

Article

Feature Selection Solution with High Dimensionality and Low-Sample Size for Land Cover Classification in Object-Based Image Analysis

Yaohuan Huang ^{1,2}, Chuanpeng Zhao ^{1,2,*}, Haijun Yang ³, Xiaoyang Song ^{4,*}, Jie Chen ¹ and Zhonghua Li ^{1,2}

¹ Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; huangyh@igsnrr.ac.cn (Y.H.); chenj@igsnrr.ac.cn (J.C.); lizh.16s@igsnrr.ac.cn (Z.L.)

² College of Resource and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

³ State Environmental Protection Key Laboratory of Satellite Remote Sensing, Beijing 100094, China; yanghj@lreis.ac.cn

⁴ College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China

* Correspondence: zhaoxp.15s@igsnrr.ac.cn (C.Z.); songxiaoyang.good@163.com (X.S.); Tel.: +86-10-6488-9681 (C.Z.)

Received: 18 July 2017; Accepted: 8 September 2017; Published: 11 September 2017

Abstract: Land cover information extraction through object-based image analysis (OBIA) has become an important trend in remote sensing, thanks to the increasing availability of high-resolution imagery. Segmented objects have a large number of features that cause high-dimension and low-sample size problems in the classification process. In this study, on the basis of a partial least squares generalized linear regression (PLSGLR), we propose a group corrected PLSGLR, known as G-PLSGLR, that aims to reduce the redundancy of object features for land cover identifications. Using Gaofen-2 images, the area of interest was segmented and sampled to generate small sample-size training datasets with 51 object features. The features selected by G-PLSGLR were compared against a guided regularized random forest (GRRF) in metrics of reduction rate, feature redundancy, and accuracy assessment of classification. Three indicators of overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) were applied for accuracy assessment in this paper. The result shows that the G-PLSGLR achieved a reduction rate of 9.27 with a feature redundancy of 0.29, and a value of OA 90.63%. The GRRF achieved a reduction rate of 1.61 with a feature redundancy of 0.42, and a value of OA 85.56%. The PA of each land cover category was more than 95% using features selected by G-PLSGLR, while the PA ranged from 77 to 96% using features selected by GRRF. The UA of G-PLSGLR-selected features ranged from 70 to 80% except for grass land and bare land, which achieved 10% higher UA than GRRF-selected features. The G-PLSGLR method we proposed has the advantages of a large reduction rate, low feature redundancy, and high classification performance, which can be applied in OBIA-based land cover classification.

Keywords: feature selection; generalized partial least squares regression; small samples; land cover; OBIA

1. Introduction

Land cover data are key inputs for modeling earth surface processes, such as climate, natural environment, ecology, food security, water resources, and soils [1–7]. Obtaining land cover information is difficult, however, because it is highly dynamic due to modification from not only human activity but also nature [8]. Even though remote sensing is the only method for efficiently

deriving land cover information at a regional or continental scale, remotely-sensed land cover classification is far from satisfactory for many modeling tasks [7,9,10].

In land cover classification, pixel-based image analysis had been dominant since the early 1970s [11]. With the development of high-resolution sensors, intra-class spectral variability of images has been increasing, which invalidates pixel-based image analysis and resulted in the development of object-based image analysis (OBIA) [12]. When comparing OBIA with pixel-based image analysis of land cover extractions from high-resolution images, OBIA not only overcomes the “salt-and-pepper” effect, but also adds additional information on spectra, geometry, context, and texture [13–15]. For these reasons, OBIA has been used widely in high-resolution image-based land cover classification [11]. However, each object in OBIA is regarded as a basic unit, and a large number of features is calculated that may be useful in classification [16].

As the dimensionality of each object in OBIA increases, the sample size needed to support land cover classification methods, e.g., k-nearest neighbors, neural networks, decision trees, and random forests, often grows exponentially [17,18]. Consequently, feature selection is important in OBIA for dimensionality reduction, for class separation, and for the efficient use of training samples [19]. There are three types of feature subset selection approaches: filters, wrappers, and embedded approaches [20]. Filter methods evaluate the quality of features by some criteria (such as correlation criteria, mutual information) to sort features, which are independent from the classification algorithm. Some extensive surveys of various filters methods that select proper features in object-based classification can be found in the literature [21,22]. However, filter methods select features regardless of the performance of the classification algorithm [23]. Wrapper methods measure the performance of features with a classifier, including sequential selection algorithms (such as sequential backward selection, sequential forward selection, and sequential floating selection) and heuristic search algorithms (such as genetic algorithms [24] and particle swarm optimization [25]). However, wrapper models are very computationally expensive. Embedded methods that combine the advantages of previous methods optimize both the goodness-of-fit and the number of variables, which include least absolute shrinkage and selection operator [26], regularized trees, regularized random forest [27], etc. Previous works have focused on generic dimensionality reduction solutions and mainly concentrated on hyperspectral image analysis by pixels [28–30]. Another problem in OBIA is that sampling data of high dimensionality is time-consuming and expensive for large areas. Utilizing collected samples with minimized effects of intra-class variability is also a challenge. Though some research in OBIA has focused on the effects of training set size [31,32], there is still a need for feature selection on limited sample sizes to study the optimal features in the identification of different land covers.

To solve the problem of large intra-class variability and low sample size in OBIA-based land cover classification, we proposed a feature selection solution of group corrected partial least squares generalized linear regression (G-PLSGLR) on the basis of partial least squares generalized linear regression (PLSGLR), which has been applied in many disciplines with small sample sizes, such as genetics, spectroscopy, and analytical chemistry [33–35]. Our approach can be described in four steps: (1) bootstrap the sampling to construct balanced training samples and validation samples; (2) feature grouping based on the Pearson correlation coefficient and graph-based theory using the training samples; (3) removing insignificant features and ranking the importance of grouped features based on PLSGLR regression coefficients; and (4) selecting features by Bayesian information criterion (BIC) calculated on the PLSGLR model after the features are added one by one.

2. Materials and Methods

2.1. Data and Study Area

In this paper, Gaofen-2 image data collected on 26 May 2015 are used as the major data source. These data were geo-referenced and corrected to surface reflectance with the FLAASH algorithm [36]. Gaofen-2 is an optical satellite with high-resolution images that belong to a series of Chinese civilian remote sensing satellites. This satellite provides 1-m panchromatic (0.45–0.90 μm) images and 4-m

images with multispectral bands (blue—B1 (0.45–0.52 μm), green—B2 (0.52–0.59 μm), red—B3 (0.63–0.69 μm), and near infrared—B4 (0.77–0.89 μm)) on a swath of 45 km.

The study area is approximately 31 km^2 over the coast of the Bohai Sea, which covers more than 21 villages (upper left longitude 119°8'29" and latitude 39°28'42", lower right longitude 119°12'22" and latitude 39°25'42") (Figure 1). The size of Gaofen-2 image data of the study area is 5439 rows by 5754 columns. The land cover type of the study area consists of six classes including water, forest land, grass land, crops land, bare land, and residential and build-up land. There are many sub-categories of each land cover type, which result in large intra-class variability. The residential and build-up land category is comprised of such sub-categories as residential districts, industrial area, roads, and greenhouses; the forest land category is comprised of, for instance, the sparse forest alongside roads, the dense forest near the river, and juvenile woodland. The large intra-class variability poses a challenge to land cover classification.

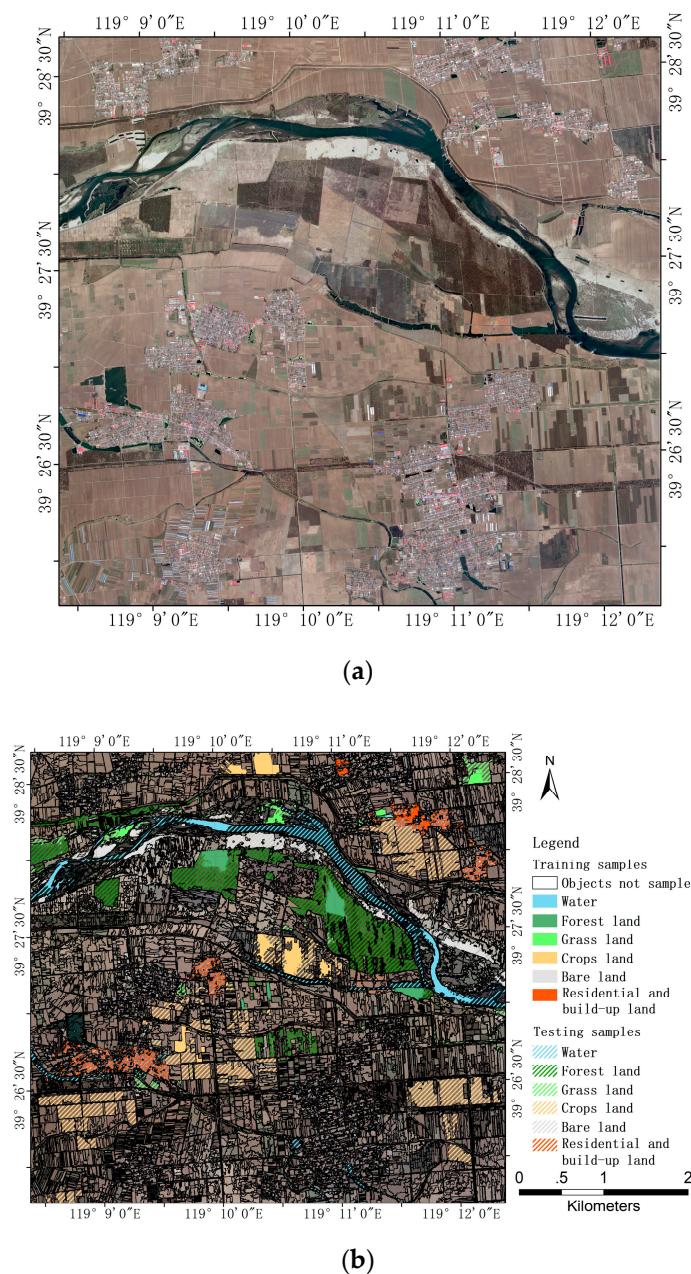


Figure 1. Study area on the eastern coast of China showing the plot acquired by the Gaofen-2 images. (a) Geo-referenced and atmospheric corrected image; and (b) image overlaying a segmentation layer with training samples and testing samples.

2.2. Segmentation and Sampling

Based on an automated multi-resolution segmentation algorithm [37], we carried out image segmentation using eCognition Developer 9.0 software (Trimble Inc., Munich, Germany). After the segmentation step, a total of 51 attributes were calculated on 10,709 objects (Table 1).

To obtain small size samples for OBIA for training, we collected samples from each land cover type, the number of which is even less than the number of features in all types. In addition, we collected testing samples based on both expert interpretation and in situ investigation. The number of samples is shown in Table 2.

Since large intra-class variability and insufficient samples result in the curse of dimensionality [38], the samples were labeled with both category and subcategory at same time in order to integrate human knowledge and avoid instability in clustering algorithms that intend to tackle such problems. Next, a bootstrap sampling strategy was applied to construct balance training samples, adjusting for the fact that most classifiers tend to favor the majority class, resulting in inaccuracy under class-imbalance [39]. For each category, we regarded samples of the current category as one part and extract the same size of samples from the remaining as the other part, considering the subtype impacts.

Table 1. Features used to identify image objects in this study.

Features Category	Object Features	Number of Features
Spectral	Mean (5), Mode (5), Median (5), Standard deviation (5), Skewness (5), Hue, Saturation, Intensity, Max. Diff.	29
Geometry	Asymmetry, Border index, Compactness, Shape index, Roundness	5
Texture	GLCM Homogeneity (all direction), GLCM Contrast (all direction), GLCM Dissimilarity (all direction), GLCM Entropy (all direction), GLCM Ang. 2nd moment (all direction), GLCM Mean (all direction), GLCM Standard Deviation (all direction), GLCM Correlation (all direction), GLDV Mean (all direction), GLDV Contrast (all direction), GLDV Entropy (all direction), GLDV Ang. 2nd moment (all direction)	12
Customized	NDVI, NDWIF, SAVI, OSAVI, DVW	5
	Total	51

Table 2. Samples of different land cover.

Land Cover	Number of Samples	
	Training Objects	Testing Objects
Water	15	113
Forest land	36	258
Grass land	20	67
Crops land	38	212
Bare land	15	85
Residential and build-up land	27	172

2.3. Group Features

Feature selection algorithms frequently do not take into account the structural effects between features, which may result in the loss of representative features, especially in OBIA analysis, where the dimensionality is high [40]. In this paper, we propose a Pearson correlation coefficient-based method for feature grouping, and derived feature groups using graph-based theory. The Pearson correlation coefficient is a measure of the linear dependence between two random variables, and has been proven to be invariant to scaling and translation [41]. The correlation can be simply expressed as:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where x_i and y_i are a series of n measurements of features X and Y, \bar{x} and \bar{y} are mean values of x_i and y_i , and r_{xy} is the correlation between features X and Y.

The feature groups deriving process can be depicted as follows:

$$E = \begin{cases} 1, & r_{xy} \geq th \\ 0, & r_{xy} < th \end{cases} \quad (2)$$

$$g = clique((V, E))$$

The correlation r_{xy} between each pair of features (denoted by X, Y) was calculated for the dataset. When the absolute value of the correlation between a pair of features is greater than threshold th , the edge E between two features is set to 1, otherwise it is set to 0. After the features are traversed, a graph will be established. The vertices V of the graph are made up of features, and the edge between vertices indicates whether the correlation between features exceeds the threshold. The parts that are connected to each other are extracted as a group g .

2.4. Feature Ranking

To rank the feature importance, we chose PLSGLR regression coefficients (β) as indicators on the basis of the PLSRGLM package [42,43] in R software. The PLSGLR method chooses latent components and considers the response variable in regression, which is different from similar methods, such as principal component regression [44].

The model of the PLSGLR with H components is written as follows:

$$P(y=1) = \frac{e^{\left(c_0 + \sum_{h=1}^H c_h t_h\right)}}{1 + e^{\left(c_0 + \sum_{h=1}^H c_h t_h\right)}} \quad (3)$$

where c_h is partial coefficient of component t_h of the logistic regression of the response variable y and c_0 is the intercept. The component t_h with p features can be written as follows:

$$t_h = \sum_{j=1}^p w_{jh}^* x_j \quad (4)$$

where w_{jh}^* is loading on feature x_j of component t_h . Hence, the model with p features we proposed can be expressed as:

$$P(y=1) = \frac{e^{\left(\beta_0 + \sum_{i=1}^p \beta_i x_i\right)}}{1 + e^{\left(\beta_0 + \sum_{i=1}^p \beta_i x_i\right)}} \quad (5)$$

where β_i is the coefficient of feature x_i , and features are sorted based on coefficients; β_0 is the intercept and is removed from later analyses.

Given a two-class dataset, it contains p features, and has a response variable with a value of 1 representing the current class and a value of 0 representing the other class. The PLSGLR method

extracts H principal components t_h in view of explanations of the response variable. Meanwhile, logistic regression is used to determine the coefficients C_h of each component. In order to extract the importance of each feature, we extract loadings w_{jh}^* of features in each component, and multiply them with the coefficient C_h of the corresponding component. The results produce a coefficient β that represents the importance of a feature. We rank the features according to the importance denoted by the absolute value of β .

To estimate the statistical significance of explanatory variables, a nonparametric framework by means of a bootstrap procedure was adopted [45]. Through a large, pre-determined number of repeated samplings with replacements, the distribution parameters of coefficients were estimated, and their confidence intervals (CI) were calculated. Features that did not pass the significance test were removed [46].

2.5. Feature Selection

To select features that fit land cover categories well under a particular model in OBIA, BIC was adopted. BICs are often used to address model selection problems, which can be defined as:

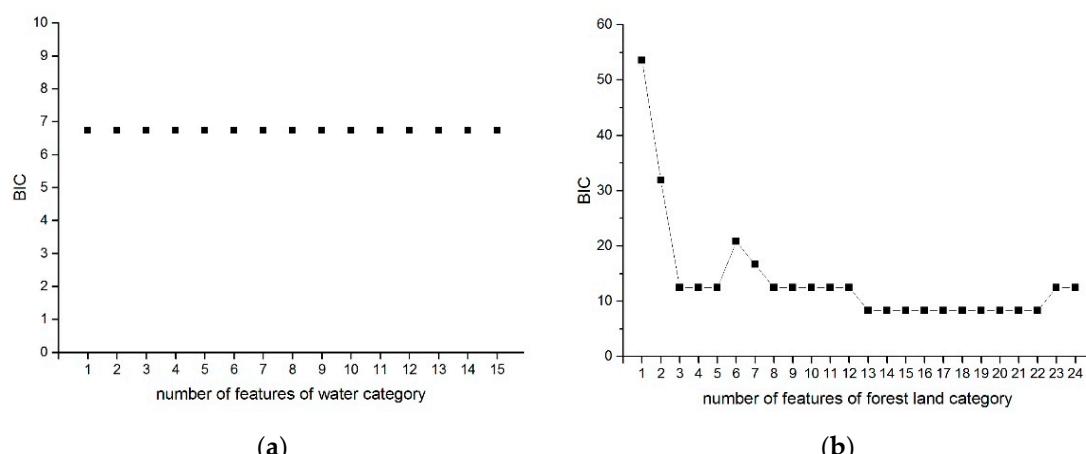
$$BIC = -2 \ln L + k \ln n \quad (6)$$

where L is the likelihood of the PLSGLR, n is the sample size, and k is the number of features added to the PLSGLR model.

Given a series of ranked features for a certain land cover category, it contains k features after the significance test. The samples with a number of n are used to calculate the BIC-based PLSGLR using the selected features. Other methods such as random forests or the support vector machine can also be used to extract BICs for the selected features, but that is beyond the scope of this study. The features are added to the PLSGLR model one by one in a ranked order and the BIC is calculated each time that a feature is added. A low BIC score signals that the features added to the model are optimal in terms of their accuracy considering their dimensionality. Hence, we chose the features corresponding to the minimum BIC or with a BIC disturbance less than 3 [47].

3. Results

According to the BIC curve (Figure 2), the optimal number of features for water, forest land, grass land, crops land, bare land, and residential and build-up land are 1, 3, 8, 8, 9, and 4, respectively.



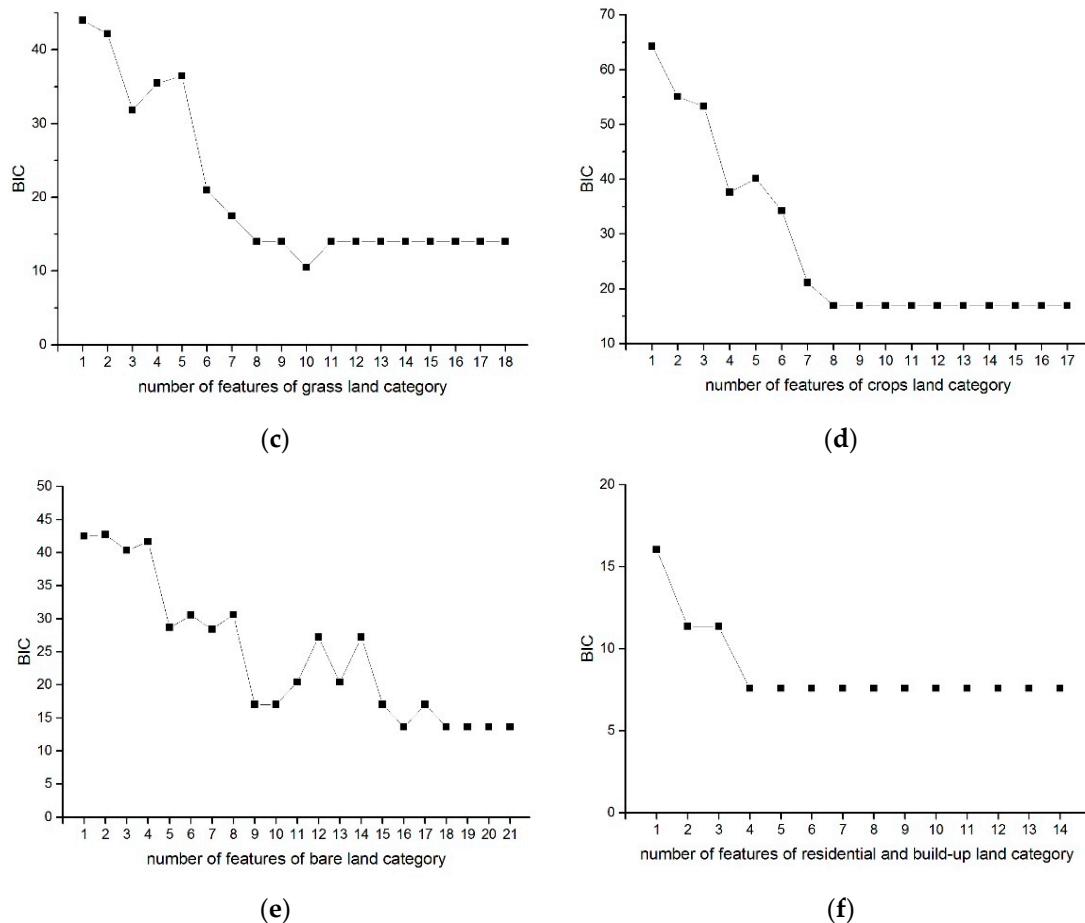
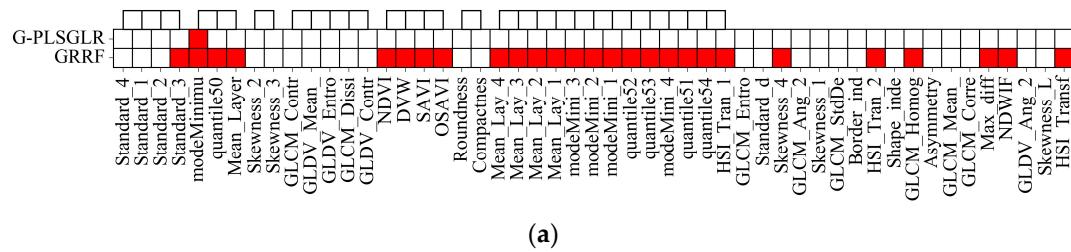


Figure 2. Graph of BIC against the number of features for the G-PLSGLR. Water—BIC vs. number of features (a); Forest land—BIC vs. number of features (b); Grass land—BIC vs. number of features (c); Crops land—BIC vs. number of features (d); Bare land—BIC vs. number of features (e); Residential and build-up land—BIC vs. number of features (f).

In comparison with our approach, guided regularized random forest (GRRF) [48] (an embedded method that optimizes both the goodness-of-fit and the number of variables) was employed to extract features from the same training datasets. The features selected by the two methods, namely, G-PLSGLR and GRRF, are shown in the Figure 3. To describe the dimensionality reduction capability, a reduction rate was defined as the ratio of the original dimensions and the post-processing dimensions of one class, and an overall reduction rate was calculated by taking the mean of each class.



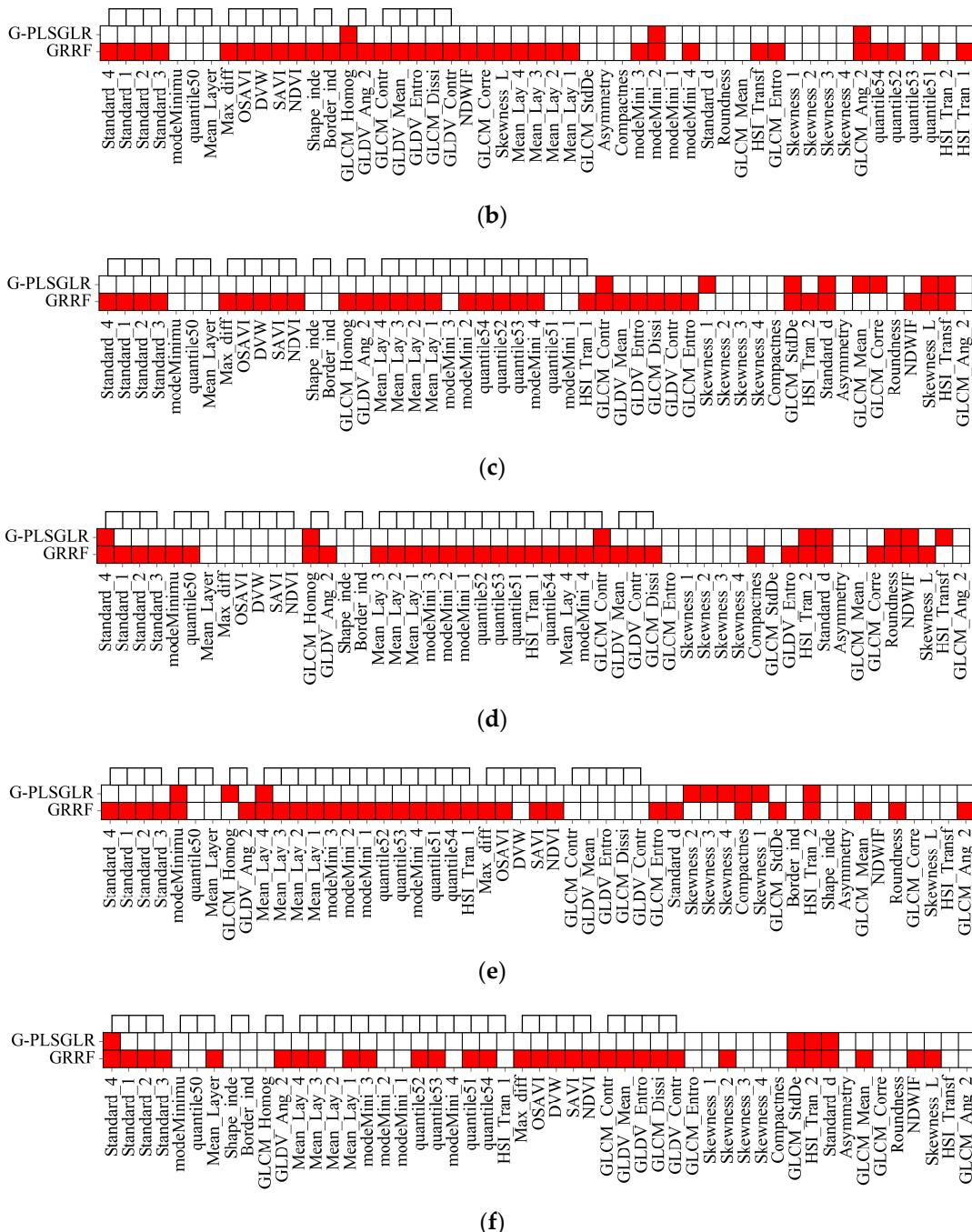


Figure 3. Tile map of the two methods, namely, G-PLSGLR and GRRF, against the selected features. The links on graphs are the feature grouping results from the G-PLSGLR. The selected features are depicted as red tiles across the x-axis. Water—selected features vs. methods (**a**); Forest land—selected features vs. methods (**b**); Grass land—selected features vs. methods (**c**); Crops land—selected features vs. methods (**d**); Bare land—selected features vs. methods (**e**); Residential and build-up land—selected features vs. methods (**f**).

In the OBIA-based land cover classification, only one feature *modeMinimu* was selected to distinguish water from other land cover types. For the extraction of forest land from other land cover types, the features *GLCM_Ang_2*, *GLCM_Homog*, *modeMini_2* were selected. The principal information used to select the features included textural and spectral characteristics, which conforms to the current consensus. For grass land, the features *GLCM_Contr*, *GLCM_Mean_Skewness_L*, *GLCM_Corre*, *GLCM_StdDe*, *HSI_Transf*, *Standard_d*, *Skewness_1* were selected. These features also captured textural and spectral characteristics but were more complex than those employed for forest

land. For identifying crops land, the features *Standard_4*, *HSI_Tran_2*, *Standard_d*, *GLCM_Homog*, *GLCM_Contr*, *Roundness*, *NDWIF*, *HSI_Transf* were selected, which introduced geometric characteristics in addition to textural and spectral characteristics. The geometric characteristics may reveal the regularity of segmented objects on crops land. For the identification of bare land, the features *Skewness_2*, *Skewness_4*, *HSI_Tran_2*, *Mean_Lay_4*, *Skewness_1*, *Skewness_3*, *modeMinimu*, *Compactnes*, *GLCM_Homog* were selected. These features primarily focus on the first-order moments and third-order moments of the segmented object, which may be caused by the high homogeneity of bare land. For the identification of residential and build-up land, the variables *Standard_4*, *HSI_Tran_2*, *Standard_d*, *GLCM_StdDe* were selected. These features primarily focus on the second-order moment information of the segmented object, which may result from the high heterogeneity of the artificial surface. As seen in Figure 3, the number of selected features for the G-PLSGLR method is much less than that of GRRF. The overall reduction rate for G-PLSGLR is 9.27, which is larger than that of GRRF, which is 1.61. The overall reduction rate for G-PLSGLR is 5.78 times higher than that of GRRF.

4. Discussion

4.1. Validation

To compare the feature redundancy of the proposed methodology, average correlation coefficients were applied. The overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) were used to test the classification performance by applying linear discriminant analysis (LDA) using testing samples. The LDA classifier was chosen on the basis of its performance in terms of minimizing the Bayes error for binary classification.

4.1.1. Evaluation of Feature Redundancy

Feature redundancy increases search space size and affects the speed and the accuracy of learning algorithms [49]. The correlation coefficient, which indicates the strength and direction of a relationship between two random variables, is used to measure the redundancy. We extracted the upper triangular matrix from the correlation coefficient matrix of selected features for each class and later calculated its average value to measure the overall redundancy. An average correlation coefficient value close to 1 indicates a strong redundancy, while a value close to 0 indicates a weak redundancy. The absolute value of the correlation coefficient is used to capture the dependency, whether it is positive or negative.

As shown in Figure 4, the GRRF method has the largest redundancy in all land cover categories. The redundancy of selected features for water by G-PLSGLR is 0; in contrast, GRRF has a redundancy of 0.53. As for forest land, the redundancy of features selected by G-PLSGLR is 0.26, while GRRF-selected features have a redundancy of 0.41. In view of grass land, features selected by G-PLSGLR have a redundancy of 0.28, and GRRF-selected features have a redundancy of 0.40. For crops land, the redundancy of features selected by G-PLSGLR is 0.29, and GRRF-selected features have a redundancy of 0.41. The redundancy of features selected by G-PLSGLR for bare land is 0.35, and GRRF-selected features have a redundancy of 0.39. In view of residential and build-up land, the redundancy of features selected by G-PLSGLR is 0.28, and GRRF-selected features have a redundancy of 0.38. Additionally, the overall redundancy for G-PLSGLR is 0.29, while GRRF has an overall redundancy of 0.42. In short, the performance of G-PLSGLR-selected features exceeds GRRF-selected features in metrics of redundancy.

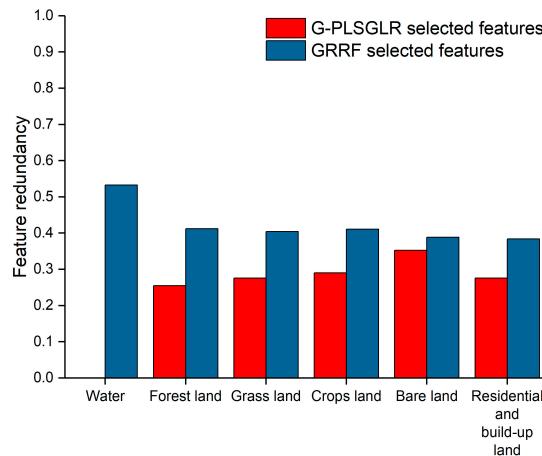


Figure 4. Feature redundancy of G-PLSGLR and GRRF against six land cover categories.

4.1.2. Accuracy Assessment on Selected Features

The accuracy assessment is derived from a confusion matrix. The OA describes the overall correctness of classification, but does not reveal the distribution of errors [50]. Therefore, UA and PA are taken into consideration. They are defined as follows:

$$\begin{aligned}
 OA &= \frac{\sum_{i=1}^n P_{ii}}{N} \\
 PA &= \frac{P_{ii}}{P_{+j}} \\
 UA &= \frac{P_{ii}}{P_{i+}}
 \end{aligned} \tag{7}$$

where P_{ii} is the number of the i -th land cover category that is correctly classified, N is the total number of testing samples, n is the number of categories, P_{+j} is sum of the j -th columns in the confusion matrix, and P_{i+} is sum of the i -th rows in the confusion matrix.

The OA of G-PLSGLR and GRRF is shown in Figure 5 and Table 3. Five land cover categories have an OA of more than 80%, while the OA of grass land is about 75%. The mean OA for features selected by G-PLSGLR has a mean value of 90.63%, while GRRF-selected features have a mean OA of 85.56% for six land cover categories. This indicates that features selected by G-PLSGLR are more representative than features selected by GRRF. The PA of each land cover category was more than 95% using features selected by G-PLSGLR, while the UA ranged from 70 to 80%, with exceptions of the grass land and bare land. For the serious landscape fragmentation and small occupation area of these two land cover categories, G-PLSGLR chooses no more than 18% features to describe them. The insufficient discrimination of these two land covers at a fixed segmentation scale may account for the low UA, which can be ignored in the accuracy assessment of this study. Based on OA, PA, and UA metrics, G-PLSGLR-selected features achieved a higher accuracy than GRRF-selected features at higher reduction rates, which indicates that G-PLSGLR-selected features are more representative for the identification of specific land cover categories.

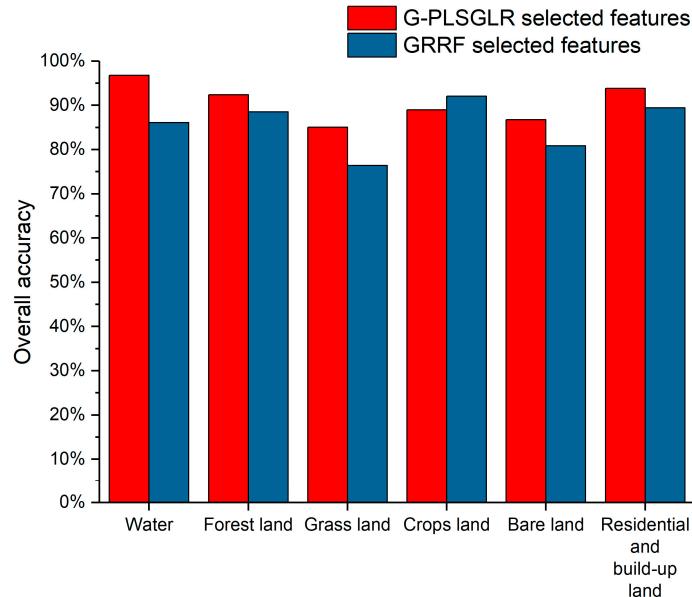


Figure 5. Overall accuracy of G-PLSGLR and GRRF against six land cover categories.

Table 3. The accuracy assessment of LDA classification using features selected by G-PLSGLR and GRRF.

	OA		UA		PA	
	G-PLSGLR	GRRF	G-PLSGLR	GRRF	G-PLSGLR	GRRF
Water	96.80%	86.11%	80.43%	47.19%	98.23%	96.46%
Forest land	92.39%	88.53%	79.44%	72.38%	98.84%	96.51%
Grass land	85.01%	76.41%	32.49%	20.72%	95.52%	77.61%
Crops land	88.97%	92.06%	68.42%	75.93%	98.11%	96.70%
Bare land	86.77%	80.82%	41.03%	30.90%	94.12%	84.71%
Residential and build-up land	93.83%	89.42%	76.85%	64.96%	96.51%	95.93%

4.2. Comparison with Other Low Sample Size Studies

Previous studies have demonstrated that when the number of training samples is very small (e.g., 20, 40 samples), algorithms including Support Vector Machine, Random Forests, Classification and Regression Tree, and Maximum-Likelihood Classification did not perform well on 24 features [31]. With the increasing number of features, the accuracy of classification using these algorithms may be even lower. This study used a number of training samples between 15 and 38 for each class of 51 features, and achieved relative high accuracy of classification after the G-PLSGLR feature selection process, which is detailed in Section 3.2. We interpret the increased performance to the following mechanisms: elimination of class imbalance, reduction of feature redundancy, and suitability of PLSGLR.

As mentioned in Section 2.2, classifiers tend to result in a classification preference towards the major class with uneven class sizes [39,51]. In a feature selection process, the selected features may be more conducive to the separation of the major class. Under-sampling, which involves removing samples of the major class, is a method of dealing with this class imbalance problem. With the high dimensionality and large intra-class variability of samples, a cluster-based under-sampling approach would be invalid [38]. In this work, the manual labeling of subtypes for training samples is acceptable due to the low sample size, which is inevitable for sample efficiency and classification accuracy. It is more reliable than automatic clustering process or synthetic sampling, with the introduction of domain knowledge.

Previous research has shown that the most important predictor variables always are highly correlated, which results in unstable classification accuracy even when the same training data are used [52,53]. Spearman rank-order correlation was used in past research to determine pair-wise

correlations. In this study, we absorbed the correlation criteria-based filter methods for feature selection to group highly correlated variables, which is a similar process to that of the above studies. However, a slight difference is that we choose an arbitrary feature as a representative of the total group. The advantages of feature grouping include: (1) the outputs of feature selection are more stable, and (2) the redundancy of the selected features is lower.

This study introduces the PLSGLR method for feature selection for OBIA-based land cover classification. The method is widely used in areas such as hyperspectral information analysis [54–56]. In these areas, the dimensionality is always higher than number of observations, reaching tens and even hundreds of times the number of observations. Our study provides a framework known as G-PLSGLR, taking class imbalance and feature redundancy into consideration for feature selection in OBIA-based land cover classification to extract main predictor variables with comparable validation accuracy.

4.3. Potential Extensions for Producing Multi-Class Land Use and Land Cover Maps

Although this study focuses on extracting the best features that identify the objects of specific land cover categories, a land use and land cover (LULC) map can be produced on basis of the binary classification results. In general, multi-class classification could be extended from binary classification by strategies such as stacking binary classifiers and hierarchical binary classifiers [57,58]. The stacking of binary classifiers, which started with the binary classification of each land cover category from image objects, can be used to produce a final LULC map by overlaying the outputs of each binary classification result and voting. In contrast, hierarchical binary classifiers are also widely used to solve the problem of combining the results of binary classifiers with multi-class classification [59]. A hierarchical binary classifier can be started by building a hierarchical classification tree from one land cover that is easier to identify or by following the maximum margin rule to split classes into two macro-classes [60]. Then, other land cover categories can be extracted from the outputs of each node. By optimizing the order of extraction, the hierachic constraints are helpful in improving the accuracy. However, there are some problems such as error accumulations, extraction orders, and category conflicts, which need further research.

5. Conclusions

The high-dimensionality and low sample size associated with land cover information extraction using OBIA is a common problem [15]. In this study, we propose an improved extended partial least squares regression method known as G-PLSGLR to reduce the feature dimensionality of OBIA-based land cover classification. To obtain a low sample size for OBIA, we collected several samples from each land cover category using results from Gaofen-2 image segmentation, and we collected a series of validation samples to evaluate the performance of the proposed method. The approach consisted of four steps: (1) labeling the sub-categories of samples and automatically sampling them into class balanced datasets, in view of the fact that the training samples have high intra-class variability, and the number of training samples is less than the dimensionality; (2) grouping features based on the Pearson correlation coefficient and graph theory to reduce the redundancy of features calculated by OBIA; (3) ranking features; the grouping results were analyzed using a PLSGLR regression, and non-significant features were removed after evaluation by a bootstrap method for confidence interval estimation. The remaining features found to be significant for explaining land cover types were sorted based on their regression coefficients; and (4) selecting the optimal feature set from the remaining sorted features according to the BIC.

For the study area, the G-PLSGLR was applied to generate the optimal feature set and was validated by comparing with the results with those of GRRF using the metrics of reduction rate, feature redundancy, OA, and Kappa statistic. The results showed that G-PLSGLR is suitable for OBIA feature selection and can significantly reduce the number and redundancy of selected features compared with GRRF. The overall reduction rate on all land cover categories of features selected by G-PLSGLR was 9.27, while GRRF had a value of 1.61. The overall feature redundancy of G-PLSGLR was 0.29, whereas GRRF had a value of 0.42. The overall OA using an LDA classifier was 90.63% for

G-PLSGLR, and 85.56% for GRRF. The PA of each land cover category was more than 95% using features selected by G-PLSGLR, while the PA ranged from 77 to 96% using features selected by GRRF. The UA of G-PLSGLR-selected features ranged from 70 to 80% except for the grass land and bare land, which achieved a 10% higher UA than GRRF-selected features. The evaluation results showed that G-PLSGLR can extract more representative features for a variety of land cover categories at a higher reduction rate with lower feature redundancy and higher classification performance than that of GRRF.

The G-PLSGLR we proposed broadens the choice of feature selection methods for OBIA-based land cover classification with high dimension and low sample size. The method is well-suited to OBIA-based land cover information extraction with a high reduction rate and comparable accuracy. G-PLSGLR can be applied to a wide range of feature selection scenarios and has the advantage of a higher reduction rate with a smaller sample size. Future work will focus on the non-linear effects in the process of grouping features.

Acknowledgments: This research was supported and funded by the National Key Research and Development Program of China (Grant No. 2017YFB0503005, Grant No. 2016YFC0401404); and National Natural Science Foundation of China (Grant No. 51309210).

Author Contributions: Jie Chen and Zhonghua Li collected the remote sensing data, Chuanpeng Zhao processed the remote sensing data, proposed the model and analyzed the results, Yaohuan Huang and Chuanpeng Zhao wrote the manuscript, Haijun Yang and Xiaoyang Song advised the research work that led to this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ochoa, P.; Fries, A.; Mejía, D.; Burneo, J.; Ruíz-Sinoga, J.; Cerdà, A. Effects of climate, land cover and topography on soil erosion risk in a semiarid basin of the Andes. *Catena* **2016**, *140*, 31–42.
- Godinho, S.; Guiomar, N.; Machado, R.; Santos, P.; Sá-Sousa, P.; Fernandes, J.P.; Neves, N.; Pinto-Correia, T. Assessment of environment, land management, and spatial variables on recent changes in *montado* land cover in southern Portugal. *Agrofor. Syst.* **2016**, *90*, 177–192.
- Zhou, G.; Wei, X.; Chen, X.; Zhou, P.; Liu, X.; Xiao, Y.; Sun, G.; Scott, D.F.; Zhou, S.; Han, L. Global pattern for the effect of climate and land cover on water yield. *Nat. Commun.* **2015**, *6*, 1–8.
- Tuanmu, M.N.; Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **2014**, *23*, 1031–1045.
- Mahmood, R.; Pielke, R.A.; Hubbard, K.G.; Niyogi, D.; Dirmeyer, P.A.; McAlpine, C.; Carleton, A.M.; Hale, R.; Gameda, S.; Beltrán-Przekurat, A.; et al. Land cover changes and their biogeophysical effects on climate. *Int. J. Climatol.* **2014**, *34*, 929–953.
- Verburg, P.H.; Mertz, O.; Erb, K.-H.; Haberl, H.; Wu, W. Land system change and food security: Towards multi-scale land system solutions. *Curr. Opin. Environ. Sustain.* **2013**, *5*, 494–502.
- Lu, X.; Zhuang, Q. Evaluating climate impacts on carbon balance of the terrestrial ecosystems in the Midwest of the United States with a process-based ecosystem model. *Mitig. Adapt. Strateg. Glob. Chang.* **2010**, *15*, 467–487.
- Ban, Y.; Gong, P.; Giri, C. Global land cover mapping using earth observation satellite data: Recent progresses and challenges. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 1–6.
- Jiang, D.; Huang, Y.; Zhuang, D.; Zhu, Y.; Xu, X.; Ren, H. A Simple Semi-Automatic Approach for Land Cover Classification from Multispectral Remote Sensing Imagery. Available online: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045889> (accessed on 8 September 2017).
- Gong, P. Remote sensing of environmental change over China: A review. *Chin. Sci. Bull.* **2012**, *57*, 2793–2801.
- Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F.; et al. Geographic object-based image analysis—towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191.
- Hay, G.; Niemann, K.; McLean, G. An object-specific image-texture analysis of H-resolution forest imagery. *Remote Sens. Environ.* **1996**, *55*, 108–122.

13. Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using random forests. *Remote Sens. Environ.* **2011**, *115*, 2564–2577.
14. Dronova, I.; Gong, P.; Wang, L. Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China. *Remote Sens. Environ.* **2011**, *115*, 3220–3236.
15. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16.
16. Formaggio, A.R.; Vieira, M.A.; Rennó, C.D. Object Based Image Analysis (OBIA) and Data Mining (DM) in Landsat time series for mapping soybean in intensive agricultural regions. In Proceedings of the Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012.
17. Huang, X.; Lu, Q.; Zhang, L. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS J. Photogramm. Remote Sens.* **2014**, *90*, 36–48.
18. Powell, W.B. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*; John Wiley & Sons: Hoboken, NJ, USA, 2007; Volume 703.
19. Jensen, J.R. *Remote Sensing of the Environment: An Earth Resource Perspective* 2/E. 2009. Available online: https://s3.amazonaws.com/academia.edu.documents/31163537/08_rs_vegetation.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1505103677&Signature=L37TljB8tcuCXSiqYYFP%2BJ8fB0%3D&response-content-disposition=inline%3B%20filename%3DRemote_Sensing_of_the_Environment_An_Ear.pdf. (accessed on 11 September 2017).
20. Tang, J.; Alelyani, S.; Liu, H. Feature Selection for Classification: A Review. 2014. Available online: http://eprints.kku.edu.sa/170/1/feature_selection_for_classification.pdf (accessed on 8 September 2017).
21. Wu, B.; Xiong, Z.-G.; Chen, Y.-Z. Classification of quickbird image with maximal mutual information feature selection and support vector machine. *Procedia Earth Planet. Sci.* **2009**, *1*, 1165–1172.
22. Ma, L.; Cheng, L.; Li, M.; Liu, Y.; Ma, X. Training set size, scale, and features in geographic object-based image analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 14–27.
23. Hall, M.A.; Smith, L.A. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference, Orlando, FL, USA, 1–5 May 1999.
24. Van Coillie, F.M.; Verbeke, L.P.; De Wulf, R.R. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sens. Environ.* **2007**, *110*, 476–487.
25. Chen, Q.; Chen, Y.; Jiang, W. Genetic particle swarm optimization-based feature selection for very-high-resolution remotely sensed imagery object change detection. *Sensors* **2016**, *16*, 1204.
26. Takayama, T.; Iwasaki, A. Optimal wavelength selection on hyperspectral data with fused lasso for biomass estimation of tropical rain forest. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-8*, 101–108.
27. Mureriwa, N.; Adam, E.; Sahu, A.; Tesfamichael, S. Examining the spectral separability of *Prosopis glandulosa* from co-existent species using field spectral measurement and guided regularized random forest. *Remote Sens.* **2016**, *8*, 144.
28. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675.
29. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122.
30. Ghamisi, P.; Couceiro, M.S.; Benediktsson, J.A. A novel feature selection approach based on FODPSO and SVM. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2935–2947.
31. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sens.* **2014**, *6*, 964–983.
32. Kavzoglu, T.; Colkesen, I. The effects of training set size for performance of support vector machines and decision trees. In Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Florianopolis-SC, Brazil, 10–13 July, 2012.
33. Otto, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*; John Wiley & Sons: Hoboken, NJ, USA, 1998.

34. Boulesteix, A.L.; Lambert-Lacroix, S.; Peyre, J.; Strimmer, K. Plsgenomics: PLS Analyses for Genomics. R Package Version. 2011. Available online: <https://rdrr.io/cran/plsgenomics> (accessed on 11 September 2017).
35. Brown, D.J.; Shepherd, K.D.; Walsh, M.G.; Mays, M.D.; Reinsch, T.G. Global soil characterization with vnir diffuse reflectance spectroscopy. *Geoderma* **2006**, *132*, 273–290.
36. Felde, G.; Anderson, G.; Cooley, T.; Matthew, M.; Berk, A.; Lee, J. Analysis of Hyperion data with the FLAASH atmospheric correction algorithm. In Proceedings of the Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003.
37. Drăguț, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127.
38. Yen, S.-J.; Lee, Y.-S. Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Syst. Appl.* **2009**, *36*, 5718–5727.
39. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
40. Friedman, J.; Hastie, T.; Tibshirani, R. A Note on the Group Lasso and a Sparse Group Lasso. 2010. Available online: <https://arxiv.org/pdf/1001.0736.pdf> (accessed on 8 September 2017).
41. Haindl, M.; Somol, P.; Ververidis, D.; Kotropoulos, C. Feature selection based on mutual correlation. In Proceedings of the 11th Iberoamerican Congress in Pattern Recognition, Cancun, Mexico, 14–17 November 2006.
42. Bertrand, F.; Maumy-Bertrand, M.; Meyer, N. Plsrglm, PLS generalized linear models for the R language. In Proceedings of the 12th International Conference on Chemometrics in Analytical Chemistry, Anvers, Belgium, 19 October 2010.
43. Bertrand, F.; Magnanensi, J.; Meyer, N.; Maumy-Bertrand, M. Plsrglm: Algorithmic Insights and Applications. 2014. Available online: <ftp://alvarestech.com/pub/plan/R/web/packages/plsRglm/vignettes/plsRglm.pdf> (accessed on 8 September 2017).
44. Boulesteix, A.-L.; Strimmer, K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.* **2007**, *8*, 32–44.
45. Bastien, P.; Vinzi, V.E.; Tenenhaus, M. PLS generalised linear regression. *Comput. Stat. Data Anal.* **2005**, *48*, 17–46.
46. Chun, H.; Keleş, S. Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection. 2010. Available online: <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2009.00723.x/full> (accessed on 8 September 2017).
47. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
48. Deng, H. Guided Random Forest in the RRF Package. 2013. Available online: <https://arxiv.org/pdf/1306.0237.pdf> (accessed on 8 September 2017).
49. Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the Science and Information Conference (SAI), London, UK, 27–29 August 2014.
50. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
51. Graves, S.J.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Colgan, M.S.; Kalantari, L.; Bohlman, S.A. Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. *Remote Sens.* **2016**, *8*, 161.
52. Millard, K.; Richardson, M. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sens.* **2015**, *7*, 8489–8515.
53. Millard, K.; Richardson, M. Wetland mapping with LIDAR derivatives, SAR polarimetric decompositions, and LIDAR–SAR fusion using a random forest classifier. *Can. J. Remote Sens.* **2013**, *39*, 290–307.
54. Fassnacht, F.E.; Neumann, C.; Förster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central european test sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2547–2561.
55. Song, K.; Li, L.; Li, S.; Tedesco, L.; Hall, B.; Li, Z. Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm–partial least squares (ga–pls) modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 368–385.
56. Wilson, M.; Ustin, S.L.; Rocke, D. Comparison of Support Vector Machine Classification to Partial Least Squares Dimension Reduction with Logistic Discrimination of Hyperspectral Data. 2003. Available Online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/4886/1/Comparison-of-support->

- vector-machine-classification-to-partial-least-squares/10.1111/12.463169.short?SSO=1 (accessed on 8 September 2017).
- 57. Sánchez-Marcano, N.; Alonso-Betanzos, A.; García-González, P.; Bolón-Canedo, V. Multiclass classifiers vs multiple binary classifiers using filters for feature selection. In Proceedings of the 2010 international joint conference on Neural networks (IJCNN), Barcelona, Spain, 18–23 July 2010.
 - 58. Tax, D.M.; Duin, R.P. Using two-class classifiers for multiclass classification. In Proceedings of the 16th International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002.
 - 59. Begum, S.; Aygun, R.S. Greedy hierarchical binary classifiers for multi-class classification of biological data. *Netw. Model. Anal. Health Inform. Bioinform.* **2014**, *3*, 53.
 - 60. Tibshirani, R.; Hastie, T. Margin trees for high-dimensional classification. *J. Mach. Learn. Res.* **2007**, *8*, 637–652.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).