Comparative study of classification methods for surficial materials in the Umiujalik Lake region using RADARSAT-2 polarimetric, Landsat-7 imagery and DEM Data

Fan Li, David A. Clausi, Alexander Wong

Abstract

A study focusing on the classification of surficial materials in the Umiujalik Lake area using multisource data including polarimetric SAR data, Landsat optical data, and DEM has been conducted. The purpose of this study is to explore improving classification performance by comparing different feature combinations and different classifiers. First, four classification methods were compared on different combinations of features of intensity and texture. Second, the effects of dimension reduction algorithms for classification were investigated. Finally, six different dimension reduction methods were used to see if they can improve or remain classification performance by using fewer dimensions. Results show that adding texture features can help improve classification accuracy; the best classification accuracy is achieved by rotation forest classification method using the combination of intensity and texture features; the classification performance remains stable using fewer features.

1. Introduction

The Geological Survey of Canada of Natural Resources Canada has launched a program named Geomapping for Energy and Minerals (GEM). This program aims at providing detailed geospatial information to assist the northern resource development in the Canadian Arctic and to minimize the cost of field work (Shelat et al., 2012). The current study follows from that of Shelat et al. (2012) that addressed the need for more accurate bedrock and surficial geology maps of the northeast Thelon River region.

Optical data can measure primarily the reflective properties of the surficial materials in visible, near infrared, and mid-infrared bands. Synthetic aperture radar (SAR) images have several advantages for surficial material mapping. SAR data can measure the geometrical structure, surface roughness, and moisture content of the target. Different combinations of incidence angle, polarization, and frequency have been used in several previous studies (Baghdadi et al., 2006; Srivastava et al., 2009). Short wavelength visible bands in optical data sets are often used for distinguishing deep and shallow water, and it has been demonstrated that the cross-polarization SAR data have good capability of recognizing both shallow areas and water areas with high sediment loads (Shelat et al., 2012).

In the pattern recognition and data mining fields, there is no classification approach that guarantees the best results for all types of data. The appropriate classifier should be selected based on the properties of data such as data distribution, number of features, and linear/nonlinear separability. This study presented an assessment of the use of RADARSAT-2 polarimetric data, Landsat multi-spectral images, and a DEM for the classification of surface types. The study first analyzed data characteristics using visualization techniques, and then compared the performance of different classifiers using different combinations of features.

2. Background

2.1. Classifier comparisons in remote sensing

There are numerous papers applying different classification techniques to remote sensing data sets. Foody and Mathur (2004) evaluated the performance of the support vector machine (SVM) classifier by comparing with other classification methods such as discriminant analysis, decision tree, and multilayer perceptron neural network for imagery acquired by an airborne thematic mapper (ATM). The most accurate classification was derived from the SVM approach, and, with the largest training set, the SVM classification was significantly more accurate than the other classifiers. Pal (2005) compared the performance of SVM and the random forest (RF) method using Landsat Enhanced Thematic Mapper Plus (ETM+) imagery of an area in the UK with seven different land covers. The results show that SVM and RF have similar performance in terms of classification accuracy and training time. Gislason et al. (2006) compared the accuracy of RF to other ensemble methods including bagging and boosting using multisource data including Landsat Multispectral Scanner (MSS), elevation, slope, and aspect data. The authors found that RF is comparable to bagging and boosting in terms of classification accuracy, but RF is much faster in training and does not overfit the data. Zou et al. (2010) proposed two strategies to select features extracted from polarimetric SAR data by various decomposition algorithms including manual selection and a simple criterion, and used extremely randomized clustering forests (ERCFs), a random forest algorithm for classification. Results show that the ERCFs achieve competitive performance while requiring less time for training and testing.

Generally speaking, SVM and RF are two state-of-the-art classification techniques in recent years. They have been successfully applied for remote sensing data and achieved comparable classification accuracy in most of data sets. In our paper, we compared both methods with an ensemble classification technique, i.e., the rotation forest algorithm (RoF) (Rodriguez et al., 2006) which has been demonstrated to be better than random forests and other ensemble classifiers for classification of hyperspectral imagery (Xia et al., 2013). Also, we not only used intensity, but also texture as well as the integration of both of them as features.

2.2. Dimension Reduction Methods

Dimension reduction techniques are usually required when the number of dimensions is high. In remote sensing, these techniques are mainly applied for hyperspectral imagery (Chen and Qian, 2008; Jimenez and Landgrebe, 1999; Plaza et al., 2009). Apart from these, there is also literature using dimension reduction methods for SAR imagery. Zou et al. (2010) used a heuristic automatic feature combination technique to select PolSAR features. It iteratively selects features with highest score measured by a selection metric. Haddadi et al. (2011) proposed a feature selection method using a combination of a genetic algorithm and an artificial neural network to select features extracted from polarimetric SAR data by various decomposition algorithms. Tu et al. (2012) chose fourteen groups of polarimetric signatures to construct a high-dimensional polarimetric feature space, and used a method based on Laplacian eigenmaps for dimension reduction. We studied the effects of dimension reduction methods on the concatenated features for different classifiers in our paper.

3. Data

3.1. Study Area

The study area is the same as in Shelat et al. (2012). The area is located in Kivalliq Region of Nunavut north of the Thelon River. The geographic boundary for the study area is between $65°22'$ and $65°41'$ North and $97°31'$ and $98°07'$ West. The north and north-eastern part of the study area is mainly covered by numerous medium-to-large lakes and ponds with relief close to 160 m, and the largest lake is the Umiujalik Lake. The southern part of the study area is mainly covered by glacial till interspersed with bedrock and boulder exposures with maximum relief of $240m$. The study is located within the Keewatin ice divide of the Laurentide ice sheet. Glacial material, including drumlinoid ridges, eskers, Rogen moraines, hummocky terrains, crag and tail structures, and striated surfaces (Thomas, 1981), are deposited along northerly trending drainage systems to the Umiujalik Lake area. Thus, the surficial materials in the study area include silty and gravelly sand (till), ice contact and fluvial deposits (sand and gravel), silt and clay, till veneer, and exposed bedrock surfaces.

3.2. Data

Three RADARSAT-2 single look complex (SLC) fine quad (FQ) polarimetric SAR imageries have been acquired for this study area during August and September of 2009. A detailed description can be found elsewhere (LaRocque et al., 2012; Shelat et al., 2012). The FQ image sets with medium incidence angle (FQ12) achieved best overall classification accuracy of 48.7% (Shelat et al., 2012). Thus, here we also focused our testing on the FQ12 image acquired on August 28, 2009. This polarimetric image has an incidence angle of $31.37°$ to $33.04°$ from near range to far range. The swath width is $25km$. The range and azimuth pixel spacing are $4.73m$ and $4.97m$, respectively. The nominal resolution is $10m$ in the near range and $9.5m$ in the far range. The local acquisition time is $13:09$.

Landsat-7 Enhanced Thematic Mapper were acquired in North American Datum 83 (NAD83) Universal Transverse Mercator (UTM) Zone 14 orthorectified format over a three-year time period starting in July, 2001 (LaRocque et al., 2012). The six optical bands including TM1-5 and TM7 in $30m$ spatial resolution were used.

A $1:50000$ digital elevation model (DEM) was also used in this study to orthorectify the RADARSAT images in the preprocessing. The DEM data used was extracted from Canadian Digital Elevation Data (CDED). It has 9.5m spatial resolution in the x direction and 23.2m in the y direction. The elevation information was also used as a feature in the classification step.

3.2.1. Supporting data for validation

The training areas and validation data used in this study were obtained from (LaRocque et al., 2012; Shelat et al., 2012). They derived these data from photo-interpretation of both aerial panchromatic photographs and Landsat-7 ETM images, for which we note there may be some deviance from those products and the requirements for this study with respect to actual land cover. In the validation, training data were excluded. A total of nine surficial material classes are used for the study, namely: (i) “bedrock (BK)”, (ii) “boulders (BR)”, (iii) “organic deposits (OD)”, (iv) “sand and gravel (SG)”, (v) “till and vegetation (TV)”, (vi) ”thick till (TK)”, (vii) “thin till (TN)”, (viii) “deep water (DW)”, and (viii) “shallow water (SW)”. The map of training areas (Fig. 1) consists of disjoint homogeneous training areas, each of which is assigned to one of the nine thematic classes. More details about the training areas can be referred to the previous papers (LaRocque et al., 2012; Shelat et al., 2012). In the validation, training data were excluded to ensure the validation was performed using mutually exclusive, independent data to provide a more reliable estimate of classification accuracy for the study area.

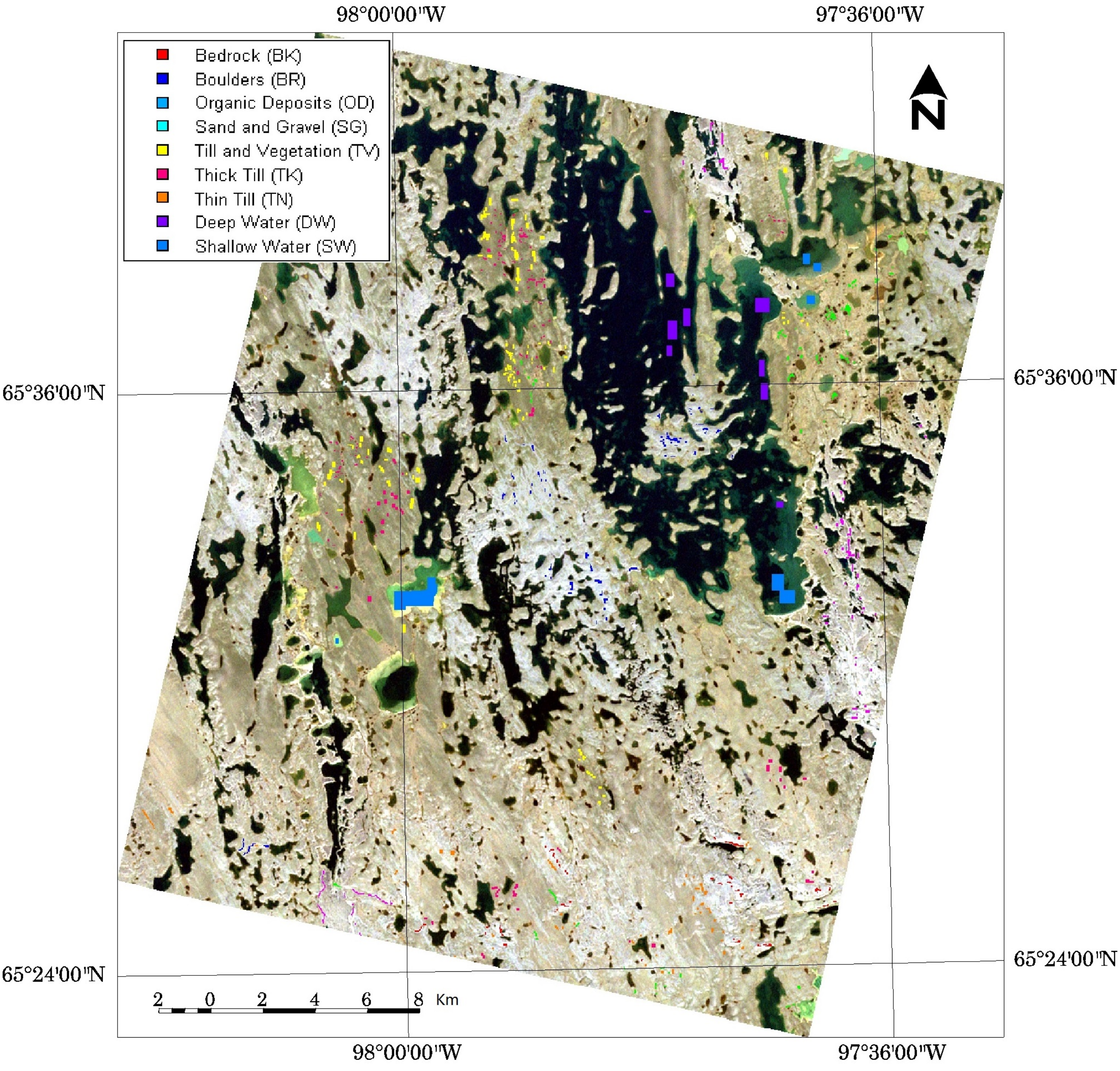

Figure 1: Training areas overlaid on optical data shown on the Landsat-7 TM 3, TM 2, and TM 1 RGB true color-composite image.

4. Research Questions and Methodology

4.1. Research Questions

Our primary objective of this study is to achieve best accuracy for surficial material classification. We focus our attention on the following three research questions:

- Question a) For different combinations of RADARSAT-2 magnitude bands, Landsat-7 bands, DEM, and their texture features, which one has the best separability? Especially, we investigated if the classification accuracy can improve by concatenating texture features to feature sets.
- Question b) Which classifier on which group of features has the best classification accuracy? We compared the following four classification methods: maximum likelihood classifier, support vector machines, random forests, and rotation forests.
- Question c) Can dimension reduction methods help improve classification accuracy? We tested on combinations of different dimension reduction techniques and different classifiers.

4.2. Methodology and Analysis

4.2.1. Feature Extraction

The features used in this research include both intensity and texture features. Intensity features include the magnitude of the four polarizations of SAR data (HH, HV, VH, and VV), the six optical bands of Landsat data, and the height of the DEM data. For texture features, we used gray-level covariance matrix (GLCM) (Haralick et al., 1973). The probabilities provide a second-order method for generating texture features which capture the correlations of pairwise pixels in the spatial window of interest (Clausi, 2002). The probability measure is defined as

$$Pr(x) = \{C_{ij} \mid (\delta, \theta)\} \tag{1}$$

where C_{ij} is the co-occurrence probability between grey levels i and j, δ is the interpixel distance and θ is the orientation. C_{ij} is defined as:

$$C_{ij} = \frac{p_{ij}}{\sum_{i,j=1}^{G} P_{ij}} \tag{2}$$

where P_{ij} is the number of occurrences of grey levels i and j, and G is the quantized number of grey levels. Common GLCM statistics include uniformity, entropy, dissimilarity, contrast, inverse difference, inverse different moment, and correlation (Clausi, 2002).

4.2.2. Data Visualization

The goal of data visualization is to have a visual interpretation of the data characteristics (e.g., distribution, separability, etc.) by projecting the data into two dimensions or three dimensions. In this paper, linear discriminative analysis (LDA) (Hastie et al., 2009) was used for visualization of data. LDA is a well-known dimension reduction method in pattern recognition that aims to maximize the intra-class variance and minimize the inter-class variance simultaneously. For a two-class problem, the objective function of LDA is

$$J(W) = \frac{\tilde{S_B}}{\tilde{S_W}} = \frac{W^T S_B W}{W^T S_W W} \tag{3}$$

where W is the transformation matrix to calculate, S_B is the intra-class covariance matrix, and S_W is the inter-class covariance matrix.

LDA can be used for visualization of high-dimensional data when the number is reduced to two or three. Compared to principal component analysis (PCA), LDA uses training data to define both intra-class and inter-class covariance matrices, so it can better reflect the class separability of the data and is more relevant to classification performance.

4.2.3. Classification

Classification of land cover types has been studied for decades in the remote sensing field. Previously a widely-used classification method was the maximum likelihood classifier (MLC) (Mather and Tso, 2003), in which the predicted label of a sample is the one that maximizes the posterior probability. However, for MLC, the number of training samples required increases exponentially as the number of feature dimensions increases. In addition, the inverse matrix of the covariance matrix becomes ill-conditioned when two features in a class are highly correlated.

Another popular classification method, support vector machines (SVM), can overcome the above limitations of MLC. The basic idea of SVM is straightforward: to maximize the margin between "support vectors" that separate different classes. In statistical learning theory, the upper bound of the expected risk can be minimized when the margin is maximized by the separating hyperplane (Vapnik, 1998). The most important reason that makes SVM popular is the "kernel trick" whereby the original data can be implicitly mapped to high-dimensional or even infinite feature space. The data which is not linearly separable in the original space can be separated by a hyperplane after it is mapped to the new space. The main disadvantage of SVM is that it cannot be directly generalized to a multi-class problem. Also, its performance decreases when there are irrelevant features (Weston et al., 2000). In practical applications, performing feature selection before using SVM might offer improvements.

Recently, random forests (RF) have been introduced in the remote sensing field(Pal, 2005)(Gislason et al., 2006). RF is an ensemble method that uses decision trees as base classifiers. The idea of ensemble methods is to combine multiple weak classifiers into one strong classifier. The predictions by weak classifiers usually have high variance, which can be reduced significantly by majority voting or averaging of the results of the weak classifiers. The key of an effective ensemble method is to lower the correlation of the results by each classifier while keeping the results unbiased. Compared to bagging (or bootstrap aggregating), random forests use a random feature subspace method which can make the results in each classifier more diverse. RF is robust for high-dimensional data even when the data are noisy. Also, RF can be inherently used for multi-class classification problems. The main disadvantage of RF is, unlike MLC and SVM, it does not have a statistical or geometric meaning, and the results are thereby difficult to interpret.

The rotation forest (RoF) algorithm (Rodriguez et al., 2006) is another type of ensemble methods. The base classifier can be any weak classifier that has low bias and high variance. Bootstrap sampling is transformed by a rotation matrix R_i (4) before using the weak classifier in order to increase diversity.

$$R_i = \begin{bmatrix} pc_{i,1}^{(1)}, pc_{i,1}^{(2)}, \cdots, pc_{i,1}^{(M_1)} & [0] & \cdots & [0] \\ [0] & & & \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \cdots & pc_{i,K}^{(1)}, \cdots, pc_{i,K}^{(M_K)} \end{bmatrix} \quad (4)$$

where $pc_{i,j}^{(1)}, \cdots, pc_{i,j}^{(M_j)}$ are coefficients of the principal components obtained from training samples. By introducing the rotation matrix, both accuracy and individual accuracy can be improved.

4.2.4. Dimension Reduction for Classification

Dimension reduction methods include feature transformation and feature selection methods. Feature transformation means performing linear or non-linear transformation on the feature space to generate new features, sometimes in lower dimension space. For redundant data, feature transformation methods are helpful because the intrinsic dimensionality of the data is typically less than the number of original features. For noisy data, feature transformation methods have the risk of mixing irrelevant features with strong features when creating new features.

In remote sensing, PCA is a popular dimension reduction method (Farrell and Mersereau, 2005) due to its simplicity and robustness. However, PCA is sensitive to scaling because it is based on global variance, and the variance is not directly related to classification. LDA uses label information of training data and is itself a classification method, but the number of dimensions has to be smaller than the number of classes due to the rank of the intra-class covariance matrix. A recently developed dimension method called "local Fisher discriminant analysis (LFDA)" (Sugiyama, 2007) can overcome this limitation by reformulating the objective function of LDA. This method has been successfully applied to hyperspectral data (Li et al., 2012).

Feature selection attempts to remove redundant or irrelevant features that might reduce classification performance. Feature selection methods can be grouped into wrapper methods, filter methods, and embedded methods (Guyon and Elisseeff, 2003). For SVM, the recursive feature elimination (RFE) algorithm (Guyon and Elisseeff, 2003) is often used for feature selection. RFE is a wrapper method that selects or eliminates features based on a predictive model. RFE uses the SVM model by eliminating the features with the smallest weights iteratively. However, RFE which can remove irrelevant features is incapable of removing redundant features because the weights of redundant features are almost the same (Xie et al., 2006). Another popular wrapper method is forward selection (FW) (Guyon and Elisseeff, 2003). It is a greedy algorithm in which features that improve the classification accuracy most are selected in each iteration until the classification accuracy is no longer improved. FW can avoid redundant features to be selected, but the classification should be performed as many times as the number of features in each iteration. So FW is not as efficient as RFE when the feature dimension is very high.

Compared to MLC and SVM, RF is robust to both redundant features and irrelevant features (Pal, 2005). RF is also an embedded method by itself because the feature importance generated in the learning step can be used for feature selection. A common method of performing feature selection for random forests is RFE (Chehata et al., 2009) which is similar to RFE using SVM. RoF has similar properties to RF. Also, RoF performs feature transformation without dimension reduction implicitly using the rotation matrix. However, there is no known study with respect to how dimension reduction affects the classification performance by RoF.

5. Tests and Analysis

5.1. Data Preparation

For this study, half of the homogeneous training area samples were randomly selected to be training samples and the other half used as test samples. This ensures that a pixel and its neighbors belong to the same training or test samples; otherwise, there would be correlation between texture features of training and test samples and thus the classification accuracy would probably be over-estimated. The number of training and test samples for each class is shown in Table 1. All feature values were normalized to between 0 and 1. GLCM texture features were extracted from all Landsat bands, RADARSAT-2 polarizations, and DEM. The window size of GLCM was set to 15×15, and

Table 1: Number of training and test samples in each class (BK-bedrock; BR-boulders; OD-organic deposits; SG-sand and gravel; TV-till and vegetation; TK-thick till; TN-thin till; DW-deep water; SW-shallow water)

Land cover class codes	BK	BR	OD	SG	TV	TK	TN	DW	SW
Number of training samples	212	339	383	388	968	632	292	701	1044
Number of test samples	216	341	386	392	983	624	294	829	1297

the number of gray levels was set to 64. Both parameters were optimized by grid search. The offset was set to $[0, 1]$ for simplicity. The intensity and texture features were concatenated together. Statistics including contrast, correlation, and entropy were used as texture features. Texture statistics were calculated for each pixel in each intensity image.

5.2. Data Visualization Results

We performed LDA on the training data. The $n \times 2$ transformation matrix was obtained by minimizing (3). The transformation matrix was then applied to the test data to obtain two-dimensional points. The visualization using different combinations of features is shown in Fig. 2. For ease of viewing, a subset of 1000 samples were randomly selected from the test samples to plot. We can see that the visualization result is consistent with the class separability result in Shelat et al. (2012). In Fig. 2 (a), water classes ("deep water" and "shallow water"), vegetation classes ("thick till", "thin till", "till and vegetation", and "organic deposits"), and non-vegetation classes ("bedrock", "boulders", and "sand and gravel") range approximately from the left side to the right side. However, the separability within those classes is low. For the non-vegetation classes, samples are fully mixed in the two-dimension plot.

In Fig. 2 (b), we can see the gap between three clusters (water classes, vegetation classes, and non-vegetation classes) is wider when using the Landsat data. Also, there is an evident gap between shallow water and deep water. In the right part of the plot, the samples are roughly more vegetated from top to bottom. For the non-vegetation cluster in the top-right side of the plot, boulder samples are approximately on top of the other non-vegetation classes, but the other two classes are still mixed together. For the vegetation cluster in the bottom-right, the sub-classes are less separable than (a).

The improvement of class separability is not significant when combining RADARSAT magnitude bands and Landsat bands with DEM, or combining both intensity and texture features, as shown in Fig. 2(c) and (d). However, data information is typically removed when mapping multi-dimensional feature sets into only two-dimension space for visualization, which might affect the class separability. Nevertheless, we also need to investigate classification performance to evaluate the class separability of the features.

5.3. Classification

For SVM, the LibSVM library (Chang and Lin, 2011) was used for testing. Grid search was performed to find the optimal set of parameters (C and γ) for each combination of features. Radial basis function (RBF) kernel was used for testing. RF, RoF, and MLC were all implemented in Matlab. The standard decision tree function was used in RF and RoF. The number of trees in RF was set to 500. The number of variables to select for each decision split was set to the square root of the number of variables. For RoF, the the number of trees was set 100, and LDA was used as the feature transformation method to obtain the rotation matrix. The number of features in the subspace was three, and eight classes were eliminated for each bootstrap data to achieve the largest diversity.

The classification results of four classifiers are shown in Table 2. For Landsat intensity bands, MLC is comparable to SVM, RF, and RoF. The classification accuracy increases by including GLCM texture features. For RADARSAT-2 imagery, the accuracy can increase by 10.4% for all the classifiers after including GLCM texture. The incorporation of DEM can also increase the overall classification accuracy, for example, from 85.5% to 89.6% using the random forest classifier, indicating that the height information can be a good feature for discriminating some land cover types. When combining both intensity and texture features, however, the performance of MLC drops greatly because of the curse of dimensionality. Classification accuracy by SVM also drops mildly potentially because irrelevant bands are included. In contrast, the overall accuracy by both RF and RoF increase with increasing features. The highest classification accuracy among all combinations of features is obtained by RoF (94.4%) while using the combination of RADARSAT quad-pol scenes, Landsat bands, the DEM, and the corresponding GLCM texture statistics. It has higher classification accuracy than SVM and RF by 1.4% and 2.1% respectively. When the number dimensions are larger than 10, the RoF algorithm outperforms the second best classifier in five out of six tests using different feature combinations, which indicates its constantly good capability of classifying high-dimensional data. Among these tests, RoF on the combination of Landsat-7 bands and their GLCM features achieves the classification accuracy 84.7%, 5.0% higher than the next best classifier, i.e., the SVM classifier.

The final classification maps by RoF are shown in Fig. 3. Generally speaking, the labels become smoother when texture features are used. Smooth labeling is often considered as an improvement by overcoming the within-class variation with the aid of spatial information, while over-smoothing might tend to lose details, especially land covers that are innately scattered. Comparing two classification maps in Fig. 3, the most significant difference between

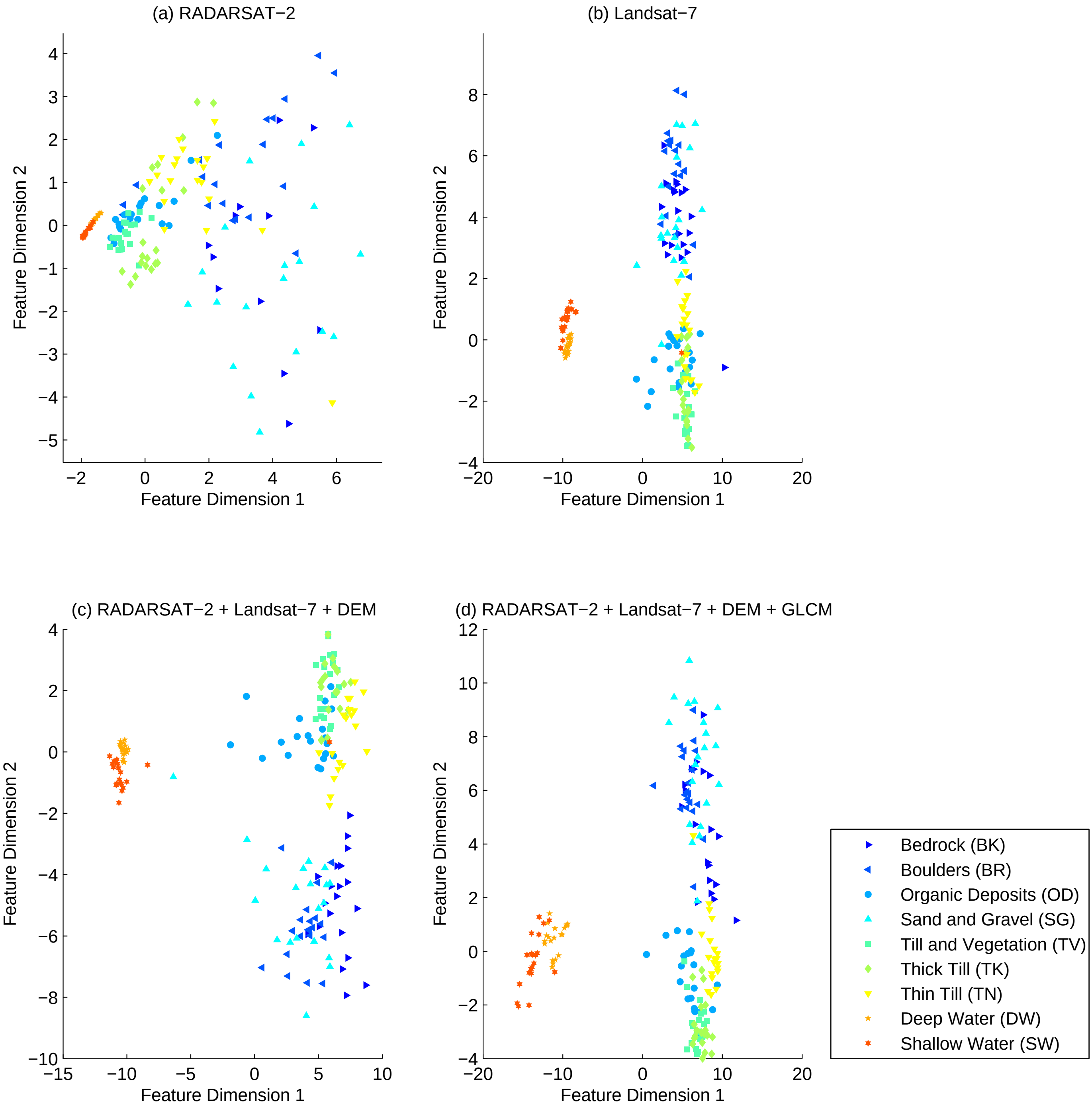

Figure 2: Projection of multi-dimensional feature sets into 2-dimensions using LDA using (a). RADARSAT-2 magnitude data; (b) Landsat-7 data; (c) Combination of RADARSAT-2 manitude, Landsat-7, and DEM data; (d) Combination of RADARSAT-2 manitude, Landsat-7, DEM data, and their corresponding texture features.

Table 2: Overall accuracy and corresponding Kappa. "GLCM" means GLCM features (entropy, contrast, and correlation) for the corresponding intensity bands. The number of dimensions of each combination is in the parentheses.

Feature Combinations	MLC	SVM	RF	RoF
RADARSAT-2 bands (RAD) (4)	69.1% (0.64)	72.8% (0.68)	**73.4%** (0.69)	70.2% (0.65)
Landsat-7 bands (LAN) (6)	76.7% (0.73)	**77.7%** (0.74)	74.9% (0.70)	73.2% (0.68)
RAD + LAN (10)	86.2% (0.84)	88.4% (0.86)	85.5% (0.83)	**88.7%** (0.87)
RAD + LAN + DEM (11)	86.5% (0.84)	91.8% (0.90)	89.6% (0.88)	**92.2% (0.91)**
RAD + GLCM (16)	74.7% (0.70)	**85.6%** (0.83)	82.7% (0.80)	84.1% (0.81)
LAN + GLCM (24)	69.8% (0.65)	79.7% (0.76)	77.6% (0.74)	**84.7% (0.82)**
RAD + LAN + GLCM (40)	66.4% (0.62)	92.6% (0.91)	91.2% (0.90)	**93.6% (0.93)**
RAD + LAN + DEM + GLCM (44)	55.6% (0.50)	93.0% (0.92)	92.3% (0.91)	**94.4% (0.93)**

two maps is that the sand and gravel and organic deposits classes which are fragmented in the left result turn into larger homogeneous regions after texture features are used, as shown inside the red circles. Even though the classification accuracy using texture features (94.4 %) is higher than that without using texture features (92.2%), it is not appropriate to assert that the right result is better than the left one. It is better to make additional validation on those regions before determining the feature combination for final classification.

5.4. Effects of Dimension Reduction Methods to Classification Accuracy

We test three feature transformation methods (PCA, LDA, and LFDA) and three feature selection methods: one FW method using SVM and two RFE methods using SVM and RF respectively on the combination of 44 features including intensity and texture, using SVM, RF, and MLC. The same data set as that in Section 5.3 is used for testing. The overall classification accuracy (OA) by different methods from 2 to 44 dimensions can be seen in Fig. 4.

For MLC, feature transformation methods produce higher classification results than using the full feature set. LDA is better than PCA and LFDA. However, at two dimensions, the classification accuracy of LDA is the highest for all the DR methods. This shows evidence for using LDA instead of other dimension reduction methods for visualization in our research. Note that LDA can be no more than eight dimensions because there are total of eight classes. Moreover, feature selection methods such as FW and RF-RFE constantly perform worse than feature transformation methods. The highest classification accuracy (87.5 %) is achieved by PCA at the dimension of 17.

For SVM, however, two feature selection methods, i.e., FW and RF-RFE, achieve higher classification accuracy and stability than other dimension reduction methods. Using the forward selection method, the overall accuracy of using only six features (91.5%) is close to using all the features (92.3%). If we wish to select features that can outperform compared to using the whole feature set, RF-RFE will be a good choice because the accuracy using 26 features (93.1%) is slightly better than using all the features. However, when the dimension is reduced to a small number, both RF-RFE and SVM-RFE will mistakenly remove useful features because they are not able to detect redundant features. When the number of dimensions grows to more than 28, the classification accuracy by most of the dimension methods except LFDA has no much difference.

Compared to SVM, RF and RoF have fewer fluctuations in classification accuracy compared to RF and RoF with increasing feature space dimensionality. Their performances are stable when the dimension is reduced to more than ten, no matter what dimension reduction method is used. However, there is little increase using fewer dimensions over using the original features. When the number of reduced dimensions is less than ten, forward selection is better than other dimension reduction methods for these two classifiers. The overall classification accuracy achieves 92.63% by RoF when only six dimensions are used.

Comparing results of all the classifiers, RoF using RADARSAT-2 polarizations, Landsat intensity, and the texture features of them can achieve the highest accuracy (94.4%), and RoF tends to have higher classification rates compared to other classifiers when increasing the number of features. Also, the performances of RF and RoF are both robust to different dimension reduction techniques.

6. Conclusions and Future Work

In this paper, the classification of surficial material in the Umiujalik Lake in Nunavut using RADARSAT polarimetric and Landsat images, and DEM data was explored. We first investigated the separability of data using data visualization techniques. Then we tested four classification methods on different combinations of features. Finally, we tested the effect of different dimension reduction methods on classification accuracy.

Experiments showed that texture features can help improve classification accuracy, especially for the RADARSAT-2 polarimetric data. The highest classification accuracy is achieved using rotation forests on the combination of intensity and texture features of RADARSAT magnitude, Landsat intensity bands, and DEM. Feature transformation methods (LDA, PCA, and LFDA) are capable of improving the classification accuracy of MLC by reducing

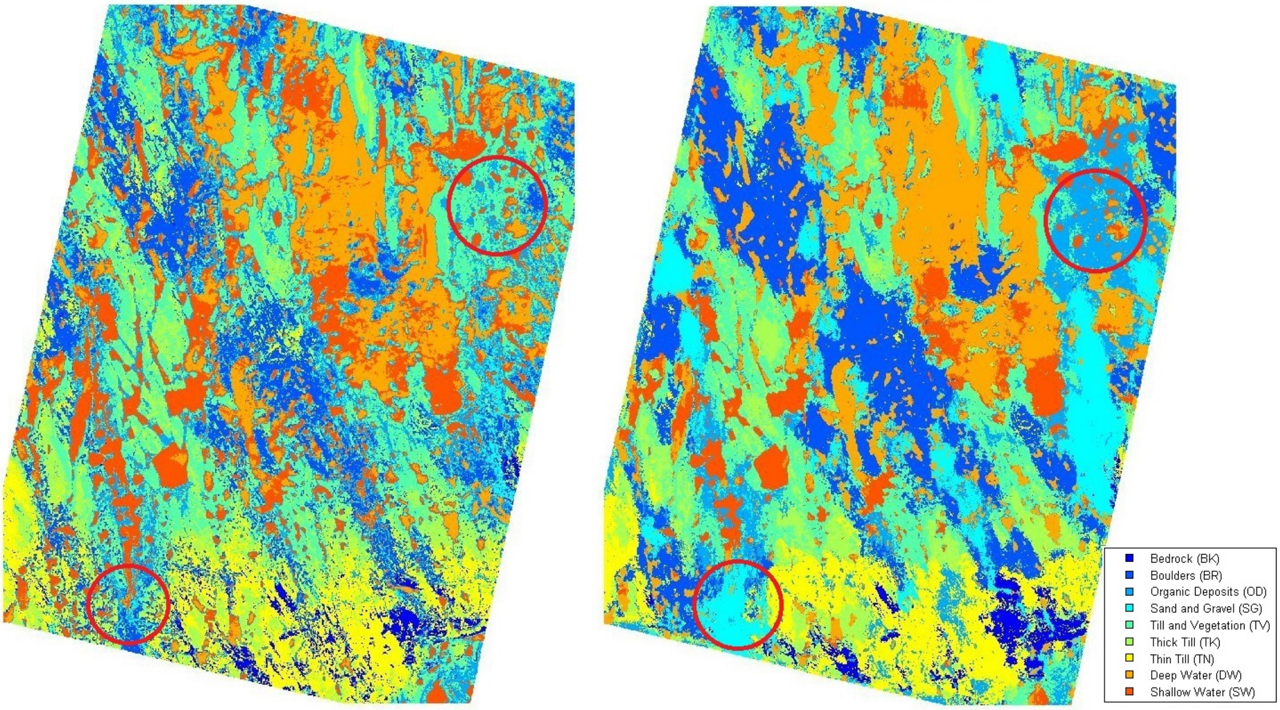

Figure 3: Classification map by the rotation forest algorithm. The left result uses RADATSAT-2 magnitude, Landsat-7 intensity and DEM bands (11 bands), and the right one uses both intensity and GLCM texture bands (44 bands).

the “curse of dimensionality”, but the classification accuracy is no higher than only using intensity and texture bands. For the feature selection methods, FW-SVM has very good performance for SVM classifier when we wish to only use a small number of features. SVM using RF-RFE, however, can achieve slightly higher accuracy than other methods. Compared to MLC and SVM, RF and RoF are less sensitive to different dimension reduction methods, and the classification accuracy become stable when the number of dimensions is more than ten. Future work will be focus on incorporating spatial context, such as using Markov random fields, to help further improve classification accuracy.

7. Acknowledgement

This work is partly funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Space Agency (CSA). This research was undertaken, in part, thanks to funding from the Canada Research Chairs program. We would like to thank Dr. Armand Larocque for providing the supporting data for validation, and thank Kara Webster for providing the RADARSAT-2 imagery. We also thank anonymous reviewers for their careful reading and helpful remarks.

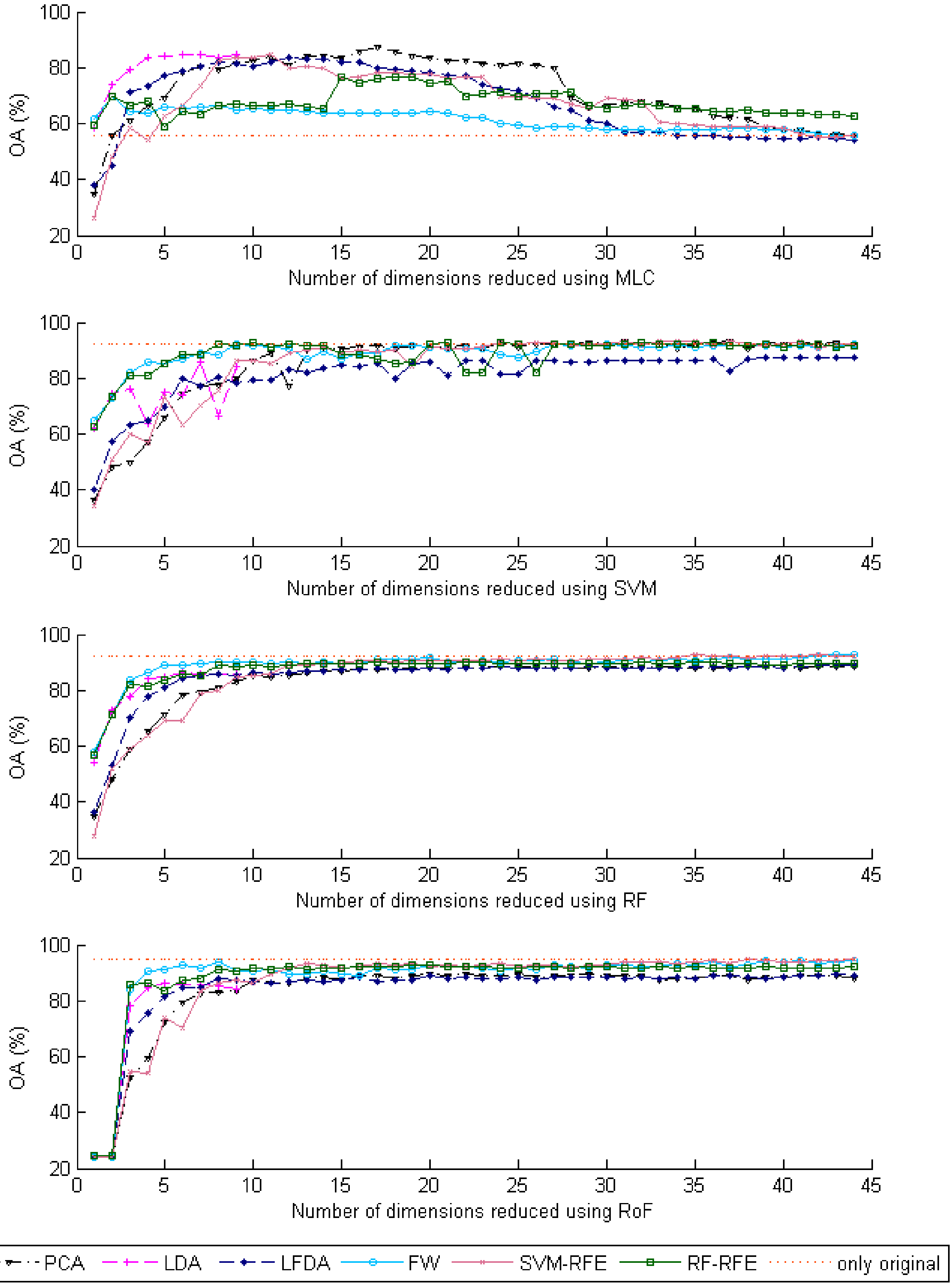

Figure 4: Overall classification accuracy (OA%)of MLC, SVM, RF, and RoF (top to bottom) using different dimension reduction methods as a function of the number of reduced dimensions (2 to 44).

8. Reference

Baghdadi, N., Holah, N., and Zribi, M. Soil moisture estimation using multi-incidence and multi-polarization asar data. *International Journal of Remote Sensing*, 27(10):1907–1920, 2006.

Chang, C.-C. and Lin, C.-J. LibSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, 2011.

Chehata, N., Guo, L., and Mallet, C. Airborne lidar feature selection for urban classification using random forests. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(Part 3/W8):207–12, 2009.

Chen, G. and Qian, S.-E. Evaluation and comparison of dimensionality reduction methods and band selection. *Canadian Journal of Remote Sensing*, 34(1):26–36, 2008.

Clausi, D. A. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.

Farrell, M. D. and Mersereau, R. M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geoscience and Remote Sensing Letters*, 2(2):192–195, 2005.

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

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300, March 2006.

Guyon, I. and Elisseeff, A. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

Haddadi, A., Reza Sahebi, M., and Mansourian, A. Polarimetric SAR feature selection using a genetic algorithm. *Canadian Journal of Remote Sensing*, 37(1):27–36, February 2011.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, (6):610–621, 1973.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. *The elements of statistical learning*, volume 2. Springer, 2009.

Jimenez, L. O. and Landgrebe, D. A. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999.

LaRocque, A., Leblon, B., Harris, J., Jefferson, C., Tschirhart, V., and Shelat, Y. Surficial materials mapping in Nunavut, Canada with multibeam RADARSAT-2 dual-polarization C-HH and C-HV, LANDSAT-7 ETM+, and DEM data. *Canadian Journal of Remote Sensing*, 38(03):281–305, 2012.

Li, W., Prasad, S., Fowler, J. E., and Bruce, L. M. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1185–1198, 2012.

Mather, P. and Tso, B. *Classification methods for remotely sensed data*. CRC press, 2003.

Pal, M. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222, January 2005.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., et al. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:110–122, 2009.

Rodriguez, J. J., Kuncheva, L. I., and Alonso, C. J. Rotation forest : a new classifier ensemble method. *IEEE Transactions on Pattern Analysis and and Machine Intelligence*, 28(10):1619–1630, 2006.

Shelat, Y., Leblon, B., Larocque, A., Harris, J., Jefferson, C., Lentz, D., and Tschirhart, V. Effects of incidence angles and image combinations on mapping accuracy of surficial materials in the Umiujalik Lake area , Nunavut , using RADARSAT-2 polarimetric and Landsat-7 images , and DEM data. *Canadian Journal of Remote Sensing*, 38(03):383–403, June 2012.

Srivastava, H. S., Patel, P., Sharma, Y., and Navalgund, R. R. Large-area soil moisture estimation using multi-incidence-angle radarsat-1 sar data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2528–2535, 2009.

Sugiyama, M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *The Journal of Machine Learning Research*, 8:1027–1061, 2007.

Thomas, R. *Surficial Geology, Amer Lake, District of Keewatin.* Geological Survey of Canada, 1981.

Tu, S. T., Chen, J. Y., Yang, W., and Sun, H. Laplacian eigenmaps-based polarimetric dimensionality reduction for SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(1):170–179, 2012.

Vapnik, V. N. Statistical learning theory. 1998.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. Feature selection for SVMs. In *Advances in Neural Information Processing Systems*, volume 12, pages 668–674, 2000.

Xia, J., Du, P., Member, S., He, X., and Chanussot, J. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Transactions on Geoscience and Remote Sensing Letters*, pages 1–5, 2013.

Xie, Z.-X., Hu, Q.-H., and Yu, D.-R. Improved feature selection algorithm based on SVM and correlation. In *Advances in Neural Networks-ISNN 2006*, pages 1373–1380. Springer, 2006.

Zou, T., Yang, W., Dai, D., and Sun, H. Polarimetric SAR image classification using multifeatures combination and extremely randomized clustering forests. *EURASIP Journal on Advances in Signal Processing*, 2010:1–10, 2010.