

An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing

Thomas Oommen · Debasmita Misra ·
Navin K.C. Twarakavi · Anupma Prakash ·
Bhaskar Sahoo · Sukumar Bandopadhyay

Received: 19 September 2006 / Accepted: 3 October 2007
© International Association for Mathematical Geology 2008

Abstract Accurate thematic classification is one of the most commonly desired outputs from remote sensing images. Recent research efforts to improve the reliability and accuracy of image classification have led to the introduction of the Support Vector Classification (SVC) scheme. SVC is a new generation of supervised learning method based on the principle of statistical learning theory, which is designed to decrease uncertainty in the model structure and the fitness of data. We have presented a comparative analysis of SVC with the Maximum Likelihood Classification (MLC) method, which is the most popular conventional supervised classification technique. SVC is an optimization technique in which the classification accuracy heavily relies on identifying the optimal parameters. Using a case study, we verify a method to obtain these optimal parameters such that SVC can be applied efficiently. We use multispectral and hyperspectral images to develop thematic classes of known lithologic units in order to compare the classification accuracy of both the methods. We have varied the training to testing data proportions to assess the relative robustness and the optimal training sample requirement of both the methods to achieve compa-

T. Oommen
Department of Civil and Environmental Engineering, Tufts University, 200 College Avenue,
Medford, MA 02155, USA

D. Misra (✉) · B. Sahoo · S. Bandopadhyay
Department of Mining and Geological Engineering, University of Alaska-Fairbanks,
P.O. Box 755800, Fairbanks, AK 99775, USA
e-mail: ffdm1@uaf.edu

N.K.C. Twarakavi
Department of Environmental Sciences, University of California-Riverside, Riverside, CA 92521,
USA

A. Prakash
Geophysical Institute, University of Alaska Fairbanks, 903 Koyukuk Dr, Fairbanks, AK 99775-7320,
USA

erable levels of accuracy. The results of our study illustrated that SVC improved the classification accuracy, was robust and did not suffer from dimensionality issues such as the Hughes Effect.

Keywords Remote sensing · Support vector machines · Maximum likelihood · Multispectral · Hyperspectral classification

1 Introduction

Remote sensing images provide a wealth of data that need to be processed in order to extract information of interest. Image classification is a component of image processing that extracts thematic information by learning the relation between the spectral signatures and the various classes or themes that are of interest to the user. Earlier, image classification techniques relied on the visual interpretation of the user that often led to systematic errors because of excessive reliance on the skill and experience of the interpreter (Schowengerdt 1983; Ikeda and Dobson 1995; Lillesand et al. 2004). Digital Image Classification helped to provide greater objectivity to the image classification process. However, it suffered from various other limitations that have failed to provide the adequate level of accuracy in classification of the images. Therefore, improving the accuracy in classification has been a crucial component in contemporary research.

Digital classification of images has been performed using either the ‘unsupervised’ or the ‘supervised classification’ methods. Both methods use an automated quantitative decision making mechanism (Schowengerdt 1983). The numerical basis for classification of remote sensing image is that the spectral signature of each pixel corresponds uniquely to a particular theme covered by the image. In ‘unsupervised classification’, images are classified automatically into different spectral classes or themes based on the differentiability of the spectral signatures of each pixel. The weakness of this method, however, is that themes of high interest to the user may have subtle spectral difference; hence, the themes identified automatically would not be of high interest to the user in most cases (Hord 1982).

In order to overcome this weakness, ‘supervised classification’ utilizes representative pixels of a known theme for a section of the data (training data) to compile an interpretation key that statistically characterizes the spectral signature for each theme of interest to the user (Lillesand et al. 2004). Subsequently, this key is compared with the spectral information of each unknown pixel (testing data) and classified to the theme to which it resembles closely. As opposed to the unsupervised classification where the themes are chosen purely based on contrasts in spectral characteristics, this method relies on the knowledge and experience of the user in choosing the training data so that the classification can be accomplished efficiently and accurately. Despite the user intervention, classification errors may be introduced due to the limitation in the ability of the particular algorithm chosen to learn from a training data and classifying the testing data (limitation of the method), and failure in choosing a representative training data for the entire span of themes (limitation of choice).

Errors introduced due to the limitation of choice depend totally on the experience and knowledge of the user. However, errors introduced due to the limitation

of the method can be reduced significantly with the right choice of an algorithm or scheme. There are several conventional supervised classification schemes available (Murai 1996), out of which the most generally accepted classification method is the Maximum Likelihood Classifier (MLC). Despite its popularity, MLC suffers from several limitations that have been discussed by Murai (1996) and are outlined later. A relatively new scheme called the Support Vector Classification (SVC) has recently been introduced for supervised image classification (Huang et al. 2002; Zhu and Blumberg 2002). SVC has been found to achieve a higher level of accuracy than contemporary conventional methods of classification (Melgani and Bruzzone 2004; Foody and Mathur 2004; Pal and Mather 2005). In this study we provide an objective analysis of the effectiveness and accuracy obtained from SVC while comparing the relative ability of SVC and MLC to accurately classify an image. We provide a brief description of both SVC and MLC in the following section and specify the objectives of this research.

2 Description of Methods and Objectives

2.1 Maximum Likelihood Classification

MLC is based on the probability that every pixel in an image belongs to a particular class and that the probabilities are equal for all classes. A variation to this assumption is known as the Bayesian Decision Rule (BDR), where the probabilities of different classes need to be specified. However, it has been suggested not to use BDR unless a priori knowledge of the probabilities of different classes is available (Hord 1982; Schrader and Pouncey 1997). We have assumed the probabilities for each class to be equal for our analyses.

The equation for the MLC/ BDR is

$$D = \ln(a_c) - [0.5 \ln(|Cov_c|)] - [0.5(\mathbf{X} - M_C)^T (Cov_c^{-1})(\mathbf{X} - M_C)], \quad (1)$$

where D is the weighted distance (likelihood), c represents a particular class of interest, \mathbf{X} is the vector of spectral signature for the candidate pixel from the testing data, M_C is the mean vector of the sample of class c from the training data, a_c is the percent probability that any candidate pixel is a member of class c (assumed to be 1 or equal probability in this analysis), Cov_c is the covariance matrix of the pixels in the sample of class c from the training data, $|Cov_c|$ is the determinant of Cov_c , Cov_c^{-1} is the inverse of Cov_c , \ln is the natural logarithm function, and T is the transposition function (Schrader and Pouncey 1997). The candidate pixel is assigned to the class c which has the most likelihood (D) or the least weighted distance.

The strength of MLC is that it takes into account the variability within each class using the covariance matrix to classify the candidate pixel. However, Murai (1996) discussed several disadvantages of the method. He found that the method relies heavily on the assumption that the data in each input band is normally distributed, and there has to be sufficient training pixels sampled to allow representative estimation of the mean vector and the variance–covariance matrix. From our analysis, the lack of sufficient training sample tends to misfit themes that have small spectral differences

and the inverse of variance–covariance matrix becomes unstable if there is a very high correlation between any two bands in the training data.

2.2 Support Vector Classification

SVC falls in the family of a multi-purpose scheme called the Support Vector Machines (SVM). SVM is a new generation of supervised learning system that is based on the principle of statistical learning theory (Vapnik 1995). For classification, SVMs operate by finding a hyper-plane in the space of possible inputs. The basic approach in support vector machines is to identify a hyper-plane that produces optimal separation between the two classes (Vapnik 1995). Usually in SVM, the hyper-plane is developed using a subset of the data called the training data set and the generalizing ability of the developed hyper-plane is validated using an independent subset called testing data set. To classify a data of N dimensions, a $(N - 1)$ dimensional hyper-plane is developed in SVC. Considering a binary case of linearly separable data (Fig. 1), it is observed that there can be an infinite number of hyper-planes that can separate this data. However, there is only one hyper-plane that has the maximum margin. This is the optimum hyper-plane and the vectors (points) that constrain the width of the margin are the support vectors (Sherrod 2003).

From Fig. 1B, it may be observed that the support vectors lie on two hyper-planes that are parallel to the optimal hyper-plane. These are defined by the function, $\omega \mathbf{x}_i + b = \pm 1$, where \mathbf{x} is a point on the hyper-plane, ω is normal to the hyper-plane, and b is the bias. Therefore, the margin between these planes is $2/\|\omega\|$ (Foody and Mathur 2004). The objective of maximizing the margin yields to a constrained optimization problem given by

$$\min \left[\frac{1}{2} \|\omega\|^2 \right], \quad (2)$$

subject to the constraints

$$y_i(\omega \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \text{and} \quad y_i \in \{1, -1\}. \quad (3)$$

However, most classification problems are not linearly separable (Fig. 2).

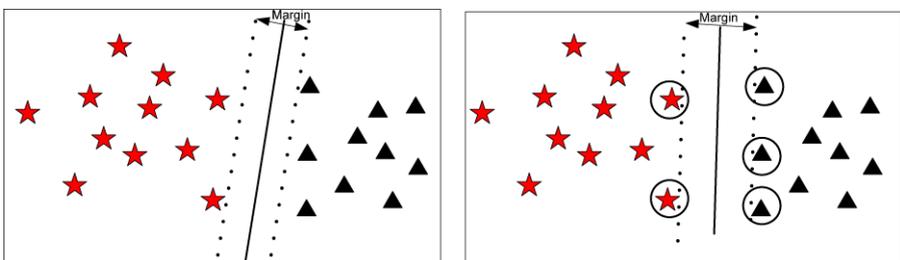


Fig. 1 Binary case of linearly separable data: **A** Due to the fact that the margin is not maximum, the hyper-plane is not optimal; **B** Optimal hyper-plane with maximum margin, the support vector is encircled and it lies on the planes that constrain the width of the margin

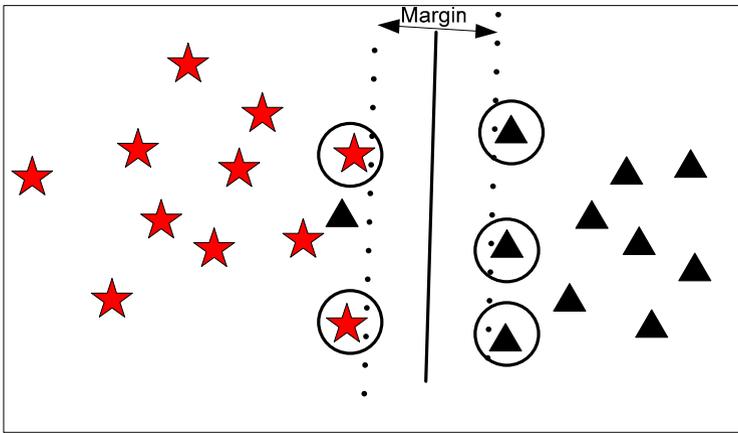


Fig. 2 Linearly nonseparable dataset

In order to deal with these situations, the data is mapped into a higher dimensional space (feature space). Although traditionally it is achieved by the construction of non-linear transformations, this suffers from the curse of dimensionality (Hughes Effect). In SVM a kernel method is used which enables substituting the nonlinear transformation by an inner product that can be defined by a kernel function ϕ (Gunn 1998). The inner product does not require evaluation in the feature space and thus helps to address the dimensionality issues (Gunn 1998). The kernel function enables the data points to spread in such a way that a linear hyper-plane can be fitted (Foody and Mathur 2004). Hence, the optimization problem for maximizing the margin becomes a combination of two criteria, the margin maximization and the error minimization. It is defined as

$$\min \left[\frac{\|\omega\|^2}{2} + C \sum_{i=1}^r \xi_i \right], \tag{4}$$

subject to the constraints

$$y_i (\omega \cdot \phi(\mathbf{x}_i) + b) - 1 \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, i = 1, \dots, N, \tag{5}$$

where C is a penalty term that controls the magnitude of penalty associated with the training samples classified on the wrong side of the hyper-plane. It is C that allows striking a balance between the two competing criteria of margin maximization and error minimization, whereas the slack variables ξ_i indicate the distance of the incorrectly classified points from the optimal hyper-plane (Foody and Mathur 2004).

The *e1071* package of the R programming language version 1.5-13 has linear, polynomial, radial basis and sigmoid kernel functions that can be used in SVM models. A linear kernel is considered to be a special case of radial basis kernel (Keerthi and Lin 2003). Also, Keerthi and Lin (2003) illustrated that the sigmoid kernel behaves like a radial basis kernel for certain parameters. Studies have also illustrated that a radial basis kernel yields the best results in remote sensing applications (Melgani and Bruzzone 2004; Foody and Mathur 2004;

Pal and Mather 2005). We chose to use the radial basis kernel for SVC in this study. The verification of the applicability of other specialized kernel functions for the classification of remote sensing data may be used in future studies. The equation for the radial basis kernel is

$$\phi = e^{-\gamma \|(\mathbf{x}-\mathbf{x}_i)\|^2}, \quad (6)$$

where γ is the width of the Gaussian kernel. The strengths and shortcomings of SVC, being a relatively recent method, have not been adequately explored. Therefore, in this study we have made a comparative analysis of MLC and SVC to classify remote sensing images.

2.3 Objectives of our Study

SVC has been found to provide better classification accuracy than other conventional supervised classification schemes. It has been noted that the accuracy of the classification using SVC largely depends on the parameters C and γ . In this paper, we illustrate the methodology for SVC classification scheme and compare the classification accuracy of SVC to MLC using multispectral and hyperspectral images.

Hyperspectral imaging is a fast growing field of remote sensing. It takes advantage of several contiguous spectral channels to uncover materials that usually cannot be resolved by multispectral sensors (Chang 2003). In the case of hyperspectral data, the classification accuracy using statistical methods is considered to reduce (due to large number of bands) (Hughes 1968; Pal and Mather 2005). This is known as the Hughes Effect/Curse of Dimensionality (Hughes 1968), which states that the increasing dimensionality decreases the reliability of the estimates of statistical parameters required to compute the probabilities. Foody et al. (1995) stated that an artificial neural network is less susceptible to these effects. We hypothesize that SVC will be robust and will not be susceptible to the Hughes Effect due to the introduction of kernel functions. We tested our hypothesis using hyperspectral image data and made a comparative analysis using both SVC and MLC. Besides testing for sensitivity to the Hughes Effect, we also verified the sensitivity to classification accuracy of SVC and MLC using different training data sizes and subset band combinations using the hyperspectral data. The robustness of these methods to classify data with small training samples (sparse data issues) and the improvement in classification accuracy by using a hyperspectral image over the multispectral image of the same area was also tested.

Therefore, the objectives of this study are:

- To illustrate the methodology for the SVC classification scheme.
- To compare the classification accuracy using SVC and MLC for multispectral images.
- To verify the susceptibility of both SVC and MLC to the Hughes Effect using hyperspectral data.
- To study the sensitivity to classification accuracy obtained from both SVC and MLC to training data size, subset band combinations and sparse data issues using hyperspectral data.
- To study the improvement in classification accuracy of the methods using hyperspectral data as compared to multispectral image data of the same area of interest.

3 Data and Methodology Used

In this paper, two multispectral and one hyperspectral images have been analyzed. The images were chosen from two locations. A hyperspectral and a multispectral image were obtained from Cuprite, Nevada. Cuprite is one of the alteration sites explored for precious metals. It is a mineralogical site that has been established as a reference site for remote sensing instruments (Barry et al. 2003). A multispectral image was also obtained from Goodnews Bay area in southwest Alaska. Goodnews Bay and vicinity is an area of economic interest for platinum and gold prospecting for which a detailed geologic map is available from Mertie (1940).

3.1 Multispectral Data

The multispectral image from Goodnews Bay is a subset of Landsat 7 Enhanced Thematic Mapper (ETM) image of September 27th, 2000. Six reflective bands of the image were used for our analysis. These were blue, green, red, near infrared (NIR), and the shortwave infrared [(SWIR-1) and (SWIR-2)] spectral regions. The width of the image was 166 pixels with a height of 84 pixels comprising a total of 13944 pixels. The lithological classes, to aid our classification, were obtained from the geologic map of Goodnews Bay (Mertie 1940). Using the geologic map, five lithological units were identified on the Landsat image spanning over 3995 pixels. This data was used for the training and testing purposes of the supervised classification schemes.

The 3995 pixels were randomized three times to avoid any bias in our analyses. These three randomized datasets have been referred to as Landsat GNB Trial-1, Trial-2, and Trial-3 in the following discussion (see Table 1). The objectives for developing these three randomized datasets were to ensure that there was no bias introduced in the selection of testing and training data, to verify the robustness of the classification technique, and to ensure that the classification results were not artifacts of the training data selection method. The classification of datasets using both SVC and MLC was repeated on all three datasets using different percentages of testing and training data.

Table 1 The summary of the overall classification accuracies obtained using the SVC and MLC methods on the multispectral data

Data and method properties	Minimum accuracy (Min)	Maximum accuracy (Max)	Mean accuracy	Max–Min
Landsat GNB Trial-1, 2, and 3 (SVC)	89.42%	96.87%	91.38%	7.45%
Landsat GNB Trial-1, 2, and 3 (MLC)	73.01%	89.02%	78.45%	16.01%
Landsat CUP Trial-1, 2, and 3 (SVC)	53.62%	65.95%	60.22%	12.33%
Landsat CUP Trial-1, 2, and 3 (MLC)	47.87%	61.35%	55.01%	13.48%

The second multispectral image chosen was a Landsat 7 ETM of July 20th, 2000, from Cuprite mining district of West-Central Nevada. Similar to the Goodnews Bay multispectral image, all 6 bands were used for the classification. The width of the image was 209 pixels with a height of 27 pixels resulting in a grand total of 5643 pixels. To aid our analysis, the lithological classes were obtained from the lithological map of Nevada developed by Stewart and Carlson (1978). Using the lithological map, five units were identified on the image. The Cuprite, Nevada, data was also randomized three times to obtain three datasets, which have been referred to as Landsat CUP Trial-1, Trial-2, and Trial-3 in the following discussion (Table 1).

3.2 Hyperspectral Data

The Hyperspectral image used for this analysis was a subset of the Hyperion EO-1 image of Cuprite mining district of West-Central Nevada, obtained on March 4th, 2002. This image had 220 unique spectral channels with a complete spectrum covering from 357–2576 nm (Barry et al. 2003). Of these 220 channels, 196 are spectrally unique and calibrated. These 196 unique bands were used for our study. The lithological units for our analysis were obtained from the same geologic map of Nevada by Stewart and Carlson (1978). From this map, five lithological units were identified on the image comprising 5643 pixels and this data was used for the training and testing purposes with our classification schemes.

The algorithms used in this study are the SVM algorithm available in the package *e1071* version 1.5-13 of the R programming language (Meyer 2001) and the MLC algorithm that was developed in MatlabTM 7.1 using (1). In this paper, overall classification accuracy was used as the measure of the performance of a classification technique. The overall classification accuracy is the total percentage of correct classifications of all themes. It was obtained from the confusion matrix (Kohavi and Provost 1998). The following methodology recommended by Hsu et al. (2003) was used for the SVC.

1. The data was arranged in a matrix form. Each row represented a particular pixel of the image with the different columns representing reflectance from the different bands, and the last column corresponded to the lithological class.
2. The data was randomly divided into subsets of training and testing data.
3. The data was scaled to prevent columns of greater numeric ranges dominating over columns of smaller numeric ranges.
4. The radial basis kernel function was used. The two parameters, C and γ , inherent to a radial basis function were obtained by a grid search using cross validation. The cross validation overcomes the overfitting problems. In this study we used a five-fold cross validation. The grid search was done in two steps. Initially a coarser grid was applied with an exponentially growing sequence of $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$. An example of the coarser grid search obtained for the SVC Trial-1 on the Goodnews Bay data is shown in Fig. 3. The coarser grid search was followed by a finer grid search in the area (values of C and γ) and produced fewer measures of error (Fig. 4). However, this approach can be time-consuming in the case of large datasets; therefore, Keerthi and Lin (2003) recommend searching the values of C and γ only for conditions that satisfy the expression $\log \gamma = \log C - \log \tilde{C}$, where \tilde{C} is the optimal C for an SVC using a linear

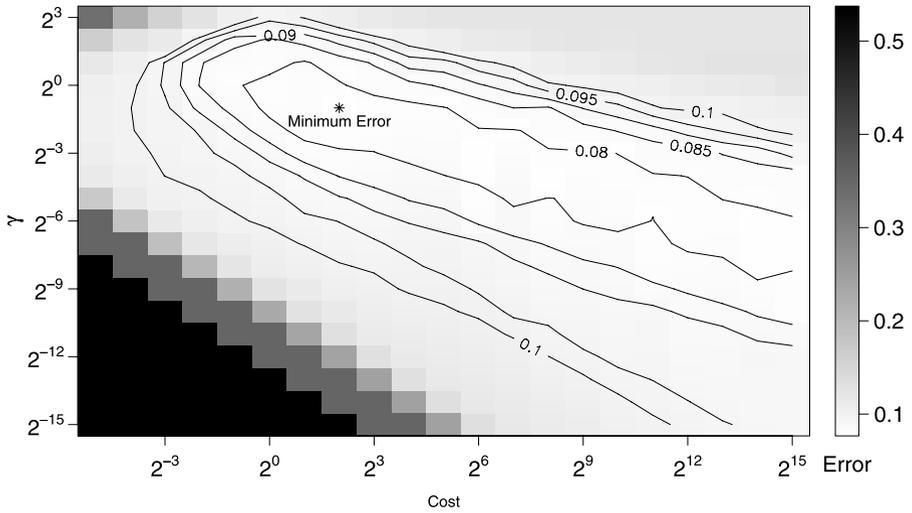


Fig. 3 The coarser grid search with an exponentially growing sequence of $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ for the SVC Trial-1 on the Goodnews Bay data

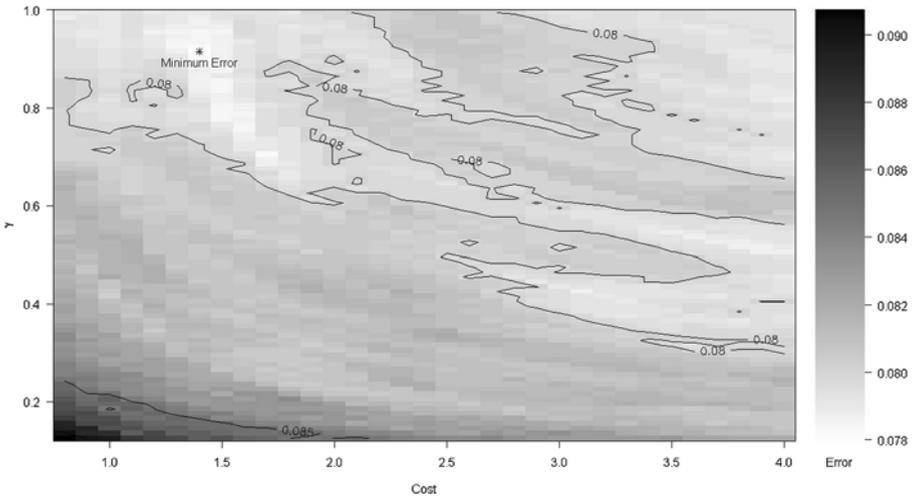


Fig. 4 The finer grid search in the area (values of C and γ) that produced the fewest measures of error from Fig. 3 in order to identify the optimal values of C and γ for the model

- kernel. The optimal C for an SVC using a linear kernel can be determined by a simple grid search. A plot of the C and the measure of error for an SVC using a linear kernel illustrates that, once C reaches an optimal value, the variation in the measure of error with increasing values of C is negligible.
5. The optimal values of C and γ from the previous step were used to train the entire training data set.
 6. Finally, the developed SVC model was tested on an independent testing dataset.

4 Results and Discussion

4.1 Comparison of the Performance of SVC and MLC Using Multispectral Data

In order to compare the classification accuracy of SVC and MLC for multispectral images, both the Goodnews Bay data and the Cuprite data were used. The overall accuracy using the SVC and MLC were calculated on both the data for several training versus testing data ratios. These ratios ranged from 5% to 96% of the overall data comprising the training data subset. The overall classification accuracy on the Goodnews Bay data is plotted in Fig. 5 and that of the Cuprite data is plotted in Fig. 6. It is evident from Figs. 5 and 6 that the overall classification accuracy achieved using SVC is higher than using the MLC.

4.2 Comparison of the performance of SVC and MLC Using Hyperspectral Data

The Hyperion EO-1 hyperspectral data has 196 unique bands. As discussed earlier, the high dimensionality (196 bands) is considered to reduce the classification accuracy using statistical methods (Hughes Effect). In order to test our hypothesis that SVC is not affected by the Hughes Effect, classification was carried out for a different number of bands and training data sizes using the Hyperion EO-1 hyperspectral data. The different band combinations were obtained by choosing the alternate bands from the previous trial. The first trial had 196 bands; the second trial had 98 bands; subsequent trials had 49, 25, 13, and 7 bands. The reason for selecting the alternate bands was to ensure that even the smallest band combination (7 bands in this case) included representative spectra from the entire range of the original data (196 bands).

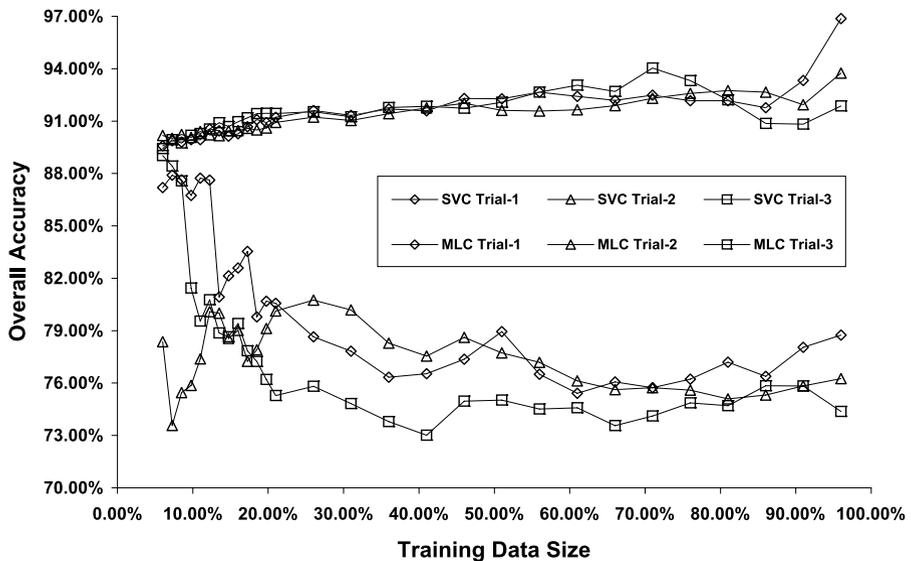


Fig. 5 Results of SVC and MLC on the testing data of Landsat Goodnews Bay image for three random trials. The training data size range from 6% to 96%

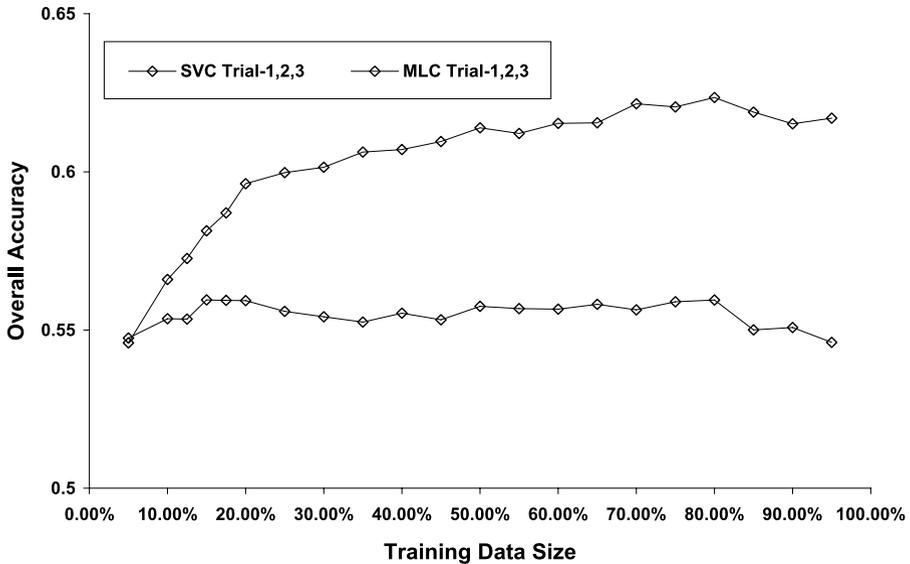


Fig. 6 The mean of the three random trials of SVC and MLC on the testing data of Landsat from Cuprite, Nevada. The training data size range from 5% to 95%

The choice of a varying number of bands was also used to determine if fewer bands could be used to obtain at least comparable classification as obtained using multispectral data. Four training data sizes were used in our analysis. They were 10%, 12.5%, 15%, and 17.5% of the entire data. The classification was repeated using MLC. The results of the classification using both SVC and MLC are given in Table 2.

From Table 2, it is evident that the overall classification accuracy using SVC increases with the increase in the number of bands. In the case of MLC, when the number of bands was increased from 13 to 25 and beyond, it was observed that the classification accuracy decreased. This indicates that, compared to MLC, SVC is not affected by the Hughes Effect. We also observe from Table 2 that the classification accuracy of SVC is higher than that of MLC.

In the case of MLC, when the number of bands was increased to 49, 98, and 196, it failed to classify and gave error (cf. Table 2). The reason for this error is that MLC relies on having a non-singular class specific covariance matrix for all classes (Benediktsson et al. 1995; Richards and Jia 1998; Chi and Bruzzone 2007). In the case of hyperspectral data, the cause for the covariance matrix being singular is high dimensionality. Also in an N (number of bands) dimensional image, at least $N + 1$ samples from each theme are required to avoid the singularity of the covariance matrix (Richards and Jia 1998). In summary, SVC is a robust method that does not suffer from the Hughes Effect and is insensitive to the number of bands and training pixels used to classify using hyperspectral data, as compared to MLC.

Table 2 Overall classification accuracy using both SVC and MLC methods on Hyperion EO-1 image of Cuprite, Nevada. The overall classification accuracy is computed for different band combinations. The first trial had all the 196 unique bands; the consequent trials had the alternate bands from the previous trial, i.e., 98, 49, 25, 13, and 7 bands. The various training data sizes considered are 10%, 12.5%, 15%, and 17.5%

Percentage of training vs testing data	Classification method	Number of bands					
		196	98	49	25	13	7
10 : 90	MLC	*	*	*	56.46	56.84	55.91
	SVC	72.01	71.45	70.07	67.36	64.98	62.16
12.5 : 87.5	MLC	*	*	*	57.06	58.12	57.11
	SVC	73.39	71.88	71.63	67.54	65.09	62.64
15 : 85	MLC	*	*	55.03	57.24	58.74	56.79
	SVC	74.84	74.59	72.80	68.75	65.94	62.88
17.5 : 82.5	MLC	*	*	55.98	58.43	59.54	57.53
	SVC	76.05	75.79	73.88	69.84	66.10	63.05

*For these training data sizes and band combinations MLC gave error due to the covariance matrix being singular

4.3 Analysis of the Number of Support Vectors

In this study, the number of support vectors used to obtain the optimum classification accuracy was analyzed. From Table 1, the variation in the overall accuracy using SVC increased from 7.45% in Goodnews Bay data to 12.33% in the Cuprite data. A scrutiny of the number of support vectors used in both the data to obtain the optimum accuracy in classification revealed that in the Goodnews Bay data, when the training data size was 6%, the percentage of support vectors was 40% of the training data. When the training data size was increased to 96%, the percentage of support vectors was reduced to 22% of the training data (Fig. 7). Similarly, in the Cuprite data, when the training size was 5%, the percentage of support vectors was 88% of the training data. When the training data size was increased to 95%, the percentage of support vectors decreased to 77% of the training data (Fig. 8). It is evident that in the case of the Cuprite data large numbers of the training pixels are being used as the support vectors compared to the Goodnews Bay data. In fact, it is observed that the larger the number of support vectors, the higher the variation in the classification accuracy as the training data sizes are varied.

The highest classification accuracy achieved using SVC on the Goodnews Bay data was 96.87%, when 96% (3835 pixels) of the data was used for the training (Fig. 9A). However, out of the 96% data in the training only 833 pixels (20.85%) were effectively contributing to the training data as support vectors (Fig. 9B). When these 833 pixels were used for training and tested against the rest of the data (79.15% testing data size), they yielded a 99.81% overall classification accuracy. It is important to note that, when the training data was selected randomly with 839 pixels (21% training data size), the classification accuracy obtained was only 91.22%. Thus, it is obvious from the above results that SVC can give higher classification accuracy

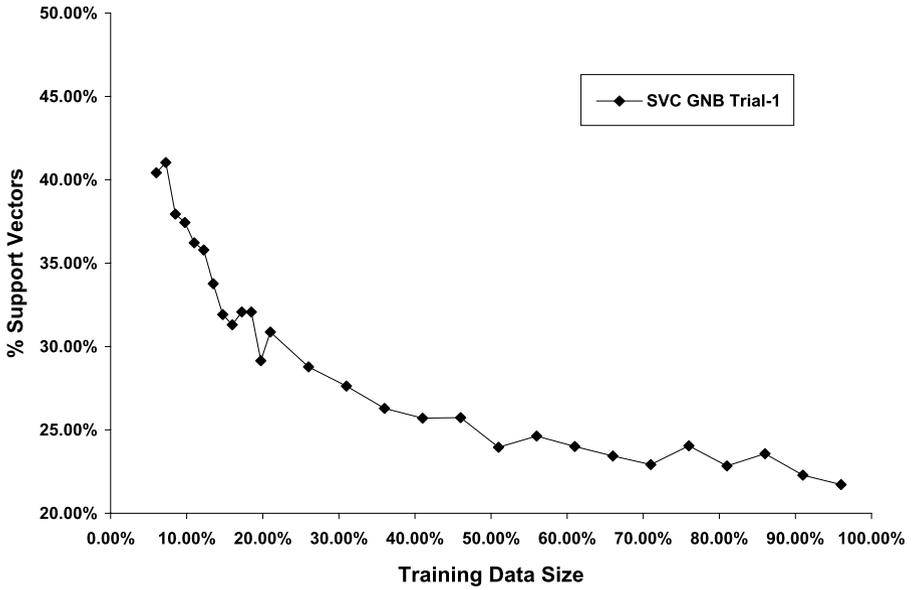


Fig. 7 The number of support vectors needed for the different training data sizes of the Landsat GNB Trial-1

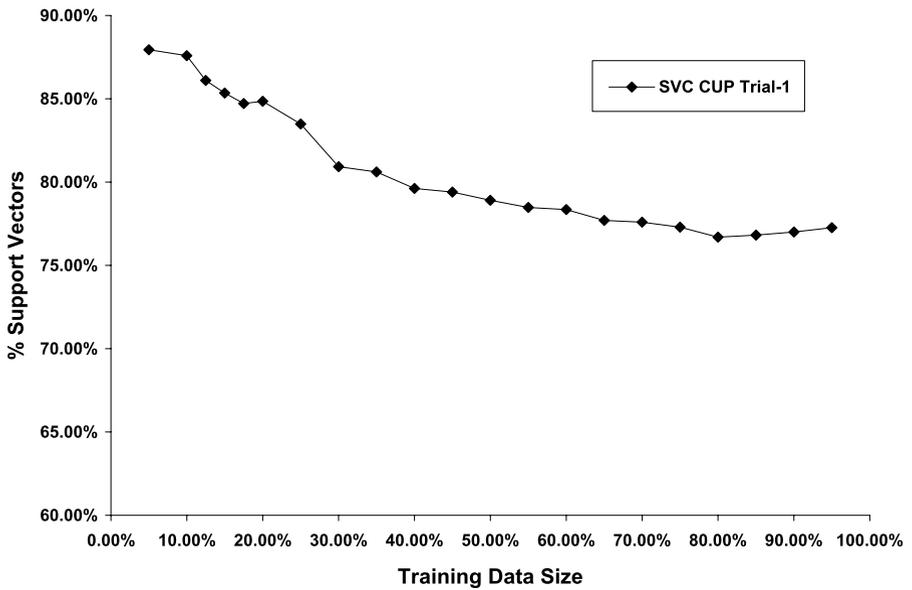


Fig. 8 The number of support vectors needed for the different training data sizes of the Landsat CUP Trial-1

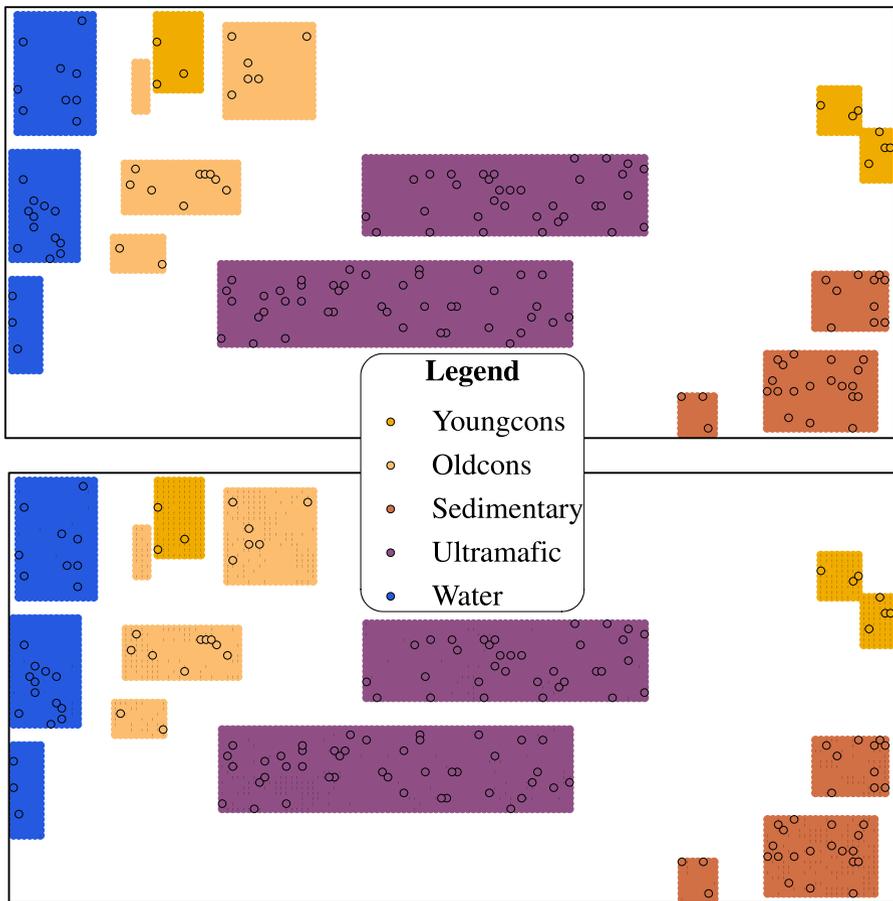


Fig. 9 Themes identified from September 27th, 2000, Landsat 7 ETM data of Goodnews Bay: **A** The pixels with a circular outline show the 160 pixels of testing data (4%) and the other 3835 pixels (96%) were used for training; **B** The pixels with a filled dot represent the 833 support vectors that are effectively contributing to the model from the 3835 training data

with a smaller training dataset. However, the strategy would be to include the useful support vectors in that small training data. In support of our observation, Foody and Mathur (2004) stated that, unless some intelligent training data acquisition process was followed, the solution to having the useful support vectors included in the critical training sample can only be achieved through a choice of a large training sample.

5 Summary and Conclusions

The two parameters that have to be determined when a radial basis kernel is used with SVC are C and γ . The classification accuracy of SVC largely depends on the optimal values of these parameters. A grid search approach recommended by Hsu et al. (2003)

was used for this study. This approach can often be time consuming and, therefore, an alternate approach has been recommended by Keerthi and Lin (2003). A comparison of the SVC against MLC was carried out for both multispectral and hyperspectral data. It is evident from the results that, in both multispectral and hyperspectral data, the overall classification accuracy using SVC outperforms MLC. Moreover, it was concluded that SVC could be applied for both multispectral and hyperspectral data without any restrictions, whereas MLC failed to classify when the covariance matrix was singular. We conducted a classification of the hyperspectral data with different band combinations to test if SVC was affected by the dimensionality issue (Hughes Effect). It was observed from the results that in the case of SVC, as the number of bands increased from 7 to 196, the classification accuracy also increased. Thus, it is evident from our results that the dimensionality issues do not affect SVC. On the other hand, as the number of bands increased, the classification accuracy using MLC decreased.

A selective training data selection was done with the Landsat GNB Trial-1 data. The criterion for the selective training data selection was to identify all the support vectors in the training data that gave the highest classification accuracy. Furthermore, a training dataset was developed which contained only these selected support vectors. By doing so, in the case of the Landsat GNB Trial-1, the classification accuracy increased from 91.22% to 99.81%. Therefore, if the useful support vectors can be included in the critical training sample by some *a priori* knowledge, the classification accuracy of SVC can be considerably increased with a small training sample. The training pixels are always limited and expensive in remote sensing (Jia 1999). Future work should develop a technique to identify useful support vectors for various remote sensing classifications and to develop a guideline for training data collection. If this can be achieved, training data would only be collected from the useful support vector locations and this in turn would help achieve high classification accuracy with a small training sample.

References

- Barry P, Shippert P, Gorodetzky D, Beck R (2003) Draft Hyperion hyperspectral mapping exercise using atmospheric correction and end members from spectral libraries and regions of interest with data from Cuprite, Nevada. EO-1 User Guide, v 2.3, 74 p
- Benediktsson JA, Sveinsson JR, Arnason K (1995) Classification and feature extraction of AVIRIS data. *IEEE Trans Geosci Remote Sens* 33(5):1194–1205
- Chi M, Bruzzone L (2007) Classification of hyperspectral remote sensing data with primal semi-supervised SVMs: 4th International Workshop on Pattern Recognition in Remote Sensing (PRRS'06), Hong Kong. *IEEE Trans Geosci Remote Sens* 45(6):1870–1880
- Chang CI (2003) *Hyperspectral imaging: techniques for spectral detection and classification*. Kluwer/Plenum, New York, 370 p
- Foody G, McCullagh MB, Yates WB (1995) The effect of training set size and composition on artificial neural net classification. *Int J Remote Sens* 16:1707–1723
- Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans Geosci Remote Sens* 42:1335–1343
- Gunn SR (1998) Support vector machines for classification and regression. Technical Report, University of Southampton, 54 p
- Hord MR (1982) *Digital image processing of remotely sensed data*. Academic Press, New York, 256 p
- Huang C, Davis LS, Townshend JRG (2002) An assessment of support vector machines for land cover classification. *Int J Remote Sens* 23:725–749

- Hughes GF (1968) On the mean accuracy of statistical pattern recognizers. *IEEE Trans Inf Theory* 14:55–63
- Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification. National Taiwan University, 12 p
- Ikeda M, Dobson FW (1995) Oceanographic applications of remote sensing. CRC Press, Boca Raton, 492 p
- Jia X (1999) Adaptable class data representation for hyperspectral image classification. <http://www.gisdevelopment.net/aars/acrs/1999/ts10/ts10109pf.htm>
- Keerthi SS, Lin CJ (2003) Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput* 15:1667–1689
- Kohavi R, Provost F (1998) Glossary of terms. *Mach Learn* 30(23):271–274
- Lillesand TM, Kiefer RW, Chipman JW (2004) Remote sensing and image interpretation, 5th edn. Wiley, New York, p 724
- Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans Geosci Remote Sens*, pp 1778–1790
- Mertie JB Jr (1940) The Goodnews platinum deposits, Alaska. *US Geol Surv Bull* 918:97
- Meyer D (2001) Support vector machines. *R News*, Volume 1/3. http://cran.r-project.org/doc/Rnews/Rnews_2001-3.pdf
- Murai S (1996) GIS workbook (fundamental course). Japan Association of Surveyors, Tokyo, 169 p
- Pal M, Mather PM (2005) Support vector machines for classification in remote sensing. *Int J Remote Sens* 26:1007–1011
- Richards JA, Jia X (1998) Remote sensing digital image analysis, 3rd edn. Springer, Berlin, 63 p
- Schrader S, Pouncey R (1997) *Erdas field guide*, 4th edn. Erdas Inc., Atlanta Georgia, 686 p
- Schwoenegerdt RA (1983) Techniques for image processing and classification in remote sensing. Academic Press, New York, p 245
- Sherrod PH (2003) Classification and regression trees and support vector machines for predictive modeling and forecasting. DTREG program manual. www.dtreg.com
- Stewart JH, Carlson JE (1978) Geologic map of Nevada. Nevada Bureau of Mines and Geology, Map
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York, 188 p
- Zhu G, Blumberg DG (2002) Classification using ASTER data and SVM algorithms – The case study of Beer Sheva, Israel. *Remote Sens Environ* 80:233–240