of the proposed approach, which obtained the best results in almost all data sets. The proposed approach has demonstrated to improve standard OPF classification results, as well as HS has shown to hold the best trade-off between effectiveness and efficiency in pruning ensembles of OPF-based classifiers. Furthermore, experiments comparing the well-known SVM classifier show that OPF ensemble pruning can be effective without degrading the performance when applied to large data sets. Also, further work is needed to address some important issues related to the SVM using ensemble pruning, such as different kernels and other strategies. Finally, it is important to note that the free parameters in OPF classifier and the effectiveness in pruning ensembles can be applied in several problems with a good data generalization.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This authors are grateful to FAPESP grants #2014/16250-9 and #2014/12236-1, CNPq grant #306166/2014-3, as well as CAPES grant.

References

- Abe, B., A. Gidudu, and T. Marwal. 2010. "Investigating the Effects of Ensemble Classification on Remotely Sensed Data for Land Cover Mapping." In *Geoscience and Remote Sensing Symposium (IGARSS)*, 2010 *IEEE International*, 2832–2835.
- Al-Ani, A., and M. Deriche. 2002. "A New Technique for Combining Multiple Classifiers Using the Dempster-Shafer Theory of Evidence." *Journal of Artificial Intelligence Research* 17 (1): 333–361.

- Allène, C., J.-Y. Audibert, M. Couprie, and R. Keriven. 2010. "Some Links between Extremum Spanning Forests, Watersheds and Min-Cuts." *Image and Vision Computing* 28 (10): 1460–1471. doi:10.1016/j.imavis.2009.06.017.
- Andrews, D. F. 1972. "Plots of High-Dimensional Data." *Biometrics* 28: 1. doi:10.2307/2528964.
- Bauer, E., and R. Kohavi. 1999. "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants." *Machine Learning* 36 (1–2): 105–139. doi:10.1023/A:1007515423169.
- Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24 (2): 123–140. doi:10.1007/BF00058655.
- Breiman, L. 1998. "Arcing Classifiers." *Annals of Statistics* 26 (3): 801–824.
- Caruana, R., and A. Niculescu-Mizil. 2006. "An Empirical Comparison of Supervised Learning Algorithms." In *Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, Pennsylvania, USA. ICML '06*, 161–168. New York, NY: ACM.
- Chandrashekar, G., and F. Sahin. 2014. A Survey on Feature Selection Methods. *Computers & Electrical Engineering* 40 (1): 16–28. 40th-year commemorative issue. doi:10.1016/j.compeleceng.2013.11.024.
- Chang, C.-C., and C.-J. Lin. 2016. "LIBSVM Data: Classification, Regression, and Multi-Label." Accessed 29 June. <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>.
- Coletta, L. F. S., E. R. Hruschka, A. Acharya, and J. Ghosh. 2015. "Using Metaheuristics to Optimize the Combination of Classifier and Cluster Ensembles." *Integrated Computer-Aided Engineering* 22 (3): 229–242. doi:10.3233/ICA-150485.
- Diao, R., and Q. Shen. 2011. "Fuzzy-Rough Classifier Ensemble Selection." *IEEE International Conference on Fuzzy Systems* 1516–1522.
- Dietterich, T. 2000. "Ensemble Methods in Machine Learning." In *Multiple Classifier Systems, Vol. 1857 of Lecture Notes in Computer Science*, edited by Josef Kittler and Fabio Roli, 1–15. London, UK: Springer-Verlag.
- Geem, Z. W. 2009. *Music-Inspired Harmony Search Algorithm: Theory and Applications*. 1st ed. Berlin Heidelberg: Springer-Verlag.
- Ho, T. K. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–844. doi:10.1109/34.709601.
- Jia, L., Y. Wang, and L. Fan. 2014. "Multiobjective Bilevel Optimization for Production-Distribution Planning Problems Using Hybrid Genetic Algorithm." *Integrated Computer-Aided Engineering* 21 (1): 77–90.
- Jodavi, M., M. Abadi, and E. Parhizkar. 2015. "JSObfusDetector: A Binary PSO-based One-Class Classifier Ensemble to Detect Obfuscated JavaScript Code." *International Symposium on Artificial Intelligence and Signal Processing* 21: 322–327.
- Joly, M. M., T. Verstraete, and G. Paniagua. 2014. "Integrated Multifidelity, Multidisciplinary Evolutionary Design Optimization of Counterrotating Compressors." *Integrative Computer-Aided Engineering* 21 (3): 249–261.
- Kennedy, J., and R. C. Eberhart. 2001. *Swarm Intelligence*. M. Kaufman.
- Kozioł, J. A., and W. Hacke. 1991. "A Bivariate Version of Andrews Plots." *IEEE Transactions on Biomedical Engineering* 38 (12): 1271–1274. doi:10.1109/10.137294.
- Krawczyk, B. 2015. "One-Class Classifier Ensemble Pruning and Weighting with Firefly Algorithm." *Neurocomputing* 150 (Part B): 490–500. doi:10.1016/j.neucom.2014.07.068.
- Kuncheva, L. I. 2002. "A Theoretical Study on Six Classifier Fusion Strategies." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2): 281–286. doi:10.1109/34.982906.
- Kuncheva, L. I. 2004. *Combining Pattern Classifiers: Methods and Algorithms*. New Jersey: Wiley-Interscience.
- Kuncheva, L. I., M. Skurichina, and R. P. W. Duin. 2002. "An Experimental Study on Diversity for Bagging and Boosting with Linear Classifiers." *Information Fusion* 3 (4): 245–258. doi:10.1016/S1566-2535(02)00093-3.
- Langdon, W. B., S. J. Barrett, and B. F. Buxton. 2002. "Combining Decision Trees and Neural Networks for Drug Discovery." In *Genetic Programming, Vol. 2278 of Lecture Notes in Computer*

- Science*, edited by J. A. Foster, E. Lutton, J. Miller, C. Ryan, and A. Tettamanzi, 60–70. Germany: Springer Berlin Heidelberg.
- Larkins, R., and M. Mayo. 2008. "Adaptive Feature Thresholding for Off-Line Signature Verification." In *Image and Vision Computing New Zealand*, 2008. IVCNZ 2008. 23rd International Conference, 1–6. Piscataway, NJ: IEEE.
- Lee, B., S. Choi, B. Oh, J. Yang, and S. Park. 2013. "A New Ensemble Learning Algorithm Using Regional Classifiers." *International Journal on Artificial Intelligence Tools* 22 (04): 1350025. doi:10.1142/S0218213013500255.
- Lee, D.-S., and S. N. Srihari. 1995. "A Theory of Classifier Combination: The Neural Network Approach." *Proceedings of the Third International Conference on Document Analysis and Recognition* 1: 42–45.
- Li, B., Q. Lin, Q.-H. M. Max, and Y. Fan. 2009. "Using Ensemble Classifier for Small Bowel Ulcer Detection in Wireless Capsule Endoscopy Images." *IEEE International Conference on Robotics and Biomimetics* 2326–2331.
- Li, C.-F., and J.-Y. Yin. 2013. Variational Bayesian Independent Component Analysis-Support Vector Machine for Remote Sensing Classification. *Computers & Electrical Engineering* 39 (3): 717–726. Special issue on Image and Video Processing/Special issue on Recent Trends in Communications and Signal Processing. doi:10.1016/j.compeleceng.2012.10.004.
- Lichman, M. 2013. "UCI Machine Learning Repository." Accessed 21 April. <http://archive.ics.uci.edu/ml>.
- Macas, M., B. Gabrys, D. Ruta, and L. Lhotska. 2007. "Particle Swarm Optimization of Multiple Classifier Systems." In *Computational and Ambient Intelligence, Vol. 4507 of Lecture Notes in Computer Science*, edited by F. Sandoval, A. Prieto, J. Cabestany, and G. Manuel, 333–340. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Mao, S., L. C. Jiao, L. Xiong, and S. Gou. 2011. "Greedy Optimization Classifiers Ensemble Based on Diversity." *Pattern Recognition* 44 (6): 1245–1261. doi:10.1016/j.patcog.2010.11.007.
- Markatopoulou, F., G. Tsoumakas, and I. Vlahavas. 2010. "Instance-Based Ensemble Pruning via Multi-Label Classification." *22nd IEEE International Conference on Tools with Artificial Intelligence* 1: 401–408.
- Martínez-Muñoz, G., D. Hernández-Lobato, and A. Suárez. 2009. "An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation." *IEEE Tpami* 31 (2): 245–259. doi:10.1109/TPAMI.2008.78.
- Mendoza, L. F., M. Vellasco, and K. Figueiredo. 2014. "Intelligent Multiagent Coordination Based on Reinforcement Hierarchical Neuro-Fuzzy Models." *International Journal of Neural Systems* 24 (08): 1450031. doi:10.1142/S0129065714500312.
- Nabavi-Kerizi, S. H., M. Abadi, and E. Kabir. 2010. "A PSO-based Weighting Method for Linear Combination of Neural Networks." *Computers & Electrical Engineering* 36 (5): 886–894. doi:10.1016/j.compeleceng.2008.04.006.
- Nemenyi, P. 1963. *Distribution-Free Multiple Comparisons*. Princeton University.
- Osaku, D., R. Y. M. Nakamura, L. A. M. Pereira, R. J. Pisani, A. L. M. Levada, F. A. M. Cappabianco, A. X. Falcão, and J. P. Papa. 2015. "Improving Land Cover Classification through Contextual-Based Optimum-Path Forest." *Information Sciences* 324: 60–87. doi:10.1016/j.ins.2015.06.020.
- Osaku, D., D. R. Pereira, A. L. M. Levada, and J. P. Papa. 2016. "Fine-Tuning Contextual-Based Optimum-Path Forest for Land-Cover Classification." *IEEE Geoscience and Remote Sensing Letters* 13 (5): 735–739. doi:10.1109/LGRS.2016.2541458.
- Papa, J. P., S. E. N. Fernandes, and A. X. Falcão. 2016. "Optimum-Path Forest Based on K-Connectivity: Theory and Applications." *Pattern Recognition Letters* 86: 117–126.
- Papa, J. P., and A. X. Falcão. 2008. "A New Variant of the Optimum-Path Forest Classifier." In *Proceedings of the 4th International Symposium on Advances in Visual Computing, Lecture Notes in Computer Science*, 935–944. Springer Berlin Heidelberg.
- Papa, J. P., A. X. Falcão, V. H. C. Albuquerque, and J. M. R. S. Tavares. 2012. "Efficient Supervised Optimum-Path Forest Classification for Large Datasets." *Pattern Recognition* 45 (1): 512–520. doi:10.1016/j.patcog.2011.07.013.

- Papa, J. P., A. X. Falcão, and C. T. N. Suzuki. 2009. "Supervised Pattern Classification Based on Optimum-Path Forest." *International Journal of Imaging Systems and Technology* 19 (2): 120–131. doi:10.1002/ima.v19:2.
- Papa, J. P., G. H. Rosa, D. Rodrigues, and X.-S. Yang. 2017. "LibOPT: An Open-Source Platform for Fast Prototyping Soft Optimization Techniques." *ArXiv e-prints*.
- Papa, J. P., C. T. N. Suzuki, and A. X. Falcão. 2009. "LibOPF: A Library for the Design of Optimum-Path Forest Classifiers." Accessed 21 April. <https://github.com/jppbsi/LibOPF>.
- Pisani, R. J. 2016. "Remote Sensing Datasets." Accessed 21 April. http://wwwwp.fc.unesp.br/papa/recogna/remote_sensing.html.
- Pisani, R. J., R. Y. M. Nakamura, P. S. Riedel, C. R. L. Zimback, A. X. Falcão, and J. P. Papa. 2014. "Toward Satellite-Based Land Cover Classification Through Optimum-Path Forest." *IEEE Transactions on Geoscience and Remote Sensing* 52 (10): 6075–6085. doi:10.1109/TGRS.2013.2294762.
- Ponti, M. P., and J. P. Papa. 2011. "Improving Accuracy and Speed of Optimum-Path Forest Classifier Using Combination of Disjoint Training Subsets." In *Multiple Classifier Systems, Vol. 6713 of Lecture Notes in Computer Science*, edited by Sansone C., J. Kittler and F. Roli, 237–248.
- Ponti, M. P., J. P. Papa, and A. L. M. Levada. 2011. "A Markov Random Field Model for Combining Optimum-Path Forest Classifiers Using Decision Graphs and Game Strategy Approach." In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Vol. 7042 of Lecture Notes in Computer Science*, edited by San Martin C. and S. W. Kim, 581–590.
- Quinlan, J. R. 1996. "Bagging, Boosting, and C4.5." In *Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 1*, 725–730. Portland, Oregon: AAAI Press.
- Rätsch, G., T. Onoda, and K.-R. Müller. 2001. "Soft Margins for AdaBoost." *Machine Learning* 42 (3): 287–320. doi:10.1023/A:1007618119488.
- Romay, M. G. 2016. "Hyperspectral Remote Sensing Scenes." Accessed 21 April. http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.
- Sheen, S. S., V. Aishwarya, R. Anitha, S. V. Raghavan, and S. M. Bhaskar. 2012. *Hybrid Artificial Intelligent Systems: 7th International Conference, HAIS 2012, Salamanca, Spain, March 28-30th, 2012. Proceedings, Part II*. Chap. Ensemble Pruning Using Harmony Search, 13–24. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sonnenburg, S. 2016. "Machine Learning Data Set Repository." Accessed 21 April. <http://mldata.org>.
- Tinoco, S. L. J. L., H. G. Santos, D. Menotti, A. B. Santos, and J. A. Dos Santos. 2013. "Ensemble of Classifiers for Remote Sensed Hyperspectral Land Cover Analysis: An Approach Based on Linear Programming and Weighted Linear Combination." In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, 4082–4085. Piscataway, NJ: IEEE.
- Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1 (6): 80–83. doi:10.2307/3001968.
- Xu, L., A. Krzyzak, and C. Y. Suen. 1992. "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition." *IEEE Transactions on Systems, Man, and Cybernetics* 22 (3): 418–435. doi:10.1109/21.155943.
- Yang, X. S. 2010. "Firefly Algorithm, Stochastic Test Functions and Design Optimisation." *International Journal of Bio-Inspired Computation* 2 (2): 78–84. doi:10.1504/IJBIC.2010.032124.
- Yang, X. S., and S. Deb. 2010. "Engineering Optimisation by Cuckoo Search." *International Journal of Mathematical Modelling and Numerical Optimisation* 1: 330–343. doi:10.1504/IJMMNO.2010.035430.
- Zhou, Z.-H., J. Wu, and W. Tang. 2002. "Ensembling Neural Networks: Many Could be Better than All." *Artificial Intelligence* 137 (1–2): 239–263. doi:10.1016/S0004-3702(02)00190-X.