





A Comparative Land-Cover Classification Feature Study of Learning Algorithms: DBM, PCA, and RF Using Multispectral LiDAR Data

Suoyan Pan, Haiyan Guan , Senior Member, IEEE, Yongtao Yu , Member, IEEE, Jonathan Li , Senior Member, IEEE, and Daifeng Peng 

Abstract—Multispectral LiDAR, characterization of completeness, and consistency of spectrum and spatial geometric data provide a new data source for land cover classification. However, how to choose the optimal features for a given set of land covers is an open problem for effective land cover classification. To address this problem, we propose a comparative scheme, which investigates a popular deep learning (deep Boltzmann machine, DBM) model for high-level feature representation and widely used machine learning methods for low-level feature extraction and selection [principal component analysis (PCA) and random forest (RF)] in land cover classification. The comparative study was conducted on the multispectral LiDAR point clouds, acquired by a Teledyne Optech's Titan airborne system. The deep learning-based high-level feature representation experimental results showed that, on an ordinary personal computer or workstation, this method required larger training samples and more computational complexity than the machine learning-based low-level feature extraction and selection methods. However, our comparative experiments demonstrated that the classification accuracies of the DBM-based method were higher than those of the RF-based and PCA-based methods using multispectral LiDAR data.

Index Terms—Feature extraction, feature selection, land cover classification, learning algorithms, multispectral LiDAR data.

I. INTRODUCTION

OVER the past two decades, studies have proven that, as a single data source for land cover classification, there

Manuscript received November 6, 2018; revised January 2, 2019 and January 26, 2019; accepted February 2, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 41671454, 61603146, and 41801386, and in part by the Natural Science Foundation of Jiangsu Province under Grants BK20160427 and BK20180797. (Corresponding author: Haiyan Guan.)

S. Pan is with the School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: pansy_nuist@163.com).

H. Guan and D. Peng are with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: guanhy.nj@nuist.edu.cn; daifeng@nuist.edu.cn).

Y. Yu is with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China (e-mail: allennessy.yu@gmail.com).

J. Li is with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2899033

exist identification problems in LiDAR data due to a lack of rich spectral information. Therefore, to improve the accuracy of object detection in different environments, many studies have integrated LiDAR data with high spatial resolution multispectral images acquired by earth observing satellite (e.g., IKONOS, QuickBird, SPOT-5, GeoEye, Worldview-2) and color aerial imagery [1]–[4] to improve land cover classification accuracy. However, when integrating LiDAR data with multispectral images for land cover classification, how to accurately register different data sources to a same spatial coordinate system is still an open problem.

In recent years, some institutes and companies have successfully introduced prototypes of multispectral and even hyperspectral LiDAR systems. For example, Teledyne Optech's Titan, the first commercial multispectral LiDAR system, was released in Canada in December 2014. As a new type of active remote sensing sensor, multispectral LiDAR has unlimited possibilities in three-dimensional (3-D) land cover classification, vegetation mapping, shallow-water bathymetry, and dense topography. Multispectral LiDAR, providing relatively complete and consistent spectral and spatial geometric data, has obvious advantages for land cover classification [5]–[9]. For example, Wichmann *et al.* [10] investigated the spectral patterns of major categories (e.g., no-asphalted ground, asphalted ground, buildings, vegetation, and water bodies) from multispectral LiDAR data. These studies showed that multispectral airborne LiDAR is suitable for traditional geometric classification and mapping, compared with single-wavelength (channel/band) LiDAR data and optical images.

Most land cover classification methods for multispectral LiDAR data are generally performed on either 3-D LiDAR points [10]–[12] or two-dimensional (2-D) feature images interpreted from 3-D points [13]–[16]. Miller *et al.* [17] demonstrated that the classification accuracies of 3-D LiDAR points were better than those of 2-D feature images. However, with the development of LiDAR technologies, the higher the LiDAR point density, the heavier the computational burden when land cover classification performed on 3-D LiDAR points [18]. Therefore, converting 3-D airborne multispectral LiDAR points into 2-D feature images is an effective way to obtain land cover maps by using established image processing algorithms (e.g., maximum likelihood classification, support vector machine, decision tree, and random forest).

Features extracted from the data play an essential role in land cover classification, therefore selecting representative features is a common task [19]. The purpose of feature selection is to remove irrelevant and/or redundant features. It is well known that LiDAR data provide height, intensity, and height-derived features, while image data provide abundant spectral and texture features. Although there are a plenty of features that can be obtained from single-wavelength/multispectral LiDAR data, or even optical images, how to robustly choose features is an open problem for effective land cover classification. Most studies on LiDAR-based land cover classification used subjectively selected features, whose classification accuracies varied with different types of input data and diverse environmental conditions. In practice, to obtain satisfactory classification accuracy, the optimal number of features and the most appropriate features are no easy way to be determined in advance, therefore it is indispensable to perform a greedy feature selection. This selection method is time-consuming and does not guarantee robust classification accuracy.

Machine learning is capable of separating main image features from complex data in the field of image processing. Machine learning involved in feature extraction and selection is mainly based on data dimensionality reduction. There are two main methods for reducing dimensions: feature extraction and feature selection. In feature extraction, a new feature set (containing k -dimensional features) is calculated/transformed from d -dimensional features ($k \leq d$). The widely used feature extraction method is principal component analysis (PCA) [20], [21], which reduces feature dimensions and redundant information between the original features, leading to the improvement of data quality and processing efficiency. In feature selection, it is common to find the k -dimensional useful features from all the d -dimensional features ($k < d$) in land cover classification. Dong *et al.* [22] demonstrated that the accuracy of urban land cover classification was improved by removing some irrelevant and/or redundant features from LiDAR-derived geometric features. Random forest (RF), an improved version of bagging (one of ensemble learning algorithms), is capable of providing an estimate of an individual variable importance index. The ability of variable importance estimation contributes to investigate the influence of each predictive feature for selecting the best features when LiDAR data are used in a classification model [23], [24]. Feature extraction and selection by machine learning methods (e.g., PCA and RF) require no users' intervention. The PCA method automatically transforms and maps features to produce a highly discriminative feature set, however, the resulting new feature set is not interpretable. The RF method uses decision trees to classify and evaluate the importance of variables to extract better features. However, there may be many similar decision trees in the classification, resulting in unreliable results.

Usually, we consider spectral, geometrical, and height-derived features extracted from LiDAR data or optical images as low-level features. PCA-based feature extraction and RF-based feature selection are generally implemented on these low-level features. To describe strong feature representation, a deep learning model has been widely used in the fields of image analysis. Compared to features abstracted by shallow learning methods, features learned by deep learning techniques are high-

TABLE I
SPECIFICATIONS OF THE TELEDYNE OPTTECH'S TITAN AIRBORNE SYSTEM

Items	Channel1	Channel2	Channel3
Wavelength(nm)	1550	1064	532
Waveband	SWIR	NIR	GREEN
Look angle(degree)	3.5° forward angle	nadir	7° forward angle
Pulse repetition frequency(kHz)	300	300	300
Flying height(m)	~1000	~1000	~1000
Point density(points/m ²)	3.6	3.6	3.6

level feature representations in an automated manner through deep-structured neural networks. By learning multilevel feature representation, deep learning models have proven to be an effective tool for fast object-oriented identification problems. Deep Boltzmann machine (DBM) [25], is an important breakthrough in the demand for powerful deep feature representation models [26]. DBM, a learning structure of a multilayer perceptron with multiple hidden layers, can form abstract high-level features to discover distributed feature representation of data.

Although machine learning methods (e.g., PCA and RF) have been proven to be capable of objectively extracting and selecting low-level features for land cover classification, the number and types of the low-level features mainly rely on the operator's prior knowledge and experience. In contrast, deep learning methods (e.g., DBM) can automatically abstract high-level feature representation from a voluminous data samples. Thus, deep learning methods have become attractive in the fields of image processing, language identification, etc. However, there are rare related studies on whether the powerful feature representation ability of DBM can be used to improve land cover classification accuracy [27].

The objective of this study is to compare two widely used machine learning methods (PCA and RF) and a representative deep learning method (DBM) for feature analysis used in multispectral LiDAR-based land cover classification, and analyze their performances on airborne multispectral LiDAR data. Section II describes the data acquired from the Teledyne Optech's Titan multispectral LiDAR system, as well as data pre-processing. Section III presents DBM-based high-level feature representation and PCA-based and RF-based low-level feature extraction and selection methods. The conducted tests are described and analyzed in Section IV. Finally, concluding remarks are presented in Section V.

II. MULTISPECTRAL LiDAR DATA AND DATA PREPROCESSING

A. Test Multispectral LiDAR Data

The study area is a small town located in Whitchurch-Stowe, Ontario, Canada, with a latitude and longitude in the center region (43°58'00", 79°15'00"). The data in this study were collected by the first commercial multispectral active sensor—Teledyne Optech's Titan, whose detailed specifications are listed in Table I. The data covered an area of about 25 square kilometers, with 19 flying strips (9 strips vertically intersecting 10 strips). Note that, in this study, the system parameters and trajectories were unavailable, the three channels of intensities

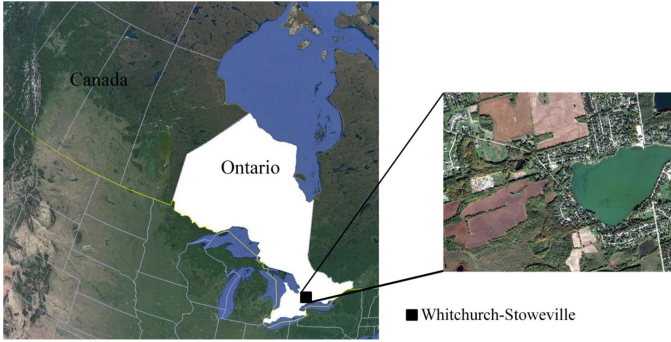


Fig. 1. Study area.

were directly used from the LiDAR outputs (as a file format of ASPRS LAS files) without intensity calibration.

A test data with an area of $2052 \text{ m} \times 1566 \text{ m}$ was selected from the collected Teledyne Optech's Titan multispectral LiDAR data. Fig. 1 shows the study area. As shown in Fig. 1, the selected study area contains a rich variety of objects, such as roads, trees, grass, and soil, which contribute to the implementation of our comparative study of feature analysis in land cover classification. To improve the quality of training sample selection and further evaluate land cover classification accuracies of the selected features, a high-resolution remote sensing image corresponding to the study area was downloaded from the Google Earth.

B. Data Preprocessing

Each channel of the Teledyne Optech's Titan multispectral LiDAR system generated a group of independent point cloud data. It is well known that point density is an essential factor in LiDAR-based land cover classification task. Therefore, we first merged the three-wavelengths point clouds into a single, high-density point cloud because three wavelengths in different scanning directions acquire more points than one wavelength. In the merged point cloud, each point contains the intensities of three wavelengths. Specifically, each of the single-wavelength point clouds was taken as the reference data, in which each point was processed to find its neighbors in the other two wavelengths of point clouds using a nearest neighbor searching algorithm. Because the average point density of single wavelength was about 3.6 points/m^2 , the searching distance in this study was set to 1.0 m to obtain sufficient points in the two wavelengths of point clouds. To obtain the intensities of the two other wavelengths, a bilinear interpolation method was used. If there were no neighboring points in the one of two wavelengths, the intensity value of this wavelength was set to zero. As such, three wavelengths were merged into a single, multispectral point cloud with higher point density than that of the single wavelength.

To efficiently perform feature extraction in land cover classification, through the inverse distance weighted interpolation method [28], the merged multispectral LiDAR data were rasterized into 2-D grid datasets, by using elevation and intensities from different single wavelength and the merged multiwavelength. Based on the fact that the average point density of each wavelength was 3.6 points/m^2 , the resolution of the 2-D image

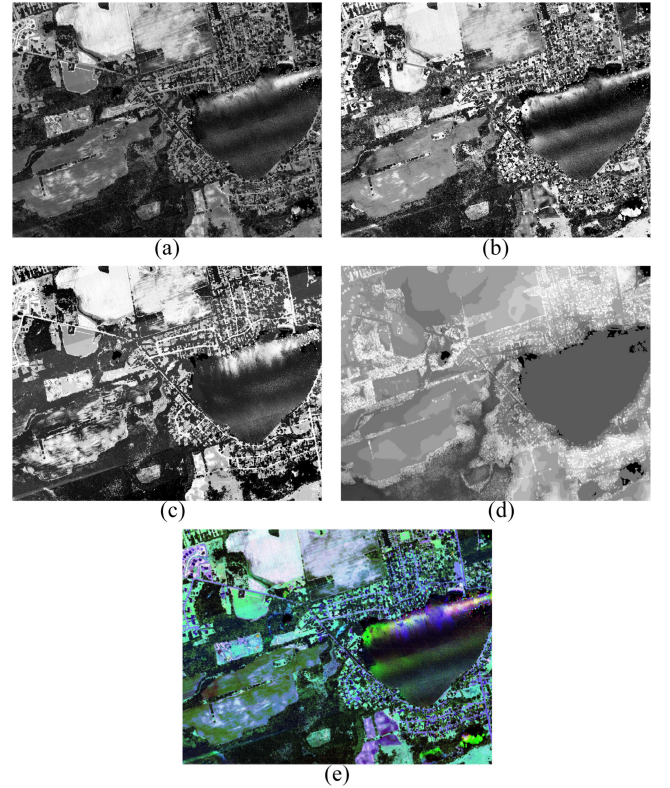


Fig. 2. LiDAR images. (a) Intensity image from SWIR band. (b) Intensity image from NIR band. (c) Intensity image from GREEN band. (d) Elevation image. (e) Pseudo-color image.

was set to 0.5 m. Fig. 2(a)–(c) show the intensity images of the three single-wavelengths, respectively. Fig. 2(d) and (e) show the elevation and multispectral image of the merged point cloud, respectively.

III. PROPOSED SCHEME

The proposed scheme is to compare the performance of deep-learning-based high-level feature representation and machine learning-based low-level feature extraction and selection in land cover classification. Fig. 3 shows the workflow of the proposed scheme. The comparative study for feature analysis in land cover classification consists of three parts, including DBM-based high-level feature representation, machine learning-based low-level feature analysis, and land cover classification for comparison. The input data are the rasterized multispectral LiDAR images. High-level feature representation, obtained by the DBM, is input to a support vector machine (SVM) classifier, which is termed as Case 1. PCA-based low-level feature extraction and RF-based low-level feature selection methods are also tested by the SVM classifier, termed as Cases 2 and 3, respectively. The output data are the classification results for these cases, which were used to compare the performance of different feature analysis methods.

A. DBM-Based High-Level Feature Representation

A DBM is a layer-wise extension of the restricted Boltzmann machines with multiple hidden layers [29]–[31]. In this study,

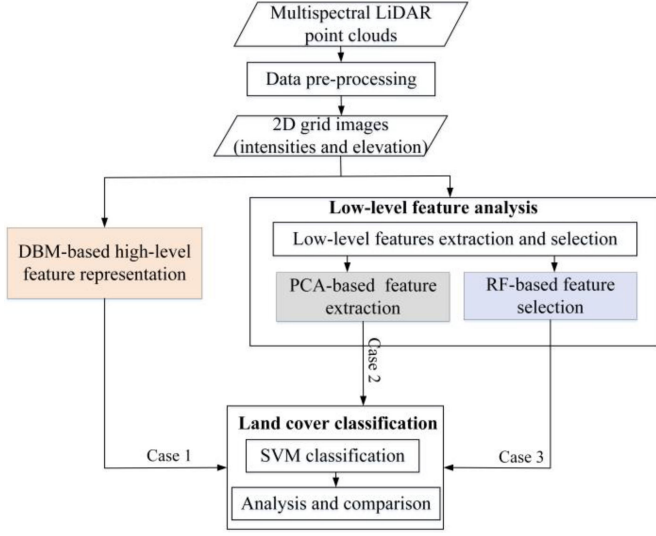


Fig. 3. Workflow of the proposed scheme.

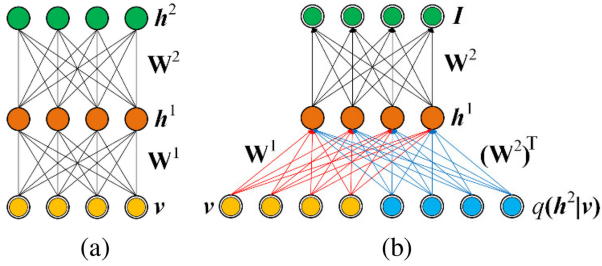


Fig. 4. (a) Two-layers DBM model. (b) High-level feature representation model.

we construct a deep feature generation model for generating high-level abstract features of multispectral LiDAR data using a DBM model.

As shown in Fig. 4(a), we construct a two-layer DBM model. Denote $v \in [0, 1]^n$ as the real-valued visible units representing a multispectral LiDAR pixel. Denote $h^1 \in \{0, 1\}^{N_{h^1}^1}$ and $h^2 \in \{0, 1\}^{N_{h^2}^2}$ as the first and second layer hidden units, respectively. $N_{h^1}^1$ and $N_{h^2}^2$ are the numbers of hidden units in the first and second hidden layers, respectively. Then, the energy of the configuration, $\{v, h^1, h^2\}$, is defined as follows:

$$\begin{aligned} E(v, h^1, h^2; \phi) = & \frac{1}{2} \sum_{i=1}^n \frac{v_i^2}{\sigma_i^2} - \sum_{i=1}^n \sum_{j=1}^{N_{h^1}^1} \frac{v_i}{\sigma_i} w_{ij}^1 h_j^1 \\ & - \sum_{j=1}^{N_{h^1}^1} \sum_{m=1}^{N_{h^2}^2} h_j^1 w_{jm}^2 h_m^2 \end{aligned} \quad (1)$$

where $\phi = \{W^1, W^2, \sigma\}$ are the model parameters; v_i is the i th element of v ; h_j^1 is the j th element of h^1 ; h_m^2 is the m th element of h^2 ; W^1 and W^2 are the visible-to-hidden and hidden-to-hidden symmetric interaction terms, respectively; w_{ij}^1 is the element on the i th row and j th column of W^1 ; w_{jm}^2 is the element on the j th row and m th column of W^2 ; σ^2 represents the variances of the visible units; σ_i is the i th element of σ . The conditional distributions over the visible and two sets of hidden units are

expressed as follows:

$$p(h_j^1 = 1 | v, h^2) = g \left(\sum_{i=1}^n \frac{v_i}{\sigma_i} w_{ij}^1 + \sum_{m=1}^{N_{h^2}^2} h_m^2 w_{jm}^2 \right) \quad (2)$$

$$p(h_m^2 = 1 | h^1) = g \left(\sum_{j=1}^{N_{h^1}^1} h_j^1 w_{jm}^2 \right) \quad (3)$$

$$p(v_i = x | h^1) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left(-\frac{\left(x - \sigma_i \sum_{j=1}^{N_{h^1}^1} h_j^1 w_{ij}^1 \right)^2}{2\sigma_i^2} \right) \quad (4)$$

where $g(x) = 1/(1 + e^{-x})$ is the logistic function [26].

To rapidly and effectively train the model parameters, ϕ , first, a greedy layer-wise pretraining is performed to initialize the model parameters. Then, an iterative training algorithm combined with variable and stochastic approximation approaches [26] is used to fine-tune the model parameters.

Once the model parameters are trained, the stochastic activities of binary features in the hidden layers of the DBM are replaced by real-valued probability estimations to construct a deep feature generation model [see Fig. 4(b)]. Considering the feedbacks from hidden layers, for each visible vector v , mean-field inference [29] is adopted to generate an approximate posterior distribution, $Q(h^2 | v)$. Then, the marginal, $q(h^2 | v)$ of the approximate posterior serves as an augment to the deep feature generation model. Finally, the last layer of the deep feature generation model produces the following high-level feature representation:

$$I^T = g \left(g \left(\frac{v^T}{\sigma^T} W^1 + q(h^2 | v)^T (W^2)^T \right) W^2 \right) \epsilon [0, 1]^{N_{h^2}^2}. \quad (5)$$

In this study, the 2-D gridded multispectral LiDAR data (intensity and elevation images of SWIR, NIR, and GREEN) were input into the DBM model to abstract high-level feature representation. Afterward, the abstracted high-level features are then input into the SVM classifier for land cover classification.

However, for a DBM model, there is no strict theoretical guidance on how many hidden layers (N_{hL}) and hidden units (N_{hU}) for each layer might be used to obtain the optimal performance. Previous experiments demonstrated that a large number of hidden layers leads to time-consuming and helps little to improve the accuracy. Therefore, this study used a two-layers DBM model, which has been demonstrated to be able to provide acceptable high-level features to an SVM classifier. Before performing the proposed DBM model, we first resized the training samples with the neighborhood size of N_s , and normalized their pixel values into the range $[0, 1]$. Afterward, the processed training samples are linearly arranged into a real-valued vector as an input to the DBM model.

B. Machine Learning-Based Low-Level Feature Analysis

1) *Low-Level Features*: To obtain useful information for identifying land covers, features are first extracted from the

TABLE II
FEATURES AND DESCRIPTION OF FEATURES

Data	Features	Description of Features	Number of Features
	Spectral features of three wavelengths: <i>GREEN, NIR, MIR</i>	Teledyne Optech's Titan multi-spectral LiDAR data provide spectral information in three bands, which can be used to identify objects to be interest.	3
Intensity images	Vegetation indexes: <i>GDVI, GRVI, GNDVI</i> ; Water index: <i>MNDWI</i>	(1) GDVI = NIR-GREEN. It is sensitive to changes in soil background and suitable for detecting low-coverage vegetation [32]. (2) GRVI = NIR/GREEN. It is sensitive to green plants, and its sensitivity is proportional to the size of vegetation coverage [32]. (3) GNDVI = (NIR-GREEN) / (NIR+GREEN). It is usually used to detect vegetation coverage, vegetation growth status and eliminate some radiation errors [33]. (4) MNDWI=(GREEN-SWIR)/(GREEN +SWIR). It enhances and extracts water from a background that is dominated by build-up land areas [34].	4
	GLCM measures: <i>GREEN_corre, GREEN_sedmom, GREEN_entropy, GREEN_dissimi, GREEN_cont, GREEN_homo, GREEN_vari, GREEN_mean, NIR_corre,....., MIR_mean</i>	Gray level co-occurrence matrix (GLCM) is used to calculate eight feature components, including Correlation, Second.Moment, Entropy, Dissimilarity, Contrast, Homogeneity, Variance, and Mean.	8×3=24
Elevation image	height information: <i>elevation</i> GLCM measures: <i>ele_corre, ele_sedmom, ele_mean</i>	The 2D Elevation image is considered as single band. The height-texture-based features is calculated by GLCM.	1 8×1=8

multispectral LiDAR data. An overview of features used in this study is listed in Table II, including spectral-based and LiDAR height-based features.

a) Spectral-based features: The three wavelengths were used as three individual features after a process of low-pass filtering or smoothing. In addition to the spectral information of the 2-D multispectral LiDAR data, a set of texture information, calculated by a grey-level co-occurrence matrix (GLCM), is used for describing the spatial properties and relationships between grey levels in neighboring pixels of an image region. More importantly, according to the three wavelengths of the multispectral LiDAR data, we can derive three vegetation and water indexes, including green difference vegetation index (GDVI) [32], green ratio vegetation index (GRVI) [32], green normalized difference vegetation index (GNDVI) [33], and modified normalized difference water index (MNDWI) [34].

b) LiDAR height-based features: Our previous studies [35], [36] summarized LiDAR height-based features from single-wavelength LiDAR point clouds. In this study, due to the shortage of multiple echo information provided by the Teledyne Optech's Titan system, the height-based features related with multiecho information were not selected. Similar to spectral-based features, the 2-D elevation image is considered as one spectral band after a process of low-pass filtering. Although Dong *et al.* [22] summarized a total of near 20 geometric features, including height-statistics-based, height-texture-based, fitting-plane-based, and eigenvalue-based features, our previous studies found that the height-features, directly calculated from original 3-D point clouds in a given spherical neighborhood, contributed less to land cover classification results [35], [36]. In contrast, some texture features, derived from the 2-D elevation image, played a significant role in classification. The reason behind this phenomenon is that the height-texture-based features were calculated in a given neighborhood similar to the calculation way of some eigenvalue-based and fitting-plane-based features. Thus, in this study, we only use eight GLCM-based elevation texture features.

2) PCA-Based Low-Level Feature Extraction: PCA transformation is an orthogonal linear transformation based on the

amount of information. The transformation mainly uses a linear projection method to project data into a new coordinate space, where the new generated components are distributed according to the amount of information. When the number of principal components (PCs) increases, the amount of information contained in the component decreases. That is, the first PC contains the largest amount of information. Note that each PC represents the critical, uncorrelated information of the original data. In addition, these PCs can both reduce the impact of noise on data analysis, and improve the data representation effectively. However, PCA is an unsupervised feature extraction method, which fails to utilize the information of classes of interest.

3) RF-Based Low-Level Feature Selection: RF has gained the reputation of computational efficiency, robustness to outliers and noises, and useful internal estimates of error, strength, correlation, and variable importance because multiple classifiers have a better classification performance than a single classifier [37]. In an RF, two-thirds of the training data (called in-bag data) are used to construct the decision tree, and the remaining one-third of the training data (called out-of-bag (*OOB*) data) are used to test the decision tree to evaluate its classification performance. Each decision tree in the RF model produces an *OOB* error, and the average *OOB* error obtained by all the decision trees is called the *OOB* error estimation.

In a RF, there are two parameters: the number of variables ($F_{m\text{try}}$) in the random subset at each node and the number of trees ($F_{n\text{tree}}$) in the forest. Usually, $F_{m\text{try}}$ is set to the square root of the number of features [38]. $F_{n\text{tree}}$ can be as large as possible because RF is fast and not overfitting. Considering the computer computation performance, $F_{n\text{tree}}$ is usually several hundred [39].

One important characteristic of the RF model is that it provides the variable importance of classification results. The specific process is as follows:

- 1) a decision tree is created from the randomly-resampled in-bag data, and then verified by the out-of-bag data;
- 2) $F_{m\text{try}}$ features are randomly selected from the low-level features (F_m features totally) extracted from multispectral LiDAR data. With those $F_{m\text{try}}$ features, each decision tree

is tested based on Gini impurities to determine the optimal segmentation method for each node;

- 3) the importance of the variables is finally estimated by the majority voting method.

To select the best and uncorrelated features, an iterative backward feature selection method [40] is used. First, the feature importance is calculated according to the *OOB* error. Then, the feature variables are sorted with a descending order according to the numerical value, and the least important feature variables are iteratively eliminated. For each iteration, the last 20% of the features are eliminated. Afterward, the eliminated new features are re-input into the SVM classifier, and the importance of the new feature variables to the classification results is calculated and sorted according to the *OOB* error. The iteration process is terminated until the feature set with the highest classification accuracy is obtained.

C. Land Cover Classification

As shown in Fig. 1, the study area covers a variety of land cover objects on the ground, such as buildings, roads, trees, and open grassy spaces. Considering the study area, we selected the most important objects for land cover classification: building, tree, road, grass, soil, and water. These object classes can be easily identified by means of the Google image (as shown in Fig. 1). For Cases 2 and 3, this study manually selected about 800 randomly distributed sample pixels for each class. For Case 1, we manually selected 30 000 pixels for each class due to the characteristics of demanding for a great training samples for deep learning-based methods. For feature analysis comparison in this study, we used an SVM classifier that is based on small samples, which avoided the traditional classification process from induction to deduction. In addition, SVM could deal with the solution where few known samples are in LiDAR data, therefore this classification method has been widely used in the field of LiDAR data feature classification. The kernel function and gamma value of the SVM in this study were the radial basis function and 0.2, respectively.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The experiments in this study compared the performance of the deep learning-based high-level representation (Case 1) and machine learning-based low-level feature analysis (Cases 2 and 3) in land cover classification. To quantify the performance of the feature analysis methods, the LibSVM toolbox was used in land cover classification. All the methods were implemented under MATLAB R2016b. We used overall accuracy, user's accuracy, producer's accuracy, and kappa coefficient based on confusion matrixes for each land cover.

A. DBM-Based High-Level Feature Representation

In this study, there are three important parameters— N_{hU}^1 (the number of hidden units at the first layer), N_{hU}^2 (the number of hidden units at the second layer), and N_s (the neighborhood size)—in the DBM-based high-level feature representation model. In this section, we designed three groups of

experiments to investigate the sensitivity of the DBM-based high-level feature representation to the selection of these three parameters.

In the first group (i.e., Case 1–1), we maintained $N_s = 9 \times 9$ pixels, $N_{hU}^2 = 80$, and varied N_{hU}^1 from 100 to 500 at an interval of 100. Table III(a) shows the experimental results. As shown in Table III(a), the kappa coefficients and overall accuracies of the SVM land cover classification results dramatically vary with increase of the parameter, N_{hU}^1 . Specifically, the classification accuracies slightly increase as the parameter, N_{hU}^1 changes from 100 to 300. However, the classification accuracies drop down when N_{hU}^1 changes from 300 to 500. Moreover, as seen in Table III(a), the producer's accuracies and user's accuracies of all land covers reach the highest when the N_{hU}^1 is 300. The meaningful and acceptable results demonstrate the superior performance of the DBM model. Specifically, grass and water are classified better than the other land covers. Therefore, in this study, the N_{hU}^1 value of 300 obtains the best land cover classification accuracies.

In the second group (i.e., Case 1–2), we maintained $N_s = 9 \times 9$ pixels, $N_{hU}^1 = 300$, and varied N_{hU}^2 from 10 to 120 at an interval of 10. As shown in Table III(b), both the classification overall accuracies and kappa coefficients generally increase in a mountain-like way as N_{hU}^2 changes from 10 to 120. Moreover, the accuracies dramatically increase when N_{hU}^2 increases from 10 to 80. In this study, the N_{hU}^2 value of 80 obtains the best land cover classification accuracy. The reason behind this phenomenon might be that with relatively small values of N_{hU}^1 and N_{hU}^2 , the DBM model cannot be fully learned due to insufficient connections between neurons, leading to less information in the abstracted high-level feature representation, and reducing land cover classification accuracies. In contrast, when the values of N_{hU}^1 and N_{hU}^2 are relatively large, it may cause the overfitting problems, resulting in redundant information contained in the abstracted high-level feature representation, and also reducing the classification accuracies. In addition, as seen in Table III(b), the user's accuracies and producer's accuracies of all land covers show the same increasing rules. As such, in our study, the best land cover classification accuracy (89.3% of overall accuracy) is obtained at $N_{hU}^1 = 300$ and $N_{hU}^2 = 80$.

In the third group (i.e., Case 1–3), we used $N_{hU}^1 = 300$, $N_{hU}^2 = 80$, and varied N_s from 5 to 15 pixels with an interval of 2 pixel. As shown in Table III(c), when the neighborhood size increases from 5 to 9 pixels, the overall accuracies grow accordingly. However, the overall accuracies greatly decrease as the parameter N_s changes from 9 to 15 pixels. This is because when the neighborhood size value is smaller, the neighborhood information of the objects to be classified is insufficient to extract high-level feature representation of the object, resulting in relatively lower land cover classification. Conversely, a larger value of the neighborhood size generates redundant information of an object class, also leading to the reduction of land cover classification accuracies. As seen in Table III(c), when $N_s = 9$, the DBM model achieved the maximum user's accuracies and producer's accuracies of all land covers. The phenomenon is consistent with the overall accuracy and kappa coefficient assessment.

TABLE III
SENSITIVITY TESTS WITH PARAMETERS

Experiment	N_{hU}^1	Overall Accuracy (%)	Kappa	Producer's Accuracy (%)						User's Accuracy (%)					
				building	tree	road	grass	soil	water	building	tree	road	grass	soil	water
Case1-1-1	100	70.4	0.711	68.7	75.6	71.2	78.4	74.3	88.7	68.7	70.9	69.5	80.3	81.6	86.9
Case1-1-2	200	81.5	0.793	75.3	86.3	79.6	88.8	84.7	92.2	79.8	82.4	76.3	90.2	90.3	89.6
Case1-1-3	300	89.3	0.866	88.9	90.5	83.6	94.1	88.7	98.1	85.3	88.2	87.8	97.9	98.6	97.3
Case1-1-4	400	83.6	0.822	82.2	87.8	80.9	89.8	86.3	95.1	81.2	82.1	81.7	91.5	79.0	95.5
Case1-1-5	500	82.2	0.785	80.5	83.3	78.4	83.3	79.1	90.3	79.2	81.2	76.2	86.3	86.6	89.8

(a)

Experiment	N_{hU}^2	Overall Accuracy (%)	Kappa	Producer's Accuracy (%)						User's Accuracy (%)					
				building	tree	road	grass	soil	water	building	tree	road	grass	soil	water
Case1-2-1	10	40.7	0.396	39.7	41.6	37.4	41.7	38.5	45.9	37.8	42.7	39.9	45.6	48.4	44.4
Case1-2-2	20	55.9	0.504	53.9	55.2	50.7	56.6	52.5	58.7	50.1	51.2	52.1	58.1	63.6	55.2
Case1-2-3	30	62.2	0.632	62.3	67.4	61.2	69.6	60.9	70.6	61.9	67.1	61.4	72.2	76.8	69.8
Case1-2-4	40	75.7	0.714	74.6	77.6	69.8	78.7	72.3	78.4	72.6	76.6	70.2	80.6	83.3	79.9
Case1-2-5	50	80.6	0.745	79.5	80.7	74.1	82.4	78.4	85.7	77.6	80.4	76.6	85.7	88.2	82.2
Case1-2-6	60	84.3	0.796	83.6	85.6	78.7	86.3	82.7	88.8	80.7	84.7	79.3	88.4	90.4	87.4
Case1-2-7	70	87.1	0.821	87.2	88.7	80.1	90.7	86.6	95.6	84.7	87.3	82.1	94.6	95.1	93.6
Case1-2-8	80	89.3	0.866	88.9	90.5	83.6	94.1	88.7	98.1	85.3	90.2	87.8	97.9	98.6	97.3
Case1-2-9	90	88.4	0.854	87.6	89.8	80.7	90.6	86.8	94.8	84.3	88.7	83.3	95.1	96.9	93.8
Case1-2-10	100	84.6	0.832	82.7	85.4	78.9	88.7	80.9	89.7	80.5	85.3	80.1	91.2	93.4	90.7
Case1-2-11	110	80.5	0.801	78.8	80.3	74.4	82.3	76.7	85.3	75.7	79.8	76.6	85.7	87.6	84.9
Case1-2-12	120	79.1	0.787	76.5	77.9	71.2	80.1	74.4	82.7	72.1	76.8	73.9	82.4	85.5	81.1

(b)

Experiment	N_s	Overall Accuracy (%)	Kappa	Producer's Accuracy (%)						User's Accuracy (%)					
				building	tree	road	grass	soil	water	building	tree	road	grass	soil	water
Case1-3-1	5	68.3	0.735	68.1	78.9	66.5	79.6	75.6	89.7	62.3	77.5	67.9	82.3	85.6	87.4
Case1-3-2	7	78.2	0.813	77.2	85.5	75.6	89.7	82.1	94.3	74.6	84.7	78.8	90.2	90.5	93.1
Case1-3-3	9	89.3	0.866	88.9	90.5	83.6	94.1	88.7	98.1	85.3	90.2	87.8	97.9	98.6	97.3
Case1-3-4	11	81.9	0.844	80.2	86.6	76.4	90.4	83.9	94.6	75.6	85.4	77.4	91.4	91.6	92.4
Case1-3-5	13	77.6	0.805	78.3	80.3	72.2	86.2	77.1	90.1	72.4	77.9	72.7	87.1	88.4	88.7
Case1-3-6	15	71.4	0.756	72.6	77.4	69.7	78.6	70.4	85.7	69.5	70.7	69.8	79.6	83.7	83.6

(c)

(a) N_{hU}^1 , (b) N_{hU}^2 , and (c) N_s .

B. Machine Learning-Based Low-Level Feature Analysis Methods

Multispectral LiDAR data contain spectral features at different wavelengths and geometrical features (such as height and height-derived texture measures). It is well known that different features describe the characteristics of the objects to be classified. Therefore, sufficient features contribute to the improvement of land cover classification accuracies. However, redundant information exists between these extracted features, which aggravates the complexity of the classification process, and increases the possibility of reducing land cover classification accuracies. In this section, to evaluate the capabilities and feasibilities of machine learning-based feature extraction and selection for land cover classification, we, based on the low-level features, described in Table II, designed two groups of experiments: PCA-based feature extraction and RF-based feature selection in Sections IV-B1 and IV-B2, respectively.

1) *PCA-Based Feature Extraction*: In this group of experiments (i.e., Case 2), the widely used PCA algorithm was used to transform the extracted 40 features in Table II to obtain the PCs for land cover classification. The transformed 40 feature components were arranged by contribution rates. Fig. 5(a)–(h) show the first eight components by PCA transformation, respectively.

As shown in Fig. 5, the first six components contain 99.60% of the information. Visual inspection shows that the first six components contain the majority of information, while the seventh and eighth components contain less information. To further confirm the efficiency of the PCA-based feature extraction, we conducted eight experiments on the Teledyne Optech's Titan multispectral LiDAR data (i.e., Cases 2–1 to 2–8). By using

the SVM classifier, different numbers of the PCA feature PCs were tested to compare their land cover classification accuracies. We summarized the overall accuracy and kappa coefficient in different numbers of the feature PCs (see Table IV).

As shown in Table IV, the number of the PCs has a significant impact on the classification results. When the number of the PCs increases, the overall accuracies and kappa coefficients grow accordingly. Although Case 2–1 used the first six components containing nearly 99.60% information (see Fig. 5), its land cover classification accuracy is lowest in all experiments (see Table IV). Probably, the information contained by the first six components might be not necessarily beneficial to this land cover classification task. When these 40 PCs were input into the SVM classifier, we achieved the best classification overall accuracy of 70.2% and kappa coefficient of 0.642, respectively. As seen in Fig. 6, for most land covers (e.g., building, tree, grass, soil, and water), both the user's and producer's accuracies generally increase when the number of the transformed PCs increases, and they reach the highest values when the number of PCs is 40. The reason behind this phenomenon is that the PCs transformed by PCA might lose clear physical meanings of geometrical and spectral features in the original multispectral LiDAR data. Therefore, for the Teledyne Optech's Titan multispectral LiDAR data, the obtained PCs are uncertain and failed to directly reduce redundant features in land cover classification. In addition, PCA is based on the assumption that the embedded subspace of high-dimensional data is linear, which might be robust to deal with because the multispectral LiDAR data contain many nonlinear structures, resulting in the poor classification results.

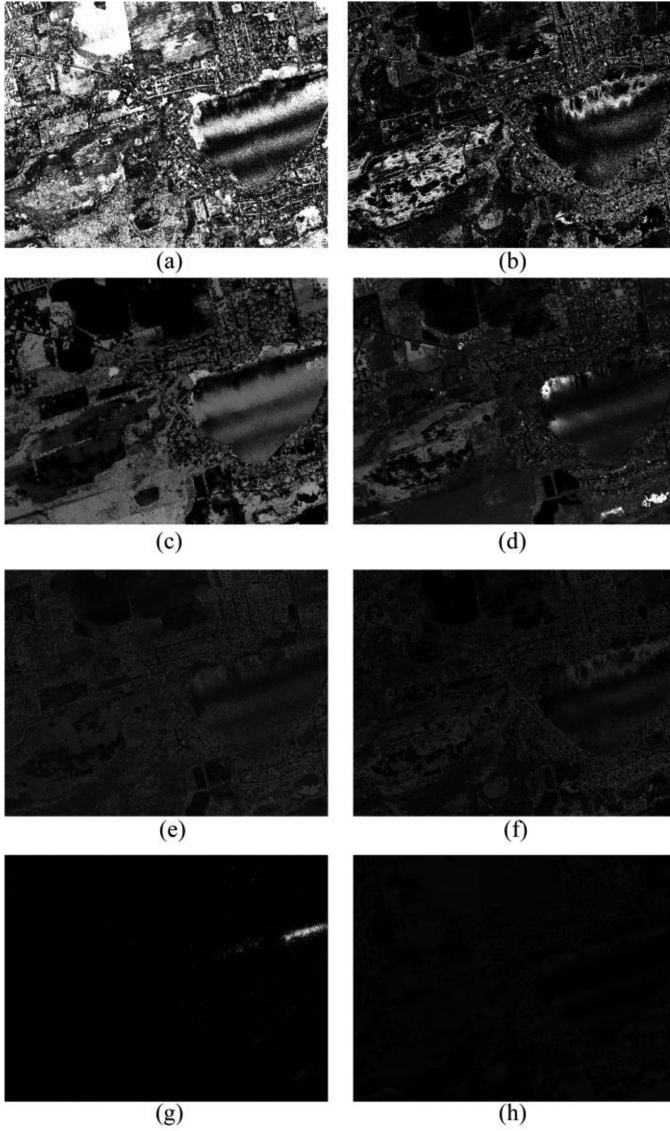


Fig. 5. First eight components by PCA transformation. (a) First component (with contribute rate of 62.32%). (b) Second component (22.70%). (c) Third component (6.68%). (d) Fourth component (3.72%). (e) Fifth component (2.94%). (f) Sixth component (1.24%). (g) Seventh component (0.30%). (h) Eighth component (0.04%).

2) *RF-Based Feature Selection*: As introduced in Section III-B3, when constructing an RF model, there are two important parameters: F_{ntree} and F_{mtry} . Based on the characteristics of multispectral LiDAR data and the number of low-level features listed in Table II, we set $F_{ntree} = 100$ and $F_{mtry} = 6$.

The RF algorithm can provide the importance of the feature variables for the land cover classification model and each land cover. Table V shows the importance of the contribution of the first eight variables to the RF model and demonstrates how those features impact the land cover separability of the RF classification. As seen in Table V, we found that several vegetation indexes are the most important variables for land covers (e.g., tree and grass), as those indexes enable us to divide an image into vegetation and nonvegetation objects. Thus, classification confusion between elevationally identical and spectrally

TABLE IV
CLASSIFICATION ACCURACY OF DIFFERENT PCs

Experiment	No. of PCs	Overall Accuracy (%)	Kappa
Case2-1	6	60.8	0.529
Case2-2	12	62.9	0.555
Case2-3	15	64.0	0.569
Case2-4	20	66.7	0.600
Case2-5	25	67.2	0.607
Case2-6	30	68.7	0.624
Case2-7	35	69.2	0.631
Case2-8	40	70.2	0.642

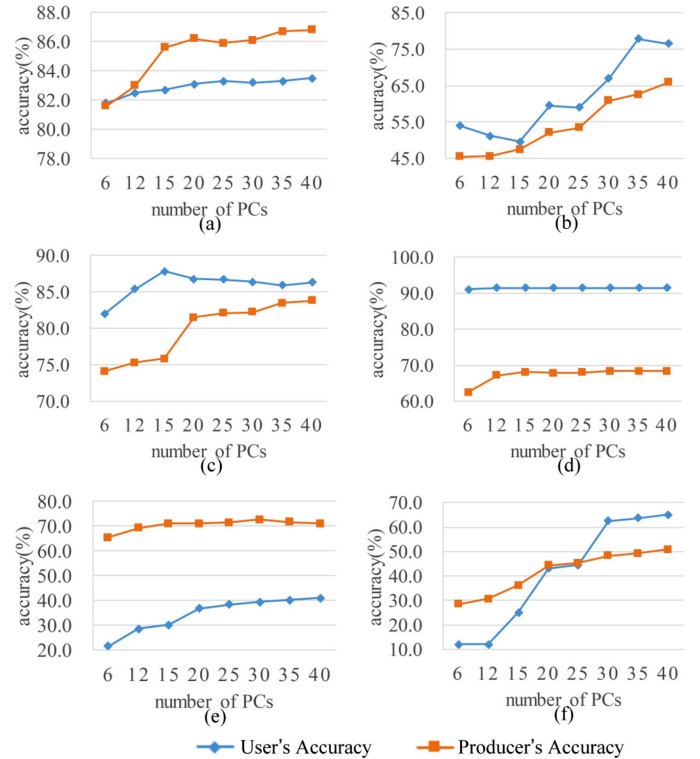


Fig. 6. User's accuracy and producer's accuracy for six land covers on different PCs, (a) building, (b) tree, (c) road, (d) grass, (e) soil, and (f) water.

different objects can be reduced, and classification accuracy is significantly improved. Moreover, elevation is another important variable in the classification of high-rise objects (e.g., building and tree) and low-rise objects (e.g., road, grass and soil). In Table V, variables NIR and MIR also contribute to the classification of soil and vegetation according to object spectral characteristics. Accordingly, to the RF classification model, we arranged these 40 features according to their importance of variable by the calculated *OOB* mean decrease accuracy. As shown in Fig. 7, the first eight feature variables that contribute greatly to the classification results include *GREEN_mean*, *NIR_mean*, *elevation*, *GNDVI*, *GDVI*, *ele_mean*, *MIR*, *ele_corre*. The eight most importance variables correspond to the variables shown in Table V.

To eliminate less important and more correlated features, an iterative backward elimination scheme was used according to the *OOB* mean decrease accuracy. Table VI shows the

TABLE V
VARIABLE IMPORTANCE FOR EACH LAND COVER IN TERMS OF MEAN *OOB* DECREASE, RANKING OF IMPORTANCE FROM TOP TO BOTTOM

building	tree	road	grass	soil	water
<i>NIR_mean</i>	<i>GREEN_mean</i>	<i>GREEN_mean</i>	<i>GDVI</i>	<i>elevation</i>	<i>ele_mean</i>
<i>ele_mean</i>	<i>ele_mean</i>	<i>elevation</i>	<i>NIR_mean</i>	<i>MIR_corre</i>	<i>GREEN_mean</i>
<i>GREEN_mean</i>	<i>elevation</i>	<i>GDVI</i>	<i>elevation</i>	<i>GREEN_mean</i>	<i>elevation</i>
<i>elevation</i>	<i>GNDVI</i>	<i>MIR_corre</i>	<i>Green_mean</i>	<i>NIR_mean</i>	<i>ele_corre</i>
<i>MIR</i>	<i>GREEN_cont</i>	<i>GREEN</i>	<i>MIR_corre</i>	<i>ele_mean</i>	<i>MIR_corre</i>
<i>MIR_corre</i>	<i>GRVI</i>	<i>NIR_mean</i>	<i>GRVI</i>	<i>GDVI</i>	<i>GREEN</i>
<i>ele_corre</i>	<i>GREEN_homo</i>	<i>ele_mean</i>	<i>GNDVI</i>	<i>NIR</i>	<i>ele_sedmom</i>
<i>GREEN_vari</i>	<i>ele_sedmom</i>	<i>GNDVI</i>	<i>NIR</i>	<i>MIR_mean</i>	<i>GREEN_dissimi</i>

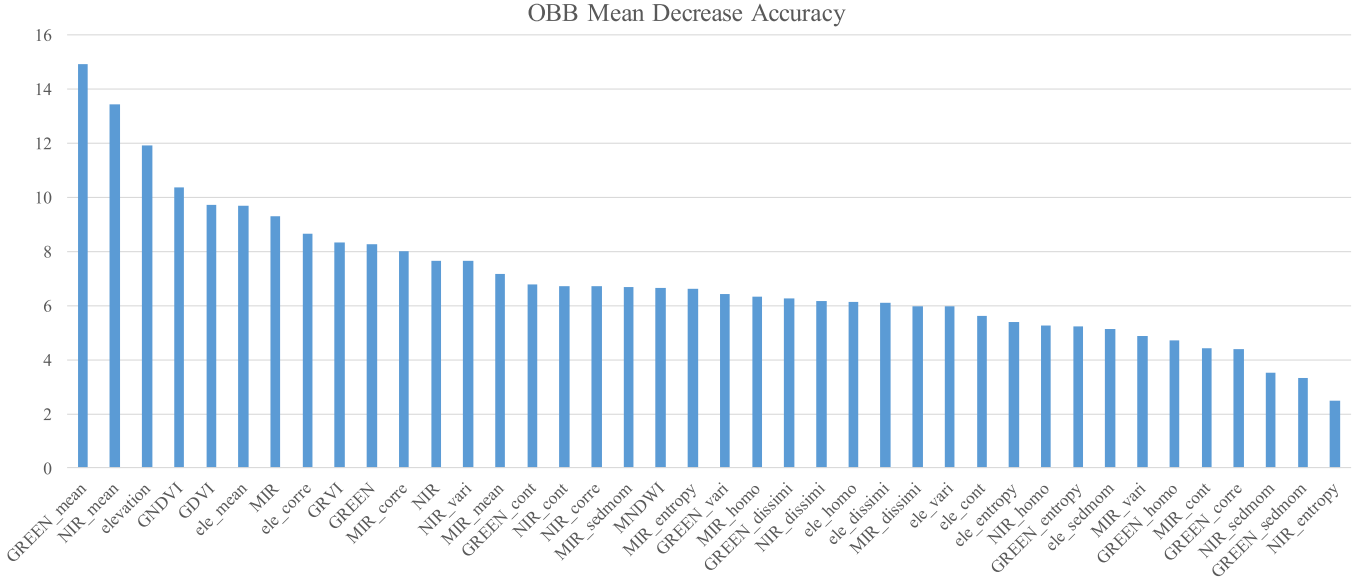


Fig. 7. Variable importance of the classification model demonstrated by mean *OOB* decrease.

TABLE VI
CLASSIFICATION ACCURACIES OF THE RF-BASED
BACKWARD FEATURE SELECTION METHOD

Experiment	No. of back selection features	Overall Accuracy (%)	Kappa
Case3-1	40	77.7	0.732
Case3-2	32	80.4	0.765
Case3-3	26	80.8	0.770
Case3-4	21	80.1	0.761
Case3-5	17	78.7	0.744
Case3-6	14	76.8	0.721

overall accuracies and kappa coefficients of land cover classification by the recursive feature elimination process for the Titian multispectral LiDAR data. When the features are reduced from 40 to 26, we found, in Table VI, that the overall accuracy and kappa coefficient stably increase from 77.7% to 80.8%, and from 0.732 to 0.770, respectively. This is because the backward feature selection method might eliminate some redundant and partially correlated features. However, the overall accuracy and kappa coefficient slowly decrease when the number of feature is below 26, indicating that an excessive elimination of features can be counterproductive. Fig. 8 shows both the user's and producer's accuracies for each land cover. As shown in Fig. 8, for tree and road, when the number of features varied

from 40 to 14, the user's accuracies first greatly increase and then decrease, while the producer's accuracies maintain stable. The reason behind this phenomenon might be that the most important feature variables (see in Table V) include elevation and vegetation information, which contributes to classifying tree and road from other land covers. On the contrary, for grass and soil, the user's accuracies are relatively stable, while the producer's accuracies first grow up and then drop down. According to the analysis of the most important feature variables in Table V, the vegetation indexes are useful for maintaining the correctness (around 93%) and slightly improving the completeness (from 84% to 87%) of grass. For building, there are less changes in user's and producer's accuracies when the number of features varies from 40 to 14. The RF-based backward feature selection method generates less the user's and producer's improvement for water. Overall, the RF-based method improves recognition rates and reliabilities for all land covers.

C. Computational Performance Analysis

To verify the feasibility of deep learning and machine learning algorithms for feature analysis, computational complexity was evaluated on these three learning algorithms by comparing the time required to extract the features, as shown in Table VII. All

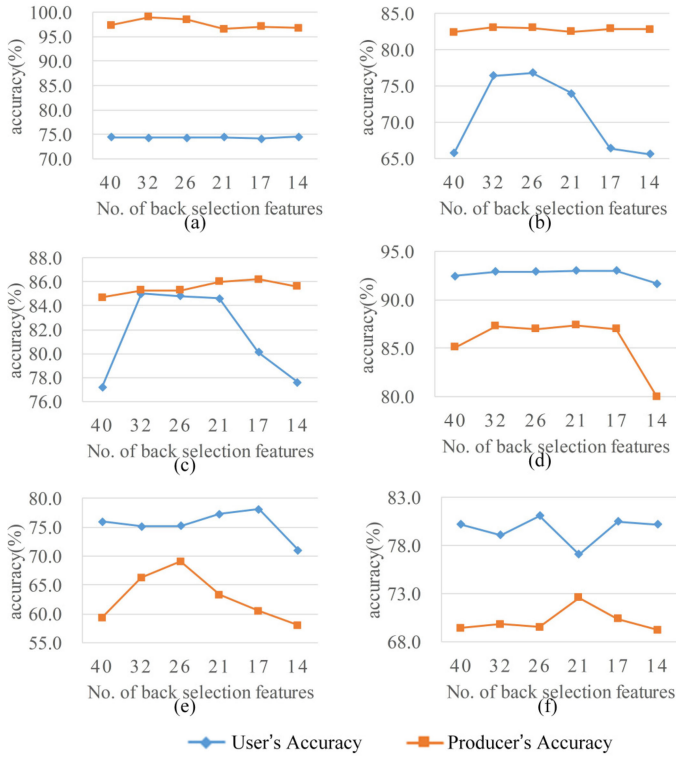


Fig. 8. User's accuracy and producer's accuracy for six land covers, (a) building, (b) tree, (c) road, (d) grass, (e) soil, and (f) water.

TABLE VII
COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN DEEP LEARNING AND MACHINE LEARNING-BASED FEATURE ANALYSIS METHODS

Experiment	Method	Description	Time (min)
Case1-1-3	DBM-based	model with two hidden layers	6.783
Case2-8	PCA-based	model with 40 PCs	5.317
Case3-2	RF-based	model with 32 features (1 st iteration)	2.248
Case3-3		model with 26 features (2 nd iteration)	1.943
Case3-4		model with 21 features (3 rd iteration)	1.701
Case3-5		model with 17 features (4 th iteration)	1.304
Case3-6		model with 14 features (5 th iteration)	1.113

these experiments were conducted on the HP EliteDesk 880 G2 TWR computer with Intel Core TM i5-6500 CPU 3.20 GHz. The DBM-based method requires a large number of training samples to train the model. In this study, 30 000 training samples for each land cover were input to the DBM model. The training process took 34.657 h for creating the DBM model with two hidden layers. Once the DBM model is constructed, it can perform very fast toward feature abstraction. On the given data, the feature extraction process took approximately 6.783 min. For the two machine learning-based feature analysis methods, the PCA-based method took 5.317 min to obtain the transformed features from

the original low-level features listed in Table II; the RF-based method took a total of 8.309 min (2.248, 1.943, 1.701, 1.304, and 1.113 min for each iteration, respectively) by performing five iterations on the original extracted low-level features. As seen in Table VII, with the decrease of the number of features (i.e., increase of iteration numbers), this method required less computational time. Therefore, as shown in the computational analysis results, the PCA-based and DBM-based methods performed significantly faster than the RF-based method.

D. Discussion

Our comparative experiments on land cover classification feature analysis were conducted on the Teledyne Optech's Titan multispectral LiDAR data. To explore the potential of the DBM model, one of the popular deep learning methods, in high-level feature representation, we compared it with two widely used machine learning methods—PCA and RF. All experiments for these three methods (i.e., Case 1 for the DBM-based high-level feature presentation, Case 2 for PCA-based low-level feature extraction, and Case 3 for RF-based low-level feature selection) can be seen in Tables III, IV, and VI.

In the DBM-based high-level feature representation method, the number of the hidden layer units and the neighborhood size had a certain impact on the land cover classification results. At the first hidden layer, the classification accuracy first increased and then decreased with the increase of the parameter, N_{hU}^1 . The change of the classification accuracies was relatively smooth, and the optimal classification accuracy occurred at $N_{hU}^1 = 300$. At the second layer, the land cover classification accuracy fluctuated greatly with the change of N_{hU}^2 . The better performance of land cover classification occurred at the hidden layer with 80 units. Although the two groups of experiments indicated that there was no obvious rules to determine the values of the parameters N_{hU}^1 and N_{hU}^2 , we empirically concluded that the smaller the values of the two parameters, the fewer information the model learned due to insufficient connections between neurons, whereas excessive large values of the two parameters generated the redundant information learned from the training samples. On the contrary, the value of the neighborhood size showed a distinct change pattern on the land cover classification accuracy. The classification accuracy first increased and then decreased when the neighborhood size increased. The overall accuracies appropriately changed from 68.3% to 89.3%. The better performance occurred when $N_s = 9$. It should be noted that an image region with a small neighborhood size could contain less useful contextual or spatial information for sufficiently representing the high-level features, which might lead to the decrease of land cover classification. Similarly, an image region with an excessive neighborhood size could include redundant contextual or spatial information, which also might result in the decrease of land cover classification.

As a widely used machine learning method, PCA gains a reputation on data dimensionality reduction. In the PCA-based feature extraction, all 40 low-level features, including three-wavelength intensities and height-derived information of multispectral LiDAR data, were transformed and generated new,

orthogonal, and uncorrelated components. The first six PCs contained almost 96% image information. However, when the first six PCs were input into the SVM classifier, the overall classification accuracy was 60.8%, the lowest of all experiments in Case 2. When all 40 components were used, the overall accuracy improved nearly 10%. Although PCA has a good projection dimensionality reduction processing for data samples, it also has disadvantages. For one thing, when the original dataset has a nonlinear structure, the low-dimensional intrinsic geometric structure of the data cannot be characterized by the dimension reduction result, and the dimension reduction results lose the physical meaning of the original band. For another thing, the dimensionality reduction process of PCA is realized by coordinate transformation, projecting the high-dimensional dataset to the direction of maximum variance, satisfying the conditions of minimum projection error and maximum variance. However, the best direction of classification is not necessarily the direction with the largest variance. Moreover, the information of class labels is not fully considered, and it is also one of the important reasons for dimension reduction errors.

In RF-based feature selection method, based on the feature importance, one characteristic of RF, an iterative backward feature elimination method was used to eliminate the redundant and correlation features. We performed five iterations, in which 20% of features with the lowest *OOB* errors were eliminated. Accordingly, the land cover classification accuracies were steadily improved when the number of features changed from 40 to 26. This is because the method of iterative feature elimination reduces some redundant and partially correlated features. Afterward, the land cover classification accuracies were decreased, indicating that an excessive elimination of features can be counterproductive. We achieved the best classification accuracies, with overall accuracy of 80.8% and kappa coefficient of 0.770, when taking 26 features into account.

Quantitatively, the DBM-based high-level feature representation method was superior to the RF-based feature selection method and PCA-based feature extraction method in this study. For the DBM-based method, the overall accuracy and kappa coefficient were approximately 8.5% and 0.096 higher than the RF-based method, and about 19.1% and 0.224 higher than the PCA-based method. Moreover, the best classification accuracies using the low-level features transformed by PCA were considerably lower than those of the RF. Specifically, the DBM model took much more time at the model training stage. The GPU can be used to accelerate the training process. However, once the model is constructed, it can perform very fast toward feature extraction. The experimental results demonstrated that, due to the capability of representing high-level features, the DBM model performed very effectively and obtained superior performance. Therefore, the DBM model was appropriate for land cover classification using multispectral LiDAR data.

Moreover, the computational complexity analysis showed that the PCA-based feature extraction method took 5.317 min to transform the 40 low-level features into the PCs, which was faster than the other two methods. The RF-based feature selection method took a total of 8.309 min for five iterations to select 14 low-level features, which was slower than the DBM-

based method. From the classification accuracies, we found 2 or 3 iterations generated the best performance in our land cover classification task, indicating that around 2–3 iterations were required for feature selection. However, the DBM model required considerably much more time at the training stage on an ordinary personal computer or workstation. Comparatively, our two machine learning-based methods—PCA and RF—were less computationally complex even under an ordinary personal computer or workstation.

V. CONCLUSION

As a new LiDAR technology, a multispectral LiDAR system provides both geometrical and multiwavelength information, which contributes to identify different land covers. In land cover classification tasks, one of the major issues is to obtain optimal multiple features, extracted from multispectral LiDAR data, and thus achieve stable classification accuracies. The study discusses and compares the deep learning and machine learning-based feature analysis methods for multispectral LiDAR data in land cover classification. The proposed scheme is tested on the Teledyne Optech's Titan multispectral LiDAR data. The study area is classified into six land covers: building, tree, road, grass, soil, and water. The major contributions of this study are to:

- 1) explore a popular deep learning method—DBM—for high-level feature representation;
- 2) investigate two widely used machine learning methods—PCA and RF—for low-level feature extraction and selection;
- 3) compare the advantages and disadvantages of these learning algorithms in land cover classification using multispectral LiDAR data.

Deep learning methods have attracted much attention in recent years because of their superior performance in learning hierarchical features from high-dimensional data. In the study, the DBM model can obtain high-level feature representation and show superior performance in land cover classification using multispectral LiDAR data. Moreover, compared to the low-level features extracted and selected by the PCA and RF methods, the DBM-based method outperforms them and shows acceptable time complexity in land cover classification.

In the future, we will explore different deep learning models in high-level feature representation for land cover classification using multispectral LiDAR data. Especially, we will focus on object-based land cover classification based on the deep learning models for multispectral LiDAR data. Moreover, we may use GPU equipment to carry out the experiment to shorten the experiment time and improve the experiment efficiency.

ACKNOWLEDGMENT

The authors would like to acknowledge Teledyne Optech for providing the datasets acquired by a Titan multi-spectral airborne LiDAR system to support this study. The authors would also like to acknowledge the anonymous reviewers for their valuable comments.

REFERENCES

- [1] Z. Kang, J. Yang, and R. Zhong, "A Bayesian-network-based classification method integrating airborne LiDAR data with optical images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1651–1661, Apr. 2017.
- [2] A. Wang, S. Zhao, H. K. Zhou, Y. X. Luo, and L. Tan, "Object-based classification using LiDAR-derived metrics and QuickBird imagery," in *Proc. Int. Workshop. Earth Observ. Remote Sens. Appl.*, 2012, pp. 181–185.
- [3] V. Saeidi, B. Pradhan, M. O. Idrees, and Z. A. Latif, "Fusion of airborne LiDAR with multispectral SPOT 5 image for enhancement of feature extraction using Dempster–Shafer theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6017–6025, Oct. 2014.
- [4] K. Kandare, M. Dalponte, H. O. Ørka, L. Frizzera, and E. Næsset, "Prediction of species-specific volume using different inventory approaches by fusing airborne laser scanning and hyperspectral data," *Remote Sens.*, vol. 9, no. 5, pp. 400–400, Apr. 2017.
- [5] J. Vauhkonen *et al.*, "Classification of spruce and pine trees using active hyperspectral LiDAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1138–1141, Sep. 2013.
- [6] E. Puttonen *et al.*, "Artificial target detection with a hyperspectral LiDAR over 26-h measurement," *Opt. Eng.*, vol. 54, no. 1, pp. 013105–013120, Jan. 2015.
- [7] H. W. Leigh and L. A. Magruder, "Using dual-wavelength, full-waveform airborne lidar for surface classification and vegetation characterization," *J. Appl. Remote Sens.*, vol. 10, no. 4, pp. 045001–045020, Oct. 2016.
- [8] L. Matikainen, J. Hyypää, and P. Litkey, "Multispectral airborne laser scanning for automated map updating," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 323–330, Jul. 2016.
- [9] L. Matikainen, K. Karila, J. Hyypää, P. Litkey, E. Puttonen, and E. Ahokas, "Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 298–313, 2017.
- [10] V. Wichmann, M. Bremer, J. Lindenberger, M. Rutzinger, C. Georges, and F. Petroni-Monteferrì, "Evaluating the potential of multispectral airborne LiDAR for topographic mapping and land cover classification," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. II-3/W5, pp. 113–119, 2015.
- [11] S. Morsy, A. Shaker, and A. El-Rabbany, "Multispectral LiDAR data for land cover classification of urban areas," *Sensors*, vol. 17, no. 5, pp. 958–979, Apr. 2017.
- [12] Z. Kang, L. Zhang, B. Wang, Z. Li, and F. Jia, "An optimized BaySAC algorithm for efficient fitting of primitives in point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1096–1100, Jun. 2014.
- [13] J. C. Fernandez-Diaz *et al.*, "Capability assessment and performance metrics for the titan multispectral mapping lidar," *Remote Sens.*, vol. 8, no. 11, pp. 936–970, Nov. 2016.
- [14] C. Hopkinson, L. Chasmer, C. Gyman, C. Mahoney, and M. Sitar, "Multisensor and multispectral LiDAR characterization and classification of a forest environment," *Can. J. Remote Sens.*, vol. 42, pp. 501–520, Jun. 2016.
- [15] T. A. Teo and H. M. Wu, "Analysis of land cover classification using multi-wavelength LiDAR system," *Appl. Sci.*, vol. 7, no. 7, pp. 663–683, Jun. 2017.
- [16] X. Chen, Y. E. Chengming, J. Li, and M. Chapman, "Quantifying the carbon storage in urban trees using multispectral ALS data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3358–3365, Sep. 2018.
- [17] C. I. Miller, J. J. Thomas, A. M. Kim, J. P. Metcalf, and R. C. Olsen, "Application of image classification techniques to multispectral lidar point cloud data," *Proc. SPIE*, vol. 9832, May 2016, Art. no. 98320X.
- [18] Z. Kang and J. Yang, "A probabilistic graphical model for the classification of mobile LiDAR point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 143, pp. 108–123, May 2018.
- [19] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 3, pp. 131–156, Mar. 1997.
- [20] V. Kumar, P. Agrawal, and S. Agrawal, "ALOS PALSAR and hyperion data fusion for land use land cover feature extraction," *J. Indian Soc. Remote Sens.*, vol. 45, no. 3, pp. 407–416, Jun. 2017.
- [21] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Oct. 2016.
- [22] W. Dong, J. Lan, S. Liang, W. Yao, and Z. Zhan, "Selection of LiDAR geometric features with adaptive neighborhood size for urban land cover classification," *Int. J. Appl. Earth Observ. Geoinformat.*, vol. 60, pp. 99–110, May 2017.
- [23] L. Guo, N. Chehata, C. Mallet, and S. Boukir, "Relevance of airborne lidar and multispectral image data for urban scene classification using random forests," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, pp. 56–66, Sep. 2011.
- [24] R. Huang and J. Zhu, "Using random forests to integrate LiDAR data and hyperspectral imagery for land cover classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 3978–3981.
- [25] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [26] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba, "Learning with hierarchical-deep models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1958–1971, Aug. 2013.
- [27] D. Li, Y. Yu, H. Guan, and L. Zhong, "Aeroplane detection in very-high-resolution images using deep feature representation and rotation-invariant Hough forests," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6882–6893, Aug. 2017.
- [28] J. Li and A. D. Heap, "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecol. Informat.*, vol. 6, nos. 3/4, pp. 228–241, Jul. 2011.
- [29] R. Salakhutdinov and G. E. Hinton, "Deep boltzmann machines," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2009, vol. 12, pp. 448–455.
- [30] R. Salakhutdinov and G. E. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Comput.*, vol. 24, no. 8, pp. 1967–2006, 2012.
- [31] T. Tieleman, "Training restricted Boltzmann machines using approximations to the likelihood gradient," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 1064–1071.
- [32] R. P. Sripada, R. W. Heiniger, J. G. White, and A. D. Meijer, "Aerial color infrared photography for determining early in-season nitrogen requirements in corn," *Agron J.*, vol. 97, no. 5, pp. 1443–1451, Sep./Oct. 2005.
- [33] A. A. Gitelson and M. N. Merzlyak, "Remote Sensing of Chlorophyll Concentration in Higher Plant Leaves," *Adv. Space Res.*, vol. 22, no. 5, pp. 689–692, 1998.
- [34] H. Xu, "A study on information extraction of water body with the modified normalized difference water index (MNDWI)," *Int. J. Remote Sens.*, vol. 9, no. 5, pp. 589–595, Sep. 2005.
- [35] H. Guan, J. Li, M. Chapman, F. Deng, Z. Ji, and X. Yang, "Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests," *Int. J. Remote Sens.*, vol. 34, no. 14, pp. 5166–5186, Apr. 2013.
- [36] H. Guan, Z. Ji, L. Zhong, J. Li, and Q. Ren, "Partially supervised hierarchical classification for urban features from lidar data with aerial imagery," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 190–210, Jan. 2013.
- [37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, pp. 294–300, Oct. 2005.
- [39] N. Horning, "Random forests: An algorithm for image classification and generation of continuous fields data sets," in *Proc. Int. Conf. Geoinformat. Spatial Infrastructure Develop. Earth Allied Sci.*, 2010, pp. 1–6.
- [40] R. Díaz-Uriarte and S. A. De Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinform.*, vol. 7, no. 1, pp. 3–15, Jan. 2006.



Suoyan Pan received the bachelor's degree in remote sensing science and technology from the Nanjing University of Information Science & Technology, Nanjing, China, in 2017. She is currently working toward the master's degree with the School of Geographic Science, Nanjing University of Information Science & Technology.

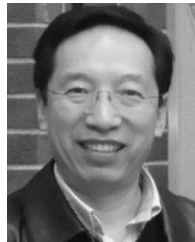
Her current research interests include information extraction from multispectral LiDAR point clouds and deep learning.



Haiyan Guan (M'15–SM'18) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009, and in geomatics from the University of Waterloo, Waterloo, ON, Canada, in 2014.

She is currently a Professor with the School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, China. She has published more than 40 research papers in refereed journals, books, and proceedings. Her current research interests include

information extraction from LiDAR point clouds and from earth observation images.



Jonathan Li (M'00–SM'11) received the Ph.D. degree in remote sensing and GIS from the University of Cape Town, Cape Town, South Africa, in 2000.

He is a Professor of Geomatics and Head of the Geospatial Technology and Remote Sensing (GeoSTARS) group in the Department of Geography & Environmental Management, University of Waterloo, Waterloo, ON, Canada. He is also an Adjunct Professor of York University in Canada and Guest Professor of several top-ranked universities in China.

His current research is focused on use of mobile laser scanning point clouds for 3-D surface modeling. He has published extensively in leading remote sensing journals.

Dr. Li is the recipient of several prestigious awards. He was the Chair of ISPRS ICWG I/Va on Mobile Scanning and Imaging Systems (2012–2016), the Vice Chair of ICA Commission on Mapping from Remote Sensor Imagery (2011–2015), the Vice Chair of FIG Commission IV on Hydrography as well as the Chair of FIG WG4.4 on Maritime and Marine Spatial Information Management (2014–2018). He has been the Remote Sensing Editor of *GEOMATICA* since 2007.

He has been the Remote Sensing Editor of *GEOMATICA* since 2007.



Yongtao Yu (M'16) received the Ph.D. degree in computer science and technology from Xiamen University, Xiamen, China, in 2015.

He is currently an Assistant Professor with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, China. He has coauthored more than 30 research papers in refereed journals, books, and proceedings. His research interests include pattern recognition, machine learning, intelligent interpretation of 3-D point clouds, and remotely sensed imagery.



Daifeng Peng received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017.

He is currently a Lecturer with the School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, China. His research interests include pattern recognition, machine learning, and intelligent interpretation and information extraction from high-resolution remote-sensing imagery.