COMBINATION TECHNIQUES FOR HYPERSPECTRAL IMAGE INTERPRETATION

A. B. Santos, A. de A. Araújo,
J. A. dos Santos, W. R. Schwartz

UFMG - Federal University of Minas Gerais
Computer Science Department
Belo Horizonte, MG, Brazil

D. Menotti

UFPR - Federal University of Paraná
Informatics Department
Curitiba, PR, Brazil

ABSTRACT

In this work, we propose two main contributions to hyperspectral image interpretation. Firstly, while the traditional Weighted Linear Combination optimized by Genetic Algorithms (WLC-GA) [1] intends to give more discriminant power to those classification approaches contributing the most, we extend it to make a fine tuning over the class probabilities within the combination process. Then, we compare both methods (WLC-GA and its extension) with a more complex non-linear meta learning strategy called *Stacked Generalization* in which *Support Vector Machines* with *Radial Basis Function* kernel was used as combiner [2]. The experimental results, considering two widely used data sets, the *Indian Pines* and the *Pavia University*, are conducted in three different scenarios. Results show that both WLC-GA and its extended version achieve the best overall accuracy, and the proposed classification approach overcomes the accuracies of the other traditional ones used in this study.

***Index Terms*—** Hyperspectral image; multiple classification systems; consensual combination; stacked generalization; genetic algorithms.

1. INTRODUCTION

Remote sensed hyperspectral images have been used for many purposes and have become one of the most important tools in remote sensing.Typically, hyperspectral images are acquired using an airborne or satellite sensor at short, medium or long distance [3] and the analysis and exploitation of such data can bring many advances to the field of land cover classification such as urban planning, ecology, precision agriculture, forestry and military [4].

Aiming at producing a thematic map as accurate as possible, this study is based on the idea of multi-source classification, exploiting different strategies to classify the same data by using different feature representations and different learning algorithms. This way, we expect to build a more reliable classification system, therefore producing more accurate thematic maps.

In this work, four classification approaches are used in the combination process and are not formally presented here due to space constraints. Two are based only on spectral information and other two based on spectral-spatial information. Three of them come from the literature (Pixelwise-SVM [3], FEGA-kNN [5], and EMP-MLP [6]), and one, spectral-spatial based, is a work of us [7].

This study is carried out through the combination of multiple classification methods by means of consensual methods and the Stacked Generalization idea. We apply two different linear combinations to perform the consensual fusion or combination of classification approach. In the first case, we use a single set of values for weighting the i-th classification approach according to its contribution in the combination process (this version appears in [8, 1]). In the second case, each classification approach has its own set of weights for each class in order do adjust the confidence value of j-th class for each i-th classification approach. This second fusion scheme can be seen as an extension of [8, 1] and is a contribution of this work. In both cases, we apply *Genetic Algorithms* (GA) to optimize the set of weights. A non-linear combination is also investigated using the Stacked Generalization procedure, which is performed by the SVM learning algorithm using the RBF kernel. The results obtained on two widely used data sets (*Indian Pines* and *Pavia University* data sets) in three different scenarios show that the linear combinations using GA optimization have the best accuracy indexes.

2. COMBINATION METHODS

The use of learning algorithms that produce a set of probabilities, certainties, support degrees or fuzzy values as output has proven to be more suitable for post processing [9]. Such estimate is known as *soft* output [10] and provides a degree of confidence as posterior probability of a given sample belongs to a specific category, which provides more information to combine multiple learners. There are several methods to combine classifiers in the technical literature [10, 9]. We make use of a structure called *Decision Profile* (DP). A DP is a $L \times C$ matrix: $DP(x) = [D_1(x), D_2(x), ...D_L(x)]^T$, in which L is the number of classifiers and C the number of categories or classes. In this scheme, each row composes a set of probabilities and it is represented by a vector $D_i(x) = [d_{i,1}(x), ..., d_{i,c}(x)]$, in which $d_{i,j}(x)$ is the degree of support of classifier D_i for class j. More details can be found in [10].

2.1. Weighted Linear Combination Optimized by GA

To generate a consensus among all classification methods, a weighted linear combination is performed. In [11], this is known as *Linear Opinion Pool* (LOP). Since the task of finding a suitable set of weights for this combination is considered computational intensive, Genetic Algorithms (GA) have been used in previous works of ours for this purpose [8, 1].

The main idea of the WLC-GA method is to give more discriminant power for the classification approaches which contribute the most in the combination/classification process. This is done by weighting the i-th classification approach with values optimized by a genetic algorithm. New degrees of confidence for each class (often called support) are produced by the weighted classification approach. The weighted linear combination is defined as $\mu_j(x) = \sum_{i=1}^{L} w_i \times d_{i,j}(x)$, satisfying $\forall i, 0 \leq w_i \leq 1, \sum_{i=1}^{L} w_i = 1$, in which L is the number of classifiers, w_i is the weight of the i-th classifier, $d_{i,j}(x)$ is the probability given by the i-th classifier for class j, and $\mu_j(x)$ is the new support value for the class j. Then, for a given sample x, we choose the class with highest probability which is the index of the maximum support value $\mu_j(x)$, *i.e.*,

$$j^* = \arg\max_j \mu_j(x) \quad (1)$$

The objective function, or *fitness*, of the GA is based on the Overall Accuracy (OA) over a given traing set. In this case we attempt to maximize the OA as much as possible. The GA framework used in this study is provided by GALib [12] library. We kept the parameter setup the same as in our previous work [1].

2.2. WLC-GA-2

Similar to WLC-GA, we perform a weighted combination of the classification approach outputs. However, from WLC-GA definition, we have only one weight for each classification approach which does not change the differences among the class probabilities for a given sample. In contrast, in the WLC-GA-2 a set of weights for each classification method is used aiming to adjust the class probabilities given by the i-th classification approach. Then, the new linear combination (WLC-GA-2) is defined$\mu_j(x) = \sum_{i=1}^{L} w_{i,j} \times d_{i,j}(x)$, satisfying $\forall i, j, 0 \leq w_{i,j} \leq 1$, in which $w_{i,j}$ is the i-th classification approach weight for the j-th class. The class is chosen with the highest support value $\mu_j(x)$, similarly to WLC-GA (Equation 1). With this modification, each classification approach has its own set of weights which have the ability of tuning the soft output of a given classifier within the combination process. Although we do not ensure that the weights have the same distribution as before, we expect to tune more precisely the predictions made by the classifiers so that they are able to contribute the most with the whole combination process.

Since we have more weights to adjust, the problem becomes harder and finding an optimal (or sub-optimal) set of weights is not trivial. Again, we choose GA algorithm to find the set of weights. The goal, or *fitness* function, is the same as in WLC-GA, that is, maximizing the OA in a give training set.

2.3. Stacking-SVM

Proposed by Wolper [13], the *Stacked Generalization* is designed to perform a more elaborated, and not necessarily linear, combination of the classifier outputs. Instead of using simple rules, the combination rule is learned through a new learning algorithm. The classifier outputs of each classification approach used here are concatenated together to form a new feature vector, which spans a new feature space (*meta-feature-space*). This is often called *meta-data*, and consequently the learning algorithm used to perform the combination is known as *meta-learner*.

For a given sample x, each classification approach D_i yields probabilities $d_{i,j}(x)$ for the j-th class ($D_i(x) = \{d_{i,1}(x), ..., d_{i,c}(x)\}$). The meta-data $Z(x)$ for each sample x is formed by stacking these sets of probabilities $Z(x) = [D_1(x), D_i(x), ..., D_L(x)]$. Hence, the number of meta-features within the meta-data is equal to $L \times c$, in which L is the number of classification approaches used and c is the number of classes. Note that this technique can raise some issues, one of them is the significantly increase of the dimensionality as the number of classification approaches and classes increases.

In this work, we propose to use the SVM learning algorithm with RBF kernel as the meta-learner. This classifier has proven its promising capability when dealing with remote sensed hyperspectral data, specially when small training sets with large feature vectors are available [3].

3. EXPERIMENTS

3.1. Datasets

Experiments were performed on two widely used data sets, *Indian Pines* and the *Pavia University*, taken by the *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)* and *Reflective Optics System Imaging Spectrometer (ROSIS)*, respectively. More details regarding these data sets can be found in [14, 15] and also in [1]. As the size of the training data is an important issue in remote sensed image classification, experiments were carried out by using different training sizes to evaluate the robustness of our proposed approach. We setup our experiments as follows:

Training phase. We train the single classification approaches with training sample sizes of 10%, 15%, 20%, and 25% (full training set), where the samples were randomly chosen. The models produced here will be used in the final testing phase. Instead of using an additional training set for estimating the weights of the WLC-GA and WLC-GA-2, and the models (hylperplanes) of the Stacking-SVM, we split the full training set into two sets: the first one (50% of the training set) for temporarily training the single classifiers and the remaining training set for estimating the parameters of the combiners (*i.e.*, the weights for the WLC-GA and WLC-GA-2 and the hyperplanes for the Stacking-SVM) by using the temporarily trained single classifiers.

Testing phase. We use the weights estimated on half of the training sets and the single classifiers *re*trained on the full training set such that a fair comparison with the non-trainable combiners, *i.e.*, majority vote and average rules, could be performed. For each of these configurations, we ran the experiments 30 times, and then Confidence Intervals (CI) (obtained from the mean and standard deviation) of the Overall Accuracy (OA), Average Accuracy (AA), and $F-score$ for each class were reported. The respective sample size of the test sets are 90%, 85%, 80% and 75%.

To provide benchmark results, we compare the studied methods with two simple combination rules: Majority Vote (MV) and Average (Aveg.) [10]. A *paired t-test* [16] was also performed to confirm the statistical differences among all methods. A confidence level of 95% was used. This is the same procedure employed in [1].

Furthermore, aiming a deeply investigation regarding the effectiveness and behaviour of the combination methods, we performed the experiments in three different scenarios. **Scenario 1**: All classification approaches mentioned in this work were used in the combination process. **Scenario 2**: We removed the classification approach which stands with the best classification accuracy indexes. Only Pixelwise-SVM, FEGA-kNN and EMP-MLP were used. **Scenario 3**: We removed the classification approach which has the worst accuracy indexes. Only Pixelwise-SVM, EMP-MLP and SpecEmpPLS-SVM were used in the combination process.

3.2. Results and Discussion

Besides its simplicity, the WLC-GA combination method is the one which stands with the highest accuracy indexes in all scenarios for Indian Pines data sets (see Figure 1). Table 1 shows the combination results when all classification approaches considered here are used in the whole combination process. The OA presented by the WLC-GA is higher than its modified version, including when compared to

Table 1. Accuracy and $F-score$ results for the *Indian Pines* data set using only 10% for training.

			Individual Classification Approaches				Combination Methods				
	number of samples	**train samples (%)**	Pixelwise SVM (200 features)	FEGA kNN (95 features)	EMP MLP (18 features)	SpecEmpPLS SVM (16+16 features)	MV	Aveg.	WLC-GA-2	Stacking SVM (64 meta-features)	WLC-GA
OA(%)			80.24(±0.20)	61.24(±0.36)	84.60(±0.78)	93.95(±0.17)	88.77(±0.23)	91.82(±0.22)	93.55(±0.27)	91.36(±0.24)	**94.79**(±0.15)
AA(%)			67.56(±0.72)	50.03(±0.87)	80.42(±1.00)	85.82(±0.90)	79.42(±0.70)	83.40(±0.69)	87.53(±0.92)	78.52(±0.81)	**89.62**(±0.61)
Classes			**F − score**								
Alfalfa	54	10	0.22(±0.08)	0.21(±0.06)	0.72(±0.08)	**0.93**(±0.01)	0.81(±0.06)	0.75(±0.06)	0.84(±0.05)	0.06(±0.07)	0.92(±0.01)
Corn-notill	1434	10	0.77(±0.00)	0.50(±0.01)	0.78(±0.03)	0.90(±0.00)	0.84(±0.01)	0.89(±0.01)	0.91(±0.00)	0.88(±0.01)	**0.92**(±0.00)
Corn-min	834	10	0.68(±0.01)	0.41(±0.01)	0.77(±0.01)	0.93(±0.01)	0.87(±0.01)	0.91(±0.01)	0.92(±0.01)	0.89(±0.01)	**0.94**(±0.00)
Corn	234	10	0.53(±0.04)	0.40(±0.03)	0.53(±0.05)	0.86(±0.01)	0.69(±0.03)	0.80(±0.02)	0.84(±0.02)	0.78(±0.02)	**0.88**(±0.01)
Grass/Pasture	497	10	0.89(±0.01)	0.65(±0.02)	0.83(±0.02)	0.95(±0.00)	0.93(±0.00)	0.93(±0.00)	0.94(±0.01)	0.90(±0.01)	**0.96**(±0.00)
Grass/Trees	747	10	0.93(±0.00)	0.67(±0.01)	0.90(±0.01)	**0.98**(±0.00)	0.95(±0.00)	0.95(±0.00)	0.97(±0.01)	0.96(±0.01)	**0.98**(±0.00)
Grass/pasture-mowed	26	10	0.54(±0.08)	0.17(±0.06)	0.79(±0.06)	0.67(±0.12)	0.72(±0.08)	0.74(±0.05)	0.78(±0.06)	0.63(±0.09)	**0.89**(±0.03)
Hay-windrowed	489	10	0.93(±0.00)	0.91(±0.00)	0.91(±0.02)	**1.00**(±0.00)	0.97(±0.00)	0.96(±0.00)	0.98(±0.01)	0.98(±0.01)	0.99(±0.00)
Oats	20	10	0.02(±0.02)	0.01(±0.01)	0.38(±0.06)	0.12(±0.06)	0.12(±0.04)	0.28(±0.06)	0.40(±0.07)	0.32(±0.07)	**0.45**(±0.07)
Soybeans-notill	968	10	0.73(±0.01)	0.53(±0.01)	0.77(±0.01)	0.89(±0.00)	0.85(±0.01)	0.87(±0.01)	0.89(±0.01)	0.85(±0.01)	**0.90**(±0.00)
Soybeans-min	2468	10	0.78(±0.00)	0.64(±0.00)	0.87(±0.01)	0.94(±0.00)	0.89(±0.00)	0.92(±0.00)	0.94(±0.00)	0.92(±0.00)	**0.95**(±0.00)
Soybean-clean	614	10	0.73(±0.01)	0.31(±0.01)	0.73(±0.04)	0.91(±0.01)	0.83(±0.01)	0.89(±0.01)	0.90(±0.01)	0.88(±0.01)	**0.93**(±0.01)
Wheat	212	10	0.96(±0.01)	0.81(±0.01)	0.99(±0.00)	**1.00**(±0.00)	0.98(±0.00)	0.99(±0.00)	0.99(±0.00)	0.99(±0.00)	**1.00**(±0.00)
Woods	1294	10	0.94(±0.00)	0.82(±0.00)	0.98(±0.00)	**0.99**(±0.00)	0.96(±0.00)	0.98(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)
Bldg-Grass-Trees-Drives	380	10	0.67(±0.01)	0.24(±0.02)	0.94(±0.01)	0.95(±0.01)	0.79(±0.01)	0.93(±0.01)	0.95(±0.01)	0.94(±0.01)	**0.96**(±0.01)
Stone-steel towers	95	10	0.83(±0.02)	0.83(±0.01)	0.91(±0.03)	0.92(±0.02)	0.90(±0.01)	0.92(±0.01)	0.91(±0.02)	0.88(±0.03)	**0.93**(±0.01)

Table 2. Accuracy and $F-score$ results for the *Pavia University* data set using only 10% for training.

			Individual Classification Approaches				Combination Methods				
	number of samples	**train samples (%)**	Pixelwise SVM (103 features)	FEGA kNN (34 features)	EMP MLP (18 features)	SpecEmpPLS SVM (4+8 features)	MV	Aveg.	WLC-GA-2	Stacking SVM (36 meta-features)	WLC-GA
OA(%)			93.12(±0.07)	88.32(±0.08)	95.73(±0.25)	98.94(±0.03)	96.67(±0.07)	98.33(±0.04)	98.89(±0.06)	98.69(±0.06)	**99.03**(±0.03)
AA(%)			90.54(±0.12)	85.53(±0.13)	93.67(±0.39)	98.18(±0.06)	94.85(±0.10)	97.01(±0.06)	98.13(±0.10)	97.63(±0.12)	**98.22**(±0.07)
Classes			**F − score**								
Asphalt	6631	10	0.94(±0.00)	0.89(±0.00)	0.97(±0.00)	**0.99**(±0.00)	0.98(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)	0.98(±0.00)	**0.99**(±0.00)
Meadow	18649	10	0.97(±0.00)	0.94(±0.00)	0.98(±0.00)	**1.00**(±0.00)	0.98(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)
Gravel	2099	10	0.78(±0.00)	0.69(±0.00)	0.80(±0.02)	**0.95**(±0.00)	0.89(±0.00)	0.91(±0.00)	**0.95**(±0.00)	**0.95**(±0.00)	**0.95**(±0.00)
Trees	3064	10	0.94(±0.00)	0.88(±0.00)	0.97(±0.00)	**0.99**(±0.00)	0.96(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)	**0.99**(±0.00)
Metal Sheets	1345	10	0.99(±0.00)	0.99(±0.00)	0.98(±0.01)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)
Bare Soil	5029	10	0.90(±0.00)	0.79(±0.00)	0.96(±0.01)	**1.00**(±0.00)	0.95(±0.00)	0.99(±0.00)	0.99(±0.00)	0.99(±0.00)	**1.00**(±0.00)
Bitumen	1330	10	0.85(±0.00)	0.77(±0.00)	0.86(±0.01)	**0.96**(±0.00)	0.94(±0.00)	0.95(±0.00)	0.95(±0.00)	0.94(±0.00)	**0.96**(±0.00)
Bricks	3682	10	0.85(±0.00)	0.78(±0.00)	0.90(±0.01)	**0.97**(±0.00)	0.93(±0.00)	0.95(±0.00)	**0.97**(±0.00)	**0.97**(±0.00)	**0.97**(±0.00)
Shadow	947	10	0.99(±0.00)	0.99(±0.00)	0.98(±0.01)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)	**1.00**(±0.00)

a more sophisticated method such as Stacking-SVM. As the *Pavia University* result charts present similar conclusions to the ones of the *Indian Pines*, they were omitted.

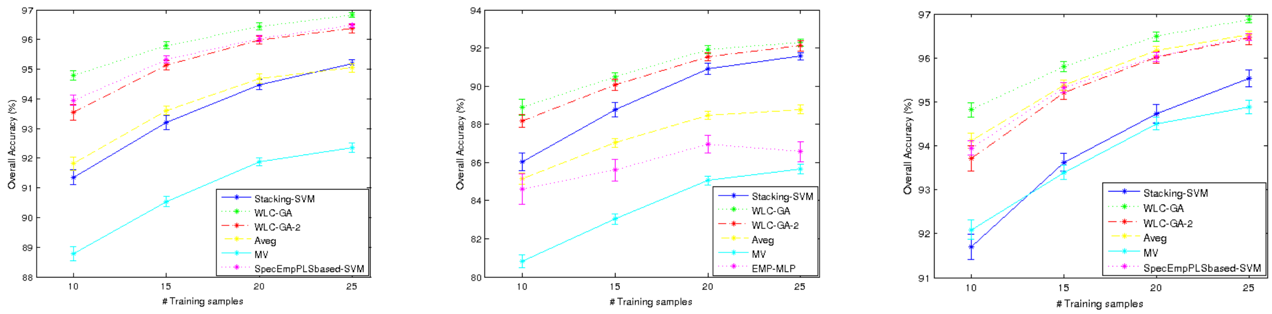
Based on the results in Figure 1, let us present a discussion about each studied scenario. In scenario 1 (Figure 1(a)), using any portion of training samples (*i.e.*, 10%, 15%, 20%, and 25%), it is important to observe that despite the t-test has found statistical difference at 95% of confidence level, the OA difference between the best classification approach SpecEmpPLS-SVM and the best combination method WLC-GA is less than 1%, that it, the accuracy gain is not so expressive/significant. Hence, the combination methods were not able to capture significant additional information, consequently, not making further improvements compared to the best single classification approach. This result indicates that the best single classification approaches tend to agree with each other, not providing useful information for expressive accuracy improvements in the combination process. This negative effect is even more stressed for WLC-GA-2 and the other combiners, which produced similar OA (WLC-GA-2) or worse (Stacking-SVM, Aveg., and MV) than the SpecEmpPLS-SVM. Thus, we hypothesize that the suitability of the combination methods depends on how much the single classification approaches are uncorrelated with each other.

In scenario 2 (Figure 1(b)), in which the SpecEmpPLS-SVM approach is out of the combination process, the WLC-GA-2 and Stacking-SVM methods produce higher OA than the best classification approach present in the evaluation. Note that these two methods achieve OA very close to the traditional WLC-GA. In addition, note that for 25% of training samples, the t-test reveals that there is no statistical difference between WLC-GA and WLC-GA-2. Moreover, note that the OA difference between the best single classification approach in scenarios 1 (SpecEmpPLS-SVM) and 2 (EMP-MLP) and the best combiner (WLC-GA), when 10% of training samples are used, is about 1% and 4%, respectively. In scenario 2, we suppose that the single classification approaches better complement each other, in other words, we suggest that they are able to generate uncorrelated errors so that the final gain is greater than the one obtained in scenario 1. It is important to note that although the results obtained by the combiners in scenario 2 are smaller than the ones achieved in scenario 1, the space for improvement in scenario 2 is higher than the one in scenario 1.

In scenario 3 (Figure 1(c)), SpecEmpPLS-SVM is again considered in the whole combination process. Only the classification approach with the worst OA was removed and, as it was expected, the classification results are close to the ones obtained in scenario 1 for the WLC-GA, WLC-GA-2 and Stacking-SVM methods. On the other hand, Average and MV methods achieved significant OA improvements if compared with their performance in scenario 1. This result suggests that Average and MV methods are negative biased by poor classification approaches.

Regarding the experiments in the three scenarios, we can observe that the Average method has demonstrated to be a very simple and useful method, specially in the last scenario. In contrast, the simple MV method provides the worst OA in almost all cases. One can also note that the complexity of finding the weights in WLC-GA-2 is much higher than its simple version. This suggests the need for a deep study of the GA parameters to search for a better setting.

Although we do not show the computation cost, the combination methods based on GA for the optimization are the most computational expensive due to the time to train the single classification approaches plus the time expended to adjust the weights through N generations of the GA method. Nevertheless, other optimization algorithms can be successfully applied to perform such task. For instance, one could employ *integer linear programming*, which has demonstrated to achieve the same accuracy indexes with further improvements of the computational time (10×) [17].



(a) Scenario 1 (all approaches). (b) Scenario 2 (best off). (c) Scenario 3 (worst off).

Fig. 1. Results for the *Indian Pines* data set on the three scenarios with training set with different sizes. The star symbols (*) are the averages (over 30 runs) and the vertical bars are the respective CI.

4. CONCLUSIONS

In this work, we conducted a study concerning the combination of multiple classification approaches for the problem of interpretation of remote sensed hyperspectral images. We extended, implemented and tested consensual rules based on linear combinations and also a non-linear combination method called Stacking-SVM. We compared the combination methods in some scenarios to investigate where combination methods are useful for the problem of hyperspectral data interpretation.

According to the experimental results, we draw the following conclusions. The extended WLC-GA (WLC-GA-2) and the Stacking-SVM combination methods are more suitable when classifiers have different opinions, *e.g.*, complementary classification approaches are used. Whereas, the traditional WLC-GA could also improve the accuracy indexes even when the single classification approaches tend to agree with each other. Despite of its simplicity, the WLC-GA method yields the highest Overall Accuracy on both used data sets. Our study hypothesized that the employment of different feature representation and learning algorithms to create an ensemble of classification approaches is able to generate uncorrelated errors, and thus, useful information for the combination process.

Acknowledgments: The authors would like to thank UFOP, UFPR, FAPEMIG, CAPES and CNPq for the financial support.

5. REFERENCES

[1] A. B. Santos, A. de A. Araújo, and D. Menotti, "Combining multiple classification methods for hyperspectral data interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, pp. 1450–1459, 2013.

[2] Jin Chen, Cheng Wang, and Runsheng Wang, "Using stacked generalization to combine svms in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE TGARS*, vol. 47, no. 7, pp. 2193–2205, 2009.

[3] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing Environmet*, vol. 113, no. 1, pp. 110–122, 2009.

[4] Saurabh Prasad, Lori M. Bruce, and Jocelyn Chanussot, *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, vol. 3, Springer, 2011.

[5] A. B. Santos, C. S. F. de S. Celes, A. de A. Araújo, and D. Menotti, "Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithms and cluster validity," in *IPCV*, 2012, pp. 675–681.

[6] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson, "Classification of hyperspectral data from urban areas based on extrended morphological profiles," *IEEE TGARS*, vol. 43, no. 3, pp. 480–491, 2005.

[7] A. B. Santos, A. de A. Araújo, D. Menotti, and W. R. Schwartz, "Hyperspectral image interpretation based on partial least squares," in *IEEE ICIP*, 2015, submitted.

[8] A. B. Santos, A. de A. Araújo, and D. Menotti, "Combiner of classifiers using genetic algorithms for classification of remote sensed hyperspectral images," in *IEEE IGARSS*, 2012, pp. 4146 –4149.

[9] Ethem Alpaydin, *Introduction to Machine Learning*, The MIT Press, 2010.

[10] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms. 2004*, Wiley-Interscience, 2004.

[11] J.A. Benediktsson and P.H. Swain, "Consensus theoretic classification methods," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 22, no. 4, pp. 688 –704, jul/aug 1992.

[12] Matthew Wall, "Galib: A c++ library of genetic algorithm components," *Mechanical Engineering Department, Massachusetts Institute of Technology*, vol. 87, pp. 54, 1996.

[13] D.H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[14] M. Fauvel, *Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data*, Ph.D. thesis, University of Iceland and Grenoble Institute of Technology, 2007.

[15] Y. Tarabalka, *Classification of Hyperspectral Data Using Spectral-Spatial Approaches*, Ph.D. thesis, University of Iceland and Grenoble Institute of Technology, 2010.

[16] Raj Jain, *Art of Computer Systems Performance Analysis Techniques For Experimental Design Measurements Simulation And Modeling*, John Wiley & Sons, Inc., 1991.

[17] S. L. J. L. Tinoco, H. G. Santos, D. Menotti, A. B. Santos, and J. A. dos Santos, "Ensemble of classifiers for remote sensed hyperspectral land cover analysis: An approach based on linear programming and weighted linear combination," in *IEEE IGARSS*, 2013.