

# UNSUPERVISED CLASSIFIER SELECTION APPROACH FOR HYPERSPECTRAL IMAGE CLASSIFICATION

*Bharath Bhushan Damodaran, Nicolas Courty, Sébastien Lefèvre*

Univ. Bretagne-Sud, UMR 6074, IRISA, F-56000 Vannes, France  
{firstname.lastname}@irisa.fr

## ABSTRACT

Generating accurate and robust classification maps from hyperspectral imagery (HSI) depends on the users choice of the classifiers and input data sources. Choosing the appropriate classifier for a problem at hand is a tedious task. Multiple classifier system (MCS) combines the relative merits of the various classifiers to generate robust classification maps. However, the presence of inaccurate classifiers may degrade the classification performance of MCS. In this paper, we propose a unsupervised classifier selection strategy to select an appropriate subset of accurate classifiers for the multiple classifier combination from a large pool of classifiers. The experimental results with two HSI show that the proposed classifier selection method overcomes the impact of inaccurate classifiers and increases the classification accuracy significantly.

**Index Terms**— Hyperspectral image classification, Multiple classifier system, Classifier selection, Classifier combination, Ensemble learning

## 1. INTRODUCTION

Supervised image classification is employed to translate the pixel information into a useful information in the form of a classification map. The high dimensionality, Hughes phenomena are the well-known limitations which affect the classification performance of hyperspectral imagery (HSI) [1]. Furthermore, the choice of the classification method plays a significant role in determining the accuracy of the resulting classification map.

With the plethora of classifiers available in the literature, finding an appropriate classifier for a given problem is always a challenging and time consuming task. It is established that there is no single best classifier which is optimal across all the datasets and applications. Therefore, human expert involvement and pre-image classification tasks are necessary to identify the optimal classifier.

Recently, MCS has proven to be an effective strategy for generating accurate and robust classification map for HSI [2–4]. MCS generates a large pool of classifiers and combines the decision values of these numerous classifiers in an effective manner to generate accurate classification map. However,

the presence of inaccurate classifiers may degrade the classification performance of MCS and the resulting classification accuracy might not be better than the best individual classifier. In order to alleviate this, a subset of reasonably accurate classifiers is selected from a large of pool classifiers known as classifier selection or ensemble selection [5, 6]. Sparse optimization approach, and the joint accuracy and diversity measures are employed to select the subset of classifiers for HSI classification [7, 8]. In these studies, the criteria for selecting the classifiers are dependent on the evaluation measures of the training samples.

Evaluating the classifier using training accuracy measurement is not recommended, since the training samples do not provide a fair global representation of HSI, especially when a very limited number of training samples is available. This calls for the development of classifier selection strategies in an unsupervised fashion. Though MCS has been extensively studied in remote sensing [3], only few attempts have been made in studying classifier selection strategies [7, 8] and so far no attempt has been reported to develop classifier selection approach in an unsupervised manner for HSI classification. In this paper, we propose such an approach to select the subset of accurate classifiers in MCS for effective HSI classification. The proposed unsupervised classifier selection approach is developed based on the minimum energy framework.

## 2. MATERIALS AND METHODS

In this section, we first present the generation of multiple classifier system and the classifiers employed in this study, and later we present the proposed method.

Let  $\Psi = \{\psi_1, \psi_2, \dots, \psi_L\}$  be the base classifiers forming a MCS, and each classifier  $\psi_l, l = 1, 2, \dots, L$  be a function  $\psi_l : \chi \rightarrow \Omega$  from an input space  $\chi \subseteq \mathbb{R}^d$  to a set of class labels  $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$  ( $d$  is the number of bands in HSI, and  $M$  is the number of classes). For any given  $\mathbf{x} \in \chi$ , classifier  $\psi_l$  produces a vector of probability scores  $P = [P_{l1}, P_{l2}, \dots, P_{lM}]$  and  $\mathbf{x}$  is assigned to the class which has the maximum probability (decision) value. For the MCS to be successful, the base classifiers forming MCS has to commit different types of error in their predictions.

## 2.1. Random Subspace Method

Random subspace method (RSM) is a popular ensemble generation technique to generate multiple input data sources from a single input data, and thus creating diversity among the base classifiers in the MCS [4]. RSM randomly selects a subset of features from the original spectral bands  $L$  times, where  $L$  is the number of base classifiers in the ensemble. In order to introduce good diversity between each subspace, it is recommended to select  $\sqrt{d}$  number of features in each subspace. Each input data source generated from the RSM is returned as the input to the supervised learning algorithm  $\psi$ . In order to obtain the final classification results, the base classifiers decision values or probabilistic scores are combined using three popular combination functions such as majority voting (MV), Bayesian average (Avg), and product (Prod) in the MCS [9].

Support vector machine (SVM) [10], and extreme learning machine (ELM) [11] are used as the learning algorithms in the RSM. The minimum number of hidden neurons (100 hidden neurons) are used with the ELM classifier to introduce more bias among the base classifiers.

## 2.2. Proposed Unsupervised Classifier Selection Approach

To mitigate the negative impact of inaccurate classifiers in multiple classifier combination, here we develop a novel classifier selection method for the MCS in the framework of HSI classification.

The proposed unsupervised classifier selection (UCS) method is based on the calculation of the energy term related to the classification results. It is reasonable to assume that the accurate classifiers will always have a minimum energy. More specifically, the proposed method builds upon two following assumptions: (a) the accurate classifiers will have good spatial consistency among the neighbours. In other words, the neighbouring pixels will belong to the same class and we call this energy as the spatial energy (SE); (b) the probability of correct classification will be sufficiently greater than the probability of incorrect classification. That is,  $P(\frac{\omega_i}{\mathbf{x}}) \gg P(\frac{\omega_j}{\mathbf{x}}), \forall i \neq j$  ( $\omega_i$  is the correct label for  $\mathbf{x}$ ), and we call this energy as the potential energy (PE). Therefore, the classifier selection problem is modeled as computing the energy of the base classifiers and selecting the subset of appropriate classifiers which have the minimum energy. The energy of the base classifiers is calculated as follows.

The spatial energy is given by

$$SE_l = \sum_{i=1}^T \sum_{j \in N(\mathbf{x})} (1 - \delta(\omega_i, \omega_j)), l = 1, \dots, L. \quad (1)$$

The spectral energy is given by

$$PE_l = \sum_{i=1}^T -\log(P(\omega_i/x_i)), l = 1, \dots, L, \quad (2)$$

and the combined spatial and spectral energy is given by

$$CE_l = PE_l + SE_l, l = 1, \dots, L. \quad (3)$$

$T$  is the total number of pixels in the HSI,  $\delta(a, b) = 1$  when  $a = b$ , and  $P(\omega_i/x_i)$  is the posterior probability of the pixel  $\mathbf{x}_i \in \omega_i$ . The energies are sorted and the first  $K$  classifiers are selected as the appropriate classifiers for the classifier combination.

## 3. EXPERIMENTAL RESULTS AND DISCUSSION

### 3.1. Experimental Datasets

In order to study the potential of the proposed unsupervised classifier selection method for HSI classification, we adopted two benchmark HSI with different land cover settings.

*Pavia University:* The first hyperspectral dataset was collected over the University of Pavia by the ROSIS sensor. The image contains  $610 \times 340$  pixels with high spatial resolution of 1.3 m/pixel and 103 spectral channels. Nine classes are presented in the image and ground truth reference map contains 42,776 labeled pixels.

*Indian Pines:* The second hyperspectral image used was collected by the AVIRIS sensor over the Indian pines site in Northwestern Indiana. This image contains  $145 \times 145$  pixels with the spatial resolution of 20 m/pixel and 200 spectral channels. It consists of sixteen land cover classes and ground truth reference map contains 10,249 labeled pixels.

### 3.2. Experimental Design

From the available ground truth reference samples, we randomly choose 20 samples per class for training and remaining samples are used for testing. The experiments are repeated five times, and the reported accuracy measurements in Sec. 3.3 are averaged over the five runs. In the MCS, we generated 50 base classifiers or ensembles ( $L = 50$ ). The parameters of the SVM classifier  $C$  (cost function) and  $\gamma$  (band width parameter of the Gaussian Radial basis function) are both automatically tuned by 10 fold cross-validation from the range of  $[2^{-5}, 2^{15}]$ , and  $[2^{-15}, 2^5]$  respectively. The parameter  $\gamma$  of the ELM classifier is also tuned in a similar manner. The experiments are conducted independently with SVM as a base classifier, and ELM as a base classifier in the MCS.

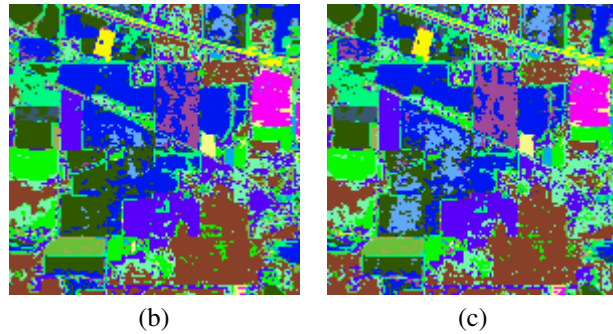
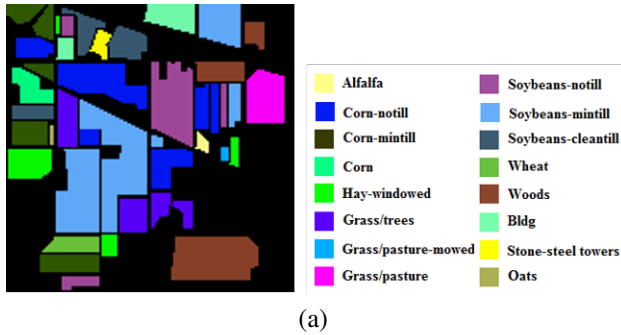
### 3.3. Results and Discussion

#### 3.3.1. MCS with SVM as a base classifier

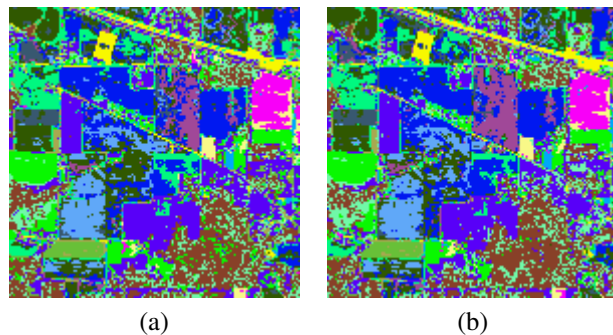
Table 1 summaries the classification accuracies (mean accuracy and standard deviation) for both datasets when SVM classifier is used as the base classifier in the MCS. The classification accuracy of the proposed classifier selection (UCS-CF) method is compared with random forest classifier (RF)

**Table 1:** Classification results when SVM is used as the base classifier in the MCS. The numbers in the brackets indicate the selected number of classifiers (NC) by the proposed UCS method.

Image	RF	FB	BIC	MV	MCS-CF		ET(NC)	UCS-CF		
					Avg	Prod		MV	Avg	Prod
Indian Pines	56.24±1.8	52.88±3	55.63±4	55.15±2.8	55.99±3.3	56.20±3.3	SE(3)	57.29±3.9	58.41±4	58.51±4
							PE(7)	59.29±3.6	60.36±3.9	60.60±3.9
							SE+PE(1)	60.34±4.4	60.34±4.4	60.34±4.4
Pavia University	64.51±3.8	73.35±3.4	71.25±3.7	71.18±2.7	71.25±2.7	71.30±2.7	SE(32)	71.00±3.3	71.07±3	71.14±3
							PE(2)	76.05±1	73.53±2.3	73.55±2.3
							SE+PE(32)	71.12±3.1	71.18±2.8	71.26±2.8



**Fig. 1:** *Indian Pines* image. (a) Ground truth reference map and legends. (b) SVM-MCS-CF classification map. (c) SVM-UCS-CF classification map



**Fig. 2:** *Indian Pines* image. (a) ELM-MCS-CF classification map. (b) ELM-UCS-CF classification map

[12], fullband classification (FB) [10,11], best individual classifier (BIC), and combination of all the classifiers in the MCS (MCS-CF) [4]. In the classifier selection, we have evaluated the three versions of energy term (ET), see Eqns. 1-3.

As seen from Tab. 1, the results are coherent for both images. The proposed UCS-CF yields significantly higher classification accuracies when compared with the baseline approaches. For *Indian Pines*, the classification accuracy is improved by 4-5% when compared with MCS-CF, BIC, RF, and by 7.7% with FB classification respectively. Similarly, for *Pavia University*, there exists a 2-5% improvement when compared with MCS-CF and BIC, 3% with FB classification, and 11% with RF classification.

It is interesting to see that there is no significant accuracy difference between the BIC and MCS-CF. The presence of inaccurate classifiers (and possibly redundant classifiers) might be the reason for this phenomenon and as a result MCS-CF does not yield expected improvement in classification accuracy. This observation emphasizes the significance of our proposed method in the MCS framework.

### 3.3.2. MCS with ELM as the base classifier

Table 2 summarizes the classification accuracies (mean accuracy and standard deviation) when ELM is used as the base classifier. As observed with Tab. 1, the proposed UCS-CF has yielded highest classification accuracy for both images when compared with baseline approaches. The proposed UCS-CF improved classification accuracy by 5% when compared with MCS-CF and RF, 10% with BIC, and 35% with FB classification for the *Indian Pines* image.

The comparison of Tab. 1 and Tab. 2 reveals some interesting observations. On contrary to Tab. 1, there is a significant accuracy difference between BIC and MCS-CF. This states that ELM classifier with minimum number of hidden neurons creates more diversity among the base classifiers than the SVM classifier in the MCS. The accuracy of the FB classification and BIC is lower with ELM classifier when compared to the SVM classifier, but the UCS-CF in Tab. 2 has provided higher classification accuracies than in Tab. 1. Furthermore, The classification accuracies with ELM classifier have small variations compared to accuracies with SVM classifier.

**Table 2:** Classification results when ELM is used as the base classifier in the MCS. The numbers in the brackets indicate the selected number of classifiers (NC) by the proposed UCS method.

Image	FB	BIC	MCS-CF			ET(NC)		USC-CF		
			MV	Avg	Prod			MV	Avg	Prod
Indian Pines	26.39±1.5	51.03±4.1	55.70±1.2	56.58±0.9	56.72±0.8	SE(7)	59.74±1.4	61.60±1.3	61.69±1.3	
						PE(9)	59.94±1.5	60.60±1.1	60.79±1	
						SE+PE(7)	59.74±1.4	61.60±1.3	61.69±1.3	
University	58.10±1.8	67.29±2.3	75.83±1.97	75.87±1.94	75.75±1.97	SE(16)	76.53±1.2	76.55±1.8	76.48±1.7	
						PE(23)	76.17±2.1	76.33±1.9	76.23±1.9	
						SE+PE(16)	76.53±1.9	76.55±1.8	76.48±1.7	

Figures. 1(b-c), and 2 compare the classification maps of MCS-CF and UCS-CF for the *Indian Pines* image. It can be seen that the UCS-CF yielded better quality classification maps. Furthermore, we have evaluated the impact of three energy terms (ET). Higher classification accuracies are obtained with PE in Tab. 1 and on the other hand there is no huge difference among different energy terms in Tab. 2. The joint energy term does not provide any additional information to classifier selection, thus the selected classifiers are similar to SE term. The analysis of Fig. 1(b) and 2 (a) reveals that most of base classifiers (SVM) misclassifies the entire region of particular classes (for e.g. soybeans-mintill), and thus having minimum SE. This might be the reason for the inferior performance of SE term with SVM classifier in 1. Furthermore, the obtained improvement in classification accuracy by the proposed method is achieved with few number of selected classifiers. This proves that the combining few classifiers is better than all.

#### 4. CONCLUSION

In this paper, we addressed the problem of classifier selection in the framework of multiple classifier system for HSI classification. We presented a new unsupervised classifier selection method based on the minimum energy framework to select a subset of reasonably accurate classifiers for the multiple classifier combination. We generated two types of MCS, one with the SVM as a base classifier and another with ELM as a base classifier. The experimental results on two hyperspectral images showed that the subset of classifiers selected by our proposed method outperformed all the baseline approaches. Furthermore, we showed that ELM classifier with a minimum number of hidden neurons has appealing properties in the MCS with respect to SVM classifier.

#### Acknowledgment

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-13-JS02-0005-01 (Asterix project). The authors also would like to thank Prof. D. Landgrebe and Prof. P. Gamba for providing AVIRIS and ROSIS images.

#### 5. REFERENCES

- [1] L.O. Jimenez and D.A. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Tran. on Sys, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 1, pp. 39–54, 1998.
- [2] B.B. Damodaran, R.R. Nidamanuri, and Y. Tarabalka, "Dynamic Ensemble Selection Approach for Hyperspectral Image Classification With Joint Spectral and Spatial Information," *IEEE JSTARS*, vol. 8, no. 6, pp. 2405–2417, jun 2015.
- [3] J. Xia et.al, "SpectralSpatial Classification for Hyperspectral Data Using Rotation Forests With Local Feature Extraction and Markov Random Fields," *IEEE TGRS*, vol. 53, no. 5, pp. 2532–2546, may 2015.
- [4] X. Ceamanos et.al, "A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data," *Int. J. Image Data Fusion*, vol. 1, no. 4, pp. 293–307, dec 2010.
- [5] Z. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Art. Int.*, vol. 137, no. 1-2, pp. 239–263, may 2002.
- [6] Z. Yi, B. Samuel, and W. N. Street, "Ensemble pruning via semi-definite programming," *JMLR*, vol. 7, no. Jul, pp. 1315–1338, 2006.
- [7] P. Gurram and H. Kwon, "Sparse Kernel-Based Ensemble Learning With Fully Optimized Kernel Parameters for Hyperspectral Classification Problems," *IEEE TGRS*, vol. 51, no. 2, pp. 787–802, feb 2013.
- [8] B.B. Damodaran and R.R. Nidamanuri, "Dynamic Linear Classifier System for Hyperspectral Image Classification for Land Cover Mapping," *IEEE JSTARS*, vol. 7, no. 6, pp. 2080–2093, 2014.
- [9] J. Kittler et.al, "On combining classifiers," *IEEE Trans. on PAMI*, vol. 20, no. 3, pp. 226–239, mar 1998.
- [10] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE TGRS*, vol. 43, no. 6, pp. 1351–1362, jun 2005.
- [11] G. Huang et.al, "Extreme learning machine for regression and multiclass classification," *IEEE Tran. on Sys., Man, and Cybernetics. Part B*, vol. 42, no. 2, pp. 513–29, apr 2012.
- [12] J. Ham, M.M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE TGRS*, vol. 43, no. 3, pp. 492–501, mar 2005.