Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data

Björn Waske, *Member, IEEE*, Sebastian van der Linden,
Jón Atli Benediktsson, *Fellow, IEEE*, Andreas Rabe, and Patrick Hostert

***Abstract*—The accuracy of supervised land cover classifications depends on factors such as the chosen classification algorithm, adequate training data, the input data characteristics, and the selection of features. Hyperspectral imaging provides more detailed spectral and spatial information on the land cover than other remote sensing resources. Over the past ten years, traditional and formerly widely accepted statistical classification methods have been superseded by more recent machine learning algorithms, e.g., support vector machines (SVMs), or by multiple classifier systems (MCS). This can be explained by limitations of statistical approaches with regard to high-dimensional data, multimodal classes, and often limited availability of training data. In the presented study, MCSs based on SVM and random feature selection (RFS) are applied to explore the potential of a synergetic use of the two concepts. We investigated how the number of selected features and the size of the MCS influence classification accuracy using two hyperspectral data sets, from different environmental settings. In addition, experiments were conducted with a varying number of training samples. Accuracies are compared with regular SVM and random forests. Experimental results clearly demonstrate that the generation of an SVM-based classifier system with RFS significantly improves overall classification accuracy as well as producer's and user's accuracies. In addition, the ensemble strategy results in smoother, i.e., more realistic, classification maps than those from stand-alone SVM. Findings from the experiments were successfully transferred onto an additional hyperspectral data set.**

***Index Terms*—Classifier ensembles, hyperspectral data, multiple classifier systems (MCSs), random feature selection (RFS), support vector machines (SVMs).**

I. Introduction

REMOTE sensing applications, such as land cover classification, provide a variety of important information for decision support and environmental monitoring systems. When complex environments are mapped or when very detailed analyses are performed, the spectral and spatial resolution requirements can be very high, e.g., in urban area mapping, the characterization of mineral composition, or in plant-type differentiation [1], [2]. In such situations, airborne hyperspectral sensors are—at the moment—probably the most valuable single data source. Data from these sensors provide detailed and spectrally continuous spatial information on the land surface, ranging from the visible to the short-wave infrared regions of the electromagnetic spectrum. They enable discriminating between spectrally similar land cover classes that occur at highly frequent spatial patterns [3].

However, it is well known that increasing data dimensionality and high redundancy between features might cause problems during data analysis, e.g., in the context of supervised classification: The overall map accuracy can decrease when only a limited number of training samples are available [4]. Against this background, machine learning algorithms such as support vector machines (SVMs) and concepts like multiple classifier systems (MCSs) have emerged over the past decade [5]–[9]. SVMs construct an optimal separating hyperplane between training samples of two classes within the multidimensional feature space. In linear nonseparable cases, the data are mapped using a kernel function into a higher dimensional feature space. This enables the definition of a separating hyperplane, which appears nonlinear in the original feature space. Based on this so-called kernel trick, SVM can describe complex classes with multimodal distributions in the feature space. Despite their relatively good performance when high-dimensional data are classified with only small training sets, a sufficient number of samples should be considered to ensure that adequate training samples are included during supervised classification [7]. Recent studies discuss the use of SVM for spectral–spatial classification of urban hyperspectral data [10]–[12] and multisource classification [13]–[15] and extend the supervised SVM techniques by semisupervised concepts [16], [17].

The concept of MCS, on the other hand, does not refer to a specific algorithm but to the idea of combining outputs from more than one classifier to enhance classification accuracy [18]. These outputs may result from either a set of different classification algorithms or independent variants of the same algorithms (i.e., the so-called base classifier). The latter are achieved by modifying aspects of the input data, to which the base classifier is sensitive during separate training processes. This includes the following: the generation of training sample subsets, named bagging [19], the sequential reweighting of training samples, boosting [20], and the generation of feature subsets, e.g., by

Manuscript received July 2, 2009; revised November 19, 2009. Date of publication March 29, 2010; date of current version June 23, 2010. This work was supported in part by the Research Fund of the University of Iceland and in part by the Icelandic Research Fund.

B. Waske is with the Institute of Geodesy and Geoinformation, Faculty of Agricultural, University of Bonn, 53115 Bonn, Germany (e-mail: bwaske@uni-bonn.de).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland.

S. van der Linden, A. Rabe, and P. Hostert are with the Geomatics Laboratory, Geography Department, Humboldt-Universität zu Berlin, 10099 Berlin, Germany.

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/TGRS.2010.2041784

random feature selection (RFS) [21]. Afterward, outputs from the various instances of the same base classifier are combined to create the final class decision [18]. Many MCS applications employ base classifiers of relatively little computational demand, usually self-learning decision trees (DTs), which construct many rather simple decision boundaries that are parallel to the feature axis. MCSs were recently reviewed in the context of remote sensing [22] and yield excellent results when dealing with hyperspectral and multisource data sets [8], [9], [23]–[25].

In summary, the concepts of SVM and MCS are based on different ideas: While the first focuses on optimizing a single processing step, i.e., the fitting of the presumably optimal separating hyperplane, the latter relies on an ideally positive influence of a combined decision derived from several suboptimal yet sometimes computationally simple outputs. Nevertheless, the two approaches are not exclusive, and it appears desirable to combine them in a complementary approach. Studies trying to employ computationally more complex classifiers such as neural networks [26] and SVMs [27] in MCSs exist, but they are very limited, particularly in the remote sensing context. In [28], the use of an SVM ensemble for the spectral–contextual classification of a single high-spatial resolution remote sensing image was shown to increase the accuracy of individual results. Results in a non-remote-sensing context are more controversial, though: Whereas bagging and boosting were successfully used with SVMs in [27], the results of an empirical analysis in [29] demonstrate that ensembles based on bagging and boosting are not generally better than a single SVM in terms of accuracy. Given these discrepancies and the significantly higher computational demand of SVM compared to DT, the idea of an MCS based on SVM requires well thought and systematic approaches to develop an SVM ensemble concept that leads to a general trend of increasing accuracy and to give appropriate guidelines to achieve efficient processing, i.e., reliable default parameters.

We present an SVM ensemble that uses RFS to generate independent variants of SVM results. This novel combination of the two classifier concepts appears particularly useful with regard to the high dimensionality and redundancy of hyperspectral information. We expect the results of this MCS to show clearer trends than those reported in previous studies that use bagging or boosting [27]–[29]. In order to evaluate the potential of such a concept, we focus on three main research questions.

1) Is there a significant increase in accuracy compared to regular SVM and advanced DT classification, when RFS is performed to construct SVM ensembles?
2) What is the impact of the two parameters, namely, feature subset size and ensemble size, on the accuracy and stability of ensembles in terms of the classification result?
3) Is it possible to derive default values or recommendations for the parameters in (2) in order to make the use of SVM ensembles with RFS feasible?

To answer these three research questions, the specific objective of our study is the classification of two different hyperspectral data sets, i.e., an urban area from the city of Pavia, Italy, and a volcanic area from Hekla volcano, Iceland, with various SVM ensembles. These are generated by systematically increasing the number of randomly selected features before SVM classification as well as the number of repetitions that are combined for the final class decision. Moreover, the size of the training sets is varied to investigate the possible influences of the number of training samples on the classification accuracy.

This paper is organized as follows. Section II introduces the conceptual and algorithmic framework of SVM and MCS. The used SVM ensemble strategy is explained in Section III. The experimental setup and experimental results are presented in Section IV. Section V discusses results, followed by the conclusion in Section VI.

II. Background

A. SVMs

The SVM is a universal learning machine for solving binary classification problems and can be seen as an approximate implementation of Vapnik's Structural Risk Minimisation principle, which has been shown superior to traditional Empirical Risk Minimisation principle [30]. SVMs are able to separate complex (e.g., multimodal) class distributions in high-dimensional feature spaces by using nonlinear kernel functions and to deal with noise and class confusion via a regularization technique. A detailed introduction on the general concept of SVM is given in [31]; an overview in the context of remote sensing can be found, e.g., in [5] and [6].

In this paper, a Gaussian radial basis function (RBF) kernel function is used, which is widely accepted in remote sensing applications. A one-against-one (OAO) strategy is used to handle multiclass problems with the originally binary SVM. The OAO class decision is determined by a majority vote using classwise decisions [6], [7].

B. MCSs

The concept of classifier ensembles or MCSs is based on the hypothesis that independent classifiers generate individual errors, which are not produced by the majority of the other classifiers within the ensemble. The basic aim in MCS is therefore to generate a diverse set of classifiers, making each individual classifier as unique as possible [18].

Among MCS approaches based on iterative and independent variations of a base classifier, bagging and boosting are perhaps the widest used concepts for the construction of classifier ensembles. Boosting [20] consecutively modifies the training data by adaptively weighting the training samples after each individual iteration. Accurately classified samples are assigned a lower weight than those samples classified incorrectly. Bagging, on the other hand, randomly generates a set of training sample subsets. Each sample subset is used to train an individual base classifier, and the outputs are combined to generate a final class decision. While boosting is performed in series, bagging [19] can be performed in parallel and may result in lower computation time.

The generation of classifier ensembles by *RFS*—also known as the random subspace method or *attribute bagging*—is another method used to train an MCS [21], [32]. Each base classifier is trained on randomly generated and independently drawn feature subsets. The diverse outputs are combined to define the final MCS class decision, often by a majority vote. In [33], the concept was successfully applied to a set of multitemporal SAR images using a DT as a base classifier.

Breiman [34] introduced random forests (RF) which is a DT-based classifier concept based on training sample and

feature subsets. Each tree within the ensemble is trained on bootstraps of the training samples (i.e., the same as in bagging). In addition, at each internal node of the DT, a randomly generated feature subset is used.

III. SVM Ensemble Strategy

In [29], SVM ensembles were constructed by bagging and different boosting variants. On average, these ensembles outperformed a standard SVM in terms of accuracy. Nevertheless, a general improvement of MCS strategies by incorporating SVM cannot be stated based on the results in [29], and greater computational time does not necessarily result in higher classification accuracies. On the other hand, increased computations may sometimes, in fact, reduce the accuracy. In [35], the stability of SVM in terms of classification performance was investigated following a bagging strategy. The obtained experimental results underline that bagging does not improve nor even slightly decrease the classification results of SVM. Thus, Buciu *et al.* [35] consider SVM as a stable classifier with regard to bagging. However, the instability of the base classifier, i.e., a small change in the training samples leads to varying classification results, is an important requirement for a bagging ensemble. Boosting, on the other hand, is computationally very demanding, because it must be processed in series and performs less efficient than bagging in the context of SVM [29]. In regard to these findings and the fact that ensembles generated by RFS can outperform bagging and boosting in terms of accuracy [21], [32], the construction of SVM ensembles by RFS seems more adequate.

Whereas bagging is based on the selection of training samples, RFS modifies the d-dimensional feature space $\Re^d$, and each classifier within the ensemble is generated by randomly choosing a feature subspace $\Re^{d'}$ of user-defined size d'. Ho [21] recommends the selection of $d' \approx d/2$, features for a DT-based classifier system, where d is the total number of features.

By using RFS for generating an ensemble of size z for solving a classification problem with n classes, each classifier with class decision $Y_k(\mathbf{x}) \in \{1, \ldots, n\}$ and $1 \leq k \leq z$ is trained on a randomly generated feature subset. The feature subsets are usually generated *without replacement*, and d' features are selected out of the whole feature set of size d. As in bagging, the different classifications are often combined to the final class decision $Y(\mathbf{x})$ by a simple majority vote using classwise scores

$$Y(\mathbf{x}) = \underset{i \in \{1,\ldots,n\}}{\arg\max}\, S_i(\mathbf{x}) \tag{1}$$

$$S_i(\mathbf{x}) = \sum_{k=1}^{z} \begin{cases} 1, & \text{if } Y_k(\mathbf{x}) = i \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Our proposed strategy for the MCS is shown in Fig. 1. As a first step after preprocessing, an RFS is performed to perform various sets of feature subspaces of size d' (the feature subset size). Afterward, an individual SVM is applied on the feature subset, providing an individual classification result. These steps are performed z times, either parallel or in series, with z being the number of classifiers within the ensemble (i.e., ensemble size). The z classification outputs are combined with a majority vote, according to (1). The proposed fusion strategy has been compared to alternative approaches, e.g., majority vote on binary decision and class probability values, which, in no case, further improved the results. We believe that a detailed analysis of more sophisticated fusion strategies may be worthwhile. It requires a separate analysis that goes beyond the scope of this paper, however.

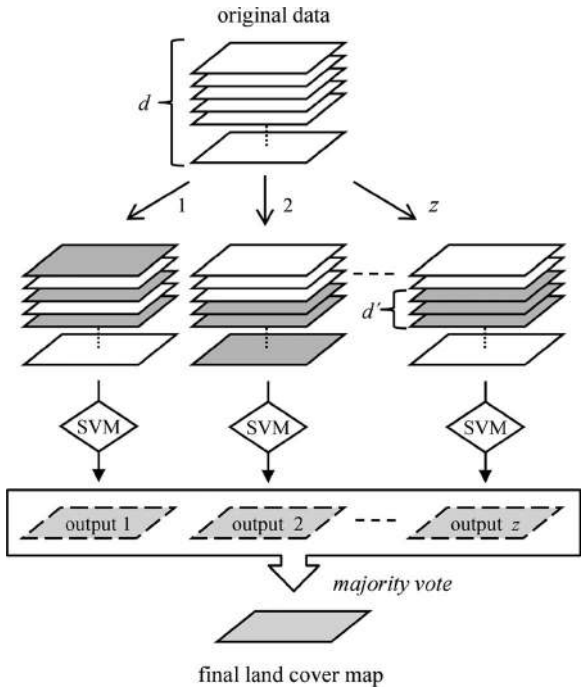

Fig. 1. Schematic diagram of the RFS using z iterations and selecting d' out of d available features. It should be noted that the classification of each individual SVM already requires a voting strategy (Section II-A). This is not shown in the diagram, and the individual outputs refer to a land cover map containing class labels between 1 and c.

IV. Experimental Setup and Results

A. Experimental Setup

Two hyperspectral data sets from study sites with different environmental setting were used in this study. Both classification problems are set up in ways that require hyperspectral information for an appropriate description of target surface types, i.e., lava types from different mineral composition and age as well as urban surfaces. The latter study also requires a very high spatial resolution in order to avoid a high fraction of mixed pixels.

The first study site lies around the central-volcano Hekla, one of the most active volcanoes in Iceland (Fig. 2). The image was collected by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) on a cloud-free day (June 17, 1991). AVIRIS operates from the visible to the short-wave infrared regions of the electromagnetic spectrum, ranging from 0.4 to 2.4 μm. Due to a malfunction, spectrometer 4 operating in the wavelength range from 1.84 to 2.4 μm was working incorrectly. Thus, the bands from spectrometer 4 were deleted from the image data along with the first channels of each of the three spectrometers, which contained noise. Finally, 157 data channels were left. The image strip is 2048 $\times$ 614 pixels, with a spatial resolution of 20 m [36]. The classification aims on 22 land cover classes, mainly lava flows from different eruptions and older hyaloclastites.

The second data set was acquired by the ROSIS-3 (Reflective Optics System Imaging Spectrometer) sensor over the city

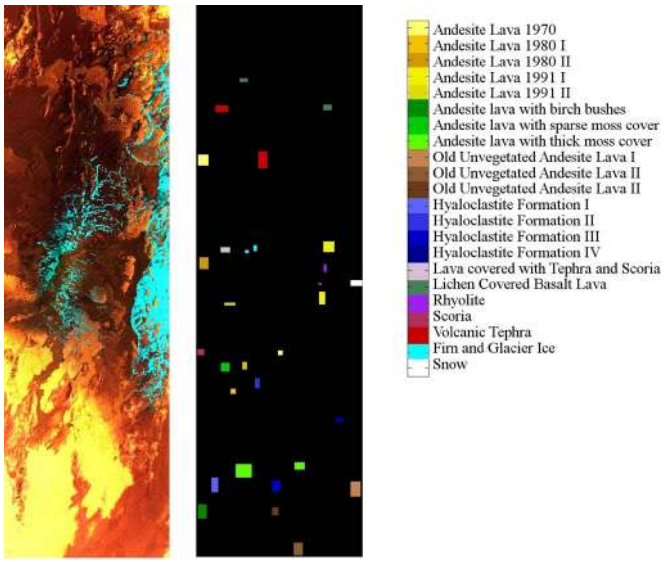

Fig. 2. AVIRIS data, Hekla, Iceland. False-color composite and corresponding ground truth areas representing 22 land cover classes.

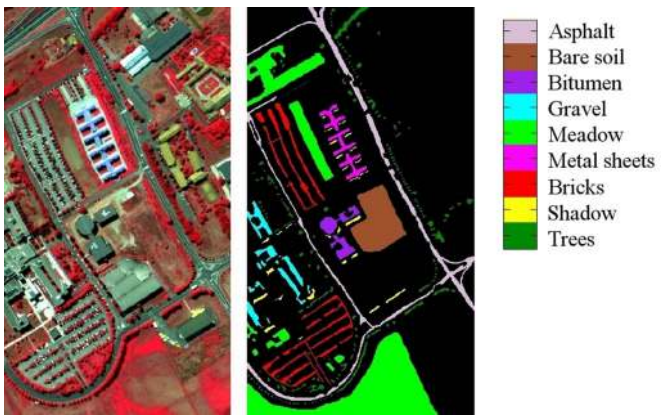

Fig. 3. ROSIS-3 data, Pavia, Italy. False-color composite and corresponding ground truth areas representing nine land cover classes.

of Pavia, Italy (Fig. 3). The 115 bands of the sensor cover the 0.43–0.86-μm range of the electromagnetic spectrum. The spatial resolution of the data set is 1.3 m per pixel. The image strip is 610 $\times$ 340 pixels, sounding the University of Pavia. Some bands have been removed due to noise; the remaining 103 channels have been used in the classification. The classification is aiming nine land cover classes.

For both data sets, ground truth information was used for generating training and validation sample sets, using expert knowledge in image interpretation. An equalized random sampling was performed, guaranteeing that all classes are equally included in the sample set. To investigate the possible influence of the number of training samples on the performance of the proposed method, training sets with different size were generated, containing 25, 50, 100, and 200 training samples per class, respectively (from now on referred to as tr#25, tr#50, ...). For each data set, an independent test set was available, containing 14 966 and 40 002 samples, respectively. Several experiments were conducted to investigate the sensitivity of SVM classifiers to RFS. Diverse SVM ensembles were generated for the two data sets using the following: 1) feature subsets of different sizes ($10\%, 20\%, \ldots, 90\%$ of all available features) and 2) different numbers of classifiers within the ensemble (10, 25, and 50). Aside from this, a standard SVM, which is based on the whole data set, and an RF classifier [34] were applied to the images using the same training sets.

TABLE I
DATA SET 1—HEKLA. OVERALL ACCURACY (IN PERCENT) USING DIFFERENT METHODS AND NUMBER OF TRAINING SAMPLES. THE SVM ENSEMBLES ARE BASED ON THE RANDOMLY SELECTION OF 30% OF ALL FEATURES. *INDICATES A SIGNIFICANT DIFFERENCE ($\alpha = 0.01$) IN COMPARISON TO THE REGULAR SVM

Method	Training sample sets			
	#25	#50	#100	#200
SVM	82.2	90.2	94.8	96.9
RF	83.3	88.1	90.5	92.7
SVM ensemble 50 iterations	89.7*	94.1*	96.6*	97.7*
SVM ensemble 25 iterations	89.4*	94.0*	96.5*	97.6*
SVM ensemble 10 iterations	88.3*	93.0*	95.7*	97.1

The training and classification were performed using imageSVM [37], which is a freely available IDL/ENVI implementation. ImageSVM[1] is based on the LIBSVM approach by Chang and Lin [38] for the training of the SVM. A Gaussian RBF kernel was used, and a regularization parameter C and a kernel parameter g are determined by a grid search using a threefold cross validation. Training and grid search are performed for each SVM classifier in the ensemble, i.e., in the case of an ensemble size of 50, individual SVMs are trained on 50 different feature subsets with each training including its own grid search.

The RF classification was performed with WEKA [39] using 100 iterations. First experiments have shown that an increased number of iterations do not significantly improve the total accuracy. The number of features at each node was set to the square root of the number of input features. This value is considered adequate in literature [24] and proved reliable in a previous studies by the authors [25].

Accuracy assessment was performed, giving overall accuracies and confusion matrices that were used to calculate the producer's and user's accuracies. Based on 99% confidence intervals, a statistical comparison of ensemble-based results and maps from regular SVM was performed.

B. Results for Data From Hekla, Iceland

The results demonstrate that the SVM ensemble outperformed the regular SVM and RF classifier in terms of overall accuracy for the four training sample sets. The RF achieved an overall accuracy between 83.3% and 92.7%; a regular SVM, on the other hand, achieved accuracies between 82.2% and 96.9%. In contrast to this, an SVM ensemble that is based on RFS achieves overall accuracies between 88.3% and 97.7% (Table I).

These results clearly underline that the overall accuracy can be increased by the SVM ensemble. However, a strong

[1]Software available at http://www.hu-geomatics.de.

influence of the number of training samples on classification accuracy can be observed: For all methods, accuracies increase monotonically with sample set size. The rate of this increase is highest for the regular SVM (14.7% difference between tr#25 and tr#200) and lowest for SVM ensembles with 50 iterations (8%). SVM ensembles with ten iterations outperform RF and a regular SVM in terms of accuracy for all training sample sets. The overall accuracy was further improved by increasing the number of iterations from 10 to 25. No significant further increase takes place for larger ensembles.

Feature subset size significantly affected the overall accuracy (Fig. 4).

Irrespective of the number of training samples, the use of only 10% of the features (i.e., 16 bands) was ineffective in terms of accuracies (e.g., 93.4% accuracy with 50 iterations and tr#100). In many cases, the accuracies were below the accuracies for the regular SVM. In contrast to this, the use of 20% of the available features increased the accuracy and outperformed a regular SVM classifier in terms of accuracy, or performed at last equally well. The maximum accuracy is achieved by generating an ensemble with 30% of the features (e.g., 96.6% with 50 iterations and tr#100). The use of additional features did not further improve the overall accuracy. In the case of tr#25 and tr#50, the overall accuracy decreases significantly when feature subset size is further increased.

With regard to the class accuracies, the proposed strategy outperformed the other methods in the majority of the cases. In Fig. 5, the differences between the producer's and users's accuracy, achieved by the SVM ensemble and a regular SVM, are shown. While some classes show almost no differences, such as the three vegetation-covered *Andesite Lava* classes (classes 6, 7, and 8) as well as *Firn and Glacier Ice* and *Snow* (classes 21 and 22), the difference tends significantly toward the positive in the majority of the classes, i.e., the ensemble achieves higher class accuracies. As for the overall accuracy, this effect is reduced with an increasing number of training samples (Fig. 4). In the case of tr#50, the class accuracies are often improved by 5% or even more using the ensemble approach. On the other hand, the improvement in the case of the training set tr#200 is less significant and usually below 5%. Only the user accuracies for classes 5 and 14 are improved by more than 5% (*Andesite Lava 1991 II* and *Hyaloclastite Formation III*) due to the ensemble classifier (see Figs. 5a and 5b). The two classification maps achieved by the regular SVM and ensemble appear similar in many regions (Figs. 6 and 7). Nevertheless, some differences exists and the map achieved by the ensemble appears more homogenous. Classification accuracies achieved by the SVM ensemble were significantly higher than those produced by the regular SVM classifier with the respective number of training samples based on a test with a 99% confidence interval $(\alpha = 0.01)$ (Table I).

C. Results for Data From Pavia, Italy

As for the Hekla data set, the results achieved with the SVM ensemble show higher overall accuracies than those for the regular SVM and RF for three training sample sets (Table II). Again, SVM ensembles with ten iterations yield higher accuracies than the regular SVM. Accuracies for 25 and 50 iterations are even higher but do not show relevant differences.

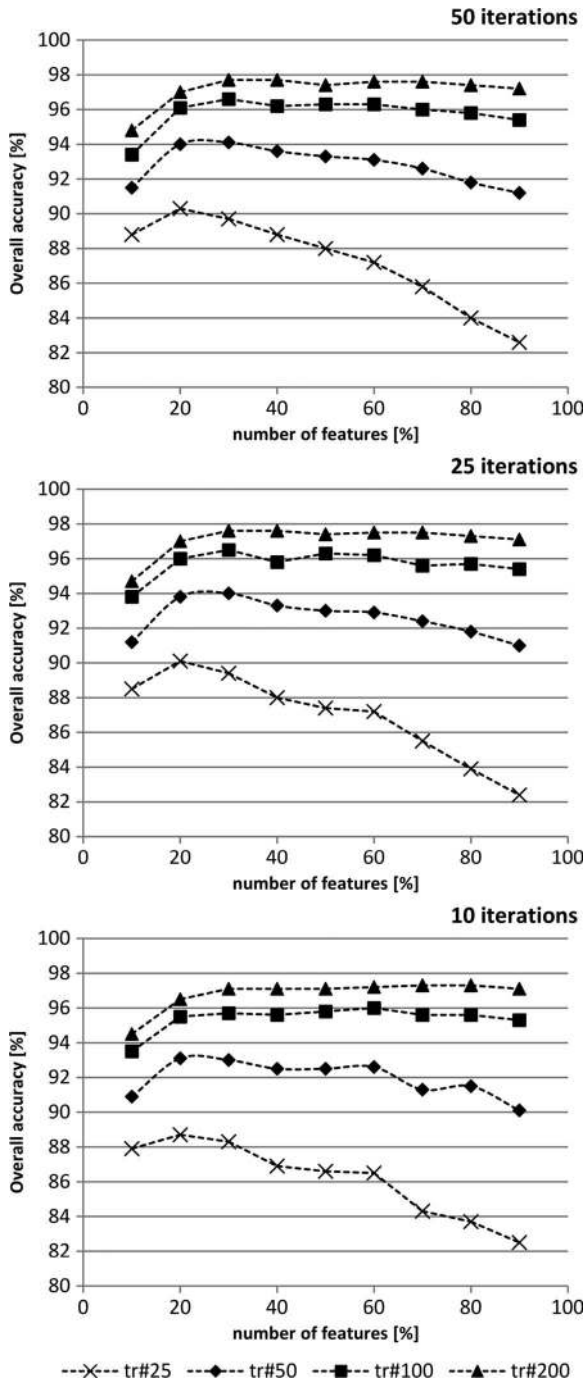

Fig. 4. Hekla data. Overall accuracy (in percent) achieved by the SVM ensemble using different number of iterations, input features, and training samples.

The experimental results again show that SVM ensembles perform well with small training sample sets: Accuracies achieved with the regular SVM and tr#25 are 4.5% and 4.9% below accuracies obtained with the ensemble approach in 25 or 50 iterations. For tr#200, this difference is only 2.3%.

The investigation of the impact of the number of features on the overall accuracy clearly showed that the use of only 10% of the features was ineffective and resulted in lower overall accuracies than a regular SVM classifier does. The adequate number of features is 30%, using the training sets tr#100 and tr#200, and an increased number of features do not improve the overall

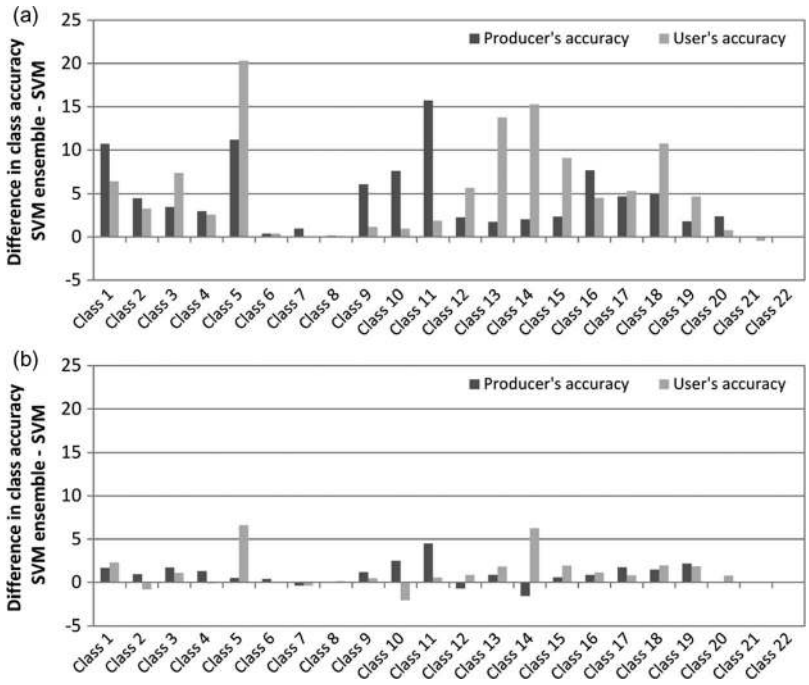

Fig. 5. Hekla data. Difference between producer's and user's accuracy achieved by SVM ensemble with 25 iterations and a regular SVM classifier using (a) tr#50 and (b) tr#200.

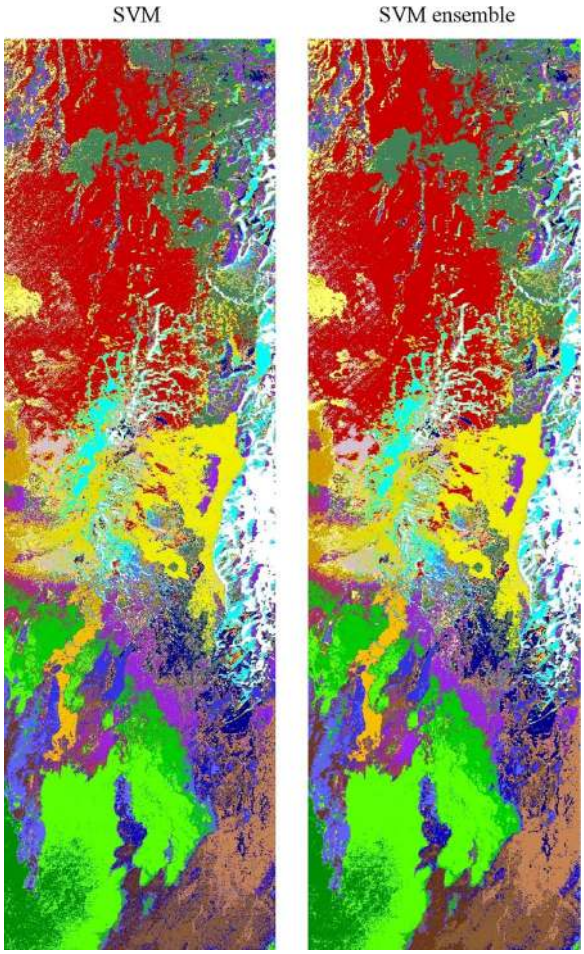

Fig. 6. Hekla data. Classification maps achieved by SVM and SVM ensemble (number of features/iterations: 30%/25) with training set tr#50.

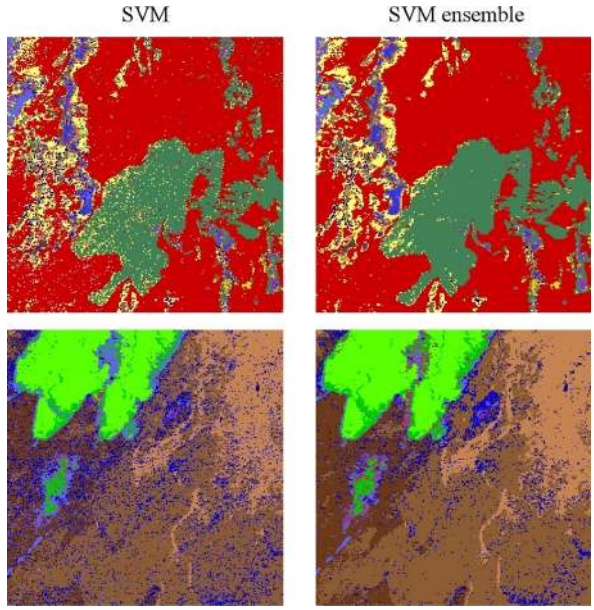

Fig. 7. Hekla data. Detailed visual comparison of classification maps achieved by SVM and SVM ensemble (number of features/iterations: 30%/25) with training set tr#50.

accuracy. In contrast to all other results, the SVM ensemble with tr#50 shows a steady increase in overall accuracy with an increasing number of randomly selected features. The most accurate result was attained with 80% of the whole feature set.

Irrespective of the number of classifiers that is used in the classifier system, an ensemble that is based on a smaller quantity of the available features (i.e., 10%) is relatively ineffective in terms of accuracy. In some cases, it is even less accurate than a regular SVM. On the other hand, the use of 25 iterations outperformed the regular SVM classifier in terms of accuracy, irrespective of the training sample set size (Fig. 8).

TABLE II
DATA SET 2—PAVIA. OVERALL ACCURACY (IN PERCENT) USING DIFFERENT METHODS AND NUMBER OF TRAINING SAMPLES. *INDICATES A SIGNIFICANT DIFFERENCE ($\alpha = 0.01$) IN COMPARISON TO THE REGULAR SVM

Method	Training sample sets			
	#25	#50	#100	#200
SVM	61.9	72.5	77.8	78.9
RF	61.9	64.6	66	68.1
SVM ensemble 50 iterations	66.8*	76.8*	80.2*	81.2*
SVM ensemble 25 iterations	66.4*	76.8*	80.4*	81.6*
SVM ensemble 10 iterations	62.8	75.8*	79.0*	80.5*

Regarding the producer's and user's accuracies, the ensemble strategy performs more accurate in most cases (Fig. 9). The differences between the SVM ensemble and the regular SVM tend clearly toward being positive, i.e., the accuracy achieved by the ensemble is higher.

The producer's accuracies for *Asphalt*, *Bare soil*, and *Bricks* are increased by more than 5%, as do the user's accuracies for *Bitumen*, *Meadow*, and *Trees*. For the Hekla data set, this effect is less obvious, when a higher number of training samples are available. In this case, the users's accuracy of *Bitumen* is even slightly reduced by the ensemble, compared to the accuracy achieved by a standard SVM.

The findings of the accuracy assessment are confirmed by a significance test using a confidence interval of 99% ($\alpha = 0.01$). The test shows that the results achieved with the ensemble approach significantly outperform the results of a standard SVM classifier in terms of accuracy.

The two classification maps achieved by the regular SVM and ensemble appear very similar in most of the regions (Fig. 10). Some differences exist in the north–west (i.e., bottom of the image), which is classified as mixture between *Meadow* and *Bare soil*. In this region, the ensemble approach assigns more pixels to the *Meadow* class, while the conventional SVM classification clearly tends toward the *Bare soil*.

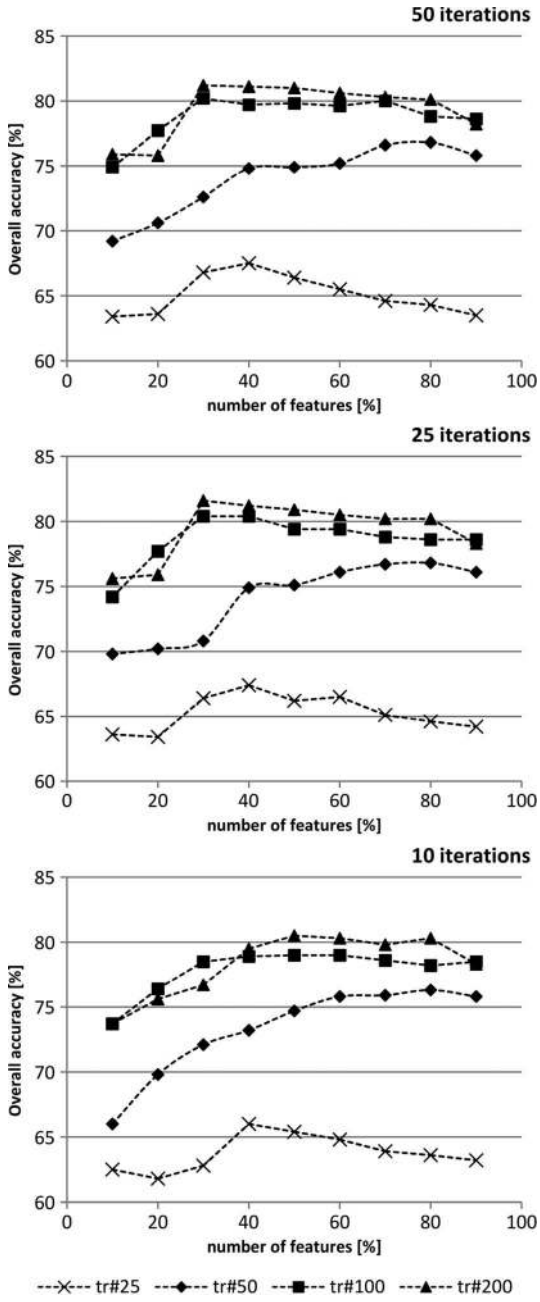

Fig. 8. Pavia data. Overall accuracy in percentage achieved by the SVM ensemble using different number of iterations, input features, and training samples.

V. DISCUSSION

The generation of SVM ensembles by RFS was shown to be generally positive in experiments. Obtained overall accuracies are significantly higher than those achieved by a regular SVM and by an RF, and classification maps appear more homogeneous in visual analysis. This general trend exists independently from specific classes, as shown by the producer's and user's accuracies.

Experiments showed the influence of training sample size on the obtained results. Despite the positive effect of the ensemble strategy for all training sets, the large surplus in accuracy achieved by SVM ensembles for small training data sets needs to be mentioned. One reason for this might be that, in the case of small suboptimal training sample sets, the SVM classifier is affected by the curse of dimensionality, even though SVMs usually perform well in high-dimensional feature space and with small training sets. Another effective method to handle the problem of small training sets is the use of semisupervised SVM [17]. Related to this, the discussion in [40] is interesting. The author argues that classifier combination (i.e., MCS) is similarly useful in a semisupervised context, as unlabeled data (i.e., semisupervised approach) are valuable for ensemble methods.

With respect to the feature subset size, the accuracy assessment demonstrates that a very small number of features (i.e., 10%) are ineffective and generally result in a lower overall accuracy than for the regular SVM. One reason for this could be that each SVM requires some sort of minimum information (i.e., randomly selected features) to be effective in separation. Moreover, each classifier within an ensemble must at least perform slightly better than random guessing in terms of OA. A smaller number of features would result in a very weak

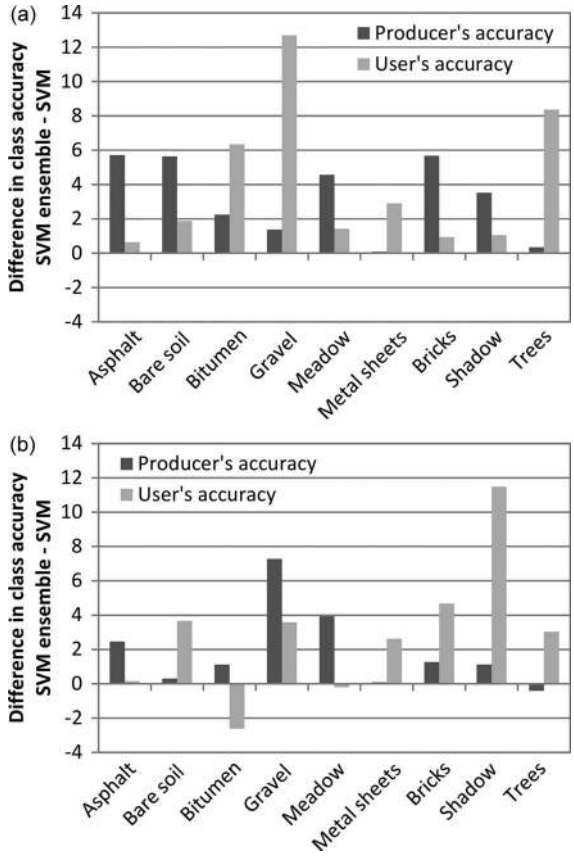

Fig. 9. Pavia data. Difference between producer's and user's accuracy achieved by SVM ensemble with 25 iterations and a regular SVM classifier using (a) tr#50 and (b) tr#200.

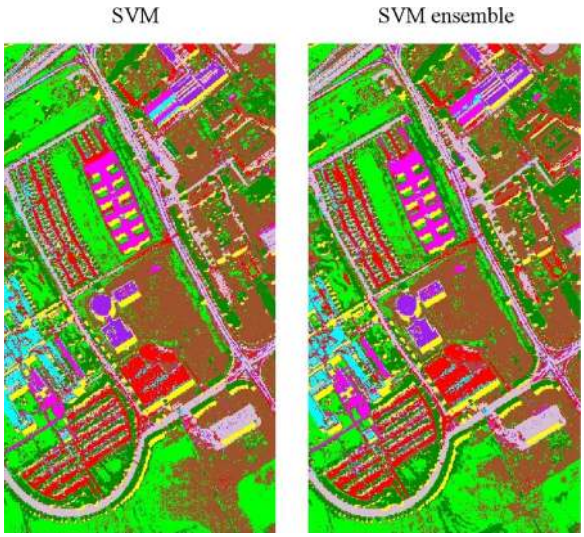

Fig. 10. Data set 2, Pavia. Classification maps achieved by SVM and SVM ensemble (number of features/iterations: 30%/25) using training set tr#100.

performance of each single classifier within the ensemble, and consequently, the overall accuracy cannot be improved. Based on the experimental results, a feature subset size between 20% and 50% is adequate. However, the selection of a small value (i.e., 20%) results in lower accuracies in some cases, whereas the selection of a higher number (i.e., 50%) does not improve the OA, compared to an ensemble, which is based on 30% of the features. Thus, a random selection of 30% of all available features might be a good value in terms of the overall accuracy. This is not in accordance with the recommendations given in [21], which suggest a higher number of features in the case of DT-based classifier systems. Nevertheless, this follows recommendations from Ho [21], who states that selecting 50% of the features can achieve similar classification accuracies, compared to the ensemble, which is based on 30% of the features. On the other hand, the use of a high number of features of the data set results in an increased processing time. Considering the necessary number of iterations, the experimental results show that an ensemble with 25 iterations is already very accurate.

To test transferability of our findings, a third hyperspectral data set (i.e., AVIRIS Indian Pines[2]) was classified using an SVM ensemble with 30% of the features and 25 iterations. The Indian Pines data set was collected by the AVIRIS instrument over an agricultural study site in northwest Indiana and used in several remote sensing studies, e.g., [1], [6], and [41]. The scene consists of 145 $\times$ 145 pixels, and 14 land cover classes were considered in our experiments, ranging from 54 to 2466 pixels in size. Classifications were performed using 25 randomly selected training samples per class. In comparison, the ensemble-based result is 15% more accurate than the classification result achieved with a regular SVM.

Despite the clear increase in classification accuracy by SVM ensembles, the matter of computational complexity appears as a main drawback and needs to be discussed against the background of possible strategies to reduce processing times. On the one hand, the idea of parallel SVM implementation, which was recently discussed in the context of hyperspectral remote sensing [1], [42], will be useful for SVM ensembles. Moreover, the reduced-set approach, which was originally developed for binary SVM [43] and later extended to the multiclass case [44], can be directly adapted, because an ensemble of SVM can be seen as an ensemble of binary SVM. On the other hand, the lower data dimensionality achieved by the suggested strategy for ensemble construction levels out some of the additional complexity, particularly when small feature subsets appear optimal.

VI. Conclusion

In the presented study, an SVM ensemble strategy based on RFS was proposed, tested in experiments, and discussed in the context of classifying hyperspectral data. Three research questions were stated in the beginning regarding the classification accuracy, ensemble size, and computational complexity. Based on our experiments, it can be assessed that the classification accuracy can be significantly increased by the proposed ensemble strategy, compared to classification accuracy achieved by a standard SVM or an RF. Both ensemble parameters, namely, feature subset size and ensemble size, have a significant impact on the accuracy and the stability of SVM ensembles in terms of the classification results. Given this sensitivity of the ensembles and the high computational complexity of an MCS strategy with an SVM as base classifier, parameter selection appears critical. However, our experimental results allow some guidelines with regard to reliable ranges for the two parameters, i.e., according to the results, about 20%–30% of the available feature should be

[2]Data are available online from http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/.

used to construct the classifier system, and an ensembles size of 25 appears sufficient. These recommendations proved effective in an additional experiment on an independent data set. Due to the fact that even values outside these ranges yield results superior to those from regular SVM and the relatively small values for ensemble size and feature subset size, the use of SVM ensembles appears worthwhile, and efficient implementation strategies should be investigated. Particularly for small training sample sets, the presented SVM ensemble strategy by RFS constitutes a feasible approach and useful modification of the regular SVM.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments that helped improve this paper.

REFERENCES

[1] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, Sep. 2009, Supplement 1.

[2] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. 5–16, Sep. 2009, Supplement 1.

[3] M. Herold, M. E Gardner, and D. A. Roberts, "Spectral resolution requirements for mapping urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1907–1919, Sep. 2003.

[4] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[5] C. Huang, L. S. Davis, and J. R. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.

[6] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[7] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification of support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.

[8] J. Ham, Y. Chen, and M. M. Crawford, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[9] J. C.-W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.

[10] S. van der Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert, "Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines," *J. Appl. Remote Sens.*, vol. 1, p. 013 543, 2007.

[11] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[12] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[13] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[14] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[15] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramo, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.

[16] N. Ghoggali, F. Melgani, and Y. Bazi, "A multiobjective genetic SVM approach for classification problems with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1707–1718, Jun. 2009.

[17] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[18] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 2006.

[19] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[20] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learn.*, L. Saitta, Ed., Bari, Italy, Jul. 3–6, 1996.

[21] 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.

[22] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: From basics to recent developments," in *Proc. 7th Int. Workshop MCS*, vol. 4472, *Lecture Notes in Computer Science*, M. Haindl, J. Kittler, and F. Roli, Eds., Prague, Czech Republic, 2007.

[23] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2291–2299, Oct. 2002.

[24] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, Mar. 2006.

[25] B. Waske and S. van der Linden, "Classifying multilevel imagery from SAR and optical sensors by decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1457–1466, May 2008.

[26] L. Hansen and O. Salomon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.

[27] H. C. Kim, S. Pang, H. M. Je, D. Kim, and S. Yang Bang, "Constructing support vector machine ensemble," *Pattern Recognit.*, vol. 36, no. 12, pp. 2757–2767, Dec. 2003.

[28] M. Zortea, M. De Martino, and S. Serpico, "A SVM ensemble approach for spectral-contextual classification of optical high spatial resolution imagery," in *Proc. IGARSS*, Boston, MA, 2008, pp. 1489–1492.

[29] S.-J. Wang, A. Mathew, Y. Chen, L.-F. Xi, L. Ma, and J. Lee, "Empirical analysis of support vector machine ensemble classifiers," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6466–6476, Apr. 2009.

[30] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1999.

[31] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[32] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets," *Pattern Recognit.*, vol. 36, no. 6, pp. 1291–1302, Jun. 2003.

[33] B. Waske, S. Schiefer, and M. Braun, "Random feature selection for decision tree classification of multi-temporal SAR data," in *Proc. IGARSS*, Denver, CO, 2006, pp. 168–171.

[34] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[35] I. Buciu, C. Kotropoulos, and I. Pitas, "Demonstrating the stability of support vector machines for classification," *Signal Process.*, vol. 86, no. 9, pp. 2364–2380, Sep. 2006.

[36] 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.

[37] A. Janz, S. Schiefer, B. Waske, and P. Hostert, "imageSVM—A user-oriented tool for advanced classification of hyperspectral data using support vector machines," in *Proc. 5th Workshop EARSeL Special Interest Group Imaging Spectrosc.*, Bruges, Belgium, 2007.

[38] C.-C. Chang and C.-J. Lin, LIBSVM: A Library for Support Vector Machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[39] I. H. Witten and F. Eibe, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005.

[40] Z-H. Zhou, "When semi-supervised learning meets ensemble learning," in *Proc. 8th Int. Workshop, MCS*, vol. 4472, *Lecture Notes in Computer Science*, J. A. Benediktsson, J. Kittler, and F. Roli, Eds., Reykjavik, Iceland, 2009, pp. 529–538.

[41] J. Chen, C. Wang, and R. Wang, "Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2193–2205, Jul. 2009.

[42] J. Muñoz, A. Plaza, J. A. Gualtieri, and G. Camps-Valls, "Parallel implementation of SVM in Earth observation applications," in *Parallel Programming, Models and Applications in Grid and P2P Systems*, F. Xhafa, Ed. Amsterdam, The Netherlands: IOS Press, 2009, pp. 292–312.
[43] C. J. Burges, "Simplified support vector decision rules," in *Proc. 13th Int. Conf. Mach. Learn.*, L. Saitta, Ed., Bari, Italy, 1996, pp. 71–77.
[44] B. Tang and D. Mazzoni, "Multiclass reduced-set support vector machines," in *Proc. 23rd Int. Conf. Mach. Learn.*, vol. 148, *ACM International Conference Proceeding Series*, W. W. Cohen and A. Moore, Eds., Pittsburgh, PA, Jun. 25–29, 2006, pp. 921–928.

Björn Waske (M'08) received the degree in applied environmental sciences, with a major in remote sensing, from Trier University, Trier, Germany, in 2002 and the Ph.D. degree in geography from the Center for Remote Sensing of Land Surfaces, University of Bonn, Bonn, Germany, in 2007.

From 2002 to mid-2004, he was a Research Assistant with the Department of Geosciences, Munich University, Munich, Germany, doing research on the use of remote sensing data for flood forecast modeling. In 2006, he visited the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík, Iceland, for three months, where he was a Postdoctoral Researcher from beginning of 2008 until August 2009 and has been a (Junior) Professor for remote sensing in agriculture since September 2009. His current research activities concentrate on advanced concepts for image classification and data fusion, with a focus on agricultural applications.

Dr. Waske is a Reviewer for different international journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Sebastian van der Linden studied at the University of Edinburgh, Edinburgh, U.K., from 1999 to 2000. He received the M.Sc. degree in applied environmental sciences from the University of Trier, Trier, Germany, in 2002 and the Ph.D. degree from Humboldt-Universität zu Berlin, Berlin, Germany, in 2008.

In 2006, he spent three months as a Visiting Scholar with the Lamont-Doherty Earth Observatory, Columbia University, New York. He is currently a Postdoctoral Researcher with the Geomatics Laboratory, Humboldt-Universität zu Berlin. His scientific interests include the development, implementation, and application of machine-learning-based approaches for qualitative and quantitative analysis of imaging spectroscopy data. His current research is closely linked to the German hyperspectral satellite mission EnMAP, and he coordinates the development of the EnMAP Toolbox. He is the author and coauthor of more than 15 journal publications and book chapters ranging from methodological to applied remote sensing.

Dr. van der Linden is a Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1987 and 1990, respectively.

He is currently a Pro Rector for academic affairs and a Professor of electrical and computer engineering with the University of Iceland. He has held visiting positions with the Department of Information and Communication Technology, University of Trento, Trento, Italy, since 2002; the School of Computing and Information Systems, Kingston University, Kingston upon Thames, U.K., during 1999–2004; the Joint Research Centre of the European Commission, Ispra, Italy, in 1998; the Technical University of Denmark, Lyngby, Denmark, in 1998; and the School of Electrical and Computer Engineering, Purdue University, in 1995. In August 1997, he was a Fellow with the Australian Defence Force Academy, Canberra, A.C.T., Australia. From 1999 to 2004, he was the Chairman of the energy company Metan Ltd. He was also the Chairman of the University of Iceland's Science and Research Committee from 1999 to 2005. Since 2006, he has been the Chairman of the University of Iceland's Quality Assurance Committee. His research interests include remote sensing, pattern recognition, machine learning, image processing, biomedical engineering, and signal processing, and he has published extensively in those fields.

Dr. Benediktsson was the Past Editor and is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGARS) and an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He coedited (with Prof. D. A. Landgrebe) a Special Issue on Data Fusion of TGARS (May 1999) and (with P. Gamba and G. G. Wilkinson) a Special Issue on Urban Remote Sensing from Satellite (October 2003). From 1999 to 2002, prior to being an Editor, he was an Associate Editor for TGARS. From 1996 to 1999, he was the Chairman of GRSS Technical Committee on Data Fusion and was elected to the Administrative Committee of the GRSS for the terms 2000–2002 and 2008–2010. He is currently, the Executive Vice President of GRSS. In 2002, he was appointed as Vice President of Technical Activities of GRSS, and in 2008, he was appointed as Vice President of Professional Activities of GRSS. He was the founding Chairman of the IEEE Iceland Section and served as its Chairman from 2000 to 2003. He was a member of a NATO Advisory Panel of the Physical and Engineering Science and Technology Subprogram (2002–2003). He was a member of Iceland's Science and Technology Council (2003–2006) and a member of the Nordic Research Policy Council (2004). He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award. In 2000, he was granted the IEEE Third Millennium Medal. In 2004, he was a corecipient of the University of Iceland's Technology Innovation Award. In 2006, he received the yearly research award from the Engineering Research Institute, University of Iceland. In 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He is a member of Societas Scinetiarum Islandica and Tau Beta Pi.

Andreas Rabe received the M.Sc. degree in computer science from Humboldt-Universität zu Berlin, Berlin, Germany, in 2007.

He is a Researcher with the Geomatics Laboratory, Humboldt-Universität zu Berlin. Currently, he is incorporated in the development of the EnMAP Toolbox, which is part of the German hyperspectral satellite mission EnMAP. His scientific interests include image and signal processing, with a focus on machine learning applications.

Patrick Hostert received the B.Sc. degree in physical geography from Trier University, Trier, Germany, in 1994, the M.Sc. degree in geographical information science (GIS) from Edinburgh University, Edinburgh, U.K., in 1995, and the Ph.D. degree from Trier University, in 2001.

He was an Assistant Professor with Humboldt-Universität zu Berlin, Berlin, Germany, from 2002 to 2006, where he has been a Full Professor in geomatics, with a focus on remote sensing, since 2006. He is the Founder and Head of the Geomatics Laboratory, Humboldt-Universität zu Berlin, where he has been the Director of the Geography Department since 2008. He is a member of the scientific advisory group of the Environmental Monitoring and Analysis Program—Germany's hyperspectral EnMAP mission. His research in remote sensing focuses on hyperspectral image analysis and analyzing remote sensing data with machine learning algorithms, specifically support vector machine classifiers and regression.