APPLYING BOOSTING FOR HYPERSPECTRAL CLASSIFICATION OF ORE-BEARING ROCKS

Sildomar T. Monteiro, Richard J. Murphy, Fabio Ramos, Juan Nieto

Australian Centre for Field Robotics, School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, NSW 2006, Australia
s.monteiro@acfr.usyd.edu.au

ABSTRACT

Hyperspectral sensors provide a powerful tool for non-destructive analysis of rocks. While classification of spectrally distinct materials can be performed by traditional methods, identification of different rock types or grades composed of similar materials remains a challenge because spectra are in many cases similar. In this paper, we investigate the application of boosting algorithms to classify hyperspectral data of ore rock samples into multiple discrete categories. Two variants of boosting, GentleBoost and LogitBoost, were implemented and compared with Support Vector Machines as benchmark. Two pre-processing transformations that may improve classification accuracy were investigated: derivative analysis and smoothing, both calculated by the Savitzky-Golay method. To assess the performance of the algorithms over noisy data, white Gaussian noise was added at various levels to the data set. We present experimental results using hyperspectral data collected from rock samples from an iron ore mine.

1. INTRODUCTION

Characterizing surface geology from hyperspectral data can be of enormous value for the mining industry. The accurate assessment of lithology can be used during several phases of the mining process, from exploration to processing and reconciliation. Despite the constraint that hyperspectral data only provides information from the surface of rocks [1], it can be useful in open pit mine operations where the rocks of interest are exposed. It has the potential to provide fast assessment of the location and distribution of waste and ore on the bench, resulting in more efficient mining and providing a higher-value product.

The authors are thankful to Mr. James Batchelor, Rio Tinto Technology and Innovation, for his assistance in collecting the hyperspectral data and analysing the rock samples.

This work has been supported in part by the Rio Tinto Centre for Mine Automation and in part by the ARC Centres of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government.

Hyperspectral sensors acquire data in hundreds of narrow, contiguous bands and provide a powerful tool for non-destructive analysis of remote samples [2]. Spectral signature analysis of hyperspectral data can be applied to classify samples into categories and produce land cover maps [3]. Conventional land cover classification methods allow easy distinction among different materials, e.g., bare soil, vegetation and minerals. However, there are still challenges in providing robust hyperspectral classification algorithms [4]. There are issues caused by the high dimensionality of hyperspectral data and the correlation between spectral bands. It is difficult and laborious to produce labeled samples for ground-truth, which may lead to correlation between training and validation data sets. There may also be significant amount of noise in the data due to the narrow bandpasses which are sampled, decreasing solar irradiance, particularly towards longer wavelengths, sensor induced effects and errors in calibration. Therefore, robust material identification is still a significant challenge, especially when targets present a high degree of spectral similarity [5].

In this paper, we investigate the performance of machine learning techniques to classify hyperspectral data of ore-bearing samples into discrete rock categories. The hyperspectral classification problem is characterized by having multiple categories (rock types), high-dimensional features (hyperspectral bands), and limited number of labelled samples (ground-truth). Boosting is a machine learning technique for supervised classification that has become very popular due to its sound theoretical foundation, and also due to many empirical studies showing that it tends to yield smaller classification error rates and be more robust to overfitting than competing methods, e.g., Neural Networks or Support Vector Machines (SVMs) [6]. We propose to apply a version of Boosting called LogitBoost, which can efficiently classify multiple categories directly. We present experimental results comparing LogitBoost with two benchmark algorithms: another Boosting variant called GentleBoost and an extension to SVMs called least squares-SVMs. The effect on the classifiers' performance of pre-processing the data using Savitzky-Golay smoothing and derivatives is also in-

978-1-4244-4948-4/09/$25.00 © 2009 IEEE

vestigated. The algorithms were assessed using hyperspectral data sets of ore-bearing samples collected from an open pit mine in Western Australia.

2. METHODOLOGY

Let us consider that the hyperspectral data is given as a vector $x_i \in \Re^d$ comprising d spectral bands. The training set is composed of pairs $\langle (x_1, y_1), \ldots, (x_n, y_n) \rangle$ of n labelled examples, where each sample $i = \{1, \ldots, n\}$ can be assigned to a label y. For the binary classification problem, the target label set is defined as $y_i \in \{-1, +1\}$. In the multi-class case, each label is assigned an integer $y_i \in \{1, 2, \cdots, C\}$ with the number of classes $C \geq 3$.

Given a binary classifier, there are several schemes for coding the outputs to solve the multi-class problem [7]. The two most widely used strategies are the one-versus-all and the one-versus-one approaches [8]. The present study employs the one-versus-all approach which learns a set of binary classifiers $\{f_1, f_2, \cdots, f_C\}$, where the c-th class is assigned to the positive class, while the others are assigned to the negative class. The prediction of the set of classifiers is given by majority voting $y_i^* = \arg\max_{c=1,2,\cdots,C} \{f_c(x_i)\}$. The one-versus-one approach did not perform as well on this problem, probably due to the high number of classes in relation to the small number of samples.

2.1. Boosting

The idea of Boosting is to train many "weak" learners on various distributions (or set of weights) of the input data and then combine the resulting classifiers into a single "committee" [9]. A weak learner can be any classifier whose performance is guaranteed to be better than a random guess. There are many different variants of boosting algorithms. In this study, we investigate two versions called GentleBoost and Logitboost [10].

2.1.1. GentleBoost

GentleBoost is a "gentle" version of the popular AdaBoost algorithm, which is more robust numerically and has shown to outperform the latter in experiment tests [10]. GentleBoost is a binary classifier and can be extended to handle multi-class problems using the one-versus-all scheme. GentleBoost uses adaptive Newton steps to optimize the cost function of the classifier following an iterative procedure.

Let us assume a set of weighting coefficients w initialized as $w_i = 1/n$, for $i = 1, \ldots, n$, and a classification function $F(x) = 0$. Let M be the number of weak learners. Repeat for $m = 1, \ldots, M$:

a) fit the regression function f_m by minimizing the weighted least squares
$$\sum_{i=1}^{n} \left(w_i (y_i - f_m(x_i))\right)^2 ;$$
b) update $F(x) \leftarrow F(x) + f_m(x)$;
c) update weights $w_i \leftarrow w_i \exp(-y_i f_m(x_i))$;
d) re-normalize weights $w_i^m = \dfrac{w_i^m}{\sum_{j=1}^{n} w_j^m}$.

Finally, return predictions of the final ensemble
$$sign\left[F(x)\right] = sign\left[\sum_{m=1}^{M} f_m(x)\right] . \qquad (1)$$

2.1.2. LogitBoost

LogitBoost is a Boosting variant that can directly classify multiple classes. Logitboost, similarly to GentleBoost, uses stagewise optimization of the maximum likelihood through adaptive Newton steps to fit additive logistic regression models [10].

To solve the multi-class problem, the LogitBoost algorithm defines a symmetric multiple logistic transformation
$$p_c(x) = \frac{\exp\left(F_c(x)\right)}{\sum_{c=1}^{C} \exp\left(F_c(x)\right)} , \quad \sum_{c=1}^{C} F_c(x) = 0 , \qquad (2)$$
where p_c is the probability of assigning class c among C classes.

Let $y_c^* = \{-1, 1\}$ be the indicator response vector. For all classes $c = 1, \ldots, C$ and training samples $i = 1, \ldots, n$, we initialize the weighting coefficients w as $w_{ic} = 1/n$, the classification function as $F_c(x) = 0$, and the class probabilities as $p_c(x) = 1/C$. For a given number of weak learners $m = 1, \ldots, M$, repeat:

a) repeat for all classes $c = 1, \ldots, C$:
   i) compute working responses
   $$z_{ic} = \frac{y_{ic}^* - p_c(x_i)}{p_c(x_i)\left(1 - p_c(x_i)\right)} ;$$
   ii) compute weights $w_{ic} = p_c(x_i)\left(1 - p_c(x_i)\right)$;
   iii) fit the regression function f_{mc} by minimizing the weighted least squares
   $$\sum_{i=1}^{n} \left(w_{ic}(z_{ic} - f_{mc}(x_i))\right)^2 ;$$
b) set $f_{mc}(x) \leftarrow \dfrac{C-1}{C}\left(f_{mc}(x) - \dfrac{1}{C}\sum_{c=1}^{C} f_{mc}(x)\right)$;
c) update $F_c(x) \leftarrow F_c(x) + f_{mc}(x)$;
d) update $p_c(x)$ using Eq.(2) .

The prediction of the classifier is given by
$$F(x) = \arg\max_c F_c(x) . \qquad (3)$$

2.1.3. Regression stumps

As for the weak learners, the present study utilizes regression stumps, which can be viewed as binary decision trees with only one node. A regression stump learns an optimal

threshold θ that takes a feature ϕ such that the minimum number of examples x is misclassifed. It can be defined as

$$f(x, \phi, \theta, a, b) = a\delta\left[\phi(x) > \theta\right] + b, \tag{4}$$

where δ is an indicator function, and a and b are regression parameters. The parameters $\{\phi, \theta, a, b\}$ are optimized by minimizing the weighted squared error w.r.t. f. This can be performed efficiently by best-first search and forward selection [11].

2.2. Support Vector Machines

SVMs have been shown to be effective for nonlinear classification, regression and density estimation problems [12]. SVMs were introduced for the binary classification problem by fitting an optimal separating hyperplane between the positive and negative classes with the maximal margin [13]. The classical SVM algorithm is based on convex optimization theory, typically quadratic programming involving inequality constraints. We focus on a different formulation called Least Squares Support Vector Machines (LS-SVMs) [14], which present lower computational complexity and may scale better for high-dimensional problems. In a LS-SVM classifier the problem is simplified because of the use of equality constraints instead of inequalities. The solution can then be obtained in a finite number of steps by solving a set of linear equations.

In a LS-SVM classifier, its primal weight space is defined as

$$\hat{y}(x) = w^T\varphi(x) + b, \tag{5}$$

where φ is a non linear function which maps the input space into a higher dimensional space, the weight vector w and bias b are parameters of the hyperplane. The LS-SVM optimization problem can be formulated as the minimization of a function J defined as

$$J(w, b, e) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{i=1}^{n} e_i^2, \tag{6}$$

where γ is the regularization factor. The optimization is subject to the constraints $y_i = w^T\varphi(x_i) + b + e_i$, where $e_i = y_i - \hat{y}_i$. Solving this optimization problem in dual space using the kernel trick leads to finding the coefficients of the function

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b, \tag{7}$$

where the kernel function $K(x, x_i)$ is the dot product between the $\varphi(x)^T$ and $\varphi(x)$ mappings. The present study employs a radial basis function (RBF) as the kernel function, which can be defined as

$$K(x, y) = \exp\varphi\left(-\sigma\left\|x - y\right\|^2\right), \tag{8}$$

where σ is a free model selection parameter that controls the widths of the Gaussian or RBF functions.

2.3. Savitzky-Golay Filtering and Derivatives

Empirical studies have shown that derivative analysis increase classification accuracy of Artificial Neural Networks in some cases [15, 16]. However, due to increased amount of noise at each higher derivative, a careful noise reduction pre-processing step is required. A common approach to calculate spectrum derivatives is to first apply a smoothing filter to reduce the noise and then perform the derivatives using a finite differentiation scheme [17]. The Savitzky-Golay filter, also known as least-squares smoothing filter, is a more elaborate approach that fits polynomials and differentiates them analytically [18]. Savitzky-Golay filters are attractive for spectropy because they are effective at preserving the relative widths and heights of spectral signatures in noisy spectrometric data [19].

In the Savitzky-Golay method, a frame of data points surrounding the current point (spectral band) is fit to a polynomial using local least-squares regression. The function value of the current band is retained, while the function values of the others in the frame are discarded. The simplified least square convolution can be used to calculate the qth-order derivative $\bar{u}$ of a band λ according to

$$\frac{d^q\bar{u}_\lambda}{dx^q} = \sum_{s=-r}^{r} P_s^{(q)} u_{s+\lambda}(x), \tag{9}$$

where P is the set of filter coefficients, and r is the half-width of the filter size, which correspond to a smoothing window of $2r + 1$. Note that $\lambda = \{r + 1, \cdots, n - r + 1\}$, where n is the number of bands. Equation (9) allows calculation of the smoothed signal from zeroth order to the sixth order of derivatives [20]. The present study implements Savitzky-Golay smoothing using cubic (order 3) polynomials, for a moderate level of smoothing and capable of providing up to the third derivative.

3. RESULTS

The hyperspectral data was collected using a portable field spectrometer manufactured by Analytical Spectral Devices (ASD) Inc. (Boulder, Colorado). The sensor acquires hyperspectral data from the visible (350 nm) to the short-wave-infrared (SWIR) (2500 nm) region of the spectrum at 1 nm intervals. The data were downsampled to 2 nm intervals on the visible region and to 6.5 nm in the SWIR in order to match the resolution of bandpasses which are consistent with many commercially available hyperspectral imaging systems; the total number of bands used was 429.

For this study, 14 rock samples from an iron ore mine located in the Pilbara region of Western Australia were collected. The samples comprise several rock types representative of the mineralogy commonly found in that region: banded ironstone formation (BIF), martite, goethite, kaolinite, and mixtures of those. The classification of the samples

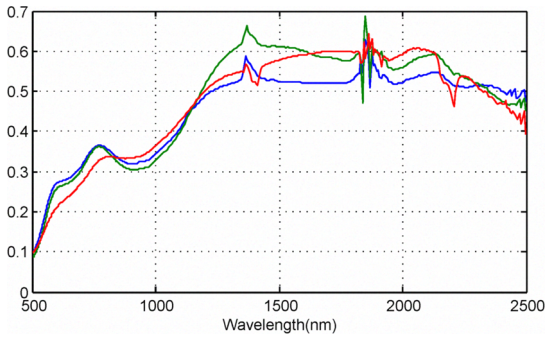


Fig. 1: Spectral signatures of three typical ore rocks: BIF (in blue), goethite (in green) and kaolinite (in red). The effect of light absorption from water vapour present in the atmosphere can be observed at wavelengths around 1450, 1950 and 2500 nm

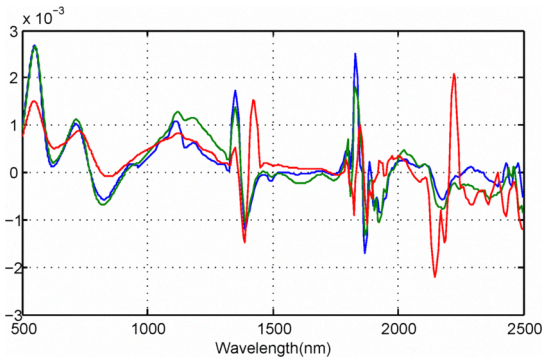
was provided by an experienced geologist. A total of 10 spectral measurements were taken from each rock type. As an example, the spectral reflectance signatures of three ore rocks are shown in Fig. 1. The resulting curves of the first and second derivatives are illustrated in Fig. 2a and Fig. 2b, respectively. We decided to keep the noisy bands due to water vapour absorption to provide a rigorous test for the machine learning algorithms.

In order to evaluate the performance of the algorithms over different conditions, artificial noise at various levels was added to the reflectance data. White Gaussian noise was added to produce spectra with different signal-to-noise ratios (SNRs) ranging from 60 dB to 10 dB, at 10 dB intervals. Figure 3 shows an example of the resulting spectral signatures after this procedure.

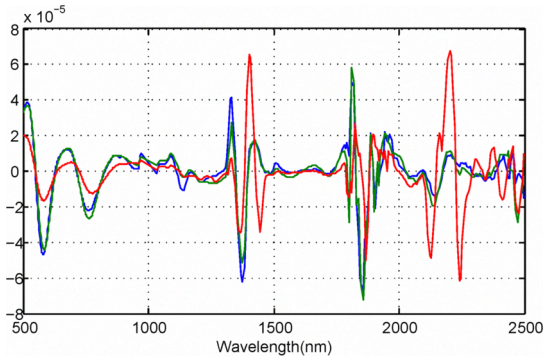
For LogitBoost and GentleBoost using decision stumps as weak learners, the only parameter that needs to be defined is the number of weak learners. We have chosen the number of weak learners that provided the better classification for the unprocessed reflectance data set, 35 for the LogitBoost and 33 for the GentleBoost. For LS-SVMs using a radial basis function kernel, we have to adjust the kernel scale factor. Using the same criterion as used in Boosting, the scale factor was set to 3 for all experiments.

The algorithms were tested using k-fold cross-validation. For the present data set, a 5-fold approach was implemented. At each turn, 2 spectral samples of each class are taken out of the training set. The model is then tested on this validation set. The results of all turns are grouped together and the final results are calculated based on the confusion matrix of the whole validation set. The model accuracy is the percentage of correctly classified samples over the total number of samples.

The experimental results for LogitBoost, GentleBoost and LS-SVMs are presented in Table 1. The rows correspond to the different pre-processing techniques, smoothing and derivatives, applied to the data set; the indices 20, 40, 80 indicate the width of the smoothing window in nanometers. The columns correspond to the different levels of noise added to the data, and the last column shows the no-noise-added data set; the best results for each SNR level are highlighted in bold letters. We have also separated the results into three groups based on the type of the data set: raw reflectance, first and second derivatives. These results are presented in Fig. 4 as box-and-whisker diagrams.



(a) First Derivative



(b) Second Derivative

Fig. 2: Derivative spectra of three rock samples: BIF (in blue), goethite (in green) and kaolinite (in red)

In the experiments, LogitBoost presented better performance than GentleBoost overall, although with slightly more variance on the derivatives data. The LS-SVMs achieved higher accuracy for the raw reflectance data, but presented higher variance and performed very poorly on the derivatives and noisy data. This agrees with a previous comparative study of SVMs and also AdaBoost [21], which indicate that the cost function for hyperspectral classification leads to sparser solutions by the Boosting algorithm. Nevertheless, LS-SVMs could yield better results on all cases if its parameters were fine tuned for each case, a procedure that is not required for Boosting, which performs well with fixed parameters. Another advantage of LogitBoost is that it can provide class probability estimates, which can be useful for

Table 1: Comparison of results of algorithms for ore classification using hyperspectral data

Method	Preprocessing	10dB	20dB	30dB	40dB	50dB	60dB	No-noise
	Raw Reflectance	0.1643	0.3857	0.7214	0.8714	0.9071	0.9214	0.9214
	Smoothing 20	0.3571	0.7643	0.9071	0.9286	0.9214	0.9214	0.9143
	Smoothing 40	**0.4429**	0.7857	**0.9357**	**0.9286**	**0.9357**	0.9143	0.9214
	Smoothing 80	0.4286	**0.8500**	0.9071	0.9286	0.9143	0.9286	0.9214
LogitBoost	1st Derivative 20	0.0214	0.0571	0.1214	0.2786	0.6929	0.8357	0.9571
	1st Derivative 40	0.0643	0.1643	0.3357	0.6357	0.8929	0.9000	0.9429
	1st Derivative 80	0.0857	0.2357	0.6714	0.8500	0.9071	0.9571	0.9643
	2nd Derivative 20	0.0786	0.1143	0.0786	0.1714	0.2857	0.5571	0.9571
	2nd Derivative 40	0.0500	0.1357	0.1857	0.4500	0.6000	0.8786	0.9643
	2nd Derivative 80	0.0929	0.1714	0.4429	0.6929	0.8714	**0.9643**	**0.9714**
	Raw Reflectance	0.1500	0.3643	0.6500	0.7929	0.8786	0.9071	0.9071
	Smoothing 20	0.3143	0.6643	0.8500	0.8786	0.9143	0.8929	0.8929
	Smoothing 40	0.3857	0.7714	0.8571	**0.9214**	**0.9286**	**0.9143**	0.9143
	Smoothing 80	**0.4071**	**0.8000**	**0.8786**	0.9214	0.9143	0.9000	0.9071
GentleBoost	1st Derivative 20	0.0500	0.0500	0.0643	0.3500	0.5786	0.8286	0.8929
	1st Derivative 40	0.0929	0.1429	0.3929	0.6000	0.8286	0.8571	0.9071
	1st Derivative 80	0.1429	0.2429	0.6143	0.8000	0.8714	0.9143	**0.9429**
	2nd Derivative 20	0.0857	0.0286	0.1500	0.2071	0.3071	0.4643	0.8929
	2nd Derivative 40	0.0643	0.0857	0.1571	0.3286	0.5857	0.8643	0.9071
	2nd Derivative 80	0.1214	0.2000	0.4357	0.6357	0.8000	0.8929	0.9357
	Raw Reflectance	0.0500	0.0786	0.0429	0.7571	0.9429	0.9500	0.9429
	Smoothing 20	0.0286	0.0500	0.7571	**0.9429**	0.9429	0.9429	0.9429
	Smoothing 40	**0.1071**	0.0500	0.9143	0.9429	**0.9500**	0.9429	**0.9500**
	Smoothing 80	0.0643	**0.1714**	**0.9429**	0.9429	0.9500	**0.9571**	0.9429
LS-SVMs	1st Derivative 20	0.0857	0.0500	0.1000	0.0643	0.0643	0.0571	0.0714
	1st Derivative 40	0.0500	0.0929	0.0714	0.0500	0.0643	0.0643	0.0857
	1st Derivative 80	0.1000	0.0429	0.0929	0.1000	0.0714	0.0571	0.1000
	2nd Derivative 20	0.0429	0.0714	0.0786	0.0857	0.1071	0.0643	0.0714
	2nd Derivative 40	0.0857	0.1071	0.0571	0.0571	0.0857	0.0714	0.0571
	2nd Derivative 80	0.0357	0.0643	0.0643	0.0714	0.0714	0.0357	0.0500

eventual post-processing phases; in case its results need to be combined with other sensors for example.

4. CONCLUSIONS

In this paper, we evaluated three machine learning approaches for hyperspectral classification of rock types. LogitBoost, GentleBoost and LS-SVMs are able to classify rocks with a high degree of accuracy even when dealing with samples that present high spectral similarity, which can be challenging for conventional remote sensing methods based on spectral similarity measures. The classification problem is rendered difficult by the high dimensionality (429 features) and small number of labeled samples (10 per class). In our experiments considering noisy, filtered and derivative transformed data sets, LogitBoost presented better performance overall. Smoothing the data improved accuracy for all algorithms. Derivative analysis improved accuracy of the Boosting methods on high signal-to-noise ratio data sets, but the results deteriorate as noise increases. Therefore, while smoothing and derivatives can improve classification accuracy of the Boosting algorithm, the derivative transformation should only be applied to spectra presenting a high SNR.

Despite the fact that the LS-SVM approach provided higher accuracies for unprocessed reflectance data and for data with low noise, both Boosting algorithms yielded better results on the noisiest data sets and on the derivatives using a fixed number of weak learners. While the LS-SVM parameters require fine tuning for each case, Boosting is more flexible to classify different transformations of hyperspectral data.

5. REFERENCES

[1] H. M. Rajesh, "Application of remote sensing and GIS in mineral resource mapping - An overview," *J. Mineral. Petrol. Sci.*, vol. 99, no. 3, pp. 83–103, 2004.

[2] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis, An Introduction*, Springer-Verlag, New York, fourth edition, 2005.

[3] 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," *J. Geophys. Res.*, vol. 108, no. E12, pp. 5.1–5.44, 2003.

[4] P. M. Mather, *Computer Processing of Remotely-Sensed Images, An Introduction*, John Wiley & Sons, Chichester, 2004.

[5] C. Kwan, B. Ayhan, G. Chen, J. Wang, B. Ji, and C.-I Chang, "A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents,"

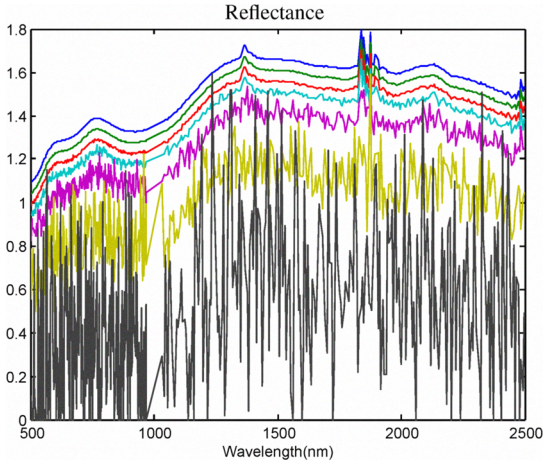


Fig. 3: Spectral signatures of a goethite sample with different SNRs. From bottom to top, the spectral curves correspond to SNRs of 10 db, 20 db, 30 db, 40 db, 50 db, 60 db, and the original raw reflectance. The curves are shifted up by a constant for visualization purposes

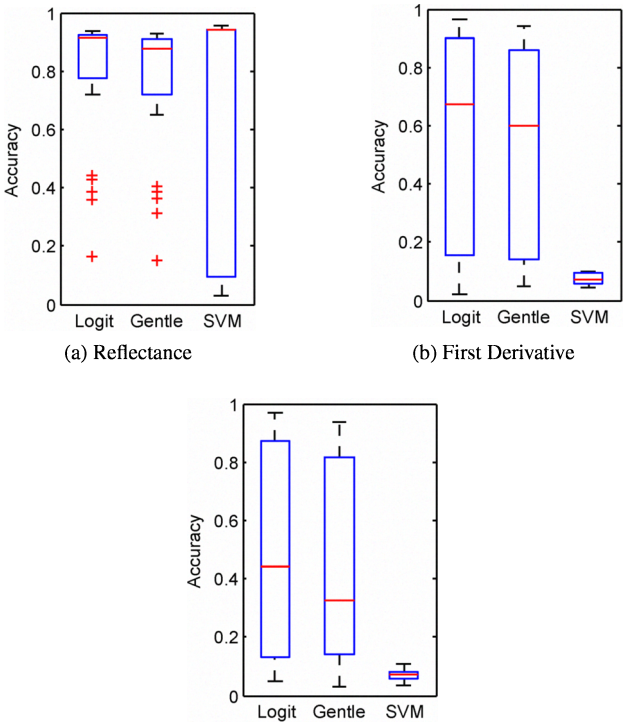


Fig. 4: Box-and-whisker diagrams comparing the results of LogitBoost, GentleBoost and LS-SVMs grouped by type of data set

IEEE Trans. Geosci. Rem. Sens., vol. 44, no. 2, pp. 409–419, 2006.

[6] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Intl. Conf. Mach. Learn.*, 2006, pp. 161–168.

[7] K. Crammer and Y. Singer, "On the learnability and design of output codes for multiclass problems," *Mach. Learn.*, vol. 47, no. 2–3, pp. 201–233, 2002.

[8] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, 2000.

[9] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Intl. Conf. Mach. Learn.*, 1996, pp. 148–156.

[10] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *Ann. Stat.*, vol. 28, no. 2, pp. 337–407, 2000.

[11] Antonio Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 29, no. 5, pp. 854–869, 2007.

[12] B. Schlkopf and A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[13] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 2nd edition, 2000.

[14] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.

[15] S.T. Monteiro, Y. Minekawa, Y. Kosugi, T. Akazawa, and K. Oda, "Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery," *ISPRS J. Photogram. Rem. Sens.*, vol. 62, no. 1, pp. 2–12, 2007.

[16] A. M. Filippi, "Derivative-neural spectroscopy for hyperspectral bathymetric inversion," *Prof. Geograph.*, vol. 59, no. 2, pp. 236–255, 2007.

[17] S. J. Orfanidis, *Introduction to Signal Processing*, Signal Processing. Prentice-Hall, Englewood Cliffs, 1996.

[18] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[19] F. Tsai and W. Philpot, "Derivative analysis of hyperspectral data," *Rem. Sens. Env.*, vol. 66, pp. 41–51, 1998.

[20] H. H. Madden, "Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data," *Anal. Chem.*, vol. 50, no. 9, pp. 1383–1386, 1978.

[21] Gustavo Camps-Valls and Lorenzo Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Rem. Sens.*, vol. 43, no. 6, pp. 1351–1362, 2005.