

Dynamic Ensemble Selection Approach for Hyperspectral Image Classification With Joint Spectral and Spatial Information

Bharath Bhushan Damodaran, Rama Rao Nidamanuri, Yuliya Tarabalka

► **To cite this version:**

Bharath Bhushan Damodaran, Rama Rao Nidamanuri, Yuliya Tarabalka. Dynamic Ensemble Selection Approach for Hyperspectral Image Classification With Joint Spectral and Spatial Information. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE, 2015, 8 (6), pp.2405-2417. 10.1109/JSTARS.2015.2407493 . hal-01184338

HAL Id: hal-01184338

<https://hal.inria.fr/hal-01184338>

Submitted on 14 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamic Ensemble Selection Approach for Hyperspectral Image Classification With Joint Spectral and Spatial Information

Bharath Bhushan Damodaran, *Student Member, IEEE*, Rama Rao Nidamanuri, *Senior Member, IEEE*,
and Yuliya Tarabalka, *Member, IEEE*

Abstract—Accurate generation of a land cover map using hyperspectral data is an important application of remote sensing. Multiple classifier system (MCS) is an effective tool for hyperspectral image classification. However, most of the research in MCS addressed the problem of classifier combination, while the potential of selecting classifiers dynamically is least explored for hyperspectral image classification. The goal of this paper is to assess the potential of dynamic classifier selection/dynamic ensemble selection (DCS/DES) for classification of hyperspectral images, which consists in selecting the best (subset of) optimal classifier(s) relative to each input pixel by exploiting the local information content of the image pixel. In order to have an accurate as well as computationally fast DCS/DES, we proposed a new DCS/DES framework based on extreme learning machine (ELM) regression and a new spectral-spatial classification model, which incorporates the spatial contextual information by using the Markov random field (MRF) with the proposed DES method. The proposed classification framework can be considered as a unified model to exploit the full spectral and spatial information. Classification experiments carried out on two different airborne hyperspectral images demonstrate that the proposed method yields a significant increase in the accuracy when compared to the state-of-the-art approaches.

Index Terms—Dynamic classifier selection, dynamic ensemble selection, hyperspectral image classification, markov random field model, multiple classifier system, spectral-spatial classification.

I. INTRODUCTION

HYPERSPECTRAL image provides detailed spectral information in numerous narrow contiguous bands of the electromagnetic spectrum. This capability has led to the widespread use of hyperspectral images as an important data source for a range of applications, such as environmental monitoring, vegetation health monitoring, mineral exploration, military, and defence, etc. [1]–[3]. Supervised image classification has been extensively used to analyze hyperspectral data. However, several factors such as high dimensionality, spatial

and spectral redundancy, interclass variability, noisy bands, and limited labeled samples make the information exploitation of hyperspectral images a very challenging task [4], [5]. Accurate hyperspectral image classification depends on the ability of the chosen classifiers to trade upon the relationship among available labeled samples, data dimensionality, and information classes. There is a high risk that the selected classifier is suboptimal for a problem and data at hand. Multiple classifier system (MCS) has been recently explored to improve the performance of hyperspectral image classification by combining the predictions of multiple classifiers [6]–[8], thereby reducing the dependence on the performance of a single classifier. For the MCS to perform better than the single best (SB) classifier, the classifiers used in the MCS construction have to be diverse, because combining similar classification results may not improve accuracy [9].

Diversity in the MCS can be created explicitly and implicitly. Explicitly, the diversity in the MCS is created by defining a diversity measure and optimizing it. Implicitly, diversity can be introduced by selecting a subset of features [10]–[12], training samples manipulation, selecting classifiers from different categories, and different feature extraction methods [13], [14]. However, the diversity constraint alone does not guarantee that the MCS always performs better. The possibility of inaccurate base classifiers and the incompatible combinations of the classifiers may instead end up the MCS with the suboptimal performance. An ensemble pruning approach has been proposed to select reasonably accurate base classifiers [15]–[17]. In this method, instead of combining all the available base classifiers in the MCS, a subset of classifiers is selected based on the criteria like diversity measures and performance measures for decision fusion. This approach has been expanded by proposing a unified framework [18] which consists of both diversity creation (implicit and explicit) and performance measures of base classifiers, as well as selecting the classifiers with nonzero weights by sparse optimization methods [19] to form an effective MCS. However, the selection of classifiers in this method is independent of the location of the image pixel in the feature space; hence, all the classifiers take part in classifying each image pixel. On the other hand, the optimal subset of classifiers varies for different spatial locations in the image. Therefore, the performance of the MCS can be improved by selecting the best classifier or a subset of classifiers dynamically relative to each image pixel, known as dynamic classifier/ensemble selection (DCS/DES) [20]–[22].

Manuscript received September 04, 2014; revised February 06, 2015; accepted February 11, 2015. Date of publication March 25, 2015; date of current version July 30, 2015.

B. B. Damodaran and R. R. Nidamanuri are with the Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala 695 547, India (e-mail: bhushabhusha@gmail.com; rao@iist.ac.in).

Y. Tarabalka is with the INRIA Sophia Antipolis, 06902 Cedex, France (e-mail: yuliya.tarabalka@inria.fr).

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

Digital Object Identifier 10.1109/JSTARS.2015.2407493

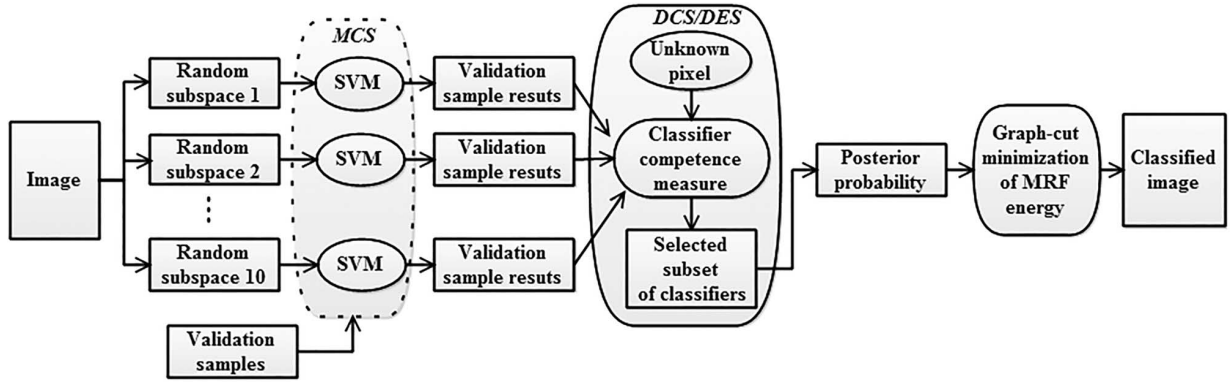


Fig. 1. Flowchart of the proposed spectral-spatial classification method.

The success of DCS depends on an accurate estimation of classifiers competence for a given image pixel. The classifiers forming a dynamic subset are chosen based on estimating the accuracy (competence) of each base classifier in a local region around the image pixel, and the selection is based on the highest accuracy criterion. The classifier competence can be computed by local accuracy (LA) estimation methods [22]–[24] and probabilistic model-based methods [25]. Recently, Du *et al.* have studied the capability of DCS for hyperspectral image classification using LA estimation [26]. However, their study is limited with two optimal classifiers as the base classifiers. When the number of base classifiers increases and in the presence of inaccurate classifiers, the performance of DCS based on LA is uncertain. Both the LA estimation methods and a probabilistic model approach compute the distance between the test pixel and the training (validation) samples. Hence, these techniques are computationally expensive for large-scale problems. The regression-based probabilistic model reduces the computational factor at the cost of the accuracy [25]. However, in practical situations, the accuracy is an important criterion in mapping applications. Recently, extreme learning machine (ELM) has shown good performance in terms of both computational time and accuracy for the classification and regression problems [27]. In this paper, we modeled the DCS problem as the classification problem by mapping the validation samples to the classifier competence measure based on ELM regression. The proposed DCS/DES method based on ELM increases the efficiency in computation time and classification accuracy for the large-scale problems. Our extensive literature review reveals that the potential of the DCS/DES approaches for the hyperspectral image classification is not well studied. Hence, it is highly desirable to study the potential of the DCS/DES approaches, and develop an accurate and computationally efficient DCS/DES methodology for hyperspectral image classification.

Apart from the spectral content, airborne hyperspectral sensors also provide rich spatial information, which has been extensively utilized in recent studies for hyperspectral image classification [28]–[30]. Markov random field (MRF) model is a powerful method for modeling the spatial contextual information, which assumes that the neighboring pixels are likely to belong to the same class [31]–[33]. Most of the state-of-the-art

studies deal with MRF regularization for a single classifier, yielding significant improvement in the classification performance [32], [34], [35]. The performance of MRF regularization depends on the accuracy of the classifier’s probability estimates. It has been shown that the combination of several classifiers yields reliable probability estimates when compared to the single classifier. Thus, the use of an MCS to derive the data energy term for the MRF regularization is likely to improve the classification accuracy. However, only few studies tested the application of an MRF model to the MCS-based image classification [36], [37]. Furthermore, DCS/DES overcomes the structural limitation of a single classifier, as well as classifier combination or classifier fusion (CF) methods and provides more reliable probability estimates. Thus, DCS/DES emerges as a strong candidate to capture the spectral information of a hyperspectral image. There are no studies available in literature which test the application of an MRF model to the DES based image classification. Hence, the combination of the DES to extract spectral information with the MRF model to exploit spatial context into a unified framework would yield a powerful tool for hyperspectral image classification.

The main contributions of this paper are as follows. 1) We tested the performance of different DCS and DES methods, and compared the performance between DCS and DES methods for hyperspectral image classification. 2) We propose an extended version of the probabilistic model-based DES based on ELM approach. 3) We propose a new unified framework to exploit both spectral and spatial information based on DES and MRF models for hyperspectral image classification.

The flowchart of the proposed spectral-spatial DES method is shown in Fig. 1. This framework extracts the spectral information by using the DES method and the spatial information by applying the MRF model. Experiments had been conducted with two multisite airborne hyperspectral images, and results showed that proposed method yielded the improved classification accuracies when compared to the previously proposed techniques.

The remainder of this paper is organized as follows. Section II gives an overall view of the DCS and the proposed approaches. In Section III, we report the experimental results and we discuss and conclude them in Section IV.

II. METHODOLOGY

In this section, we first describe about different DCS/DES methods which we propose to apply for hyperspectral image classification and later these methods are used for comparison with our proposed method. Then, we present the proposed DCS/DES-ELM method and spectral-spatial DES approach. The term DCS indicates that only the best classifier is selected relative to each image pixel, whereas DES indicates that the subset of best classifiers is selected relative to each image pixel.

A. MCS

Let $\Psi = \{\psi_1, \psi_2, \dots, \psi_L\}$ be the base classifiers forming an MCS, and each classifier $\psi_l, l = 1, 2, \dots, L$ be a function $\psi_l : \mathcal{X} \rightarrow \Omega$ from an input space $\mathcal{X} \subseteq R^n$ to a set of class labels $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$ (M is the number of classes). For any given $\mathbf{x} \in \mathcal{X}$, a classifier ψ_l produces a vector of decision values $\mathbf{d} = [d_{l1}, d_{l2}, \dots, d_{lM}]$ and \mathbf{x} is assigned to the class that has the maximum probability (decision) value. The base classifiers have to commit different types of errors in their predications on different parts of the input space, so that the MCS produces more accurate results when compared to individual classifiers. The random subspace method (RSM) is a popular ensemble generation technique to generate multiple input data sources from a single input data, thus creating diversity among the classifiers in an MCS [12].

The RSM partitions hyperspectral image bands into L subsets and each subset contains $\frac{P}{L}$ number of bands, where P denotes the number of bands in the original hyperspectral image. Each input data source generated from the RSM is returned as the input to the learning algorithm ψ . Support vector machines (SVM) have gained interest in hyperspectral image classification due to their ability to deal effectively with high-dimensional data and small training sets [38], [39]. The performance of SVM varies across different input data sources, thus introducing diversity in the MCS. Apart from SVM, RSM also has the capability to mitigate the small sample size problem and offers good classification accuracies in the heterogeneous environment. The SVM coupled with the output from the RSM were used as base classifiers in the MCS. The concept of combining all the base classifiers available in the MCS to obtain the final classified image, known as classifier combination or CF method, was extensively explored. Therefore, here we focus on the DCS/DES approach, which dynamically selects the best (subset of) classifier(s) for a given image pixel. In the following section, we describe the various dynamic classifier approaches that we propose to apply for hyperspectral image classification.

B. DCS and DES Approaches

The basic idea of the DCS is to find the classifier with the highest probability of being correct for a given unseen sample. The selection of correct classifiers and hence the success of the DCS depends on the estimation of the classifiers competence for a given sample. The classifier competence measure is estimated from validation samples (different from training samples and test samples). Apart from the training and test samples, validation samples are also generated

for estimating the classifier competence in DCS. Let $\mathbf{V} = \{(\mathbf{v}_1, j_1), (\mathbf{v}_2, j_2), \dots, (\mathbf{v}_N, j_N)\}$ be the validation set containing pairs of validation samples and their corresponding class labels. A brief description of different methods used to estimate the classifier competence is given below.

1) *DCS/DES by LA Estimate (DCS/DES-LA)*: The DCS/DES-LA estimates accuracy of each classifier in a local surrounding region of the image pixel and selects the classifier that exhibits higher LA [23]. Let \mathbf{x} be an image pixel to be classified and let us consider k -nearest neighbors of \mathbf{x} in the validation set, denoted as $Q(\mathbf{x}) \in \mathbf{V}$. Without loss of generality, we assume that the classifier ψ_l assigns a class label ω_m to the image pixel \mathbf{x} (i.e., $\psi_l(\mathbf{x}) = \omega_m$). Then, the LA of a classifier ψ_l (LA is known as the classifier competence measure) is denoted by

$$LA(\psi_l|\mathbf{x}) = \frac{N_m}{\sum_{i=1}^M N_{im}}, \quad Q(\mathbf{x}) \in \mathbf{V}|\psi_l(\mathbf{v}_j) = \omega_m, \\ j = 1, 2, \dots, k \quad (1)$$

where N_m is the number of correctly classified samples by the classifier ψ_l to the class ω_m in the neighborhood $Q(\mathbf{x})$, and $\sum_{i=1}^M N_{im}$ is the number of the k -nearest samples of \mathbf{x} in \mathbf{V} that have been assigned to the class ω_m by the classifier ψ_l . The classifier which exhibits the highest LA is selected as the adaptive classifier for image pixel \mathbf{x} :

$$l = \arg \max_i LA(\psi_i, \mathbf{x}). \quad (2)$$

However, in this approach, all the neighboring samples are given equal significance and probability values of the classifiers have not been considered. The validation samples that are closer to the image pixel may have more impact than the samples that are farther away. The classifiers probability values are weighted based on the distance to the neighboring samples, to improve the estimation of LA [22], labeled as posterior LA (PLA). The PLA is estimated as

$$PLA(\psi_l, \mathbf{x}) = \frac{\sum_{\mathbf{v}_j \in \omega_m} P(\omega_m|\mathbf{v}_j, \psi_l) w_j}{\sum_{i=1}^M \sum_{\mathbf{v}_j \in \omega_i} P(\omega_i|\mathbf{v}_j, \psi_i) w_j}, \\ Q(\mathbf{x}) \in \mathbf{V}|\psi_l(\mathbf{v}_j) = \omega_m \quad (3)$$

where $\mathbf{v}_j \in Q(\mathbf{x})$, $P(\omega_m|\mathbf{v}_j, \psi_l)$ is the posterior probability value of the validation sample \mathbf{v}_j assigned to the class ω_m by the classifier ψ_l and $w_j = 1/d_j$, d_j is the Euclidean distance between the image pixels \mathbf{x} and \mathbf{v}_j . The classifier that has the maximum PLA is selected for classifying the image pixel \mathbf{x} similar to (2). In order to select a subset of T classifiers, the classifier competence values (PLA, LA) are arranged in descending order and the first T classifiers are selected. The classification process is then performed by using weighted Bayesian average methods [40] and is called DES

$$P(\omega_i/\mathbf{x}) = \sum_{t=1}^T \eta_t p_t(\omega_i/\mathbf{x}), \quad i = 1, 2, \dots, M. \quad (4)$$

The class label is obtained as $\mathbf{x} \in \omega_m, m = \arg \max_i P(\omega_i/\mathbf{x})$, where η_t is the weight of the classifier ψ_t [for instance, it is obtained as $\eta_t = PLA(\psi_t, \mathbf{x})$], and $p_t(\omega_i/\mathbf{x})$ is the resulting posterior probability of class ω_i for a classifier ψ_t .

2) *DCS/DES With Modified LA (DCS/DES-MLA)*: This approach is similar to DCS/DES-LA, except that the LA is estimated using weighted nearest neighbors of the image pixel x . Motivated by the performance of the distance weighted k -NN classifier, Smits [24] used generalized Dudani's weighting scheme for scaling distance with the s th nearest neighbor for scaling the distances as

$$w_r(\mathbf{x}) = \begin{cases} \frac{d_k - d_r}{d_s - d_1} & d_s \neq d_1, s = 3k \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

where d_k is the distance between k th sample and the image pixel x , and d_r is the distance between the r th of the k th nearest neighbor and the image pixel x . The MLA is then estimated as

$$MLA(\psi_l, \mathbf{x}) = \frac{1}{k} \sum_{j \in N(\mathbf{x}) | \psi_l(\mathbf{x}) = \omega_m} w_j. \quad (6)$$

The classifier which exhibits the maximum LA is chosen to classify each image pixel. If a subset of classifiers is selected, then classification is performed similar to (4) with $\eta_t = MLA(\psi_t, \mathbf{x})$. Furthermore, we modified (6) by incorporating classifiers posterior probability values of the neighboring samples for better LA estimation

$$MPLA(\psi_l, \mathbf{x}) = \frac{\sum_{\mathbf{v}_j \in \omega_m} p(\omega_m | \mathbf{v}_j, \psi_l) \cdot w_j}{\sum_{i=1}^M \sum_{\mathbf{v}_j \in \omega_i} p(\omega_m | \mathbf{v}_j, \psi_l) \cdot w_j}, \quad (7)$$

$$Q(\mathbf{x}) \in V | \psi_l(\mathbf{v}_j) = \omega_m$$

where w_j is the weight obtained from (5), and $p(\omega_m | \mathbf{v}_j, \psi_l)$ is the posterior probability value of the validation sample \mathbf{v}_j assigned to the class ω_m .

3) *DCS/DES-Beta Probabilistic Model (DCS/DES-Beta)*: The third employed method to estimate the classifier competence is based on the beta probabilistic model [25]. The classifier competence is modeled as the probability of correct classification of a random reference classifier (RRC). The RRC produces a randomized vector of class supports, such that its expected value is equal to the vector of class supports produced by the classifier ψ_l for each of the samples $\mathbf{v}_j, j = 1, 2, \dots, N$ in the validation set. The RRC depends on the beta probability distribution with the parameters $\alpha_m, \beta_m, m = 1, \dots, M$. The parameters α_m and β_m are derived from the vector of class supports produced by the classifier ψ_l .

Let ω_j be an original class label of the sample $\mathbf{v}_j \in V$, and the classifier ψ_l produces a vector of class supports as $[d_1(\mathbf{v}_j), d_2(\mathbf{v}_j), \dots, d_M(\mathbf{v}_j)]$. The estimation of classifier competence can be summarized as

- 1) estimate the parameters of beta distribution as $\alpha_m = Md_m(\mathbf{v}_j), \beta_m = M[1 - d_m(\mathbf{v}_j)]$;
- 2) construct the RRC and compute its conditional probability of correct classification as

$$P_c(RRC | \mathbf{v}_j) = \int_0^1 b(u, \alpha_{m_j}(\mathbf{v}_j), \beta_{m_j}(\mathbf{v}_j)) \left[\prod_{m=1, m \neq \omega_j}^M B(u, \alpha_m(\mathbf{v}_j), \beta_m(\mathbf{v}_j)) \right] du \quad (8)$$

where $b(u, \alpha_{m_j}(\mathbf{v}_j), \beta_{m_j}(\mathbf{v}_j))$ is the beta probability distribution and $B(u, \alpha_m(\mathbf{v}_j), \beta_m(\mathbf{v}_j)) = \int_0^u b(w, \alpha_m(\mathbf{v}_j), \beta_m(\mathbf{v}_j)) dw$ is the beta cumulative distribution function. The classifier competence (C) for each validation sample is estimated as

$$C(\psi_l, \mathbf{v}_j) = P_c(RRC | \mathbf{v}_j), \quad j = 1, 2, \dots, N; \quad (9)$$

$$l = 1, 2, \dots, L.$$

The classifier competence is computed for all the validation samples, which essentially indicates which classifier is most suited for the validation samples. In order to choose the optimal classifier for a given image pixel, the classifier competence set is generalized to the entire feature space as follows:

$$c(\psi_l, \mathbf{x}) = \frac{\sum_{j=1}^N C(\psi_l, \mathbf{v}_j) \exp(-dist(\mathbf{x}, \mathbf{v}_j)^2)}{\sum_{j=1}^N \exp(-dist(\mathbf{x}, \mathbf{v}_j)^2)} \quad (10)$$

where $dist$ is the Euclidean distance between the image pixel \mathbf{x} and the validation samples. The most competing classifier is selected for each pixel similar to (2). If the subset of competent classifiers is selected, then image pixels are classified by using (4). Criterion (10) is known as potential model (DCS/DES-beta potential). This method eliminates the necessity of finding the nearest neighbors for each image pixel \mathbf{x} ; instead, it weights the validation samples that are closer to \mathbf{x} with high weights and the validation samples that are farther away with low weights. However, it is required to compute N distances for each image pixel, yielding high computational complexity, especially when the image size is large.

In order to reduce the computational complexity, the classifier selection problem can be formulated as the regression problem. Let us consider L classifiers as L classes; the objective is to learn a function that selects a classifier for each of the image pixels. In other words, the classifier selector is a function $f: V \rightarrow C$ that maps from the validation data set to the competence set of validation samples.

Let $\{(\mathbf{v}_1, C(\psi_l, \mathbf{v}_1)), \dots, (\mathbf{v}_j, C(\psi_l, \mathbf{v}_j))\}, l = 1, \dots, L$ be pairs of a validation sample and its corresponding classifier competence value of the classifier ψ_l . For simplicity, let $C(\psi_l) = [C(\psi_l, \mathbf{v}_1), C(\psi_l, \mathbf{v}_2), \dots, C(\psi_l, \mathbf{v}_N)]$, now

$$f(V; \beta_l) = \beta_l^t V \Rightarrow \beta_l^t V = C(\psi_l) \quad (11)$$

where β_l is the parameter to be estimated for the classifier ψ_l . The classifier competence of the image pixel \mathbf{x} can be obtained by

$$c(\psi_l, \mathbf{x}) = \beta_l^t \phi(\mathbf{x}) \quad (12)$$

and the parameter β_l can be obtained by pseudoinverse as

$$\beta_l = (\Phi^t \Phi)^{-1} \Phi^t C(\psi_l) \quad (13)$$

where $\Phi = [\phi(v_1), \dots, \phi(v_n)]$ and $\phi(\mathbf{v})$ is the polynomial transformation of the sample \mathbf{v} as $\sum_{i=0}^r v^i, r = 2, 3, 5$. We empirically set $r = 3$. The classifier that has the maximum competence value in (12) is selected to classify the image pixel

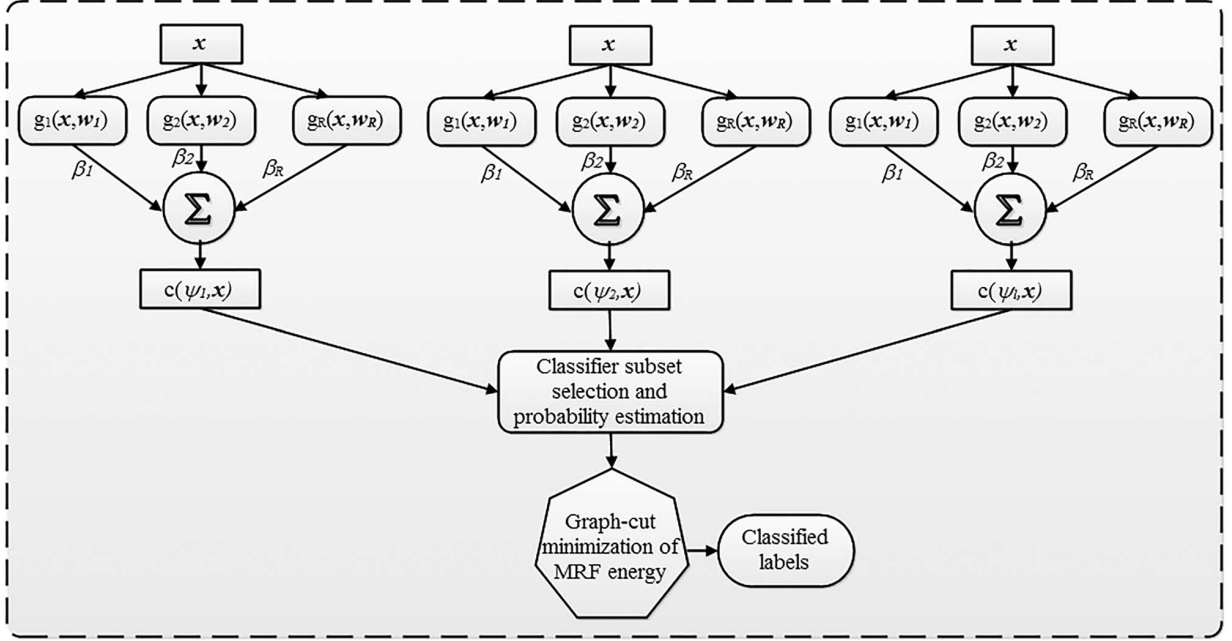


Fig. 2. Flowchart of the proposed DCS/DES-ELM and DES-ELM + MRF method.

\mathbf{x} . We call this method as the DCS-beta least square regression (DCS-beta-LSR). When a subset of classifiers is considered, then the classification is performed by applying (4), and we call it DES-beta-LSR.

C. Proposed DCS/DES-ELM Method and Spectral-Spatial DES Approach (DES-ELM + MRF)

In this section, we propose an extended version of the probabilistic model-based DCS/DES based on ELM approach. Furthermore, this proposed method is regularized with MRF model to develop spectral-spatial DES approach for hyperspectral image classification. The flowchart of the proposed DCS/DES-ELM and spectral-spatial DES is shown in Fig. 2

The performance of the DCS/DES-beta-LSR depends on the feature transformation of the validation and input samples. It has been shown that the DCS/DES-beta-LSR has a suboptimal performance compared to the DCS/DES-beta potential model [25]. Hence, it would be beneficial to have a DCS/DES approach which is independent of feature transformation. Recently, ELM has demonstrated its superior capability to offer better generalization ability and fast training speed for classification and regression problems [27], [41], [42]. In this paper, we propose DCS/DES based on ELM for accurate and fast hyperspectral image classification, labeled as DCS/DES-ELM. The ELM method has the inherent ability to transform the input samples. In addition to that, the performance of the ELM is independent of its parameters, and a wide variety of transformation functions can be used.

The DCS/DES-ELM approach can be modeled as

$$\sum_{i=1}^R \beta_i g_i(\mathbf{w}_i, \mathbf{v}_j) = C(\psi_l, \mathbf{v}_j) \Rightarrow \mathbf{h}(\mathbf{v}_j) \boldsymbol{\beta} = C(\psi_l, \mathbf{v}_j), \quad j = 1, 2, \dots, N \quad (14)$$

where R is number of the hidden nodes, $\mathbf{h}(\mathbf{v}_j) = [g_1(\mathbf{w}_1, \mathbf{v}_j), \dots, g_R(\mathbf{w}_R, \mathbf{v}_j)]$ is the output row vector of hidden layer for input \mathbf{v}_j , $g_i(\mathbf{w}_i, \mathbf{v}_j)$ is the output of the transformation function in the i th hidden node [radial basis function (RBF) is used as the transformation function, and the input weights \mathbf{w}_i are randomly chosen], $\boldsymbol{\beta} = [\beta_1, \dots, \beta_R]^t$ is the output weight between the hidden layer nodes and the output nodes, and $C(\psi_l, \mathbf{v}_j)$ is competence value of the j th validation sample of the classifier ψ_l obtained from (9).

For all the validation samples j , (14) can be represented as

$$\mathbf{H} \boldsymbol{\beta}_l = \mathbf{C}(\psi_l) \quad (15)$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{v}_1) \\ \vdots \\ \mathbf{h}(\mathbf{v}_N) \end{bmatrix}$, the parameter $\boldsymbol{\beta}_l$ is the weight vector of the hidden layer matrix and classifier competence value of the validation samples for the classifier ψ_l , and this can be obtained as

$$\boldsymbol{\beta}_l = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{C}(\psi_l). \quad (16)$$

The classifier competence for an image pixel \mathbf{x} is computed as

$$c(\psi_l, \mathbf{x}) = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{C}(\psi_l) \mathbf{h}(\mathbf{x}). \quad (17)$$

The competence values are arranged in the descending order and the first T classifiers are selected as adaptive classifiers for the image pixel \mathbf{x} . Then, classification is performed by computing the weighted Bayesian average (4), with $\eta_l = c(\psi_l, \mathbf{x})$, and we call it as DES-ELM. The class label is obtained as $\mathbf{x} \in \omega_m, m = \arg \max_i P(\omega_i / \mathbf{x})$ (Fig. 2).

Spectral-Spatial DES approach: In the proposed method, the spatial contextual information is incorporated into the

DES-ELM classification by using the MRF-based regularization model. The DES method is only regularized with MRF model, as the DES method provides better class posterior probability estimates than the DCS method. In the MRF framework, the classification task is formulated as an energy minimization problem on the graph of image pixels. The energy to optimize is computed as a sum of spectral and spatial energy terms and assumes that a pixel belonging to a specific class tends to have neighboring pixels belonging to the same class. The MRF model can be written as

$$\hat{\omega} = \arg \min_{\omega} \left(- \sum_{i \in \mathcal{S}} \log P(\omega_i / \mathbf{x}_i) + \gamma \sum_{j \in N(\mathbf{x}_i)} (1 - \delta(\omega_i, \omega_j)) \right) \quad (18)$$

where $\delta(\cdot)$ is the Kronecker function [$\delta(\omega_i, \omega_j) = 1$ for $\omega_i = \omega_j$; $\delta(\omega_i, \omega_j) = 0$ for $\omega_i \neq \omega_j$], $N(\mathbf{x}_i)$ is the neighboring pixels of \mathbf{x}_i , $\hat{\omega}$ is the resulting class labels from the MRF regularization, \mathcal{S} is the set of all image pixels, and γ is a positive constant parameter that controls the importance of spatial smoothing. The first term $P(\omega_i / \mathbf{x}_i)$ characterizes the spectral information and it is derived from the DES-ELM by employing (4). The second term is expressed by using a Potts model, which favors spatially adjacent pixels to belong to the same land cover class [43]. This MRF regularization is solved by applying an efficient α -expansion graph-cut-based algorithm described in [44].

III. EXPERIMENTAL RESULTS

A. Hyperspectral Image Description

In order to study the potential of DCS/DES for hyperspectral image classification, we adopted two benchmark hyperspectral images with different land cover settings (one in the urban area and one in the agricultural area) captured by two different sensors (ROSIS and AVIRIS).

ROSIS University: The first hyperspectral data set was collected over the University of Pavia, Italy by the ROSIS airborne hyperspectral sensor in the framework of HySens project managed by DLR (German national aerospace agency). The ROSIS sensor collects images in 115 spectral bands in the spectral range from 0.43 to 0.86 μm with a spatial resolution of 1.3 m/pixel. After the removal of noisy bands, 103 bands were selected for experiments. The image contains 610×340 pixels with nine classes of interest. Fig. 3 shows a false color composite (FCC) image and its corresponding ground truth map.

AVIRIS Indian Pines: The second hyperspectral image was collected by the AVIRIS sensor over the Indian Pines site in the Northwestern Indiana. The AVIRIS sensor collects images in 220 spectral bands in the spectral range from 0.43 to 0.86 μm at 20-m spatial resolution. Twenty water absorption bands were removed, and 200 bands were used for experiments. This image contains 145×145 pixels with 16 classes of interest. Fig. 4 shows the FCC image and its corresponding ground truth map.

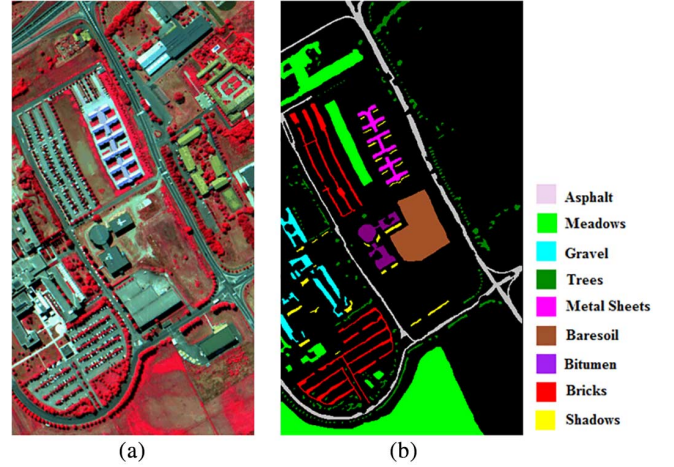


Fig. 3. (a) FCC of the ROSIS University image (R: 0.8340 μm G: 0.6500 μm B: 0.5500 μm). (b) Ground truth image and its corresponding class labels.

B. Design of Experiments

From the available ground truth samples, we randomly selected 100 samples for training, 100 samples for validation, and the remaining samples were used for testing (see Tables I and II). If the total number of available reference samples was lower than 300 samples per class, then 25% of samples were selected for training, another 25% of samples for validation, and remaining samples were used as the testing samples. The experimental results were assessed by overall accuracy (OA), average accuracy (AA), and producer accuracy (PA). In order to avoid the bias induced by random sampling of the training and validation samples, ten independent Monte Carlo runs are performed and the accuracies (OA, AA, PA) are averaged over the ten runs.

In each of the RSM, multiclass pair-wise probabilistic SVM classification with the Gaussian RBF kernel was performed [44]. The SVM parameters in all experiments were automatically tuned by using fivefold cross-validation with $C = 2^\alpha$, $\alpha = \{-5, -4, \dots, 15\}$ and $\gamma = 2^\beta$, $\beta = \{-15, -13, \dots, 3\}$ (C is the cost function and γ is the width of the RBF kernel). When using the DCS-LA and DCS-MLA methods, the classifier competence was estimated based on both strategies (1) and (3), and the best results were retained. The performance of the DCS-LA and DCS-MLA approaches depends on the k -nearest neighbors of the test sample in the validation data set. Hence, we varied the value of k from 3 to 25 and only the best classification accuracies were retained. When more than one classifier was selected, the performance of the DES depended on the number of classifiers (T) included in (4). Hence, in the experiment, we varied the number of classifiers from 2 to 7 and only the best accuracy is reported. However, in most of the Monte Carlo runs, the optimal results were obtained with four and five classifiers. The parameters of ELM regression were automatically tuned using fivefold cross-validation method. In the MRF model (18), the parameter γ that controls the spatial smoothness was tuned empirically for the optimal classification results.

Furthermore, the classifier combination or CF method using Bayesian average algorithm was adapted to combine all the

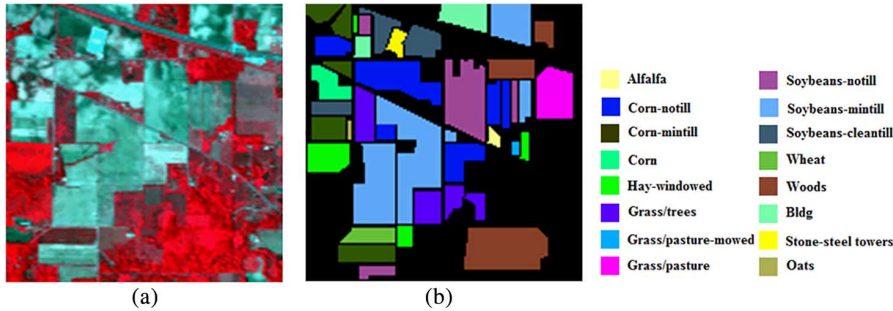


Fig. 4. (a) FCC of the AVIRIS Indian Pines image (R: 0.8314 μm G: 0.6566 μm B: 0.5574 μm). (b) Ground truth image and its corresponding class labels.

base classifiers in the MCS [40] to compare with the DCS/DES-ELM approaches. The results of the proposed spectral-spatial DES (DES-ELM + MRF) were compared with the results of the full-band SVM classification, full-band ELM classification, SB classifier, and CF. Apart from this, the proposed spectral-spatial DES method was compared with the four state-of-the-art approaches, full-band SVM + MRF [32], full-band ELM + MRF, MCS + MRF [36], composite kernels (CK) [30], and SVM ensemble fusion [12]. In the CK [30], the spatial and spectral kernels are combined to exploit the spectral and spatial information of the hyperspectral image. The spatial component of the kernel was derived from the mean of the spectral bands over a spatial neighborhood. The spectral component of the kernel was derived from the spectral information of the individual pixels. The Gaussian RBF and the polynomial kernel were considered in the CK. The experiments were conducted with different combination of the kernels (for instance, polynomial kernel for spectral information and RBF kernel for spatial information), and only the best accuracies were reported (for both images, RBF kernel was used for both spectral and spatial information). We tuned hyperparameters of the composite kernel γ , and C by using fivefold cross-validation and we varied the parameter μ between 0 and 1.

In the SVM ensemble fusion method [12], the hyperspectral image was partitioned into different subsets (i.e., 4 and 6 subsets for the ROSIS University and AVIRIS Indiana Pines hyperspectral images) using correlation coefficient, and each of these subsets was classified by SVM classifier. The decision function values of the individual SVM classifier were fused by one more SVM classifier (the hyperparameters are optimally tuned by fivefold cross-validation). The same number of training and testing samples were used for computing the accuracy of the state-of-the-art approaches.

C. Classification Results of RSM

Table III shows the classification accuracies of SVM classification relative to each random subspace. The results show that there is a considerable variability among the base classifiers in the MCS in terms of OA and class-specific accuracies, thus indicating the suitability of RSM for forming the MCS. The variability of classifiers accuracy is less significant for the University image. There is a 2% accuracy difference between the maximum and minimum overall classification accuracy in the MCS, whereas it is about 8% difference for the AVIRIS Indian Pines hyperspectral image.

TABLE I
NUMBER OF REFERENCE SAMPLES CONSIDERED FOR THE EXPERIMENT OF UNIVERSITY IMAGE

Class name	Training	Validation	Testing
1. Asphalt	100	100	6431
2. Meadows	100	100	18 449
3. Gravel	100	100	1899
4. Trees	100	100	2864
5. Metal sheets	100	100	1145
6. Bare soil	100	100	4829
7. Bitumen	100	100	1130
8. Bricks	100	100	3482
9. Shadows	100	100	747
Total	900	900	40 976

D. Classification Results of the DCS and DES

In this section, the classification results of the different DCS/DES approaches and the proposed DES-ELM, DES-ELM + MRF methods are presented. Tables IV and V summarize accuracies of the DCS/DES for both hyperspectral images. DCS indicates that only the most competent classifier is selected for each image pixel, whereas DES indicates that a subset of classifiers was selected for each image pixel. The DCS, DCS-LA, and DCS-ELM methods yield a marginal increase in overall classification accuracies. The remaining methods resulted in 2% decrease in classification accuracies when compared to the SB classifier for both images. This observation highlights the need to select an adaptive subset of classifiers for each image pixel, instead of choosing only the most competent classifier. This is analogous to the case of selecting multiple classifiers instead of one classifier to avoid the risk of the suboptimal performance.

When a subset of classifiers is chosen for each image pixel and combined by the weighted Bayesian average method, the classification accuracy is significantly improved with all the DES approaches (except the DES-LSR method for the University image). There is a significant improvement in classification accuracy (about 5–6 percentage points) for the Indian Pines image, and a moderate improvement (about 2–3 percentage points) for the University image. Among the DES approaches, the DES-ELM achieved the highest accuracy for the University image and the DES-LA, DES-potential, and DES-ELM achieved the highest accuracies for the Indian Pines image.

The per-class and average-class accuracies have also been improved. There is about 7%–8% improvement in per-class

TABLE II
NUMBER OF REFERENCE SAMPLES CONSIDERED FOR THE INDIAN PINES IMAGE

Class name	Training	Validation	Testing	Class name	Training	Validation	Testing
1. Alfalfa	12	11	23	9. Oats	5	5	10
2. Corn-no till	100	100	1228	10. Soybeans-no till	100	100	772
3. Corn-min till	100	100	630	11. Soybeans-min till	100	100	2255
4. Corn	60	59	118	12. Soybeans- clean till	100	100	393
5. Hay-windowed	100	100	283	13. Wheat	53	52	102
6. Grass/Trees	100	100	530	14. Woods	100	100	1065
7. Grass/pasture-mowed	7	7	14	15. Bldg-grass-trees-drives	100	100	186
8. Grass/pasture	100	100	278	16. Stone-steel towers	24	23	46
Total					1158	1158	7933

TABLE III
OA AND AA (IN PERCENTAGE) OF THE SVM CLASSIFICATION RELATIVE TO EACH RANDOM SUBSPACE AND FULL-BAND HYPERSPECTRAL IMAGE

Image		Full band	RSM 1	RSM 2	RSM 3	RSM 4	RSM 5	RSM 6	RSM 7	RSM 8	RSM 9	RSM10
University	OA	87.12	85.36	84.74	85.47	83.80	84.05	85.35	85.97	82.95	86.19	85.49
	AA	90.16	88.81	88.46	89.19	88.20	88.67	88.63	89.37	87.66	89.40	88.78
Indian Pines	OA	79.04	76.97	76.75	72.32	75.39	75.90	68.23	75.40	72.74	75.25	77.06
	AA	79.43	74.97	76.05	74.74	76.26	76.63	64.92	77.86	72.61	72.42	77.38

TABLE IV
PA, OA, AND AA (IN PERCENTAGE) OF THE DCS AND DES METHODS FOR THE UNIVERSITY IMAGE

Class name	Dynamic classifier selection (DCS)						Dynamic ensemble selection (DES)					
	SB	LA	MLA	Beta-Pot	Beta-LSR	ELM	LA	MLA	Beta-Pot	Beta-LSR	ELM	ELM+MRF
Asphalt	77.66	77.09	79.53	78.0	78.65	78.31	78.39	80.5	79.51	79.97	79.94	97.17
Meadows	86.66	87.05	84.19	81.72	82.24	85.34	88.75	89.22	88.96	69.47	89.78	98.92
Gravel	82.47	86.91	81.6	83.22	82.09	83.4	87.04	86.22	86.29	86.12	86.60	97.3
Trees	94.54	96.31	91.43	92.12	94.14	94.78	96.04	94.79	95.17	78.72	95.51	95.1
Metal sheets	99.48	99.66	99.48	99.55	99.59	99.62	99.62	99.57	99.58	99.48	99.60	99.84
Bare soil	87.48	89.26	85.93	84.23	86.33	88.26	91.29	90.64	90.22	88.26	91.83	99.85
Bitumen	93.12	94.7	91.85	92.8	92.63	93.43	94.87	94.11	94.38	94.27	94.40	98.38
bricks	83.28	79.39	81.38	83.18	81.84	82.02	84.56	85.4	85.75	85.23	85.87	98.59
Shadow	99.93	99.95	99.88	99.79	99.87	99.88	99.95	99.92	99.85	99.93	99.96	99.96
AA	89.40	90.04	88.36	88.29	88.6	89.45	91.17	91.15	91.08	86.83	91.5	98.35
OA	86.19	86.53	84.74	83.48	83.96	86.25	88.18	88.56	88.38	78.16	89.82	98.41

The best accuracies are in bold.

accuracy for most of the classes in the Indians Pines image, while it is moderate with the University image. This observation supports the need of adopting the adaptive classifiers based on local pixel information for enhanced classification performance. However, the poor per-class accuracy is observed with the classes *oats* and *alfalfa*. This is because the classifier fails to characterize the class information due to the presence of the insufficient number of training samples. Furthermore, our proposed DES-ELM approach has outperformed other DES methods in terms of both accuracies and computational time (see Table VI).

From the above observations, we can conclude that DES-ELM better characterizes the spectral information and provides reliable probability estimates and class labels when compared to the other considered methods. The inclusion of spatial contextual information in DES-ELM by the MRF model further significantly increases classification performance. In this case (DES-ELM + MRF), the overall and average classification

accuracies are improved by 12%–15% and by 9%–12% over the SB classifier, respectively. When compared to its earlier version (DES-ELM), about 9% enhancement in the classification accuracy is observed. Furthermore, the class-specific accuracies exceed 95% for medium and large spatial structures, and are less than 95% for small spatial structures (e.g., *trees*, *alfalfa*, and *oats*). The lower per-class classification accuracy of *oats* and *alfalfa* might be due to insufficient number of training samples. The classification maps of the SB, DES-ELM, and DES-ELM + MRF are shown in Figs. 5 and 6. Visual inspection of Figs. 5(a) and (b), and 6(a) and (b) reveal that DES produced smoother classification maps than the SB classifier. Figs. 5(c) and 6(c) confirm a significant increase in classification accuracies and highlight the potential of the MRF model to produce a smooth classification map with spatially connected regions.

Computational time analysis: Table VI shows the computational time of the different DES approaches for both

TABLE V
PA, OA, AND AA (IN PERCENTAGE) OF THE DCS AND DES METHODS FOR THE INDIAN PINES IMAGE

Class name	Dynamic classifier selection (DCS)						Dynamic ensemble selection (DES)					
	SB	LA	MLA	Beta-Pot	Beta-LSR	ELM	LA	MLA	Beta-Pot	Beta-LSR	ELM	ELM+MRF
Alfalfa	62.17	59.13	61.74	63.48	51.3	48.7	56.96	55.65	67.39	50.87	58.7	81.74
Corn-no till	70.2	75.29	67.55	70.98	65.58	69.41	77.5	76.47	77.6	75.52	78.12	81.74
Corn-min till	74.06	75.46	70.33	71.51	71.89	72.89	78.22	78.33	79.52	79.38	79.49	94.05
Corn	79.92	85.17	76.95	74.75	76.44	80.76	88.47	86.69	86.86	86.53	87.37	99.07
Hay windowed	93.22	94.95	90.21	91.31	90.35	92.08	94.35	93.5	93.99	93.22	93.85	97.03
Grass/trees	90.81	97.62	91.17	92.49	92.08	94.94	97.38	96.28	96.62	92.94	96.94	99.55
Grass/pasture-mow	62.86	63.57	47.14	56.43	57.14	50.71	52.14	46.43	65.71	57.86	62.86	82.14
Grass/pasture	99.1	99.39	98.24	98.67	98.96	99.17	99.53	99.5	99.42	99.17	99.46	100
Oats	32	5	2	29	44	39	10	12	39	36	35	38
Soybeans-no till	79.29	89	76.72	75.69	77.95	80.97	89.84	87.09	87.64	86.99	87.33	98.34
Soybeans-min till	67.01	61.66	62.16	64.85	65.46	65.31	72.36	72.79	73.54	73.43	70.36	91.17
Soybeans-clean till	80.84	85.06	77.89	81.93	78.98	83.66	91.12	90.03	90.59	89.59	88.8	98.37
wheat	97.64	99.22	98.43	98.14	97.35	98.04	99.41	99.41	99.51	99.02	99.31	99.51
Woods	87.5	92.07	87.29	87.17	87.88	89.49	93.46	92.66	92.92	91.46	93.16	94.19
Bldy	69.14	69.62	71.67	73.01	69.68	69.95	79.78	81.18	81.88	73.49	76.72	99.68
Stone-steel towers	92.39	96.74	87.17	87.39	89.35	93.26	95.0	92.61	92.17	93.7	93.7	98.91
AA	77.38	78.06	72.92	76.05	75.9	76.77	79.72	78.79	82.77	79.95	81.32	90.84
OA	77.06	78.82	74.37	76.03	75.45	77.16	83.23	82.65	83.4	82.18	83.2	93.12

The best accuracies are in bold.

TABLE VI
CPU PROCESSING TIME (IN SECONDS) OF THE DIFFERENT DES APPROACHES

Image	DES-LA	DES-MLA	DES Beta-Pot	DES Beta-LSR	DES-ELM (proposed approach)
University	2368.35	2164.86	1373.6	8.36	8.53
Indian Pines	169.62	159.92	61.26	8.17	8.27

Since the RSM generation includes for all the methods, the CPU processing time of RSM generation is discarded.

hyperspectral images. The computational time complexity of the DES-LA, DES-MLA, and DES beta potential methods is very high and it grows with the number of pixels to be classified, thus impeding the use of the DES approaches for the large-scale image classification. On the other hand, the DES-LSR and DES-ELM methods are independent of the number of pixels in the image and thus perform very fast image classification. It can be seen from Tables IV to VI that the proposed DES-ELM method outperforms the other DCS/DES approaches in terms of accuracy and computational time.

E. Comparative Performance of the Proposed Methods With the State-of-the Art Approaches

The accuracy of the DES-ELM and the spectral-spatial DES (DES-ELM + MRF) are compared with the state-of-the-art pixel-wise classification methods such as full-band ELM, full-band SVM, CF or MCS, and SVM ensemble fusion method (see Table VII). The proposed DES-ELM approach outperforms the state-of-the-art approaches by 2%–3% for the University image and 4%–6% for the Indian Pines hyperspectral images.

Furthermore, there is a higher magnitude of improvement in accuracy about 12% for University image and 27% for Indian Pines image over the full-band ELM classifier. When compared with the SVM ensemble fusion [12], DES-ELM + MRF yields improvement of the OA by 11 percentage points for the University image and by 16 percentage points for the Indian Pines image. When compared with CF (MCS), it yields about 9% improvement for both images. In order to have a fair comparison, we also compared the performance of the proposed spectral-spatial DES with the state-of-the-art spectral-spatial classification approaches, and the results are reported in Table VII. The proposed DES-ELM + MRF method outperforms SVM + MRF [32], SB + MRF, and MCS + MRF (CF + MRF) [36] techniques by 2%–3.5% for the University image and 1.6%–2.5% for the Indian Pines image. When compared with the CK [30] and full-band ELM + MRF, the DES-ELM + MRF improved accuracy of about 4.8% and 8.2% for the University image, and 8.8% and 11.3% for Indian Pines image. Furthermore, the proposed DES-ELM + MRF approach also yields the highest class-specific accuracies. This observation highlights the potential of merging the advantages of the two different approaches into a unified framework.

In order to examine statistical significance of the results, we have conducted two-tail kappa statistical significance test and the results are shown in Table VIII. The results are statistically significant at 95% confidence interval, if the tabulated value $|Z| > 1.96$. As can be seen from Table VIII, the accuracy differences of the proposed DES-ELM are statistically significant when compared to the SB classifier, CF (MCS), full-band SVM, and SVM ensemble fusion method. However, there is no significant difference between DES-ELM and CF

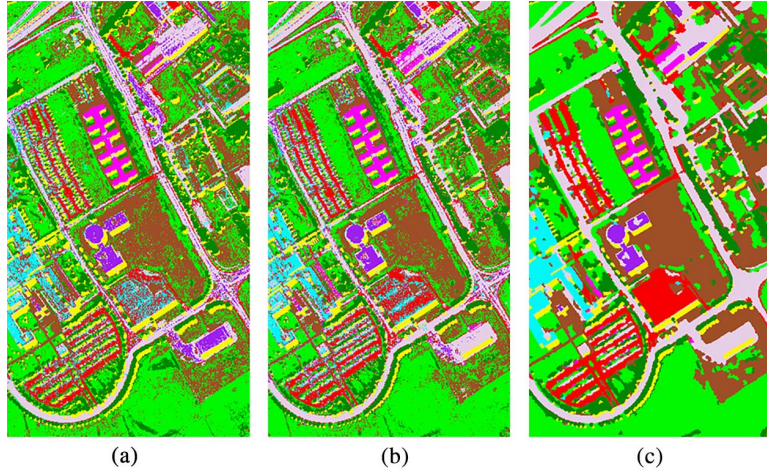


Fig. 5. Classified images of the University image. (a) SB classifier; (b) DES-ELM; and (c) DES-ELM-MRF.

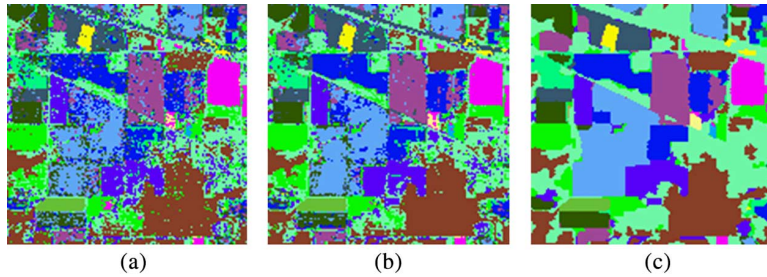


Fig. 6. Classified images of the Indian Pines image. (a) SB classifier; (b) DES-ELM; and (c) DES-ELM-MRF.

TABLE VII
OA, AA (IN PERCENTAGE) OF THE PIXEL-WISE CLASSIFICATION METHODS (FULL-BAND ELM, FULL-BAND SVM, SB, AND CF) AND THE SPECTRAL-SPATIAL CLASSIFICATION METHODS (FULL-BAND ELM + MRF, FULL-BAND SVM + MRF, SB + MRF, AND MCS + MRF)

Image		Full band ELM	Full-band SVM	SB	CF	SVM Fusion [12]	Full band ELM+MRF	Full band SVM+MRF [32]	SB+MRF	MCS+MRF [36]	CK [30]	DES-ELM+MRF
University	OA	77.35	87.12	86.19	88.65	86.76	90.18	95.41	94.73	96.47	93.54	98.41
	AA	83.35	90.16	89.40	91.31	90.24	93.09	95.51	95.55	96.14	94.81	98.35
Indian Pines	OA	50.11	79.04	77.06	83.22	76.59	81.73	90.91	90.5	91.5	84.31	93.12
	AA	47.39	79.43	77.38	81.34	80.97	72.25	85.86	84.74	86.56	85.32	90.84

for the Indian Pines image. The high statistical significance values are observed when the spatial contextual information is incorporated with the DES-ELM approach. This observation confirms the advantage of the spatial contextual models to obtain accurate classification image over the pixel-wise classification methods. Furthermore, the accuracy improvement offered by the proposed DES-ELM + MRF approach is statistically significant when compared with the state-of-the-art spectral-spatial classification models.

IV. DISCUSSION AND CONCLUSION

MCS has evolved as a promising approach for hyperspectral image classification. The classifiers in the MCS are combined in two ways by CF and classifier selection. Many studies have demonstrated that combining multiple classifiers (for instance, Bayesian average) has the potential to deliver significant performance for hyperspectral image classification [8], [46].

However, the classifiers forming the MCS have to be diverse in order to get enhanced performance; otherwise, the result may be suboptimal. It is understood that along with the diversity constraint, the classifiers forming the MCS should also be accurate enough to enrich the performance of the MCS. This requirement is often met by developing methodologies, which select both diverse and high performance classifiers. However, most often all the selected classifiers take part in the decision-making and do not account for local class diversity and distribution variations within the image. Dynamic classifier (ensemble) selection is an alternative way of combining multiple classifiers in the MCS, by selecting a classifier (or a subset of classifiers) relative to each image pixel [25]. Most of the previous studies using MCS for hyperspectral image classification are focused on the classifier combination or CF, while little or no attention has been paid to the classifier selection mechanism.

In this paper, we proposed a new method for hyperspectral image classification, which explores the potential of the

TABLE VIII
KAPPA STATISTICAL SIGNIFICANCE TEST OF DIFFERENT PIXEL-WISE CLASSIFICATION METHODS
AND SPATIAL CONTEXTUAL METHODS OF UNIVERSITY AND INDIAN PINES IMAGE

(a)									
Methods	SB	SVM	CF (MCS)	SVM Fusion	SB+ MRF	SVM+ MRF	MCS+ MRF	CK	DES-ELM+ MRF
DES-ELM	15.09	8.30	6.58	9.98	37.34	35.97	32.95	20.05	42.73
DES-ELM+ MRF	67.67	65.96	46.52	65.29	16.36	15.22	8.34	38.39	0
(b)									
DES-ELM	9.60	6.65	0.23	10.16	15.73	16.87	17.44	3.65	19.29
DES-ELM+ MRF	28.74	25.80	19.26	29.29	5.69	3.94	2.98	18.05	0

The results are considered as significant at 95% confidence interval if the tabulated value $|Z| > 1.96$.

DCS/DES. The LA-based methods and DCS/DES-beta potential compute the distance between each image (test) pixel and the whole set of validation samples, resulting in a computational burden. On the other hand, DCS/DES-beta LSR finds a function that maps the validation data samples to the classifier competence and reduces the computational burden at the cost of accuracy. Hence, it will be beneficial to have an effective framework which reduces the computational burden without reducing the accuracy. We proposed an ELM-based regression framework, which estimates the function mapping validation samples to the classifier competence measure, thus reducing the computational burden without degrading the accuracy. Furthermore, ignoring the spatial correlation among the neighboring pixels yields poor classification. Our proposed spectral-spatial classification framework combines both the spectral information from the DES-ELM and spatial contextual information, resulting in accurate and smooth classification maps. Experimental results show that selecting one best classifier is not an optimal choice and it could end up with the accuracy no better or less than SB classifier. On the other hand, when the subset of classifiers is selected, DES offers 2%–6% increase in classification accuracy. The proposed DES-ELM method outperforms the existing DES methods in terms of accuracy and computational aspects.

Compared to the single classifier, DES-ELM provides reliable probability estimates by alleviating the limitation of the single classifiers and the CF, and emerges as a strong candidate to extract the spectral information. The incorporation of the spatial contextual information shows remarkable performance of about 9% in OA over the pixel-wise classification of DES-ELM. Compared to the pixel-based classification methods (full-band ELM, full-band SVM, SB, CF, and SVM ensemble fusion [12]), there is 9%–27% increase in OA. Similarly, the proposed spectral-spatial DES method shows very high performance when compared to the state-of-the-art spectral-spatial classification approaches (full-band ELM + MRF, full-band SVM + MRF [32], MCS + MRF [36], and CK [30]). Furthermore, Table VIII indicates that there is no significant difference in classification accuracy between CF (MCS) and DES-ELM for the Indian Pines image, but there is a significant accuracy difference when the spatial information is incorporated by MRF model. This indicates the superior capability of

DES-ELM to better characterize the spectral information and provide reliable probability estimates to be used with MRF regularization, when compared to the CF and SB classifiers. In addition, the experiments are performed with few training samples per class (around 5% of total reference samples for the University image and around 20% of total reference samples for the Indian Pines image). The limitation of the DES methods is that the number of classifiers to be selected is fixed and uniform across all the image pixels. In our future work, we would like to propose a strategy to adaptively determine the number of classifiers to be selected relative to each image pixel, so that it could further improve the classification accuracy.

ACKNOWLEDGMENT

The authors gratefully acknowledge Prof. Paolo Gamba, Department of Electronics, University of Pavia, Italy for providing us with ROSIS hyperspectral image and ground truth map used in this study. B. B. Damodaran acknowledges the Doctoral Fellowship provided by the Indian Institute of Space Science and Technology, Department of Space, Government of India. The highly constructive criticism and suggestions of the associate editor and anonymous reviewers are gratefully acknowledged.

REFERENCES

- [1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [2] A. Ghiyammat and H. Z. M. Shafri, "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment," *Int. J. Remote Sens.*, vol. 31, no. 7, pp. 1837–1856, Apr. 2010.
- [3] J. L. Boggs, T. D. Tsegaye, T. L. Coleman, K. C. Reddy, and A. Fahsi, "Relationship between hyperspectral reflectance, soil nitrate-nitrogen, cotton leaf chlorophyll, and cotton yield: A step toward precision agriculture," *J. Sustain. Agric.*, vol. 22, no. 3, pp. 5–16, Jul. 2003.
- [4] L. O. Jimenez and D. A. Landgrebe, "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 28, no. 1, pp. 39–54, Feb. 1998.
- [5] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Atli Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
- [6] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

- [7] A. B. Santos, A. De A. Araujo, and D. Menotti, "Combining multiple classification methods for hyperspectral data interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1450–1459, Jun. 2013.
- [8] S. Samiappan, S. Prasad, and L. M. Bruce, "Non-uniform random feature selection and Kernel density scoring with SVM based ensemble classification for hyperspectral image analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 792–800, Apr. 2013.
- [9] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.
- [10] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
- [11] J. M. Yang, B. C. Kuo, P. T. Yu, and C. H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2840–2853, Jul. 2010.
- [12] 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.
- [13] K. L. Bakos and P. Gamba, "Combining hyperspectral data processing chains for robust mapping using hierarchical trees and class memberships," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 968–972, Sep. 2011.
- [14] B. B. Damodaran and R. R. Nidamanuri, "Assessment of the impact of dimensionality reduction methods on information classes and classifiers for hyperspectral image classification by multiple classifier system," *Adv. Space Res.*, vol. 53, no. 12, pp. 1720–1734, Nov. 2014.
- [15] L. Zhang and W.-D. Zhou, "Sparse ensembles using weighted combination methods based on linear programming," *Pattern Recognit.*, vol. 44, no. 1, pp. 97–106, Jan. 2011.
- [16] Z. Yi, B. Samuel, and W. N. Street, "Ensemble pruning via semi-definite programming," *J. Mach. Learn. Res.*, vol. 7, pp. 1315–1338, 2006.
- [17] Z. H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, no. 1–2, pp. 239–263, May 2002.
- [18] B. B. Damodaran and R. R. Nidamanuri, "Dynamic linear classifier system for hyperspectral image classification for land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2080–2093, Jun. 2014.
- [19] P. Gurram and H. Kwon, "Sparse Kernel-based ensemble learning with fully optimized Kernel parameters for hyperspectral classification problems," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 787–802, Feb. 2013.
- [20] K. Kirchhoff and J. A. Bilmes, "Dynamic classifier combination in hybrid speech recognition systems using utterance-level confidence values," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP'99)*, 1999, vol. 2, pp. 693–696.
- [21] R. Singh, M. Vatsa, and A. Noore, "Multiclass mv-granular soft support vector machine: A case study in dynamic classifier selection for multispectral face recognition," in *Proc. 19th Int. Conf. Pattern Recog.*, 2008, pp. 1–4.
- [22] L. Didaci, G. Giacinto, F. Roli, and G. L. Marcialis, "A study on the performances of dynamic classifier selection based on local accuracy estimation," *Pattern Recognit.*, vol. 38, no. 11, pp. 2188–2191, Nov. 2005.
- [23] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.
- [24] P. C. Smits, "Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 801–813, Apr. 2002.
- [25] T. Woloszynski and M. Kurzynski, "A probabilistic model of classifier competence for dynamic ensemble selection," *Pattern Recognit.*, vol. 44, no. 10–11, pp. 2656–2668, Oct. 2011.
- [26] P. Du, J. Xia, J. Chanussot, and X. He, "Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 174–177.
- [27] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
- [28] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioned clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [29] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012.
- [30] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite Kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [31] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.
- [32] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.
- [33] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.
- [34] L. Xu and J. Li, "Bayesian classification of hyperspectral imagery based on probabilistic sparse representation and Markov random field," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 823–827, Apr. 2014.
- [35] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 803–817, Feb. 2013.
- [36] X. He, J. Xia, and P. Du, "MRF-based multiple classifier system for hyperspectral remote sensing image classification," in *Multiple Classifier Systems*. Berlin, Germany: Springer, vol. 7872, 2013, pp. 343–351.
- [37] M. Khodadadzadeh *et al.*, "Spectral-spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6298–6314, Oct. 2014.
- [38] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.
- [39] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [40] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
- [41] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E²ELM: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
- [42] M. A. Bencherif *et al.*, "Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 527–531, Mar. 2015.
- [43] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 631–651, Mar. 2013.
- [44] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [45] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [46] G. Thoonen, Z. Mahmood, S. Peeters, and P. Scheunders, "Multisource classification of color and hyperspectral images using color attribute profiles and composite decision fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 510–521, Apr. 2012.



Bharath Bhushan Damodaran (S'12) received the M.Sc degree in mathematics from Bharathiar University, Coimbatore, India, in 2008, the M.Tech degree in remote sensing and wireless sensor networks from Amrita Vishwa Vidyapeetham, Coimbatore, India, in 2010, and the Ph.D degree in earth and space sciences from the Indian Institute of Space Science and Technology, Trivandrum, India, in 2015.

Currently, he is a Project Associate with the Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Trivandrum, India. His research interests include multiple classifier system, hyperspectral/multispectral image analysis, machine learning, and image processing.



Rama Rao Nidamanuri (SM'12) received the B.S. degree in mathematics and computer sciences from Nagarjuna University, Guntur, India, in 1996, the M.S. degree in space physics from Andhra University, Visakhapatnam, India, in 1998, the M.Tech degree in remote sensing from Birla Institute of Technology, Ranchi, India, in 2002, and the Ph.D. degree in remote sensing from the Indian Institute of Technology, Roorkee, India, in 2006.

Currently, he is an Associate Professor in Remote Sensing and Image Processing with the Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Trivandrum, India. Since 2008, he has been a Research Fellow of the Alexander von Humboldt Foundation, Bonn, Germany, and a Guest Researcher with the Leibniz Centre for Agricultural Landscape Research, Muencheberg, Germany. His research interests include multiple classifier systems, multi-spectral/hyperspectral image analyses methods and algorithms, UAV remote sensing, and search based classification methods.



Yuliya Tarabalka (S'08–M'10) received the B.S. degree in computer science from Ternopil Ivan Pul'uj State Technical University, Ukraine, in 2005, and the M.Sc. degree in signal and image processing from the Grenoble Institute of Technology (INPG), Grenoble, France, in 2007. She received a joint Ph.D. degree in signal and image processing from INPG and in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 2010.

From July 2007 to January 2008, she was a Researcher with the Norwegian Defence Research Establishment, Norway. From September 2010 to December 2011, she was a Postdoctoral Research Fellow with the Computational and Information Sciences and Technology Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA. From January to August 2012, she was a Postdoctoral Research Fellow with the French Space Agency (CNES) and Inria Sophia Antipolis-Méditerranée, France. Currently, she is a Researcher with the TITANE team of Inria Sophia Antipolis-Méditerranée. Her research interests include image processing, pattern recognition, hyperspectral imaging, and development of efficient algorithms.