

# ENSEMBLE OF CLASSIFIERS FOR REMOTE SENSED HYPERSPECTRAL LAND COVER ANALYSIS: AN APPROACH BASED ON LINEAR PROGRAMMING AND WEIGHTED LINEAR COMBINATION

*S. L. J. L. Tinoco, H. G. Santos, D. Menotti*

*A. B. Santos*

*J. A. dos Santos*

Federal University of Ouro Preto  
Computing Department  
Ouro Preto, MG, Brazil

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

University of Campinas  
Institute of Computing  
Campinas, SP, Brazil

## ABSTRACT

Hyperspectral images have been considered as one of the most important tool for remote sensed land cover analysis. Such images have information about materials on earth's surface expressed in many wavelengths that allow us to identify and classify those materials with more accuracy. In this work we used a combination of several classification methods in order to produce an accurate thematic map based on the remote sensed hyperspectral image classification. To perform the combination, three types of feature representation and two learning algorithms (Support Vector Machines (SVM) and Backpropagation Multilayer Perceptron Neural Network (MLP)) were used yielding six classification methods. Our approach proposal is based on Weighted Linear Combination (WLC), in which weights are found using Linear Programming (LP) - WLC-LP. Experiments are carried out using two well-known databases: Indian Pines, acquired by AVIRIS sensor; and Pavia University, acquired by ROSIS sensor. Results show the efficiency of our proposed approach which significantly reduces the time required to found optimal weights for the combiner compared to a previous approach based on Genetic Algorithm.

**Index Terms**— Ensemble of classifiers, conscious combiners, hyperspectral images, classification, linear programming solver, cplex.

## 1. INTRODUCTION

Hyperspectral image systems are capable of gathering information about materials on earth surface in a larger wavelength than traditional Multispectral ones [1]. Such information allows us to make a more accurate interpretation of those materials. However, the large amount of spectral information and its small set of referenced data become the classification of hyperspectral images a still challenge task [2, 3, 1]. In order to improve the power of generalization and reduce the computational overhead in the learning process, feature extraction/selection techniques are often applied [4, 1]. Furthermore, several authors join spectral-spatial information in their classification systems [2, 3, 5] such that the final classification accuracy is still improved.

Several works have been done in the field of combining multiple learners [6, 7] and have shown better results when compared to the best available learner. Recently, in [8] Majority Voting scheme (MV) has been employed for the combination of several classification schemes in order to further accuracy improvement for the hyperspectral image interpretation task. However traditional MV [8] does not guarantee any improvement [6] and so other techniques have been exploited. In [9], the authors investigated the advantage

of using a weighted linear combination (WLC) of soft outputs produced by different classification methods. Instead of using an heuristic (soft computing) optimization search technique based on genetic algorithms (GA) to find the weights and build the final decision as done in [9], we employed an exact (hard computing) approach using an Integer Linear Programming (ILP) [10] model which is solved using a state-of-art ILP solver in order to find the best weights which optimizes the final interpretation. We employ three feature extraction techniques with two different classification algorithms (Support Vector Machines (SVM) and a Multi-layer Perceptron (MLP) Artificial Network), resulting in six methods of classification to perform the combination. Experiments were carried out in well-known datasets: Indian Pines and Pavia obtained by AVIRIS and ROSIS sensors, respectively [2], showing that this new approach can overcome the results of traditional combining approaches and produce similar accuracy to [9] when the run time is ten times smaller than the one used by the GA in [9]. Moreover, in our approach there is no need of parameter tuning of the solver.

The remainder of this work is organized as follows. In Section 2, the feature representations used by the learning algorithms are briefly presented. Section 3 introduces our ensemble method based on a Weighted Linear Combination of classifiers which determines the weights using Linear Programming. Section 4 presents the experiments performed in order to evaluate our proposal and compare the accuracy and run time with a previous version of the combiner [9]. Finally, in Section 5, conclusions are pointed out

## 2. FEATURE REPRESENTATION TECHNIQUES

In the following, we present the most simple spectral representation, one spectral-spatial, and another spectral one based on feature selection.

### 2.1. Pixelwise Representation

The simplest feature representation is the pixelwise. It is composed of all hyperspectral bands, *i.e.*, the raw data. Note that the so-called pixelwise representation does not take into account any spatial information. That is, it only uses all spectral responses/bands (features) of the pixel. In other words, the prediction (classification) happens using all pixel features but without “inspecting” its neighborhood.

### 2.2. Extended Morphological Profiles - EMP

In order to obtain an accurate classification system, spectral and spatial information should be joined [2]. For this purpose, morphological operations are applied to build Morphological Profiles (MP) [3].

A feature reduction technique, such as Principal Component Analysis (PCA), is applied on hyperspectral data, and the principal components (PCs) are selected. Then, the MPs are built applying openings and closings by reconstruction on the PCs with increasing size structure elements (see details in [3]). Finally, PCs and MPs are put together to form the Extended Morphological Profile (EMP). This technique can connect similar structures of hyperspectral information making the data of the same class more homogeneous [3].

### 2.3. Feature Selection by Genetic Algorithms

In some cases, reducing data dimensionality is very important to avoid Hughes effect. Nowadays, many feature extraction/selection techniques are available [4, 11]. The method used here is a filter-based feature selection using Genetic Algorithms (FEGA - see [12] for more details). Since its creation, GA has been used to solve many optimization, search, and also machine learning problems [13]. The applied feature selection technique is based on the optimization of a clustering measurement that computes the “quality” of the yielded clusters by a subset of features. Such quality is done by computing the *Dunn’s index* of the evaluated clusters. Large values of Dunn’s index indicates compact and well-formed clusters. Similar to [4], the individual representation (*i.e.*, the chromosome) is a bit string that encodes absence or presence of a feature. The “optimal” subset is that one which maximizes the Dunn’s index. One can empirically observe that this optimization produces denser clusters and more distant from each other, in feature space. Hence, it is expected that from these subset of features it can be easier to build decision boundaries for a given classifier.

## 3. COMBINATION METHODS

In general, multiple classifiers yield soft outputs which means that for each class we have a certain degree of support [6]. To obtain this output, each classifier also needs to produce soft outputs. These outputs can be fuzzy, posterior probabilities, certainty, or possibility values [14]. That is, for an input sample there are membership values associated to classes. From these soft outputs one can build a Decision Profile (DP) [6]. Mathematically, a DP for a given sample  $x$  can be defined as a  $|L| \times |C|$  matrix  $DP(x) = [D_1(x), D_2(x), \dots, D_L(x)]$  in which  $D_i(x) = [d_{i,1}(x), \dots, d_{i,|C|}(x)]^T$ ,  $|L|$  is the number of classifiers,  $|C|$  is the number of classes, and  $d_{i,j}(x)$  is a degree of support given by a classifier  $D_i$  [6, 14] to the class  $j$ . After building support degrees for each input sample, a crisp value (the final label) can be assigned by using the maximum support value in the set, for instance. Combiners are methods which use all predictions made by two or more classifiers to build an accurate final decision [6].

### 3.1. Nontrainable and Trainable Combiners

Popularly known as “nontrainable”, by the fact that these types of combiners have no need of training any parameter, they perform some basic operations, such as average, maximum, minimum and product, in DP to produce new support values and hence a final decision [6]. Among these combiners, majority vote is the simplest one [14] and largely used [8]. In contrast, there are other types of combiners. It is reasonable to give more discriminant power to classifiers that have greater accuracy [6] when classifiers have different accuracies. Weighted Average, Weighted Majority Vote (WMV),

and other weighted approaches are based on this idea. Some authors suggest the use of probabilistic methods and global search algorithms [6], which are in general too computational expensive. The combiners in this category are known as trainable because they need to find the best set of weights to produce the best set of support, and our proposal combiner fits in this category.

### 3.2. Proposed WLC-LP

Here, and similar to [9], we use a Weighted Linear Combination (WLC) of the DP, in order to combine classifiers. Let  $\mu_{c,x} = \sum_{l=1}^L w_l \times d_{l,c,x}$  be the support for the class  $c$  given a sample  $x$ ,  $|L|$  be the number of classifiers,  $w_l$  be the weight of the  $l$ -th classifier and  $d_{l,c,x}$  be the support of  $l$ -th classifier for the class  $c$ . The task of finding the best weights is as an *Integer Linear Programming* optimization problem, so our proposed weighted linear combination linear programming approach (WLC-LP). And the IBM CPLEX solver [15], a state-of-art ILP solver, is used as optimization routine. This problem requires the minimization (or maximization) of a linear form subject to linear inequality constraints, and it is defined as follows.

Initially, let us define the following *inputs*:  $X$ : set of train samples;  $L$ : set of classification approaches;  $C$ : set of classes;  $d_{l,c,x}$ : degree of confidence that the classification approach  $l \in L$  has for item  $x \in X$  to be of class  $c \in C$ ;  $cl_{c,x} \in \{0, 1\}$  is one if the true class for  $x \in X$  is  $c \in C$ , and zero otherwise. Note that  $cl$  is obtained from the six classification methods which are generated from the feature representation methods links to the learning algorithms. Then, let the *main decision variables*  $w_l \in [0, 1]$  be the weight of the classifier  $l$ . Also, let us define the following *auxiliary variables*:  $\mu_{c,x} \in \mathbb{R}^+$  is the degree of confidence that the sample  $x$  belongs to class  $c$ ;  $s_{c,x} = 1$ , if for  $x \in X$  the class with the highest confidence in accordance with the weights is  $c$  and  $s_{c,x} = 0$ , otherwise. And finally the *objective function and constraints*:

$$Max : \sum_{c \in C} \sum_{x \in X} cl_{c,x} s_{c,x}.$$

Subject to:

$$\mu_{c,x} = \sum_{l \in L} w_l d_{l,c,x}, \forall c \in C, x \in X;$$

$$\sum_{l \in L} w_l = 1;$$

$$\mu_{c,x} + 1 - s_{c,x} \geq \mu_{c',x} \forall c \in C, c' \in C, x \in X : c \neq c'.$$

New supports for each class are built using the WLC and the weights found by running the simplex method. Then, a label is assigned, for a given sample  $x$ , as the index of the maximum support  $\mu_{c,x}$ .

## 4. EXPERIMENTS

Experiments are performed using the *Indian Pines* datasets, acquired by AVIRIS airborne sensor data [2], which cover an area of agriculture and forest in northeastern Indiana, USA, represented by  $145 \times 145 \times 220$  pixels. Noise bands are removed, that is, the indexed from 104 to 108, from 150 to 163 and 220, remaining a total of 200 bands. This image presents sixteen classes or categories. In order to verify the degree of generalization of our approach, tests are

**Table 1:** Results using the well-known Pavia dataset for testing (only 5% of data training).

	# samples	Pixelwise MLP	FEGA MLP	EMP MLP	Pixelwise SVM	FEGA SVM	EMP SVM	Average	Majority Vote	WLC-GA 5%	WLC-LP 5%
OA (%)		66.88(± 01.63)	68.90(± 02.05)	81.21(± 01.55)	79.69(± 01.07)	75.81(± 01.29)	83.92(± 01.02)	87.89(± 00.08)	86.47(± 00.10)	89.52(± 00.93)	89.47(± 00.98)
AA (%)		54.65(± 02.02)	59.08(± 03.02)	76.41(± 03.19)	62.62(± 01.82)	58.77(± 01.49)	66.98(± 01.60)	75.22(± 00.16)	71.51(± 00.14)	76.76(± 03.16)	77.01(± 02.26)
Alfafa	54	19.08(± 13.79)	29.80(± 16.07)	68.56(± 26.91)	10.59(± 13.47)	02.42(± 05.93)	09.15(± 16.16)	41.37(± 1.40)	41.24(± 1.50)	44.71(± 15.57)	41.89(± 16.61)
Corn-notill	1434	60.70(± 03.93)	60.06(± 04.41)	76.05(± 05.51)	77.71(± 03.10)	65.85(± 05.75)	80.23(± 03.17)	81.08(± 0.18)	84.73(± 0.15)	85.29(± 01.26)	85.40(± 01.65)
Corn-mintill	834	46.37(± 05.92)	51.18(± 06.25)	63.55(± 09.98)	64.65(± 04.75)	52.94(± 05.41)	64.12(± 04.74)	79.15(± 0.31)	80.11(± 0.34)	81.87(± 03.31)	81.00(± 03.70)
Corn	234	28.35(± 11.95)	35.59(± 15.10)	41.34(± 22.00)	30.02(± 15.64)	21.98(± 15.30)	39.13(± 15.27)	51.19(± 0.60)	42.66(± 0.81)	41.32(± 04.93)	41.41(± 04.64)
Grass-pasture	497	66.96(± 09.87)	67.25(± 12.65)	71.57(± 10.39)	88.39(± 03.73)	86.91(± 04.09)	88.12(± 04.16)	91.74(± 0.27)	92.81(± 0.30)	90.90(± 01.60)	91.27(± 01.60)
Grass-trees	747	85.79(± 05.18)	85.06(± 05.76)	93.68(± 02.92)	93.28(± 03.20)	92.11(± 02.97)	93.94(± 02.61)	97.56(± 0.11)	97.97(± 0.11)	97.54(± 00.50)	97.48(± 00.49)
Grass-pasture-mowed	26	05.60(± 06.09)	15.33(± 13.58)	67.73(± 30.96)	00.00(± 00.00)	00.00(± 00.00)	00.00(± 00.00)	00.00(± 0.00)	03.73(± 1.01)	15.60(± 13.83)	13.46(± 10.94)
Hay-windrowed	489	88.18(± 06.35)	86.69(± 06.37)	93.31(± 10.72)	98.65(± 00.87)	98.77(± 00.64)	97.43(± 02.19)	99.38(± 0.07)	99.20(± 0.10)	99.38(± 00.09)	99.38(± 00.07)
Oats	20	05.61(± 05.85)	23.33(± 20.58)	59.47(± 30.13)	00.00(± 00.00)	00.00(± 00.00)	00.00(± 00.00)	35.44(± 2.37)	05.09(± 0.96)	32.28(± 20.11)	27.36(± 21.20)
Soybean-notill	968	60.51(± 05.65)	64.62(± 06.77)	66.71(± 04.30)	65.01(± 08.05)	63.38(± 06.46)	82.79(± 03.42)	77.82(± 0.30)	77.92(± 0.34)	79.56(± 02.86)	79.95(± 03.66)
Soybean-mintill	2468	72.25(± 02.45)	72.95(± 03.64)	89.56(± 03.22)	86.53(± 02.07)	85.34(± 02.89)	91.10(± 02.35)	93.30(± 0.11)	90.21(± 0.15)	94.72(± 00.84)	94.58(± 00.93)
Soybean-clean	614	38.46(± 10.21)	55.61(± 11.22)	58.83(± 13.70)	72.44(± 06.53)	72.61(± 05.83)	53.65(± 10.47)	86.55(± 0.33)	79.98(± 0.38)	88.38(± 02.71)	87.87(± 02.80)
Whea	212	85.09(± 08.18)	89.10(± 08.56)	98.47(± 01.44)	93.70(± 03.95)	93.05(± 03.92)	96.19(± 02.69)	98.59(± 0.19)	99.05(± 0.15)	98.54(± 00.58)	98.52(± 00.84)
Woods	1294	90.87(± 02.91)	91.41(± 02.10)	98.44(± 01.84)	97.33(± 01.25)	97.30(± 01.14)	99.40(± 00.35)	99.01(± 0.08)	98.78(± 0.08)	99.68(± 00.11)	99.66(± 00.11)
Build.-Grass...	380	40.54(± 08.58)	35.72(± 08.72)	89.79(± 06.25)	44.61(± 06.45)	32.28(± 06.65)	92.55(± 03.99)	79.93(± 0.38)	61.31(± 0.55)	90.37(± 06.49)	89.58(± 06.06)
Stone-Steel-Towers	95	80.00(± 10.40)	81.56(± 16.68)	85.48(± 11.25)	79.04(± 09.59)	75.44(± 08.35)	83.93(± 09.66)	91.33(± 0.54)	89.30(± 0.80)	88.11(± 03.78)	87.85(± 03.18)

**Table 2:** Results using the well-known Indian Pines and Pavia datasets for testing (only 5% of data training).

	# samples	Pixelwise MLP	FEGA MLP	EMP MLP	Pixelwise SVM	FEGA SVM	EMP SVM	Average	Majority Vote	WLC-GA 5%	WLC-LP 5%
OA (%)		66.88(± 01.63)	68.90(± 02.05)	81.21(± 1.55)	79.69(± 1.07)	75.81(± 1.29)	83.92(± 1.02)	087.89(± 0.08)	086.47(± 0.10)	089.52(± 0.93)	89.47(± 0.98)
AA (%)		54.65(± 02.02)	59.08(± 03.02)	76.41(± 3.19)	62.62(± 1.82)	58.77(± 1.49)	66.98(± 1.60)	075.22(± 0.16)	071.51(± 0.14)	076.76(± 3.16)	77.01(± 2.26)
Asphalt	6631	89.81(± 01.32)	87.66(± 01.65)	97.54(± 0.62)	93.28(± 0.78)	91.09(± 1.36)	97.91(± 0.67)	098.21(± 0.03)	097.93(± 0.04)	098.78(± 0.20)	98.79(± 0.20)
Meadows	18649	97.12(± 00.43)	96.85(± 00.41)	99.18(± 0.28)	97.77(± 0.38)	97.39(± 0.38)	99.03(± 0.29)	099.62(± 0.01)	099.51(± 0.01)	099.78(± 0.05)	99.78(± 0.02)
Gravel	2099	69.36(± 03.69)	64.67(± 06.60)	65.26(± 6.78)	77.42(± 2.01)	74.13(± 2.76)	74.86(± 2.28)	080.41(± 0.20)	082.39(± 0.19)	080.55(± 2.82)	81.70(± 2.36)
Trees	3064	89.57(± 02.42)	88.18(± 02.53)	96.36(± 1.03)	93.32(± 1.41)	91.87(± 1.82)	97.07(± 0.95)	097.13(± 0.07)	094.42(± 0.09)	098.68(± 0.32)	98.62(± 0.24)
Painted metal sheets	1345	99.05(± 00.74)	98.99(± 00.68)	99.21(± 0.54)	98.78(± 0.73)	98.67(± 0.79)	99.61(± 0.15)	099.63(± 0.03)	099.63(± 0.03)	099.66(± 0.05)	99.61(± 0.06)
Bare Soil	5029	83.04(± 04.55)	76.12(± 04.88)	96.42(± 1.70)	88.59(± 1.50)	84.63(± 1.46)	97.08(± 0.77)	093.02(± 0.08)	087.15(± 0.13)	098.12(± 0.78)	97.94(± 0.75)
Bitumen	1330	75.76(± 04.65)	50.77(± 22.87)	88.98(± 2.48)	82.44(± 2.71)	71.18(± 7.81)	94.27(± 0.97)	089.35(± 0.18)	082.00(± 0.22)	091.72(± 0.37)	91.73(± 0.42)
Self-Blocking Bricks	3682	85.23(± 01.73)	85.15(± 02.01)	93.14(± 2.54)	89.55(± 1.37)	87.34(± 1.76)	95.05(± 1.41)	095.44(± 0.08)	091.90(± 0.10)	096.42(± 0.27)	96.25(± 0.28)
Shadows	947	97.08(± 14.73)	99.64(± 00.57)	98.39(± 4.21)	99.57(± 0.33)	99.60(± 0.26)	99.99(± 0.05)	100.00(± 0.00)	100.00(± 0.00)	100.00(± 0.00)	99.99(± 0.02)

performed with a second training set, the *University of Pavia (Italy)* dataset, acquired by the sensor ROSIS, depicted by  $610 \times 340 \times 103$  pixels [2]. This image presents nine classes. The experiments are performed with all 103 image bands.

To find the best weights for our WLP-LP, we use the CPLEX 12.5, optimization software package, that implements the simplex method. In the first two numerical rows of Tables 1 and 2, we show the overall (OA) and average (AA) accuracy obtained by the six classification methods, the feature representations described above with SVM and MLP, and our proposed combiner: WLC-LP, the one in [9]: WLC-GA and other trainable combiners (average and majority vote). Observe that the accuracies for each class of Indian Pines and University of Pavia are shown in their respective tables in the remainder numerical rows. Also observe that we used 5% of data for training both the classifications methods and the combiners, the remainder data (95%) is used for testing. This scheme is randomly performed 30 times and so the mean and standard deviation obtained are reported, *i.e.*,  $\mu \pm \sigma$ . One can see that our approach achieved the highest accuracies for both datasets together with the one in [9].

In Fig. 1, we show the groundtruth images of the two datasets used (Fig. 1(a) and Fig. 1(c)) and their respective Thematic Maps using our WLC-LP approach (Fig. 1(b) and Fig. 1(d)). These images were obtained by the combiner whose OA is the closest to the average OA obtained in the 30 runs.

We limit the simplex to run ten times less than the time required for the GA in [9]. Simplex run time was fixed to 2 and 5 seconds for Indian Pines and Pavia University datasets, respectively since the GA average run time was 23.93 and 47.33 seconds, respectively. It is important to note that if the CPLEX run time achieves the limits, the found solution is not demonstrably optimal. In Table 3, we show the testing and training time required for each combiner approach and classification method. It is noticeable that our approach can find for both datasets the best weights ten times faster in average than the WLC-GA (*i.e.*, the training time) as shown in Table 3 and still pro-

vide the same accuracy (Tables 1 and 2). Note that the total run time required for each combiner is the sum of the classification methods run time plus its run time.

## 5. CONCLUSIONS

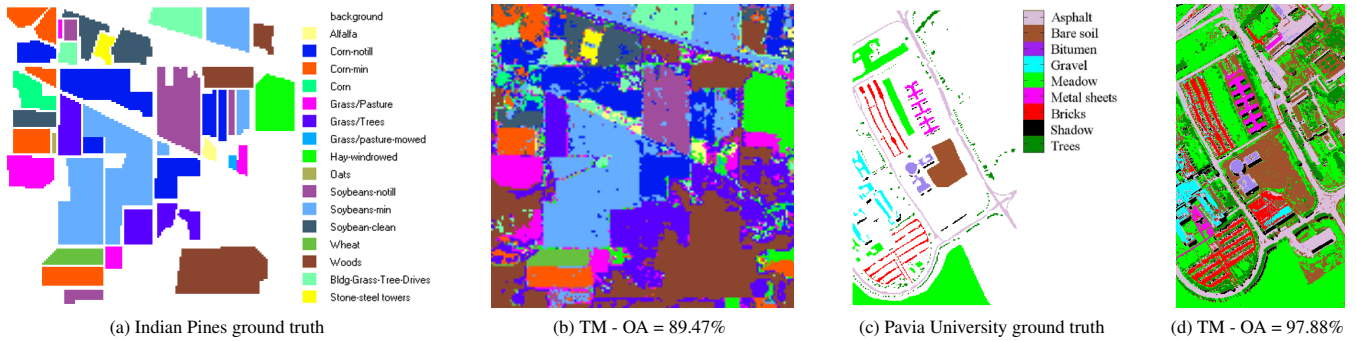
In this paper we proposed, implemented, and tested a combiner/ensemble of classifiers for remote sensed hyperspectral land cover analysis. We used three types of feature representation methods and two classification algorithms yielding six classifier methods. In the experiments, the weighted linear combiner adjusted by a linear programming model was compared with the most well-known combiners in the literature and it has shown better performance in terms of overall accuracy and average accuracy in the two datasets used. We also compare our method with the one proposed in [9] (WLC-GA). We observed that they produce similar accuracy results, when we limit our approach to run in ten times less than the WLC-GA. Moreover in contrast to [9] in which GA parameters such as number of iterations, size of individuals, crossover and mutation probabilities, *etc.* need to be tuned, we do not need to tune any parameters of the ILP solver.

## 6. REFERENCES

- [1] C.I. Chang, *Hyperspectral data exploitation: theory and applications*, Wiley-Blackwell, 2007.
- [2] A. Plaza, J.A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing Environmet*, vol. 113, pp. 110–122, 2009.
- [3] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson, "Classification of hyperspectral data from urban areas based on ex-

**Table 3:** Run time (in seconds) using the well-known Indian Pines and Pavia datasets for testing and training.

dataset		Pixelwise MLP	FEGA MLP	EMP MLP	Pixelwise SVM	FEGA SVM	EMP SVM	Average	Majority Vote	WLC-GA 5%	WLC-LP 5%
Indian Pines	Training	21.40	14.88	10.95	0.60	0.46	0.19	-	-	23.93	1.98
	Testing	0.10	0.06	0.02	2.25	1.30	0.60	3.55	0.39	0.02	0.02
Pavia	Training	32.78	20.00	19.08	1.76	0.87	0.50	-	-	47.33	4.98
	Testing	0.20	0.06	0.56	7.72	3.39	1.60	8.08	0.94	0.05	0.05



**Fig. 1:** Ground Truth and respective Thematic Maps (TM) obtained by the WLC-LP approach for Indian Pines and Pavia University datasets

tended morphological profiles,” *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 43, pp. 480–491, 2005.

- [4] Y. Bazi and F. Melgani, “Toward an optimal svm classification system for hyperspectral remote sensing images,” *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 44, no. 11, pp. 3374–3385, 2006.
- [5] Y. Tarabalka, J.A. Benediktsson, J. Chanussot, and J.C. Tilton, “Multiple spectral-spatial classification approach for hyperspectral data,” *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 48, pp. 4122–4132, 2010.
- [6] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms. 2004*, Wiley-Interscience, 2004.
- [7] B. Gabrys and D. Ruta, “Genetic algorithms in classifier fusion,” *Applied soft computing*, vol. 6, no. 4, pp. 337–347, 2006.
- [8] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Chritophe, J. Chanussot, and P. Gamba, “Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest,” *IEEE Trans. on Geoscience and Remote Sensing (TGARS)*, vol. 47, pp. 3857–3865, 2009.
- [9] A. B. Santos, A. de A. Araújo, and D. Menotti, “Combiner of classifiers using genetic algorithm for classification of remote sensed hyperspectral images,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012, pp. 4146–4149.
- [10] L.A. Wolsey, *Integer Programming*, Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1998.
- [11] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 2 edition, 1995.
- [12] 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 algorithm and cluster validity,” in *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV’12)*, 2012, vol. 2, pp. 675–681.
- [13] D.E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-wesley, 1989.
- [14] A. Ghosh, B.U. Shankar, L. Bruzzone, and S.K. Meher, “Neuro-fuzzy-combiner: an effective multiple classifier system,” *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 2, no. 2, pp. 107–129, 2010.
- [15] ILOG S.A., *CPLEX 12.5 User’s Manual*, 2012.