Feature Selection with PSO and Kernel Methods for Hyperspectral Classification

Anthony S. J. Tjiong and Sildomar T. Monteiro
Australian Centre for Field Robotics
School of Aerospace, Mechanical and Mechatronic Engineering
University of Sydney, NSW 2006, Australia
{FirstName.LastName}@sydney.edu.au

Abstract—Hyperspectral image data has great potential to identify and classify the chemical composition of materials remotely. Factors limiting the use of hyperspectral sensors in practical land-based applications, such as robotics and mining, are the complexity and cost of data acquisition, and the processing time required for the subsequent analysis. This is mainly due to the high dimensional and high volume nature of hyperspectral image data. In this paper, we propose to combine a feature selection method, based on particle swarm optimization (PSO), with a kernel method, support vector machines (SVM), to reduce the dimensionality of hyperspectral data for classification. We evaluate several different kernels, including some optimized for hyperspectral analysis. In particular, a recent kernel called observation angle dependent (OAD) kernel, originally designed for Gaussian Process regression, was extended for SVM classification. The SVM with the optimized kernel was then applied to induce the feature selection of a binary version of PSO. We validate the method using hyperspectral data sets acquired of rock samples from Western Australia. The empirical results demonstrate that our method is able to efficiently reduce the number of features while keeping, or even improving, the performance of the SVM classifier.

I. INTRODUCTION

Mining involves heavy machinery and the processing of vast quantities of material. Through the use of automated processes, it is possible to improve the safety of employees by reducing the human element from mining and, thus, reducing the risk of severe injury or death. Collection and analysis of data about the geology is crucial in order to make decisions on the best options to spend resources. Identifying the location of minerals of interest in a mine site can be complex and require lengthy interpretation of geophysical data. Hyperspectral data has been successfully used to classify materials on the Earth's surface [1]. Hyperspectral image data provides a large amount of data about the chemical composition and spatial distribution of materials. Different materials present different level of reflectance at each wavelength in the spectrum, thus allowing the remote identification of their composition without destroying the sample. One issue with hyperspectral data, however, is the inherent large number of bands, which can make the analysis and processing very time consuming or intractable. Our goal is to find the key spectral features that allow the successfully classification of the data, in order to reduce the number of spectral bands required and reduce the data processing complexity. This has the potential to reduce the requirements for data acquisition to limited number of bands. As a consequence, remotely sensed images could be obtained using less expensive multispectral cameras, calibrated to acquire data from these key spectral bands.

In this paper, we propose a method that combines particle swarm optimization (PSO) with support vector machines (SVMs) to perform the selection of optimal hyperspectral features. Our feature selection method attempts to reduce the dimensionality of the data by selecting only the features that are required for an accurate classification. PSO is a type of evolutionary optimization algorithm that has shown to perform well in a range of applications [2]. We apply PSO to the problem of feature selection and compare its performance to a conventional method, sequential selection, as a baseline.

In order to select the most relevant features for mapping geology, we use the support vector machine (SVM) [3]. SVM is a maximum margin classifier that calculates a decision boundary to separate data. It can be used to take the features obtained from hyperspectral data and classify the data according to its material composition. SVM has been shown to provide state-of-the-art results for hyperspectral classification [4]. We present an empirical analysis of different kernel functions that are used in SVMs. In particular, we introduce the novel observation angle dependent (OAD) covariance function [5], as a kernel for SVM classification. The OAD kernel was originally proposed as a covariance function for Gaussian processes.

We validate the proposed method using hyperspectral data sets collected of rock samples from an iron ore mine in Western Australia. Three data sets were used in the experiments: a training set acquired under artificial illumination, an independent test set acquired under varying conditions of illumination, and hyperspectral image data of a vertical mine face.

The main contributions of this paper are: i) proposing a novel approach combining PSO and SVMs to select optimal hyperspectral bands; ii) introducing the application of the OAD kernel for SVM classification of hyperspectral image data; iii) presenting an empirical comparison of several kernels for hyperspectral classification; iv) presenting an empirical analysis of the optimal set of spectral bands for classifying ore-bearing rocks.

II. Related Work

There has been a great deal of research done into using SVM as well as its application for classifying hyperspectral data. There has been work done looking at comparing SVM to other feature extraction or feature reduction approaches [6]. Feature selection has also been used in the classification of hyperspectral data. A method of feature selection based on linear separability and spatial invariance has been examined in [7]. Distance metrics commonly used in hyperspectral data processing, such as the spectral angle mapper (SAM) and the spectral information divergence (SID), have been extended as kernels for classifying hyperspectral data in [8], [9]. However, most of the papers above presented results using only well-studied hyperspectral data sets, such as the Indian Pine data set for example, in which training and test pixels come from the same image.

There are several feature selection methods that can help to reduce the number of necessary features required for classification [10]. Genetic algorithms, such as in the approach proposed in [11], is one of the most popular methods. PSO has been applied in combination with SVM [12]. However, their approach uses a different variant of PSO and they use only the basic binary SVM classifier. Moreover, they test their method only on publicly available benchmark data sets. Another PSO-based method has been proposed to perform feature extraction from hyperspectral data [13]. However, they combine PSO with a neural network for performing regression. To the best of our knowledge, our method is the first application of PSO and SVM for hyperspectral feature selection using a specialized kernel optimized for spectral classification.

III. Support Vector Machines

SVM, introduced in [14], is a binary classifier that has shown robust and accurate performance for classification in a range of applications. The SVM algorithm finds the best hyperplane that separates the data by maximizing the margin between the hyperplane and the support vectors. Since SVMs have been widely used in recent studies, in this paper we omit details on the training and optimization algorithm, which can be found in standard textbooks on the subject, such as [3]. Implementation of the SVM algorithm is also readily available in open source code or embedded in commercial software packages.

Because the SVM is based on a hard decision boundary, it cannot provide the probabilities that its classifications are correct. It can only decide whether a sample belongs to one category or not. Nevertheless, probabilistic outputs can be estimated from the SVM output by fitting a parametric model [15]. We use an improved version of this method to calculate the class probabilities, as provided in [16].

A. Kernels

In cases where the data is not linearly separable, SVM requires the use of a kernel function to map the data into a higher dimensional space. The SVM can then find the best linear separation of the transformed data in this higher dimensional space, which will correspond to a nonlinear classification in the input space. In this paper, we investigate several kernel functions, including standard kernel functions commonly used for SVM classification and specialized kernel functions proposed specifically to process hyperspectral data.

The polynomial kernel and the Gaussian radial basis function (RBF) kernel are very popular choices of kernel functions for SVM. The polynomial kernel maps the feature space into a higher dimension defined by the polynomial order. It is written as

$$k\left(x, x'\right) = \left(x^T x' + 1\right)^n. \tag{1}$$

The Gaussian RBF produces a mapping equivalent to an infinite dimensional Hilbert space. It therefore allows for the mapping of a wider variety of data sets. It is written as

$$k\left(x, x'\right) = \exp\left(-\frac{1}{2\sigma^2} \|x - x'\|\right). \tag{2}$$

The spectral angle mapper (SAM) and the spectral information divergence (SID) are functions specifically designed for analysing spectral signature information [9]. Their kernelized version incorporate prior knowledge and information about spectral signatures into SVM. The SAM kernel is written as

$$k\left(x, x'\right) = acos\left(\frac{x^T x'}{\|x\| \|x'\|}\right), \tag{3}$$

and the SID kernel is written as

$$k\left(x, x'\right) = \sum_{i=1}^{d} p_i \log\left(\frac{p_i}{q_i}\right) + \sum_{i=1}^{d} q_i \log\left(\frac{q_i}{p_i}\right), \tag{4}$$

where $p_i = x_i / \sum_{i=1}^{d} x_i$, $q_i = x_i' / \sum_{i=1}^{d} x_i'$, and the parameter d is the total number of features, or dimensions.

The modified observation angle dependent (OAD) nonstationary covariance was proposed in the context of Gaussian processes [5] but, since it satisfies the Mercer's conditions, it can be used as a kernel function in SVM. The OAD kernel is written as

$$k\left(x, x'\right) = \sigma_0^2 \left(1 - \frac{1 - sin\phi}{\pi} acos\left(\frac{x^T x'}{\|x\| \|x'\|}\right)\right), \tag{5}$$

where σ_0 and ϕ are scalar hyper-parameters of the kernel function. Note that the OAD kernel can be seen as a reparametrization of the SAM kernel, which turns out to be very efficient and robust.

IV. Particle Swarm Optimization

The particle swarm optimization (PSO) algorithm was developed while trying to simulate the behavior of animals that exhibit both individual and group behavior, such as birds, bees and fish [2]. PSO can be used to find near-optimal solutions in search spaces containing local minima. Each particle is considered to be one possible solution to the problem. The swarm of particles are given a random initial location and velocity and are updated according to the following equations

$$v_{i,j}^{t+1} = W v_{i,j}^t + c_1 r_1 \left(p_{i,j} - x_{i,j}^t\right) + c_2 r_2 \left(p_{g,j} - x_{i,j}^t\right); \tag{6}$$

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1}, \qquad (7)$$

where x and v are the position and velocity of the particle i for dimension j, respectively, at time t. W is the inertial weight and represents how much of the previous velocity is retained while exploring. Parameters c_1 and c_2, and r_1 and r_2 are, respectively, the weighting and random factor associated with the local best particle $p_{i,j}$ and the global best particle $p_{g,j}$.

The local best of a particle can be considered the cognitive aspect of PSO, whereas the global best can be considered as the social aspect. Due to this 'communication' between the particles, PSO is able to "fly" over the local minima in the direction of the global optimum. However, the PSO algorithm is stochastic in nature and, therefore, depending on the initial conditions of the search, the evolution of the particles may vary and the global optimum may not be found.

A. PSO feature selection

We apply PSO as a feature selection technique and determine how applicable it is for selecting features from hyperspectral data. Because PSO is designed to search through continuous spaces, it needs to be discretized for use in feature selection. We use a binary version of the PSO algorithm, as in [17].

By converting the position vectors of PSO into probabilities, we can use roulette wheel selection to select the appropriate features. The probablities are calculated as follows

$$P_{i,j} = \frac{x_{i,j}^{\alpha}}{\sum\limits_{j=1}^{n} x_{i,j}^{\alpha}}, \qquad (8)$$

where α is a scaling factor known as the selection pressure. Each feature is assigned with a probability and the roulette wheel marks the feature to be selected. The roulette is spun repeatedly until the desired number of features is selected.

V. EXPERIMENTAL RESULTS

A. Hyperspectral Data Sets

The data used in our experiment was comprised of spectra taken from rock samples obtained from an iron ore mine in Western Australia, as described in [18]. Reflectance spectra were taken using an ASD spectrometer. There are 409 spectral bands in the data set. The data set is composed of 9 rock types typically found in the region, as listed in Table I. The spectrometer data set was separated into a training set—pure spectra acquired under artificial light—and a test set—spectra acquired under various conditions of illumination, to simulate the real scenario of classifying data in the field. The number of samples used for training and testing was 228 and 297, respectively.

In addition, we also tested the applicability of the SVM classifier using the reduced number of features to classify hyperspectral images. Hyperspectral image data was acquired using two separate visible infrared (VNIR: 400–970 nm) and shortwave infrared (SWIR: 907–2516 nm) Specim sensors

TABLE I
ROCK TYPES CONTAINED IN THE HYPERSPECTRAL DATA SET

Code	Description
WRC	Water Reactive Clay
BIF	Banded Iron Formation
GOE	Goethite
MAR	Martite
SHL	West Angelas Shale
SHN	Manganiferous Shale
NS4	Marker Shales
NS3	Volcanic Shales
CHT	Chert

Fig. 1. Hyperspectral camera used to acquire data for the experiments; note that there are two sensors, one for the VNIR and other for the SWIR.

mounted adjacently on a rotating stage, as shown in Fig. 1. The hyperspectral image data set was acquired from a vertical geology, typical of mining environments, as shown in Fig. 6.

B. Performance Metrics

To evaluate the performance of our method, we present an analysis of various statistical metrics including accuracy, precision, recall, F-measure, kappa and area under the ROC curve (AUC). Each of the metrics was calculated as follows:

$$Accuracy = \frac{TP + TN}{P} \qquad (9)$$

$$Precision = \frac{TP}{TP + FP} \qquad (10)$$

$$Recall = \frac{TP}{TP + FN} \qquad (11)$$

$$F-measure = \frac{2 \times precision \times recall}{precision + recall} \qquad (12)$$

$$Kappa = \frac{(P \times (TP + TN) - C)}{(P^2 - C)} \qquad (13)$$

$$P = (TP + TN + FP + FN)$$

$$C = (TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN)$$

Fig. 2. RGB image of the study area

	Polynomial	Gaussian RBF	SAM	SID	OAD
Accuracy	0.9192	0.8965	0.8679	0.8627	0.9426
Precision	0.7156	0.6444	0.4525	0.3360	0.8659
Recall	0.7533	0.6661	0.3883	0.2644	0.7350
F-measure	0.6945	0.5963	0.5037	0.4423	0.7879
Kappa	0.6553	0.5496	0.2342	0.1376	0.7312
AUC	0.9323	0.8998	0.7953	0.6920	0.9367

Parameter	Value
Max. Epochs (iterations)	200
Population Size	40
Max. Particle Velocity	4
c_1	2
c_2	2
Initial ψ	0.95
Final ψ	0.2
Epochs to reach final ψ	90

where TP, TN, FP, FN are the number of true and false positives and negatives.

Each different metric captures a different aspect of the behavior of the model. Accuracy is by far the most widely used metric in classification studies. However, as our results will show, accuracy alone might not be the best indication of good performance of the classifier. Because of this, in fields such as information retrieval or bioinformatics, other metrics, such precision, recall and F-measure, are normally used. Kappa is a metric that is popular in hyperspectral (remote sensing) papers and is included here to allow comparison. The ROC curve and the corresponding AUC are also important metrics used in machine learning. For all metrics, higher values represent better performance, and their maximum value is 1, by definition.

C. Kernel Selection

The SVM algorithm was implemented using the following optimization techniques: least squares minimization [19], sequential minimal optimization [15] and quadratic programming [3]. The best performance of each kernel was kept, regardless of optimization method. Since the different metrics may give different results, we chose to optimize the F-measure, based on the observation that its results suffer less positive bias than accuracy. The data set used for kernel optimization was the training data and independent test data, i.e., performing out-of-sample analysis. The parameters of each kernel were optimized using a grid search. Table II shows a summary of the best results for all kernels.

D. Hyperspectral Band Selection

After choosing the kernel found to be the best, which, in this case, was the OAD kernel, feature selection techniques

were then used to reduce the select the most relevant spectral bands to classify the data. We compare our PSO-based method against a widely used technique for dimensionality reduction, sequential selection [10]. Sequential selection is a feature selection technique that involves comparing different subsets of the data based on a criterion function. It can be performed either forward or backward by adding or removing features, one feature at each step. Depending on the direction of the selection, features will be added, sequential forward selection (SFS), or removed, sequential backward selection (SBS), from the feature subset based on the criterion function. As in the case with PSO, SBS and SFS the performance of the SVM classifier as the criterion function. Again, F-measure was the chosen metric to induce the feature selection.

During feature selection, the SVM algorithm was trained using the least squares method [19], which is the most computationally efficient method. Because PSO is stochastic by nature, PSO was run multiple times and the best results, highest F-measure, were kept. The parameters used in the PSO algorithm are shown in Table III. Since we had no a priori knowledge of the optimal number of features, tests were run iteratively, each time searching for a different number of features. Figure 3 shows results for up to 50 features being selected. Note that scores in the range of 5–15 features provide comparable results to higher numbers of features.

The best performing subset selected by each method contained different number of features. Tables IV, V and VI show the features selected by PSO, SFS and SBS, respectively, as well as the average wavelengths corresponding to the spectral band that the feature represents. Figure 4 show a representation of the wavelengths of selected features for PSO, SFS and SBS; the plots also show the spectral curve of martite, to
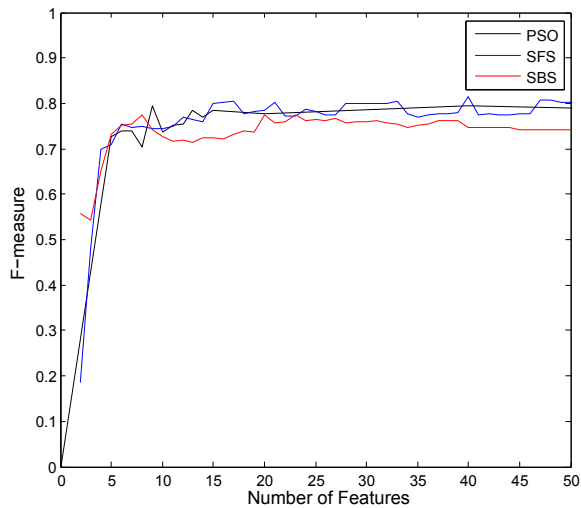
Fig. 3. Comparison of performance of PSO, SFS and SBS for the number of features ranging from 1 to 50.

TABLE IV
PSO BEST SELECTED FEATURES

Method	Band Number	Corresponding Wavelength (nm)
PSO	3	408.54
	46	506.41
	69	559.70
	88	604.91
	116	671.60
	155	765.06
	318	1603.96
	339	1736.5
	388	2202.52

demonstrate the wavelengths that seem to have significance in classifying spectral data.

The features selected by PSO, SFS and SBS were also compared against the entire feature set to determine whether the accuracy of classification was comparable for a small number of features when compared to all the features. A summary of the results is shown in Table VII. In addition, Fig. 5 shows the ROC curve for the features selected by PSO, SFS and SBS.

TABLE V
SFS BEST SELECTED FEATURES

Method	Band Number	Corresponding Wavelength (nm)
SFS	4	410.77
	32	474.26
	68	557.32
	106	647.76
	129	702.61
	174	811.57
	210	899.33
	257	1135.81
	299	1483.91
	386	2189.94

TABLE VI
SBS BEST SELECTED FEATURES

Method	Band Number	Corresponding Wavelength (nm)
SBS	2	406.32
	51	517.89
	89	607.29
	104	642.99
	178	821.36
	207	892.02
	253	1110.46
	298	1477.59

E. Classifing Hyperspectral Image Data

We present a qualitative test of the selected features to classify hyperspectral images. The selected features were used to train SVM using the two data sets acquired with the spectrometer—both the training and test data sets used in the previous sections combined. The multiclass classification was obtained using a one-against-all approach in which the most likely class based on the SVM probabilistic estimate was assigned for each pixel.

We compared the features selected by PSO, SFS and SBS, to the full feature set. Figure 6 show a visual representation of the classification of the various rock types performed on the hyperspectral image data set; each different color represents a different rock type; the background sky is blacked out in the image; and N/D in the legend, Fig. 6(e), refers to "no data." Note that there are no detailed ground-truth labels for the hyperspectral image data. Therefore we cannot provide numerical results on this test. However, the results coincide with the known geology distribution in the region.

VI. DISCUSSION

Considering the various metrics used to assess performance, the OAD kernel clearly outperformed the other kernels with significantly higher scores across all metrics. The polynomial kernel appeared as the next best kernel in terms of F-measure and accuracy. The Gaussian RBF kernel was competitive, but required a great deal of parameter tuning to achieve its best performance. The poor performance displayed by both SAM and SID was unexpected, as they are kernels specifically designed for hyperspectral analysis. One thing to notice is how the range of accuracy values in the results was relatively narrow compared to F-measures. In this case, accuracy alone could produce ambiguous or even misleading results. F-measure, combined with the other metrics, allowed a more robust assessment of performance.

The results of the feature selection demonstrate that even using a reduced data set by selecting several key features we can still retain an acceptable level of accurate classification. The fact that by using only 9 features outperformed using all features suggests that, in the latter, the SVM is probably having convergence difficulties due to the high dimensionality of the data (curse of dimensionality). Even if the improvement in classification performance is statistically small, the reduction in training and test time for the SVM using only the selected

|	PSO: 9 Features	SFS: 10 Features	SBS: 8 Features	All 409 Features
Accuracy	0.9465	0.9508	0.9417	0.9426
Precision	0.8418	0.8571	0.8511	0.8659
Recall	0.7603	0.7538	0.7203	0.7350
F-measure	0.7975	0.7778	0.7738	0.7879
Kappa	0.7381	0.7539	0.7150	0.7312
AUC	0.9523	0.9370	0.9219	0.9367
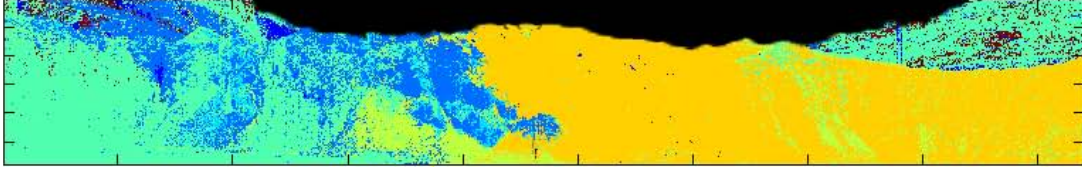
(a) All 409 features
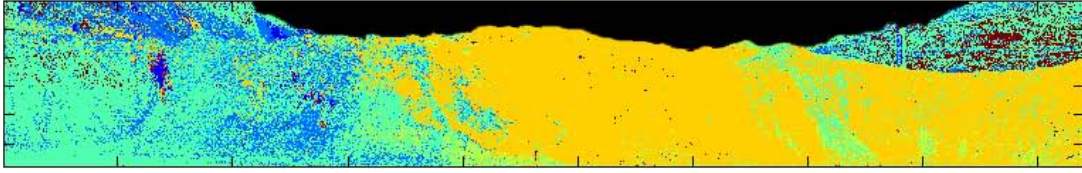
(b) PSO selection: 9 features
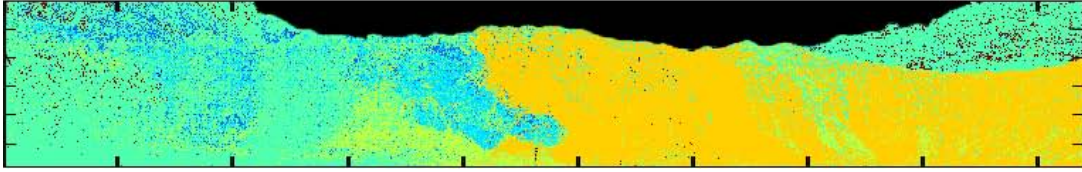
(c) SFS selection: 10 features

(d) SBS selection: 8 features

N/D WRC BIF GOL MAR SHL SHN NS4 NS3 CHT
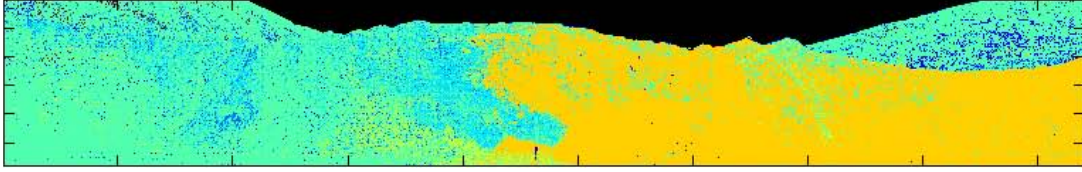
(e) Legend for figures

Fig. 6. Comparison of multi-class classification results produced by the different methods showing the spatial distribution of rock types on a vertical mine face.

feature set is remarkable. PSO provided the best performing selection of hyperspectral features. Probably because PSO has more flexibility to explore the search space of possible features, compared to SFS and SBS.

The selected features correspond to particular bands and as such the physical meaning of the bandwidths in that range can be identified. The results indicate that many of the important wavelengths for classifying the rock types are located at the lower end of the spectrum. Although none of the three selection techniques selected the same feature, there is evidence to suggest particular regions where the features selected are only a few bands apart. These particular wavelength ranges seem to hold a particular significance for the classification of minerals.

By limiting the number of bands to a selected few, we can design optical filters tuned in the selected spectral ranges. Such a multispectral system, would allow accurate classification of minerals without the burden cost of acquiring and processing unnecessary spectral bands.

It is worth noting that the experiment using the hyperspectral image data was very challenging, since the training data and the test images were acquired with different sensors (the ASD and the Specim sensors, respectively) and under different environmental conditions (illumination, geometry, field of view, etc). The good qualitative results can be attributed to the generalization capabilities of the SVM-OAD classifier and the careful calibration performed on the hyperspectral data sets.
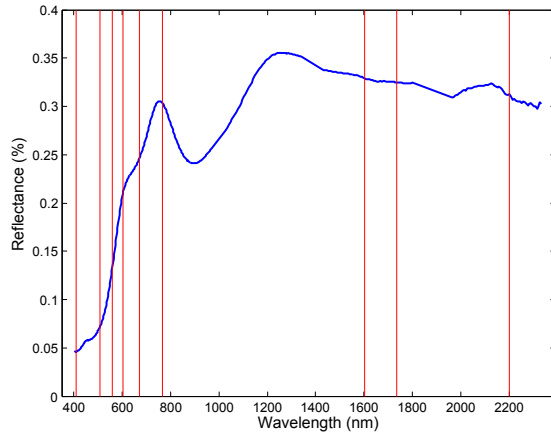
(a) PSO selection: 9 features
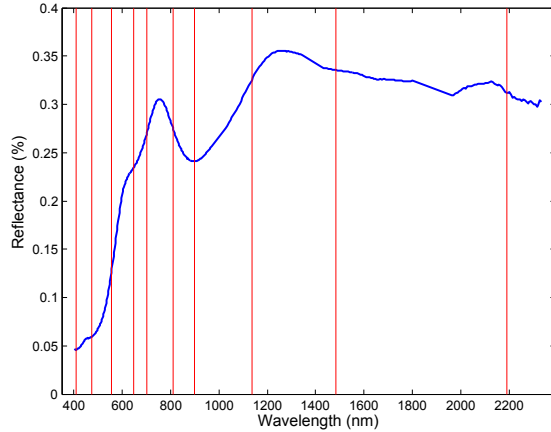
(b) SFS selection: 10 features
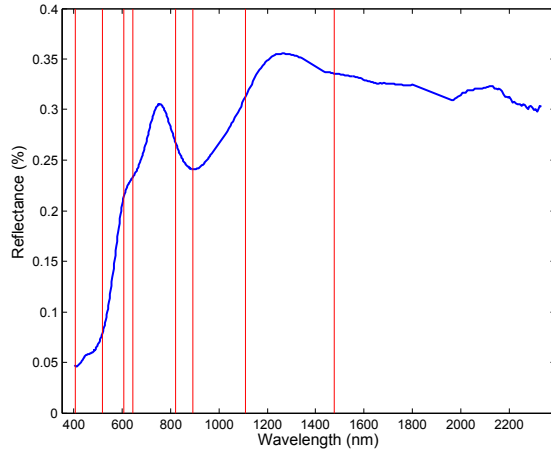
(c) SBS selection: 8 features

Fig. 4. Comparison of spectral features (vertical red lines) selected by the different methods for classifying martite (spectral curve in blue).
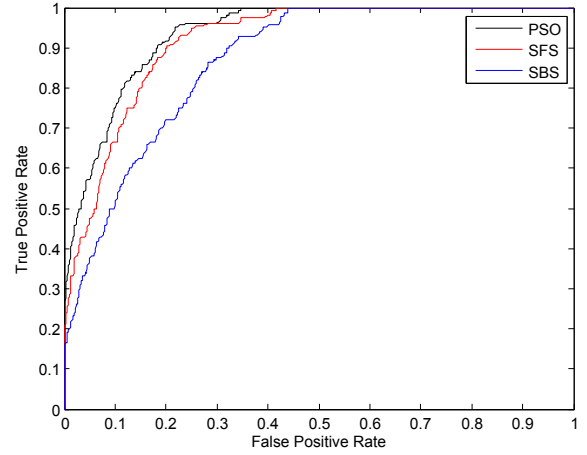

Fig. 5. ROC curves for the best selected features of PSO, SFS and SBS.

number of features optimized for classifying hyperspectral data. The effectiveness of the methods was demonstrated using hyperspectral data sets of rock samples; a challenging real-world problem, since the spectral signatures of ore-bearing rocks are very similar, which make classification particularly difficult. PSO was very effective in selecting features and outperformed conventional sequential selection techniques. Our experiments with SVM kernel selection reveal that the OAD kernel is highly applicable to the classification of hyperspectral data. It outperformed other widely used kernels, such as the RBF kernel, and also kernels specifically designed to process hyperspectral data.

Nevertheless, the PSO-based approach is not restricted to SVMs and can be easily extended to work in combination with other kernel methods, such as Gaussian process regression and classification. Future works include improving the convergence properties of PSO on the feature selection problem, perhaps investigating an alternative selection strategy, such as in [20].

ACKNOWLEDGMENT

VII. CONCLUSION

In this paper, we presented a method for reducing the number of features required to classify rock types from hyperspectral image data. Using a combination of a kernel method, SVM with a specialized kernel, and an evolutionary optimization technique, PSO, it is possible to efficiently select a limited

REFERENCES

[1] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougal, and C. A. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the usgs tetracorder and expert systems," *Journal of Geophysical Research*, vol. 108, no. E12, pp. 5.1–5.44, 2003.

[2] J. Kennedy, "Swarm intelligence," in *Handbook of Nature-Inspired and Innovative Computing*, A. Zomaya, Ed. Springer, 2006, pp. 187–219.

[3] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, ser. Adaptive computation and machine learning. Cambridge, MA, USA: MIT Press, 2001.

[4] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, no. Supplement 1, pp. S110 –S122, 2009.

[5] S. Schneider, A. Melkumyan, R. Murphy, and E. Nettleton, "Gaussian processes with OAD covariance function for hyperspectral data classification," in *IEEE International Conference on Tools with Artificial Intelligence*, 2010, pp. 393–400.

[6] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience And Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[7] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 9, pp. 3180–3191, 2009.

[8] Y. Du, C. Chang, H. Ren, C. Chang, J. Jensen, and F. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, no. 8, pp. 1777–1786, 2004.

[9] M. Kohram and M. Sap, "Composite kernels for support vector classification of hyper-spectral data," in *Mexican International Conference on Artificial Intelligence*, 2008, pp. 360–370.

[10] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[11] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 2, pp. 44–49, 1998.

[12] E. Alba, J. Garcia-Nieto, L. Jourdan, and E. Talbi, "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms," in *IEEE Congress on Evolutionary Computation*, 2007, pp. 284–290.

[13] S. T. Monteiro and Y. Kosugi, "Particle swarms for feature extraction of hyperspectral data," *IEICE Transactions on Information and Systems*, vol. 90, no. 7, pp. 1038–1046, 2007.

[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer-Verlag, 2000.

[15] J. C. Platt, *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, ser. Advances in kernel methods: Support Vector Learning. MIT Press, 1999, ch. 12, pp. 185–208.

[16] H. T. Lin, C. J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Machine Learning*, vol. 68, no. 3, pp. 267–276, 2007.

[17] S. T. Monteiro and Y. Kosugi, "A particle swarm optimization-based approach for hyperspectral band selection," in *IEEE Congress on Evolutionary Computation*. IEEE, 2007, pp. 3335–3340.

[18] S. Schneider, R. Murphy, S. T. Monteiro, and E. W. Nettleton, "On the development of a hyperspectral library for autonomous mining," in *Australasian Conference on Robotics and Automation*, 2009.

[19] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

[20] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459 – 471, 2007.