

A Land Cover Classification Method for Antarctica Using Support Vector Machine and Decision Tree

Ping Wang^{*,1,2}, Hui Fu¹, Zhaokun Zhai¹, Jing Wei¹, and Zhiling Zhao¹

¹School of Geomatics, Shandong University of Science and Technology, Qingdao, Shandong, 266590, P.R. China

²Key Laboratory of Marine Surveying and Mapping in Universities of Shandong, P.R. China

Abstract: Global land cover data are fundamental for applications, especially ecological environmental assessment and climate change research. Currently available global land cover data products show some deficiencies in data accuracy and spatial and temporal resolution. So we discuss fast automatic classification methods for the study area in Antarctica. A classification method based on a Support vector machine (SVM) and a decision tree (DT) model is proposed. We compare the land cover classification using four common kernel functions for a SVM. The experiment indicates that the SVM classification method using a radial basis function (RBF) leads to the optimal accuracy and running time. In view of existing phenomenon that surface features in shadow areas are easily confused, classification is further improved by using a DT model, at last a majority analysis of the above classification result removes small polygon artifacts to form the final land cover data product. The overall accuracy is 95.82%, higher than the SVM alone and the maximum likelihood method. Land cover classification in Antarctica can be conducted more reliably through our proposed classification method.

Keywords: Antarctica, classification, DT model, land cover, RBF, SVM.

1. INTRODUCTION

Land cover, a main topic in global change research, received increased international attention in recent decades [1, 2]. Efficient and precise remote sensing mapping of global land cover has become a leading technology in global change research. This development level of remote sensing science attests to the comprehensive scientific and technological strength of a country [3]. For this reason, in 2009, China's 863 Program (State High-Tech Development Plan) approved the sub-topic "Global Land Cover Remote Sensing Data Products Development" with the automatic extraction of land cover classification being the key technology.

This paper selects Antarctica as the study area for classification. Antarctica's geographical location is too remote for most people to reach. There are fewer types of land cover and other significant differences distinguishing Antarctica from other locations, and the scope of our study area is very wide, and the classification workload is very heavy. Thus, the reduced number of land cover types appropriately simplifies the classification task for Antarctica. Warren *et al.* (in 1980, 1982, and 1986) studied the optical properties of snow in Antarctica and used satellite data for glaciological research. However, the interpretation of the images was hampered by the lack of simultaneously recorded ground data [4-6]. The classification was also slow and the manual work

was time consuming. Therefore, it is necessary to study faster and more efficient classification methods.

We use a support vector machine (SVM) classification method because our number of samples is limited and SVM is less sensitive to this than other algorithms. Furthermore, it is easy to select the appropriate number of samples. The classification error can be minimized by choosing the training sample data and the kernel function to create a hyper-plane. Mantero *et al.* (2005) estimated probability densities of thematic classes using SVM. Their approach used a recursive procedure to generate prior probability estimates for known and unknown classes by adapting the Bayesian minimum-error decision rule [7]. Foody and Mathur (2006) proposed a focus on mixed pixel training samples instead of the more tedious conventional pure pixel acquisition method using an SVM classifier [8]. Unfortunately, this method does not provide advantages in spatial analysis.

Then our work builds a DT model by using human judgment and automated spatial analysis to distinguish between water and shadows. DT algorithm, a test method for variable interaction, is applied to remote sensing in various fields. Jingfeng Yu (2007) and Chen Ye *et al.* (2012) used a thematic mapper (TM) remote sensing image as the data source. They established a set of DT classification rules using the spectral characteristics and geographical features. The accuracy of their DT classifier reached 90% [9, 10]. Thoreau Rory Tooke *et al.* (2009) classified the urban vegetation characteristics for the City of Vancouver, Canada within the spectral mixing space derived from high spatial resolution Quickbird imagery and a LiDAR surface model. Their ap-

proach is based on spectral mixing and statistically developed decision trees [11]. Milap Punia *et al.* (2011) explored the potential of multi-temporal IRS P6 (Resourcesat) Advanced Wide Field Sensor (AWiFS) data for mapping of land use/land cover (LULC) for Delhi, India using DT classification. The results show that a temporal data set with well-defined training sites can result in a high overall accuracy (91.81%) as well as high individual classification accuracies [12]. Generally speaking, the DT model combining human judgment and automated spatial analysis can improve the classification accuracy.

2. DATA AND METHODS

2.1. Location and Data

Antarctica is the continent with the highest latitude, and it encompasses most longitudes. At 14 million km², it accounts for 9.4% of the world's total land area. It has the world's coldest weather and is windy and dry. The average annual temperature is -25 °C across Antarctica, and around -56°C in the inland plateau. The South Pole has reached a minimum temperature of -89.8°C. Only 2% of the entire continent is not covered by snow, and this is the area where most of the animals and plants thrive. Precipitation is rare in Antarctica - only about 30 mm annually in the interior with almost no precipitation near the South Pole. It has abundant natural resources mainly distributed in the Antarctic Peninsula and the coastal islands. Antarctica is also an important storage site of freshwater in the world. The weather supports only a few species of plants: moss, algae, and some flowering plants. We select only a particular area for land cover classification as the entire continent is very large. The study area is located within: 163°45'20"E - 166°24'23"E, 77°43'49"S - 78°40'0"S. The Enhanced Thematic Mapper Plus (ETM+) image of the study area is shown in Fig. (1).

We use Landsat 7 ETM+ images with a resolution of 30 m for land cover classification. The images were taken on

December 5th, 2002. Aerial photography data, high-resolution satellite images (Quickbird), and high-resolution land cover data with the same phase are used for sample selection and accuracy verification. These data are part of the 863 Program Project. The geometric correction and radiometric calibration for the ETM+ image are conducted by ENVI software.

2.2. Support Vector Machines

SVM is an artificial intelligence algorithm that has been used for many years in the field of remote sensing. It is based on statistical learning theory and was developed by Vapnik [13]. Its many advantages include computational effectiveness, robustness, and statistical stability. These advantages are unique to SVM when solving problems with limited samples, or applying nonlinear and high-dimensional pattern recognition. Thus, SVM has been widely applied to the field of pattern recognition and classification. SVM is meant for binary classification but multi-class problems are common in practical applications. Therefore, extending SVM to multi-class classification has become a popular research.

Based on the structural risk minimization criterion, SVM tries to improve the generalization ability of the classifier by minimizing the classification error of the training samples [14]. Low-dimensional space vector sets are often difficult to divide. The solution is to map them to a higher dimensional space, then construct a hyper-plane as the decision surface. That is, we can increase the separation using the kernel function.

Different kernel functions lead to essentially different SVM algorithms. Thus, the kernel function crucial to the classification result. Currently, the commonly used kernel functions are the linear kernel function (formula 1), polynomial kernel function (formula 2), radial basis function (RBF) kernel function (formula 3), and the sigmoid function (formula 4).

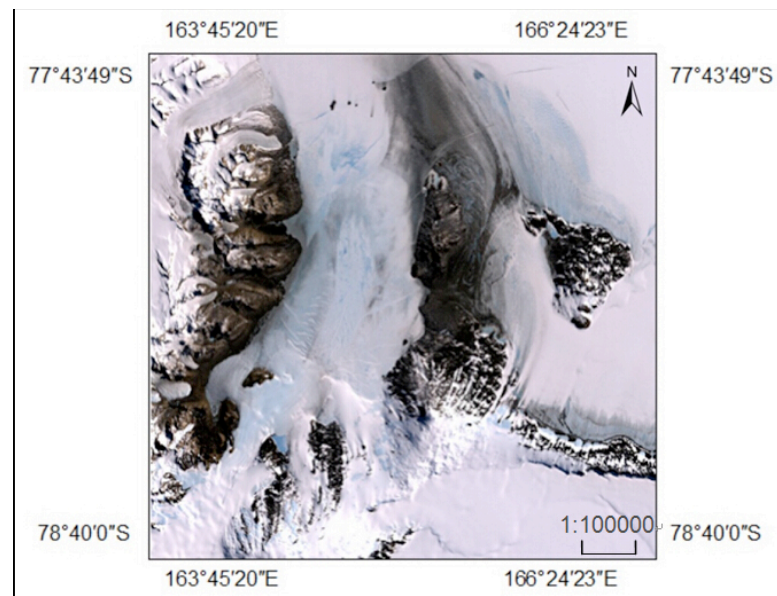


Fig. (1). ETM+ image of the study area.

(1) Linear kernel function

$$K(x_i, x_j) = (x_i \bullet x_j) \tag{1}$$

(2) Polynomial kernel function

$$K(x_i, x_j) = [\gamma(x_i \bullet x_j + coef)]^p \tag{2}$$

(3) RBF kernel function

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{3}$$

(4) Sigmoid function

$$K(x_i, x_j) = \tanh(\gamma(x_i, x_j) + coef) \tag{4}$$

where, $x_i \in R^n$ is the sample value, $x_j \in \{-1,1\}$ is the class value, γ is the width parameter between two classes, p is the degree of polynomial, and *coef* is the bias coefficient.

2.3. Decision Trees

DT is an approach to generate decision rules based on training samples and then classifying new data by

using the DT model and rules. This paper mainly uses a binary tree structure comprising a root node, a series of internal nodes, and leaf nodes. The root node is divided into two sub-nodes with optimal separation rules. The leaf nodes of DT correspond to the class attribute values. Different leaf nodes can correspond to the same class attribute value.

DT analysis is a promising classification approach complementing basic remote-sensing analysis to produce increased accuracy and efficiency. In our study, the shadows and the water are difficult to distinguish using SVM. We thus combine the spectral characteristics and the decision models and rules through comprehensive analysis and judgment.

2.4. Experimental Procedure

Our experimental procedure includes ETM+ image pre-processing (geometric correction, radiometric correction), training sample selection, SVM kernel function selection, SVM classification based on RBF, reclassification based on a DT model, majority analysis, and accuracy evaluation. The main steps are presented in detail in Fig. (2).

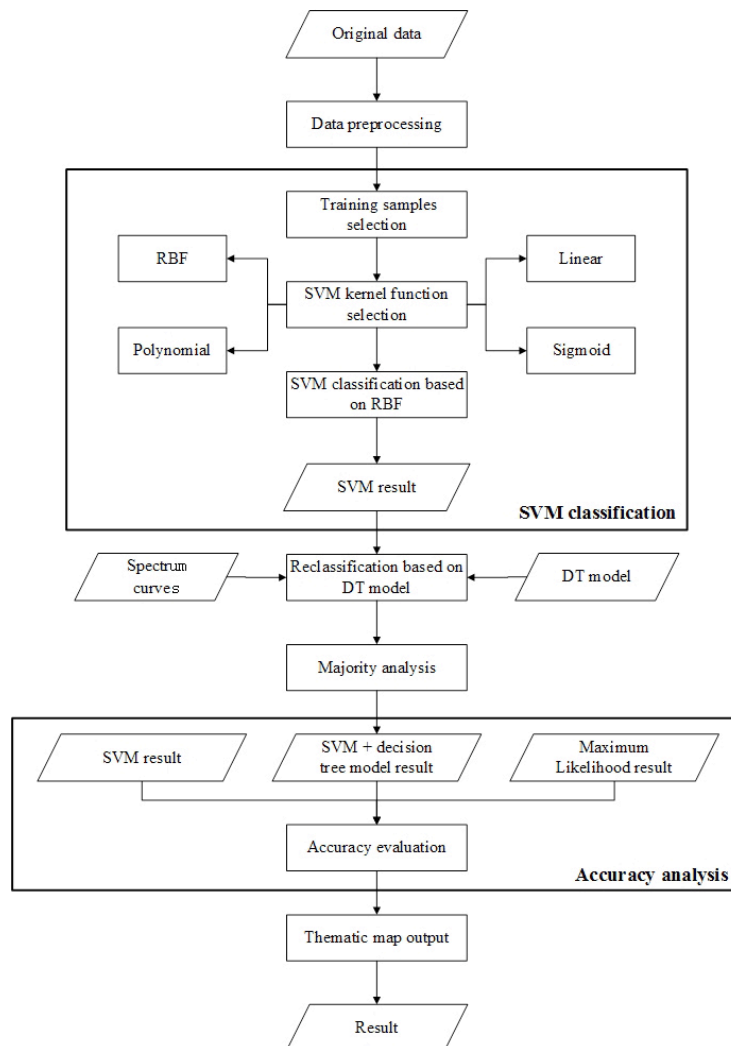


Fig. (2). Flow chart of the classification experiment.

2.5. Selection of Training Samples

Antarctica is located at high latitudes, and the surface is covered by snow and ice causing a high reflectance. The Landsat satellites are polar-orbiting satellites. This causes some particular characteristics in the Antarctica Landsat ETM+ images, including a high overlapping rate and a great disparity in gray levels between images. We propose a land cover classification system in the study area according to the actual landform condition in Antarctic and the current domestic and international classification systems for large-area land cover. Thus, the land cover in the study area is grouped into four classes: ice/snow, water, rock/soil, and vegetation. The study area labels are established based on the above classification system through image interpretation.

Image interpretation labels the target based on basic interpretation elements. Remote sensing image interpretation has eight basic elements: tone or color, shadow, size, shape, texture, pattern, site, and association. The targets in the image can then be labeled according to the chosen interpretation elements, the date of the image capture, the geographical area, and the type of image. For example, glacier has high gray value and its texture is not obvious. The bare land and glacier in the shadow area can be distinguished according to the tone and texture. The interpretation signs on different ETM+ synthetic images for the study area are as shown in Fig. (3).

The training and testing samples in the study area are extracted according to the above interpretation signs corresponding to target color. In this study, a total of 2661 samples are selected. Then, 1344 of these samples are used as training samples and the remaining 1317 as test samples for accuracy evaluation (Table 1).

2.6. Selection of SVM Kernel Function

There is no standard stipulating the selection of the kernel function. We select a 600 by 600 pixel experimental area (Fig. (4)), and conduct the remote sensing image classification with different kernel functions and parameters. We then select the optimal kernel function and its parameters based on the resulting accuracy. The key parameters of the kernel function are the degree of the polynomial (polynomial only), the bias coefficient (polynomial and sigmoid), the penalty factor, and the width of the kernel function. In this paper, values are restricted to {1,2,3}, and are set to {-1,0,...,3} and {10⁻⁴,10⁻³,...,10⁴}, respectively. can be obtained through cross validation of the training sample. The accuracy increases at first and then decreases when varied over its range. The accuracy is highest when = 100. We thus acquire the optimal parameters and their corresponding accuracies with the four types of kernel functions. The results are shown in Table 2.

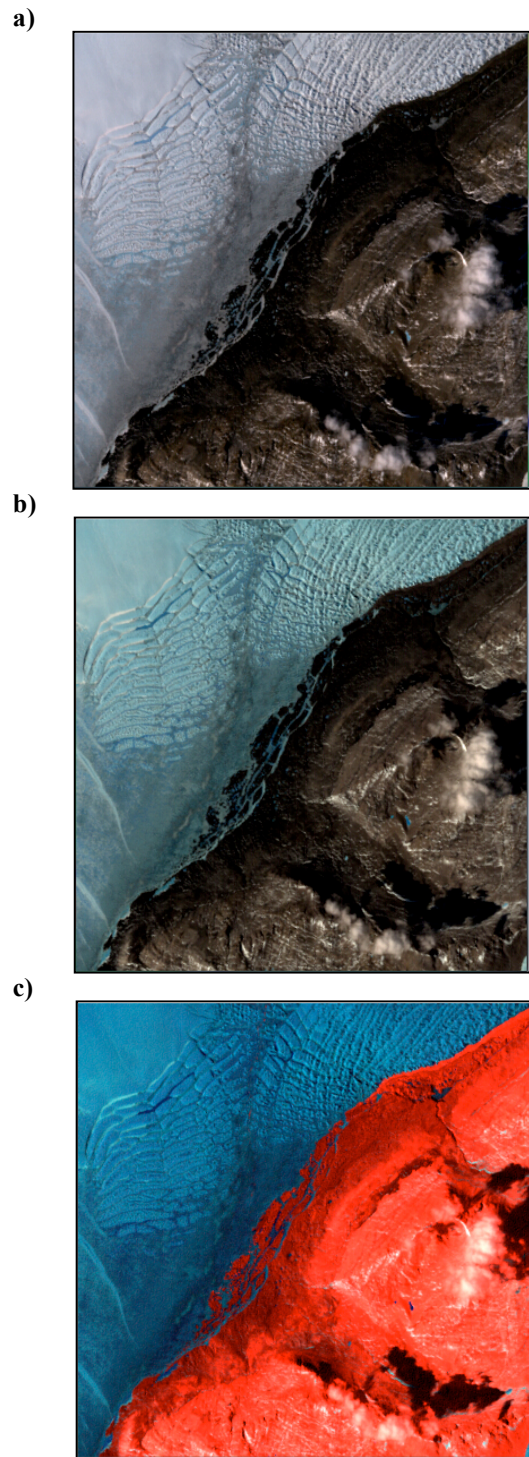


Fig. (3). ETM+ Synthetic images of different bands: true color image (a); b4,b3,b2 synthetic image (b); b7,b4,b3 synthetic image (c).

Table 1. Sample set of different classes.

Class	Glacier	Vegeta-tion	Water	Bare land	Sum
Training samples	582	319	96	347	1344
Test samples	542	265	126	384	1317

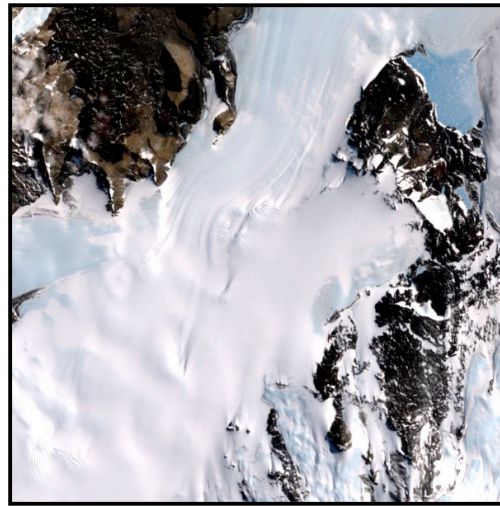


Fig. (4). ETM+ image of the experimental area.

Table 2. Parameters and accuracy of the four types of kernel functions.

Kernel Type	p	$coef$	Overall accuracy (%)	Classification time (s)
RBF			95.08	6.5
Polynomial	3	1	91.99	8.6
Linear			91.25	5.3
Sigmoid		1	91.62	8.7

The classification results of the four types of kernel functions with different parameters are shown in Fig. (5).

We select the optimal classification kernel function by comparing the resulting overall classification accuracy and running time. The RBF kernel function leads to the highest classification accuracy and the classification time is relatively short (Table 2). Therefore, we perform SVM classification with an RBF kernel in experiments.

3. RESULTS AND ANALYSIS

3.1. SVM Classification with RBF Kernel

As described above, we first classify land cover in the study area using SVM with an RBF kernel. The RBF SVM classification result for a part of the study area is shown in Fig. (6). Fig. (7) shows the area classified in Fig. (6).

3.2. Decision Tree Model

SVM is effective in classifying areas with simple spectral information (e.g., glaciers). However, the reflectance of Antarctica's very pure water is relatively low at each band. Moreover, the solar elevation angle causes shadow noise (blue areas on the left side of Fig. (6)). Thus, water and shadows have similar spectral signatures and are easily confused. Through the analysis we find that all the blue areas are not water, part of which is shadow, there are also misclassifications between bare land and glaciers in shadow areas. Knowledge and ancillary data are required to improve the classification image based on established expert rules [15]. To better classify shadow areas, this paper conducts a reclassification using a DT model by comparing and analyzing the spectral values of the sample surface features in different bands (Fig. (8)).

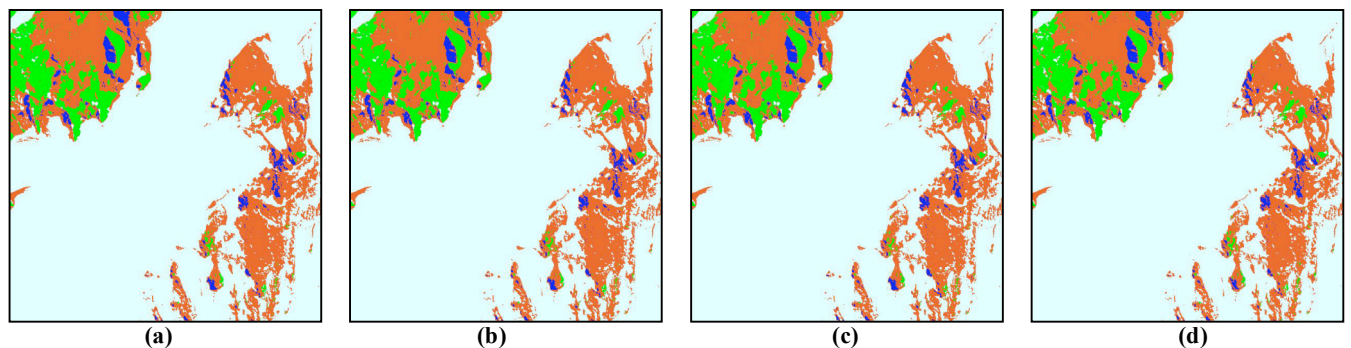


Fig. (5). Classification results of four types of kernel functions: RBF result (a); polynomial result (b); linear result (c); sigmoid result (d).

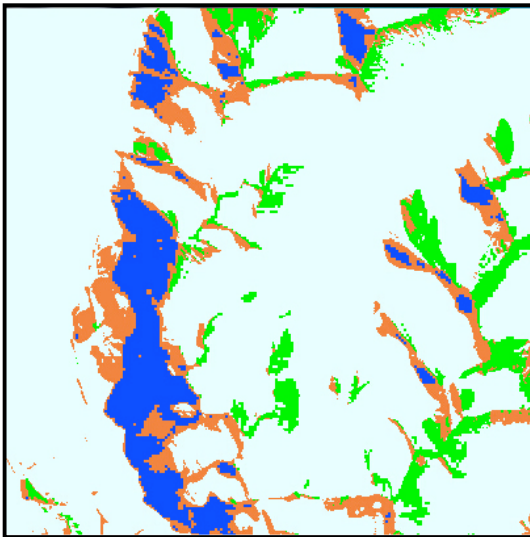


Fig. (6). RBF SVM classification result.

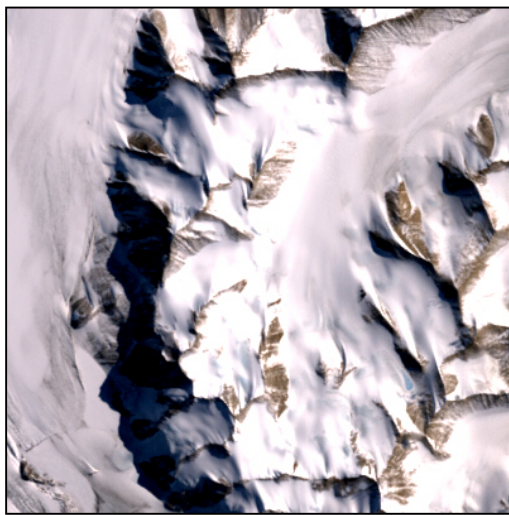


Fig. (7). ETM+ image of the area classified in Fig. (6).

Fig. (8) shows that shadow 1 (shadowed glaciers that are misclassified to water) has a higher reflectance at band 1 than water. Shadow 3 (shadowed glaciers that are misclassified to bare land) has a higher reflectance at band 1 than bare land. Thus, within band 1, if a pixel value is greater than or equal to 69 and less than 84, this pixel is classified as bare land. Otherwise, if the value is greater than or equal to 84, this pixel is classified as glacier. This reclassification approach was decided based on consultation of related litera-

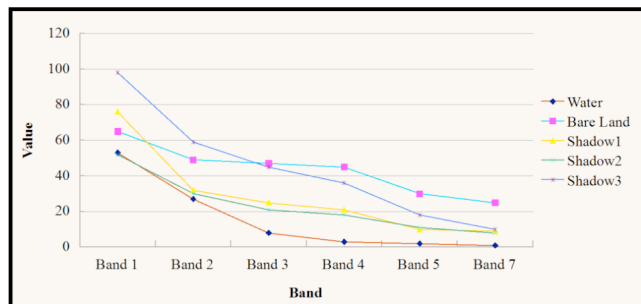


Fig. (8). Spectrum curves of different surface features.

ture and repeated trials. Shadow 2 (shadowed bare land misclassified as water) has a higher reflectance at band 4 than water. When the threshold value is set to 8, the two can be distinguished. That is, at band 4, if the value of a pixel is greater than or equal to 8, this pixel is classified as bare land. Based on the above, we build the DT model shown in Fig. (9), and obtain the classification result shown in Fig. (10).

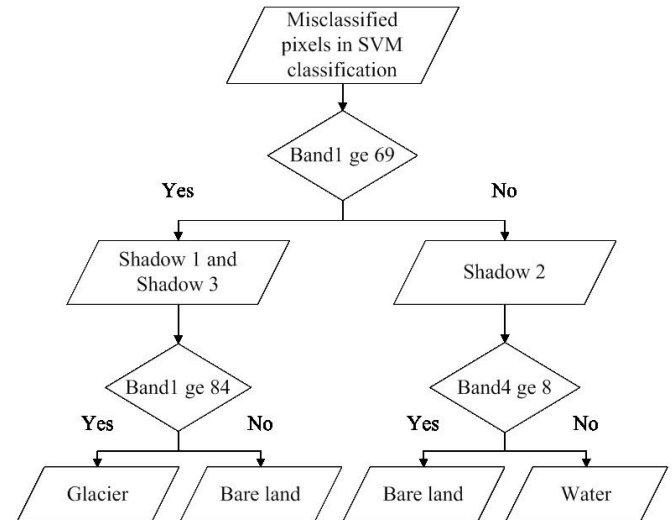


Fig. (9). Schematic of the DT model.

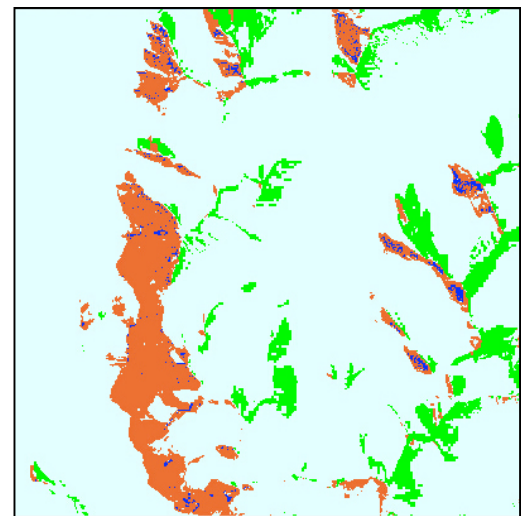


Fig. (10). Classification result of the DT model.

After this DT reclassification, the parts that were misclassified as water are mostly corrected to glacier and bare land, and the result is thus significantly improved compared with the original SVM classification.

3.3. Majority Analysis

Previous research indicates that post-classification processing is an important step in improving the quality of the overall result. Applying supervised classification as described above inevitably leads to some small polygon artifacts. To create a result that can be used for thematic map-

ping or other practical applications, these small polygons must either be removed or reclassified. Common methods to achieve this are majority/minority analysis, cluster analysis, and sieve analysis. The experiments show that Majority analysis is better than other analysis method for the removal of the scattered small polygons. So we use majority analysis to remove small polygons and obtain the optimal parameters empirically. Majority analysis classifies the false pixels in a large class by using a transform kernel of a particular size and a method similar to the convolution filter. It uses the majority class (the class describing the most pixels) instead of the center pixel class. We test the majority analysis parameters as shown in Table 3.

The classification result of majority analysis using different parameters is shown in Fig. (11).

The majority analysis is most effective in removing small polygons when the size of the transform kernel is set to 7×7 and the weight of center pixel is 1. The final land cover classification result for the study area is shown in Fig. (12).

3.5. Discussion

There were some studies on classification of ice and snow. L. Boresoj e Bronge and C. Bronge (1999) studied the spectral characteristics of different types of snow and ice in the Vestfold Hills, East Antarctica by way of principle component analysis and maximum-likelihood classification [16].

Table 3. Parameters of majority analysis.

	Classes	Kernel Size	Center Pixel Weight
a	4	3×3	1
b	4	3×3	2
c	4	5×5	1
d	4	7×7	1

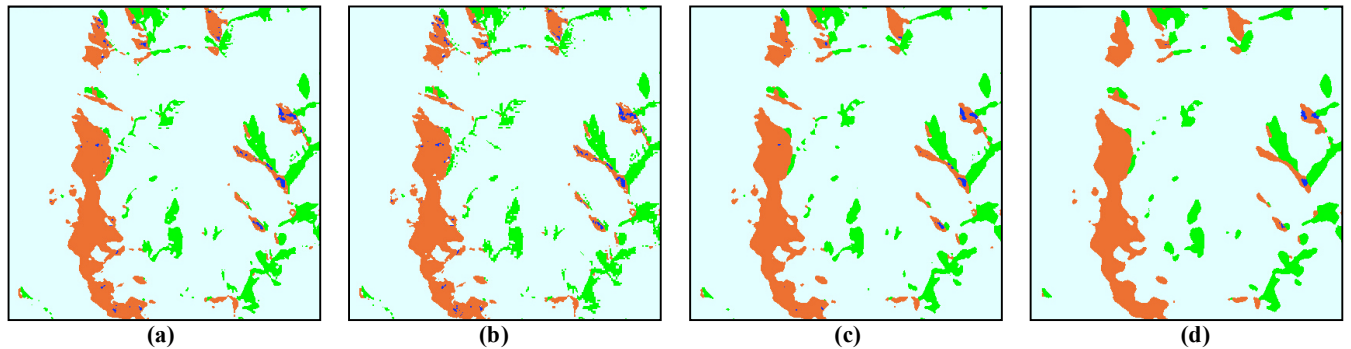


Fig. (11). Classification result of majority analysis with parameter settings from Table 3.

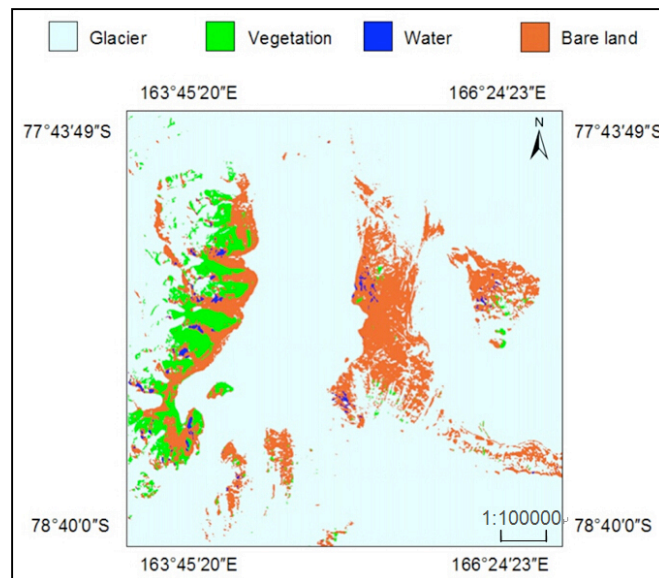


Fig. (12). Land cover classification result of the study area.

There were little studies on land cover classification mapping in Antarctica. Cheng Xiao *et al.* (2010) made a classification map of land use in Antarctica with a scale of 1:100000 by combining visual interpretation with unsupervised classification [17]. Beijing Normal University has done the study in this area before. They mainly used the hand-drawn method and their classification system was also different from ours. The classification accuracy was about 85%. In addition, the classification was also slow and the manual work was time consuming. Therefore, it is necessary to study faster and more efficient classification methods.

SVM has been widely applied to remote sensing image classification. The experiment indicates that the four kernel functions of SVM have higher classification accuracy (Table 2) than traditional methods. In particular, the RBF leads to the optimal classification accuracy and the classification time is relatively short. SVM is effective in the experimental area with obvious different spectral information. However, there exists shadow noise over a large scale of the image in study area. Water and shadows have similar spectral signatures and are easily confused (Fig. (6)). SVM has no spatial analysis capability. Thus, a DT model (Fig. (9)) is established to resolve misclassifications caused by shadows. The experiment indicates that the reclassification result is significantly improved compared with the SVM-only classification (Fig. (10)). Finally, we select Majority analysis and its parameters (Fig. (11)) to remove small polygon artifacts. All of these form the final land cover classification result for the study area (Fig. (12)).

In this paper, the proposed classification method is compared with SVM-only and the maximum likelihood classifiers (Fig. (13)), which perform classification according to the likelihood between unknown and known pixels in the sample data using statistical method. The classification accuracies

using these different classification methods are shown in Table 4. The comparison shows that our proposed method using SVM and DT has better classification accuracy than both SVM-only classification and maximum likelihood classification. In general, our proposed method shows the highest accuracy of 95.82% while maximum likelihood results in the lowest accuracy of 85.50%.

CONCLUSION

This study uses a moderate-resolution Landsat ETM+ image in classifying the land cover of an area in Antarctica. We also use remote sensing image processing software (ENVI) and geographic information system software (ArcGIS). We thus have sufficient data for accuracy verification.

- (1) We establish the classification system and interpretation signs for the focus area in Antarctica. The land cover classification system is based on glacier, vegetation, water, and bare land. The corresponding interpretation signs of the four classes are established by a synthetic scheme;
- (2) The experimental results show that RBF is an optimal kernel function to use with SVM classification for our purposes. This conclusion is reached by conducting a land cover classification and accuracy verification based on the established interpretation signs with four common kernel functions. SVM with an RBF kernel achieved the best accuracy with a reasonable running time;
- (3) A DT model is built to resolve misclassifications caused by shadows. The preliminary classification map is obtained through SVM classification based on

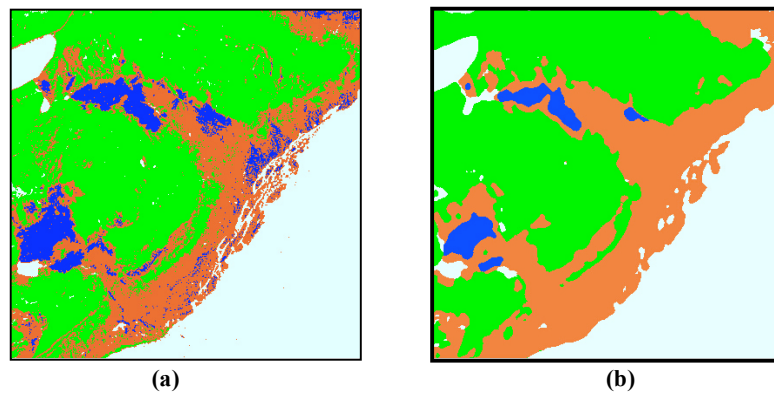


Fig. (13). Comparison of different classification methods: maximum likelihood result (a); SVM and DT result (b).

Table 4. Classification accuracy of different methods.

Classification method	Overall accuracy (%)	Kappa coefficient
SVM + DT	95.82	0.9403
SVM-only	92.03	0.8867
Maximum likelihood	85.50	0.7941

RBF. This original result shows consistent misclassifications (e.g., water and glacier). A DT model is built for further classification. We analyze the spectrum curves of the land cover classes and determine the segmentation threshold at the ETM+1 and ETM+4 bands for bare land, glacier, and water in shadow areas. This indicates that the shadows originally misclassified as water are mostly bare land. A reclassification with a DT model based on the threshold rules leads to an improved result;

- (4) The final classification result is obtained by performing majority analysis to remove small polygon artifacts. We compare our overall method with the original SVM-only classification and the maximum likelihood method. The overall accuracy of our proposed method is 95.82% and the Kappa coefficient is 0.9403. This accuracy is significantly higher than the other two methods. Our method is superior for land cover classification in the presence of undulating terrain and shadows as in this area of Antarctica.

The accuracy and the level of detail in our Antarctica land cover classification are restricted. The field verification of the land cover in the study area is limited to the resolution of the ETM+ image. In future work, we aim to obtain multi-source remote sensing data for the specific area, such as radar data, then attempt to use the other variety of classification methods to further improve the accuracy, efficiency, and robustness of classification.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors thank 863 Program Project Oceania/Antarctica Land Cover Data Products Development, Geological Hazards Conditions Analysis and Prediction in Mining Area and the Key Laboratory of Marine Surveying and Mapping, Shandong University of Science and Technology 2013B01.

REFERENCES

- [1] T. Tashpolat, and J. L. Ding, "Study on the progress of land use/land cover change", *Journal of Xinjiang University (Natural Science Edition)*, vol. 23, pp. 5-15, 2006.
- [2] J. Z. Gong, and B. C. Xia, "Effects of spatial grain size on landscape pattern of land-cover types in the rapidly urbanized region", *Acta Ecologica Sinica*, vol. 26, pp. 2198-2206, 2006.
- [3] K. Kitada and K. Fukuyama, "Land-use and land-cover mapping using a gradable classification method", *Remote Sensing*, vol. 4, pp. 1544-1558, 2012.
- [4] S. G. Warren, and W. J. Wiscombe, "A model for the spectral albedo of snow. II: Snow containing atmospheric aerosols", *Journal of the Atmospheric Sciences*, vol. 37, pp. 2734-2745, 1980.
- [5] S. G. Warren, "Optical properties of snow", *Reviews of Geophysics and Space Physics*, vol. 20, pp. 67-89, 1982.
- [6] S. G. Warren, T. C. Grenfell and P. C. Mullen, "Optical properties of Antarctic snow", *Antarctic Journal of the United States*, vol. 21, pp. 247-248, 1986.
- [7] P. Mantero, G. Moser and S. B. Serpico, "Partially supervised classification of remote sensing images through SVM-based probability density estimation", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, pp. 559-570, 2005.
- [8] G. M. Foody, and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM", *Remote Sensing of Environment*, vol. 103, pp. 179-189, 2006.
- [9] J. F. Yu, "Decision tree classification in identifying land-use types", *Earth Science*, vol. 35, pp. 269-272, 2012.
- [10] Y. Chen, Z. W. He, and B. Gong, "Application of decision tree method in remote sensing information extraction", *Geospatial Information*, vol. 10, pp. 67-72, 2012.
- [11] R. T. Thoreau, C. C. Nicholas, R. G. Nicholas, and A. V. James, "Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications", *Remote sensing of Environment*, vol. 113, pp. 398-407, 2009.
- [12] P. Milap, P. K. Joshi and M.C. Porwal, "Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data", *Expert Systems with Applications*, vol. 38, pp. 5577-5583, 2011.
- [13] V. Vapnik, "The Nature of Statistical Learning Theory," Springer: New York, USA, 1995.
- [14] R. Zhang, and J. W. Ma, "State of the art on remotely sensed data classification based on support vector machines", *Advances in Earth Science*, vol. 24, pp. 555-562, 2009.
- [15] D. S. Lu, and H. W. Qi, "A survey of image classification methods and techniques for improving classification performance", *International Journal of Remote sensing*, vol. 28, pp. 823-870, 2007.
- [16] L. B. Bronge, and C. Bronge, "Ice and snow-type classification in the Vestfold Hills, East Antarctica, using Landsat-TM data and ground radiometer measurements", *International Journal of Remote Sensing*, vol. 20, pp. 225-240, 1999.
- [17] F. M. Hui, G. X. Chen, Y. Liu, and C. Y. He, "Improved landsat image mosaic map of Antarctica", *Earth Science*, vol. 43, 131-142, 2013.