

A Multiple SVM System for Classification of Hyperspectral Remote Sensing Data

Behnaz Bigdeli · Farhad Samadzadegan · Peter Reinartz

Received: 9 February 2013 / Accepted: 22 April 2013
© Indian Society of Remote Sensing 2013

Abstract With recent technological advances in remote sensing sensors and systems, very high-dimensional hyperspectral data are available for a better discrimination among different complex land-cover classes. However, the large number of spectral bands, but limited availability of training samples creates the problem of Hughes phenomenon or ‘curse of dimensionality’ in hyperspectral data sets. Moreover, these high numbers of bands are usually highly correlated. Because of these complexities of hyperspectral data, traditional classification strategies have often limited performance in classification of hyperspectral imagery. Referring to the limitation of single classifier in these situations, Multiple Classifier Systems (MCS) may have better performance than single classifier. This paper presents a new method for classification of hyperspectral data based on a band clustering strategy through a multiple Support Vector

Machine system. The proposed method uses the band grouping process based on a modified mutual information strategy to split data into few band groups. After the band grouping step, the proposed algorithm aims at benefiting from the capabilities of SVM as classification method. So, the proposed approach applies SVM on each band group that is produced in a previous step. Finally, Naive Bayes (NB) as a classifier fusion method combines decisions of SVM classifiers. Experimental results on two common hyperspectral data sets show that the proposed method improves the classification accuracy in comparison with the standard SVM on entire bands of data and feature selection methods.

Keywords Hyperspectral · Support Vector Machine · Multiple Classifier System · Bayesian Theory

B. Bigdeli (✉) · F. Samadzadegan
Department of Geomatics Engineering, Faculty of Engineering, University of Tehran, North Kargar Street, Tehran, Iran
e-mail: bigdeli@ut.ac.ir

F. Samadzadegan
e-mail: samadz@ut.ac.ir

P. Reinartz
Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, German Aerospace Centre (DLR), Oberpfaffenhofen, P.O Box 1116, 82230 Weßling, Germany
e-mail: peter.reinartz@dlr.de

Introduction

With the development of the remote-sensing imaging technology and hyperspectral sensors, classification of hyperspectral image is becoming more and more widespread in different applications (Jia 2002; Goel et al. 2003; Li et al. 2011). These data cover in most cases a wide spectral range from the visible to the short-wave infrared with a narrow band width for each single channel, resulting in hundreds of data channels. Thanks to this amount of information, it is feasible to deal with applications that require a precise discrimination in the spectral domain. In this context, hyperspectral images

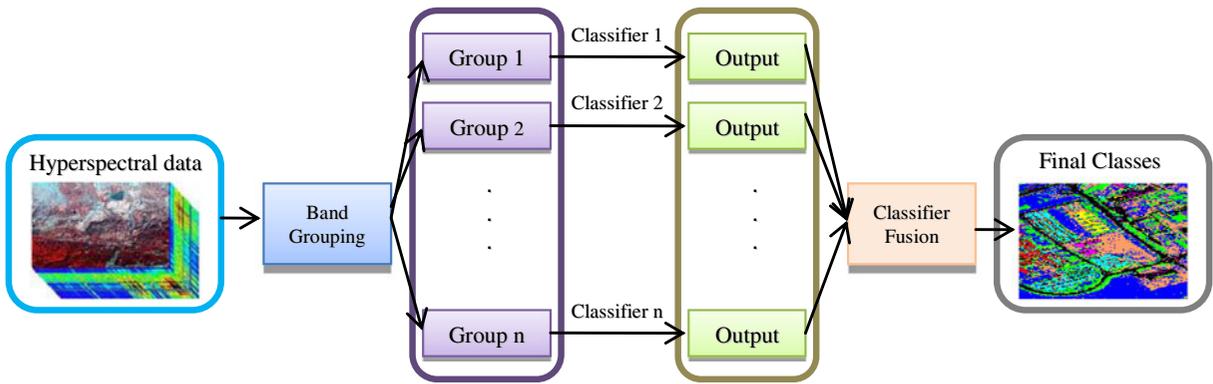


Fig. 1 A multiple SVM system based on band grouping of hyperspectral images

have been successfully used for supervised classification problems that require very precise description in spectral feature space.

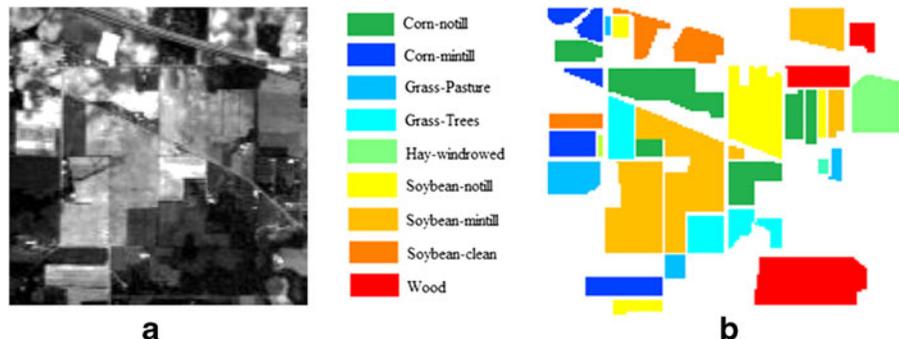
An extensive literature is available on the classification of hyperspectral images. Maximum likelihood or Bayesian estimation methods (Jia 2002), decision trees (Goel et al. 2003), neural networks (Del frate et al. 2007), genetic algorithms (Vaiphasa 2003), and kernel-based techniques (Müller et al. 2001; Camps-Valls and Bruzzone 2005) have been widely investigated in this direction. One of the most popular classification methods is Support Vector Machines (SVM) defined by Vapnik, a large margin based classifier with a good generalization capacity in the small-size training set problem with high-dimensional input space (Vapnik 1998). Recently, SVMs have been successfully applied in the classification of hyperspectral remote-sensing data. Camps-Valls and Bruzzone (2005) demonstrated that SVMs perform equal or better than other classifiers in terms of accuracy on hyperspectral data.

At the same time, hyperspectral images are usually composed of tens or hundreds of close spectral bands, which result in high redundancy and great amount of

computation time for image classification. Large number of features can become a curse in terms of accuracy if enough training samples are not available, i.e. due to the Hughes phenomenon in most of traditional classification techniques (Li et al. 2011). Hughes phenomenon means that when the training sample number is a constant, the precision of classification will be decreased with the increasing of the dimensionality. It implies that the required number of training samples for supervised classification increases as a function of dimensionality.

Conventional classification strategies often cannot overcome mentioned problem. Alternatives like Multiple Classifier Systems (MCS) are successfully applied on various types of data to improve single classifiers results. Multiple Classifier System (MCS) can improve classification accuracy in comparison to a single classifier by combining different classification algorithms or variants of the same classifier (Kuncheva 2004). In such systems a set of classifiers is first produced and then combined by a specific fusion method. The resulting classifier is generally more accurate than any of the individual classifiers that make up the ensemble.

Fig. 2 AVIRIS Indian Pine Data, a) Original data and b) Ground truth



Multiple classifier systems can be used to improve classification accuracy in remote sensing data sets (Benediktsson and Kanellopoulos 1999).

The first step of this paper lies in the problem formulation of the extraction of band groups from hyperspectral data to produce a multiple classifier system. This method tries to split the entire high dimensional hyperspectral space into few band groups for classification while it can overcome Hughes phenomenon or curse of dimensionality. The proposed approach decomposes high number of spectral bands into few uncorrelated groups. After that SVMs are applied on each group which was produced in the previous step. After producing an ensemble of classifiers, a classifier fusion method based on the Bayesian theory is applied in the multiple classifier system to fuse the outputs of SVM classifiers.

Band Grouping of Hyperspectral Imagery

Basic principle of the band grouping of hyperspectral imageries is that the adjacent bands which have high correlation are be grouped into one group and the ones with little redundancy should be separated into different groups. Band grouping of hyperspectral imageries as primary step of feature selection is investigated in a wide range of investigations.

Feature selection (Band Selection) algorithms suitably select a (sub)optimal subset of the original set of features while discarding the remaining features to the classification problem of hyperspectral images. Feature selection techniques generally involve both a search algorithm and a criterion function. The search

Table 1 AVIRIS Indian Pine lands cover classes and available reference samples

Class	Land Cover Class	Samples
1	Corn-no till	1434
2	Corn-minimum till	834
3	Grass/pasture	497
4	Grass/trees	747
5	Hay-windrowed	489
6	Soybeans-no till	968
7	Soybeans-minimum till	2468
8	Soybeans-clean till	614
9	Woods	1294

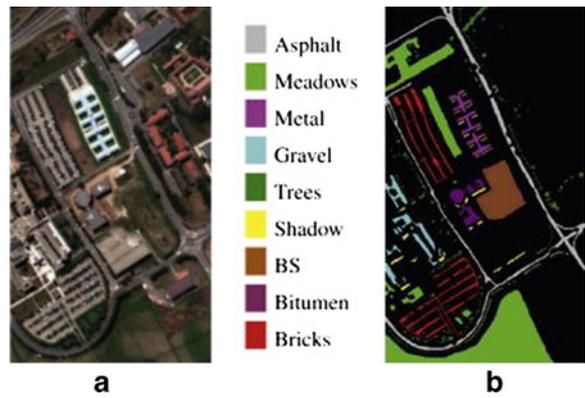


Fig. 3 ROSIS Pavia University Data, a) Original data and b) Ground truth

algorithm generates possible “solutions” of the feature selection problem (i.e., subsets of features) and compares them by applying the criterion function as a measure of the effectiveness of each solution.

Benediktsson and Kanellopoulos (1999) proposed absolute correlation as a measure of the spectral bands similarity. After computing correlation matrix between bands, they applied a manual clustering to split hyperspectral image. Prasad and Bruce (2008) proposed a divide and conquer approach that partitions the hyperspectral space into contiguous subspaces using the reflectance information. In another article (Martinez-Usó et al. 2006) they used a clustering method in relationship to hyperspectral, multi-temporal classification. After partitioning hyperspectral data using reflectance information, they used Linear Discriminant Analysis (LDA) on each subspace that ensures good class separation in their clustering method.

Table 2 ROSIS Pavia University lands cover classes and available reference samples

Class	Land Cover Class	Samples
1	Trees	524
2	Asphalt	548
3	Bitumen	375
4	Gravel	392
5	Painted metal sheets	265
6	Shadows	231
7	Self-Blocking Bricks	514
8	Meadows	540
9	Bare Soil	532

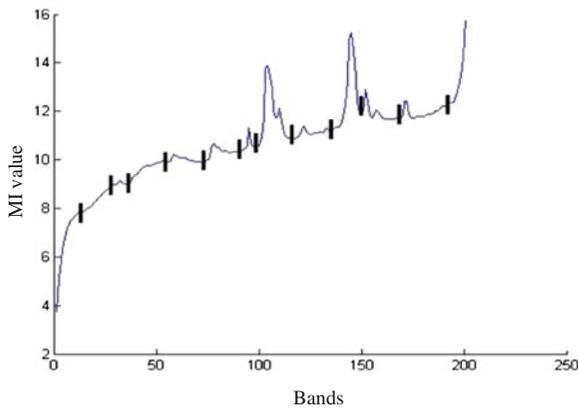


Fig. 4 Mutual Information results of AVIRIS data

The resulting system is capable of performing reliable classification even when relatively few training samples are available for a given date.

Martinez-Use et al. (2006) applied a grouping-based band selection using information measures on hyperspectral data. They used band grouping as a primary step for a band selection technique. Guo et al. (2006) found that the grouping based on the simple criterion of only retaining features with high associated mutual information (MI) values is problematic when the bands were highly correlated. It is also presented that mutual information by itself would not be suitable as a similarity measure. The reason is that it can be low because either the two bands present a weak relation (such as it should be desirable) or the entropies of these variables are small (in such a case, the variables contribute with little information).

Thus, it is convenient to define a strategy that modifies the results of band grouping using mutual information. This paper applies a hybrid band grouping strategy based on genetic algorithm and support vector machine on the primary results of mutual information. Furthermore Li et al. (2011) applied this method for feature selection on hyperspectral data. Experimental results on two reference data sets have shown that this approach is very competitive and effective.

Classification of Hyperspectral Imagery Using Support Vector Machine (SVM)

SVMs separate two classes by fitting an optimal linear separating hyper plane to the training samples of the two classes in a multidimensional feature space. The optimization problem being solved is based on structural risk minimization and aims to maximize the margins between the optimal separating hyper plane and the closest training samples also called support vectors (Weston and Watkins 1999; Scholkopf and Smola 2002). Let, for a binary classification problem in a d -dimensional feature space x_i be a training data set of L samples with their corresponding class labels $y_i \in \{1, -1\}$. The hyper plane $f(x)$ is defined by the normal vector w and the bias b where $|b|/\|w\|$ is the distance between the hyper plane and the origin,

$$f(x) = w \cdot x + b \tag{1}$$

For linearly not separable cases, the input data are mapped into a high-dimensional space in which the new distribution of the samples enables the fitting of a linear hyper plane. The computationally extensive mapping in a high dimensional space is reduced by using a positive definite kernel k , which meets Mercers conditions (Scholkopf and Smola 2002).

$$\phi(x_i)\phi(x_j) = k(x_i, x_j) \tag{2}$$

where ϕ is mapping function. The final hyper plane decision function can be defined as:

$$f(x) = \sum_{i=1}^L \alpha_i y_i k(x_i, x_j) + \gamma \tag{3}$$

where α_i are Lagrange multipliers.

Recently, SVMs have attracted increasing attention in remote-sensed hyperspectral data classification tasks and an extensive literature is available. Melgani and Bruzzone (2004) applied SVM for classification of hyperspectral data. They obtained better classification results compared to other common classification algorithms. In Watanachaturaporn and Arora (2004)

Table 3 Final band grouping results on AVIRIS data

Groups	1	2	3	4	5	6	7	8	9	10	11	12
Bands	1–18	19–33	34–44	45–57	58–77	78–105	106–125	126–131	132–147	148–157	158–170	171–202

study the aim is to investigate the effect of some factors on the accuracy of SVM classification. The factors considered are selection of multiclass method, choice of the optimizer and the type of kernel function. Tarabalka et al. (2010) present a novel method for accurate spectral-spatial classification of hyperspectral images using support vector machines. Their proposed method, improved classification accuracies in comparison to other classification approaches.

Multiple Classifier System (MCS)

Combining classifiers to achieve higher accuracy is an important research topic with different names such as combination of multiple classifiers, Multiple Classifier System (MCS), classifier ensembles and classifier fusion. In such systems a set of classifiers is first produced and then combined by a specific fusion method. The resulting classifier is generally more accurate than any of the individual classifiers that make up the ensemble (Kuncheva 2004; Kuncheva and Whitaker 2003).

The possible ways of combining the outputs of the L classifiers in a MCS depend on what information can be obtained from the individual members. Kuncheva (2004) distinguishes between two types of classifier outputs which can be used in classifier combination methods. The first types are classifiers that produce crisp outputs. In this category each classifier only outputs a unique class and finally a vector of classes is produced for each sample. The second type of classifier produces fuzzy output which means that in this case the classifier associates a confidence measurement for each class and finally produces a vector for every classifier and a matrix for ensemble of classifier.

The key to the success of classifier fusion is that, intuitively at least, a multiple classifier system should build diverse and partially uncorrelated classifiers. Diversity among classifiers is the notion describing the level to which classifiers vary in data representation, concepts, strategy etc. Consequently, this should be reflected in different classifiers making errors for different data samples. As shown in many papers, such phenomenon of disagreement to errors is highly beneficial for combining purposes (Kuncheva and Whitaker 2003). Most of the diversity measures have already been studied for artificial data by Kuncheva (2004) and real world data sets by Ruta and Gabrys (2000). In the simplest case, a measure can be applied for examining

Table 4 Overall Accuracy of SVM classifiers on bands groups of AVIRIS Data for different training size

Training size	Band Group#1		Band Group#2		Band Group#3		Band Group#4		Band Group#5		Band Group#6		Band Group#7		Band Group#8		Band Group#9		Band Group#10		Band Group#11		Band Group#12		
	5	10	20	5	10	20	5	10	20	5	10	20	5	10	20	5	10	20	5	10	20	5	10	20	
%5	53.778	57.778	57.333	62.222	64	64.667	61.556	49.778	51.333	59.778	61.556	66	52.889	46.889	50.222	49.778	51.111	50.222	49.778	51.111	50.889	51.111	51.111	51.111	63.111
%10	57.778	57.778	57.333	64	64.444	64.667	64.444	51.333	57.556	60.444	60.444	66	53.778	46.667	50	50.889	47.556	50	50.889	47.556	50.889	49.333	49.333	49.333	56.444
%20	57.333	57.333	57.333	64.667	65.333	65.333	53.556	53.556	62	62	66	66	54	48.889	50.222	51.111	48.889	50.222	51.111	49.333	49.333	49.333	49.333	49.333	53.111

diversity between exactly two classifiers. Such measures are usually referred to as pair-wise diversity measures (PDM). For more than two classifiers, PDM is typically obtained by averaging the PDM's calculated for all pairs of classifiers from the considered pool of classifiers. Disagreement (*Diss*) and Double Fault (*DF*) are two important diversity measures. Disagreement takes the form of a ratio between the numbers of samples for which the classifiers disagreed, to the total number of observations. This can be written as:

$$Diss = \frac{N^{FT} + N^{TF}}{N} \tag{4}$$

Where N^{FF} represents the number of elements which both classifiers classified incorrectly, N^{TF} is the number of elements which the 1st classifier classified correctly and the 2nd classifier classified incorrectly, and N^{FT} stands for the 2nd classifier classified correctly and the 1st classifier classified incorrectly. Second, the ‘‘Double Fault’’ estimates the probability of coincident errors for a pair of classifiers, which is

$$DF = \frac{N^{FF}}{N} \tag{5}$$

These two measures vary between [0–1]. If the disagreement measure is greater the diversity is greater however if the double fault measure is lower the diversity is greater. (This means that the relationship between disagreement measure and diversity of classifiers is straight but the relationship between DF and diversity is reverse).

The performance of a multiple classifier system essentially depends on another major factor related to the classifier’s pool: correlation. The correlation between the classifiers to be fused needs to be small to allow performance improvement in classifier fusion. Goebel et al. (2002) introduced a simple computational method for evaluating the correlation between two classifiers which can be extended for more than two classifiers. Kuncheva

and Whitaker (2003) investigated the effects of independence between individual classifiers on classifier fusion. They showed that independent classifiers offer a dramatic improvement over the accuracy of fusion correlated results. For each classifier, a confusion matrix *M* can be generated using the labelled training data. The confusion matrix lists the true classes versus the estimated classes. Goebel et al. (2002), describe a classifier correlation analysis for two classifiers.

$$\rho = \frac{2 \times N^{FF}}{N^{TF} + N^{FT} + 2 \times N^{FF}} \tag{6}$$

Goebel et al., proposed an extension of the 2 class correlation coefficient to *n* different classifiers as follows:

$$\rho_n = \frac{nN^f}{N - N^f - N^t + nN^f} \tag{7}$$

where *n* is the number of classifiers, *N* represents the number of samples, N^f is the number of samples for which all classifiers had a right answer and N^t is the number of samples for which all classifiers had a wrong answer.

In recent years, more studies applied a classifier fusion concept to improve classification results on remotely sensed data (Waske and Van der Linden 2008; Ceamanos et al. 2010). Producing a multiple classifier system with low correlation and high diversity between single classifiers can be useful for improving the classification accuracy in particular for multisource data sets and hyper dimensional imagery.

SVM-based MCS for Classification of Hyperspectral Image

A multiple SVM system based on the band grouping for classification of hyperspectral images is introduced in this paper. Figure 1 shows the general structure of the proposed methodology. The proposed method

Table 5 Comparison of classification strategies on AVIRIS data set

Training Size	Fusion methods			Full data	Feature Selection Methods	
	WMV	SVM	NB		SFSS	SFBS
%5	84.222	85.111	86	82	83.10	84
%10	90.444	91.556	92.889	89.778	90.24	90.40
%20	93.111	93.556	94	91.222	91.8	93

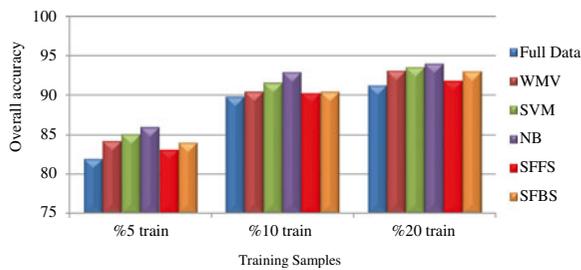


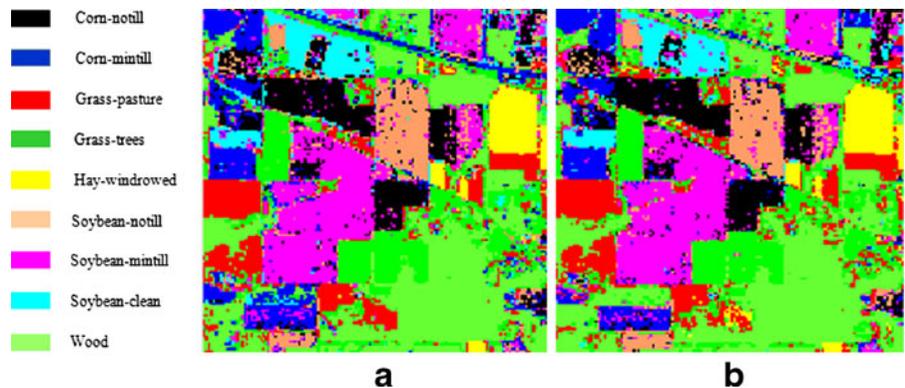
Fig. 5 Comparison between overall accuracies of different classification strategies on AVIRIS data

starts by splitting the hyperspectral image into few band groups based on the modified mutual information strategy.

First, the adjacent bands which exhibit a high mutual information measure are grouped into one group through a computation of the similarity measure of the spectral information. The major benefit of the proposed method is related to this step. All researches in feature selection techniques tried to select just useful bands while the proposed method tries to prevent losing information in feature selection by a system that enables the use of the entire high dimensional hyperspectral in few band clusters.

Second, the proposed methodology applies a SVM classifier for classification of each band group which is produced in the previous step. While conventional methods use SVM for the whole hyperspectral data by definition of one single kernel function which may not be adapted to the whole diversity of information, the proposed method uses one SVM for each band group. In fact, the kernel of each individual classifier applied on each band group is adjusted according to the corresponding data. Finally, generated classification results fused to improve classification accuracy. In order

Fig. 6 Comparison between classification maps of AVIRIS data for, **a**) Standard SVM **b**) Naive Bayes fusion method



to show the performance of the proposed method, results compared with two common classification strategies on hyperspectral data: feature selection method and standard SVM on entire bands.

Band Grouping Based on Mutual Information

As stated in the previous section, Mutual Information is applied to split hyperspectral data into few band groups. The entropy is a measure of uncertainty of random variables, and is a frequently-used evaluation criterion of feature selection (Martinez-Usó et al. 2006; Guo et al. 2006). If a discrete random variable X has Φ alphabets and the probability density function is $p(x)$, $x \in \Phi$, the entropy of X is defined as:

$$H(X) = - \sum_{x \in \Phi} p(x) \log p(x) \tag{8}$$

In the task of band grouping, the entropy of each band is computed by using all spectral information of this band. For two discrete random variables X and Y , which have Φ and Ψ alphabets and their joint probability density function is $p(x, y)$, $x \in \Phi$, $y \in \Psi$, the joint entropy of X and Y is defined as:

$$H(X, Y) = - \sum_{x \in \Phi} \sum_{y \in \Psi} p(x, y) \log p(x, y) \tag{9}$$

The mutual information is usually used to measure the correlation between two random variables and it is defined as:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) = H(X) - H(X/Y) \tag{10}$$

In Eqs. 9 and 10, X and Y represent pixel value of two adjacent bands.

The basic principle of the band grouping is that the adjacent bands which have high correlation should be grouped into one group and the ones with little redundancy should be separated into different groups. Proposed method used Mutual Information to measure the correlation between adjacent bands.

The redundancy between two bands is greater when the value of MI is larger (Li et al. 2011). During the process of band grouping based on the MI, the basic principle is that the bands are divided into groups according to local minima points of bands' MI. These local minima points can be obtained automatically by comparing the neighbourhoods of every point.

After this initial band grouping, since the MI only considers the correlation between bands, Genetic Algorithm–Support Vector Machine (GA-SVM) searches for the best combination of bands with more similar information. Since there are hundreds of bands in the hyperspectral imagery, the search space for GA directly on the original band space will be too huge. First, mutual information is employed to partition the bands into disjoint subspace, thus getting the irredundant set of bands and reducing the search space at the same time. Second, GA–SVM is adopted to search for the optimal combination of bands (Li et al. 2011).

Classifier Fusion Based on Naive Bayes

Naive Bayes is a statistical classifier fusion method that can be used for fusing the outputs of individual classifiers. The essence of NB is based on the Bayesian theory (Kuncheva 2004). Denote by $p(\cdot)$ the probability. In Eqs. 11, 12 and 13 D_j , ($j=1, \dots, L$) is ensemble of classifiers where $s=[s_1, \dots, s_L]$ denote the output labels vector of the ensemble for unknown sample x . Also, ω_k , ($k=1, \dots, c$) denote the class labels and c is the number of classes.

$$p(S/\omega_k) = p(s_1, s_2, \dots, s_L/\omega_k) = \prod_{i=1}^L p(s_i/\omega_k) \quad (11)$$

Then the posterior probability needed to label x is

$$p(\omega_k/S) = \frac{p(\omega_k)p(S/\omega_k)}{p(S)} = \frac{p(\omega_k) \prod_{i=1}^L p(s_i/\omega_k)}{p(S)}, k = 1, \dots, c \quad (12)$$

The denominator does not depend on ω_k and can be ignored, so the final support for class ω_k is

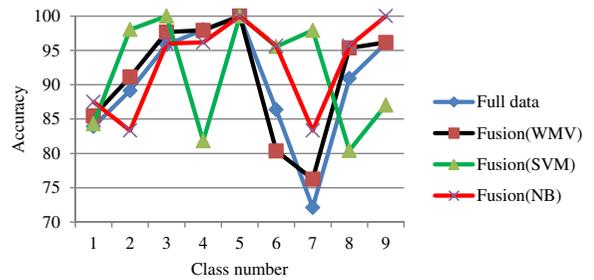


Fig. 7 Comparison of class accuracies between standard SVM and fusion methods for AVIRIS data

$$\mu_k(x) \propto p(\omega_k) \prod_{i=1}^L p(s_i/\omega_k) \quad (13)$$

Where x is the sample of data with unknown class label. The maximum membership rule (μ) will label x in ω_k class (winner class).

The practical implementation of the Naive Bayes (NB) method on a data set with cardinality N is explained below. For each classifier, a $c \times c$ Confusion Matrix CM_i is calculated by testing data set (Kuncheva 2004). The (g, h)th entry of this matrix, $cm_{k,h}^i$ is the number of elements of the data set whose true class label was ω_k and were assigned by the classifier to class ω_h . By N_h we denote the total number of elements of data set from class ω_h . Taking $cm_{k,h}^i/N_k$ as an estimate of the posterior probability, and N_k/N as an estimate of the prior probability, the final support of class ω_k for unknown sample x is

$$\mu_k(x) \propto \frac{1}{N_k^{L-1}} \prod_{i=1}^L cm_{k,h_i}^i \quad (14)$$

The maximum membership rule will label x in ω_k class.

The Bayes classifier has been found to be surprisingly accurate and efficient in many experimental studies. Kuncheva applied NB combination method on artificial data as classifier fusion strategy (Kuncheva 2004). The NB classifiers have been successfully applied in text classification for example: Xu et al. (1992) applied NB as classifier fusion method in applications to handwriting recognition. These researches have indicated the

Table 6 Correlation and diversity measures for AVIRIS data set

Data Set	Correlation	Disagreement (Diss) ↓	Double Fault (DF) ↑
Indian Pine	0.180	0.008	0.991

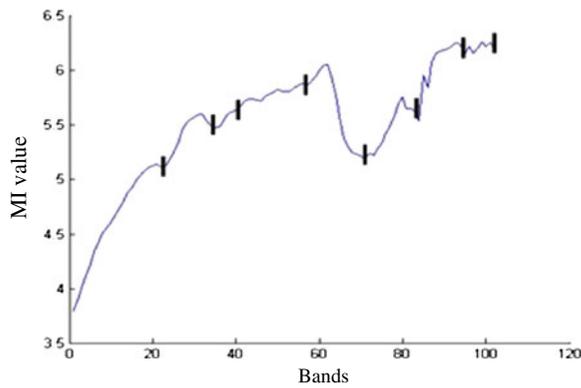


Fig. 8 Mutual Information results on ROSIS data

considerable potential of Naive Bayes approach for the supervised classification of various types of data.

SVM and Fusion Process

As mentioned in SVM-based MCS for Classification of Hyperspectral Image section after band grouping, SVM classifiers are separately applied to each group. It is worth underlining that the kernel-based implementation of SVMs involves the problem of the selection of multiple parameters, including the kernel parameters (e.g., parameters for the Gaussian and polynomial kernels) and the regularization parameter *C*.

In our proposed method, the kernel of each individual classifier is adjusted according to the corresponding band group properties. This paper utilized one-against-one multi class SVM with Radial Basis Function (RBF) kernel (see Eq. 15) as base classifier (Imbault and Lebart 2004).

$$K(x, x') = \exp(-\lambda \|x - x'\|^2) \tag{15}$$

Parameter *C* is the cost of the penalty. The choice of value for parameter *C* influences the classification outcome. If *C* is too large, then the classification accuracy rate is very high in the training stage, but very low in the testing stage. If *C* is too small, then the classification accuracy rate is unsatisfactory, making the model useless. Parameter γ has a much stronger impact than

parameter *C* on classification outcomes, because its value influences the partitioning outcome in the feature space. An excessive value for parameter γ leads to overfitting, while a disproportionately small value results in underfitting. This paper utilized Grid search as a technique to adjust parameters of kernels. The search range for *C* (SVM parameter) is in $[2^{-2}, 2^{10}]$, and $[2^{-10}, 2^2]$ for γ (Kernel Parameter). Grid search is the simplest way to determine the values for parameters *C* and γ . Sets of values for parameters *C* and γ that produce the highest classification accuracy rate in this interval are found by setting the upper and lower limits (search interval) for parameters *C* and γ and the jumping interval in the search. Various pairs of (*C*, γ) values are tried and the one with the best cross-validation accuracy is picked. Methods for obtaining the optimal parameters in the SVM are currently still under development. In this paper we applied this simple method to select the best parameters of SVM classifiers. You can see more details in Hsu et al. (2010). After producing of single classifiers for the MCS, the proposed method applies three fusion strategies. The first one is a fusion strategy based on the Naive Bayes. The second method is Weighted Majority Voting (WMV) which is based on the voting strategies and can be applied to a multiple classifier system assuming that each classifier gives a single class label as an output and is proposed by Kuncheva (2004). In this fusion method, overall accuracies of classifiers are introduced as the weights of the classifiers. In addition, the final fusion strategy used an additional SVM on the outputs of SVM classifiers. Outputs of each primary SVM on each band group used as new feature vector for new classification. All results from fusion strategies are compared to a standard SVM which is applied on full data with all hyperspectral bands.

Experimental Results

Data Sets

The proposed method was tested on two well-known hyperspectral data sets. The first data set is made up of

Table 7 Final band grouping results on ROSIS data

Groups	1	2	3	4	5	6	7	8
Bands	1–22	23–33	34–46	47–57	58–73	74–84	85–93	96–103

Table 8 Overall Accuracy of SVM classifiers on bands groups of ROSIS data for different training size

Training size	Band Group#1	Band Group#2	Band Group#3	Band Group#4	Band Group#5	Band Group#6	Band Group#7	Band Group#8
%5	60.778	52.333	52.446	49.778	60.778	61.556	55.348	45.231
%10	61.778	54	63.642	51.333	61.556	60.444	56.284	46.468
%20	62.333	54.866	64.111	53.556	62	66	57	47.021

a 145*145 pixel portion of the AVIRIS image acquired over north-western US, Indian Pine, captured in June1992. The Indian Pine data is available in Purdue University site. The second data set is from Pavia University which is another common hyperspectral data set. This data has been captured by the German ROSIS sensor during a flight campaign over Pavia, northern Italy.

AVIRIS data contains 220 spectral bands in wavelength range 0.4–2.5 μm but not all of the 220 original bands are utilized in the experiments since 18 bands are affected by atmosphere absorption phenomena and are consequently discarded. Hence, the dimensionality for the AVIRIS Indian Pine data set is considered 202. In this experiment Fig. 2 shows original data and ground truth of the AVIRIS Indian Pine data. From the 16 different land-cover classes available in the original ground truth, seven are discarded; since only a few training samples related to these classes are available. The remaining nine land-cover classes are used to generate the training and test data sets (Table 1).

In ROSIS data, there exist 103 spectral bands which covering the wavelength range from 0.43 to 0.86 μm . This data set exhibits 610*340 pixels with 1.3 m per pixel geometric resolution. Pavia University data is available in Pavia University site. Figure 3 and Table 2 show ROSIS Pavia University data set.

Experimental Results on AVIRIS Data Set

In the first step of the proposed method, it is necessary to perform band grouping process based on the mutual

information processing in order to split hyperspectral data into band groups. Figure 4 shows the obtained results for AVIRIS data by using mutual information as similarity measure between adjacent bands. In this figure, local minima points correspond to the bands with low redundancy. Initial band groups would be produced based on these points. Moreover, a threshold is considered related to the minimum number of bands in each cluster. Table 3 shows the final band grouping results after pruning MI-based groups by applying GA-SVM. This table shows that the proposed band grouping method produced 12 combinations of bands on AVIRIS hyperspectral image.

After band grouping, one-against-one SVM was applied on 12 band clusters. As proposed in 5.2, proposed strategy applied grid search as the model selection of SVM classifier. The search range for C is in $[2^{-2}, 2^{10}]$, and $[2^{-10}, 2^2]$ for γ .

Table 4 represents the overall accuracy of SVM classifiers which are applied to each band combination. In order to investigate the impact of the number of labelled data on the classifier performance, all experiments were applied to different percentage of training and testing data sets. After producing the multiple of classifiers for AVIRIS data, three decision fusion strategies (i.e. Naive Bayes, Weighted Majority Voting and SVM) are applied to the results of band group's classification. In order to show the merits of the proposed methodology, this paper compares a standard SVM on entire bands of AVIRIS data and two feature selection methods. The feature selection

Table 9 Comparison between classification strategies on ROSIS data

Training Size	Fusion methods			Full data	Feature Selection Methods	
	WMV	SVM	NB		SFSS	SFBS
%5	88.667	89.778	90.222	87.778	88.7	89
%10	91.778	92	92.444	90.667	91.8	92
%20	93.111	93.333	95.111	93.111	93.8	93.4

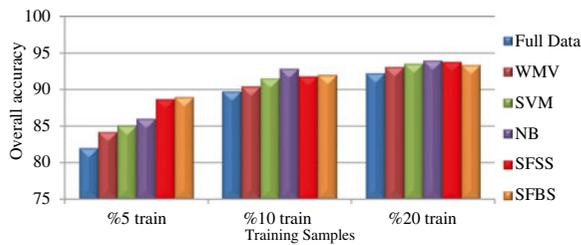


Fig. 9 Comparison between overall accuracies of different classification strategies on ROSIS data

strategies are the “sequential forward floating selection” (SFSS) and the “sequential backward floating selection” (SBFS) techniques, which identify the best feature subset that can be obtained by starting from an empty set SFSS or from the complete set of features SFBS and adding to SFS or removing from SBS the current feature subset one feature at a time until the desired number of features are achieved. More details have been provided by Pudil et al., in (1994).

Table 5 compares the results of the three fusion strategies with standard SVM and with the feature selection methods for different percentage of training samples. In terms of classification performance, this table shows that the resulting classification after classifier fusion is generally more accurate than standard SVM.

The overall results in Fig. 5 clearly demonstrate that the proposed multiple classifier system outperforms the feature selection strategies in terms of accuracy, irrespective of the number of training samples. This improvement benefits from splitting all bands of the

hyperspectral data into some band groups and applying the multiple classifier system on the produced band groups.

From the classification accuracy viewpoint, all three fusion strategies resulted in satisfactory results when compared with the standard SVM. In more detail, the Naive Bayes fusion strategy represented the best accuracy with a gain in overall accuracy of 94 % with 20 % training samples that caused accuracy improvement of standard SVM up to 3.2 %. The analysis of Fig. 5 shows two other fusion methods, Weighted Majority Voting and SVM-based fusion; perform better than the standard SVM up to 0.88 % and 1.3 % respectively. Figure 6 shows the classification map for NB fusion method and standard SVM on AVIRIS data. This visual interpretation supports the results of the statistical accuracy assessment.

Figure 7 demonstrates the accuracies of different classification strategies for all nine classes of the AVIRIS data set. For some classes, one (e.g. Class 9#Woods), two (e.g. Class 6#Soybeans-no till) or all three (e.g. Class 7#Soybeans-minimum till) fusion algorithms perform better than the results of standard SVM in terms of classification accuracy. This suggests that the decomposition of hyperspectral classification problem into a multiple of classifiers represents an effective way of improving the overall discrimination capability.

As shown in Fig. 7, the NB fusion method outperforms most standard SVM class accuracies or at least achieves similar results. Although the NB

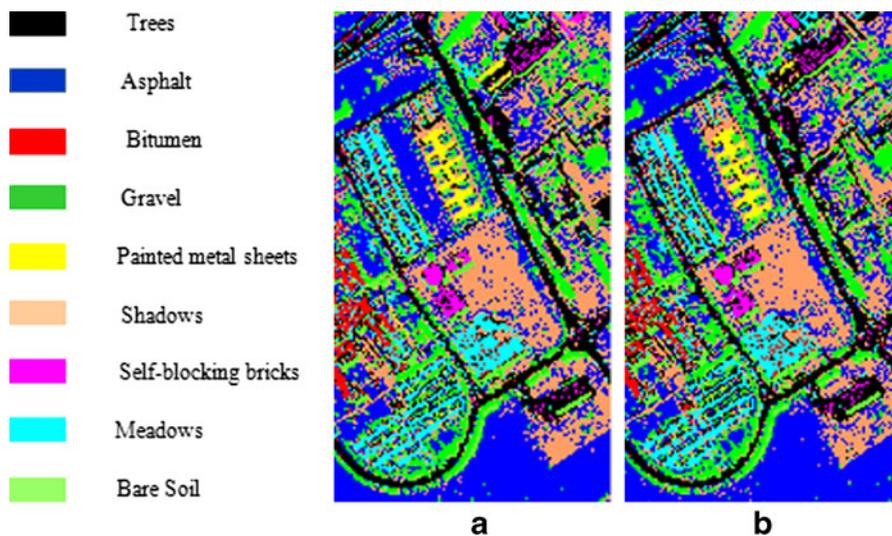


Fig. 10 Comparison between classification maps of ROSIS data for, a) Standard SVM and b) Naive Bayes fusion method

method improves the overall accuracy of standard SVM and other fusion methods, there are still some classes for which this method produced lower accuracies than other methods specially for class 2 (Corn-minimum till) and class 4 (Grass/trees). Since the diversity and correlation between classifiers are the basic assumption of an adequate classifier combination, this paper also computed the measures Disagreement and Double Fault for the AVIRIS data set (Table 6).

Results show that the MCS applied to AVIRIS data has low correlation and high diversity between classifiers. This is the most important cause of high performance of multiple classifier system in the proposed methodology.

Experimental Results on ROSIS Data Set

In order to prove the efficiency of the proposed methodology, further experiments are performed on the second hyperspectral data set. Regarding the band grouping similar results are obtained on ROSIS Pavia University data set as for AVIRIS data set but with fewer groups. Figure 8 shows the mutual information result on 103 spectral bands of ROSIS data.

The initial band groups are generated based on the MI measures pruned by using the proposed GA-SVM strategy. Table 7 represents final 8 band groups on this data set. The results of SVM classifier for this data set are shown in Table 8 in terms of overall accuracies.

For comparative purposes, all three decision fusion methods, standard SVM and feature selection methods are applied on the ROSIS data set.

Similar to Table 5 for AVIRIS data, Table 9 compares results of fusion methods and standard classifier as well as feature selection strategies for ROSIS data set. It can be observed that all three decision fusion methods specially NB improve the results of traditional classification results on full data. Figure 9 represents overall accuracies of all applied classification strategies

The comparison of accuracy improvement of Bayesian fusion algorithm with respect to the standard SVM on two hyperspectral images illustrates that the AVIRIS data exhibited higher performance. The comparison of the results of Tables 5 and 9 show that this improvement for AVIRIS data is 3.2 % while it is 2 % for the ROSIS data.

Figure 10 represents the visual classification results of standard classification and NB fusion methods for ROSIS data. In order to compare the classification

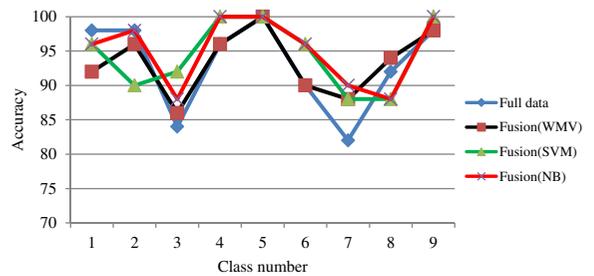


Fig. 11 Comparison of class accuracies between standard SVM and fusion methods for ROSIS data

methods in terms of single class accuracies, Fig. 11 illustrates class accuracies for all classification strategies. Similar to results on first hyperspectral image, Bayesian fusion strategy (NB) outperforms standard SVM. However, for class 1 (Trees) and class 8 (Self-Blocking Bricks), NB method exhibits lower accuracies in comparison with the standard classifier. Finally, Table 10 demonstrates the correlation and diversity measures of the classifiers in MCS for the ROSIS data. The comparison between the results of the two data sets shows that the MCS for the first data set is superior with respect to correlation and diversity measures. It means that multiple classifier system for the AVIRIS data set exhibits higher diversity and lower correlation measures in comparison ROSIS data.

Discussion and Conclusion

In this paper, the performance of a SVM based multiple classifier system for classification of hyperspectral imageries is assessed. The proposed approach applies a band grouping system based on the modified mutual information on hyperspectral image, in order to split it into few band groups. After that SVM classifiers are trained on each group to produce a multiple classifier system. Finally decision fusion strategies are applied to fuse the decisions of all the classifiers.

The first objective of the proposed method concerns the effectiveness of the band grouping strategy

Table 10 Correlation and diversity measures for ROSIS data set

Data Set	Correlation	Disagreement (Diss) ↓	Double Fault (DF) ↑
Pavia University	0.312	0.219	0.781

to solve the high dimensionality problem of hyperspectral data. Some previous researches only tried to select useful bands in dimension reduction techniques to overcome data redundancies. Nevertheless, the main drawback of dimension reduction techniques is related to the loss information through elimination of some bands. Using band grouping, the proposed method tries to overcome this weakness by enabling the use of the entire high dimensional hyperspectral image space. Using the conventional SVM for the whole heterogeneous data requires the definition of one single kernel function, which may not be suitably adapted to the whole diversity of information. It might be more adequate to take advantage of this heterogeneity by splitting the data into a few distinct subsets, defining a kernel function that is adapted for each data source separately and fuses the derived outputs. To achieve this, band grouping step of the proposed method overcomes this difficulty. In fact, the kernel of each individual classifier on each band cluster is adjusted according to the corresponding properties of those band clusters.

The second objective of the work is concerned with combining different classification results to improve the classification accuracy. Multiple classifier systems – combining the results of a set of base learners – have demonstrated promising capabilities in improving classification accuracy. As result, in this paper all decision fusion algorithms outperformed standard SVMs which use the entire set of bands of the hyperspectral image. Comparing the results of all the experiments carried out on the two considered datasets show that the proposed SVM-based MCS provided higher accuracy than an SVM standard classifier and feature selection strategies. Because of the high robustness and accuracy of Bayesian decision fusion method (NB) this method outperforms the two other fusion strategies.

In comparison with the other research papers in terms of classification accuracy, Ceamanos and his colleagues (2010) received approximately 91 % overall accuracy for 25 % of training samples with 6 classifiers on Indiana data set in a SVM ensemble system while the results were improved by our proposed method to be up to 3 % for 12 classifiers. The results obtained by the proposed MCS classification approach gave both better classification accuracies and a higher robustness compared to the traditional classifiers as conclusion of this paper. Further studies will

focus on the new decision fusion methods, novel band grouping strategies and using of new classification methods especially fuzzy classifiers. In addition, to improve classification results after solving Hughes problem, the spatial information can be integrated.

References

- Benediktsson, J. A., & Kanellopoulos, I. (1999). Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), 1367–1377.
- Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transaction on Geoscience and Remote Sensing*, 43(6), 1351–1362.
- Ceamanos, X., Waske, B., Benediktsson, J., Chanussot, J., Fauvel, M., & Sveinsson, J. (2010). A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4), 293–307.
- Del Frate, F., Pacifici, F., Schiavon, G., & Solimini, C. (2007). Use of neural networks for automatic classification from high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), 800–809.
- Goebel, K., Yan, W., & Cheatham, W. (2002). A Method to Calculate Classifier Correlation for Decision Fusion, Proceedings of IDC (Information, Decision Control Conference) 2002, Adelaide, pp. 135–140.
- Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnel, R. B., & Viau, A. A. (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture*, 39(2), 67–93.
- Guo, B., Gunn, S. R., Damper, R. I., & Nelson, J. D. B. (2006). Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4), 522–526.
- Hsu, C.-W., Chung, C.-C., & Lin, C.-J. (2010). A Practical Guide to Support Vector Classification. National Taiwan University, March 13, 2010 [Online]. Available: www.csie.ntu.edu.tw/~cjlin
- Imbault, F., & Lebart, K. (2004). A stochastic optimization approach for parameter tuning of support vector machines, Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), 4, 597–600.
- Jia, X. (2002). Simplified maximum likelihood classification for hyperspectral data in cluster space. *IEEE International Geoscience and Remote Sensing Symposium, 2002 (IGARSS '02)*, 5, 2578–2580.
- Kuncheva, L. (2004). *Combining Pattern Classifiers methods and algorithms*. Hoboken: John Wiley & Sons, INC. Publication.
- Kuncheva, L., & Whitaker, C. (2003). Measures of diversity in classifier ensemble and their relationship with the ensemble accuracy. *Machine Learning*, 51(2), 181–207.

- Li, S., Wu, H., Wan, D., & Zhu, J. (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, 24(1), 40–48.
- Martinez-Uso, A., Pla, F., Sotoca, J.M., & Garcia-Sevilla, P. (2006). Clustering based multispectral band selection using mutual information. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR), 2, 760–763.
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transaction on Geosciences and Remote Sensing*, 42(8), 1778–1790.
- Müller, K. L., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transaction on Neural Network*, 12(2), 181–202.
- Prasad, S., & Bruce, L. M. (2008). Decision fusion with confidence based weight assignment for hyperspectral target recognition. *IEEE Transaction on Geosciences and Remote Sensing*, 46(5), 1448–1456.
- Pudil, P., Novovicova, P., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letter*, 15(11), 1119–1125.
- Ruta, D., & Gabrys, B. (2000). An overview of classifier fusion methods. *Computing and Information Systems*, 7(1), 1–10.
- Scholkopf, B., & Smola, A. J. (2002). *Learning with kernels, support vector machines, regularization, optimization and beyond*. Cambridge: MIT Press.
- Tarabalka, Y., Fauvel, M., Chanussot, J., & Benediktsson, J. (2010). SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosciences and Remote Sensing letters*, 7(4), 736–740.
- Vaiphasa, C. (2003). Innovative genetic algorithm for hyperspectral image classification, In *Proceeding of International Conference of Map Asia*, pp. 20.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.
- Waske, B., & Van der Linden, S. (2008). Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transactions on Geosciences and Remote Sensing*, 46(5), 1457–1466.
- Watanachaturaporn, P., & Arora, M. K. (2004). Support vector machines for classification of multi- and hyperspectral data. In P. K. Varshney & M. K. Arora (Eds.), *Advanced image processing techniques for remotely sensed hyperspectral data* (pp. 237–255): Springer-Verlag.
- Weston, J., & Watkins, C. (1999). Support vector machines for multiclass pattern recognition. In *The Seventh European Symposium on Artificial Neural Networks*, pp. 219–224.
- Xu, L., Krzyzak, A., & Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3), 418–435.