Multisource Classification Using Support Vector Machines: An Empirical Comparison with Decision Tree and Neural Network Classifiers

Pakorn Watanachaturaporn, Manoj K. Arora, and Pramod K. Varshney

Abstract

Remote sensing image classification has proven to be attractive for extracting useful thematic information such as land-cover. However, often for a given application, spectral information acquired by a remote sensing sensor may not be sufficient to derive accurate information. Incorporation of data from other sources such as a digital elevation model (DEM), and geophysical and geological data may assist in achieving more accurate land-cover classification from remote sensing images. Recently, support vector machines (SVM) have been proposed as an alternative for classification of remote sensing data, and the results are promising. In this paper, we employ the SVM algorithm to perform multi-source classification. An IRS–1C LISS III image along with normalized differenced vegetation index (NDVI) image and DEM are used to produce a land-cover classification for a region in the Himalayas. The accuracy of SVM-based multi-source classification is compared with several other non-parametric algorithms namely a decision tree classifier, and back propagation and radial basis function neural network classifiers. The well-known kappa coefficient of agreement is used to assess classification accuracy. The differences in the kappa coefficient of classifiers have been statistically evaluated using a pairwise Z-test. The results show a significant increase in the accuracy of the SVM based classifier on incorporation of ancillary data over classification performed solely on the basis of spectral data from remote sensing sensors.

Introduction

Over the years, numerous studies have clearly demonstrated the utility of remote sensing data for extracting accurate thematic information. A variety of classifiers to produce remote sensing classification are in vogue. In general, the mapping from these classifiers is based only on the spectral response of the classes. However, in areas particularly mountainous regions where there is large variation in the spectral response of classes due to high relief and shadow, mapping solely on the basis of spectral response may not be appropriate (Arora and Mathur, 2001) Moreover, information from an individual sensor may be incomplete, inconsistent, and imprecise (Rao and Arora, 2004; Simone *et al.*, 2002) in some circumstances. Therefore, incorporation of additional or ancillary data sources in the process of remote sensing classification may result in better understanding and achievement of higher accuracy than utilizing spectral data from a remote sensing sensor alone. The ancillary data from various sources may be available in different forms and contexts, and at different frequencies, time, and spatial domains. Integration of data from different sources may also be referred to as image or data fusion (Pohl and van Genderen, 1998). Depending on the nature of data sources and methodology used, data fusion may be categorized as multi-source, multi-sensor, multi-temporal, multi-frequency, multi-polarization, or multi-resolution fusion (Arora and Mathur, 2001; Rao and Arora, 2004; Simone *et al.*, 2002). The classification of remote sensing data along with data from other sources has generally been referred to as multisource classification.

In the past, several studies (e.g., Benediktsson and Sveinsson, 2003; Bruzzone *et al.*, 1999; Fitzgerald and Lees, 1994; Peddle *et al.*, 1994) were conducted on multi-source classification, and significant improvement in classification accuracy was achieved. These studies indicate that conventional parametric statistical classifiers have limitations in performing multisource classification. One major limitation of parametric classifiers is their inability to classify data at different measurement scales and units (Peddle *et al.*, 1994). In these cases, the assumption that every dataset has the same statistical distribution may not be valid. Therefore, a non-parametric classifier may be appropriate, as it does not depend on the statistical distribution of the dataset. Examples of non-parametric classifiers for multisource classification in the remote sensing

Pakorn Watanachaturaporn is with the Department of Computer Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand.

Manoj K. Arora is with the Department of Civil Engineering, Indian Institute of Technology Roorkee, Roorkee, 247 667 India (manojfce@iitr.ernet.in).

Pramod K. Varshney is with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, 13244.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 2, February 2008, pp. 239–246.

0099-1112/08/7402–0000/$3.00/0

literature are artificial neural networks (Arora and Mathur, 2001; Foody, 1995; Foody and Arora, 1997), knowledge based classifiers (Srinivasan and Richards, 1990), fuzzy set based classifiers (Key *et al.*, 1989), classifiers based on consensus theory (Benediktsson and Sveinsson, 2003) and boosting methods (Schapire, 1999).

Recently, support vector machines (SVM) have shown a great promise in remote sensing classification. For instance, the studies by Guilliateri *et al.* (1999), Huang (2002), Halldorsson *et al.* (2003), Foody and Mathur (2004a and 2004b), Pal and Mather (2005), Watanachaturaporn *et al.* (2004 and 2005), Bruzzone *et al.* (2006) have shown the potential of SVM for classification of both multi spectral and hyperspectral data with high accuracies. In most of these studies, however, the classifications have been performed solely on the basis of spectral data. In this paper, we investigate the application of SVM classifiers for multisource classification by incorporating additional data sources.

Multisource Image Classification

Multisource data can be incorporated into the classification process using the stacked vector or the logical channel approach (Arora and Mathur, 2001; Mather, 1999; Richards and Jia, 1999; Wolpert, 1992). It is the most straightforward approach where the data from each source are treated as augmented dimensions of an input data vector. For example, if one has four spectral bands of remote sensing data, one band in the form of digital elevation model (DEM), and one band of a raster geological data, then all these six bands can be stacked together to form the input to a classifier. The fusion paradigm employed here is a data or pixel-level fusion as the raw data is fused for classification (Pohl and van Genderen, 1998).

Although the stacked vector approach is straightforward and easy to implement, some points require attention if the approach is used with a parametric classifier such as the maximum likelihood classifier (Tso and Mather, 2001). First, the scale of measurement of each data source may be different. Consequently, all data need to be normalized to bring them into the same scale. Second, when the dimension of the data vector becomes large, the stacked vector approach may be computationally inefficient. For example, when a maximum likelihood classifier is used, the computational cost is proportional to N^2 for classifying an N dimensional data vector. The severity of the problem may be reduced by using a feature reduction procedure. However, it may result in discarding of data sources that may contain useful information. Finally, the stacked vector approach is based on the assumption that each data source is equally reliable thereby equally contributing to the process of determining decision boundaries, which may not always be true.

However, when using the stacked vector approach with a non-parametric classifier, some limitations may be overcome. For instance, a decision tree classifier (DTC) can take inputs of different types; e.g., digital numbers from a remote sensing image (ratio/interval data) and class attributes (categorical data). Support Vector Machines (SVMs) perform very well with high dimensional data (Watanachaturaporn *et al.*, 2004). The computational cost of SVMs does not depend on data dimensionality, and as such, no feature selection is required. Thus, classification result for multisource classification from a non-parametric classifier is likely to be better than that obtained from a parametric classifier since a non-parametric classifier can overcome some of the limitations of a stacked vector approach.

Support Vector Machines (SVM) Based Classification

The construction of SVMs has been sufficiently described in the literature (Vapnik, 1995; Watanachaturaporn and Arora, 2004; Watanachaturaporn *et al.*, 2004, Bruzzone *et al.*, 2006). Therefore, only a brief description is provided here. Consider a binary classification problem, where the data is partitioned into two classes, which can be separated by a linear hyperplane. Assume that the training data consists of k samples represented by $(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_k,y_k)$, where $\mathbf{x}_i \in \Re^N$ is an N-dimensional data vector with each sample belonging to either of the two classes labeled as $y_i \in \{-1, +1\}$. The goal of an SVM is to find a linear decision function defined by $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$, where $\mathbf{w} \in \Re^N$ determines the orientation of the discriminating hyperplane, and $b \in \Re$ is a bias. The hyperplanes for the two classes are, therefore, represented by $y_i(\mathbf{w} \cdot \mathbf{x} + b) \geq 1$. Due to the noise or mixing of classes during the selection of training samples, variables $\xi_i > 0$, called the slack variables, are often introduced to account for the effect of misclassification. The hyperplanes for the two classes, in this case, become $y_i(\mathbf{w} \cdot \mathbf{x} + b) \geq 1 - \xi_i$. The optimal separating hyperplane (i.e., $f(\mathbf{x}) = 0$) is located where the margin between the two classes is maximized, and the misclassification is minimized. They can be obtained by solving the following constrained optimization problem by the method of Lagrange multipliers (Cristianini and Shawe-Taylor, 2000):

$$\textit{Maximize:}\ \sum_{i=1}^{k}\alpha_i - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j y_i y_j\left(\mathbf{x}_i \cdot \mathbf{x}_j\right), \tag{1}$$

$$\textit{Subject to:}\ \sum_{i=1}^{k}\alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C, \textit{ for } i = 1, 2, \ldots, k$$

where $\alpha_i > 0$ are the Lagrange multipliers. The solution of the optimization problem results in the set of Lagrange multipliers, $\alpha^o = \left(\alpha_1^o, \ldots, \alpha_k^o\right)$. According to the Karush-Kuhn-Tucker (KKT) optimality condition (Fletcher, 1987), some of the multipliers may be zero. The multipliers that have non-zero values are called the support vectors. The values of $\mathbf{w}$ and b are computed from $w^o = \sum_{i=1}^{k} y_i\alpha_i^o\mathbf{x}_i$ and $b^o = \frac{1}{2}\left[\mathbf{w}^o \cdot \mathbf{x}_{+1}^o + \mathbf{w}^o \cdot \mathbf{x}_{-1}^o\right]$, where $\mathbf{x}_{+1}^o$ and $\mathbf{x}_{-1}^o$ are the support vectors corresponding to class labels $+1$ and -1, respectively. The decision rule is then applied to classify the data into the two classes, namely, $+1$ and -1:

$$f(\mathbf{x}) = \operatorname{sign}\left(\sum_{\text{support vector}} y_i\alpha_i^o\,(\mathbf{x}_i \cdot \mathbf{x}) + b^o\right). \tag{2}$$

There are instances where a linear hyperplane is not able to separate two classes without misclassification, but these classes can be separated by a nonlinear separating hyperplane. In such circumstances, the data may be mapped to a higher dimensional space using a nonlinear transformation function. In the higher dimensional space, data can be spread out (Cover, 1965), where a linear separating hyperplane may be found.

Let a nonlinear transformation function $\phi()$ map the data into a higher dimensional space. Suppose there exists a function K, called a kernel function, such that:

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j). \tag{3}$$

The kernel function may be substituted for the dot product of the transformed vectors in Equation 1 to redefine the optimization problem for a non-linear case. Accordingly, the decision function will be changed to:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{\text{support vector}} y_i \alpha_i^o K(\mathbf{x}_i.\mathbf{x}) + b^o\right). \tag{4}$$

Equation 4 can be used to classify the data into two classes.

It is thus clear that SVM can produce binary classification. However, classification of data into more than two classes, called multiclass classification, is frequent in remote sensing applications. A number of methods to generate multiclass SVMs from binary SVMs have been proposed. Details of multiclass classification methods along with their merits and demerits can be found in (Watanachaturaporn and Arora, 2004). Here, the simple and the most widely used multiclass method namely pairwise classification is used for various SVM-based classification experiments. In this method, SVM classifiers for all possible pairs of classes are created. Therefore, for M classes, there will be $\frac{1}{2}M(M-1)$ binary classifiers. The output from each classifier in the form of a class label is obtained. The class label that occurs the most is assigned to that point in the data vector. In case of a tie, a tie-breaking strategy may be adopted. A common tie-breaking strategy is to randomly select one of the class labels that are tied.

We compare the classification results obtained from the SVM-based classifier with three other non-parametric supervised classifiers namely the back-propagation neural network classifier (BPNN) (Atkinson and Tatnall, 1997; Haykin, 1999; Paola and Schowengerdt, 1995 and 1997), the radial basis function network classifier (RBFNET) (Bruzzone and Prieto, 1999; Light, 1992; Melgani and Bruzzone, 2004; Powell, 1985, 1988, and 1992), and the decision tree classifier (DTC) (Debeir *et al.*, 2001; Hansen *et al.*, 1996; Safavian and Landgrebe, 1991). We employ the BPNN and RBFNET available in the Matlab neural network toolbox and the DTC in the C4.5, release 8 (Quinlan, 1993). Lagrangian SVM (LSVM) (Mangasarian and Musicant, 2001) with pairwise classification method is employed for SVM classification using an in-house program developed in Matlab.

Since each of the selected classifier differs from each other in the methodology employed and implementation and is also dependent on a number of its own parameters, it is not practical to compare different classifiers directly. Therefore, the best performance obtained from each classifier; i.e., the classifier with the set of parameters that resulted in the highest accuracy is reported here. The classification accuracy has been evaluated using the error matrix, overall accuracy, and kappa coefficient. The most widely used standard Z-statistical test (Congalton and Green, 1999; Rosenfield and Fitzpatrick-Lins, 1986) has been used to statistically assess the difference between the kappa coefficients of two classifications at 95 percent level of significance, based on the assumption that the testing samples are independent. However, Foody (2004) has indicated that assumption of independent samples may not be satisfied and the samples may be related, a non-parametric test such as the McNamer test therefore must be used.

Experimental Data

The remote sensing image used in our experimental study was acquired from Indian Remote Sensing Satellite IRS-1C Linear Imaging Self-scanning Sensor (LISS) III having spatial resolution of 23.5 m and four spectral bands; one each in green, red, near-infrared (NIR), and shortwave infrared (SWIR) regions.

The study area covers a mountainous region in the Himalayas. Since this study area is difficult to access as a result of natural obstructions due to high altitude and rough terrain, remote sensing data is the best source to extract accurate land-cover information. A number of studies have been conducted on the use of remote sensing data to map land-cover in high altitude mountainous areas. However, the degrees of accuracy from the reported studies vary. They may be attributed to a number of factors such as presence of shadows in the image, deep narrow valleys, steep slopes, and different types of land-cover in the area. Due to variation in environmental conditions, spectral characteristics of land-cover also vary from one region to the other (Saha *et al.*, 2005). Therefore, classification on the basis of only spectral response of land-cover classes may not be sufficient to map them effectively in a mountainous region.

In the mountainous region, the effect of topography, in particular, is pronounced. To reduce the effect of topography, the remote sensing data may be rectified by applying appropriate topographic corrections, which may result in an increase in classification accuracy. Alternatively, the topographic information in the form of digital elevation model (DEM), as an additional band, may be incorporated in the classification process. Moreover, in a hilly region, because of the presence of shadows, the data may have to be radiometrically corrected, particularly when multi-temporal remote sensing data are used in the classification process.

In this study, we incorporate a digital elevation model (DEM) and a normalized differenced vegetation index (NDVI) image as additional data sources to produce the land-cover classification from the LISS III image. DEM has been used to account for the rugged topography of the region so as to eliminate the presence or absence of certain classes in some elevation zones whereas the inclusion of NDVI image may reduce the impact of shadows in the region and to enhance the separability among various vegetation classes. Thus, incorporating NDVI image in the classification process may increase the accuracy of classification, particularly in case of statistical classifiers. A detailed justification for the use of these data sources can be found in (Saha *et al.*, 2005). Fusion of information from these complementary data sources is likely to yield more accurate classification.

The LISS III image (Figure 1) consisting of 1,535 pixels × 1,370 pixels in four bands was acquired on 26 November 1998 and covers a portion of Rudraprayag and Chamoli districts of Uttarakhand state in India. The image was not radiometrically calibrated, as this was not considered necessary for this type of study based on single date image classification. The study area is approximately 730 km^2 of the Himalayas with elevation ranging from 920 m to 4,853 m above mean sea level. Two rivers flow through the northwestern and southeastern parts of the area. A dense forest covers more than one half of the area. In the northeastern portion, most of the area is barren, whereas the high mountains are covered with snow. The DEM image (Figure 2) at 23.5 spatial resolution was produced by digitizing contours from a topographic map at a scale of 1:50 000 with 40 m contour interval, as described in Saha *et al.* (2005). The NDVI image (Figure 3) was generated from red and near-infrared (NIR) bands of the LISS III image and is defined as:

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}. \tag{5}$$

The NDVI values range from −1 to +1, which are scaled from 0 to 255 for display purposes. The higher is the NDVI value, vegetation with higher biomass or more green leaf area content is expected.

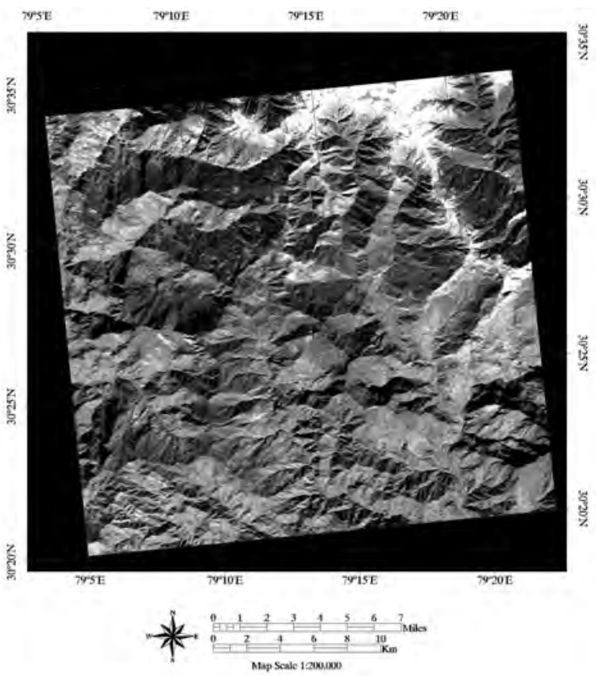

Figure 1. A color infrared composite obtained from the IRS LISS III for the Himalayan region. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

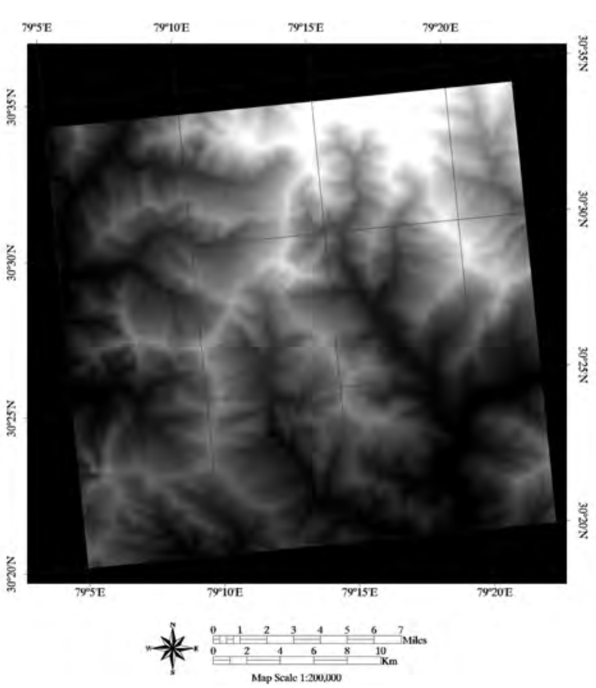

Figure 2. DEM image corresponding to the scene in Figure 1.

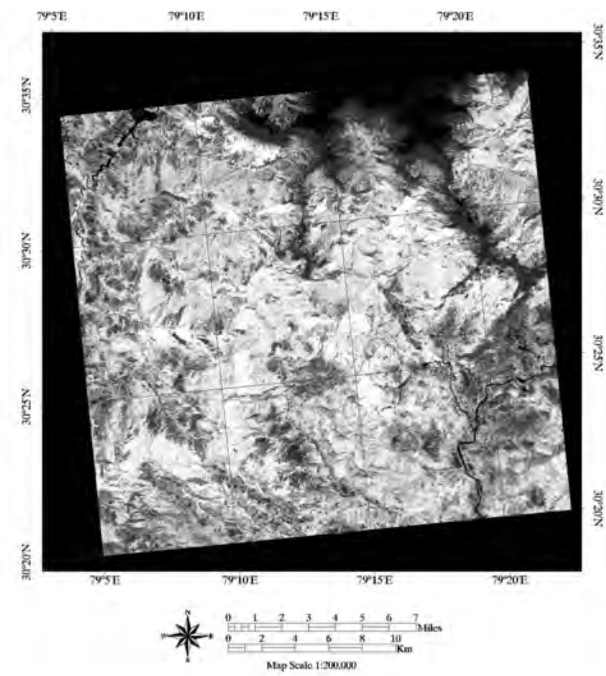

Figure 3. NDVI image corresponding to the scene in Figure 1.

TABLE 1. TOTAL NUMBER OF TESTING PIXELS PER CLASS

Class Name	Number of Testing Pixels
Dense forest	1889
Sparse vegetation	543
Agriculture	230
Fallow	667
Barren	954
Settlements	247
Fresh sediments	200
Water	309
Snow	821
Total	5860

The data for multisource classification thus consists of four spectral bands of LISS III sensors, one NDVI image, and one DEM. These data are numbered as 1 to 6 corresponding to the Green band, Red band, NIR band, SWIR band, NDVI image, and DEM, respectively. The area consists of nine land-cover classes as listed in Table 1. Two hundred pixels of each class were randomly selected from the reference data (Saha *et al.*, 2005) to train the classifiers. The number of testing pixels randomly selected for each class is given in Table 1.

A separability analysis is performed using the training dataset to know the extent of class separability between land-cover classes of interest with certain band combinations. Here, the most widely used separability measure namely average transformed divergence (TD) has been used to estimate the separability between classes (Jensen, 1996; Mather, 1999; Richards and Jia, 1999). Through separability analysis, an appropriate band combination may be found, which may be used as the input to a classifier. In our case, we do not use the separability measure to discard bands but to assess the increase in separability when ancillary data is included.

Table 2. Average TD Values of Various Band Combinations

	Band Combinations	Average TD values
Case I	1, 2, 3, 4	1940
Case II	1, 2, 3, 4, 5	1960
Case III	1, 2, 3, 4, 6	1984
Case IV	1, 2, 3, 4, 5, 6	1987

Since the goal of this study is to demonstrate the improvement in classification accuracy as additional data sources are included, we consider four cases. The first case consists of the data from the LISS III sensor only (i.e., four bands only). The second case is the combination of LISS III and NDVI images. The third case is the combination of the LISS III image and the DEM. The last case is the combination of data consisting of the LISS III image, the NDVI image, and the DEM. The average TD values obtained for various cases are given in Table 2.

It can be seen from the table that in each case the average TD values are high illustrating good separability, which increases further as the ancillary data are included. The highest average TD value is obtained for the combination consisting of all the data. This clearly shows that the inclusion of ancillary data increases the separability of the classes; therefore, is likely to produce higher classification accuracy than achieved by using original spectral data from remote sensing sensor alone.

Experimental Results and Discussion

For multisource classification, the datasets are stacked together and are used as input vectors for an SVM classifier. Classifications were performed using the datasets described as the four cases individually (see Table 2). A number of SVM classifiers were designed, each employing a different type of kernel. In this study, a linear kernel, polynomial kernels with degrees 2 to 7, and a radial basis function kernel were used. The solution of Equation 1 was obtained using the LSVM optimizer (Mangasarian and Musicant, 2001). Pairwise classification strategy was used to generate multi-class classification. The optimum penalty values (C values) of each classifier, where the overall classification accuracy is maximum, was selected from the set $\{10^n, n = -5, -4, \ldots, 0, \ldots, 4, 5\}$. Here, classification accuracies were assessed using overall accuracy, kappa coefficient of agreement, and the statistical differences in kappa coefficients were determined using the Z-test (Congalton and Green, 1999) assuming that the testing samples were independent. Since the general nature of the interpretation of the results would not have changed much by applying any other test such as the McNamer test, the latter one was not applied here. However, future studies, which aim to compare performances of different classifiers, should consider the use of this test.

The best overall accuracy obtained from different SVM classifiers based on the optimum penalty value for each case is provided in Table 3. The Z values to test for significant difference between the kappa coefficients of classifications produced from the datasets of two cases at a time for each kernel function are given in Table 4.

From Table 3 and Table 4, it can be seen that for all the cases, the classification accuracies are generally very high (i.e., in the range of 94 percent to 99 percent overall accuracy), which show the superb performance of SVM classifiers. When NDVI image is included, the classification accuracies marginally increase (Z values <1.96) over those obtained from using remote sensing data alone. The classification accuracies, however, improve significantly (Z values approximately 12 to 13) when DEM is included as an ancillary data source. An increase in classification accuracy of about 4 percent has been achieved for each kernel, which is a significant increase considering the fact that the accuracy values are over 90 percent. When both the NDVI image and the DEM are included (Case IV), although the accuracies increase significantly (Z values of approximately 11 to 12) over those obtained from Case I, they are approximately the same as obtained from (Case III). This shows that, for the classification of the mountainous region selected, DEM is the most effective ancillary data source, the inclusion of which in the classification process has resulted in significant improvement in mapping land-cover.

The SVM classifiers implemented with all the kernels resulted in high classification accuracies in the range of 95 percent to 99 percent. Since the SVM classifier with the linear kernel, which requires less computational effort, has performed competitively well in comparison to other kernels, the linear kernel may be recommended for remote sensing classification using SVM.

Table 3. Best Overall Accuracy (Percent) and Kappa Coefficient of the SVM-based Classifier as Obtained from the Use of Different Kernels

Kernel	Accuracy Measures	Case I	Case II	Case III	Case IV
Linear	Overall accuracy (%)	94.20	94.23	98.53	98.48
	Kappa Coefficient	0.93	0.93	0.98	0.98
Polynomial	Overall accuracy (%)	95.29	95.24	99.04	98.98
degree 2	Kappa Coefficient	0.94	0.94	0.99	0.99
Polynomial	Overall accuracy (%)	95.27	95.32	99.01	99.06
degree 3	Kappa Coefficient	0.94	0.94	0.99	0.99
Polynomial	Overall accuracy (%)	95.14	95.32	99.15	99.15
degree 4	Kappa Coefficient	0.94	0.94	0.99	0.99
Polynomial	Overall accuracy (%)	95.38	95.29	99.08	99.21
degree 5	Kappa Coefficient	0.94	0.94	0.99	0.99
Polynomial	Overall accuracy (%)	95.32	95.41	99.03	99.23
degree 6	Kappa Coefficient	0.94	0.94	0.99	0.99
Polynomial	Overall accuracy (%)	95.39	95.48	98.96	99.13
degree 7	Kappa Coefficient	0.94	0.95	0.99	0.99
RBF	Overall accuracy (%)	95.15	95.07	98.96	98.94
	Kappa Coefficient	0.94	0.94	0.99	0.99

Table 4. Pairwise Z-values to Test the Significant Difference Between Kappa Coefficient of Classifications Produced from Two Cases. Note that Values Greater than 1.96 are Significant at 95 Percent Degree of Confidence

	Case IV	Case III	Case II
Case I	12.51	12.70	0.08
Case II	12.43	12.63	
Case III	0.23		

(a) Linear kernel

	Case IV	Case III	Case II
Case I	12.25	12.33	0.21
Case II	12.45	12.52	
Case III	0.09		

(b) RBF kernel

	Case IV	Case III	Case II
Case I	12.09	12.39	0.13
Case II	12.21	12.50	
Case III	0.37		

(c) Polynomial kernel of degree 2

	Case IV	Case III	Case II
Case I	12.49	12.26	0.13
Case II	12.38	12.15	
Case III	0.28		

(d) Polynomial kernel of degree 3

	Case IV	Case III	Case II
Case I	13.18	13.18	0.47
Case II	12.76	12.76	
Case III	0.00		

(e) Polynomial kernel of degree 4

	Case IV	Case III	Case II
Case I	12.96	12.34	0.22
Case II	13.15	12.53	
Case III	0.81		

(f) Polynomial kernel of degree 5

	Case IV	Case III	Case II
Case I	13.16	12.23	0.22
Case II	12.96	12.03	
Case III	1.19		

(g) Polynomial kernel of degree 6

	Case IV	Case III	Case II
Case I	12.53	11.77	0.22
Case II	12.33	11.57	
Case III	0.95		

(h) Polynomial kernel of degree 7

The results from the SVM classifier were also compared with the well known maximum likelihood classifier (MLC), backpropagation neural network classifier (BPNN), radial basis function network classifier (RBFNET), and decision tree classifier (DTC) using the same training and testing datasets. The best overall accuracies obtained from these classifiers, after optimizing the parameters of respective classifiers, are given in Table 5. The Z values to statistically evaluate the classifications produced from all the classifiers are given in Table 6a through 6d.

Similar to the SVM classifier, the results from these classifiers also show that the accuracy does not increase significantly when the NDVI image is included with the remote sensing data. However, it increases significantly when DEM is included. Highest accuracies from these four classifiers are obtained from Case IV, when both NDVI and DEM data are included in the classification process.

Comparing the results of multisource classification from the SVM classifier (polynomial degree 6 kernel) with the four classifiers, it can be seen that the SVM classifier has produced significantly higher accuracy (see high Z values in Table 7) than the MLC and the RBFNET for all the cases and the DTC for most of the cases (except Case I). However, results from the BPNN classifier are not statistically different from the SVM classifier indicating that both classifiers are equally good.

Table 5. Overall Accuracy (Percent) from Various Classifiers

Classifier	Case I	Case II	Case III	Case IV
MLC	93.94	93.77	97.90	98.14
BPNN	95.07	95.10	99.02	99.06
RBFNET	87.30	86.81	90.12	90.63
DTC	94.66	94.30	97.83	98.52
SVM with polynomial kernel of degree 6	95.32	95.41	99.03	99.23

Table 6. Pairwise Z-values to Test Significant Difference Between Kappa Coefficients of Classifications Produced from all the Classifiers

	Case IV	Case III	Case II
Case I	11.79	10.95	0.39
Case II	12.14	11.31	
Case III	0.93		

(a) MLC

	Case IV	Case III	Case II
Case I	12.96	12.81	0.09
Case II	12.89	12.74	
Case III	0.19		

(b) BPNN

	Case IV	Case III	Case II
Case I	5.80	4.88	0.82
Case II	6.62	5.70	
Case III	0.92		

(c) RBFNET

	Case IV	Case III	Case II
Case I	11.61	9.11	0.85
Case II	12.40	9.93	
Case III	2.75		

(d) DTC

Table 7. Pairwise Z-values to Test Significant Difference Between Kappa Coefficient Between Classifications Produced from an SVM-based Classifier with the Polynomial of Degree 6 Kernel with Other Classifiers

Classifier	Case I	Case II	Case III	Case IV
MLC	3.34	3.94	4.96	5.20
BPNN	0.66	0.79	0.00	1.00
RBFNET	15.76	16.77	21.91	21.85
DTC	1.67	2.73	5.20	3.68

The error matrices generated from classification obtained from the SVM classifier (polynomial of degree 6 kernel) were also analyzed. For brevity, these error matrices have not been given here. However, from the inspection of the error matrix of classification produced from remote sensing data alone, it was revealed that the class dense forest, sparse vegetation, fallow land, and barren land are highly confused with other classes resulting in misclassifications and thus lowering the accuracy. However, these misclassifications were reduced on addition of NDVI data and were further reduced when the DEM data were included. In fact, the classes were mapped with

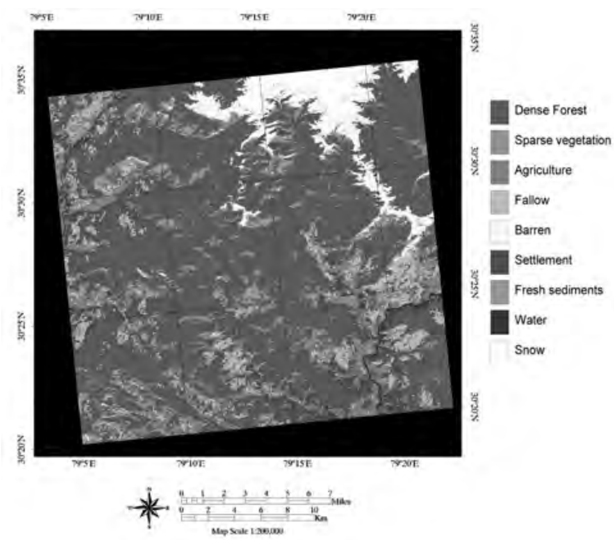

Figure 4. Classified image from SVM based classifier with the polynomial of degree 6 kernel. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

high accuracy when both the NDVI image and the DEM were included, as the misclassifications reduced significantly. For visual representation of classification, a land-cover map produced from the most accurate SVM-based classifier is shown in Figure 4. Thus, the results from this study clearly indicate that SVM can be successfully used for multisource classification yielding a high degree of accuracy.

Summary

The use of SVM based classifiers to improve the classification accuracy by fusing additional data sources was investigated in this paper. Remote sensing data was acquired over a mountainous region with high relief which resulted in shadowed areas. This might lead to inaccurate classification if only spectral data from remote sensing sensors were used. Therefore, ancillary data were included to enhance the quality of classification. An IRS-1C LISS III image was used along with NDVI and DEM data as multi-data sources. The results show that the accuracy of classification produced from SVM classifier significantly increases with the inclusion of data from other sources. This increase is observed in classification performed using different types of kernels used to construct the SVM. For comparison purposes, four other well-known classifiers namely an MLC, a DTC, a BPNN, and an RBFNET were also used to classify the same data. The accuracy of the SVM classifier was significantly higher than those obtained from MLC and the RBFNET. However, no significant difference between the accuracy of SVM-based classifier and the BPNN were observed. These results clearly show that SVM-based classifiers have the potential to produce accurate multisource classification and can be a useful classifier when data at different measurement scales and units are to be incorporated.

Acknowledgment

This work was supported by NASA Grant Number NAGS-11227.

References

Arora, M.K., and S. Mathur, 2001. Multi-source classification using artificial neural network in rugged terrain, *GeoCarto International*, 16(3):37–44.

Atkinson, P.M., and A.R.L. Tatnall, 1997. Neural network in remote sensing, *International Journal of Remote Sensing*, 18(4):699–709.

Benediktsson, J.A., and J.R. Sveinsson, 2003. Multisource remote sensing data classification based on consensus and pruning, *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):932–936.

Bruzzone, L., and D.F. Prieto, 1999. A technique for the selection of kernel function parameters in RBF neural networks for classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 37:1179–1184.

Bruzzone, L., D.F. Prieto, and S.B. Serpico, 1999. A neural statistical approach to multitemporal and multisensor and multisource remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1350–1359.

Bruzzone, L., M. Chi, and M. Marconcini, 2006. A novel transductive SVM for semisupervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 44(11): 3363–3373.

Congalton, R.G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press.

Cristianini, N., and J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*, University Press, Cambridge, U.K., New York, Cambridge.

Debeir, O., P. Latinne, and I.V.D. Steen, 2001. Remote sensing classification of spectral, spatial and contextual data using multiple classifier systems, *Image Analysis and Stereology*, 20(Supplement 1):584–589.

Fitzgerald, R.W., and B.G. Lees, 1994. Assessing the classification accuracy of multisource remote sensing data, *Remote Sensing of Environment*, 47(3):362–368.

Fletcher, R., 1987. *Practical Methods of Optimization*, Second edition, Chichester, Wiley, New York.

Foody, G.M., 1995. Land cover classification by an artificial neural network with ancillary information, *International Journal of Geographical Information System*, 9(5):527–542.

Foody, G.M., and M.K. Arora, 1997. An evaluation of some factors affecting the accuracy of an artificial neural network, *International Journal of Remote Sensing*, 18(4):789–810.

Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 70(5):627–633

Foody, G.M., and A. Mathur, 2004a. A relative evaluation of multiclass image classification by support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1335–1343.

Foody, G.M., and A. Mathur, 2004b. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification, *Remote Sensing of Environment*, 93(1–2): 107–117.

Gualtieri, J.A., S.R. Chettri, R.F. Cromp, and L.F. Johnson, 1999. Support vector machine classifiers as applied to AVIRIS data, *The Summaries of the Eighth JPL Airborne Earth Science Workshop*.

Halldorsson, G.H., J.A. Benediktsson, and J.R. Sveinsson, 2003. Support vector machines in multisource classification, *Proceedings of the Geoscience and Remote Sensing Symposium, IGARSS '03*, IEEE International.

Hansen, M., R. Dubayah, and R. DeFries, 1996. Classification trees: An alternative to traditional land-cover classifiers, *International Journal of Remote Sensing*, 17:1075–1081.

Haykin, S.S., 1999. *Neural Networks: A Comprehensive Foundation*, Second edition, Upper Saddle River, New Jersey, Prentice Hall.

Huang, C., L.S. Davis, and J.R. Townshend, 2002. An assessment of support vector machines for land-cover classification, *International Journal of Remote Sensing*, 23(4):725–749.

Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Second edition, Upper Saddle River, New Jersey, Prentice Hall.

Key, J.R., J.A. Maslanik, and R.G. Barry, 1989. Cloud classification from satellite data using a fuzzy sets algorithm: A polar example, *International Journal of Remote Sensing*, 10:1823–1842.

Light, W.A., 1992. Ridge functions, sigmoidal functions and neural networks, *Approximation Theory VII* (E.W. Cheney, C.K. Chui, and L.L. Schumaker, editors), Boston, Academic Press, pp. 163–206.

Mangasarian, O.L., and D.R. Musicant, 2001. Lagrangian support vector machines, *Journal of Machine Learning Research*, 1:161–177.

Mather, P.M., 1999. *Computer Processing of Remotely-sensed Images: An Introduction*, Second edition, Chichester, New York, Wiley.

Melgani, F., and L. Bruzzone, 2004. Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790.

Pal, M., and P.M. Mather, 2005. Support vector machines for classification in remote sensing, *International Journal of Remote Sensing*, 26(5):1007–1011.

Paola, J.D., and R.A. Schowengerdt, 1995. A review and analysis of backpropagation neural networks for classification of remote sensed multi-spectral imagery, *International Journal of Remote Sensing*, 16:3033–3058.

Paola, J.D., and R.A. Schowengerdt, 1997. The effect of neural network structure on a multispectral land-use/land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 63:535–544.

Peddle, D.R., G.M. Foody, A. Zhang, S.E. Franklin, and E.F. LeDrew, 1994. Multisource image classification II: An empirical comparison of evidential reasoning and neural network approaches, *Canadian Journal of Remote Sensing*, 20:396–407.

Pohl, C., and J.L. van Genderen, 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications, *International Journal of Remote Sensing*, 19(5):823–854.

Powell, M.J.D., 1985. Radial basis functions for multivariable interpolation: A review, *Proceedings of the IMA Conference on Algorithms for the Approximation of the Functions and Data*, RMCS, Shrivenham, England.

Powell, M.J.D., 1988. Radial basis function approximations to polynomials, *Proceedings of the Numerical Analysis 1987*, Dundee, U.K.

Powell, M.J.D., 1992. The theory of radial basis function approximation in 1990, *Advances in Numerical Analysis, Volume II: Wavelets, Subdivision Algorithms, and Radial Basis Functions* (W. Light, editor), Oxford, Oxford Science Publications, pp. 105–210.

Quinlan, J.R., 1993. *C4.5: Programs for Machine Learning*, San Mateo, California, Morgan Kaufmann Publishers.

Rao, R.M., and M.K. Arora, 2004. Overview of image processing, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data* (P.K. Varshney and M.K. Arora, editors), Springer-Verlag, pp. 51–85.

Richards, J.A., and X. Jia, 1999. *Remote Sensing Digital Image Analysis: An Introduction*, Third edition, Berlin, New York, Springer.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A measure of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.

Safavian, S.R., and D. Landgrebe, 1991. A survey of decision tree classifier methodology, *IEEE Transactions on Systems, Man, and Cybernetics*, 21:660–674.

Saha, A.K., M.K. Arora, E. Csaplovics, and R.P. Gupta, 2005. Land cover classification using IRS LISS III image and DEM in a rugged terrain: A case study in Himalayas, *GeoCarto International*.

Schapire, R.E., 1999. A brief introduction to boosting, *Proceedings of the 16th International Joint Conference on Artificial Intelligence*.

Simone, G., A. Farina, F.C. Morabito, S.B. Serpico, and L. Bruzzone, 2002. Image fusion techniques for remote sensing applications, *Information Fusion*, 3(1):3–15.

Srinivasan, A., and J.A. Richards, 1990. Knowledge-based techniques for multi-source classification, *International Journal of Remote Sensing*, 11(3):505–525.

Tso, B., and P.M. Mather, 2001. *Classification Methods for Remotely Sensed Data*, London, New York, Taylor & Francis.

Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*, New York, Springer.

Watanachaturaporn, P., and M.K. Arora, 2004. Support vector machines for classification of multi- and hyperspectral data *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data* (P.K. Varshney and M.K. Arora, editors), Springer-Verlag, pp. 237–255.

Watanachaturaporn, P., M.K. Arora, and P.K. Varshney, 2004. Evaluation of factors affecting support vector machines for hyperspectral classification, *Proceedings of the ASPRS 2004 Annual Conference*, 23–28 May, Denver, Colorado, unpaginated CD-ROM.

Watanachaturaporn, P., M.K. Arora, and P.K. Varshney, 2005. Hyperspectral image classification using support vector machines: A comparison with decision tree and neural network classifiers, *Proceedings of the ASPRS 2005 Annual Conference*, Baltimore, Maryland, unpaginated CD-ROM.

Wolpert, D.H., 1992. Stacked generalization, *Neural Networks*, 5(2):241–260.