

HYPERSENSPECTRAL IMAGE CLASSIFICATION USING SUPPORT VECTOR MACHINES: A COMPARISON WITH DECISION TREE AND NEURAL NETWORK CLASSIFIERS

Pakorn Watanachaturaporn¹

Manoj K. Arora²

Pramod K. Varshney¹

¹ Electrical Engineering and Computer Science Department, Syracuse University, Syracuse, NY 13244
{pwatanac, varshney}@ecs.syr.edu

² Department of Civil Engineering, IIT Roorkee, Roorkee, 247 667, India
manojfce@iitr.ernet.in

ABSTRACT

Recently, support vector machines (SVMs) have proven to have great potential for classification of remotely sensed hyperspectral data acquired in a large number of spectral bands. Due to Hughes' phenomenon, conventional parametric classifiers fail to classify such a high dimensional dataset. In the past, neural networks and decision tree classifiers, which are nonparametric in nature, have frequently been used for classification of multispectral remote sensing data. These, however, have marked limitations when applied to hyperspectral data. In this paper, we present the results of applying SVMs for a 16-class classification of an AVIRIS image and compare its performance with decision tree, back propagation (BP) and radial basis function (RBF) neural network classifiers. There are a number of parameters that may affect the accuracy of SVM based classifiers. The best values of these parameters have been selected on the basis of a set of hypotheses and experiments. All the SVM classifications have been performed using an in-house code developed in a Matlab environment. The Kappa coefficient of agreement has been used to assess the accuracy of classification. The differences in classification accuracy have been statistically evaluated using a pairwise Z-test. SVM classification using a polynomial kernel of degree 2 produced an accuracy of 96.94% whereas the accuracies achieved by decision tree, BP and RBF neural network classifiers were 74.75%, 38.03% and 95.30% respectively. This clearly illustrates that the accuracy of the SVM classifier is significantly higher than decision tree and BP neural network classifiers at 95% confidence level. Although, the improvement in SVM classification accuracy with respect to the RBF neural network classifier was not statistically significant, the SVM classifier was more computationally efficient than the RBF classifier for hyperspectral image classification.

INTRODUCTION

Extracting land cover information from remote sensing images is one of the key applications of remote sensing research (Campbell, 1996; Sellers et al., 1995). A great deal of research has focused on the development of advanced classification algorithms. Conventionally, parametric classifiers such as the minimum distance to means classifier and the maximum likelihood classifier have been used. However, these classifiers fail to deliver high accuracies for classifying hyperspectral images (Campbell, 1996; Jensen, 1996; Landgrebe, 2002; Richards & Jia, 1999). This may be due to the fact that parametric classifiers are based on the assumption that data follow a statistical distribution to enable an accurate estimation of the parameters. This in turn depends on the selection of a sufficient number of pure training samples, which is related to the number of bands being used in the classification process. Larger the number of bands, more is the requirement of training sample size. Otherwise, the classifiers may suffer from the problem of the *curse of dimensionality*, also referred to as the *Hughes' phenomenon* (Hughes, 1968; Landgrebe, 2002). Because of the limitations of parametric classifiers, non-parametric classifiers such as neural networks and decision trees have often been implemented to classify high dimensional hyperspectral data. In fact, neural network classifiers, particularly the one based on the back propagation algorithm, have been viewed as a substitute for the most widely used MLC. The classification accuracy achieved by non-parametric classifiers has often been found to be higher than that achieved by the conventional classifiers (Li et al., 2003; Pal & Mather, 2003; Peddle et al., 1994; Rogan et al., 2002).

Recently, support vector machines (SVMs), another non-parametric classifier, has found acceptance in the remote sensing community since they can construct class decision boundaries efficiently by directly extracting

information from the training samples. The major attraction of the SVM based classifier is that it requires relatively small training sample size than the conventional classifiers.

SVMs have originated from statistical learning theory pioneered by Boser et al. (1992) in the middle of the 1990's. Statistical learning theory establishes bounds on the error rate of a learning machine on unseen data, called the generalization error rate. These bounds are a function of the sum of the training error rate and the terms that measure classifier complexity. To minimize the bounds on the generalization error rate, both the sum of the training error rate and the classifier complexity must be minimized. Vapnik (1998) showed that the bounds on the generalization error rate can be minimized by explicitly maximizing the margin of separation. Consequently, a better classification performance on unseen data can be expected, and thus high generalization can be achieved. Moreover, since the margin of separation is not dependent on the dimensionality of the data, accurate classification from high dimensional data is possible.

SVMs aim to maximize the margin between two classes of interest and place a linear separating hyperplane between them. Moreover, SVMs can adapt themselves and operate like a nonlinear classifier in the original input space by simply mapping the data into a higher dimensional feature space that spreads the data out. As SVMs can adequately classify data in a higher dimensional feature space with a limited number of training samples, they overcome the Hughes' Phenomenon.

In this paper, our goal is to compare the results of hyperspectral classification from an SVM with those obtained from the back propagation neural network, radial basis function network, and decision tree classifiers. The results have been evaluated visually and quantitatively through Kappa coefficient. The difference in accuracy of classifiers has been statistically evaluated using a pairwise Z test. In the next section, we briefly introduce the SVM algorithm. We then describe the hyperspectral dataset, experiments and classification methods. This is followed by results, discussion and conclusions.

SUPPORT VECTOR MACHINES (SVMs)

The construction of SVMs has been described in many publications (see (Boser et al., 1992; Cortes & Vapnik, 1995; Gualtieri et al., 1999; Gualtieri & Crompton, 1998; Vapnik, 1995, 1998; Watanachaturaporn & Arora, 2004; Watanachaturaporn et al., 2004)). Consider a binary classification problem, where the dataset is partitioned into two classes with a linear hyperplane separating them. Assume that the training dataset consists of k training samples represented by $\{x_1, y_1, \dots, x_k, y_k\}$, where $x_i \in \mathbb{R}^N$ 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 SVMs is to find a linear decision function defined by $f(x) = w^T x / b$, where $w \in \mathbb{R}^N$ determines the orientation of a discriminating hyperplane, and $b \in \mathbb{R}$ is a bias. The hyperplanes for the two classes are, therefore, represented by $y_i = w^T x / b$. Due to the noise or mixture of classes during the selection of training samples, variables $T_i \geq 0$, called slack variables, are introduced to account for the effects of misclassification. The hyperplanes for the two classes become $y_i = w^T x / b + \psi \pm T_i$. The optimal hyperplane (i.e., $f(x) = 0$) is located where the margin between the two classes of interest is maximized and the error is minimized. It can be obtained by solving the following constrained optimization problem,

$$\begin{aligned} \text{Minimize} : & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k T_i, \\ \text{Subject to} : & y_i = w^T x / b + \psi \pm T_i, \quad \text{for } i = 1, 2, \dots, k. \end{aligned} \quad (1)$$

The constant $0 \leq C \leq \infty$, called the penalty value or C value is a regularization parameter. It defines the trade-off between the number of misclassifications in the training samples and the maximization of the margin. In practice, the penalty value is selected by trial and error.

The constrained optimization problem in (1) is solved by the method of Lagrange multipliers. The equivalent optimization problem becomes,

$$\begin{aligned} \text{Maximize} : & \sum_{i=1}^k \xi_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \xi_i \xi_j y_i y_j x_i^T x_j, \\ \text{Subject to} : & \sum_{i=1}^k \xi_i y_i \geq 0 \text{ and } 0 \leq \xi_i \leq C, \quad \text{for } i = 1, 2, \dots, k. \end{aligned} \quad (2)$$

where $\alpha_i \geq 0$ are the Lagrange multipliers. The solution of the optimization problem given in (2) is obtained in terms of the Lagrange multipliers α_i . According to the Karush-Kuhn-Tucker (KKT) optimality condition (Fletcher, 1987), some of the multipliers will be zero. The multipliers that have nonzero values are called the support vectors. The results from an optimizer, called an optimal solution, form the set $\alpha = [\alpha_1, \dots, \alpha_k]$. The value of w and b are calculated from $w^o = \sum_{i=1}^k y_i \alpha_i x_i$ and $b^o = \frac{1}{2} (w^o \cdot x_{+1}^o - w^o \cdot x_{-1}^o)$, where x_{+1}^o and x_{-1}^o are the support vectors of class labels $+1$ and -1 respectively. The decision rule is then applied to classify the dataset into two classes viz. $+1$ and -1 ,

$$f(x) = \text{sign} \left(\sum_{\substack{i \\ \text{support vector}}} y_i \alpha_i x_i - b^o \right) \quad (3)$$

There are instances where a linear hyperplane cannot separate classes without misclassification; however, those classes are separated by a nonlinear separating hyperplane. In this case, data may be mapped to a higher dimensional space with a nonlinear transformation function. In the higher dimensional space, data are spread out (Cover, 1965), and a linear separating hyperplane may be found.

Let a nonlinear transformation function ϕ map the data into a higher dimensional space. Suppose there exists a function K , called a kernel function, such that,

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j). \quad (4)$$

A kernel function may be substituted for the dot product of the transformed vectors. The optimization problem then becomes,

$$\begin{aligned} \text{Maximize : } & \sum_{i=1}^k \alpha_i \frac{1}{2} \sum_{j=1}^k \alpha_j y_i y_j K(x_i, x_j), \\ \text{Subject to : } & \sum_{i=1}^k \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C, \quad \text{for } i = 1, 2, \dots, k. \end{aligned} \quad (5)$$

The decision function becomes,

$$f(x) = \text{sign} \left(\sum_{\substack{i \\ \text{support vector}}} y_i \alpha_i K(x_i, x) - b^o \right) \quad (6)$$

Multiclass Classification

SVMs were originally developed to perform 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 the binary SVMs have been proposed. Details of multiclass classification methods with their merits and demerits can be found in (Varshney & Arora, 2004). Pairwise classification (Krebel, 1999) method has been used here for various SVM based classification experiments. In this method, we create SVM classifiers for all the possible pairs of classes. Therefore, for the M -class problem, there will be $M(M-1)/2$ binary classifiers. The output from each classifier is obtained in the form of a class label. The class label that occurs the most is assigned to that pixel. A tie-breaking strategy is adopted in case of a tie. A common tie-breaking strategy is to randomly select one of the class labels that are tied.

EXPERIMENTAL DATA AND CLASSIFICATION METHODS

The hyperspectral image used in the experiments is acquired from the AVIRIS sensor available at (<ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C>). One of the bands of this image is shown in Figure 1. The image was acquired on June 12, 1992 over the northern part of Indiana. Two-thirds of the scene is covered with agricultural land while one-third is forest and others. Two major dual lane highways, a smaller road, a rail line, low density housing, and other building structures can also be seen in the scene. A field-surveyed map consisting of sixteen classes and one unclassified class is also available (Figure 2) and has been used as reference data (ground data) for training and testing data creation and accuracy assessment. The availability of reference data makes this

hyperspectral image an excellent source for conducting experimental studies, and, therefore, this image had been used in many other earlier studies (e.g., (Gualtieri et al., 1999; Gualtieri & Cromp, 1998; Melgani & Bruzzone, 2002; Tadjudin & Landgrebe, 1998a, 1998b; Watanachaturaporn & Arora, 2004; Watanachaturaporn et al., 2004)). The size of the 224 bands image is 145×145 pixels. Four of the bands do not contain any data. In our experiments, similar to the other studies, twenty water absorption bands, numbered [104 – 108], [150 – 163] and 220, were removed from the original image. In addition, fifteen noisy bands [1 – 3], 103, [109 – 112], [148 – 149], [164 – 165] and [217 – 219], as observed from visual inspection, were also discarded. A pure pixel database was created from which a number of training and testing samples (pixels) for each class were randomly selected and are shown in Table 1.



Figure 1. A single band AVIRIS image (band 100 [1.3075-1.3172 μm])



Figure 2. Reference image

Table 1. The number of training and testing pixels from the AVIRIS image

Class	Training pixels	Testing pixels
Alfalfa (C1)	7	6
Corn-notill (C2)	117	117
Corn-min (C3)	70	70
Corn (C4)	30	30
Grass/pasture (C5)	47	47
Grass/trees (C6)	58	58
Grass/pasture-mowed (C7)	9	9
Hay-windrowed (C8)	80	80
Oats (C9)	13	13
Soy-notill (C10)	115	116
Soy-mintill (C11)	146	145
Soy-clean (C12)	45	46
Wheat (C13)	25	24
Woods (C14)	104	104
Bldg-grass-trees-drives (C15)	40	41
Stone-steel towers (C16)	10	9
Total	916	915

We compare the classification results obtained from the SVM based classifier with three other supervised classifiers namely back-propagation neural network (BPNN), radial basis networks, and decision tree classifier (DTC). The discriminating ability of neural networks has motivated a great deal of research effort. Many types of neural networks have been developed (Haykin, 1999; Lippman, 1987). The most widely used is a group of networks called a multi-layer perceptron (MLP) with back-propagation learning algorithm (BPNN) (cf. (Atkinson & Tatnall, 1997; Haykin, 1999; Paola & Schowengerdt, 1995, 1997)). A decision tree classifier (DTC) uses another approach for classification. It breaks a complex classification problem into multiple stages of simpler decision making processes (Safavian & Landgrebe, 1991). The DTC has frequently been used for the classification of multi-source data (Hansen et al., 1996; Huang et al., 2002). Radial basis function network (RBFNet) is a variant of neural networks, which views the design of neural network as a curve-fitting problem in a high-dimensional space instead of a stochastic approximation as in MLP. The discriminating power of the RBFNet has also drawn a lot of attention from the remote sensing community (Bruzzone & Prieto, 1999; Light, 1992; Melgani & Bruzzone, 2002; Powell,

1985, 1988, 1992). We have employed the BPNN and RBFNet available in the Matlab neural network toolbox and the DTC in the C4.5 Release 8 (Quinlan, 1993). LSVM (Mangasarian & Musicant, 2001) with pairwise classification method is employed for SVM based classification.

Since each classifier differs from others in their approach for class allocation and is dependent on a number of their own parameters, it may be impractical to compare different classifiers directly. Therefore, the best performance of each classifier, i.e. classifier with the set of parameters that resulted in the highest accuracy has been reported here. The classification accuracy was assessed using error matrix, overall accuracy, kappa coefficient, and Z statistical test (Congalton & Green, 1999; Rosenfield & Fitzpatrick-Lins, 1986).

RESULTS AND DISCUSSION

The classified images from the BPNN, RBFNet, DTC, and SVM based classifier with a polynomial of degree 2 and penalty value 1 are shown in Table 4 to Table 7. The areas in the reference image (Figure 2) and classified images (Figure 3 to Figure 6) where we do not have information are masked and are shown in black color.

By visual inspection of the classified images and the reference data, we can easily see that the BPNN classifier performs poorly on this dataset. The result from the DTC is better than the result from the BPNN classifier; however, misclassified areas appear as noise in the image. The classified images from the RBFNet and SVM based classifier appear similar and seem to show better classification than that obtained from the BPNN classifier and the DTC.

Quantitative evaluation of hyperspectral classification shows that the highest accuracy is achieved from the SVM based classifier (96.94%) even though the number of training samples is small (Table 4 - Table 7). However, at 95% confidence level this result is not statistically different from the accuracy of RBFNet classifier (95.30%). Nevertheless, the accuracy obtained from the SVM based classifier is significantly higher than that from the DTC (74.75%) and BPNN (38.03%) classifier (Table 3), at 95% confidence level. The kappa coefficients of agreement show a strong agreement between the classification result and the reference data only from the SVM and RBFNet classifiers. It shows a moderate agreement from the DTC classifier and poor agreement from the BPNN. For each individual classifier, the Z statistic values are higher than 1.96. Therefore, classifications are statistically different from random classifications at the 95% confidence level.

From the inspection of individual error matrices, it can be ascertained that most classifiers perform adequately when the number of training samples for a class were large. For example, the classes Corn-notill (Class 2), Hay-windrowed (Class 8), Soy-notill (Class 10), Soy-mintill (Class 11), and Woods (Class 14) are accurately classified by all the classifiers except the BPNN, which inaccurately classified Corn-notill (Class 2) and Soy-notill (Class 10) even though the training samples were large. When the number of training samples are small such as in case of classes Alfalfa (Class 1), Grass/pasture-mowed (Class 7), Oats (Class 9), and Stone-steel towers (Class 16), the SVM and the RBFNet still classify those accurately whereas the performance of DTC and the BPNN deteriorates in accurately classifying these classes. Thus, the SVM based classifier and the RBFNet classifier outperform DTC and BPNN classifiers for hyperspectral classification especially when the number of training samples is small.



Figure 3. Classified image from BPNN classifier

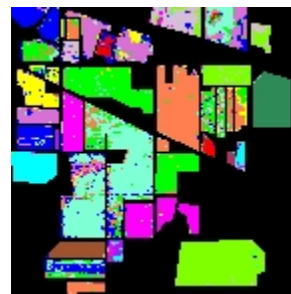


Figure 4. Classified image from RBFNet classifier

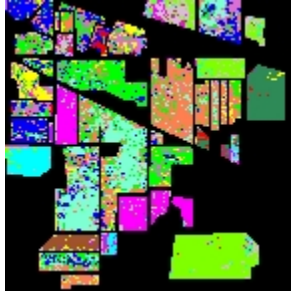


Figure 5. Classified image from DTC classifier

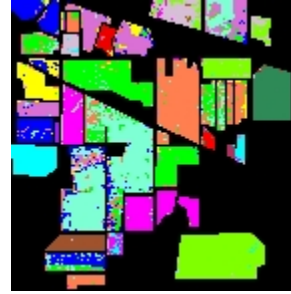


Figure 6. Classified image from SVM based classifier

Table 2. Accuracy of the hyperspectral classification from four different classifiers.

Classifier	Overall Accuracy (%)	Kappa Coefficient
BPNN	38.0328	0.2863
RBFNet	95.3005	0.9479
DTC	74.7541	0.7198
SVM	96.9399	0.9661

Table 3. Z values to indicate significant difference between kappa coefficients of hyperspectral classification obtained from any two classifiers. Values greater than 1.96 shows that the difference is significant at 95% level of confidence.

RBFNet	37.3618		
DTC	19.2594	12.8835	
SVM	39.7021	1.8199	14.3873
	BPNN	RBFNet	DTC

Table 4. Error matrix for classification obtained from the BPNN classifier. C1 to C16 are predicted classes as described in Table 1, and R1 to R16 are reference data corresponding to the classes C1 to C16.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	Row Sum
C1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C2	0	9	7	5	2	48	0	0	9	16	0	7	8	0	27	0	138
C3	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	4
C4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	6	2	5	10	4	1	9	80	2	0	2	2	0	0	1	0	124
C9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C10	0	13	25	1	0	0	0	0	2	11	2	4	0	0	0	0	58
C11	0	92	33	6	0	0	0	0	0	89	141	30	0	0	0	5	396
C12	0	1	0	4	0	0	0	0	0	0	0	3	0	0	0	0	8
C13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C14	0	0	0	0	41	9	0	0	0	0	0	0	16	104	13	4	187
C15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Column Sum	6	117	70	30	47	58	9	80	13	116	145	46	24	104	41	9	348

Table 5. Error matrix for classification obtained from the RBFNet classifier. C1 to C16 are predicted classes as described in Table 1, and R1 to R16 are reference data corresponding to the classes C1 to C16

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	Row Sum
C1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
C2	0	109	3	0	0	0	0	0	0	2	2	1	0	0	0	0	117
C3	0	0	62	2	0	0	0	0	0	0	4	1	0	0	3	1	73
C4	0	1	0	28	0	0	0	0	0	0	0	0	0	0	0	0	29
C5	0	0	0	0	47	0	0	0	0	0	0	0	0	0	0	0	47
C6	0	0	0	0	0	58	0	0	0	0	0	0	0	0	2	0	60
C7	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9
C8	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	80
C9	0	0	0	0	0	0	0	0	10	0	0	0	0	0	1	0	11
C10	0	4	0	0	0	0	0	0	0	113	1	0	0	0	0	0	118
C11	0	3	5	0	0	0	0	0	0	1	138	1	0	0	0	0	148
C12	0	0	0	0	0	0	0	0	0	0	0	43	0	0	0	1	44
C13	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	24
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	104	1	0	105
C15	0	0	0	0	0	0	0	0	3	0	0	0	0	0	34	0	37
C16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7
Column Sum	6	117	70	30	47	58	9	80	13	116	145	46	24	104	41	9	872

Table 6. Error matrix for classification obtained from DTC. C1 to C16 are predicted classes as described in Table 1, and R1 to R16 are reference data corresponding to the classes C1 to C16

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	Row Sum
C1	4	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	8
C2	0	70	8	8	0	0	0	0	0	9	16	10	0	0	0	1	122
C3	0	12	42	3	0	0	0	0	0	3	3	11	0	0	0	6	80
C4	0	5	2	13	0	0	0	4	0	0	0	5	0	0	1	0	30
C5	0	0	0	0	38	1	0	0	0	0	0	0	0	0	2	0	41
C6	0	0	0	0	0	55	0	0	1	0	0	0	0	0	8	0	64
C7	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	7
C8	2	0	0	0	0	0	2	74	0	0	0	0	0	0	0	0	78
C9	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	9
C10	0	17	6	1	0	0	0	0	0	86	8	3	0	0	0	0	121
C11	0	6	7	0	0	0	0	0	0	15	114	0	0	0	0	0	142
C12	0	7	5	5	0	0	0	0	0	0	4	17	0	0	0	0	41
C13	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	24
C14	0	0	0	0	4	0	0	0	0	0	0	0	0	102	3	0	109
C15	0	0	0	0	3	2	0	0	3	0	0	0	0	2	27	0	37
C16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
Column Sum	6	117	70	30	47	58	9	80	13	116	145	46	24	104	41	9	684

Table 7. Error matrix for classification obtained from the SVM classifier. C1 to C16 are predicted classes as described in Table 1, and R1 to R16 are reference data corresponding to the classes C1 to C16

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	Row Sum
C1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
C2	0	108	0	0	0	0	0	0	0	0	3	0	0	0	0	0	111
C3	0	0	68	0	0	0	0	0	0	0	5	1	0	0	0	0	74
C4	0	2	0	30	0	0	0	0	0	0	0	0	0	0	0	0	32
C5	0	0	0	0	47	0	0	0	0	0	0	0	0	0	0	0	47
C6	0	0	0	0	0	58	0	0	0	0	0	0	0	0	0	0	58
C7	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9
C8	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	80
C9	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	9
C10	0	2	0	0	0	0	0	0	0	115	0	0	0	0	0	1	118
C11	0	5	2	0	0	0	0	0	0	1	137	0	0	0	0	0	145
C12	0	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0	45
C13	0	0	0	0	0	0	0	0	0	0	0	0	24	0	1	0	25
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	103	0	0	103
C15	0	0	0	0	0	0	0	0	4	0	0	0	0	1	40	0	45
C16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	8
Column Sum	6	117	70	30	47	58	9	80	13	116	145	46	24	104	41	9	887

CONCLUSIONS

In this paper, we have compared the classification results of hyperspectral data obtained from the SVM based classifier with the BPNN, RBFNet, and DTC classifiers. We have qualitatively and quantitatively assessed the classification results by visual inspection and accuracy assessment including statistical testing respectively. By visual inspection, it is clear from the classified images that the results from the SVM and RBFNet are better than the results from the BPNN and DTC. This conclusion is confirmed by accuracy assessment which shows that the accuracy in the form of kappa coefficient of classification from the SVM and the RBFNet classifiers are significantly higher than that obtained from the BPNN and the DTC. Further, even though the difference in kappa coefficient of classification from the SVM and RBFNet is not statistically significant, the SVM achieved higher accuracy than RBFNet.

ACKNOWLEDGEMENTS

The work was supported by NASA under grant number NAG5 – 11227

REFERENCES

- Arora, M. K., & S. Mathur. (2001). Multi-source Classification using Artificial Neural Network in Rugged Terrain. *Geocarto International*, 16(3), 37-44.
- Arora, M. K., G. Narsimham, & B. Mohanty. (1998). Neural network approach for the classification of remotely sensed data. *Asian Pacific Remote Sensing and GIS Journal*, 10(2), 11-18.
- Arora, M. K., K. C. Tiwari, & B. Mohanty. (2000). Effect of neural network variables on image classification. *Asian Pacific Remote Sensing and GIS Journal*, 13, 1-11.
- Atkinson, P. M., & A. R. L. Tatnall. (1997). Neural network in remote sensing. *International Journal of Remote Sensing*, 18, 699-709.
- Boser, B. E., I. M. Guyon, & V. N. Vapnik. (1992). *A training algorithm for optimal margin classifiers*. In the Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, 27-29 July 1992, Pittsburgh, PA, USA.
- Bruzzzone, L., & D. F. Prieto. (1999). A technique for the selection of kernel function parameters in RBF neural networks for classification of remote sensing images. *IEEE Trans. Geosci. Remote Sensing*, 37, 1179-1184.
- Campbell, J. B. (1996). *Introduction to remote sensing* (2nd ed.). New York: Guilford Press.
- Congalton, R. G., & K. Green. (1999). *Assessing the accuracy of remotely sensed data: principles and practices*: CRC Press.
- Cortes, C., & V. Vapnik. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. on Electronic Computers*, EC-14, 326-334.
- Fletcher, R. (1987). *Practical methods of optimization* (2nd ed.). Chichester ; New York: Wiley.
- Gualtieri, J. A., S. R. Chettri, R. F. Crompt, & L. F. Johnson. (1999). *Support vector machine classifiers as applied to AVIRIS data*. In the Summaries of the Eighth JPL Airborne Earth Science Workshop.
- Gualtieri, J. A., & R. F. Crompt. (1998). *Support vector machines for hyperspectral remote sensing classification*. In the 27th AIPR Workshop, Advances in Computer Assisted Recognition, Proceeding of the SPIE.
- Hansen, M., R. Dubayah, & 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* (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.
- Huang, C., L. S. Davis, & J. R. Townshend. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725-749.
- Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognition. *IEEE Trans. Inform. Theory*, 14(55-63).
- Jensen, J. R. (1996). *Introductory digital image processing: a remote sensing perspective* (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.
- Krebel, U. (1999). Pairwise classification and support vector machines. In B. Schölkopf & C. Burges & A. J. Smola (Eds.), *Advances in kernel methods - support vector learning* (pp. 255-268). Cambridge, MA: The MIT Press.

- Landgrebe, D. (2002). Hyperspectral image data analysis. *IEEE Signal Process. Mag.*, 19, 17-28.
- Li, T. S., C. Y. Chen, & C. T. Su. (2003). Comparison of neural and statistical algorithms for supervised classification of multi-dimensional data. *Int. J. Indus. Eng. - Theory Appl. Pract.*, 10, 73-81.
- Light, W. A. (1992). Ridge functions, sigmoidal functions and neural networks. In E. W. Cheney & C. K. Chui & L. Schumaker (Eds.), *Approximation Theory VII* (pp. 163-206). Boston: Academic Press.
- Lippman, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4, 2-22.
- Mangasarian, O. L., & D. R. Musicant. (2001). Lagrangian support vector machines. *Journal of Machine Learning Research*, 1, 161-177.
- Melgani, F., & L. Bruzzone. (2002). *Support vector machines for classification of hyperspectral remote-sensing images*. In the Geoscience and Remote Sensing Symposium, IGARSS'02 IEEE International.
- Pal, M., & P. M. Mather. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86, 554-565.
- Paola, J. D., & 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., & R. A. Schowengerdt. (1997). The effect of neural network structure on a multispectral land-use/land cover classification. *Photogrammetric Engineering and Remote Sensing*, 63, 535-544.
- Peddle, D. R., G. M. Foody, A. Zhang, S. E. Franklin, & E. F. LeDrew. (1994). Multisource image classification II: An empirical comparison of evidential reasoning and neural network approaches. *Can. J. Remote Sens.*, 20, 396-407.
- Powell, M. J. D. (1985). *Radial basis functions for multivariable interpolation: A review*. In 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*. In the Numerical Analysis 1987 Proceedings, Dundee, UK.
- Powell, M. J. D. (1992). The theory of radial basis function approximation in 1990. In W. Light (Ed.), *Advances in Numerical Analysis Vol. II: Wavelets, Subdivision Algorithms, and Radial Basis Functions* (pp. 105-210). Oxford: Oxford Science Publications.
- Quinlan, J. R. (1993). *C4.5 : programs for machine learning*. San Mateo, Calif.: Morgan Kaufmann Publishers.
- Richards, J. A., & X. Jia. (1999). *Remote sensing digital image analysis: an introduction* (3rd ed.). Berlin; New York: Springer.
- Rogan, J., J. Franklin, & D. A. Roberts. (2002). A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80, 143-156.
- Rosenfield, G. H., & K. Fitzpatrick-Lins. (1986). A measure of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52, 223-227.
- Safavian, S. R., & D. Landgrebe. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21, 660-674.
- Sellers, P. J., B. W. Meeson, F. G. Hall, G. Asrar, R. E. Murphy, R. A. Schiffer, F. P. Bretherton, R. E. Dickinson, R. G. Ellingson, C. B. Field, K. F. Huemmrich, C. O. Justice, J. M. Melack, N. T. Roulet, D. S. Schimel, & P. D. Try. (1995). Remote sensing of the land surface for studies of global change: Models - algorithms - experiments. *Remote Sensing of Environment*, 51(3-26).
- Tadjudin, S., & D. Landgrebe. (1998a). *Classification of High Dimensional Data with Limited Training Samples*. Ph.D Thesis, Purdue University.
- Tadjudin, S., & D. Landgrebe. (1998b). *Covariance estimation for limited training samples*. In the International Geoscience and Remote Sensing Symposium, Seattle, WA.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.
- Varshney, P. K., & M. K. Arora. (2004). *Advanced image processing techniques for remotely sensed hyperspectral data*: Springer-Verlag.
- Watanachaturaporn, P., & M. K. Arora. (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.
- Watanachaturaporn, P., M. K. Arora, & P. K. Varshney. (2004, May 23-28, 2004). *Evaluation of factors affecting support vector machines for hyperspectral classification*. In the American Society for Photogrammetry & Remote Sensing (ASPRS) 2004 Annual Conference, Denver, CO.